CERN 96-03
May 1996

DE009922269
XC97FF998

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

CAS CERN ACCELERATOR SCHOOL

SUPERCONDUCTIVITY IN PARTICLE ACCELERATORS

Haus Rissen, Hamburg, Germany
17–24 May 1995

PROCEEDINGS
Editor: S. Turner

GENEVA
1996

VOL 28 № 24

CERN–Service d'information scientifique–RD/964–2500–mai 1996

© Copyright CERN, Genève, 1996

Propriété littéraire et scientifique réservée pour tous les pays du monde. Ce document ne peut être reproduit ou traduit en tout ou en partie sans l'autorisation écrite du Directeur général du CERN, titulaire du droit d'auteur. Dans les cas appropriés, et s'il s'agit d'utiliser le document à des fins non commerciales, cette autorisation sera volontiers accordée.

Le CERN ne revendique pas la propriété des inventions brevetables et dessins ou modèles susceptibles de dépôt qui pourraient être décrits dans le présent document; ceux-ci peuvent être librement utilisés par les instituts de recherche, les industriels et autres intéressés. Cependant, le CERN se réserve le droit de s'opposer à toute revendication qu'un usager pourrait faire de la propriété scientifique ou industrielle de toute invention et tout dessin ou modèle décrits dans le présent document.

Literary and scientific copyrights reserved in all countries of the world. This report, or any part of it, may not be reprinted or translated without written permission of the copyright holder, the Director-General of CERN. However, permission will be freely granted for appropriate non-commercial use.

If any patentable invention or registrable design is described in the report, CERN makes no claim to property rights in it but offers it for the free use of research institutions, manufacturers and others. CERN, however, may oppose any attempt by a user to claim any proprietary or patent rights in such inventions or designs as may be described in the present document.

ISSN 0007-8328
ISBN 92-9083-084-0

XC97FF998

DE009922269

ABSTRACT

These proceedings present the lectures given at the ninth specialized course organized by the CERN Accelerator School (CAS), the topic this time being 'Superconductivity in Particle Accelerators'. This course is basically a repeat of that given at the same location in 1988 whose proceedings were published as CERN 89–04. However, the opportunity was taken to improve the presentation of the various topics and to introduce the latest developments in this rapidly expanding field. First the basic theory of superconductivity is introduced. A review of the materials used for sc magnets is followed by magnet design requirements, the influence of eddy and persistent currents, and the methods used to provide quench protection. Next follows the basic theory of sc cavities, their materials, high-gradient limitations, the problem of field emission and then their power couplers. After an introduction to cryogenics and cryoplants, the theory of superfluidity is presented followed by a review of the use of superfluid helium. Finally, two seminars detail the impact of superconductors on the design of the LHC and LEP2 accelerators.

CERN ACCELERATOR SCHOOL

DEUTSCHES ELEKTRONEN SYNCHROTRON

will jointly organise a course on

SUPERCONDUCTIVITY IN PARTICLE ACCELERATORS

17–24 May 1995
at Haus Rissen, Hamburg

Photo: Tourismus-Zentrale Hamburg GmbH

Lectures:
Superconductivity
Ginsburg-Landau theory
Fields, forces and mechanics of SC magnets
Survey of SC materials
Practical superconductors for magnets
Persistent and eddy currents and their impact on accelerator performance
Quench protection
Cryogenics
Superfluidity
1.8K techniques
Basic properties of cavities
Materials for cavities
Quest for high RF fields
Field emission in cavities
Couplers for cavities
High-temperature superconductors

Seminars:
Impact of superconductors on LHC design
Operating HERA with SC magnets
SC detector magnets
Impact of SC cavities on LEP 2 design
Recirculating linacs

Visit:
DESY-HERA

For further information and application forms please contact:

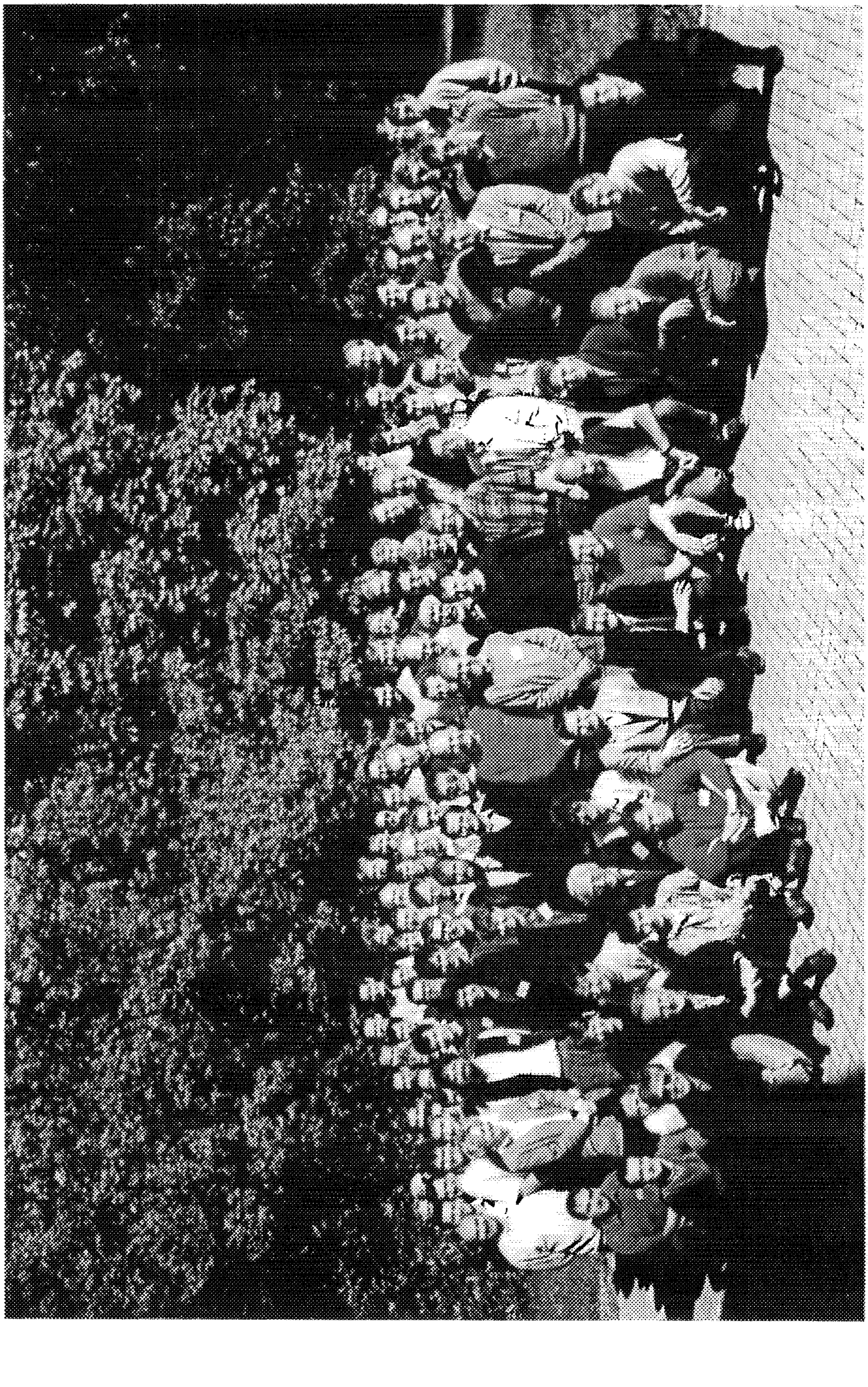
Mrs. S. von Wartburg,
CERN Accelerator School,
AC Division,
CH-1211 Geneva 23

E-Mail: CASDESY @ CERNVM.CERN.CH
Fax: + 41 22 7824836 (until 15 Jan 1995)
+ 41 22 7675460 (from 16 Jan 1995)

CLOSING DATE FOR APPLICATIONS: 1 MARCH 1995

PROGRAMME FOR THE CERN ACCELERATOR SCHOOL COURSE on *Superconductivity in Particle Accelerators*
Hamburg, 17 to 24 May 1995.

Time	**Wednesday** 17 May	**Thursday** 18 May	**Friday** 19 May	**Saturday** 20 May	**Sunday** 21 May	**Monday** 22 May	**Tuesday** 23 May	**Wednesday** 24 May
	B R E A K F A S T							
09.00 / 10.00	Superconductivity I P. Schmüser	Superconductivity II P. Schmüser	Superconductivity III P. Schmüser	Ginsburg-Landau theory M. Cyrot		Superfluidity W.F. Vinen	1.8 K P. Lebrun	High-temperature superconductors G. Müller
	C O F F E E					C O F F E E		
10.20 / 11.20	Fields, forces and mechanics of SC magnets I R. Perin	Fields, forces and mechanics of SC magnets II R. Perin	Survey of SC materials M. Wilson	Practical superconductors for magnets M. Wilson	E X	Persistent and eddy currents A. Devred	Impact of persistent and eddy currents on accelerator performances B. Holzer	Couplers for cavities II E. Haebel
	MID MORNING BREAK				C	MID-MORNING BREAK		
11.30 / 12.30	Cryogenics I J. Schmid	Cryogenics II J. Schmid	Basic properties of SC cavities II W. Weingarten	Quest for high r.f. field D. Proch	U R	Field emission for cavities B. Bonin	Couplers for cavities I E. Haebel	TESLA B. Wiik
	L U N C H					L U N C H		
14.00 / 15.00	Basic properties of SC cavities I W. Weingarten	F R	Poster session E.J.N. Wilson	Operating HERA with superconducting magnets F. Willeke	S I	Materials for cavities B. Bonin	Quench protection K.-H. Mess	TRANSPORT TO AIRPORT
	TEA		TEA			TEA		
15.30 / 17.00	Impact of superconductors on LHC design J. Gareyte	E E	SC detector magnets A. Daël	DESY VISIT + Buffet supper	O N	Impact of superconducting cavities on LEP2 design D. Boussard	Recirculating linacs D. Gräf	
17.30	COCKTAIL						BOAT TRIP	
18.30	DINNER					DINNER	GALA DINNER	



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 specialized courses listed in the Table below. The latest of these was on the topic of Superconductivity in Particle Accelerators' and was held at Haus Rissen, Hamburg, Germany, 17–24 May 1995, its proceedings forming the present volume.

List of specialized CAS courses and their proceedings

Year	Course	Proceedings
1983	Antiprotons for colliding beam facilities	CERN 84-15 (1984)
1986	Applied Geodesy for particle accelerators	CERN 87-01 (1987) also Lecture Notes in Earth Sciences 12, (Springer Verlag, 1987)
1988	Superconductivity in particle accelerators	CERN 89-04 (1989)
1989	Synchrotron radiation and free-electron lasers	CERN 90-03 (1990)
1990	Power converters for particle accelerators	CERN 90-07 (1990)
1991	RF engineering for particle accelerators	CERN 92-03 (1992)
1992	Magnetic measurement and alignment	CERN 92-05 (1992)
1993	RF engineering for particle accelerators (repeat of the 1991 course)	-
1994	Cyclotrons, linacs and their applications	CERN 96-02 (1996)
1995	Superconductivity in particle accelerators	Present volume

Over the last few years, the importance of superconductivity in the construction of particle accelerators and their detection systems, and in medical imaging and other measuring devices has grown tremendously. Superconductivity has made it possible to reduce the cost, size and power requirements of magnets, while permitting higher-energy accelerators to be constructed within the physical constraints of existing sites, typical examples being the TEVATRON at Fermilab and HERA at DESY. Machines constructed more recently such as RHIC at Brookhaven or planned for the near future, particularly the LHC at CERN, would hardly have been feasible without the use of superconductivity for both their magnets and their RF accelerating cavities. While the accelerator applications have been the prime promoter of recent superconductivity technology, the magnet size and cost reductions have made it possible for many hospitals to install imaging devices to the extent that they are presently the major user of superconducting materials. Meanwhile, serious development of the recently discovered high-temperature superconductors promises very exciting possibilities for their use in the future. With such a vigorous programme of superconducting applications and developments there was every reason not only to repeat the CAS course given on this topic in 1988 but also to publish these new proceedings.

Organization of this course required the good will and considerable effort of many people. In particular, CAS thanks the DESY and CERN Directorates for their financial and moral support, the CAS Advisory Committee for its guidance, and the Programme and Local Organizing Committees for the great attention to detail which ensured that the course was so successful. Very special thanks must go to the lecturers at the course for their enthusiasm and very considerable work in preparing, presenting and writing up their topics. The effort made by the Haus Rissen staff during our stay with them was most appreciated. Finally, we thank the participants at the course for their support and encouragement.

S. Turner
Editor

CONTENTS

Superconductivity

P. Schmüser
II. Institut für Experimentalphysik, Universität Hamburg, Germany

Abstract
Low-temperature superconductivity is treated at an introductory level. The topics include Meissner-Ochsenfeld effect and London equations, thermodynamic properties of the superconducting state, type I and II superconductors, flux quantisation and superconducting quantum interference effects. Important experiments are discussed. The basic ideas of the BCS theory and its implications are outlined.

1 INTRODUCTION

In these lectures I want to give a brief introduction into the physical principles of superconductivity and its fascinating properties. More detailed accounts can be found in the excellent text books by W. Buckel [1] and by D.R. Tilley and J. Tilley [2].
Superconductivity was discovered [3] in 1911 by the Dutch physicist H. Kamerlingh Onnes, only three years after he had succeeded in liquefying helium. During his investigations on the conductivity of metals at low temperature he found that the resistance of a mercury sample dropped to an unmeasurably small value just at the boiling temperature of liquid helium. The original measurement is shown in Fig. 1. Kamerlingh Onnes called this totally unexpected phenomenon 'superconductivity' and this name has been retained since. The temperature at which the transition took place was called the *critical temperature* T_c. Superconductivity is observed in a large variety of materials but, remarkably, not in some of the best normal conductors like copper, silver and gold, except at very high pressures. This is illustrated in Fig. 2 where the resistivity of copper, tin and the 'high-temperature' superconductor $YBa_2Cu_3O_7$ is sketched as a function of temperature. Table 1 lists some important superconductors together with their critical temperatures at vanishing magnetic field.

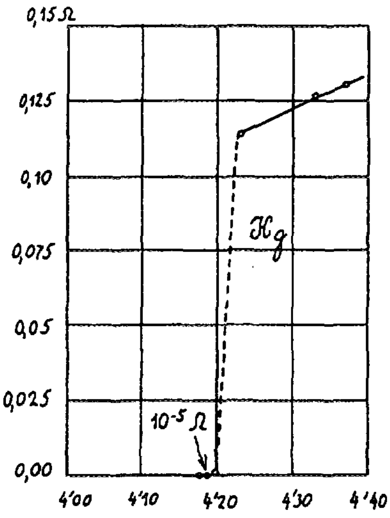


Figure 1: The first measurement of superconductivity by Kamerlingh Onnes.

Kamerlingh Onnes discovered that a conventional resistance measurement is far too insensitive to establish infinite conductivity and that a much better method consists in inducing

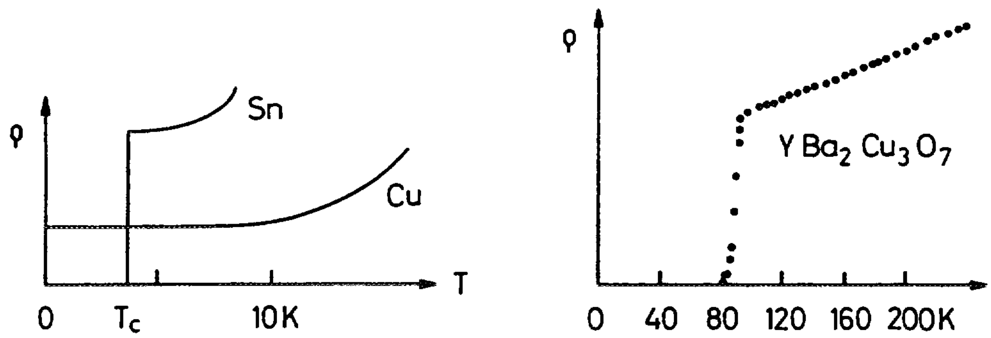


Figure 2: The low-temperature resistivity of copper, tin and $YBa_2Cu_3O_7$.

Table 1: The critical temperature of some common materials at vanishing magnetic field.

material	Ti	Al	Hg	Sn	Pb	Nb	NbTi	Nb_3Sn
T_c[K]	0.4	1.14	4.15	3.72	7.9	9.2	9.2	18

a current in a ring and determining the decay rate of the produced magnetic field. A schematic experimental setup is shown in Fig. 3. A bar magnet is inserted in the still normal-conducting ring and removed after cooldown below T_c. The induced current should decay exponentially

$$I(t) = I(0)\exp(-t/\tau)$$

with the time constant given by the ratio of inductivity and resistance, $\tau = L/R$, which for a normal metal ring is in the order of 100 μs. In superconducting rings, however, time constants of up to 10^5 years have been observed [4] so the resistance must be at least 15 orders of magnitude below that of copper and is indeed indistinguishable from zero. An important practical application of this method is the operation of solenoid coils for magnetic resonance imaging in the short-circuit mode which exhibit an extremely slow decay of the field of typically $3 \cdot 10^{-9}$ per hour [5].

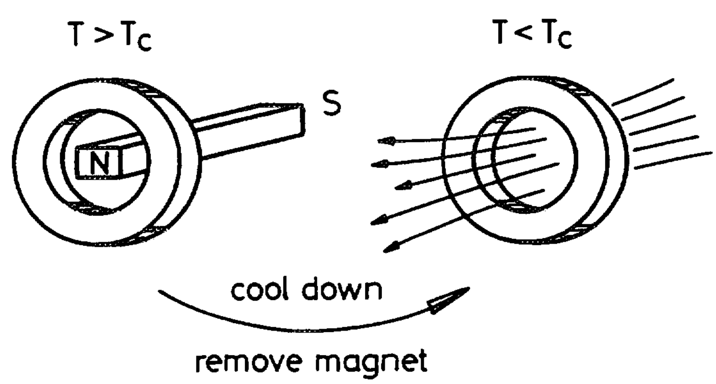


Figure 3: Induction of a persistent current in a superconducting ring.

There is an intimate relation between superconductivity and magnetic fields. W. Meissner and R. Ochsenfeld [6] discovered in 1933 that a superconducting element like lead completely expelled a weak magnetic field from its interior when cooled below T_c while in strong fields superconductivity broke down and the material went to the normal state. The spontaneous exclusion of magnetic fields upon crossing T_c could not be explained in terms of the Maxwell equations and indeed turned out to be a non-classical phenomenon. Two years later, H. and F. London [7] proposed an equation which offered a phenomenological explanation of the Meissner-Ochsenfeld effect but the justification of the London equation remained obscure until the advent of the Bardeen, Cooper and Schrieffer theory [8] of superconductivity in 1957. The BCS theory revolutionized our understanding of this fascinating phenomenon. It is based on the assumption that the supercurrent is not carried by single electrons but rather by pairs of electrons of opposite momenta and spins, the so-called *Cooper pairs*. All pairs occupy a single quantum state, the BCS ground state, whose energy is separated from the single-electron states by an energy gap which in turn can be related to the critical temperature. The BCS theory has turned out to be of enormous predictive power and many of its predictions and implications like the temperature dependence of the energy gap and its relation to the critical temperature, the quantisation of magnetic flux and the existence of quantum interference phenomena have been confirmed by experiment and, in many cases, even found practical application.

A discovery of enormous practical consequences was the finding that there exist two types of superconductors with rather different response to magnetic fields. The elements lead, mercury, tin, aluminium and others are called 'type I' superconductors. They do not admit a magnetic field in the bulk material and are in the superconducting state provided the applied field is below a *critical field* B_c which is a function of temperature. All superconducting alloys like lead-indium, niobium-titanium, niobium-tin and also the element niobium belong to the large class of 'type II' superconductors. They are characterized by two critical fields, B_{c1} and B_{c2}. Below B_{c1} these substances are in the *Meissner phase* with complete field expulsion while in the range $B_{c1} < B < B_{c2}$ a type II superconductor enters the *mixed phase* in which magnetic field can penetrate the bulk material in the form of flux tubes. The Ginzburg-Landau theory [9] provides a theoretical basis for the distinction between the two types. Around 1960 Gorkov [10] showed that the phenomenological Ginzburg-Landau theory is a limiting case of the BCS theory. Abrikosov [11] predicted that the flux tubes in a type II superconductor arrange themselves in a triangular pattern which was confirmed in a beautiful experiment by Essmann and Träuble [12]. In 1962 Josephson [13] studied the quantum theoretical tunnel effect in a system of two superconductors separated by a thin insulating layer and he predicted peculiar and fascinating properties of such a *Josephson junction* which were all confirmed by experiment and opened the way to superconducting quantum interference devices (SQUID's) with extreme sensitivity to tiny magnetic fields.

2 MEISSNER-OCHSENFELD EFFECT AND LONDON EQUATIONS

We consider a cylinder with perfect conductivity and raise a magnetic field from zero to a finite value B. A surface current is induced whose magnetic field, according to Lenz's law, exactly cancels the applied field in the interior. Since the resistance is assumed to vanish the current will continue to flow with constant strength as long as the external field is kept constant and consequently the bulk of the cylinder will stay field-free. This is exactly what happens if we expose a lead cylinder in the superconducting state ($T < T_c$) to an increasing field, see the path $(a) \rightarrow (c)$ in Fig. 4. So below T_c lead acts as a perfect diamagnetic material. There is, however, another path leading to the point (c). We start with a lead cylinder in the normal state ($T > T_c$) and expose it to a field which is increased from zero to B. Eddy currents are induced in this case as well but they decay rapidly and after a few hundred microseconds the field lines will fully penetrate the material (state (b) in Fig. 4). Now the cylinder is cooled down.

At the very instant the temperature drops below T_c , a surface current is spontaneously created and the magnetic field is expelled from the interior of the cylinder. This surprising observation is called the *Meissner-Ochsenfeld effect* after its discoverers; it cannot be explained by the law of induction because the magnetic field is kept constant.

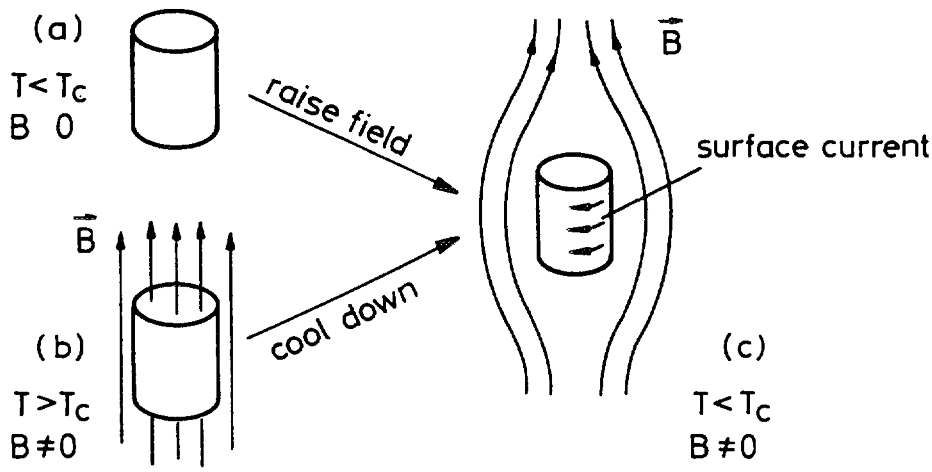


Figure 4: A lead cylinder in a magnetic field. Two possible ways to reach the superconducting final state with $B > 0$ are sketched.

In a (T, B) plane, the superconducting phase is separated from the normal phase by the curve $B_c(T)$ as sketched in Fig. 5. Also indicated are the two ways on which one can reach the point (c). It is instructive to compare this with the response of a 'normal' metal of perfect conductivity. The field increase along the path $(a) \rightarrow (c)$ would yield the same result as for the superconductor, however the cooldown along the path $(b) \rightarrow (c)$ would have no effect at all. So superconductivity means definitely more than just vanishing resistance.

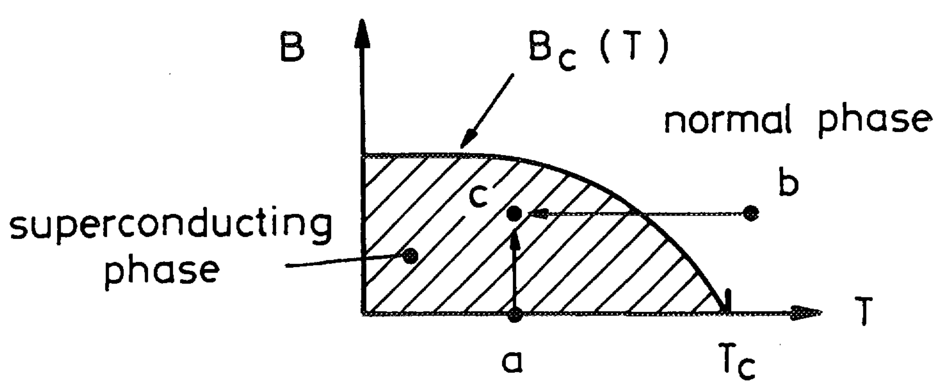


Figure 5: The phase diagram in a (T, B) plane.

I have already used the terms 'superconducting phase' and 'normal phase' to characterize the two states of lead. These are indeed phases in the thermodynamical sense, comparable to the different phases of H_2O which is in the solid, liquid or gaseous state depending on the values of the parameters temperature and pressure. Here the relevant parameters are temperature and magnetic field (for some materials also pressure). If the point (T, B) lies below the curve $B_c(T)$ the material is superconducting and expels the magnetic field, irrespective of by which path the point was reached. If (T, B) is above the curve the material is normal-conducting.

The first successful explanation of the Meissner-Ochsenfeld effect was achieved in 1935 by Heinz and Fritz London. They assumed that the supercurrent is carried by a fraction of the conduction electrons in the metal. The 'super-electrons' experience no friction, so their equation of motion in an electric field is

$$m\frac{\partial \vec{v}}{\partial t} = -e\vec{E} .$$

This leads to an accelerated motion. The supercurrent density is

$$\vec{J}_s = -en_s\vec{v}$$

where n_s is the density of the super-electrons. This immediately yields the first London equation

$$\frac{\partial \vec{J}_s}{\partial t} = \frac{n_s e^2}{m}\vec{E} . \tag{1}$$

Now one uses the Maxwell equation

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$$

and takes the curl (rotation) of (1) to obtain

$$\frac{\partial}{\partial t}\left(\frac{m}{n_s e^2}\vec{\nabla} \times \vec{J}_s + \vec{B}\right) = 0 .$$

Since the time derivative vanishes the quantity in the brackets must be a constant. Up to this point the derivation is fully compatible with classical electromagnetism, applied to the frictionless acceleration of electrons. An example might be the motion of electrons in the vacuum of a television tube or in a circular accelerator. The essential new assumption H. and F. London made is that the bracket is not an arbitrary constant but vanishes. Then one obtains the important second London equation

$$\vec{\nabla} \times \vec{J}_s = -\frac{n_s e^2}{m}\vec{B} . \tag{2}$$

It should be noted that this assumption cannot be justified within classical physics, even worse, in general it is wrong. For instance the current density in a normal metal will vanish when no electric field is applied, and whether a static magnetic field penetrates the metal is of no importance. In a superconductor, on the other hand, the situation is such that Eq. (2) applies.

Combining the fourth Maxwell equation (for time-independent fields)

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{J}_s$$

and the second London equation and making use of the relation

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{B}) = -\nabla^2\vec{B}$$

(this is valid since $\vec{\nabla} \cdot \vec{B} = 0$) we get the following equation for the magnetic field in a superconductor

$$\nabla^2\vec{B} - \frac{\mu_0 n_s e^2}{m}\vec{B} = 0 . \tag{3}$$

It is important to note that this equation is not valid in a normal conductor. In order to grasp the significance of Eq. (3) we consider a simple geometry, namely the boundary between a superconducting half space and vacuum, see Fig. 6a. Then, for a magnetic field parallel to the surface, Eq. (3) becomes

$$\frac{d^2B_y}{dx^2} - \frac{1}{\lambda_L^2}B_y = 0$$

with the solution

$$B_y(x) = B_0 \exp(-x/\lambda_L) .$$

Here we have introduced a very important superconductor parameter, the *London penetration depth*

$$\lambda_L = \sqrt{\frac{m}{\mu_0 n_s e^2}} \quad . \tag{4}$$

So the magnetic field does not stop abruptly at the superconductor surface but penetrates into the material with exponential attenuation. For typical material parameters the penetration depth is quite small, 20 – 50 nm. In the bulk of a thick superconductor there can be no magnetic field, which is just the Meissner-Ochsenfeld effect. Here it is appropriate to remark that in the BCS theory not single electrons but pairs of electrons are the carriers of the supercurrent. However, the penetration depth remains unchanged since when going from single electrons to pairs we have to make the following replacements

$$m \rightarrow 2m,\ e \rightarrow 2e,\ n_s \rightarrow n_s/2 .$$

We have now convinced ourselves that such a superconductor can tolerate a magnetic field only in a thin surface layer. An immediate consequence is that current flow is restricted to the same thin layer. Currents in the interior are forbidden as they would generate magnetic fields in the bulk. The magnetic field and the current which are caused by an external field parallel to the axis of a lead cylinder are plotted in Fig. 6b. Another interesting situation occurs if we pass a current through a lead wire (Fig. 6c). It flows only in a very thin surface sheet of about 20 nm thickness, so the overall current in the wire is small. This is a first indication that type I superconductors are not suitable for winding superconducting magnet coils.

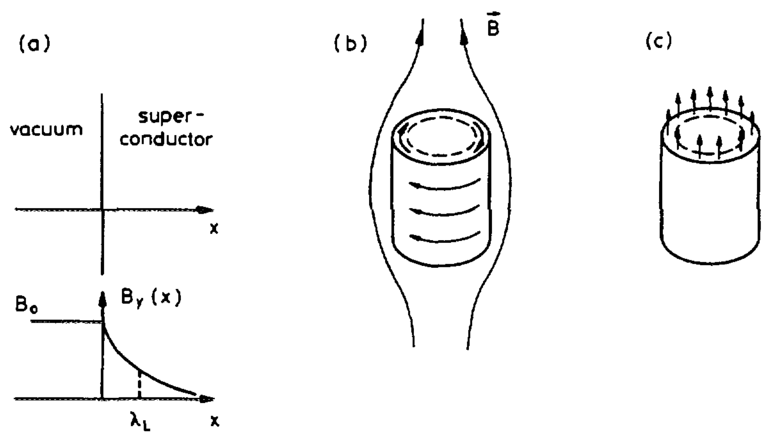


Figure 6: (a) Exponential attenuation of a magnetic field in a superconducting half plane. (b) Shielding current in a superconducting cylinder induced by a field parallel to the axis. (c) A current-carrying wire made from a type I superconductor.

The penetration depth has a temperature dependence which can be calculated in the BCS theory. When approaching the critical temperature, the density of the supercurrent carriers goes to zero, so λ_L must become infinite:

$$\lambda_L \to \infty \quad \text{for} \quad T \to T_c \, .$$

This is shown in Fig. 7. An infinite penetration depth means no attenuation of a magnetic field which is just what one observes in a normal conductor.

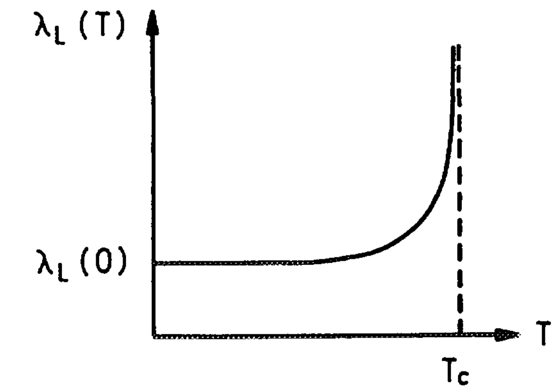


Figure 7: Temperature dependence of the London penetration depth

3 THERMODYNAMIC PROPERTIES OF SUPERCONDUCTORS

3.1 The superconducting phase

A material like lead goes from the normal into the superconducting state when it is cooled below T_c and when the magnetic field is less than $B_c(T)$. It has been mentioned already that this is a phase transition comparable to the transition from water to ice at 0°C and normal pressure. Phase transitions take place when the new state is energetically favoured. The relevant thermodynamic energy is here the free energy

$$F = U - T \cdot S \tag{5}$$

where U is the internal energy and S the entropy. The free energy of aluminium is plotted in Fig. 8. Below T_c the superconducting state has a lower free energy than the normal state and thus the transition normal → superconducting is associated with a gain in energy. The entropy of the superconducting state is lower which implies that there is a higher degree of order in this state. From the point of view of the BCS theory this is quite understandable since conduction electrons are paired and collect themselves in a single quantum state. Numerically the entropy difference is small, however, about 1 milli-Joule per mole and Kelvin, from which one can deduce that only a small fraction of the valence electrons of aluminium is condensed into Cooper pairs.

3.2 Energy balance in a magnetic field

We have argued that a lead cylinder becomes superconductive for $T < T_c$ because the free energy is reduced that way:

$$F_{\text{super}} < F_{\text{normal}} \text{ for } T < T_c \, .$$

What happens if we apply a magnetic field? A normal-conducting metal cylinder is just penetrated by the field so its free energy does not change: $F_{\text{normal}}(B) = F_{\text{normal}}(0)$. On the contrary,

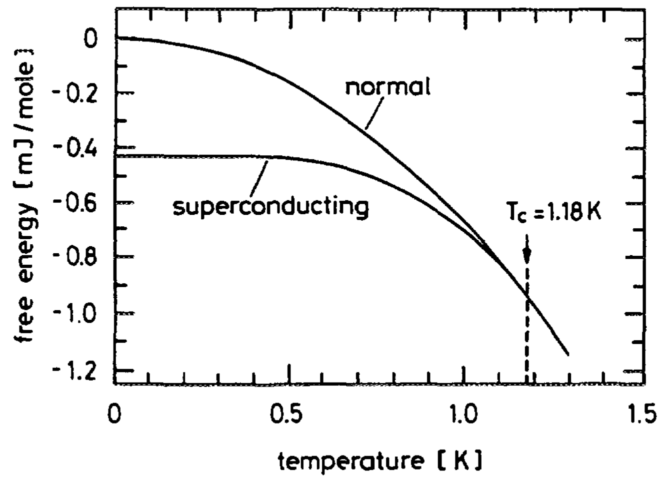


Figure 8: The free energy of normal and superconducting aluminium (after N.E. Philips). The normal state was achieved by exposing the sample to a field larger than B_c.

a superconducting cylinder is strongly affected by the field. It sets up shielding currents which generate a magnetic moment antiparallel to the applied field, see Fig. 9a. A magnetic moment $\vec{m}$ has a potential energy in the magnetic field

$$E_{\text{pot}} = -\vec{m} \cdot \vec{B} \, .$$

If $\vec{m}$ is antiparallel to $\vec{B}$ the potential energy is positive. In the following it is useful to introduce the magnetisation M as the magnetic moment per unit volume[1]. For a type I superconductor we have

$$M(B) = -B/\mu_0$$

for $B < B_c$. The potential energy per unit volume is obtained by integration

$$E_{\text{pot}} = -\int_0^B M(B')dB' = \frac{B^2}{2\mu_0} \, .$$

This is just the increase of the free energy in the superconducting state that is caused by the magnetic field. (More precisely one should use the free enthalpy instead of the free energy when a magnetic field is present but this is a detail we can disregard here.)

$$F_{\text{super}}(B) = F_{\text{super}}(0) + \frac{B^2}{2\mu_0} V_m$$

where V_m is the mole volume (10 cm^3 for Al). The *critical field* is achieved when the free energy in the superconducting state just equals the free energy in the normal state

$$\frac{B_c^2}{2\mu_0} V_m = F_{\text{normal}} - F_{\text{super}}(0) \, . \qquad (6)$$

[1]There is often a confusion whether the magnetic field should be the H or the B field. D.R. Tilley and J. Tilly [2] argue that the magnetisation of a superconductor inside a current-carrying coil resembles that of an iron core. Then the 'magnetising' field H which is generated by the coil current only is appropriate. The total field in the material is the superposition of H and the superconductor magnetisation M: $B = \mu_0(H + M(H))$. Often, however, the external field is produced by an electromagnet with iron pole shoes, and this is already a B field. Since B appears also in the Lorentz force I prefer to use this quantity. With this convention the field B_i inside a superconductor is given in terms of the external field B_e by $B_i = B_e + \mu_0 M(B_e)$. Unfortunately much of the superconductivity literature is based on the obsolete CGS system where the distinction between B and H is not very clear and the two fields have the same dimension although their units were given different names: Gauss and Oerstedt.

Since the energy density stored in a magnetic field is $B^2/(2\mu_0)$, an alternative interpretation of Eq. (6) is the following: in order to transform from the normal to the superconducting state the material has to push out the magnetic energy, and the largest amount it can push out is just the difference between the two free energies at vanishing field. Figure 9b illustrates what we have said. For $B > B_c$ the normal phase has a lower energy, so superconductivity breaks down.

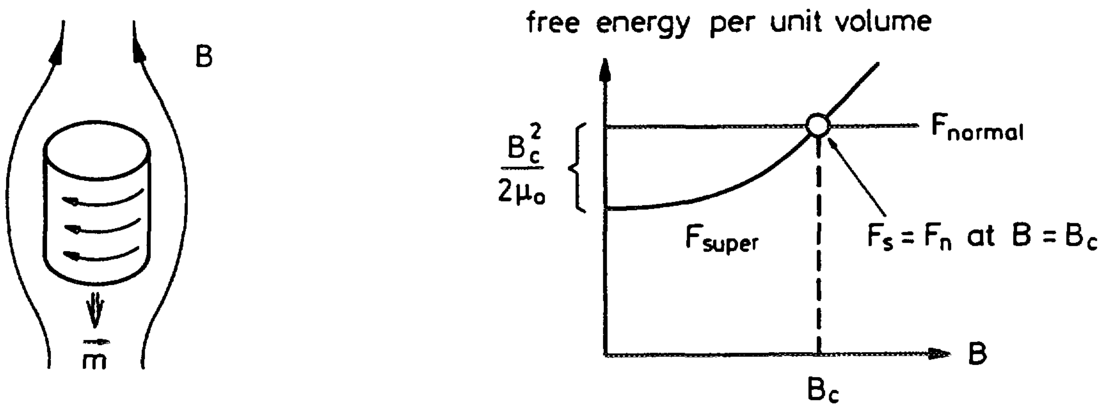


Figure 9: (a) The magnetic moment of a superconducting cylinder in an external field. (b) The free energies of the normal and the superconducting state as a function of the external field.

3.3 Type II superconductors

For practical application in magnets it would be rather unfortunate if only type I superconductors existed which permit no magnetic field and no current in the bulk material. Alloys and the element niobium are so-called type II superconductors. Their magnetisation curves exhibit a more complicated dependence on magnetic field (Fig. 10). Type II conductors are characterized by two critical fields, B_{c1} and B_{c2}, which are both temperature dependent. For fields $0 < B < B_{c1}$ the substance is in the *Meissner phase* with complete exclusion of the field from the interior. In the range $B_{c1} < B < B_{c2}$ the substance enters the *mixed phase*: part of the magnetic flux penetrates the bulk of the sample. Above B_{c2}, finally, the material is normal-conducting. The area under the curve

$$M = M(B)$$

is the same as for a type I conductor as it corresponds to the free-energy difference between the normal and the superconducting state.

It is instructive to compare measured data on pure lead (type I) and lead-indium alloys (type II) of various composition. Figure 11 shows that the upper critical field rises with increasing indium content; for Pb-In(20.4%) it is about eight times larger than the critical field of pure lead. The fact that the alloys stay superconductive up to much higher fields is easy to understand: magnetic flux is allowed to penetrate the sample and therefore less magnetic field energy has to be driven out. Under the assumption that the free-energy difference is the same for the various lead-indium alloys, the areas under the three curves A, B, C in Fig. 11 should be identical as the diagram clearly confirms.

In a (T, B) plane the three phases of a type II superconductor are separated by the curves $B_{c1}(T)$ and $B_{c2}(T)$ which meet at $T = T_c$, see Fig. 12a. The upper critical field can assume very large values which make these substances extremely interesting for magnet coils (Fig. 12b). A

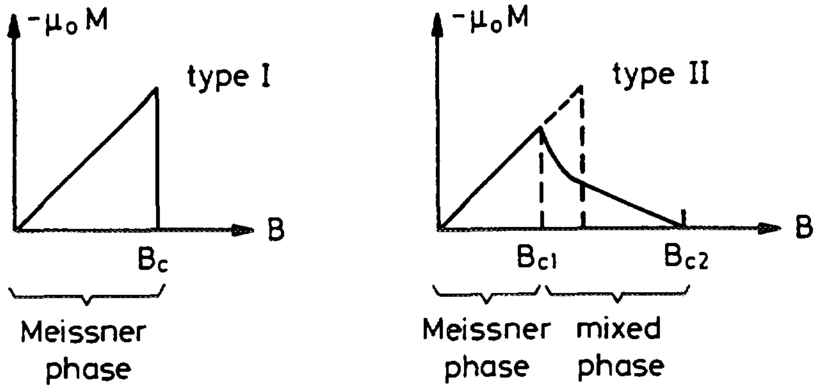


Figure 10: Magnetisation of type I and type II superconductors as a function of field.

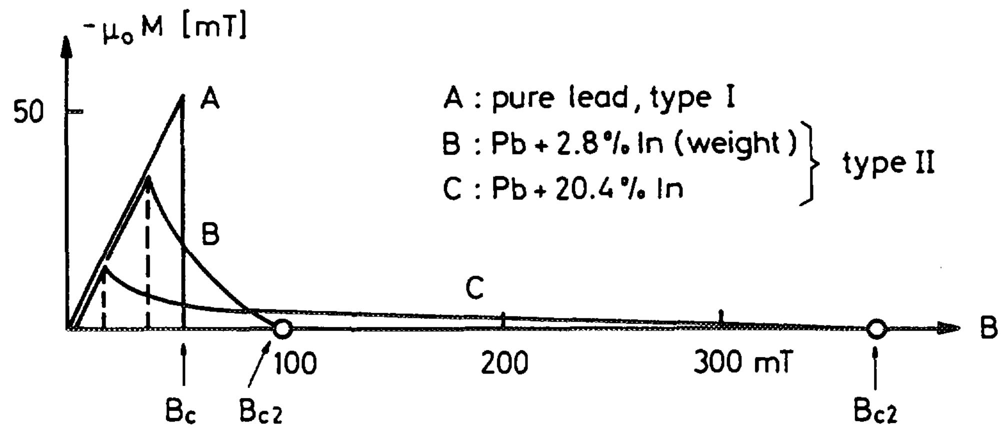


Figure 11: The measured magnetisation curves [14] of lead-indium alloys of various composition.

remarkable feature, which will be addressed in more detail in Sec. 6.3, is the observation that the magnetic flux does not penetrate the type II conductor with uniform density. Rather it is concentrated in *flux tubes* as sketched in Fig. 13. Each tube is surrounded with a super-vortex current. The material in between the tubes is field- and current-free.

3.4 Distinction between type I and type II superconductors

3.4.1 Thin sheets of type I superconductors

Let us first stick to type I conductors and compare the magnetic properties of a very thin sheet (thickness $d < \lambda_L$) to those of a thick slab. The thick slab has a vanishing field in the bulk (Fig. 14a) while in the thin sheet (Fig. 14b) the applied field does not drop to zero at the centre. Consequently less energy is needed to expel the field which implies that the critical field of the thin sheet is much larger than the B_c of a thick slab. From this point of view it might appear energetically favourable for a thick slab to subdivide itself into an alternating sequence of thin normal and superconducting slices as indicated in Fig. 14c. The magnetic energy is indeed lowered that way but there is another energy to be taken into consideration, namely the energy required to create the normal-superconductor interfaces. A subdivision is only sensible if the interface energy is less than the magnetic energy.

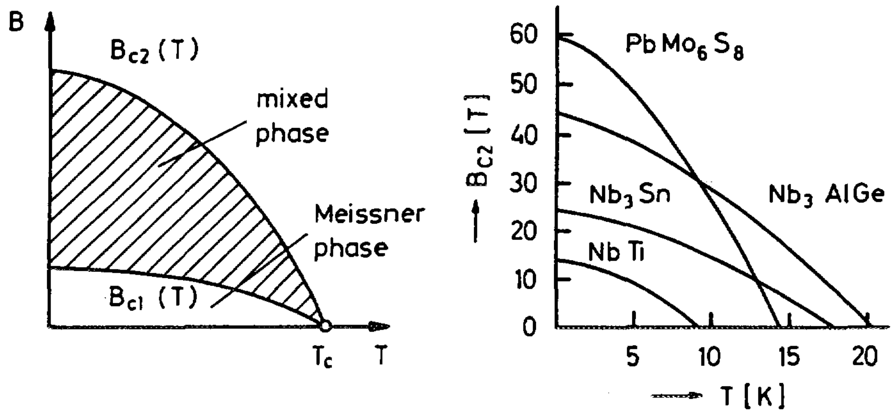


Figure 12: (a) The phase diagram of a type II superconductor. (b) The upper critical field of several high-field alloys as a function of temperature.

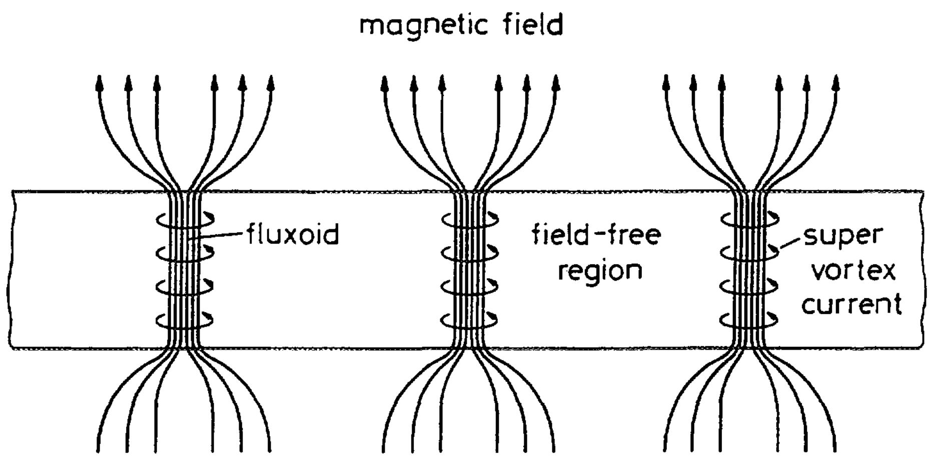


Figure 13: Flux tubes in a type II superconductor.

3.4.2 Coherence length

At a normal-superconductor boundary the density of the super charge carriers (the Cooper pairs) does not rise abruptly from zero to its value in the bulk but rises smoothly over a finite length ξ, the coherence length, see Fig. 15.

The relative size of the London penetration depth and the coherence length decides whether a material is a type I or a type II superconductor. To study this in a semi-quantitative way, we first define the *thermodynamical critical field* by the energy relation

$$\frac{B_{cth}^2}{2\mu_0} V_m = F_{\mathrm{normal}} - F_{\mathrm{super}}(0) \,. \tag{7}$$

For type I, B_{cth} coincides with B_c, see Eq. (6) while for type II conductors B_{cth} lies between B_{c1} and B_{c2}. The difference between the two free energies, $F_{\mathrm{normal}} - F_{\mathrm{super}}(0)$, can be interpreted as the Cooper-pair condensation energy. For a conductor of unit area, exposed to a field $B = B_{cth}$ parallel to the surface, the energy balance is as follows:
(a) the magnetic field penetrates a depth λ_L of the sample which corresponds to an energy gain since magnetic energy must not be driven out of this layer:

$$\Delta E_{\mathrm{magn}} = \frac{B_{cth}^2}{2\mu_0} \lambda_L \,.$$

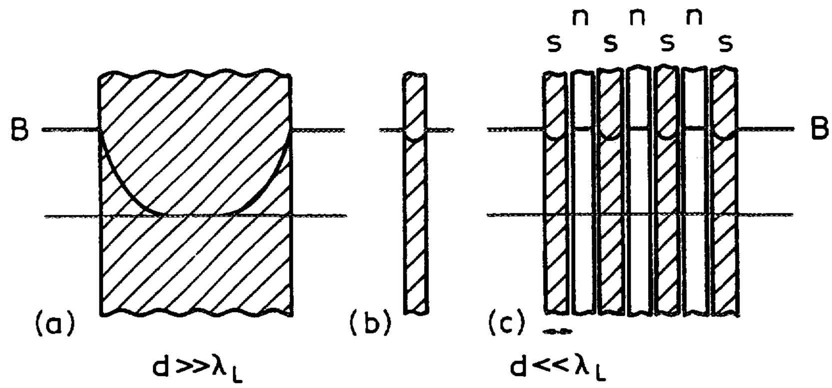


Figure 14: Attenuation of field (a) in a thick slab and (b) in thin sheet. (c) Subdivision of a thick slab into alternating layers of normal and superconducting slices.

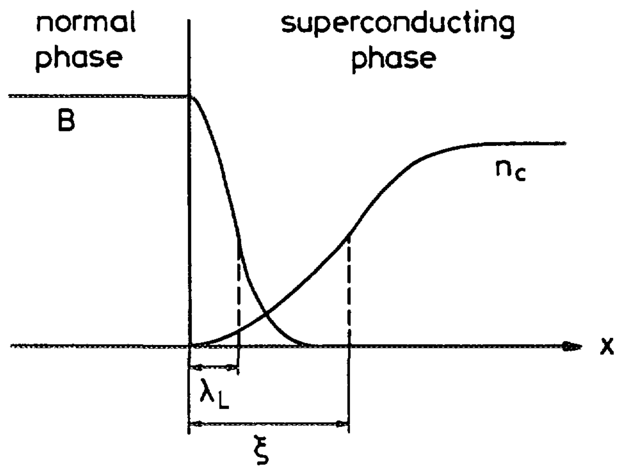


Figure 15: The decay of the magnetic field and the rise of the Cooper pair density at a normal-superconductor interface.

(b) On the other hand, the fact that the Cooper-pair density does not assume its full value right at the surface but rises smoothly over a length ξ implies a loss of condensation energy

$$\Delta E_{\text{cond}} = -\frac{B_{c\,th}^2}{2\mu_0}\,\xi\,.$$

Obviously there is a net gain if $\lambda_L > \xi$. So a subdivision of the superconductor into an alternating sequence of thin normal and superconducting slices is energetically favourable if the London penetration depth exceeds the coherence length.
A more refined treatment is provided by the Ginzburg-Landau theory [9]. Here one introduces the *Ginzburg-Landau parameter*

$$\kappa = \lambda_L/\xi\,. \tag{8}$$

The criterion for type I or II superconductivity is then

$$\begin{aligned} \text{type I:} &\quad \kappa < 1/\sqrt{2} \\ \text{type II:} &\quad \kappa > 1/\sqrt{2}. \end{aligned}$$

In reality a type II superconductor is not subdivided into thin slices but the field penetrates the sample in flux tubes which arrange themselves in a triangular pattern. The core of a flux tube is normal. The following table lists the penetration depths and coherence lengths of some important superconducting elements. Niobium is a type II conductor but close to the border to type I, while indium, lead and tin are clearly type I conductors.

material	In	Pb	Sn	Nb
λ_L [nm]	24	32	≈ 30	32
ξ [nm]	360	510	≈ 170	39

The coherence length ξ is proportional to the mean free path of the conduction electrons in the metal. In alloys the mean free path is generally much shorter than in pure metals so they are always type II conductors.
The upper critical field is given by

$$B_{c2} = \sqrt{2}\,\kappa\, B_{cth}\,. \tag{9}$$

There exists no simple expression for the lower critical field. In the limit $\kappa \gg 1$ one gets

$$B_{c1} = \frac{1}{2\kappa}(\ln\kappa + 0.08)B_{cth}\,. \tag{10}$$

3.5 Heat capacity and heat conductivity

The specific heat capacity per mole of a normal metal at low temperatures is given by the expression

$$C(T) = \gamma T + \beta T^3\,. \tag{11}$$

The linear term in T comes from the conduction electrons, the cubic term from lattice vibrations. The coefficients can be calculated within the free-electron-gas model and the Debye theory of lattice specific heat (see any standard textbook on solid state physics):

$$\gamma = \frac{\pi^2 n_m k_B^2}{2E_F}\,, \quad \beta = \frac{12\pi^4 N_A k_B}{5\Theta_D^3}\,. \tag{12}$$

Here $k_B = 1.38\cdot 10^{-23}$ J/K is the Boltzmann constant, E_F the Fermi energy, n_m the number of electrons per mole, $N_A = 6.023\cdot 10^{23}$ Avogadro's number and Θ_D the Debye temperature of the material. If one plots the ratio $C(T)/T$ as a function of T^2 a straight line is obtained as can be seen in Fig. 16a for normal-conducting gallium [15]. In the superconducting state the electronic specific heat is different because the electrons bound in Cooper pairs no longer contribute. In the BCS theory one expects an exponential rise of the electronic heat capacity with temperature

$$C_{e,s}(T) = 8.5\,\gamma T_c\,\exp(-1.44\,T_c/T) \tag{13}$$

The experimental data (Fig. 16a,b) are in remarkable agreement with this prediction. There is a resemblance to the exponential temperature dependence of the electrical conductivity in intrinsic semiconductors and these data can be taken as an indication that an energy gap exists also in superconductors.

The heat conductivity of niobium is of particular interest for superconducting RF cavities. Here the theoretical predictions are rather imprecise and measurements are indispensible. The low temperature values depend strongly on the residual resistivity ratio $RRR = R(293\,\mathrm{K})/R(4\,\mathrm{K})$ of the normal Nb and on the grain size of the crystallites. Figure 17 shows two curves obtained at DESY [16].

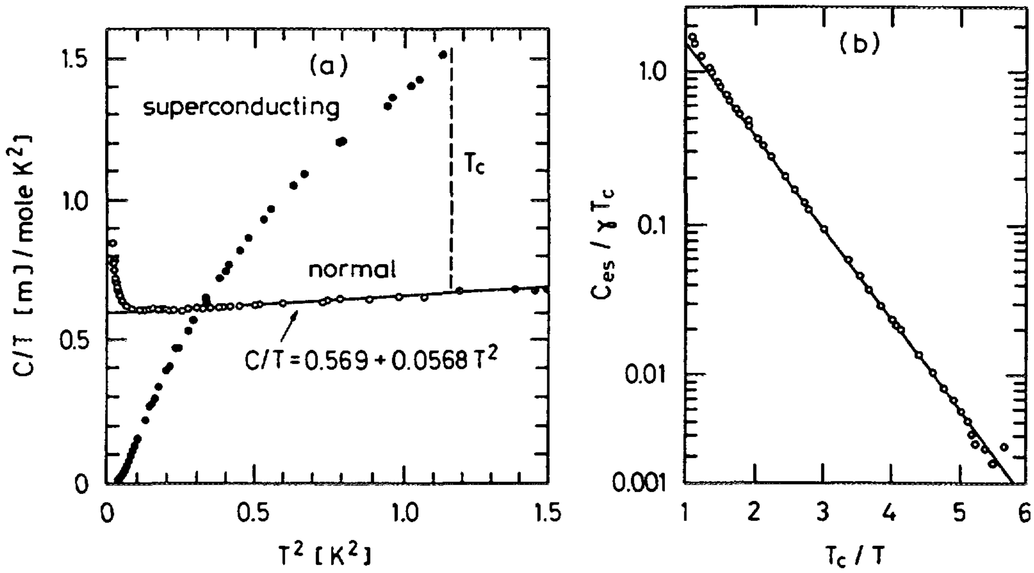


Figure 16: (a) Specific heat $C(T)/T$ of normal and superconducting gallium as a function of T^2 [15]. (b) Experimental verification of Eq. (13).

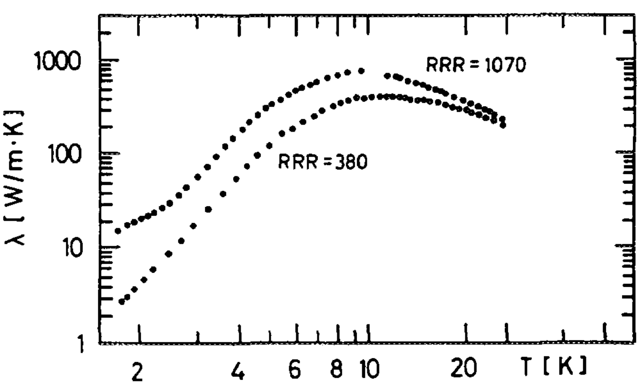


Figure 17: Measured heat conductivity in niobium [16].

4 BASIC CONCEPTS AND RESULTS OF THE BCS THEORY

4.1 The 'free electron gas' in a normal metal

4.1.1 The Fermi sphere

In a metal like copper the positively charged ions form a regular crystal lattice. The valence electrons (one per Cu atom) are not bound to specific ions but can move through the crystal. In the simplest quantum theoretical model the Coulomb attraction of the positive ions is represented by a potential well with a flat bottom, the periodic structure is neglected (this leads to the electronic band structure needed for the description of semiconductors). The energy levels are computed by solving the Schrödinger equation and then the electrons are placed on these levels paying attention to the Pauli exclusion principle: no more than two electrons of opposite spin are allowed on each level. The electrons are treated as independent and non-interacting particles, their mutual Coulomb repulsion is taken into account only globally by a suitable choice of the depth of the potential well. It is remarkable that such a simple-minded picture of a 'free electron gas' in a metal can indeed reproduce the main features of electrical and thermal conduction in metals. However, an essential prerequisite is to apply the Fermi-Dirac statistics, based on the Pauli Principle, and to avoid the classical Boltzmann statistics which one uses for normal gases. The electron gas has indeed rather peculiar properties. The average kinetic energy of the metal electrons is by no means given by the classical expression

$$\frac{m}{2}\overline{v^2} = \frac{3}{2}k_B T$$

which amounts to about 0.025 eV at room temperature ($k_B = 1.38\cdot10^{-23}\,\mathrm{J/K} = 8.62\cdot10^{-5}\,\mathrm{eV/K}$ is the Boltzmann constant). Instead, the energy levels are filled with two electrons each up to the so-called *Fermi energy* E_F. Since the electron density is very high in metals, E_F assumes large values, typically 5 eV. The average kinetic energy of an electron is $3/5E_F \approx 3$ eV and thus much larger than the average energy of a usual gas molecule. The electrons constitute a system called a 'highly degenerate' Fermi gas. The Fermi energy is given by the formula

$$E_F = \frac{\hbar^2}{2m}(3\pi^2 n_0)^{2/3} \,. \tag{14}$$

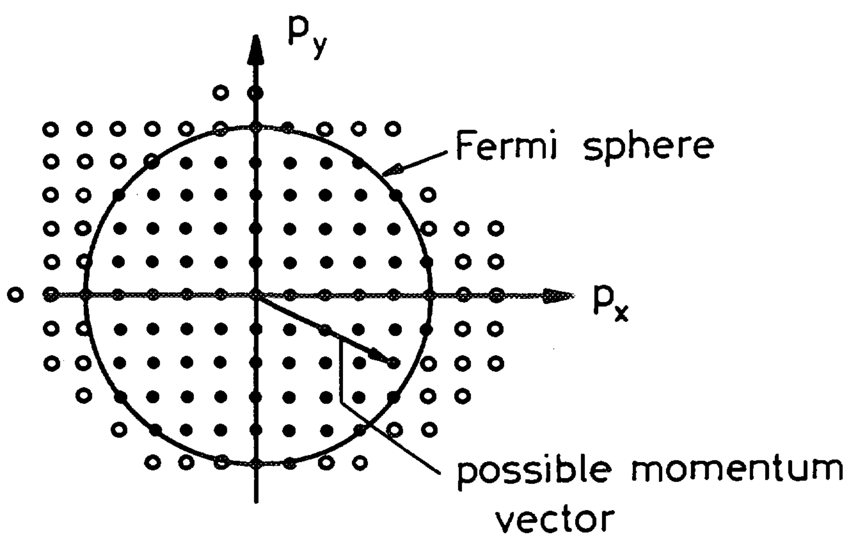


Figure 18: The allowed states for conduction electrons in the $p_x p_y$ plane and the Fermi sphere. The occupied states are drawn as full circles, the empty states as open circles.

The quantity $\hbar = h/2\pi = 1.05\cdot10^{-34}$ Js $= 6.58\cdot10^{-15}$ eVs is Planck's constant, the most important constant in quantum theory; n_0 is the density of conduction electrons. It is useful to plot the allowed quantum states of the electrons as dots in momentum space. In Fig. 18 this is drawn for two dimensions. The quantisation is usually done in a cubic box of length L; the possible momentum vectors of the electrons are then

$$\vec{p} = (p_x, p_y, p_z) = \hbar\vec{k} \quad \text{with} \quad \vec{k} = (n_x, n_y, n_z)\,2\pi/L$$

where the n_i are integer numbers $(0, \pm1, \pm2, \ldots)$. At low temperature ($T \rightarrow 0$) the energy levels below the Fermi energy are all filled while levels above E_F are empty. The highest momentum is the 'Fermi momentum' given by

$$p_F = \sqrt{2mE_F}$$

and the highest velocity is the Fermi velocity $v_F = p_F/m$ which is in the order of 10^6 m/s. The occupied states are located inside the 'Fermi sphere' of radius p_F, the empty states are outside.

What are the consequences for electrical conduction? Let us apply an electric field $\vec{E}_0$ pointing into the negative x direction. In the time δt a free electron would gain a momentum

$$\delta p_x = eE_0\,\delta t\,. \tag{15}$$

However, most of the metal electrons are unable to accept this momentum because they do not find free states in their vicinity, only those on the right rim of the Fermi sphere have free states accessible to them and can accept the additional momentum. We see that the Pauli principle has a strong impact on electrical conduction. Heat conduction is affected in the same way because the most important carriers of thermal energy are again the electrons. An anomaly is also observed in the heat capacity of the electron gas. It differs considerably from that of a mono-atomic normal gas since only the electrons in a shell of thickness k_BT near the surface of the Fermi sphere can contribute. Hence the electronic specific heat per unit volume is roughly a fraction k_BT/E_F of the classical value

$$C_e \approx \frac{3}{2}nk_B\cdot\frac{k_BT}{E_F}\,.$$

This explains the linear temperature dependence of the electronic specific heat which was discussed in Sec. 3.5.

4.1.2 The origin of Ohmic resistance

Before trying to understand the vanishing resistance of a superconductor we have to explain first why a normal metal has a resistance at all. This may appear trivial if one imagines the motion of electrons in a crystal that is densely filled with ions. Intuitively one would expect that the electrons can travel for very short distances only before hitting an ion and thereby loosing the momentum gained in the electric field. Collisions are indeed responsible for a frictional force and one can derive Ohm's law that way. What is surprising is the fact that these collisions are so rare. In an ideal crystal lattice there are no collisions whatsoever. This is impossible to understand in the particle picture, one has to treat the electrons as matter waves and solve the Schrödinger equation for a periodic potential. The Ohmic resistance is nevertheless due to collisions but the collision centres are not the ions in the regular crystal lattice but only imperfections: impurities, lattice defects and the deviations of the metal ions from their nominal position due to thermal oscillations. The third effect dominates at room temperature and gives rise to a resistivity that is roughly proportional to T while impurities and lattice defects are responsible for the residual resistivity at low temperature ($T < 20$ K). A typical curve $\rho(T)$ is plotted in Fig. 19. In very pure copper single crystals the low-temperature resistivity can become extremely small. The mean free path of the conduction electrons may be a million times larger than the distance between neighbouring ions which illustrates very well that the ions in their regular lattice positions do not act as scattering centres.

4.2 Cooper pairs

We consider a metal at $T \to 0$. All states inside the Fermi sphere are filled with electrons while all states outside are empty. In 1956 Cooper studied [18] what would happen if two electrons were added to the filled Fermi sphere with equal but opposite momenta $\vec{p}_1 = -\vec{p}_2$ whose magnitude was slightly larger than the Fermi momentum p_F (see Fig. 20). Assuming that a weak attractive force existed he was able to show that the electrons form a bound system with an energy less than twice the Fermi energy , $E_{\text{pair}} < 2E_F$. The mathematics of Cooper pair formation will be outlined in Appendix A .

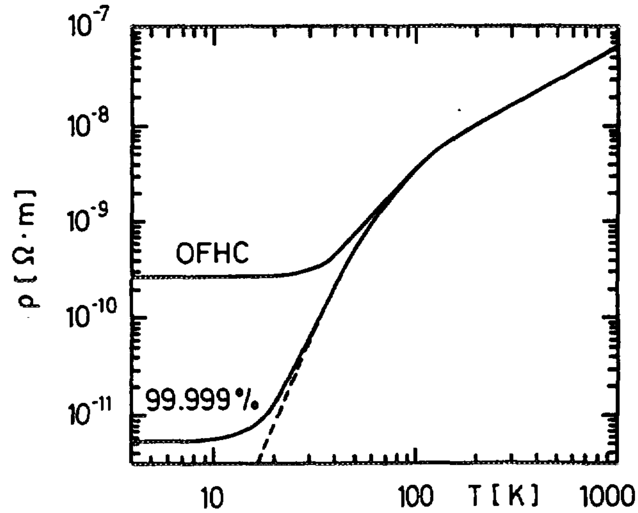


Figure 19: Temperature dependence of the resistivity of OFHC (oxygen-free high conductivity) copper and of 99.999% pure annealed copper. Plotted as a dashed line is the calculated resistivity of copper without any impurities and lattice defects (after M.N. Wilson [17]).

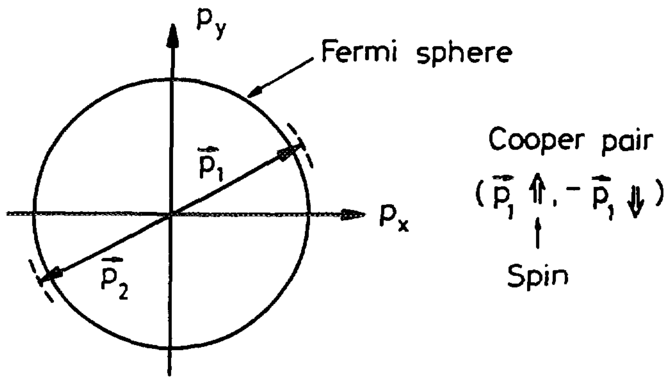


Figure 20: A pair of electrons of opposite momenta added to the full Fermi sphere.

What could be the reason for such an attractive force? First of all one has to realize that the Coulomb repulsion between the two electrons has a very short range as it is shielded by the positive ions and the other electrons in the metal. So the attractive force must not be strong if the electrons are several lattice constants apart. Already in 1950, Fröhlich and, independently, Bardeen had suggested that a dynamical lattice polarization may create a weak attractive potential. Before going into details let us look at a familiar example of attraction caused by the deformation of a medium: if you are cross-country skiing in very deep snow you will find this quite cumbersome, there is a lot of 'resistance'. Now you discover a track made by another skier, a 'Loipe', and you will immediately realize that it is much more comfortable to ski along this track than in any other direction.

This picture can be adopted for our electrons. The first electron flies through the lattice and attracts the positive ions. Because of their inertia they cannot follow immediately, the shortest response time corresponds to the highest possible lattice vibration frequency. This is called the Debye frequency ω_D. The maximum lattice deformation lags behind the electron by a distance

$$d \approx v_F \frac{2\pi}{\omega_D} \approx 100 - 1000\,\mathrm{nm}\,. \tag{16}$$

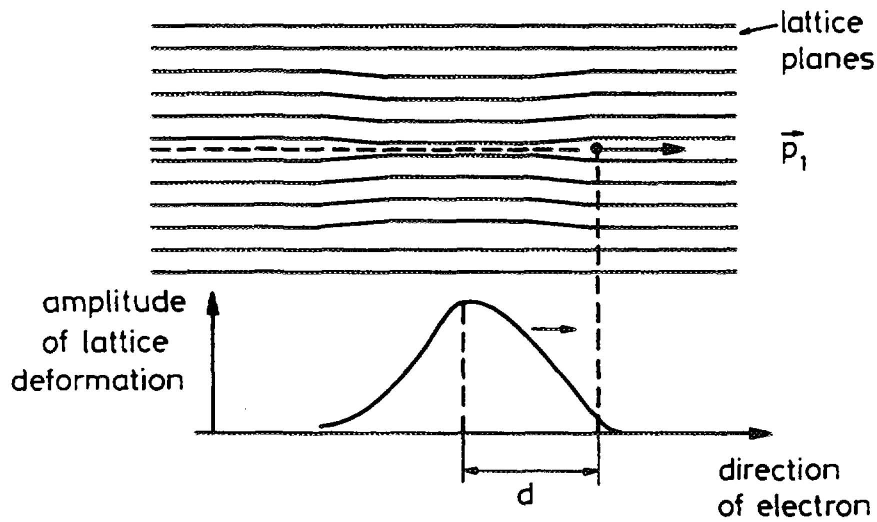


Figure 21: Dynamical deformation of the crystal lattice caused by the passage of a fast electron. (After Ibach, Lüth [19]).

Obviously, the lattice deformation attracts the second electron because there is an accumulation of positive charge. The attraction is strongest when the second electron moves right along the track of the first one and when it is a distance d behind it, see Fig. 21. This explains why a Cooper pair is a very extended object, the two electrons may be several 100 to 1000 lattice constants apart. For a simple cubic lattice, the lattice constant is the distance between adjacent atoms.

In the example of the cross-country skiers or the electrons in the crystal lattice, intuition suggests that the second partner should preferably have the same momentum, $\vec{p}_2 = \vec{p}_1$ although opposite momenta $\vec{p}_2 = -\vec{p}_1$ are not so bad either. Quantum theory makes a unique choice: only electrons of opposite momenta form a bound system, a Cooper pair. I don't know of any intuitive argument why this is so. (The quantum theoretical reason is the Pauli Principle but there exists probably no intuitive argument why electrons obey the Pauli Exclusion Principle and are thus extreme individualists while other particles like the photons in a laser or the atoms in superfluid helium do just the opposite and behave as extreme conformists. One may get used to quantum theory but certain mysteries and strange feelings will remain.)

The binding energy of a Cooper pair turns out to be small, $10^{-4} - 10^{-3}$ eV, so low temperatures are needed to preserve the binding in spite of the thermal motion. According to Heisenberg's Uncertainty Principle a weak binding is equivalent to a large extension of the composite system, in this case the above-mentioned $d = 100 - 1000$ nm. As a consequence, the Cooper pairs in a superconductor overlap each other. In the space occupied by a Cooper pair there are about a million other Cooper pairs. Figure 22 gives an illustration. The situation is totally different from composite systems like atomic nuclei or atoms which are tightly bound and well-separated objects. The strong overlap is an important prerequisite of the BCS theory because the Cooper pairs must change their partners frequently in order to provide a continuous binding.

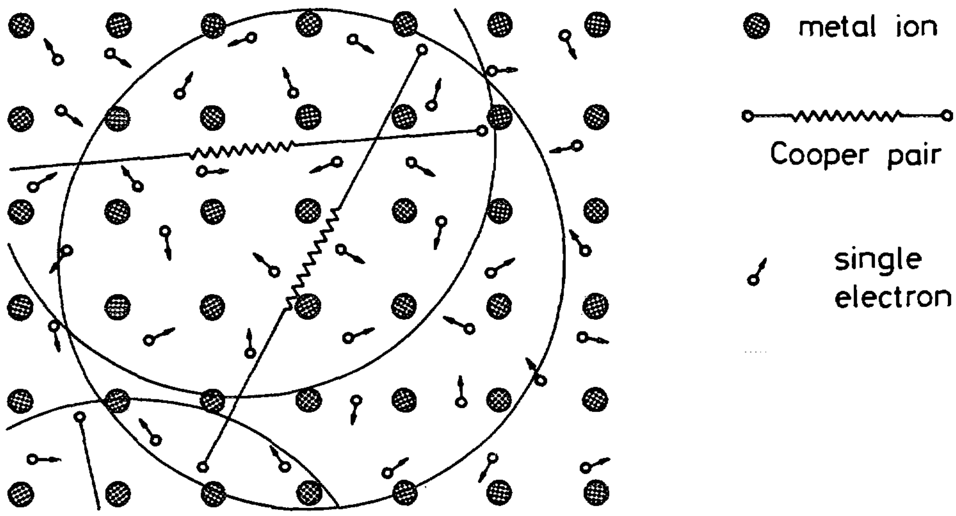


Figure 22: Cooper pairs and single electrons in the crystal lattice of a superconductor. (After Essmann and Träuble [12]).

4.3 Elements of the BCS theory

After Cooper had proved that two electrons added to the filled Fermi sphere were able to form a bound system with an energy $E_{pair} < 2E_F$, it was immediately realized by Bardeen, Cooper and Schrieffer that also the electrons inside the Fermi sphere should be able to group themselves into pairs and thereby reduce their energy. The attractive force is provided by lattice vibrations whose quanta are the phonons. The highest possible phonon energy is

$$\hbar\omega_D = k_B\Theta_D \approx 0.01 - 0.02\,\text{eV}\,. \tag{17}$$

Therefore only a small fraction of the electrons can be paired via phonon exchange, namely those in a shell of thickness $\pm\hbar\omega_D$ around the Fermi energy. This is sketched in Fig. 23. The inner electrons cannot participate in the pairing because the energy transfer by the lattice is too small. One has to keep in mind though that these electrons do not contribute to normal conduction either. For vanishing electric field a Cooper pair is a loosely bound system of two electrons whose momenta are of equal magnitude but opposite direction. All Cooper pairs have therefore the same momentum $\vec{P} = 0$ and occupy exactly the same quantum state. They can be described by a macroscopic wave function Ψ in analogy with a radio wave in which the photons are all in phase and have the same wavelength. The macroscopic photon wave function is the vector potential from which one can derive the electric and magnetic field vectors.

The reason why Cooper pairs are allowed and even prefer to enter the same quantum state is that they behave as Bose particles with spin 0. This is no contradiction to the fact that their constituents are spin 1/2 Fermi particles. Figure 23 shows very clearly that the individual electrons forming the Cooper pairs have different momentum vectors $\vec{p}, \vec{p}\,', \vec{p}\,'', \ldots$ which however cancel pairwise such that the pairs have all the same momentum zero. It should be noted though that Cooper pairs differ considerably from other Bosons such as helium nuclei or atoms: They are not 'small' but very extended objects, they exist only in the BCS ground state and there is no excited state. An excitation is equivalent to breaking them up into single electrons.

The BCS ground state is characterized by the macroscopic wave function Ψ and a ground state energy that is separated from the energy levels of the unpaired electrons by an energy gap. In order to break up a pair an energy of 2Δ is needed. This is illustrated in Fig. 24.

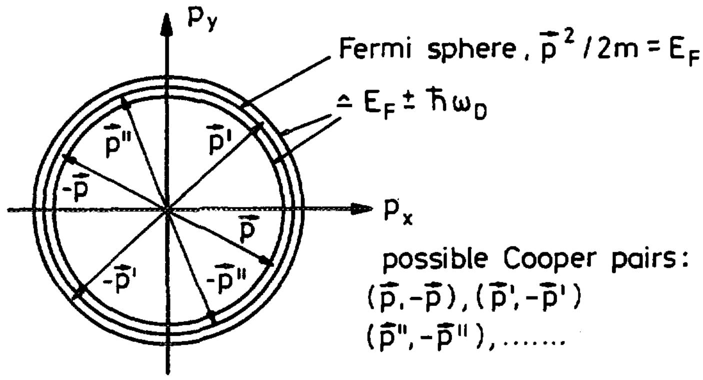


Figure 23: Various Cooper pairs $(\vec{p}, -\vec{p})$, $(\vec{p}', -\vec{p}')$, $(\vec{p}'', -\vec{p}'')$, ... in momentum space.

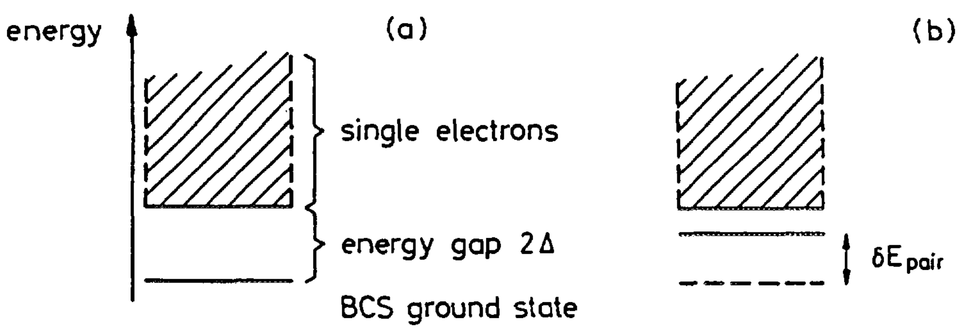


Figure 24: (a) Energy gap between the BCS ground state and the single-electron states.
(b) Reduction of energy gap in case of current flow.

There is a certain similarity with the energy gap between the valence and the conduction band in a semiconductor but one important difference is that the energy gap in a superconductor is not a constant but depends on temperature. For $T \to T_c$ one gets $\Delta(T) \to 0$. The BCS theory makes a quantitative prediction for the function $\Delta(T)$ which is plotted in Fig. 25 and agrees very well with experimental data. One of the fundamental formulae of the BCS theory is the relation between the energy gap $\Delta(0)$ at $T = 0$, the Debye frequency ω_D and the electron-lattice interaction potential V_0:

$$\Delta(0) = 2\hbar\omega_D \exp\left(-\frac{1}{V_0 \mathcal{N}(E_F)}\right) . \tag{18}$$

Here $\mathcal{N}(E_F)$ is the density of single-electron states of a given spin orientation at $E = E_F$ (the other spin orientation is not counted because a Cooper pair consists of two electrons with opposite spin). Although the interaction potential V_0 is assumed to be weak, one of the most striking observations is that the exponential function cannot be expanded in a Taylor series around $V_0 = 0$ because all coefficients vanish identically. This implies that Eq. (18) is a truely non-perturbative result. The fact that superconductivity cannot be derived from normal conductivity by introducing a 'small' interaction potential and applying perturbation theory (which is the usual method for treating problems of atomic, nuclear and solid state physics that have no analytical solution) explains why it took so many decades to find the correct theory.

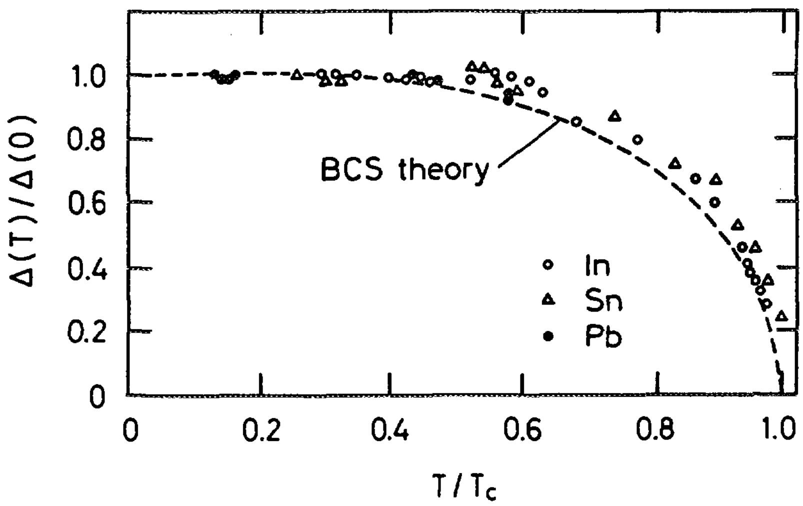


Figure 25: Temperature dependence of the energy gap according to the BCS theory and comparison with experimental data.

The critical temperature is given by a similar expression

$$k_B T_c = 1.14\, \hbar\omega_D \exp\left(-\frac{1}{V_0 \mathcal{N}(E_F)}\right) . \tag{19}$$

Combining the two equations we arrive at a relation between the energy gap and the critical temperature which does not contain the unknown interaction potential

$$\Delta(0) = 1.76\, k_B T_c . \tag{20}$$

The following table shows that this remarkable prediction is fulfilled rather well.

element	Sn	In	Tl	Ta	Nb	Hg	Pb
$\Delta(0)/k_B T_c$	1.75	1.8	1.8	1.75	1.75	2.3	2.15

In the BCS theory the underlying mechanism of superconductivity is the attractive force between pairs of electrons that is provided by lattice vibrations. It is of course highly desirable to find experimental support of this basic hypothesis. According to Eq. (19) the critical temperature is proportional to the Debye frequency which in turn is inversely proportional to the square root of the atomic mass M:

$$T_c \propto \omega_D \propto 1/\sqrt{M} .$$

If one produces samples of different isotopes of a superconducting element one can check this relation. Figure 26 shows T_c measurements on tin isotopes. The predicted $1/\sqrt{M}$ law is very well obeyed.

4.4 Supercurrent and critical current

The most important task of a theory of superconductivity is of course to explain the vanishing resistance. We have seen in Sec. 4.1 that the electrical resistance in normal metals is caused by scattering processes so the question is why Cooper pairs do not suffer from scattering while unpaired electrons do.

To start a current in the superconductor, let us apply an electric field $\vec{E}_0$ for a short time δt.

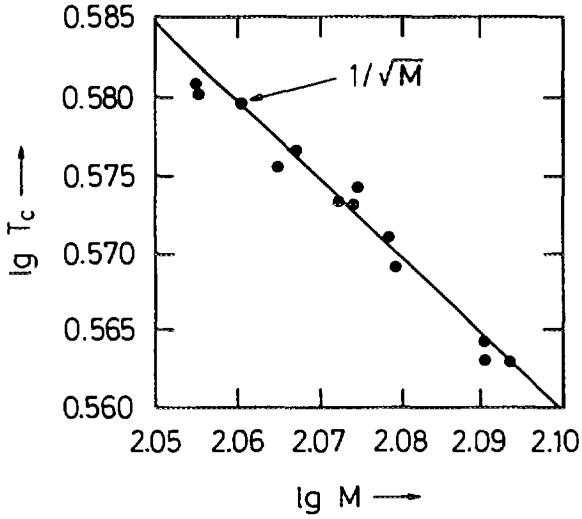


Figure 26: The critical temperature of various tin isotopes.

Both electrons of a Cooper pair receive an additional momentum $\delta\vec{p} = -e\vec{E}_0\,\delta t$ so after the action of the field all Cooper pairs have the same non-vanishing momentum

$$\vec{P} = \hbar\vec{K} = -2e\vec{E}_0\,\delta t.$$

Associated with this coherent motion of the Cooper pairs is a supercurrent density

$$\vec{J}_s = -n_c\frac{e\hbar}{m}\vec{K}\,. \tag{21}$$

Here n_c is the Cooper-pair density. It can be shown (see e.g. Ibach, Lüth [19]) that the Cooper-pair wave function with a current flowing is simply obtained by multiplying the wave function at rest with the phase factor

$$\exp(i\vec{K}\cdot\vec{R})\,.$$

Here $\vec{R} = (\vec{r}_1 + \vec{r}_2)/2$ is the coordinate of the centre of gravity of the two electrons. Moreover the electron-lattice interaction potential is not modified by the current flow. So all equations of the BCS theory remain applicable and there will be still an energy gap provided the energy gain δE_{pair} of the Cooper pair is less than 2Δ, see Fig. 24b. It is this remaining energy gap which prevents scattering. As we have seen there are two types of scattering centres: impurities and thermal lattice vibrations. Cooper pairs can only scatter when they gain sufficient energy to cross the energy gap. An impurity is a fixed heavy target and scattering cannot increase the energy of the electrons of the pair, therefore impurity scattering is prohibited for the Cooper pairs. Scattering on thermal lattice vibrations does not occur as long as the thermal energy is too small to surpass the energy gap (that means as long as the temperature is less than the critical temperature for the given current density). So we arrive at the conclusion that there is resistance-free current transport provided there is still an energy gap present ($2\Delta - \delta E_{\text{pair}} > 0$) and the temperature is sufficiently low ($T < T_c(J_s)$).
The supercurrent density is limited by the condition that the energy gain δE_{pair} must be less than the energy gap. This leads to the concept of the *critical current density.* The energy of the Cooper pair is, after application of the electric field,

$$E_{\text{pair}} = \frac{1}{2m}\left((\vec{p}+\vec{P}/2)^2 + (-\vec{p}+\vec{P}/2)^2\right) = \frac{\vec{p}^{\,2}}{m} + \delta E_{\text{pair}}$$

with $\delta E_{\text{pair}} \approx p_F P/m$. From the condition $\delta E_{\text{pair}} \leq 2\Delta$ we get

$$J_s \leq J_c \approx 2e\, n_c \Delta / p_F \,. \tag{22}$$

Coupled to a maximum value of the current density is the existence of a critical magnetic field. The current flowing in a long wire of type I superconductor is confined to a surface layer of thickness λ_L, see Fig. 6c. The maximum permissible current density J_c is related to the critical field

$$B_c(T) = \mu_0\, \lambda_L\, J_c(T) \approx \mu_0 \lambda_L\, 2en_c\, \Delta(T)/p_F \,. \tag{23}$$

The temperature dependence of the critical field is caused by the temperature dependence of the gap energy.

The above considerations on resistance-free current flow may appear a bit formal so I would like to give a more familiar example where an energy gap prevents 'resistance' in a generalized sense. We compare crystals of diamond and silicon. Diamond is transparent to visible light, silicon is not. So silicon represents a 'resistance' to light. Why is this so? Both substances have exactly the same crystal structure, namely the 'diamond lattice' that is composed of two face-centred cubic lattices which are displaced by one quarter along the spatial diagonal. The difference is that diamond is built up from carbon atoms and is one of the best electrical insulators while a silicon crystal is made from Si atoms and is a semiconductor. In the band theory of solids there is an energy gap E_g between the valence band and the conduction band. The gap energy is around 7 eV for diamond and 1 eV for silicon. Visible light has a quantum energy of about 2.5 eV. A photon impinging on a silicon crystal can lift an electron from the valence to the conduction band and is thereby absorbed. The same photon impinging on diamond is unable to supply the required energy of 7 eV, so this photon simply passes the crystal without absorption. (Quantum conditions of this kind have already been known in the Stone Age. If hunters wanted to catch an antelope that could jump 2 m high, they would dig a hole 4 m deep and then the animal could never get out because being able to jump 2 m in two successive attempts is useless for overcoming the 4 m. The essential feature of a quantum process, namely that the energy gap has to be bridged in a single event, is already apparent in this trivial example).

5 THE VECTOR POTENTIAL IN QUANTUM THEORY

Several important superconductor properties, in particular the magnetic flux quantisation, can only be explained by studying the magnetic vector potential and its impact on the so-called 'canonical momentum' of the charge carriers. Since this may not be a familiar concept I will spend some time to discuss the basic ideas and the supporting experiments which are beautiful examples of quantum interference phenomena.

5.1 The vector potential in electrodynamics

In classical electrodynamics it is often a matter of convenience to express the magnetic field as the curl (rotation) of a vector potential

$$\vec{B} = \vec{\nabla} \times \vec{A} \,.$$

The magnetic flux through an area F can be computed from the line integral of $\vec{A}$ along the rim of F by using Stoke's theorem:

$$\Phi_{\text{mag}} = \int\!\!\int \vec{B} \cdot d\vec{F} = \oint \vec{A} \cdot d\vec{s} \,. \tag{24}$$

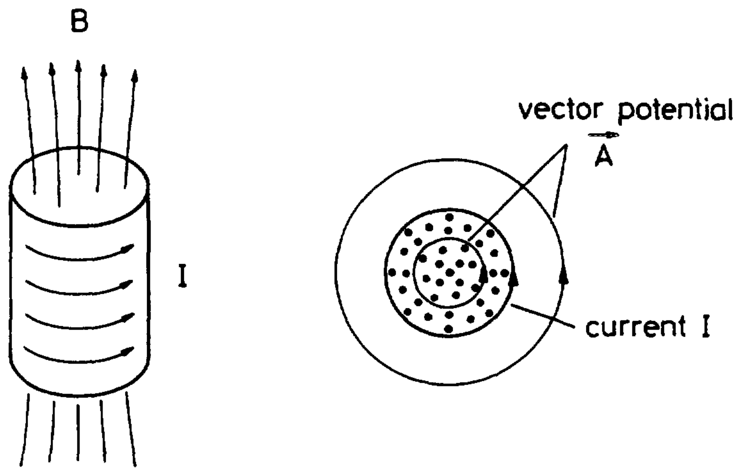


Figure 27: Magnetic field and vector potential of a solenoid.

We apply this to an interesting case namely the solenoidal coil sketched in Fig. 27.

The magnetic field has a constant value $B = B_0$ inside the solenoid and vanishes outside if we assume that the length of the coil is much larger than its radius R. The vector potential has only an azimuthal component and can be computed using Eq. (24):

$$A_\theta(r) = \begin{cases} \frac{1}{2}B_0 \cdot r & \text{for } r < R \\ \frac{1}{2}B_0\frac{R^2}{r} & \text{for } r > R \end{cases} .$$

Computing $\vec{B} = \vec{\nabla} \times \vec{A}$ in cylindrical coordinates just gives the expected result

$$B_z(r) = \begin{cases} B_0 & \text{for } r < R \\ 0 & \text{for } r > R \end{cases} .$$

What do we learn from this example?
(a) The vector potential is basically parallel to the current but perpendicular to the magnetic field.
(b) There are regions in space where the vector potential is non-zero while the magnetic field vanishes. Here it is the region $r > R$. A circular contour of radius $r > R$ definitely includes magnetic flux, namely $B_0\pi R^2$ for all $r > R$, so $\vec{A}$ must be non-zero, although $\vec{B} = 0$.

The vector potential is not uniquely defined. A new potential $\vec{A'} = \vec{A} + \vec{\nabla}\chi$ with an arbitrary function $\chi(x, y, z)$ leaves the magnetic field $\vec{B}$ invariant because the curl of a gradient vanishes identically. For this reason it is often said that the vector potential is just a useful mathematical quantity without physical significance of its own. In quantum theory this point of view is entirely wrong, the vector potential is of much deeper physical relevance than the magnetic field.

5.2 The vector potential in quantum theory

In quantum theory the vector potential is a quantity of fundamental importance:
(1) $\vec{A}$ is the wave function of the photons;
(2) in an electromagnetic field the wavelength of a charged particle is modified by the vector potential.
For the application in superconductivity we are interested in the second aspect. The *de Broglie relation* states that the wavelength we have to attribute to a particle is Planck's constant divided

by the particle momentum

$$\lambda = \frac{2\pi\hbar}{p} . \tag{25}$$

For a free particle one has to insert $p = mv$. It turns out that in the presence of an electromagnetic field this is no longer correct, instead one has to replace the mechanical momentum $m\vec{v}$ by the so-called 'canonical momentum'

$$\vec{p} = m\vec{v} + q\vec{A} \tag{26}$$

where q is the charge of the particle ($q = -e$ for an electron). The wavelength is then

$$\lambda = \frac{2\pi\hbar}{mv + qA} .$$

If one moves by a distance Δx, the phase φ of the electron wave function changes in free space by the amount

$$\Delta\varphi = \frac{2\pi}{\lambda}\Delta x = \frac{1}{\hbar}m\vec{v}\cdot\vec{\Delta x} .$$

In an electromagnetic field there is an additional phase change

$$\Delta\varphi' = -\frac{e}{\hbar}\vec{A}\cdot\vec{\Delta x} .$$

This is called the *Aharonov-Bohm effect* after the theoreticians who predicted the phenomenon [20]. The phase shift should be observable in a double-slit experiment as sketched in Fig. 28. An electron beam is split into two coherent sub-beams and a tiny solenoid coil is placed between these beams. The sub-beam 1 travels antiparallel to $\vec{A}$, beam 2 parallel to $\vec{A}$. So the two sub-beams gain a phase difference

$$\delta\varphi = \delta\varphi_0 + \frac{e}{\hbar}\oint\vec{A}\cdot d\vec{s} = \delta\varphi_0 + \frac{e}{\hbar}\Phi_{\mathrm{mag}} . \tag{27}$$

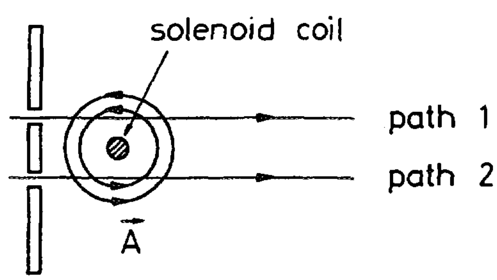


Figure 28: Schematic arrangement for observing the phase shift due to a vector potential.

Here $\delta\varphi_0$ is the phase difference for current 0 in the coil. The Aharonov-Bohm effect was verified in a beautiful experiment by Möllenstedt and Bayh in Tübingen [21]. The experimental setup and the result of the measurements are shown in Fig. 29. An electron beam is split by a metalized quartz fibre on negative potential which acts like an optical bi-prism. Two more fibres bring the two beams to interference on a photographic film. Very sharp interference fringes are observed. Between the sub-beams is a 14 μm–diameter coil wound from 4 μm thick tungsten wire. The current in this coil is first zero, then increased linearly with time and after that kept constant. The film to register the interference pattern is moved in the vertical direction.

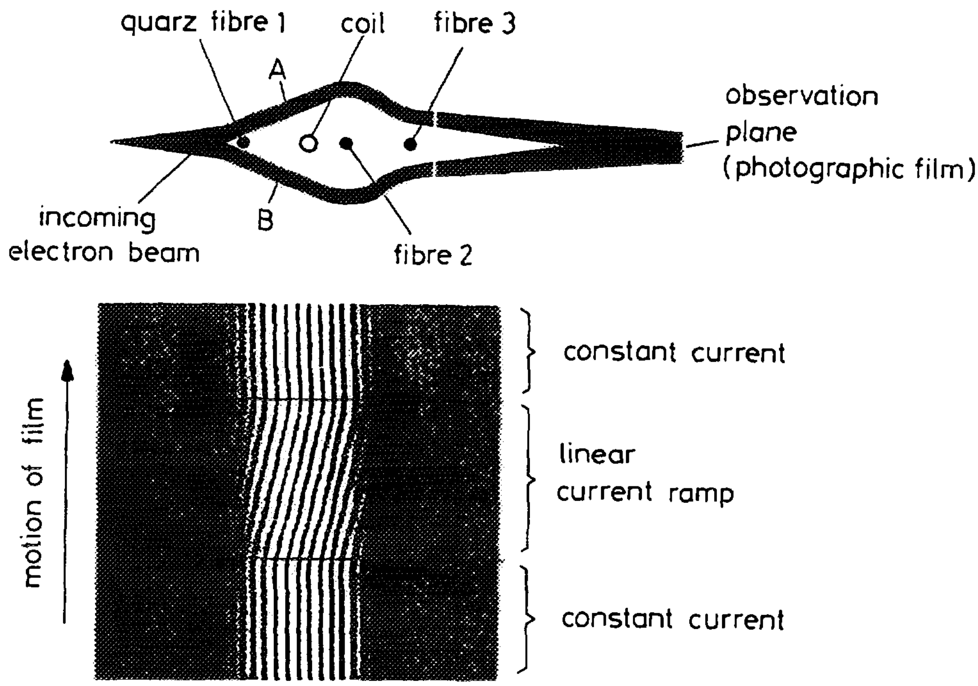


Figure 29: Sketch of the Möllenstedt-Bayh experiment and observed interference pattern.

Thereby the moving fringes are depicted as inclined lines. The observed shifts are in quantitative agreement with the prediction of Eq. (27).

An interesting special case is the phase shift $\delta\varphi = \pi$ that interchanges bright and dark fringes. According to Eq. (27) this requires a magnetic flux

$$\Phi_{\mathrm{mag}} = \pi\frac{\hbar}{e} = \frac{h}{2e}$$

which turns out to be identical to the elementary flux quantum in superconductors, see Sec. 6. In the Möllenstedt experiment however, continuous phase shifts much smaller than π are visible, so the magnetic flux through the normal-conducting tungsten coil is not quantised (there is also no theoretical reason for flux quantisation in normal conductors).
Although the magnetic field is very small outside the solenoid, and the observed phase shifts are in quantitative agreement with the expectation based on the vector potential, there have nevertheless been sceptics who tried to attribute the observed effects to some stray magnetic field. To exclude any such explanation a new version of the experiment has recently been carried out by Tonomura et al. [22] making use of electron holography (Fig. 30). A parallel electron beam is imaged by an electron microscope lens on a photographic plate. To create a holographic pattern the object is placed in the upper half of the beam while the lower half serves as a reference beam. A metalized quartz fibre (the bi-prism) brings the two-part beam to an overlap on the plate. The magnetic field is provided by a permanently magnetised ring of with a few μm diameter. The magnet is enclosed in niobium and cooled by liquid helium so the magnetic field is totally confined. The vector potential, however, is not shielded by the superconductor. The field lines of $\vec{B}$ and $\vec{A}$ are also drawn in the figure. The holographic image shows again a very clear interference pattern and a shift of the dark line in the opening of the ring which is caused by the vector potential. This experiment demonstrates beyond any doubt that it is the vector potential and not the magnetic field which influences the wavelength of the electron and the interference pattern.

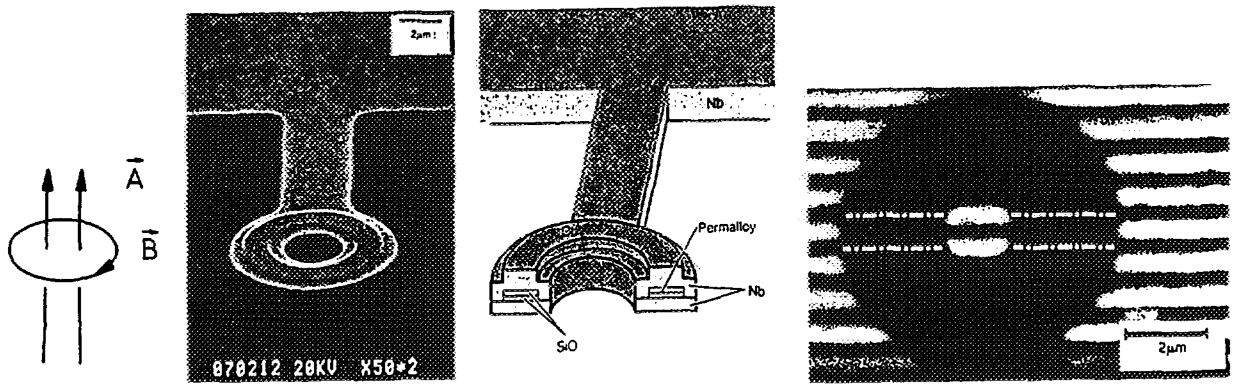


Figure 30: Observation of Aharonov-Bohm effect using electron holography (after Tonomura [22]). The permanent toroidal magnet, encapsuled in superconducting niobium, and the observed interference fringes are shown.

6 QUANTISATION OF MAGNETIC FLUX

The Meissner-Ochsenfeld effect excludes magnetic field from the bulk of a type I superconductor. An interesting situation arises if one exposes a superconducting ring to a magnetic field. Then one can obtain a trapped flux, threading the hole of the ring as shown in Fig. 31. Both the London and the BCS theory make the surprising prediction that the flux through the hole cannot assume arbitrary values but is quantised, i.e. that it is an integer multiple of an elementary flux quantum

$$\Phi_{\mathrm{mag}} = n\,\Phi_0\,, \quad n = 0, 1, 2, \ldots \tag{28}$$

The flux quantum is Planck's constant divided by the charge of the supercurrent carriers. The BCS flux quantum is thus

$$\Phi_0 = \frac{h}{2e} = \frac{\pi\hbar}{e} \tag{29}$$

while the London flux quantum is twice as big because the charge carriers in the London theory are single electrons.

6.1 Derivation of flux quantisation

The Cooper-pair wave function in the ring can be written as

$$\Psi = \sqrt{n_c}\exp(i\varphi)\,.$$

The density of Cooper pairs is denoted as n_c. The phase $\varphi = \varphi(s)$ has to change by $n \cdot 2\pi$ when going once around the ring since Ψ must be a single-valued wave function. We choose a circular path in the bulk of the ring (Fig. 31c). Then

$$\oint \frac{d\varphi}{ds} ds = n \cdot 2\pi\,.$$

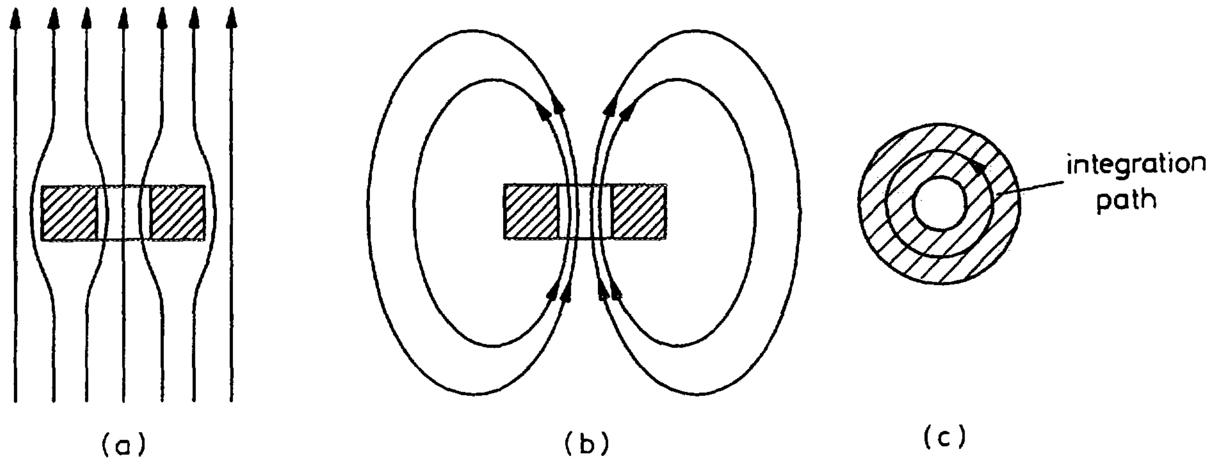


Figure 31: Trapping of magnetic flux in a ring. First the normal-conducting ring ($T > T_c$) is placed in a magnetic field, then it is cooled down (a) and finally the field is switched off (b). The integration path is shown in part (c).

In other words: the circumference must be an integer number of wavelengths. In the bulk there is no current allowed so the Cooper-pair velocity must be zero, $\vec{v} = 0$. Therefore the integrand is

$$\frac{d\varphi}{ds}ds = \frac{q}{\hbar}\vec{A}\cdot\vec{ds}\,.$$

Using Eq. (24) we see that the magnetic flux enclosed by the circular path is

$$\Phi_{mag} = \oint \vec{A}\cdot\vec{ds} = \frac{\hbar}{q}\cdot n\cdot 2\pi = n\cdot\Phi_0 \quad \Rightarrow \quad \Phi_0 = \frac{2\pi\hbar}{q}\,.$$

In the BCS theory we obtain

$$\Phi_0 = \frac{h}{2e}\,.$$

6.2 Experimental verification of flux quantisation

In 1961 two experiments on flux quantisation were carried out almost simultaneously, by Doll and Näbauer [23] in München and by Deaver and Fairbank [24] in Stanford. I describe the Doll-Näbauer experiment as it yielded the best evidence. The setup and the results are shown in Fig. 32. The superconducting ring is here a lead tube prepared by evaporation of lead on a 10 μm–thick quartz cylinder which is then suspended by a torsion fibre. Magnetic flux is captured in the tube by exposing the warm tube to a 'magnetising field' B_{mag} parallel to the axis, cooling down and switching off the field. Then a transverse oscillating field B_{osc} is applied to induce forced oscillations which are observed by light reflection from a small mirror. The resonant amplitude A_{res} is proportional to the magnetic moment of the tube, i.e. to the captured magnetic flux. Without flux quantisation the relation between resonant amplitude and magnetising field should be linear. Instead one observes a very pronounced stair-case structure which can be uniquely related to frozen-in fluxes of 0, 1 or 2 flux quanta. Both experiments proved that the magnetic flux quantum is $h/2e$ and not h/e and thus gave strong support for the Cooper-pair hypothesis.

There has been a long debate on whether the BCS theory applies also for high-T_c superconductors[2].

[2]For an excellent recent review of high-T_c superconductors see H. Rietschel [25].

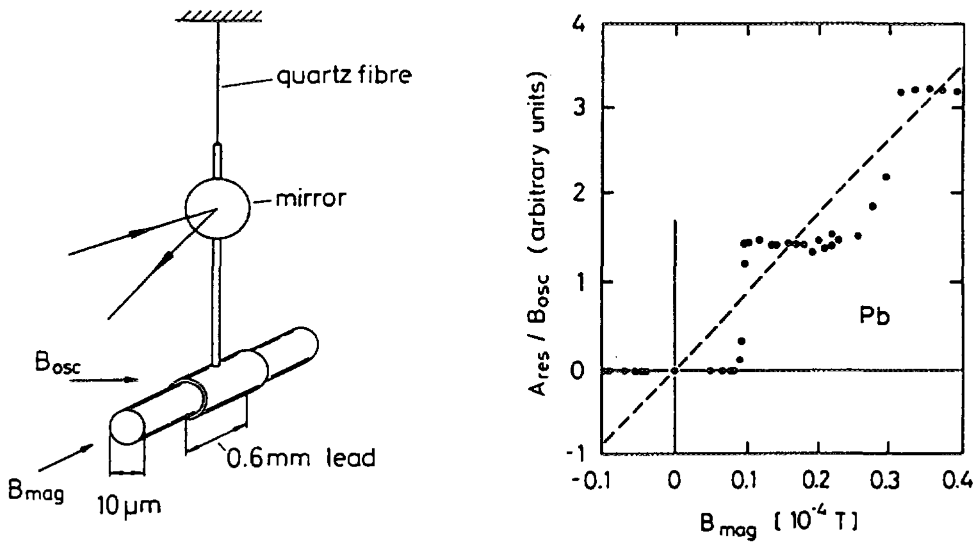


Figure 32: Observation of flux quantisation [23].

In Fig. 33 the flux through a YBaCuO ring with a weak joint is shown. Flux jumps due to external field variations occur in multiples of $h/2e$ which is a strong indication for Cooper pairing. Another interesting experiment has been performed using the setup sketched in Fig. 34. A

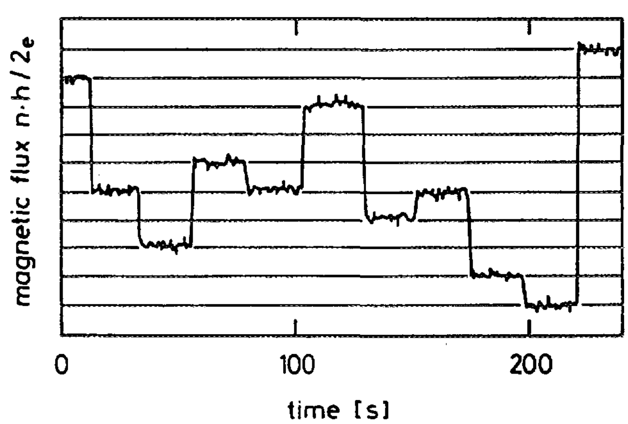


Figure 33: Flux through an $YBa_2Cu_3O_7$ ring with a weak link [26].

lead strip is bent into ring shape and closed via an intermediate $YBa_2Cu_3O_7$ piece. It has been possible to induce a permanent ring current in this combined system of a low-T_c and a high-T_c superconductor. This is additional evidence that the same charge carriers, namely Cooper pairs, are present in both superconductors.

6.3 Fluxoid pattern in type II superconductors

Abrikosov predicted that a magnetic field penetrates a type II superconductor in the form of flux tubes or fluxoids, each containing a single elementary quantum Φ_0, which arrange themselves in a triangular pattern to minimize the potential energy related to the mutual repulsion of the flux tubes. A schematic cross section of a fluxoid is presented in Fig. 35. The magnetic field lines are surrounded by a super-current vortex. The Cooper-pair density drops to zero at the centre of the vortex, so the core of a flux tube is normal-conducting.

The area occupied by a fluxoid is roughly given by $\pi\xi^2$ where ξ is the coherence length. An

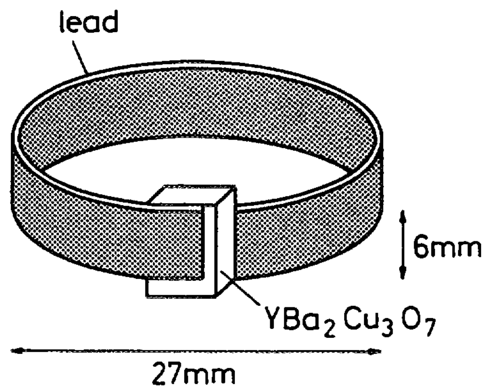


Figure 34: A lead ring with a transition piece made from $YBa_2Cu_3O_7$ [27].

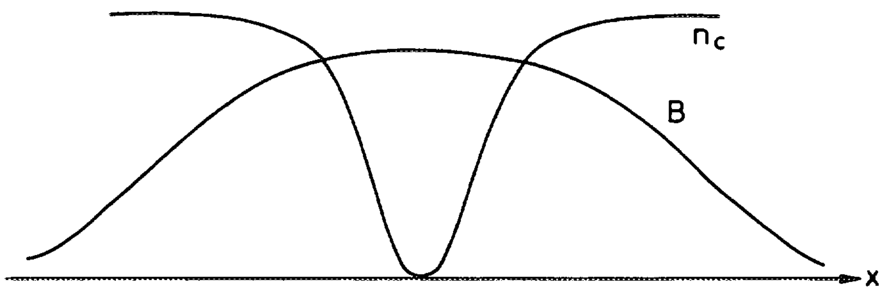


Figure 35: Schematic cross section of a fluxoid.

estimate of the upper critical field is derived from the condition that the fluxoids start touching each other:

$$B_{c2} \approx \frac{\Phi_0}{2\pi\xi^2} . \tag{30}$$

An important experimental step was the direct observation of the fluxoid pattern. Essmann and Träuble [12] developed a 'decoration' technique for this purpose. A sample of lead-indium (Pb 6% In) was cooled by liquid helium of 1.2 K. The liquid did not cover the surface of the sample. Iron was evaporated at some distance from the superconductor and in the 0.8 mbar helium atmosphere the iron atoms agglomerated to tiny crystals (about 20 nm) that were attracted by the magnetic field lines and stuck to the sample surface where the fluxoids emerged. After warming up, a thin film was sprayed on the surface to allow the iron crystals to be removed for subsequent observation in an electron microscope. The photograph in Fig. 36 shows indeed the perfect triangular pattern predicted by Abrikosov. Similar pictures have been recently obtained with high-temperature superconductors.

The electron holography setup mentioned in the last section permits direct visualization of the magnetic flux lines. Figure 37 is an impressive example of the capabilities of this advanced method.

7 HARD SUPERCONDUCTORS

7.1 Flux flow resistance and flux pinning

For application in accelerator magnets a superconducting wire must be able to carry a large current in the presence of a field in the 5 – 10 Tesla range. Type I superconductors are

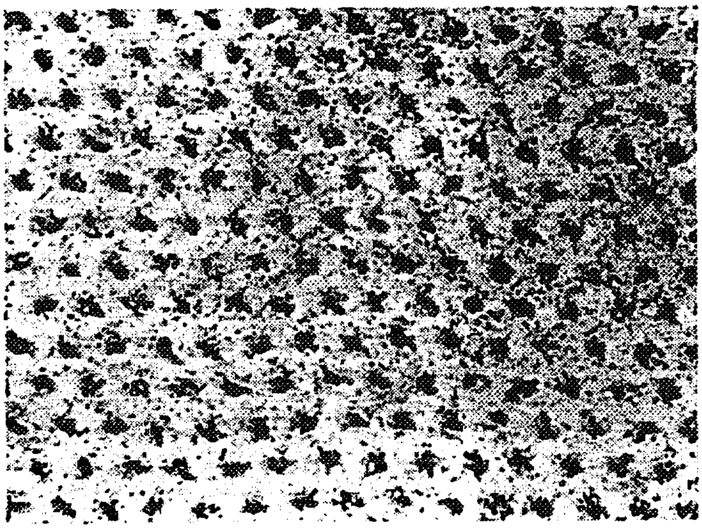


Figure 36: Observation of the fluxoid pattern [12].

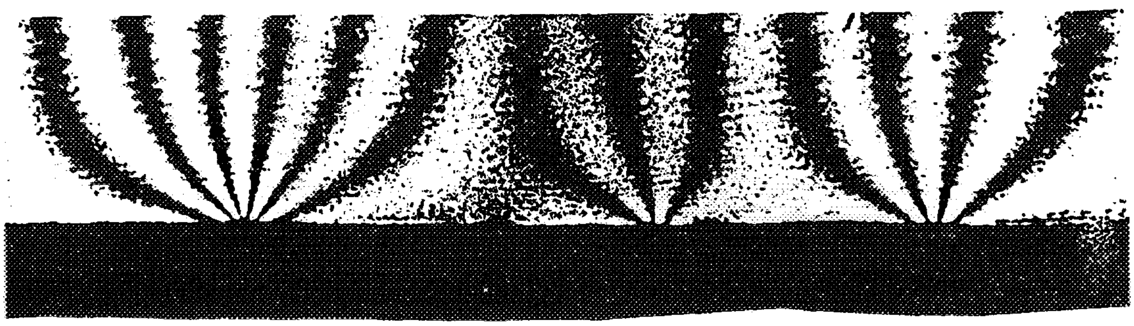


Figure 37: Holographic image of the magnetic flux lines through a thin lead plate [22].

definitely ruled out because their critical field is less than a few tenths of a Tesla and their current-carrying capacity is very small since the current is restricted to a thin surface layer (compare Fig. 6). Type II conductors appear quite appropriate on first sight: they have a large upper critical field and high currents are permitted to flow in the bulk material. Still there is a problem, it is called *flux flow resistance.* If a current flows through an ideal type II superconductor which is exposed to a magnetic field one observes heat generation. Figure 38 illustrates the mechanism. The current density $\vec{J}$ exerts a Lorentz force on the flux lines. The force density (force per unit volume) is

$$\vec{F} = \vec{J} \times \vec{B} \, .$$

The flux lines begin to move through the specimen in a direction perpendicular to the current and to the field. This is a viscous motion ($\vec{v} \propto \vec{F}$) and leads to heat generation. So although the current itself flows without dissipation the sample acts as if it had an Ohmic resistance. The statement is even formally correct. The moving fluxoids represent a moving magnetic field. According to Special Relativity this is equivalent to an electric field

$$\vec{E}_{\text{equiv}} = \frac{1}{c^2} \vec{B} \times \vec{v} \, .$$

It is easy to see that $\vec{E}_{\text{equiv}}$ and $\vec{J}$ point in the same direction just like in a normal resistor. Flux flow resistance was studied experimentally by Kim and co-workers [28].

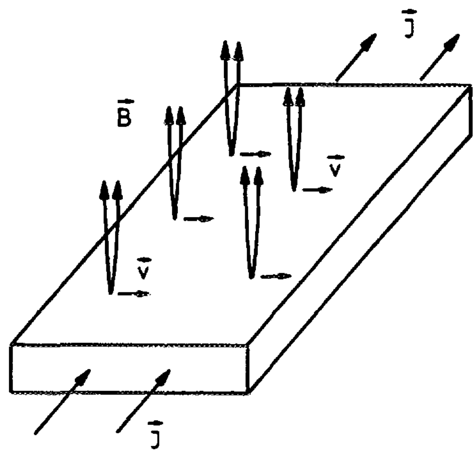


Figure 38: Fluxoid motion in a current-carrying type II superconductor.

To obtain useful wires for magnet coils the flux motion has to be inhibited. The standard method is to capture them at *pinning centres*. The most important pinning centres in niobium-titanium are normal-conducting Ti precipitates with a size in the 10 nm range. Flux pinning is discussed in detail in M.N. Wilson's lectures at this school.
A type II superconductor with strong pinning is called a *hard superconductor*. Hard superconductors are very well suited for high-field magnets, they permit dissipationless current flow in high magnetic fields. There is a penalty, however: these conductors exhibit a strong magnetic hysteresis which is the origin of the very annoying 'persistent-current' multipoles in superconducting accelerator magnets.

7.2 Magnetisation of a hard superconductor

A type I superconductor shows a completely reversible response to an external magnetic field B_e. The magnetisation M is a unique function of the field, namely the straight line $M(B_e) = -B_e/\mu_0$ for $0 < B_e < B_c$, see Fig. 10. An ideal type II conductor without any flux pinning should also react reversibly. A hard superconductor is only reversible when it is in the Meissner state because then no magnetic field enters the bulk, so no flux pinning can occur. If the field is raised beyond B_{c1} magnetic flux enters the sample and is captured at pinning centres. When the field is reduced again these flux tubes remain bound and the specimen keeps a frozen-in magnetisation even for vanishing external field. One has to invert the field polarity to achieve $M = 0$ but the starting point ($B_e = 0$, $M = 0$) can only be reached by warming up the specimen to destroy superconductivity and release any captured flux, and by cooling down again. A typical hysteresis curve is shown in Fig. 39a. There is a close resemblence with the hysteresis in iron except for the sign: the magnetisation in a superconductor is opposed to the magnetising field because the physical mechanism is diamagnetism. In an accelerator the field is usually not inverted and then the hysteresis has the shape plotted in Fig. 39b.
Detailed studies on superconductor magnetisation were performed in the HERA dipoles. The sextupole component is a good measure of M. Immediately after cooldown a dipole was excited to low fields. In Fig. 40 the sextupole field B_3 at a distance of 25 mm from the dipole axis is plotted as a function of the dipole field B_1 on the axis. One can see that the sextupole is a reversible function of B_1 up to about 25 mT (the lower critical field of NbTi is somewhat smaller, around 15 mT, but in most parts of the coil the local field is less than the value B_1 on the axis). The superconducting cable is therefore in the Meissner phase. Increasing B_1 to 50 mT already leads to a slight hysteresis so a certain amount of magnetic flux enters the NbTi filaments and is captured there. With increasing field the hysteresis widens more and more and

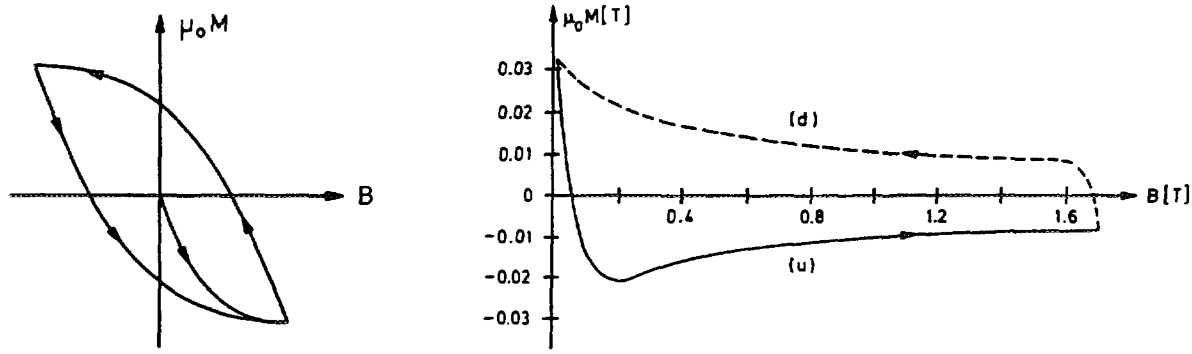


Figure 39: (a) Magnetic hysteresis of a hard superconductor. (b) Magnetisation hysteresis for the field cycle of accelerator magnets.

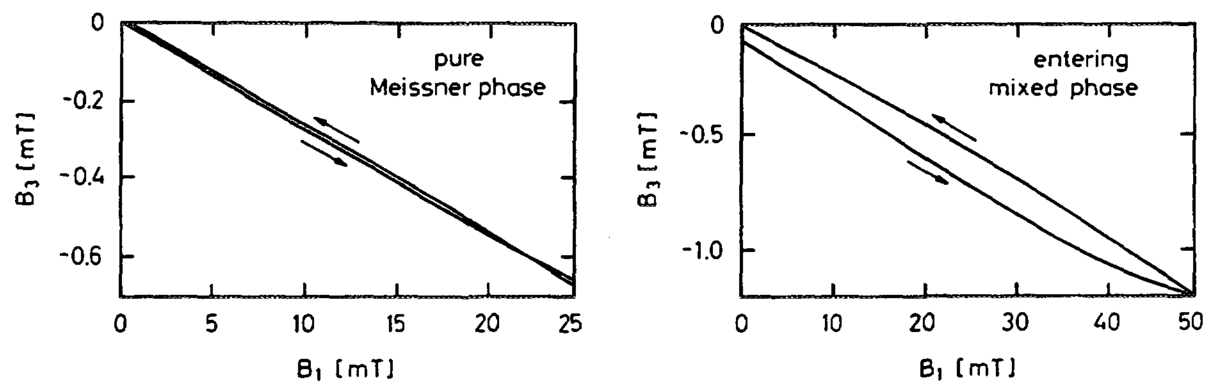


Figure 40: Sextupole field in a HERA dipole in the Meissner phase and slightly above.

is eventually nearly symmetric to the horizontal axis. The sextupole hysteresis observed in the standard field cycle at HERA is plotted in Fig. 41a. A similar curve is obtained for the 12-pole in the quadrupoles.

Only in a 'virgin' magnet, that is right after cool-down, is there the chance to influence the width of the hysteresis curve. This fact was used to advantage during the commissioning of the HERA proton storage ring. The first beam test was made with positrons of only 7 GeV since the nominal 40 GeV protons were not yet available. At the corresponding dipole field of 70 mT (coil current 42.5 A) the persistent-current sextupole component would have been two orders of magnitude larger than tolerable if the standard field cycle had been used. To eliminate the sextupole, all magnets were warmed to 20 K to extinguish any previous superconductor magnetisation and cooled back to 4.4 K. Then the current loop $0 \rightarrow 112$ A $\rightarrow 42.5$ A was performed which resulted in an almost vanishing sextupole (see Fig. 41b). A similar procedure was used in the first run with 40 GeV protons, this time with the loop $0 \rightarrow 314$ A $\rightarrow 244.5$ A. The measured chromaticity indeed proved an almost perfect sextupole cancellation. For the routine operation of HERA these procedures are of course not applicable because they require a warm-up of the whole ring. Instead, sextupole correction coils must be used to compensate the unwanted field distortions.

7.3 Flux creep

The pinning centres prevent flux flow in hard superconductors but some small *flux creep* effects remain. At finite temperatures, even as low as 4 K, a few of the flux quanta may be released from their pinning locations by thermal energy and then move out of the specimen

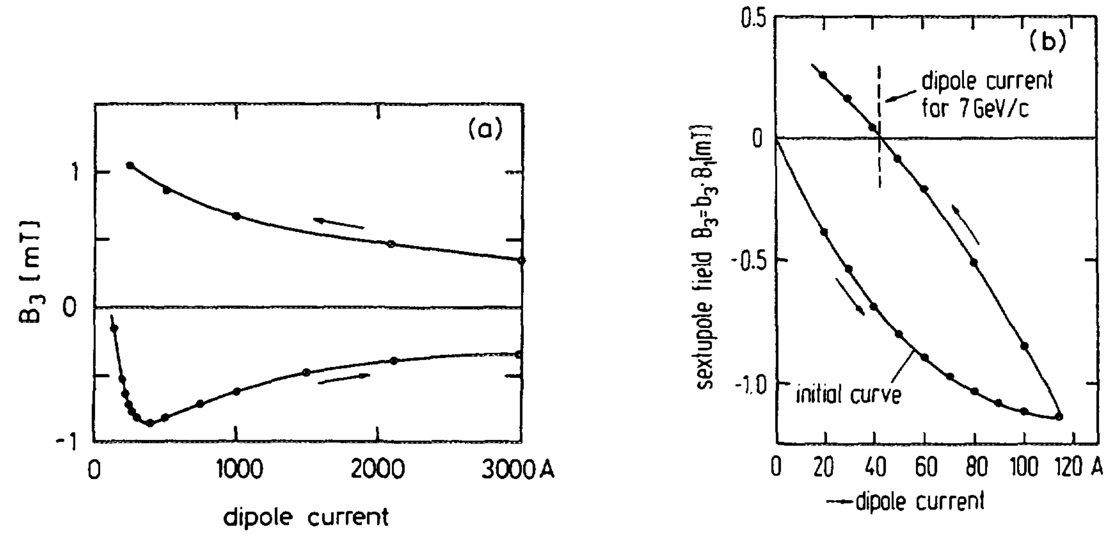


Figure 41: (a) The sextupole component in the HERA dipoles for the standard field cycle 4.7 T → 0.05 T → 4.7 T.
(b) Sextupole field for the first beam test with positrons of 7 GeV.

thereby reducing the magnetisation. The first flux creep experiment was carried out by Kim et al. [29] using a small NbZr tube. If one plots the internal field at the centre of the tube as a function of the external field the well-known hysteresis curve is obtained in which one can distinguish the shielding and the trapping branch, see Fig. 42a. Kim and co-workers realized that on the trapping branch the internal field exhibited a slow logarithmic decrease with time while on the shielding branch a similar increase was seen (Fig. 42b). Both observations can be explained by assuming a $\log t$ decay of the critical current density.

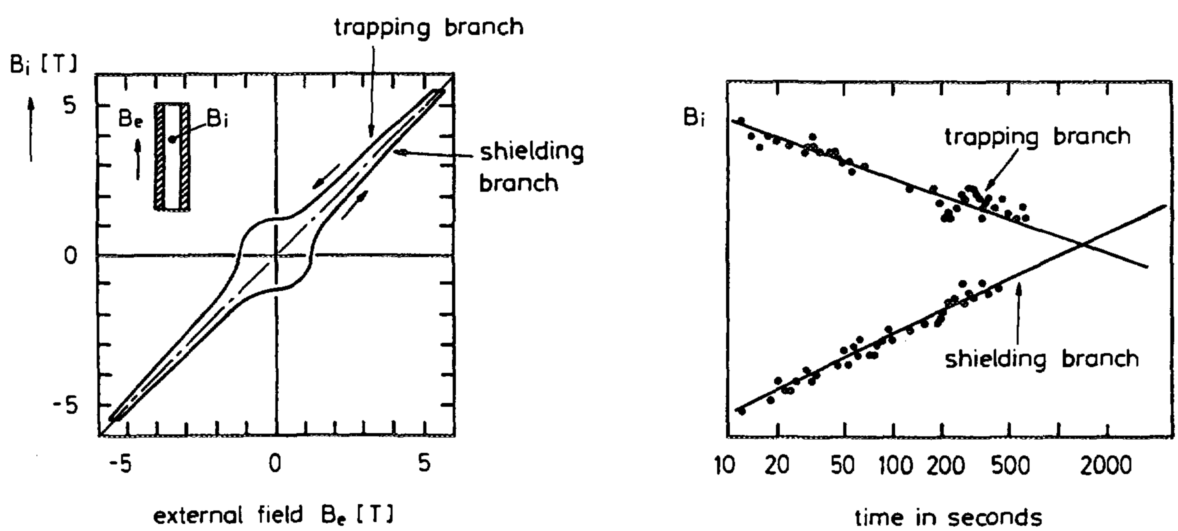


Figure 42: (a) Hysteresis of the internal field in a tube of hard superconductor. (b) Time dependence of the internal field on the trapping and the shielding branch [29].

A logarithmic time dependence is something rather unusual. In an electrical circuit with inductive and resistive components the current decays exponentially like $\exp(-t/\tau)$ with a time constant $\tau = L/R$. A theoretical model for thermally activated flux creep was proposed by Anderson [30]. The pinning centres are represented by potential wells of average depth U_0 and width a in which bundles of flux quanta with an average flux $n\Phi_0$ are captured. At zero current the probability that flux leaves a potential well is proportional to the Boltzmann factor

$$P_0 \propto \exp(-U_0/k_B T)\,.$$

When the superconductor carries a current density J the potential acquires a slope proportional to the driving force density $F \propto n\Phi_0 J$. This slope reduces the effective potential well depth to $U = U_0 - \Delta U$with $\Delta U \approx n\Phi_0 J a\, l$, see Fig. 43. Here l is the length of the flux bundle. The probability for flux escape increases

$$P = P_0 \exp(+\Delta U/k_B T)\,.$$

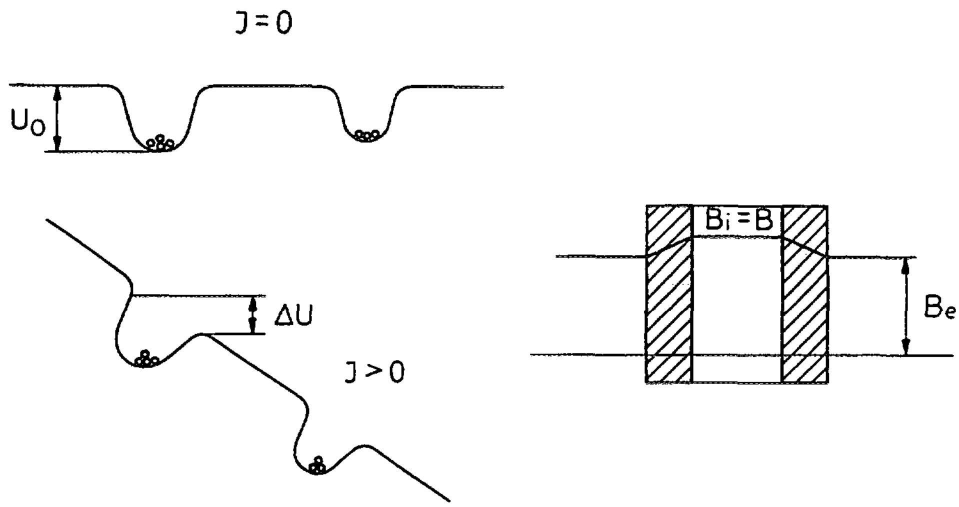


Figure 43: Sketch of the pinning potential without and with current flow and field profile across the NbZr tube.

We consider now the tube in the Kim experiment at a high external field B_{ext} on the trapping branch of the hysteresis curve. The internal field is then slightly larger, namely by the amount $B_{\text{int}} - B_{\text{ext}} = \mu_0 J_c\, w$ where J_c is the critical current density at the given temperature and magnetic field and w the wall thickness. Under the assumption $B_{\text{int}} - B_{\text{ext}} \ll B_{\text{ext}}$ both field and current density are almost constant throughout the wall. The reduction in well depth ΔU is proportional to the product of these quantities. If a bundle of flux quanta is released from its well, it will 'slide' down the slope and leave the material. In this way space is created for some magnetic flux from the bore of the cylinder which will migrate into the conductor and refill the well. As a consequence the internal field decreases and with it the critical current density in the wall. Its time derivative is roughly given by the expression

$$\frac{dJ_c}{dt} \approx -C \exp\left(\frac{\Delta U}{k_B T}\right) \approx -C \exp\left(\frac{n\Phi_0 a J_c l}{k_B T}\right) \tag{31}$$

where C is a constant. The solution of this unusual differential equation is

$$J_c(t) = J_c(0) - \frac{k_B T}{n\Phi_0 a\, l} \ln t\,. \tag{32}$$

This result implies that for given temperature and magnetic field the critical current density is not really a constant but depends slightly on time. What one usually quotes as J_c is the value obtained after the decay rate on a linear time scale has become unmeasurably small.
A nearly logarithmic time dependence is also observed in the persistent-current multipole fields of accelerator magnets, see the lecture by A. Devred at this school. So it seems tempting to

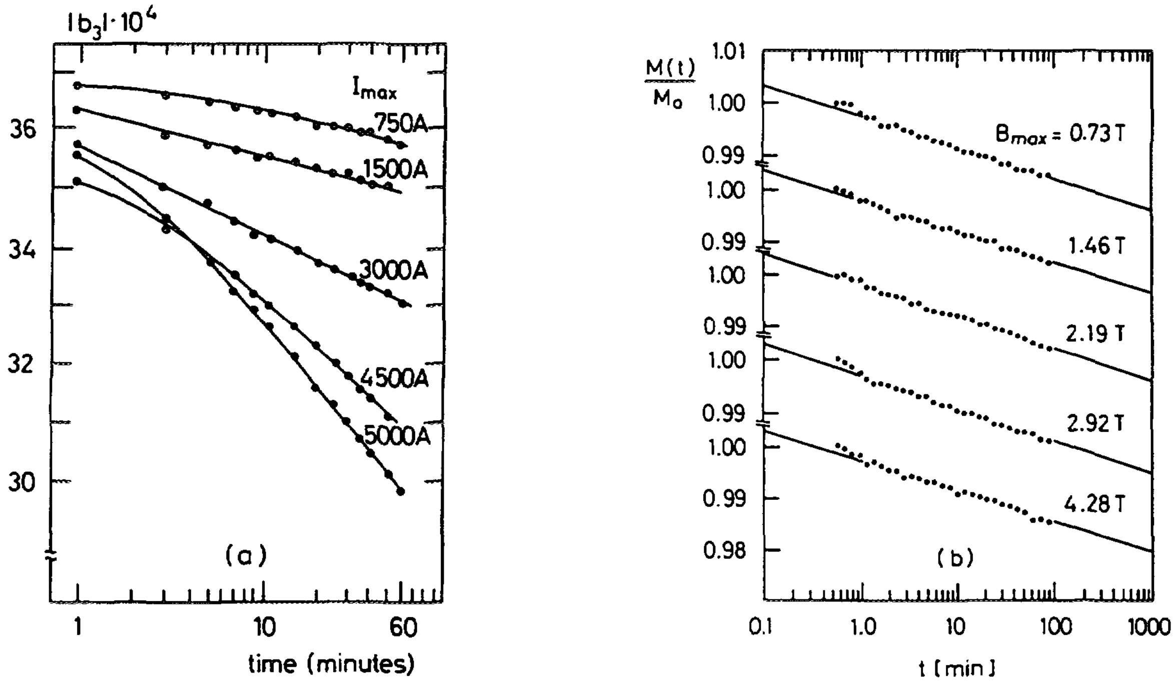


Figure 44: (a) Decay of sextupole coefficient in a HERA dipole at 0.23 Tesla for different currents in the initialising cycle ($0 \rightarrow I_{\text{max}} \rightarrow 50$ A$\rightarrow 250$ A) [31]. (b) Time dependence of the magnetisation of a superconductor sample for different fields B_{max} in the initializing cycle [32].

attribute the effect to flux creep. Surprisingly, the decay rates are generally much larger than typical flux creep rates and depend moreover on the maximum field level in a preceding excitation, see Fig. 44a. In cable samples this is not the case as is evident from Fig. 44b. The average magnetisation of a 5 m-long cable sample decays at low field ($B = 0$ in this case) by less than 1% per decade of time, and the decay rate is totally independent of the maximum field B_{max} in the preceding cycle. The observed rate agrees well with other data on flux creep in NbTi.

Obviously, superconductor properties alone cannot account for the time dependence of persistent-current fields in magnets. The field-generating transport current and position-dependent eddy current effects play a role in magnets. There may a redistribution of current among the strands in a cable which leads to reduction of superconductor magnetisation. Attempts to construct a quantitative model have failed so far.
Flux creep has become an important issue after the discovery of high-temperature superconductors. Figure 45 shows that the magnetisation of YBaCuO samples decays rapidly, in particular for single crystals. One speaks of 'giant flux creep'. This is a strong hint that flux pinning is insufficient at 77 K and implies that the presently available materials are not yet suited to building magnets cooled by liquid nitrogen.

8 JOSEPHSON EFFECTS

In 1962 B.D. Josephson made a theoretical analysis of the tunneling of Cooper pairs through a thin insulating layer from one superconductor to another and predicted two fascinating phenomena which were fully confirmed by experiment. A schematic experimental arrangement is shown in Fig. 46.
(1) DC Josephson effect. If the voltage V_0 across the junction is zero there is a dc Cooper-

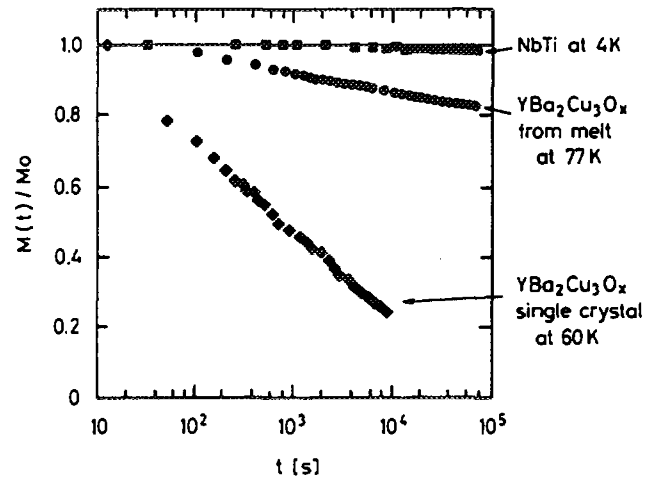


Figure 45: Comparison of superconductor magnetisation decay due to flux creep in NbTi at a temperature of 4.2 K, in oriented-grained $YBa_2Cu_3O_x$ at 77 K and in a $YBa_2Cu_3O_x$ single crystal at 60 K [33].

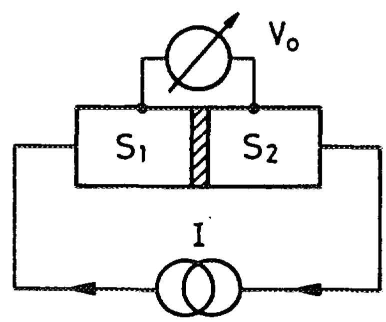


Figure 46: Schematic arrangement for studying the properties of a Josephson junction.

pair current which can assume any value in the range

$$-I_0 < I < I_0$$

where I_0 is a maximum current that depends on the Cooper-pair densities and the area of the junction.

(2) AC Josephson effect. Increasing the voltage of the power supply eventually leads to a non-vanishing voltage across the junction and then a new phenomenon arises: besides a dc current which however is now carried by single electrons there is an alternating Cooper-pair current

$$I(t) = I_0 \sin(2\pi f_J t + \varphi_0) \tag{33}$$

whose frequency, the so-called *Josephson frequency*, is given by the expression

$$f_J = \frac{2eV_0}{2\pi\hbar} \, . \tag{34}$$

For a voltage $V_0 = 1\,\mu$V one obtains a frequency of 483.6 MHz. The quantity φ_0 is an arbitrary phase. Equation (34) is the basis of extremely precise voltage measurements.

8.1 Schrödinger equation of the Josephson junction

The wave functions in the superconductors 1 and 2 are called ψ_1 and ψ_2. Due to the possibility of tunneling through the barriere the two Schrödinger equations are coupled

$$i\hbar\frac{\partial\psi_1}{\partial t} = E_1\psi_1 + K\psi_2 \, , \qquad i\hbar\frac{\partial\psi_2}{\partial t} = E_2\psi_2 + K\psi_1 \, . \tag{35}$$

The quantity K is the coupling parameter. The macroscopic wave functions can be expressed through the Cooper-pair densities n_1, n_2 and the phase factors

$$\psi_1 = \sqrt{n_1}\exp(i\varphi_1)\,, \quad \psi_2 = \sqrt{n_2}\exp(i\varphi_2)\,. \tag{36}$$

We insert this into (35) and obtain

$$\left(\frac{\dot{n}_1}{2\sqrt{n_1}} + i\sqrt{n_1}\dot{\varphi}_1\right)\exp(i\varphi_1) = -\frac{i}{\hbar}[E_1\sqrt{n_1}\exp(i\varphi_1) + K\sqrt{n_2}\exp(i\varphi_2)]$$

and

$$\left(\frac{\dot{n}_2}{2\sqrt{n_2}} + i\sqrt{n_2}\dot{\varphi}_2\right)\exp(i\varphi_2) = -\frac{i}{\hbar}[E_2\sqrt{n_2}\exp(i\varphi_2) + K\sqrt{n_1}\exp(i\varphi_1)]\,.$$

Now we multiply these equation with $\exp(-i\varphi_1)$ resp. $\exp(-i\varphi_2)$ and separate the real and imaginary parts:

$$\begin{aligned}
\dot{n}_1 &= \frac{2K}{\hbar}\sqrt{n_1 n_2}\,\sin(\varphi_2 - \varphi_1)\,, \\
\dot{n}_2 &= \frac{2K}{\hbar}\sqrt{n_1 n_2}\,\sin(\varphi_1 - \varphi_2) = -\dot{n}_1\,, \\
\dot{\varphi}_1 &= -\frac{1}{\hbar}\left[E_1 + K\sqrt{\frac{n_2}{n_1}}\cos(\varphi_2 - \varphi_1)\right]\,, \\
\dot{\varphi}_2 &= -\frac{1}{\hbar}\left[E_2 + K\sqrt{\frac{n_1}{n_2}}\cos(\varphi_1 - \varphi_2)\right]\,.
\end{aligned} \tag{37}$$

For simplicity we consider the case where the two superconductors are identical, so $n_2 = n_1$. The Cooper-pair energies E_1 and E_2 differ by the energy gained upon crossing the voltage V_0:

$$E_2 = E_1 - 2eV_0\,.$$

The equations simplify

$$\begin{aligned}
\dot{n}_1 &= \frac{2K}{\hbar}n_1\sin(\varphi_2 - \varphi_1) = -\dot{n}_2\,, \\
\frac{d}{dt}(\varphi_2 - \varphi_1) &= -\frac{1}{\hbar}(E_2 - E_1) = \frac{2eV_0}{\hbar}\,.
\end{aligned} \tag{38}$$

Integrating the second equation (38) yields the Josephson frequency

$$\varphi_2(t) - \varphi_1(t) = \frac{2eV_0}{\hbar}\cdot t + \varphi_0 = 2\pi f_J \cdot t + \varphi_0\,. \tag{39}$$

The Cooper-pair current through the junction is proportional to $\dot{n}_1$. Using (38) and (39) it can be written as

$$I(t) = I_0\sin\left(\frac{2eV_0}{\hbar}t + \varphi_0\right)\,. \tag{40}$$

There are two cases:
(1) For zero voltage across the junction we get a dc current

$$I = I_0\sin\varphi_0$$

which can assume any value between $-I_0$ and $+I_0$ since the phase φ_0 is not specified.
(2) For $V_0 \neq 0$ there is an ac Cooper–pair current with exactly the Josephson frequency.

8.2 Superconducting quantum interference

A loop with two Josephson junctions in parallel (Fig. 47) exhibits interference phenomena that are similar to the optical diffraction pattern of a double slit. Assuming zero voltage across the junctions the total current is

$$I = I_a + I_b = I_0(\sin\varphi_a + \sin\varphi_b)\,.$$

When a magnetic flux Φ_{mag} threads the area of the loop, the phases differ according to Sec. 5 by

$$\varphi_b - \varphi_a = \frac{2e}{\hbar}\oint \vec{A}\cdot d\vec{s} = \frac{2e}{\hbar}\Phi_{mag}\,.$$

With $\varphi_0 = (\varphi_a + \varphi_b)/2$ we get

$$\varphi_a = \varphi_0 + \frac{e}{\hbar}\Phi_{\mathrm{mag}}, \quad \varphi_b = \varphi_0 - \frac{e}{\hbar}\Phi_{\mathrm{mag}}$$

and the current is

$$I = I_0 \sin\varphi_0 \cos\left(\frac{e}{\hbar}\Phi_{\mathrm{mag}}\right)\,. \tag{41}$$

As a function of the magnetic flux one obtains a typical double-slit interference pattern as shown in Fig. 47. Adjacent peaks are separated by one flux quantum $\Delta\Phi_{\mathrm{mag}} = \Phi_0$, so by counting flux quanta one can measure very small magnetic fields. This is the basic principle of the Superconducting Quantum Interference Device (SQUID). Technically one often uses superconducting rings with a single weak link which acts as a Josephson junction. Flux transformers are applied to increase the effective area and improve the sensitivity.

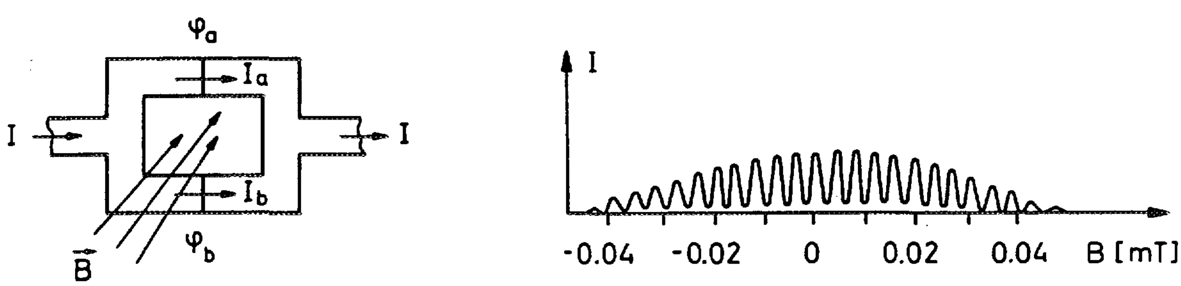


Figure 47: A loop with two Josephson junctions and the observed interference pattern [34]. The amplitude modulation is caused by the finite width of the junctions.

It is possible to capture fluxoids in superconductor tunnel contacts. The surrounding vector potential then leads to a spatial variation of the Josephson current. Huebener et al. [35] have investigated this in a scanning electron microscope. The heating due to the electron beam reduces the maximum Josephson current locally. Figure 48 shows the setup and the scanned current pattern for five captured fluxoids whose axes lie in the tunnel layer. Two-dimensional images of captured flux quanta can also be produced by this method.

I want to thank S. Turner for a critical reading of the manuscript and valuable comments.

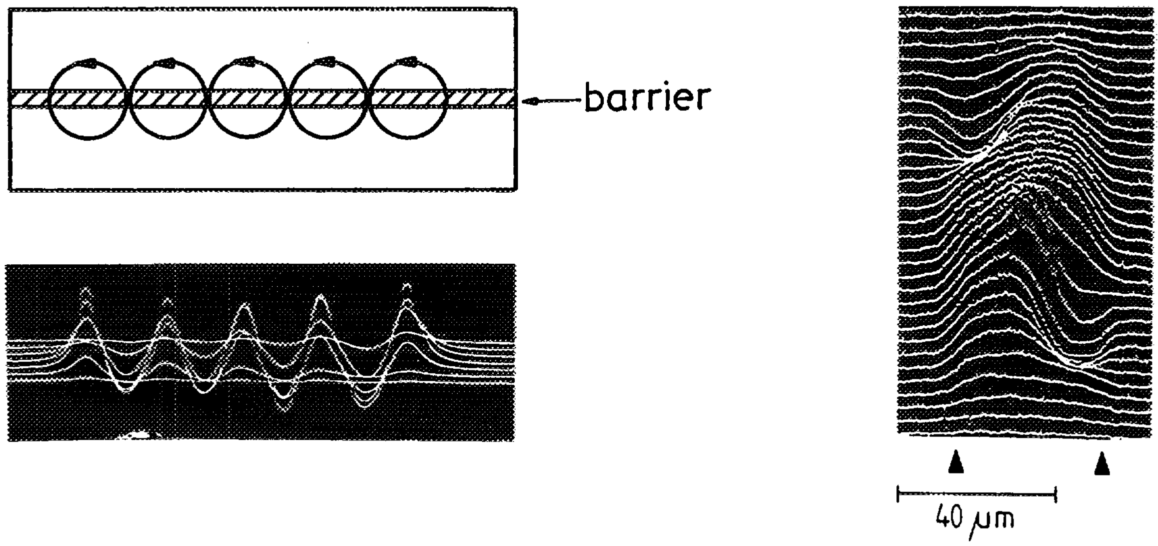


Figure 48: Observation of captured flux quanta in Josephson junctions using a scanning electron microscope [35].

REFERENCES

[1] W. Buckel, Supraleitung, 4. Auflage, VCH Verlagsgesellschaft, Weinheim 1990
[2] D.R. Tilley & J. Tilley, Superfluidity and Superconductivity, Third Edition, Institute of Physics Publishing Ltd, Bristol 1990
[3] H.K. Onnes, Akad. van Wetenschappen (Amsterdam) **14**, 113, 818 (1911)
[4] J. File and R.G. Mills, Phys. Rev. Lett. **10**, 93 (1963)
[5] M.N. Wilson, private communication
[6] W. Meissner and R Ochsenfeld, Naturwiss. **21**, 787 (1933)
[7] F. London and H. London, Z. Phys. **96**, 359 (1935)
[8] J. Bardeen, L.N. Cooper and J.R. Schrieffer, Phys. Rev. **108**, 1175 (1957)
[9] V.L. Ginzburg and L.D. Landau, JETP USSR **20**, 1064 (1950)
[10] L.P. Gorkov, Sov. Phys. JETP **9**, 1364 (1960), JETP **10**, 998 (1960)
[11] A.A. Abrikosov, Sov. Phys. JETP **5**, 1174 (1957)
[12] U. Essmann and H. Träuble, Phys. Lett. **24A**, 526 (1967) and Sci. Am. **224**, March 1971
[13] B.D. Josephson, Phys. Rev. Lett. **1**, 251 (1962)
[14] J.D. Livingston, Phys. Rev. **129**, 1943 (1963)
[15] N.E. Phillips, Phys. Rev. **134**, 385 (1965)
[16] Th. Schilcher, Diploma Thesis, Hamburg 1994, unpublished
[17] M.N. Wilson, Superconducting Magnets, Clarendon Press, Oxford 1983
[18] L.N. Cooper, Phys. Rev. **104**, 1189 (1956)
[19] H. Ibach and H. Lüth, Festkörperphysik, 2. Auflage, Springer-Verlag, Berlin 1988
[20] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1955)
[21] G. Möllenstedt and W. Bayh, Phys. Bl. **18**, 229 (1962)
[22] A. Tonomura, Phys. Today, April 1990
[23] R. Doll and M. Näbauer, Phys. Rev. Lett. **7**, 51 (1961)
[24] B.S. Deaver and W.M. Fairbank, Phys. Rev. Lett. **7**, 43 (1961)
[25] H. Rietschel, Phys. Bl. **46**, 419 (1990)
[26] C.E. Gough et al., Nature **326**, 855 (1987)
[27] G.T. Lee et al., Physica C **161**, 195 (1989)
[28] Y.B. Kim, C.F. Hempstead and A.R. Strnad, Phys. Rev. **139**, A1163 (1965)
[29] Y.B. Kim, C.F. Hempstead and A.R. Strnad, Phys. Rev. Lett. **9**, 306 (1962)

[30] P.W. Anderson, Phys. Rev. Lett. **9**, 309 (1962)
[31] H. Brück et al., Contribution to the European Accelerator Conference, Nice 1990
[32] M. Halemeyer et al., IEEE Trans. Appl. Supercond. Vol.3, No.1, 168 (1993)
[33] C. Keller et al., Cryogenics **30**, 410 (1990)
[34] R.C. Jaklevic et al. Phys. Rev. **140**, A1628 (1965)
[35] R.P. Huebener et al., Phys. Bl. **43**, 400 (1987)

APPENDIX A: THE FORMATION OF A COOPER PAIR

To illustrate the spirit of the BCS theory I will present the mathematics of Cooper-pair formation. Let us consider a metal at $T = 0$. The electrons fill all the energy levels below the Fermi energy while all levels above E_F are empty. The wave vector and the momentum of an electron are related by

$$\vec{p} = \hbar \vec{k} \, .$$

In the three-dimensional k space the Fermi sphere has a radius $k_F = \sqrt{2mE_F}/\hbar$. To the fully occupied Fermi sphere we add two electrons of opposite wave vectors $\vec{k}_1 = -\vec{k}_2$ whose energy $E_1 = E_2 = \hbar^2 k_1^2/(2m)$ is within the spherical shell (see Figs. 20, 23)

$$E_F < E_1 < E_F + \hbar\omega_D \, . \tag{42}$$

From Sec. 3 we know that $\hbar\omega_D$ is the largest energy quantum of the lattice vibrations. The interaction with the 'sea' of electrons inside the Fermi sphere is neglected except for the Pauli Principle: the two additional electrons are forbidden to go inside because all levels below E_F are occupied. Under this assumption the two electrons together have the energy $2E_1 > 2E_F$. Now the attractive force provided by the lattice deformation is taken into consideration. Following Cooper [18] we must demonstrate that the two electrons then form a bound system, a 'Cooper pair', whose energy drops below twice the Fermi energy

$$E_{\text{pair}} = 2E_F - \delta E < 2E_F \, .$$

The Schrödinger equation for the two electrons reads

$$-\frac{\hbar^2}{2m}(\nabla_1^2 + \nabla_1^2)\,\psi(\vec{r}_1, \vec{r}_2) + V(\vec{r}_1, \vec{r}_2)\psi(\vec{r}_1, \vec{r}_2) = E_{\text{pair}}\psi(\vec{r}_1, \vec{r}_2) \tag{43}$$

where V is the interaction potential due to the dynamical lattice polarisation. In the simple case of vanishing interaction, $V = 0$, the solution of (43) is the product of two plane waves

$$\psi(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{L^3}} \exp(i\vec{k}_1 \cdot \vec{r}_1) \cdot \frac{1}{\sqrt{L^3}} \exp(i\vec{k}_2 \cdot \vec{r}_2) = \frac{1}{L^3} \exp(i\vec{k} \cdot \vec{r})$$

with $\vec{k}_1 = -\vec{k}_2 = \vec{k}$ the k-vector, $\vec{r} = \vec{r}_1 - \vec{r}_2$ the relative coordinate and L^3 the normalisation volume. The most general solution of Eq. (43) with $V = 0$ is a superposition of such functions

$$\psi(\vec{r}) = \frac{1}{L^3} \sum_{\vec{k}} g(\vec{k}) \exp(i\vec{k} \cdot \vec{r}) \tag{44}$$

with the restriction that the coefficients $g(\vec{k})$ vanish unless $E_F \leq \hbar^2 k^2/2m \leq E_F + \hbar\omega_D$. This function is certainly not an exact solution of the equation (43) with $V \neq 0$ but for a weak potential it can be used to obtain the energy E_{pair} in first order perturbation theory. For this purpose we insert (44) into Eq. (43):

$$\frac{1}{L^3} \sum_{\vec{k}'} g(\vec{k}') \left[\frac{\hbar^2 k'^2}{m} + V(\vec{r}) - E_{\text{pair}} \right] \exp(i\vec{k}' \cdot \vec{r}) = 0 \, .$$

This equation is multiplied by $\exp(-i\vec{k} \cdot \vec{r})$ and integrated, using the orthogonality relations

$$\frac{1}{L^3} \int \exp\left(i(\vec{k}' - \vec{k}) \cdot \vec{r}\right) d^3r = \delta_{\vec{k}\vec{k}'} \quad \text{with} \quad \delta_{\vec{k}\vec{k}'} = \begin{cases} 1 \text{ for } \vec{k} = \vec{k}' \\ 0 \text{ otherwise.} \end{cases}$$

Introducing further the transition matrix elements of the potential V

$$V_{\vec{k}\vec{k}'} = \int \exp\left(-i(\vec{k} - \vec{k}') \cdot \vec{r}\right) V(\vec{r}) d^3r \tag{45}$$

one gets a relation among the coefficients of the expansion (44)

$$g(\vec{k}) \left[\frac{\hbar^2 k^2}{m} - E_{\text{pair}}\right] = -\frac{1}{L^3} \sum_{\vec{k}'} g(\vec{k}') V_{\vec{k}\vec{k}'} \ . \tag{46}$$

The matrix element $V_{\vec{k}\vec{k}'}$ describes the transition from the state $(\vec{k}, -\vec{k})$ to any other state $(\vec{k}', -\vec{k}')$ in the spherical shell of thickness $\hbar\omega_D$ around the Fermi sphere. Cooper and later Bardeen, Cooper and Schrieffer made the simplest conceivable assumption on these matrix elements, namely that they are all equal.

$$V_{\vec{k}\vec{k}'} = -V_0 \quad \text{for} \quad E_F < \frac{\hbar^2 k^2}{2m}, \frac{\hbar^2 k'^2}{2m} < E_F + \hbar\omega_D \tag{47}$$

and $V_{\vec{k}\vec{k}'} = 0$ elsewhere. The negative value ensures attraction.
With this extreme simplification the right-hand side of Eq. (46) is no longer $\vec{k}$ dependent but becomes a constant

$$-\frac{1}{L^3} \sum_{\vec{k}'} g(\vec{k}') V_{\vec{k}\vec{k}'} = \frac{V_0}{L^3} \sum_{\vec{k}'} g(\vec{k}') = A \ . \tag{48}$$

Then Eq. (46) yields for the coefficients

$$g(\vec{k}) = \frac{A}{\hbar^2 k^2/m - E_{\text{pair}}} = \frac{A}{\hbar^2 k^2/m - 2E_F + \delta E} \ .$$

The constant A is still unknown. We can eliminate it by summing this expression over all $\vec{k}$ and using (48) once more

$$\sum_{\vec{k}} g(\vec{k}) = A \frac{L^3}{V_0}$$

from which follows

$$A \frac{L^3}{V_0} = \sum_{\vec{k}} \frac{A}{\hbar^2 k^2/m - 2E_F + \delta E} \ .$$

Dividing by A leads to the important consistency relation

$$1 = \frac{V_0}{L^3} \sum_{\vec{k}} \frac{1}{\hbar^2 k^2/m - 2E_F + \delta E} \ . \tag{49}$$

The sum extends over all $\vec{k}$ vectors in the shell between E_F and $E_F + \hbar\omega_D$. Since the states are very densely spaced one can replace the summation by an integration

$$\frac{1}{L^3} \sum_{\vec{k}} \to \frac{1}{(2\pi)^3} \int d^3k \to \int \mathcal{N}(E) dE$$

where $\mathcal{N}(E)$ is the density of single-electron states for a definite spin orientation. (The states with opposite spin orientation must not be counted because a Cooper pair consists of two electrons of opposite spin). The integration spans the narrow energy range $[E_F, E_F + \hbar\omega_D]$ so

$\mathcal{N}(E)$ can be replaced by $\mathcal{N}(E_F)$ and taken out of the integral. Introducing a scaled energy variable

$$\xi = E - E_F = \frac{\hbar^2 k^2}{2m} - E_F$$

formula (49) becomes

$$1 = V_0 \mathcal{N}(E_F) \int_0^{\hbar\omega_D} \frac{d\xi}{2\xi + \delta E} . \tag{50}$$

The integral yields

$$\frac{1}{2} \ln \left(\frac{\delta E + 2\hbar\omega_D}{\delta E} \right) .$$

The energy shift is

$$\delta E = \frac{2\hbar\omega_D}{\exp(2/(V_0 \mathcal{N}(E_F)) - 1} .$$

For small interaction potentials ($V_0 \mathcal{N}(E_F) \ll 1$) this leads to the famous Cooper formula

$$\delta E = 2\hbar\omega_D \exp \left(- \frac{2}{V_0 \mathcal{N}(E_F)} \right) . \tag{51}$$

Except for a factor of 2 the same exponential appears in the BCS equations for the energy gap and the critical temperature.

APPENDIX B: SURFACE RESISTANCE IN MICROWAVE FIELDS

In high-frequency electromagnetic fields a superconductor is no longer free of energy dissipation. The reason is that the a.c. magnetic field penetrates into the material about one London-penetration-length deep and induces an oscillating electric field. As a consequence eddy currents of the unpaired electrons result leading to Ohmic heating. The amplitude of the electric field derives from the Maxwell equation

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} .$$

So, for a periodic surface magnetic field $B_s \cos(\omega t)$, the electric field in the London penetration layer is proportional to the frequency and the B field: $E_s \propto \omega B_s$. The single-electron current density J_e is proportional to the density n_e of unpaired electrons and to the electric field, so the dissipated power density is

$$P_{\text{diss}} = J_e E_s \propto n_e \, \omega^2 B_s^2 .$$

The probability for finding an unpaired electron with an energy $E \geq E_F + \Delta$ (compare Fig. 24) is calculated using the Fermi-Dirac distribution function. The number of unpaired electrons is

$$n_e \propto \frac{1}{\exp((E - E_F)/(k_B T)) + 1} \approx \exp(-\Delta/(k_B T)) .$$

The approximation is valid for $T < T_c/2$. Using Eq. (20) to express the energy gap in terms of the critical temperature we obtain for the dissipated power per unit volume

$$P_{\text{diss}} \propto \omega^2 \, B_s \, \exp(-1.76 T_c/T) . \tag{52}$$

For the surface resistance one obtains the expression

$$R_s = \frac{A}{T} \omega^2 \exp(-1.76 T_c/T) + R_{\text{res}} . \tag{53}$$

The quantity A depends on the mean free path of the unpaired electrons; R_{res} is the residual resistance caused by impurities or defects.

NEXT PAGE(S) left BLANK

SUPERCONDUCTING MATERIALS FOR MAGNETS

*Martin N. Wilson**
CERN, Geneva, Switzerland

Abstract
Superconducting materials for magnets will be discussed from three viewpoints: firstly the basic physical properties of the materials themselves, secondly the constraints and preferences imposed by the requirements of magnet building and finally, the manufacture of magnet conductors as a commercial product.

1. INTRODUCTION

Superconducting materials for magnets have been manufactured on a commercial basis for about 30 years and can now be regarded as a mature, but still evolving, industrial product. Accelerators such as the Tevatron, HERA and RHIC have so far constituted the largest individual units of consumption, but they are not the largest market. The largest total sales per annum have, over the last decade, been for the construction of magnets for medical MRI imaging.

With the exception of a few small developmental magnets, all useable magnets built to date have used the low temperature, high field, *Type II* superconducting materials. The newer high temperature materials are described elsewhere in these proceedings [1]. Niobium titanium alloy NbTi is currently the standard 'work horse' of the superconducting magnet industry. It has been used in all accelerators so far constructed or planned, for all MRI systems and most other magnets. It is a ductile alloy with mechanical properties which make it easy to fabricate and use.

Figure 1 shows the critical surface of NbTi: superconductivity prevails everywhere below the surface and normal resistivity everywhere above it. Superconducting magnets usually operate at a fixed temperature, generally the boiling point of liquid helium ~ 4.2 K. To see the performance of the material at this temperature, one takes a slice through the surface at 4.2 K, thereby obtaining the critical current density versus field, as shown in Fig. 2. Also shown in Fig. 2 is the typical operating domain for conventional magnets, in which the current density is limited by ohmic heating and field is limited by the saturation of iron. Superconducting magnets do not need iron yokes, although iron is often used to screen the fringe field — and of course they have no ohmic heating. Current density is thus set only by the properties of the superconducting material, although we shall see later that there are many practical considerations which cause this very high current density to be diluted in practice. Nevertheless, the operating current densities of superconducting magnets are typically 1 – 2 orders of magnitude higher than conventional magnets, which means that higher fields may be produced with more compact windings. This compactness, together with the lack of ohmic dissipation and the capacity for higher fields, is why superconducting magnets are so attractive for use in particle accelerators.

At 4.2 K the superconducting properties of NbTi are adequate up to fields of about 9 T, after which they fall off steeply. For magnets above this field, one of the brittle intermetallic compounds, usually niobium tin Nb_3Sn, must be used. As shown in Fig. 2, Nb_3Sn is able to reach ~ 20 T, with even higher current density than NbTi, although in practice the dilution of current density is greater in Nb_3Sn conductors than NbTi. For even higher fields, there are several other intermetallic compounds which offer the possibility of higher performance than Nb_3Sn, but they have not yet been developed commercially.

---

* On secondment from Oxford Instruments, Accelerator Technology Group, Osney Mead, Oxford, OX2 0DX, England

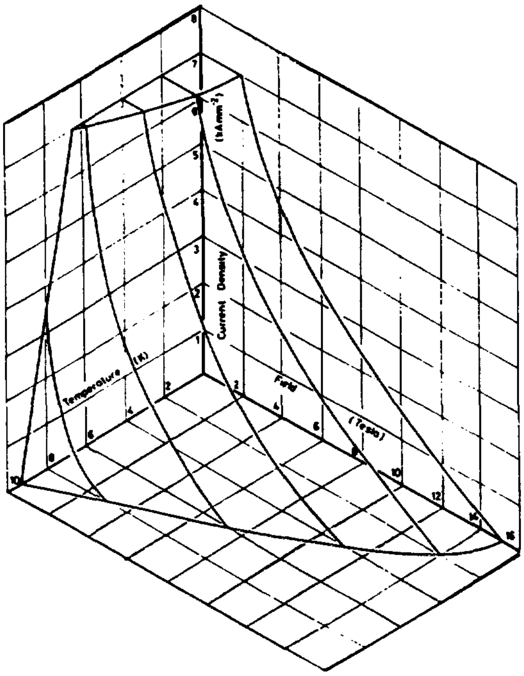


Fig. 1 The critical surface of niobium titanium: superconductivity prevails everywhere below the surface and normal resistivity everywhere above it.

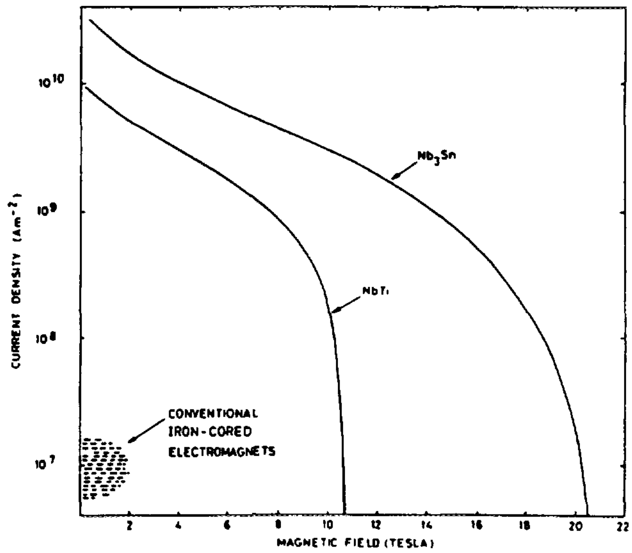


Fig. 2 The critical current densities at 4.2 K of NbTi and Nb_3Sn. Also shown for comparison is the usual operating region of conventional iron-yoke magnets with water-cooled copper windings.

In the following three sections, we first describe the physical properties of the main magnet materials. We then move on to discuss the particular constraints, problems and performance requirements which arise when these conductors are used in magnets. On the basis of these constraints, we define a set of design criteria for magnet conductors and then describe the practical industrial production of magnet materials.

2. SURVEY OF MATERIALS

2.1 Properties of the technical type II superconductors

As described in other lectures [2], the key parameters of critical field, temperature and current have rather different origins. Critical temperature and field are essentially determined by the chemistry of the material, whereas critical current is determined by its microstructure. Critical temperature of the material in zero field is simply related to the superconducting energy gap by:

$$3.5 k_B \theta_c = 2\Delta(0) \tag{1}$$

where k_B is Boltzmann's constant and D(0) is the energy gap at zero degrees. For Type II superconductors in the so called 'dirty limit', the upper critical field at zero temperature (and zero current density) is approximately given by [3]:

$$B_{c2}(0) \approx 3.1 \times 10^3 \gamma \rho_n \theta_c \tag{2}$$

where γ is the Sommerfeld coefficient of electronic specific heat and ρ_n is the resistivity in the normal state. Thus the section through Fig. 1 in the plane of zero current density is totally determined by the chemistry of that material.

Critical current density is quite a different matter. In their pure uniform annealed state, Type II superconductors carry no current in the bulk, only on their surface. As described in earlier lectures, the magnetic field penetrates these superconductors in the form of quantized fluxoids. For a uniform material, the fluxoids arrange themselves into a regular lattice, as shown in Fig. 3. A uniform array means a uniform field and hence, from $\mathrm{curl}\, B = \mu_0 J$, the current density is zero.

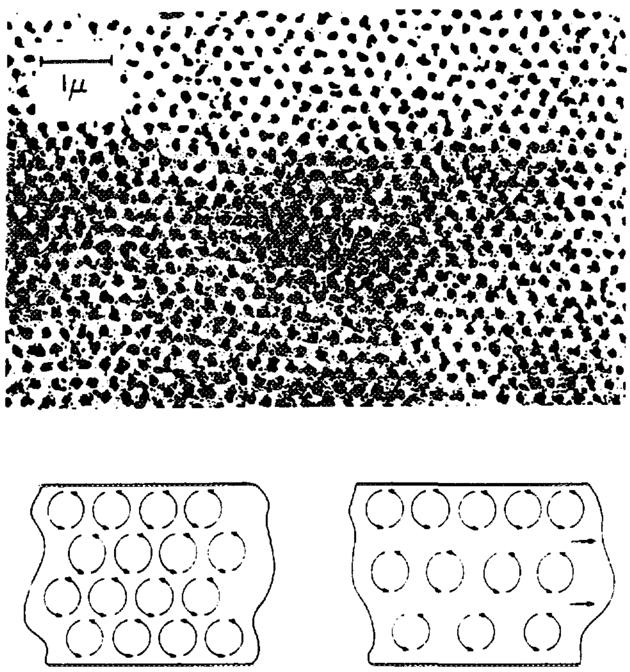


Fig. 3 Fluxoid lattice in a Type II superconductor, sketch shows firstly a uniform lattice and secondly a gradient, which results in a bulk current density

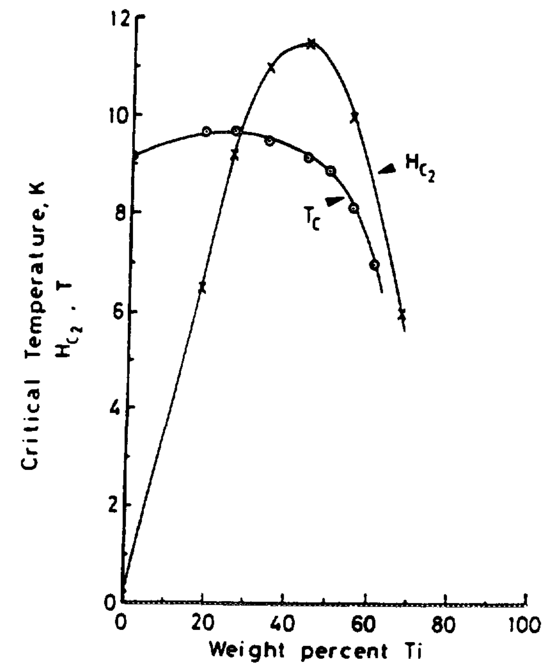


Fig. 4 The critical temperature and upper critical field of NbTi at 4.2 K as a function of alloy composition [4]

In order to promote the flow of currents within the bulk material, it is necessary to produce a non-uniform fluxoid lattice, with a gradient in density. Such density gradients may be produced by means of *pinning centres*, i.e. non-uniformities in the microstructure of the material which cause the energy of each fluxoid to vary with position, thereby giving rise to minimum energy configurations. In this way, the natural tendency of the fluxoids to form a uniform lattice is opposed and concentration gradients can be built up. There are a variety of microstructural features which can serve as pinning centres. As described in [2], the flux carried by an individual fluxoid is:

$$\phi_0 = \frac{h}{2e} = 2\times10^{-15} \quad \text{Webers} \tag{3}$$

This is a small quantity, it is approximately the flux enclosed by the cross section of human hair in the earth's magnetic field. For fluxoids arranged in a regular triangular lattice, producing an overall average field B, the spacing between fluxoids is:

$$d = \left\{ \frac{2}{\sqrt{3}} \frac{\phi_o}{B} \right\}^{1/2} \tag{4}$$

For a field of 5 T, the spacing is thus 22 nm. To provide effective flux pinning at 5 T, a material must therefore have microstructural features on a scale of ~ 20 nm.

Unlike critical field and temperature, critical current density is thus a parameter which must be 'engineered' by modifying the microstructure. The characteristic curve of J_c versus B shown in Fig. 2 is a standard design tool of all magnet builders and maximum current carrying capacity at a given field is their usual criterion of goodness. Much effort has accordingly been devoted by materials scientists and conductor manufacturers to understanding and improving the critical current density, with the result that it has steadily improved over the years.

2.2 Niobium titanium

Niobium and titanium are mutually soluble to form ductile alloys over a wide range of compositions. One may thus choose the composition for optimum critical properties. Figure 4 shows how the critical temperature and upper critical field of NbTi at 4.2 K vary as a function of composition. Although the broad optimum in critical temperature occurs towards the lower end of titanium content, commercial alloys are almost always formulated for optimum critical field, in the range Nb46.5 to 50wt%Ti. The optimum critical field and critical temperature do not occur at the same composition because the normal state resistivity increases with titanium content, see Eq. (2). A ternary element, most commonly tantalum, may also be added to the alloy to produce a modest increase in B_{c2} of ~ 0.3 T at 4.2 K, rising to 1.3 T at 1.8 K [5].

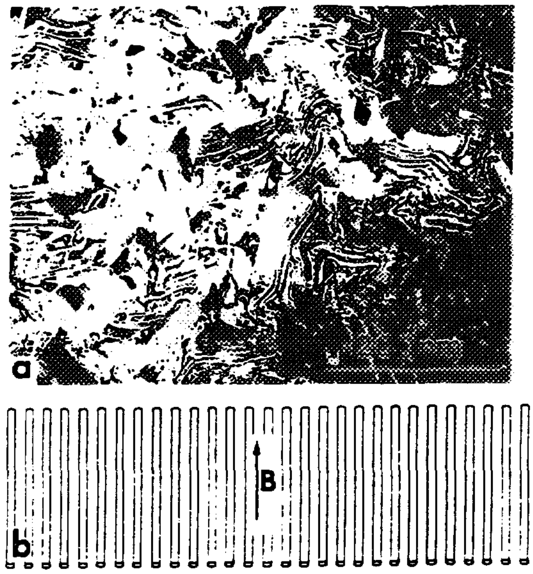
Originally it was thought that the principle source of flux pinning in NbTi came from the dislocation cell structure in the wire, which is heavily cold worked in order to reduce it to final size. Thanks to the work of Larbalestier and the Wisconsin group however it has now become clear that most of the flux pinning is provided by finely divided deposits of the 'a phase', a titanium-rich phase which is precipitated on the dislocation cell boundaries as a result of the heat treatments which are applied during manufacture. As noted above, for optimum pinning the precipitate cell size should be matched to the fluxoid spacing. Figure 5 shows an electron micrograph of a NbTi filament which has been processed in such a way as to achieve one of the highest current densities reported to date: 3700 A/mm^2 at 5 T. Figure 6 provides convincing support for the theory that a Ti is the main source of pinning by plotting J_c versus α Ti content for this particular conductor.

A recent development in the technology of NbTi has been the use of Artificial Pinning Centres APC. These pinning centres are produced by incorporating fine fibres of a different metal, such as copper or niobium, into the NbTi and then drawing down until the APC fibres have a size and spacing comparable with the fluxoid lattice spacing. To date, some very high current densities have been obtained at moderate fields [6], but less so at high fields.

The continuing push towards higher fields has lead to NbTi being used at temperatures lower than 4.2 K. The LHC is designed to operate in superfluid helium at 1.9 K, where the upper critical field of NbTi is increased from its 4.2 K value of ~10.5 T to ~14.2 T. High current densities can thus be achieved up to 10 T and the use of superfluid brings many other benefits in terms of better cooling and stability.

2.3 Niobium tin

At present, Nb_3Sn is the only other magnet conductor to be produced on a regular commercial basis. Unlike NbTi, niobium tin is a brittle intermetallic compound having a well defined stoichiometry Nb_3Sn. The crystal structure is of the type A15, which is shown in Fig. 7. This structure is shared by many other high field Type II intermetallic compounds.



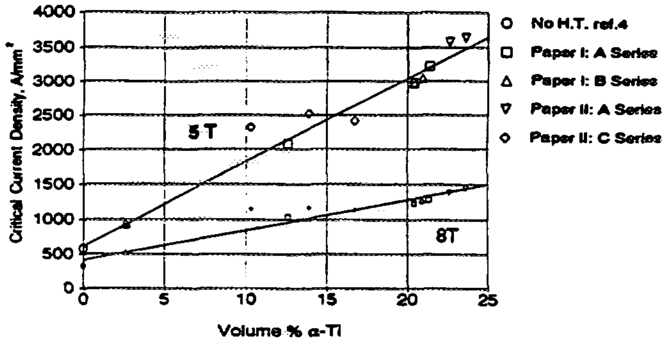


Fig. 5 a) Transmission electron micrograph of a Nb-45wt%Ti superconductor manufactured by Oxford Superconducting Technology. The white regions are precipitates of α titanium, which are thought to provide the major source of flux pinning. b) Sketch of the fluxoid lattice at 5 T on the same scale [4].

Fig. 6 Critical current density versus α Ti content for the NbTi filament shown in Fig. 5

A_3B

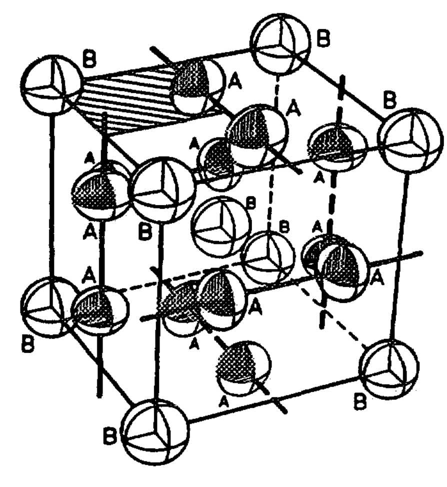


Fig. 7 The A15 crystal structure

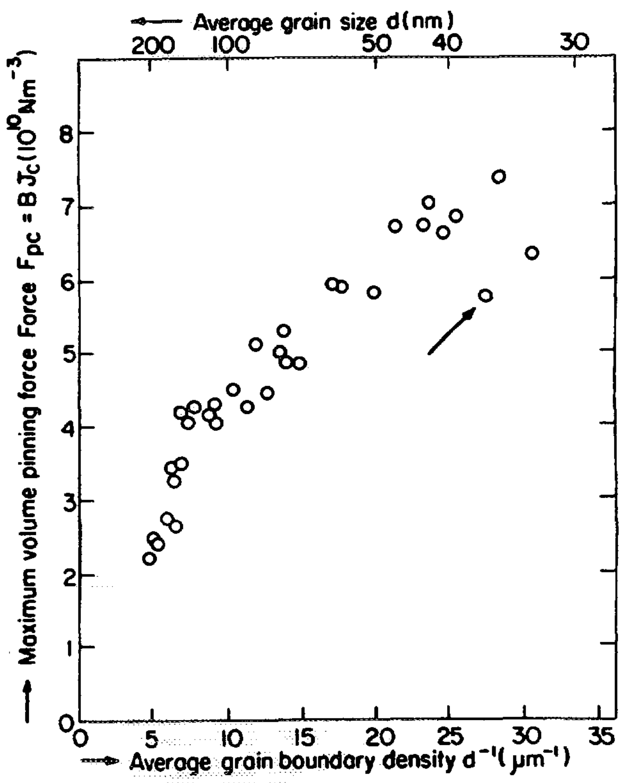


Fig. 8 The maximum pinning force per unit volume $F = B^*J_c$ versus effective grain size for a Nb_3Sn conductor produced via the bronze process

Because it is brittle, Nb_3Sn as such cannot be fabricated by a wire-drawing process, it must be formed in situ at its final size. Originally it was produced in the form of tape by either a chemical vapour deposition process or by a diffusion process. As we shall see in section 3 however, there are several good reasons for wanting the superconductor to be subdivided into fine filaments and it has accordingly been necessary to develop processing techniques for making Nb_3Sn in this form. The most popular of these techniques is the bronze process [7] in which filaments of pure Nb are drawn down in a matrix of CuSn, bronze. When the wire has been drawn down to final size, it is given a heat treatment, during the course of which the tin diffuses through the bronze and reacts with the niobium to produce Nb_3Sn.

The dominant source of flux pinning in Nb_3Sn seems to be the grain boundaries. Figure 8 illustrates this by plotting, as a function of grain size, the maximum pinning force per unit volume, which is defined as the product of critical current density and field. In general, the maximum pinning force occurs at about half the upper critical field.

Producing the best critical current density in Nb_3Sn is thus a matter of producing the finest grain structure. Unfortunately this requirement is in direct conflict with the equally important requirement for optimum B_c and q_c, which are only achieved by material of the right stoichoimetry, i.e. Nb_3Sn and not $Nb_{3.2}Sn$. In material which is produced by diffusion, good stoichiometry is obtained by long heat treatments at high temperature, but unfortunately such treatments also produce grain growth. The optimum heat treatment is thus always a compromise between the conflicting requirements of good stoichiometry and fine grain size. In this regard, the bronze process is helpful because the copper enables Nb_3Sn to form at much lower temperatures than with pure tin. Other methods for achieving good stoichiometry without excessive grain growth include fine subdivision and ternary additions.

If the bronze is to remain ductile enough for wire drawing, it must contain no more than ~15% of tin, in fact most practical conductors have so far been made with bronze containing ~12% tin. At the end of the reaction, the depleted bronze remaining occupies at least ~ 4x the volume of the niobium tin. Unfortunately the electrical and thermal conductivity of this bronze are not sufficient to play a useful role in stabilizing the conductor (see next section) and so the space it occupies is essentially wasted, bringing an unwelcome dilution of overall current density. One expedient for reducing this dilution is to incorporate reservoirs of pure tin within the cross section of the wire, the so called internal-tin process [8]. During heat treatment, this additional tin migrates through the bronze to react with the niobium filament. In this way, the cross section of bronze needed to supply the required amount of tin may be considerably reduced.

An alternative approach to the production of Nb_3Sn in filamentary form is the so called ECN method [9]. In this process, niobium tubes are filled with $NbSn_2$ powder and are then drawn down in a copper matrix. A heat treatment at final size causes the $NbSn_2$ to react with the niobium forming Nb_3Sn.

2.4 Strain sensitivity

Like all the A15 compounds, Nb_3Sn is very brittle. In bulk form it fractures at a tensile strain of ~ 0.3%. In finely divided filamentary form, when supported by a surrounding matrix, it can be strained to ~ 0.7% before fracture, but is nevertheless very fragile. After reaction, the wires must be handled with extreme care. In fact the most reliable method of making magnets is to wind the coils and then react the whole coil in its final configuration.

Even if it is not strained to fracture, Nb_3Sn exhibits strong strain sensitivity. Figure 9 (a) shows the change in critical current of several different Nb_3Sn filamentary composites at various fields as a function of applied strain. On closer investigation, it is found that much of the complexity in this behaviour is caused by a precompression of the filaments by the bronze matrix, which contracts much more than the Nb_3Sn when the conductor cools from its reaction temperature to 4.2 K. In fact, the rising part of the curve is actually a region of reducing compressive strain in the Nb_3Sn as the applied tension counteracts the thermally induced precompression. If one eliminates the differential contraction element and plots I/I_c as a function of intrinsic strain one finds that all Nb_3Sn conductors at the same field show the same behaviour: critical current always reduces with strain, whether it is tensile or compressive. Furthermore, the reduction becomes stronger at higher fields, indicating that the effects is related to a reduction in B_{c2}. Figure 9 (b) compares the strain behaviour of Nb_3Sn made with different ternary additions with some of the other A15 compounds mentioned in the next section. It may be seen that Nb_3Sn is among the most strain sensitive.

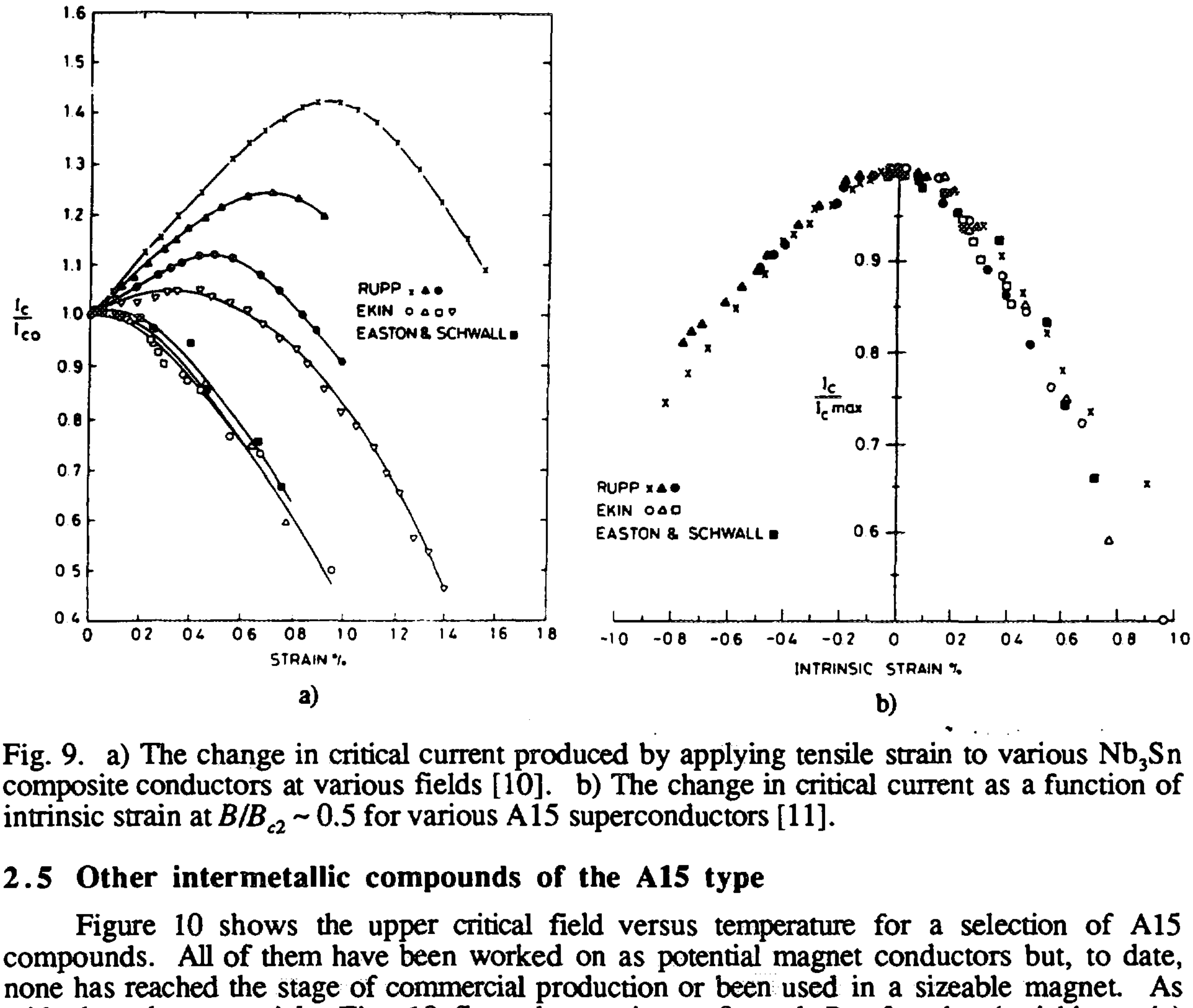


Fig. 9. a) The change in critical current produced by applying tensile strain to various Nb_3Sn composite conductors at various fields [10]. b) The change in critical current as a function of intrinsic strain at $B/B_{c2} \sim 0.5$ for various A15 superconductors [11].

2.5 Other intermetallic compounds of the A15 type

Figure 10 shows the upper critical field versus temperature for a selection of A15 compounds. All of them have been worked on as potential magnet conductors but, to date, none has reached the stage of commercial production or been used in a sizeable magnet. As with the other materials, Fig. 10 fixes the maximum θ_c and B_{c2} for the (stoichiometric) compound and the materials scientist must then use his ingenuity to produce a microstructure to give the best flux pinning and hence the highest J_c.

Of all the candidate materials shown in Fig. 10, perhaps the most promising is niobium aluminium because, as shown in Fig. 9 (b), it has the lowest strain sensitivity. For this reason, it has recently excited quite a lot of interest as a potential alternative to Nb_3Sn in the project ITER for a large-scale fusion reactor. Unfortunately there appears to be no equivalent of the bronze process for Nb_3Al, which means that the conflict between stoichiometry and grain size is severe. Nevertheless, some promising results have recently been obtained by processing elemental Nb and Sn in finely subdivided form. At final size, the wire is given a very rapid high temperature heat treatment followed by a quench to about room temperature, which forms stoichiometric, but amorphous, Nb_3Al. A subsequent heat treatment at lower temperatures is then used to grow fine grains of the A15 phase [12].

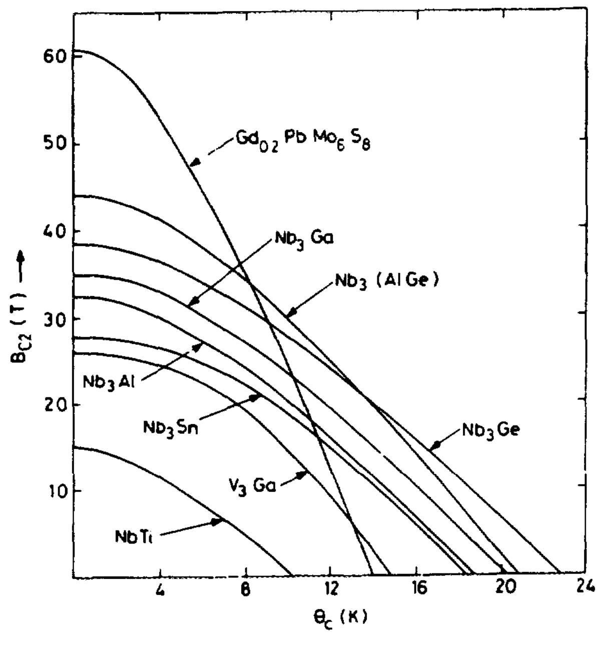


Fig. 10 Critical fields and temperatures of various high field superconducting materials

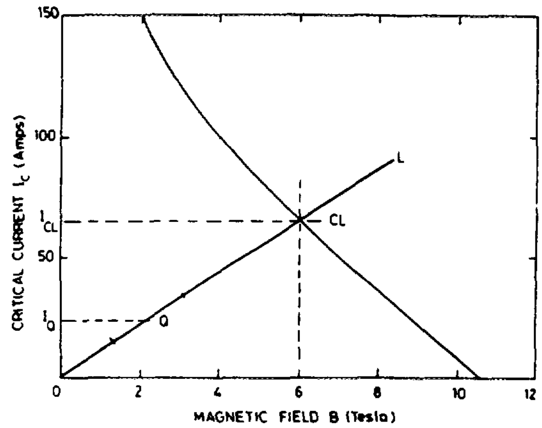


Fig. 11 Performance of a superconducting magnet showing critical current of the conductor and the magnet 'load line' which relates peak field on the magnet to the current flowing through its windings

3. CONDUCTORS FOR USE IN MAGNETS

3.1 Degradation and training

An early problem of superconducting magnets, and one which still remains to some extent, was the fact that conductors often failed to achieve in magnets the performance which they reached in short samples. Figure 11 shows a type of diagram frequently used in magnet design. The upper curve is the critical current of the conductor as a function of field and the load line OQL relates the peak field seen by the magnet to the current flowing in its windings (load lines are straight for air cored magnets but curved if iron is used). The point at which the peak field part of the magnet would be expected to reach its critical state is given by the intersection CL, but in fact it was originally found that the magnets went into the resistive state at much lower currents, typically in the region Q. This transition of the magnet from superconducting to resistive state is known as a *quench*. It can be a very spectacular process, because most of the stored inductive energy $1/2\, L\, I^2$ is dissipated as resistive heating in that part of the conductor which has become resistive. Quenching will be treated in another lecture [13]; suffice here to note that it is an irreversible process and, once it has happened, one can only turn off the magnet power supply, wait until the magnet has cooled down and then try again.

Quenching at lower currents than the critical point CL of Fig. 11 is generally known as degradation. It is often accompanied by a related behaviour known as training and illustrated in Fig. 12. After successive quenches, the magnet is able to achieve progressively higher currents, but still some variability in performance remains and very rarely does the magnet achieve its full critical current.

Degradation and training are clearly undesirable in any magnet, but are particularly to be avoided in an accelerator system, where up to 1000 magnets may be connected in series, such

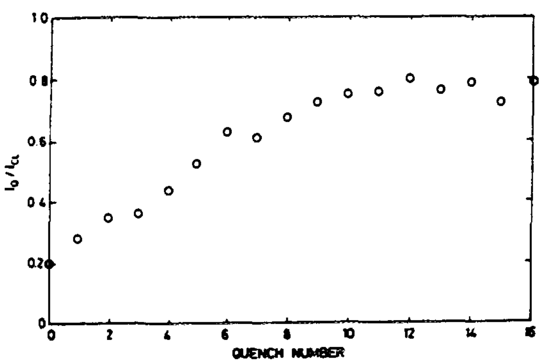


Fig. 12 Training: the progressive increase in magnet quench current as a fraction of the load line critical current after repeated quenching

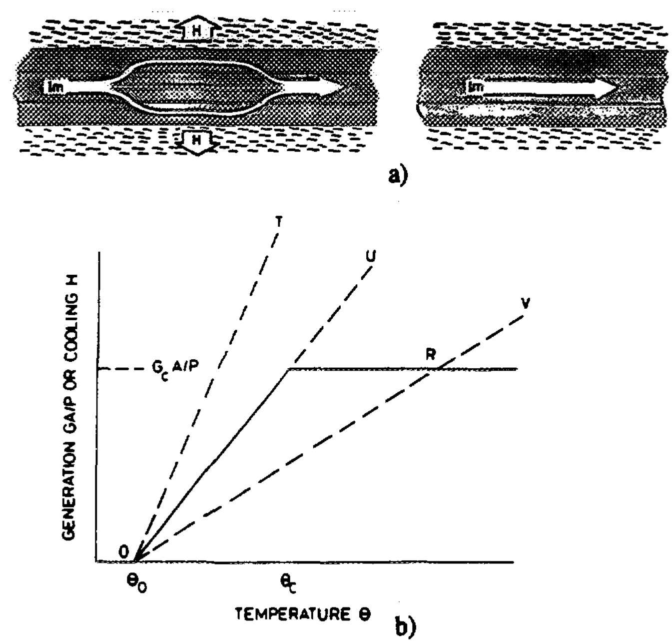


Fig. 13 Cryogenic stabilization (a) the general principle: if the superconductor becomes resistive, current switches to the parallel path in copper, ohmic heating is removed via heat transfer to the liquid helium. (b) quantitative: in units of power per unit cooled area of the conductor surface, for stability the generation line OQR must lie below the cooling line OT.

that a single magnet quenching will bring the whole system down. Strategies to avoid or minimize degradation go under the generic name of stabilization; they have important consequences for conductor design, the most important of which are described in the following sections.

3.2 Cryogenic stabilization

Historically, cryogenic stabilization was the earliest successful technique; for the first time it enabled magnets to work reliably at their critical field [14]. Figure 13 (a) shows the general principle; the superconductor is joined along its entire length to a conductor of low resistivity (usually copper, sometimes high-purity aluminium) with a much greater cross sectional area. If for any reason the superconductor stops conducting, current switches to the copper, where it generates heat. By means of cooling channels in the magnet, the copper is sufficiently well cooled for the ohmic heating to be dissipated without excessive temperature rise. Provided this temperature rise is less than a certain value, the superconductor eventually resumes its full superconducting state, current transfers back to it from the copper and the ohmic heating ceases.

Figure 13 (b) shows the quantitative aspects of cryogenic stabilization. The line OQR plots ohmic power generation versus temperature; the shape arises as follows. For most Type II materials at fixed field, the critical current density falls off linearly with temperature. Thus in the section OQ of Fig. 12 (b), the current carrying capacity of the superconductor is reducing with increasing temperature and the excess current is transferring to the copper. At point Q, all the current has transferred to the copper and further increases in temperature do not change the power generation. The variation of power generation with temperature is thus:

$$G(\theta) = \frac{\lambda^2 J_c^2 \rho\,(\theta - \theta_o)}{(1-\lambda).(\theta_c - \theta_o)} \tag{5}$$

where λ is the proportion of superconductor in the conductor cross section (the rest being copper), J_c is the critical current density, q is the conductor temperature, θ_o is the temperature of the cryogenic bath and θ_c is the superconductor critical temperature. The variation of cooling with temperature is:

$$H(\theta) = h\,(\theta - \theta_o) \tag{6}$$

where h is the heat transfer coefficient. Note that $G(\theta)$ is defined per unit volume and $H(\theta)$ per unit cooled area; to bring them into line we need a factor A/P where A is the area of cross section and P is the cooled perimeter. The condition for stability may then be written.

$$\alpha = \frac{\lambda^2 J_c^2 \rho A}{(1-\lambda) P h (\theta_c - \theta_o)} \leq 1 \tag{7}$$

Further complications arise when the heat transfer coefficient is not constant, but the general principle remains the same.

Cryogenic stabilization works well and indeed has made possible all the large superconducting magnets in operation today. Its big drawback is the large amount of copper and cooling channel required to satisfy Eq. (7), despite the considerable reduction in ρ which pure metals show when cooled to low temperature. Typical dilutions of current density are in the range 20 to 50. For large magnet systems such as detector magnets this dilution is not a problem, but for accelerator magnets it is out of the question. What is needed is a way of reducing degradation without diluting the current density too much. To achieve this, we need to understand a little more about the causes of degradation.

3.3 Flux jumping

The first cause of degradation to be investigated systematically was a phenomenon known as flux jumping. To understand what happens, let us consider a superconductor of simplified 'slab' geometry as shown in Fig. 14. When a magnetic field is applied parallel to the broad face of the slab, it induces currents which try to screen the inside of the slab – very much like eddy currents except that they do not decay. Figure 14 (b) is a plot of the magnetic field amplitude across a section of the slab for various values of the external field.

To see how these field shapes arise, let us consider the effect of raising the external field from zero in increments DB. The first DB initially induces a current to flow in the surface of the slab at very high current density. Because this density is higher than the critical current density J_c, the surface current decays resistively and magnetic field starts to penetrate the interior of the slab. As soon as the current density falls to J_c however decay ceases and the slab is left with oppositely directed currents flowing at critical density on either face. From curl $B = \mu_o J$ with no variation in the y and z directions, we have $dB/dx = \mu_o J_c = \Delta B/p$ where p is the penetration depth of the field. A second increment ΔB produces similar effects, with penetration to a depth $2\Delta B/\mu_o J_c$. This process continues until the whole slab is carrying critical current, at which point the field has fully penetrated. Further increases in field will penetrate the whole of the slab without causing any change in the screening current pattern, which flows at critical density in each half of the slab.

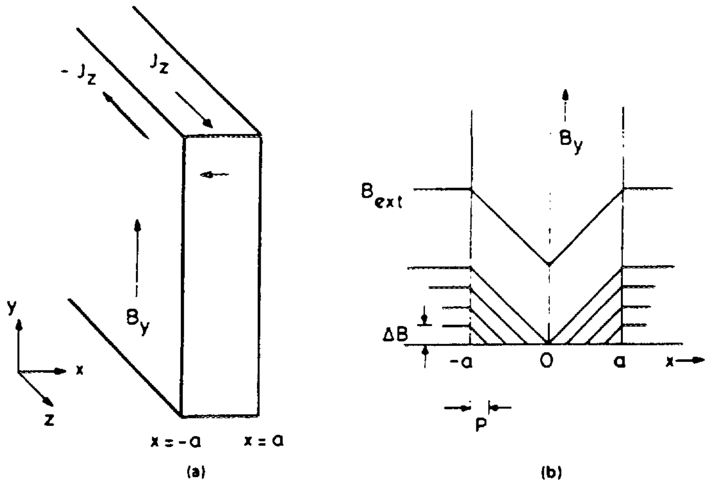


Fig. 14 (a) Screening currents induced to flow in a superconducting slab by a magnetic field parallel to the slab surface. (b) Profile of magnetic field across the slab showing the reduction of internal field caused by the screening currents.

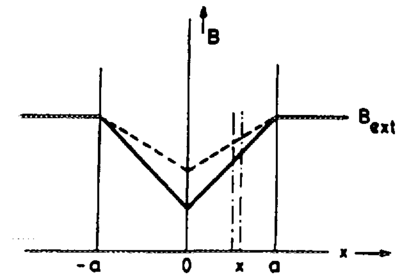


Fig. 15 The change in screening currents and internal field distribution caused by a rise in temperature

Screening currents of this kind flow in any superconductor placed in an increasing field; they flow in the windings of all magnets and are in addition to the *transport current*, i.e. the current supplied to the magnet terminals from the power supply. They work in such a way as to ensure that the superconductor is always filled with current flowing at critical density. Screening currents may become unstable because of the interaction between two inherent properties of high-field superconductors

a) critical current densities fall with increasing temperature;
b) flux motion within the superconductor generates heat.

To see how the instability arises, let us carry out an imaginary experiment to measure the specific heat of our slab while it is carrying screening currents in the external field. We thermally isolate the slab, supply a quantity of heat ΔQ_s, and observe the temperature rise $\Delta\theta$ The critical current density will be reduced by the temperature rise so that the screening currents will encounter some resistance and decay to a lower J_c, illustrated by the dashed line in Fig. 15. During decay the resistive voltage drop will be driven by the flux change within the slab, generating heat. In the region between x and $x+\delta x$ this heat is:

$$\delta q(x) = \int I(x)E(x)dt = J_c\,\delta x\,\delta\phi(x) \tag{8}$$

where $\delta\phi(x)$ is the flux change enclosed between the planes x and the centre line $x = 0$ (where, by symmetry the electric field is zero) thus:

$$\delta\phi(x) = \int_0^x \Delta B(x)\,dx = \int_0^x \mu_o\,\Delta J_c(a-x)\,dx = \mu_o\,\Delta J_c(ax - \frac{x^2}{2}) \tag{9}$$

and the heat per unit volume averaged over the whole slab is:

$$\Delta Q = \frac{1}{a}\int_0^a \delta q(x)\,dx = \frac{1}{a}\int_0^a \mu_o J_c \Delta J_c (ax - \frac{x^2}{2})\,dx = \mu_o J_c \Delta J_c \frac{a^2}{3} \tag{10}$$

Assuming a linear fall in J_c with θ:

$$\Delta J_c = -J_c \frac{\Delta\theta}{(\theta_c - \theta_o)} \tag{11}$$

we may now write the heat balance for the slab:

$$\Delta Q_s + \frac{\mu_o J_c^2 a^2}{3(\theta_c - \theta_o)}\Delta\theta = \gamma C \Delta\theta \tag{12}$$

where C is the actual specific heat and g is the density. However the effective specific heat measured by the experiment is:

$$\gamma C_e = \frac{\Delta Q_s}{\Delta\theta} = \gamma C - \frac{\mu_o J_c^2 a^2}{3(\theta_c - \theta_o)} \tag{13}$$

The energy stored in the screening currents, represented by the last term in Eq. (13), has reduced the effective heat capacity so that the temperature rise caused by a given ΔQ_s is increased. The ultimate catastrophe occurs when this last term is equal to γC: the effective heat capacity goes to zero and the smallest disturbance will cause the temperature to rise without limit — a flux jump. If this happens in a magnet winding, the magnet will quench. From Eq. (13), it may be seen that the condition for flux jumps to be avoided is:

$$\frac{\mu_o J_c^2 a^2}{\gamma C(\theta_c - \theta_o)} \leq 3 \tag{14}$$

Because of the assumption that no heat is exchanged with the surroundings, this condition is known as the adiabatic stability criterion. Fulfilment of the criterion in technical superconductors demands a fairly fine degree of subdivision. For example, niobium titanium at 4.2 K and 2 T typically has the following properties:

critical current density	$J_c = 6 \times 10^9$ A m^{-2}
density	$\gamma = 6.2 \times 10^3$ kg m^{-3}
specific heat	$C = 0.89$ J kg^{-1}
critical temperature	$\theta_c = 8.6$ K

so that flux jumps will occur when the half-width, a, exceeds 40 mm. Spontaneous flux jumping will not occur when a is reduced below 40 mm but the effective specific heat will still be markedly reduced, leaving the slab very susceptible to small energy inputs. A safer criterion is therefore to choose about half this dimension i.e. a = 20 mm which gives an effective specific heat 75% of its usual value. Note that we have taken a low field value of J_c because this gives a more conservative criterion — and there is always a low field region somewhere in the magnet winding.

The adiabatic theory of jumping has been checked experimentally by a number of workers. A common approach has been to take a sample which is much larger than a_{FJ} so that flux jumping occurs long before the field has fully penetrated. Theory similar to that outlined above shows that a flux jump is to be expected when the external field reaches B_{FJ} given by:

$$B_{FJ} = \mu_0 J_c p = \{3\mu_o \gamma C(\theta_c - \theta_c)\}^{1/2} \tag{15}$$

Figure 16 shows that there is reasonable agreement between Eq. (15) and experiment.

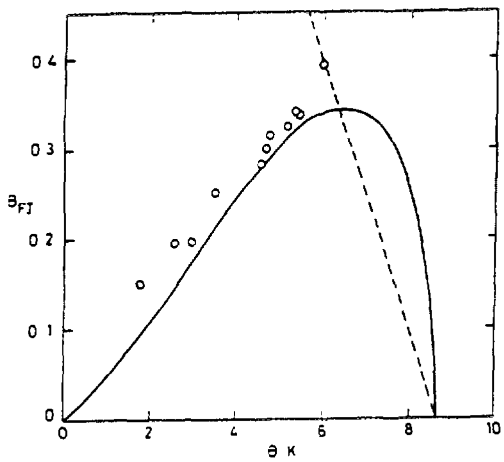


Fig. 16 Comparison between Eq. 15 and experimentally observed flux jumping fields (data from [15])

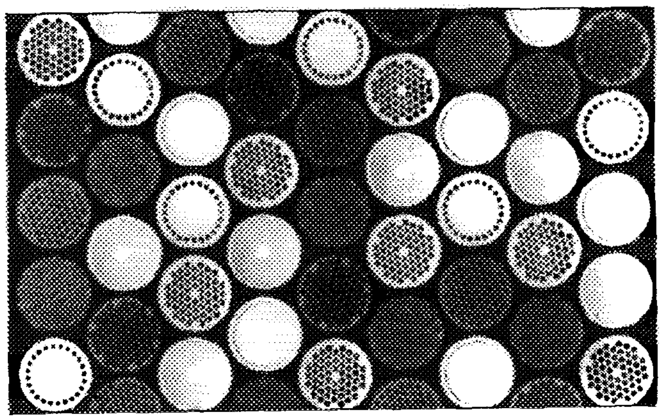


Fig. 17 Cross sections of some filamentary composite superconducting wires for various applications, wire diameter is typically ~ 1 mm (Oxford Superconducting Technology)

Equation (14) is the fundamental reason why most superconductors for magnets are made as fine filaments of size < 50 mm. Because individual filaments of this size are difficult to handle, conductors are always made in the form of multifilamentary composites, i.e. many fine filaments of superconductor embedded in a matrix of normal conductor. If the normal conductor has low resistivity, (e.g. copper) it can provide an additional stabilization against flux jumping by slowing down the motion of flux and by conducting away the resulting heat. This effect is known as *dynamic stabilization*. A theory similar to that outlined above [16,17] gives a maximum stable radius for a round filament embedded in normal metal as

$$a_{DFJ} = \left\{ \frac{8k(\theta_c - \theta_o).(1-\lambda)}{\lambda J_c^2 \rho} \right\} \tag{16}$$

where k is the thermal conductivity of the normal metal, ρ is its resistivity and λ is the fraction of superconductor in the composite. Typical numbers give filament sizes similar to the adiabatic criterion.

3.4 Filamentary composites

Figure 17 shows a selection of filamentary NbTi superconductors which have been produced commercially for various applications. The matrix material is almost invariably copper, which is easy to process with NbTi and, as shown by Eq. (16), contributes towards dynamic stabilization of flux jumping by virtue of its low resistivity and high thermal conductivity (both of which improve at low temperatures). However, the use of a high conductivity matrix does create one significant problem, it couples the filaments together magnetically. Figure 18 illustrates this by showing two composites; the first has an insulating matrix, where the filaments behave quite independently. In the second however, which has a conducting matrix, screening currents are induced which cross over the matrix and are thus able to screen more of the external field than currents which 'go and return' in the same filament. Over a period of time, these currents will decay, but a simple calculation serves to show that, for a magnet conductor of 100's metres length, the time will be years.

The result of coupling is to make the composite behave as a single large filament, the advantages of subdivision are lost and the whole composite once again becomes unstable to flux jumping. Fortunately this problem has a simple solution: twist the composite as shown in Fig. 19. A changing external field still induces screening currents to flow, but they must now cross the matrix twice per twist pitch, thereby encountering sufficient resistance to limit their amplitudes. The screening currents flow along paths as shown by the dashed lines in Fig. 19, i.e. along a filament and vertically down (parallel to the field) through the matrix. It may be shown [15] that the effect of these currents is to reduce the field inside the composite:

$$B_e - B_i = \dot{B}_i \tau \tag{17}$$

where B_e and B_i are the internal and internal fields, B'_i is the rate of change of internal field and

$$\tau = \frac{\mu_o}{2\rho_t} \cdot \left\{ \frac{L}{2\pi} \right\}^2 \tag{18}$$

where ρ_t is the effective transverse resistivity across the matrix and L is the twist pitch. It is interesting to note that τ is just the classical time constant for magnetic diffusion a distance $L/(4\sqrt{2})$ into a bulk conductor of resistivity ρ_t. Equation (17) gives us a criterion for stability: if the screened field (B_e - B_i) is less than B_{FJ} of Eq. (15), coupling between the filaments will not cause flux jumping.

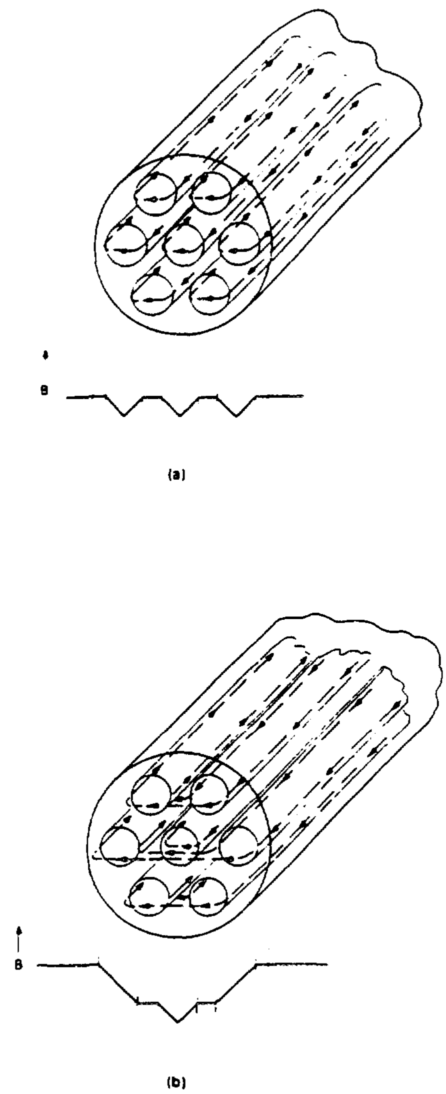
3.5 Stability against mechanical movement

Flux jumps, although originally the most serious, are unfortunately not the only source of degradation in magnets. Another problem comes from the possibility of sudden release of mechanical energy within the magnet as the current and field are increased. To get an idea of the orders of magnitude, consider a conductor in a field B, carrying current density J in superconductor which occupies a fraction l of the cross section. If this conductor moves a distance d under the influence of the Lorentz force, the work done per unit volume of conductor is:

$$W = B\lambda J\delta \tag{19}$$

The mostly likely outcome is that this work will appear as frictional heating. Taking some typical values of B = 6 T, J = 1.5 x 10^9 A m^{-2} and l = 0.3, we find that a movement of just 10 mm releases heat Q = 2.7 x 10^4 J m^{-3} which for a typical mix of copper and NbTi, will raise the temperature by ~3 K. Sudden movements of this kind are thus quite likely to quench the magnet.

The problems of designing magnets to minimize the possibility of movement under electromagnetic stress have already been dealt with in other lectures [18]. Suffice to say here that it is quite difficult to engineer a magnet winding against movements of ~10 µm and that attempts do so often result in other sources of mechanical energy release. For example, if a substance like epoxy resin is used to fill little holes in the winding, there is a danger that the epoxy may crack under stress (almost all polymers are brittle at 4.2 K) and the process of cracking can release energies at least as great as the 10 µm movement.



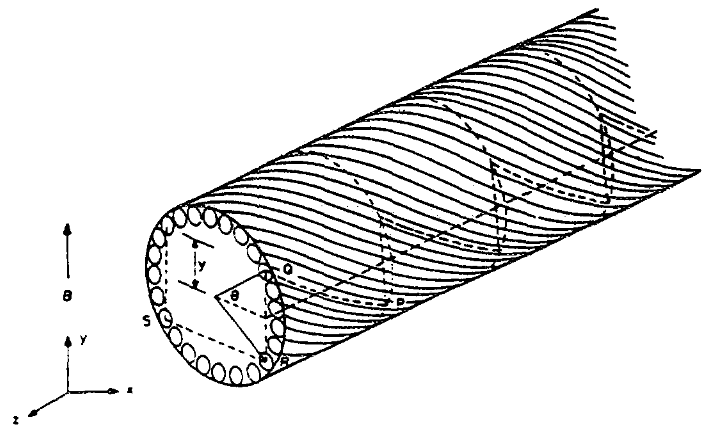


Fig. 19 Twisted filamentary composite schematic

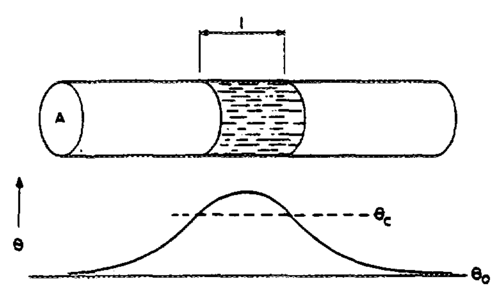


Fig. 18 Composite conductor (a) with an insulating matrix (b) with a conducting matrix which couples the filaments together

Fig. 20 A minimum propagating zone

In general we may conclude that conductors in magnets will be subject to mechanical energy inputs as the current and field are increased. What most of these mechanical inputs seem to have in common is that they are sudden and fairly localised. In thinking about how to design a conductor to cope with such disturbances, two very useful concepts are the ideas of the Minimum Propagating Zone MPZ [19] and Minimum Quench Energy MQE. Figure 20 illustrates the idea. It is assumed that a sudden localized energy input has created a resistive zone of length l in a conductor of cross sectional area A carrying current density J in the superconductor which occupies a fraction l of the cross section, the remainder being copper of thermal conductivity k and resistivity r. If the zone is in steady state we may equate the heat generated within it to the heat conducted out, very approximately:

$$\frac{2k\,A(1-\lambda).(\theta_c-\theta_o)}{l} \approx \frac{\lambda^2 J^2 \rho A l}{(1-\lambda)} \qquad (20)$$

and

$$l \approx \left\{ \frac{2k(1-\lambda)^2.(\theta_c - \theta_o)}{\lambda^2 J^2 \rho} \right\}^{1/2} \tag{21}$$

Thus l is the length of a normal zone in which the ohmic heat generated is just equal to the heat conducted out. It is a state of unstable equilibrium, zones slightly smaller will collapse and the superconducting state will be fully recovered, zones slightly larger will grow without limit and quench the magnet. The energy needed to set up the MPZ is the MQE — very approximately for a conductor of cross sectional area A:

$$\text{MQE} = Al\gamma\ C(\theta_c - \theta_o) = A\gamma\ C(\theta_c - \theta_o)^{3/2} \frac{(1-\lambda)}{\lambda J} \left\{ \frac{2k}{\rho} \right\}^{1/2} \tag{22}$$

Clearly a large MQE is to be preferred because it means that the magnet will not quench so easily. Although Eq. (22) is only very approximate, it does give a useful idea of the general scaling laws and the factors which should be optimized, i.e. a matrix with low ρ, high k and γC.

A good way of increasing γC is to make the winding slightly porous to liquid helium, because it has by far the highest heat capacity at this temperature. Most accelerator magnets are made this way. Obviously reducing λ and J would also increase MQE, but this is usually the last thing a magnet designer wishes to do!

3.6 Magnetization

The topic of persistent currents and field distortion will be covered in another lecture [20]. Here we only mention that the screening currents shown in Fig. 13 are equivalent to a magnetization:

$$M_f = \mu_o J_c \frac{a}{2} \tag{23}$$

This magnetization causes field error in the magnet aperture and should therefore be minimized. For this reason, it is often desirable to make the filament radius a much less than the size needed for stability against flux jumping, for example, accelerator magnet conductors generally have filament diameters of 5 – 10 mm.

Another source of magnetization is the coupling currents shown in Fig. 19, which give rise to an additional component of magnetization:

$$M_c = 2B_i'\tau \tag{24}$$

Here again, for accelerators it may be desirable to minimize M_c in the ramping field by twisting the conductor more tightly than needed for stability against flux jumping.

In some conductors with extremely fine filaments, a new effect has come to light whereby the copper matrix between the filaments becomes slightly superconducting due to the proximity effect. This means that the filaments are coupled together by superconducting currents [21] which can increase the magnetization to the extent of causing unacceptable field errors in accelerator magnets at injection. The effect may be controlled by increasing the interfilament spacing or by adding small amounts of Ni, Si or Mn to the copper matrix [22].

4. MANUFACTURE OF MAGNET CONDUCTORS

4.1 Summary of requirements

Before moving on to the practical manufacture of magnet conductors, we begin by recapitulating some of the requirements outlined in the previous sections.

a) Critical field and temperature: depend on the material chemistry, it is important to get the optimum composition for alloys and the correct stoichiometry for compounds, ternary additions can help

b) Critical current: depends on the microstructure, a Ti precipitates on dislocation cell boundaries in NbTi, grain boundaries in Nb_3Sn

c) Mechanical properties: only NbTi is ductile, Nb_3Sn is brittle, better in filamentary form when supported by matrix, but still strain sensitive, Nb_3Al is less so

d) For cryogenic stability, the superconductor must be combined with large quantities of copper or aluminium — only suitable for low current density magnets

e) For stability against flux jumping, one must make a filamentary composite wire with filament diameters < 50 mm

f) To decouple the filaments, the wire must be twisted

g) For low magnetization (field distortion), the filaments must be ~ 5 – 10 mm diameter

h) To provide dynamic stability against flux jumping and also to maximize the MQE, the conductor must have a low resistivity, high thermal conductivity matrix. As discussed in a later lecture [12], low resistivity is also necessary for quench protection.

4.2 Niobium titanium

Figure 21 summarizes the main steps in the production of filamentary NbTi composite wire. The cylindrical starting billet of NbTi is prepared by consumable arc vacuum melting. Production of the alloy is outlined in [4]; it is important to make the alloy composition as homogeneous as possible over the whole billet. The NbTi billet is machined to size, cleaned and fitted inside a copper extrusion can. If the composite is destined to have very fine filaments, a thin diffusion barrier of pure niobium may be interposed between the NbTi and copper. The purpose of this barrier is to prevent the formation of a CuTi intermetallic phase during the heat treatments which are applied at intermediate stages throughout the manufacturing process. The intermetallic, which is hard and brittle, does not draw down with the filament but instead breaks up into hard particles. At the later stages of drawing, when the filament size becomes comparable with these particles, the filament is broken by the particle. If the composite is destined for very fine filament size, it may be better to make the can of copper doped for example with Mn, to suppress proximity effect coupling at final size. For one of the new experimental APC composites, the Cu or Nb pinning centres may be inserted as rods into holes drilled in the NbTi billet.

Assembly of the billet should be carried out in clean conditions, after which it is evacuated, sealed by electron beam welding, heated and extruded. After cold drawing to size, the rod is drawn through a hexagonal die and then cut into many lengths. After cleaning, these lengths are stacked into another copper can which is again sealed, extruded and drawn down to final size. For accelerator magnets, which may have up to 10^4 filaments, a double stack process is often used in which the rods are again drawn into hexagonal sections and stacked in another can. Certain numbers and arrangements of filaments are preferred for fitting into the round can. In wires intended for cabling, it has been found that a central core of pure copper is beneficial in helping the wire to resist the deformation of the cabling process.

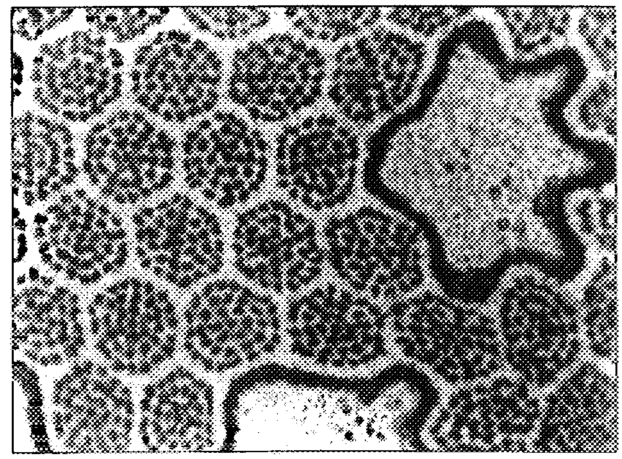


Fig. 22 Cross section of a bronze route Nb_3Sn composite after reaction heat treatment showing the fine filaments of Nb_3Sn and the islands of pure copper surrounded by a diffusion barrier

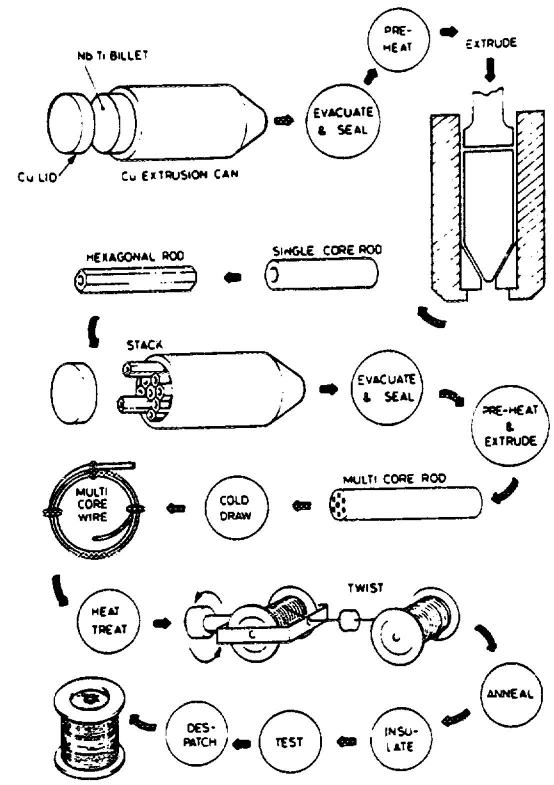


Fig. 21 Production of filamentary NbTi/copper composite wire

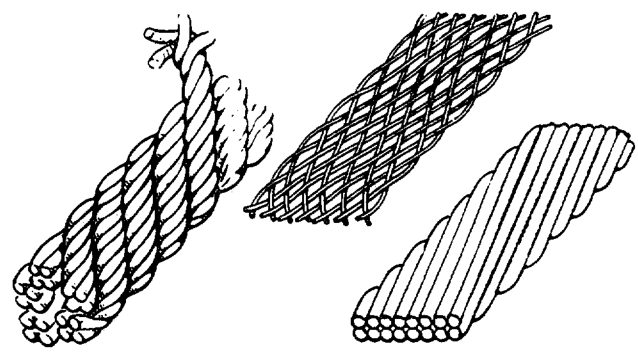


Fig. 23 Three different types of transposed twisted cable—rope, braid and Rutherford cable

Multiple heat treatments are applied throughout the process in a defined sequence of alternating cold work and heat treatment, which has been found to produce the best configuration of α Ti precipitate and hence the best flux pinning. After reaching final size, the wire is twisted, typically with a twist pitch of 25 mm, but tighter if destined for use in high ramp rate fields

4.3 Niobium tin

Production of filamentary Nb_3Sn by the bronze route [4] is similar to the process described above for NbTi, with pure Nb filaments in a matrix of CuSn bronze. For optimum current density at high field, the Nb is often doped with another metal, most commonly Ti. In one variation of the multi-stacking process, many niobium rods are put into holes drilled in a single bronze billet, which is then extruded, drawn and twisted as before. The reaction heat treatment is usually applied after the wire has been wound into its final coil shape. A typical heat treatment is in vacuum at ~ 700°C for about a week.

Many variants of the bronze process have been tried. For the internal tin process, holes larger than the filaments are drilled in the extrusion billet and are filled with pure tin or tin doped pure copper. In the so-called jelly-roll process, the niobium is made in the form of a sheet of expanded metal. It is rolled up with a sheet of bronze — like a jelly or Swiss roll — and then packed and sealed in a can for extrusion and drawing down in the usual way. One problem with both these processes is a tendency for the filaments to be joined together after reaction heat treatment, thereby increasing the magnetization and reducing the stability against flux jumping.

In the ECN [8] process, $NbSn_2$ powder of particle size < 3 μm is mixed with tin powder and packed into a thin-walled copper tube. This tube is then reduced somewhat in diameter to densify the powder and is put into an Nb tube surrounded by a copper tube. The assembly is extruded and drawn down to hexagonal section, which is then cut into lengths and multi-stacked in a copper can and re-extruded. The copper liner of the Nb tube is there to ensure that Nb_3Sn forms at low temperature (via the bronze process) during the final heat treatment.

An advantage of the ECN process is that, provided the tin does not react right through the Nb tube, the outer copper stays pure, thereby fulfilling requirement 4.1(h). In simple bronze composites, the bronze remaining after reaction still contains sufficient tin to give it a high resistivity at 4.2 K, no use for stabilization or quench protection. Additional islands of pure copper must therefore be included in the bronze matrix. To protect the copper from tin contamination during reaction, the islands must be surrounded by an inert barrier, usually Ta, as shown in Fig. 22.

4.4 Cabling and insulation

A single composite wire of ~ 1 mm diameter will typically carry ~ 500 A in field, but accelerator magnets are designed to work at 5000 – 10000 A. It is therefore necessary to use many wires in parallel. To ensure that the wires share current equally, they must be combined in the form of a twisted, fully transposed cable (remember that, with no resistance, current sharing is dominated by inductive effects). Figure 22 shows three different styles of transposed cable, all of which have been used in accelerator magnets. Because of its superior mechanical properties, the Rutherford cable has become predominant and has been chosen for all accelerators constructed or proposed so far.

Figure 24 shows a schematic of the machine used for making Rutherford cable. Between 20 and 40 strands are twisted together around the mandrel which is shaped rather like a screwdriver. As the cable leaves the blade of the screwdriver, it enters the 'Turk's-head' roller die, which consists of four rollers arranged at right angles to form a rectangular orifice. The

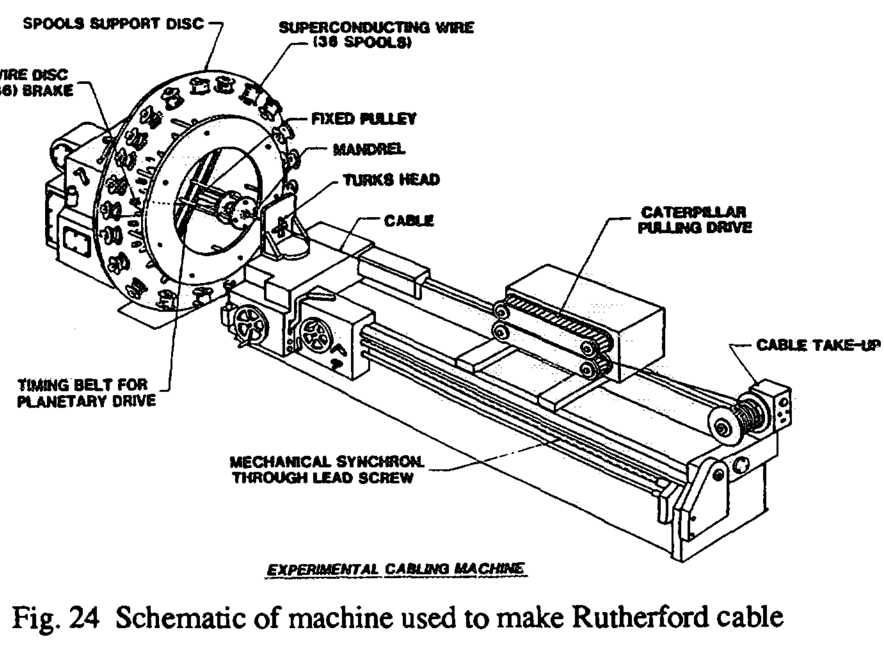


Fig. 24 Schematic of machine used to make Rutherford cable

Turk's head die squeezes the cable into its final configuration, which typically has the cross section shown in Fig. 25. Note that this cable has been given a 'keystone' shape by angling the rollers slightly so that turns will fit together around the circular magnet aperture.

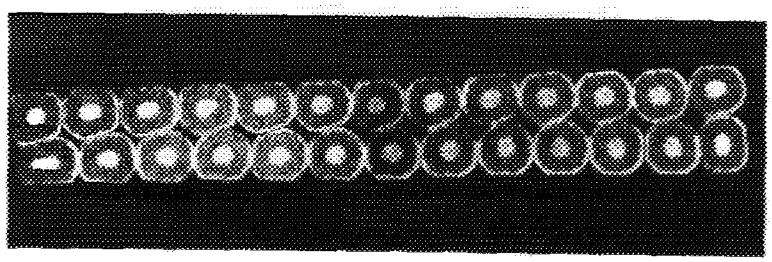


Fig. 25 Cross section of the LHC inner cable

Rutherford cable is usually kept slightly porous in magnet windings so that the helium may penetrate to provide additional stability by increasing the MQE. The insulation scheme sketched in Fig. 26 is designed to maintain some porosity; it consists of two layers of Kapton tape and an outer layer of glass-fibre tape impregnated with 'B stage' epoxy resin. After winding, the coil is compressed and heated to consolidate the turns and reduce the possibility of mechanical movement, but surprisingly it still remains slightly porous.

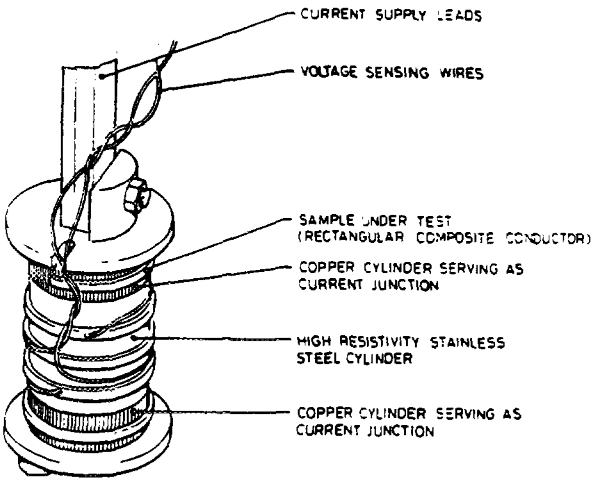


Fig. 26 Sample holder for measuring the critical current of wires or cables

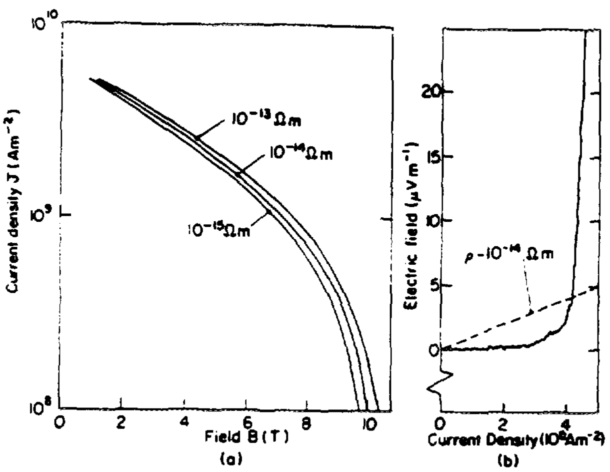


Fig. 27 Typical result of a short sample measurement showing (a) J_c at three levels of effective resistivity and (b) a typical transition

NbTi wires intended for use as single strands are usually insulated by a standard electrical varnish process (as used routinely for small motors and transformers) typically by applying 5 – 10 coats of polyvinyl acetal varnish. For niobium tin, which must be reacted after winding, the usual insulation is a braid of dry glass fibre, which is woven around the wire. After reaction, the whole winding is vacuum impregnated with epoxy resin.

4.5 Measurements

For quality control in production and to ensure that magnets perform as expected, routine measurements must be made on samples of superconducting wire and cable. The simplest and most common of these measurements is the critical current versus field characteristic at constant temperature, as used in Fig. 10. Figure 26 shows a sample holder used at Rutherford

Laboratory and elsewhere for measuring critical current; the spiral sample is soldered to current lead contacts at the top and bottom, voltage is measured across two turns of the spiral, the holder fits inside the bore of a superconducting solenoid.

For filamentary composites, it is found that the transition to the resistive state is not sharp, but progressive as shown in Fig. 27 (b). For this reason, it is usual to plot the critical current at a certain level of effective resistivity as shown in Fig. 27 (a). By considering the thermal environment inside a magnet, one may calculate the maximum resistivity which can be sustained within the winding before thermal runaway occurs [15]. For most windings, this resistivity is ~ 10^{-14} Ωm and this value has accordingly become the accepted standard for critical current. Empirically it has been found that the resistive transition may be fitted by a simple power law:

$$\frac{\rho}{\rho_o}=\left\{\frac{J}{J_o}\right\}^n \tag{25}$$

where ρ_o and J_o are defined at some arbitrary value, usually taken to be 10^{-14} Ωm. The index n is often taken as a measure of quality; it is high for composites in which the filaments are uniform along their length and low for non-uniform composites in which the filaments show 'sausaging' along their length.

Less frequently performed, but nevertheless valuable in providing additional data about critical current and filament coupling, is the measurement of magnetization. Figure 28 shows the apparatus, which consists of a pair of balanced search coils inside the bore of a superconducting solenoid. The coils are connected in series opposition and the small trim coil is rotated to compensate for any residual differences so that, with no sample in the coils, the output is zero when the magnetic field is ramped. Then, with a sample in one of the coils, the integrated output provides a measure of sample magnetization. By calibration with a known sample or by computation of the geometry, the signal may be made absolute. Magnetization gives a measure of the field distortion which will be produced by the conductor when wound into a magnet. It gives an indication of the coupling between filaments and of the effective filament size (which may be in doubt for some Nb_3Sn composites). It may also be used as a measure of J_c at very low fields, where stability and self field problems can make it difficult to perform direct transport-current measurements.

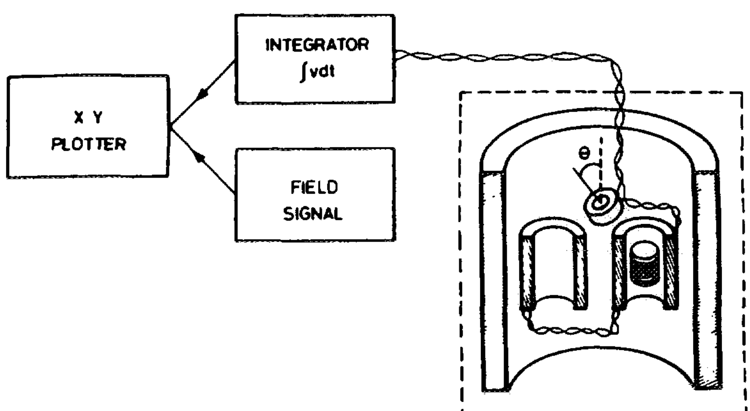


Fig. 28 Apparatus for measuring the magnetization of superconducting samples

5. CONCLUDING REMARKS

The two technical superconductors NbTi and Nb_3Sn have been the subject of continuous development over the last 30 years. Great improvements have been made in the critical current density of these materials and the technology of utilization has also been developed by making fine filaments, cables etc. Because of its excellent mechanical properties, NbTi is by far the most popular choice for all applications below ~ 10 T, but Nb_3Sn is now being used routinely for higher field magnets, particularly in NMR spectroscopy and general research. All accelerators constructed or planned so far have used NbTi, but the push towards higher fields continues and Nb_3Sn is now being seriously assessed as a candidate material for future accelerators. Another large scale potential user of Nb_3Sn is the proposed tokamak reactor ITER, which has recently been driving much of the development work on high field materials. One interesting outcome from this work has been the renewed interest in Nb_3Al, stimulated by its much lower strain sensitivity. Nb_3Al has recently been achieving current densities similar to Nb_3Sn, with slightly higher critical field and temperature. Even the old faithful NbTi has recently been exciting new interest, with the realisation that artificial pinning centres offer the possibility of substantial increases in critical current density. In the light of this continuing development effort, there seems little doubt that the performance of the low temperature Type II superconductors will continue to improve steadily. However, it seems more likely that any spectacular developments at high field will come from the new high temperature materials.

REFERENCES

[1] G.M. Muller, 'High Temperature Superconductors', these proceedings.

[2] P. Schmüser, 'Superconductivity', these proceedings

[3] B.B. Goodman Rep. Prog. Phys. **39** (2) pp 445 (1966)

[4] P.J. Lee et al, Chapter 5 in 'Composite Superconductors' ed. K. Osamura, pub Marcel Dekker 1994

[5] P.J. Lee, D.C. Larbalestier, J.C. McKinnell, McInturff AD: IEEE Trans Appl Supercon. Vol 3 No. 1 pp 1354 (1993)

[6] D.C. Larbalestier, 'New Developments in Superconductors for Particle Accelerators', Proc. 1995 Particle Accelerator Conference, to be published.

[7] D.C. Larbalestier, P.E. Madsen, J. Lee, M.N. Wilson, J.P. Charlesworth, IEEE Trans MAG-11 pp 247 (1974)

[8] N. Higuchi, K. Tsuchiya, C.J. Klamut, M. Suenaga, Adv Cryo Eng Vol 30 pp 739 (1984)

[9] E.M. Hornsveld, J.D. Elen, C.A.M Van Beijnen, P. Hoogendam, Adv Cryo Eng Vol 34 pp 493 (1987)

[10] J.W. Ekin, Adv Cryo Eng Vol 24 pp 306 (1978)

[11] R. Flukiger, R. Isernhgen, W. Goldacker, W. Specking, Adv Cryo Eng Vol 30 pp 851 (1984)

[12] Y. Iijima, M. Kosuge, T. Takeuchi, K. Inoue, Adv Cryo Eng Vol 40 pp 899 (1994)

[13] K.H. Mess, 'Quench Protection', these proceedings

[14] Z.J.J. Stekly, J.L. Zar, IEEE Trans NS 12 pp 367 (1965)

[15] T.Akachi, T. Ogasawara, K. Yasukochi, Jap Jnl App Phys 20 (8) pp 1559 (1981)

[16] Superconducting Applications Group, Rutherford Laboratory, Brit Jnl Phys D (App Phys) Vol 3 pp 1517 (1970)

[17] Wilson MN, Superconducting Magnets, pub Oxford University Press ISBN 0-019-854805-2

[18] R. Perin, 'Fields, forces and mechanics of SC magnets', these proceedings

[19] S.L. Wipf and A.P. Martinelli, Proc 1972 Appl. Supercon. Conf. pp 331, Annapolis (1972)

[20] A. Devred, 'Persistent and eddy currents', these proceedings

[21] A. Harada, Y. Mawatari, O. Miura, Y. Tanaka, K. Yamafuji, Cryogenics Vol 31 pp 183 (1991)

[22] E. Gregory, H. Liu, J.M. Seuntjens, Adv Cryo Eng Vol 40 pp 779 (1994)

BIBLIOGRAPHY

Proc Applied Superconductivity Conference, IEEE Trans Applied Superconductivity, Mar 93, Mar 95? and IEEE Trans Magnetics Mar 91,89 87,85,etc

Advances in Cryogenic Engineering, Plenum Press, materials are in Vols 30, 32.........40.

Proc. Magnet Technology Conference, pub as IEEE Trans Magnetics, biannually from Sept 81 onwards (previous pub stand alone)

Y. Iwasa, Case Studies in Superconducting Magnets, Plenum Press, New York 94, ISBN 0-306-44881-5.

M.N. Wilson Superconducting Magnets, Oxford University Press 83 ISBN 0-019-854805-2

CERN Accelerator School: Superconductivity in Particle Accelerators, Hamburg 88, CERN 89-04.

Cryogenics: published monthly by Butterworths

Applied Superconductivity: published bi-monthly by Pergamon

Superconductor Science and Technology, published monthly by Institute of Physics (UK).

D.R. Tilley and J. Tilley, Superfluidity and Superconductivity: Adam Hilger, 86, ISBN 0-85274-791-8

A.C. Rose-Innes and E.H. Rhoderick, Introduction to Superconductivity, Pergamon, 78, ISBN 0-08-021652-8

G. Vidali, Superconductivity, the Next Revolution? Cambridge University Press, 93, ISBN 0-521-37757-9

NEXT PAGE(S)
left BLANK

FIELD, FORCES AND MECHANICS OF SUPERCONDUCTING MAGNETS

Romeo Perin
CERN, Geneva, Switzerland

Abstract
The chapter starts with a brief recall of the peculiar characteristics of superconducting magnets for accelerators, and of the evolution in performance of the main magnets of the most important projects. An outline of the concepts and methods used in the magnetic and mechanical design of the magnets is then presented together with some examples of applications and indications of the manufacturing accuracies necessary to achieve the required performance and field quality.

1. INTRODUCTION

Superconducting magnets have become essential components of hadron accelerator/colliders and of compact electron accelerators. For almost three decades high-energy physics has been the prime promoter of their development as the means to reach higher beam energy while minimising cost and space requirements.

There are some fundamental differences between superconducting magnets for accelerators and other superconducting magnets (e.g. large solenoids for particle detectors, toroidal coils for nuclear fusion machines, etc.) that make their design and construction a very special branch of the technology:

a) the need to use very high current densities to economically produce the required high bending/focusing fields;

b) the more complex and non-uniform repartition of electro-magnetic forces;

c) the extreme precision of the magnetic field distribution in small apertures;

d) the high degree of reproducibility and reliability.

Some of these characteristics make the task of building superconducting accelerator magnets singularly complex and difficult. In particular:

point a precludes cryogenic stabilization;

point b leads to elaborate mechanical structures;

point c demands an unprecedented dimensional precision in components and fabrication;

point d imposes fabrication methods adapted to large scale industrial production.

As in circular machines, the attainable beam energy is proportional to the machine radius and bending magnetic field, there is a great incentive to develop dipoles for higher and higher fields.

As a consequence the dipole magnets, which are the most important components of the accelerators from the cost point of view, also become the most critical technologically. In parallel, the focusing elements (quadrupoles) have to become stronger while being kept as short as possible.

The evolution in field and field gradient of superconducting main magnets for accelerators is recalled in Table 1.

Table 1
Design parameters of superconducting main magnets in accelerator/colliders

	Dipoles			Quadrupoles			Operation
	Central field (T)	Coil aperture (mm)	Eff. unit length (m)	Field gradient (T/m)	Coil aperture (mm)	Eff. length (m)	temperature (K)
TEVATRON	4.4	76.2	6.1	75.8	88.9	1.7	4.6
HERA	4.7	75	8.8	91.2	75	1.9/1.7	4.5
RHIC	3.5	80	9.5	71.8	80	1.1	4.6
UNK	5.0	80	5.7	96.1	80	3.0	4.6
SSC	6.6	50	15.2	206	40	5.2	4.35
LHC	8.4	56	14.2	223	56	3.1	1.9

Improvements in performance of superconductors, better insulation systems, force containment structures and refinements in manufacturing have permitted the field and gradient to be increased. A bold step is being made with the LHC using the superfluid helium technique, thus enhancing the performance of the traditional NbTi superconductor, and by the adoption of the two-in-one configuration leading to a considerable reduction of costs and size.

In modern proton or heavier particle high-energy accelerators the share of the magnet system in the cost of the facility is important. In the case of the LHC [1], which will be installed in the existing LEP tunnel, it is predominant (Fig. 1), the main dipoles and quadrupoles taking the largest share (Fig. 2). This chapter will mainly deal with these two types of magnets. The design and construction of the corrector magnets which are in general less demanding and can, therefore, be dimensioned with larger margins, follow the same pattern.

For a general introduction to superconducting magnets the book "Superconducting Magnets" by M.N. Wilson [2] is recommended. More specifically on accelerator magnets, an excellent treatment may be found in the lectures given by K.H. Mess and P. Schmüser at the CERN-DESY Accelerator School in 1988 [3]. An outline on high-field accelerator magnets can be found in "New Techniques for Future Accelerators III" [4].

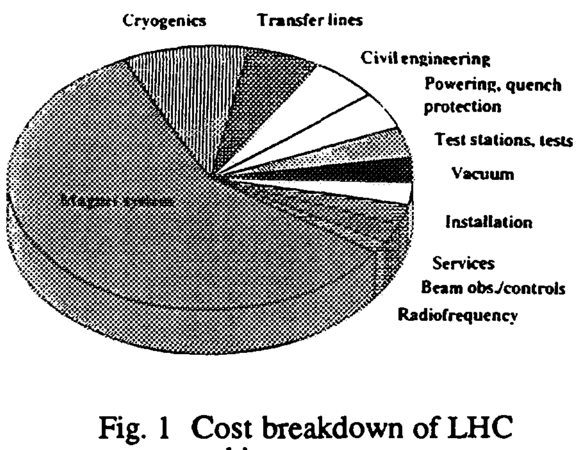


Fig. 1 Cost breakdown of LHC machine components

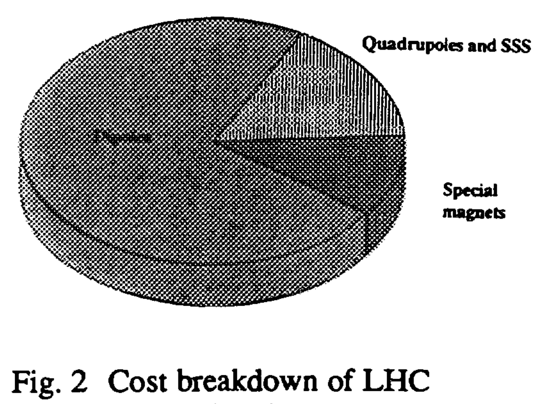


Fig. 2 Cost breakdown of LHC magnet system

2. CATEGORIES OF SUPERCONDUCTING MAGNETS FOR ACCELERATORS

Three different classes of magnets can be characterized with respect to the way in which the wanted field quality is achieved:

a) Magnets in which the field distribution is dominated by the coil configuration. The ISR and LEP quadrupoles, the Tevatron, HERA, RHIC, (SSC), LHC main dipoles and quadrupoles pertain to this class (Fig. 3).

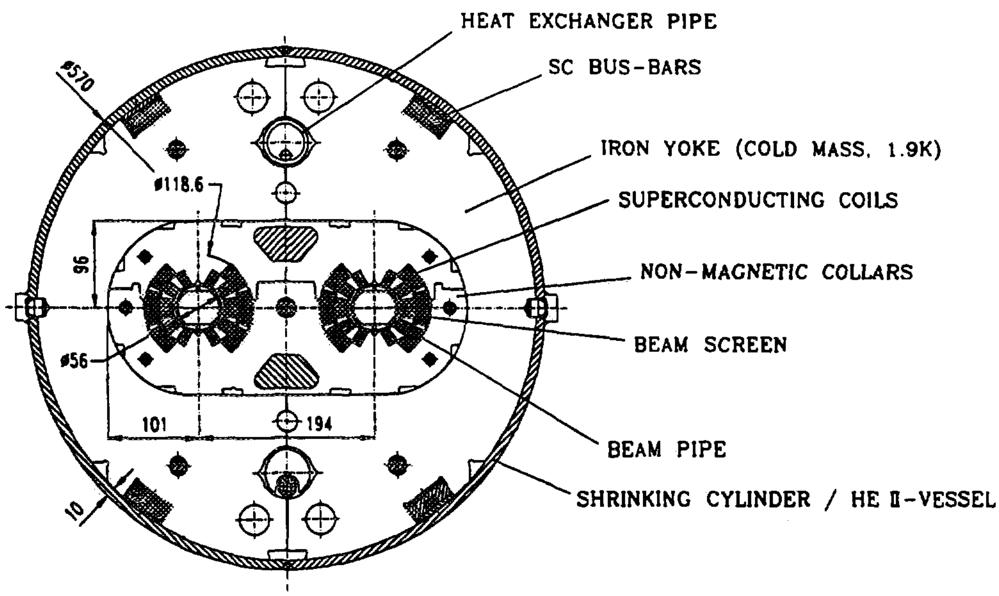


Fig. 3 Example of coil dominated magnet: the LHC dipole [1]

b) Iron dominated magnets, also called superferric, in which the iron (steel) pole shape determines the field pattern (examples of these are the HERA correctors, the RHIC sextupoles [5], etc.) (Fig. 4).

c) Magnets in which the field distribution results from the interplay of coils and yoke, both strongly contributing to produce the required field. Examples are some model magnets built at TAC and the configurations recently proposed by Texas A&M University and LBL of the "Block-Coil Dual Dipole" and of the "Pipe" magnet [6] (Fig. 5).

3. MAGNETIC DESIGN

3.1 General

In beam guiding magnets, the problem is essentially bidimensional because, apart from the localized regions of the ends, the field distribution has to be the same in all planes perpendicular to the beam axis. Type b) magnets are designed and built much in the same way as classical resistive magnets, the main difference being that they allow much higher current densities in the windings. So, for the magnetic design, use is made of the well developed finite differences and finite elements computer codes. For type c), the same methods apply, since the iron yoke plays an important role.

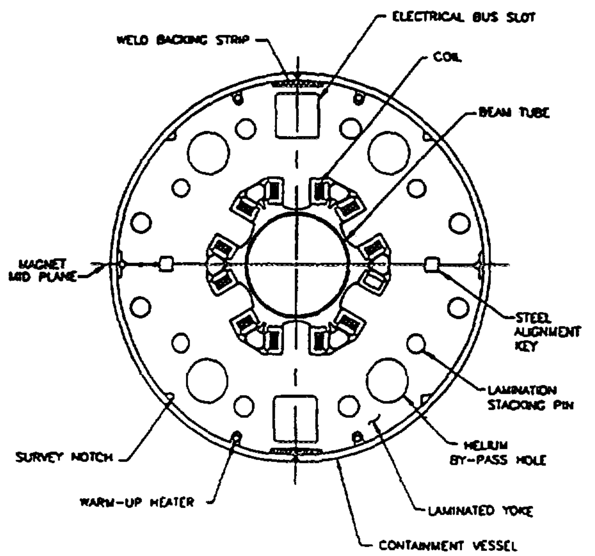


Fig. 4 Example of iron-dominated (super ferric) magnet: the RHIC arc sextupole

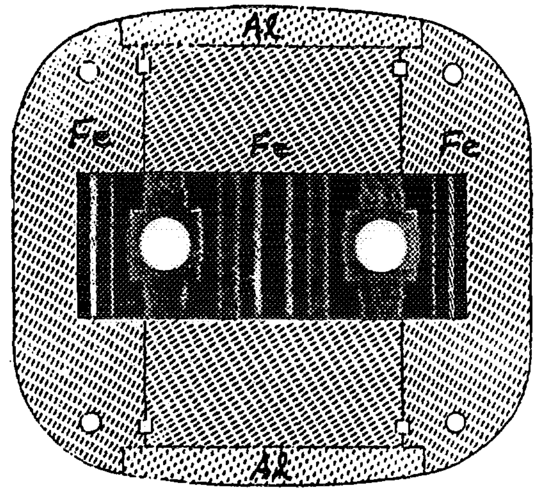


Fig. 5 Example of "mixed" magnet in which the field configuration is determined by conductor and iron: the Block-Coil Dual Dipole

In this lecture we treat only the magnetic design of type a) magnets as it is more typical of superconducting accelerator magnets. Generally the design starts with *analytical methods*, integrating the Biot-Savart law which describes the field (induction) induced by a current flowing through a (infinitely) thin wire:

$$\vec{B} = \frac{\mu_o}{4\pi} \, I \int_c \vec{d\ell} \times \frac{\vec{r}}{r^3} \tag{1}$$

When the windings are not of simple geometry, they are subdivided into small regular parts whose contributions to the field can be computed analytically and summed by means of relatively simple computer programs.

In this first phase of the design, iron is in general assumed to have infinite magnetic permeability ($\mu = \infty$). Only in a second phase does one analyse the effects of the real iron characteristics, i.e. remanence, variable permeability and saturation. This is done by means of computer programs solving the partial differential equations by finite differences or finite elements numerical methods (e.g. POISSON, MARE/MAGNET, TOSCA [7–9], etc.).

3.2 Ideal current distributions generating typical field configurations for accelerators (dipole, quadrupole, sextupole, etc.) and practical approximations

A line current distributed on a circular cylindrical surface as a function of the azimuthal angle ϑ according to:

$$I(\vartheta) = I_o \cos(n\vartheta) \tag{2}$$

generates a pure 2 n-pole field inside the circular aperture (Fig. 6).

Pure 2 n-pole fields are also obtained by means of uniformly distributed currents flowing in conductors of particular shape. The simplest of these ideal shapes, producing a dipole field, is formed by the intersection of two cylinders with their centres spaced apart and carrying oppositely directed currents with uniform density.

More generally pure multipole fields can be obtained in the aperture formed by the overlap region of equal conductors of elliptical cross-section suitably spaced and having uniform current densities of convenient sign. In fact, it has been shown by Beth [10] that the internal field of an elliptical conductor with semi-axis a and b in the x and y directions respectively and uniform current density J is:

$$\vec{B} = \vec{B}_x + \vec{B}_y \qquad B_x = -\frac{\mu o\, J}{a+b}\, a\, y \qquad B_y = \frac{\mu o}{a+b}\, b\, x \tag{3}$$

Examples are shown in Fig. 6. If two ellipses are aligned along their major axis, a pure dipole is produced, if they overlap at right angles the field in the overlap is a pure quadrupole.

The practical approximations to the above ideal configuration are shown in Fig. 7 for the dipoles and Fig. 8 for the quadrupoles.

3.3 Design of the coil configuration using an analytic/numerical approach

The field distribution produced by the practical approximations to the ideal coil cross-sections can be computed by means of analytic programs, which can take into account the presence of a circular steel screen, considered as an equipotential surface ($\mu = \infty$), around the windings.

The treatment presented here follows the note of Ref. [11]. The field produced by a current flowing through the surface S (Fig. 9) perpendicularly to the x, y plane is given by the Biot-Savart law in two dimensions:

$$\vec{B} = \frac{\mu o}{2\pi} \iint \frac{\vec{J}\,(x, y) \times \vec{r}}{r^2}\, ds \tag{4}$$

where the current density $\vec{J}$ is flowing perpendicularly to the plane of $\vec{B}$ in infinitely long conductors.

Field generated by the current $I = \int_S J(x, y)\, ds$:

$$|B| = \frac{\mu o}{2\pi} \iint \frac{J\,(x, y)}{r}\, ds = \frac{\mu o}{2\pi} \iint \frac{J\,(x, y)}{\sqrt{(x-x_c)^2 + (y-y_c)^2}}\, ds \tag{5}$$

$$B_x = |B| \sin\alpha = -\frac{\mu o}{2\pi} \iint \frac{J(x, y)}{(x-x_c)^2 + (y-y_c)^2} .(y-y_c)\, ds \tag{6}$$

$$B_y = |B| \cos\alpha = \frac{\mu o}{2\pi} \iint \frac{J(x, y)}{(x-x_c)^2 + (y-y_c)^2} .(x-x_c)\, ds \tag{7}$$

The treatment is particularly simple using the analytic function* F

$$\vec{F} = B_y + iB_x = \frac{\mu o}{2\pi} \iint \frac{J(x, y)\,(x-x_c)}{(x-x_c)^2 + (y-y_c)^2}\, ds + i\,(-\frac{\mu o}{2\pi}) \iint \frac{J(x, y)\,(y-y_c)}{(x-x_c)^2 + (y-y_c)^2}\, ds \tag{8}$$

$$= \frac{\mu o}{2\pi} \iint \frac{J(x, y)}{z - z_c}\, ds \qquad \text{valid for all } z \neq z_c$$

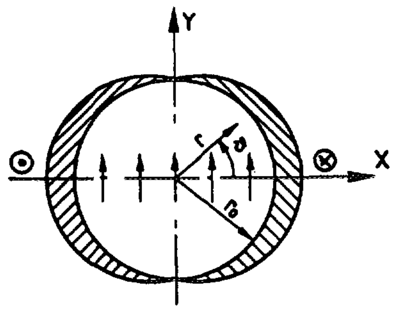
If a circular steel screen with $\mu = \infty$ is present around the coil, its contribution can be calculated with the image current method:

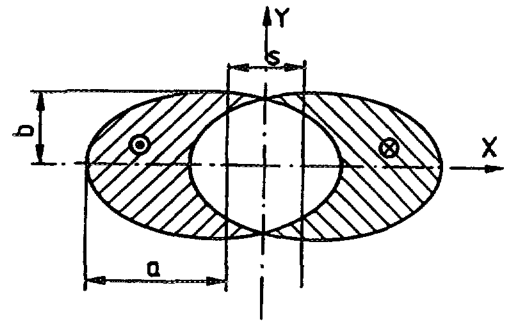
---

* Note: The magnetic field (induction) $\vec{B} = B_X + iB_Y$ is not an analytic function as it does not satisfy the Cauchy-Riemann conditions.

Cos nϑ | **Intersecting ellipses**

Dipole





$I = I_o \cos\vartheta$

$B_\vartheta = \dfrac{\mu o\, I_o}{2\, r_o} \cos\vartheta \qquad B_x = o$

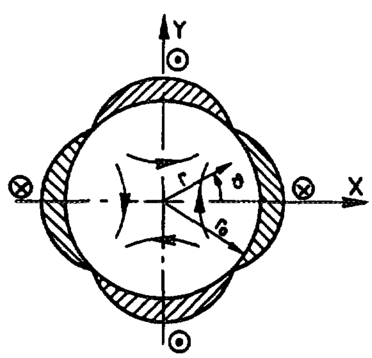
$B_\vartheta = \dfrac{\mu o\, I_o}{2\, r_o} \sin\vartheta \qquad B_y = \dfrac{\mu o\, I_o}{2\, r_o}$

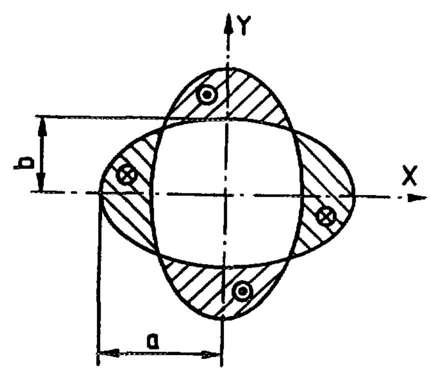
J uniform in the hatched areas

$B_x = o$

$B_y = \mu o\, J \dfrac{sb}{a + b}$

Quadrupole





$I = I_o \cos 2\vartheta$

$B_x = \dfrac{\mu o\, I_o}{2\, r_o^2}\, y$

$B_y = \dfrac{\mu o\, I_o}{2\, r_o^2}\, x$

$G = \dfrac{\mu o\, I_o}{2\, r_o^2}$

J uniform in the hatched areas

$B_x = \mu o\; J \dfrac{a-b}{a+b}\, y$

$B_y = \mu o\; J \dfrac{a-b}{a+b}\, x$

$G = \mu o\; J \dfrac{a-b}{a+b}$

Fig. 6 Ideal current distributions producing pure dipole and quadrupole fields

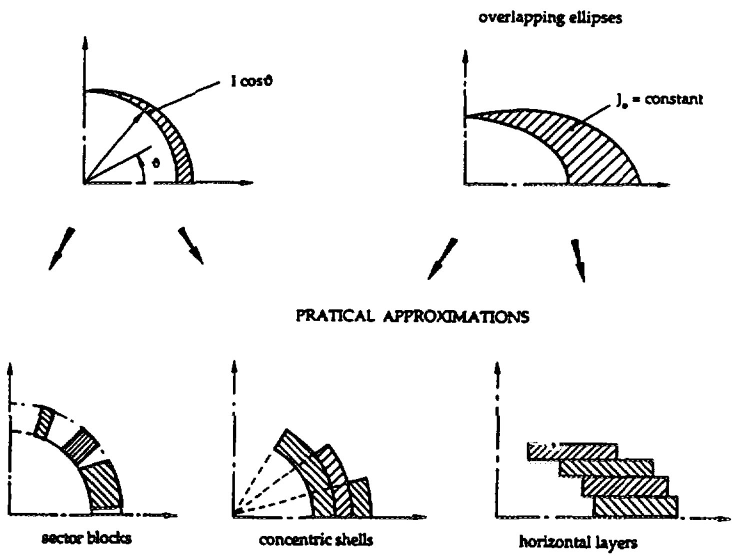


Fig. 7 Dipole ideal configurations and practical approximations

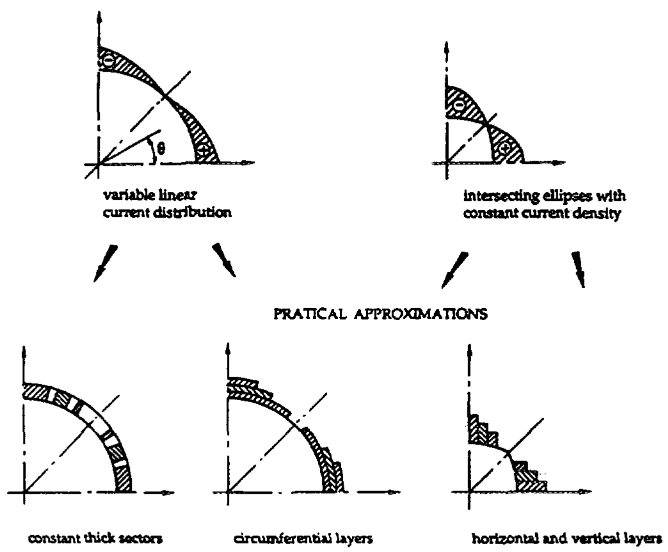


Fig. 8 Quadrupole ideal configurations and practical approximations

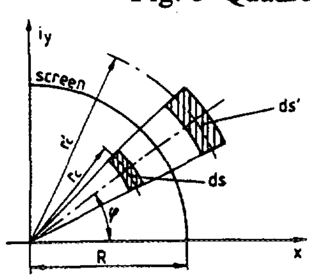


Fig. 9

J directed upwards
In the complex plane $z = x + iy$,
$z = x + iy$ coordinate of the point where the field B is calculated
$z_c = x_c + iy_c$ coordinate of conductor element ds

$$r'_c = \frac{R^2}{r_c} \qquad J' = J\left(\frac{R}{r'_c}\right)^4 \qquad ds' = \frac{R^4}{r_c^4}\, ds \qquad z'_c = \frac{R^2}{r_c}\, e^{i\phi}$$

$$F = B_y + iB_x = \frac{\mu o}{2\pi}\iint \frac{J}{z - z_c}\, ds + \frac{\mu o}{2\pi}\iint \frac{J}{z - \frac{R^2}{z_c^*}}\, ds \qquad (9)$$

$$\frac{1}{z-z_c} = -\frac{1}{z_c\,(1-\frac{z}{z_c})} = -\frac{1}{z_c}\sum_{o}^{\infty} k\,(\frac{z}{z_c})^k = -\frac{1}{z_c}\sum_{1}^{\infty}{}_n(\frac{z}{z_c})^{n-1} = -\sum_{1}^{\infty}\frac{z^{n-1}}{z_c{}^n}$$

$$\frac{1}{z-\frac{R^2}{z_c^*}} = -\frac{1}{\frac{R^2}{z_c^*}\,(1-\frac{z\,z_c^*}{R^2})} = -\frac{z_c^*}{R^2}\sum_{1}^{\infty}{}_n\,(\frac{z\,z_c^*}{R^2})^{n-1} = -\sum_{1}^{\infty}{}_n\; z^{n-1}\,\frac{z_c^{*n}}{R^{2n}}$$

$$\boxed{F = -\sum_{1}^{\infty}(c_n + f_n)\; z^{n-1}} \qquad (10)$$

with

$c_n = \frac{\mu o}{2\pi}\iint J\; z_c^{\;-n}\; ds$	Direct contribution of the coil to the n component	(11)
$f_n = \frac{\mu o}{2\pi}\iint J\;\frac{z*_c^n}{R^{2^n}}\; ds$	Contribution of the iron screen to the n component.	(12)

<u>For circular sector windings</u> defined by the radii r_1 and r_2 and the angles ϕ_1 and ϕ_2

$$z_c = r_c\; e^{i\phi} \qquad c_n = \frac{\mu o}{2\pi}\int_{r_1}^{r_2}\int_{\phi 1}^{\phi 2} J\; r^{-n}\; e^{-in\phi}\; rdrd\phi = \frac{\mu o}{2\pi}\int_{r_1}^{r_2}\int_{\phi_1}^{\phi_2} J\; r^{1-n}\; e^{-in\phi}\; drd\phi \qquad (13)$$

If J is constant in the element $r_2 - r_1$, $\phi_2 - \phi_1$

$$c_n = \frac{\mu o}{2\pi} J \left[\frac{r_c^{2-n}}{2-n}\right]_{r_1}^{r_2}\left[e^{-in\phi}\right]_{\phi_1}^{\phi_2}\frac{1}{-in} = i\;\frac{\mu o}{2\pi}\;\frac{J}{(2-n)n}\;(r_2^{2-n} - r_1^{2-n})\;(e^{-in\phi_2} - e^{-in\phi_1}) \qquad (14)$$

For n = 2,

$$c_2 = i\frac{\mu o}{4\pi}\; J\;\ell n\;\frac{r_2}{r_1}\;(e^{-in\phi_2} - e^{-in\phi_1}) \qquad (15)$$

For the case of <u>symmetrical windings with respect to the horizontal median plane</u> (Fig. 10): $\phi_1 = -\phi$ $\qquad \phi_2 = \phi$

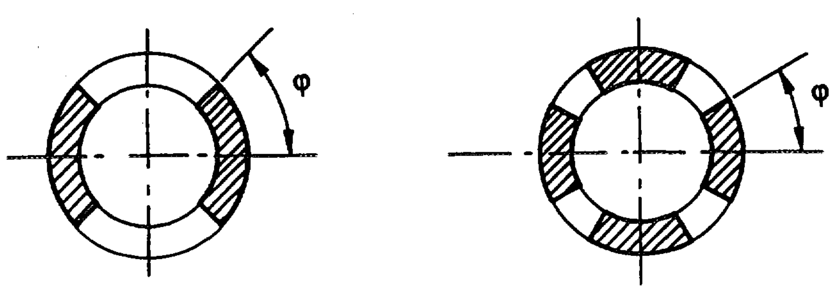


Fig. 10 Symmetries: dipole and quadrupole

$$cn = i \frac{\mu o}{2\pi} \frac{J}{(2-n)n} (r_2^{2-n} - r_1^{2-n}) (e^{-in\phi} - e^{+in\phi})$$

For n ≠ 2,

$$cn = \frac{\mu o}{\pi} \frac{J}{n(2-n)} (r_2^{2-n} - r_1^{2-n}) \sin n \phi \quad (16)$$

$$fn = \frac{\mu o}{2\pi} \frac{1}{r^{2n}} \int_{r_1}^{r_2}\int_{\phi_1}^{\phi_2} J\, r_c^{n+1}\, e^{-in\phi}\, dr d\phi = \frac{\mu o}{\pi} \frac{J}{R^{2n}\, n(n+2)} (r_2^{n+2} - r_1^{n+2}) \sin n \phi \quad (17)$$

For n = 2,

$$c_2 = \frac{\mu o}{2\pi} \int_{r_1}^{r_2} \frac{1}{r}\, dr \int_{-\phi}^{+\phi} e^{-i\,2\phi}\, d\phi = \frac{\mu o J}{2\pi}\, \ell n \frac{r_2}{r_1} \sin 2\phi \quad (18)$$

$$f_2 = \frac{\mu o}{\pi} J \frac{r_2^4 - r_1^4}{4R^4} \sin 2\phi \quad (19)$$

For a dipole, it is convenient to add the sector $\pi + \phi$, $\pi - \phi$.

For n = 1, 3, 5, 7, 9 ...,

$$\boxed{c_{n_d} = \frac{2\mu o}{\pi} \frac{J\,(r_2^{2-n} - r_1^{2-n})}{n(2-n)} \sin n \phi} \quad (20)$$

$$f_{n_d} = \frac{2\mu o}{\pi} \frac{J}{R^{2n}} \frac{(r_2^{n+2} - r_1^{n+2})}{n(n+2)} \sin n \phi \quad (21)$$

For a quadrupole, it is convenient to add the sectors $\pi + \phi$, $\pi - \phi$ and $\frac{\pi}{2} + \phi$, $\frac{\pi}{2} - \phi$ and $\frac{3}{2}\pi + \phi$, $\frac{3}{2}\pi - \phi$:

For n ≠ 2,

$$c_{n_q} = \frac{4\mu o}{\pi} \frac{J\,(r_2^{2-n} - r_1^{2-n})}{n(2-n)} \sin n \phi \quad (22)$$

For n = 2,

$$c_{2_q} = \frac{2\mu o}{\pi} J\, \ell n \frac{r_2}{r_1} \sin 2\phi \quad (23)$$

For n = 2, 6, 10, 14, 18...,

$$f_{n_q} = \frac{4\mu o}{\pi} \frac{J}{R^{2n}} \frac{r_2^{n+2} - r_1^{n+2}}{n(n+2)} \sin n \phi \quad (24)$$

Elimination of multipoles

Unwanted higher-order multipoles can be eliminated by conveniently choosing the angles, the thickness and the current density in the different parts of the windings. For example, for coils composed of m sector blocks with the same inner and outer radii (r_1, r_2) and uniform current density, the condition to be satisfied is

$$\sum_{1}^{m}{}_{K} \; J \sin n \, \phi_k = 0 \tag{25}$$

For a dipole, to eliminate the first higher multipole, the 6-pole, the condition J sin 3 f = 0 has to be satisfied, which leads to f = 60°. Therefore one block of conductors extending from the median plane to the 60° angle is sufficient.

If the coils are composed of two sector blocks of conductors (per quadrant), the first three higher multipoles, 6-pole, 10-pole, 14-pole, can be eliminated.

For a quadrupole, a single block sector winding (per octant) is sufficient to suppress the first higher-order multipole, the 12-pole. For this, the block has to cover the sector from the median plane to the angle $\phi = 30°$. If one wants to suppress the 12-pole, the 20-pole and the 28-pole, a two-block winding is necessary with angles $\phi_1 = 21.59°$, $\phi_2 = 26.075°$, $\phi_3 = 33.635°$. A three-block winding also allows the 36- and the 44-pole to be eliminated: for this the angles are $\phi_1 = 16.657°$, $\phi_2 = 18.548°$, $\phi_3 = 26.564°$, $\phi_4 = 31.682°$, $\phi_5 = 35.915°$.

If the coils are composed of two or more shells, one can adjust the different current densities, angles and subdivision in the conductor blocks in each layer in order to reduce the unwanted multipoles to acceptable values.

Real geometry

In general, the real geometry of the current carrying area does *not* consist of *perfect circular sectors*, because the conductors are rectangular or of a keystone shape and are surrounded and separated by insulation. To treat this, one subdivides each winding block or shell into a number of small circular sectors of uniform current density. The field produced by each element is separately computed and then added up with the field produced by all other elements. This is normally done by means of relatively simple computer programs, such as part of ROXIE [12].

Real iron screen (variable μ)

At this point, one applies one of the existing two-dimensional computer programs, such as MARE/MAGNET, POISSON, etc., in order to optimize yoke thickness, yoke-coil distance, and yoke shape.

In this way the disposition of the conductor around the aperture and in general the magnet cross-section are determined. Figures 3 and 11 show some examples of coil configurations and magnet cross-sections determined by the above method.

3.4 End shaping

In superconducting magnets the coil end is usually the most problematic part in terms of design, winding and mechanical support. The conductors are bent in a saddle fashion around the cylindrical aperture and this is particularly demanding in dipole magnets where the outer turns have to be bent through 180°.

The coil-end shape is optimized, taking into account various requirements:

- ease of fabrication: maximization of the smallest radius of curvature of the conductor in each block;
- mechanical stability;

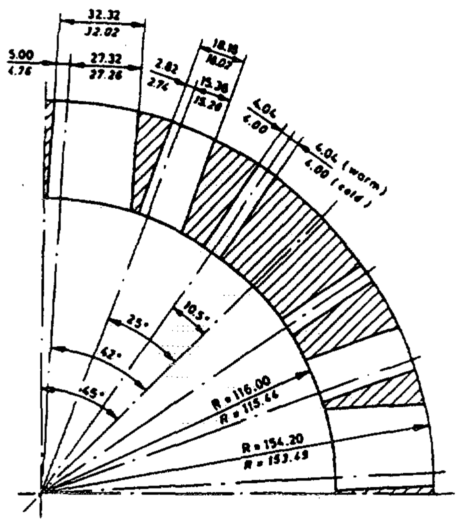


Fig. 11 ISR quadrupoles: cross-section of coil with warm and cold dimensions [13]

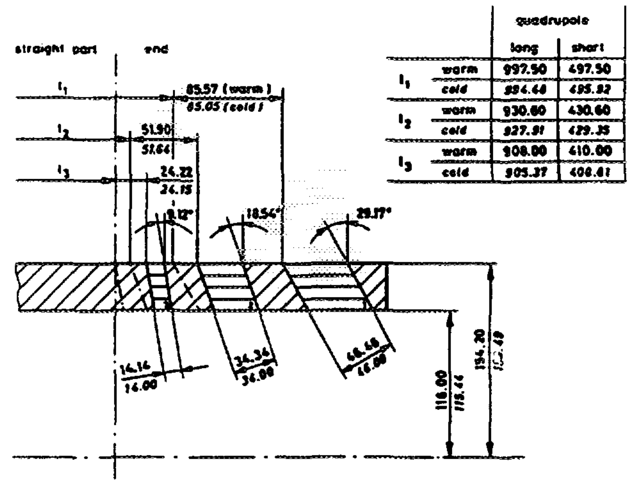


Fig. 12 ISR quadrupoles: coil-end cut along the 45° plane with warm and cold dimensions [13]

- limitation of maximum field in the coils to a value below the max. field in the straight part in order not to limit the magnet performance;
- field quality, i.e. minimum multipole content integrated over the coil end;
- minimization of the end length to limit the loss of magnetic length.

In general, at the ends the coil is spread out in conductor blocks separated by spacers whose number, shape and dimensions are determined by the above requirements. Concerning the shape of the conductor bends, two types are mainly used: conductors upright with their wide face perpendicular to the aperture cylindrical surface, e.g. the HERA dipole coils, and the quasi-isoperimetric form approximating the shape that a thin tape would naturally take (e.g. ISR quadrupoles, LHC magnets). The first configuration entails considerable deformation of the cable, but has the advantage of "filling" the same cylindrical space as in the straight part of the magnet. The other configuration is more adapted to large conductors as the resulting deformation is smaller, but the conductors take an inclination with respect to the magnet axis. Consequently some filler material is needed to restore the cylindrical surfaces through which the electro-magnetic forces are transferred to the supporting structure. Several variants of these configurations have been used, examples are shown in Figs. 12–15.

a) spaces to be filled to restore the outer cylinder of each coil layer

b) spaces to be filled to restore the inner cylinder of each coil layer

Fig. 13 Cut through two quasi-isoperimetric ends (LHC dipole models)

The calculations are carried out with the aid of computer programs. These may be analytical and based on the formulae outlined in section 3.3, taking advantage of the fact that the field integrated from $-\infty$ to $+\infty$ or from the magnet centre to ∞ satisfies Laplace's equation in two dimensions, such as the programs ENDEF [14] or ROXIE. Other programs integrate the Biot-Savart law point by point [15]. With these programs however, the iron is considered as an equipotential boundary ($\mu = \infty$). To take into account the effects of iron saturation more complex finite-element programs are used (e.g. TOSCA) [16].

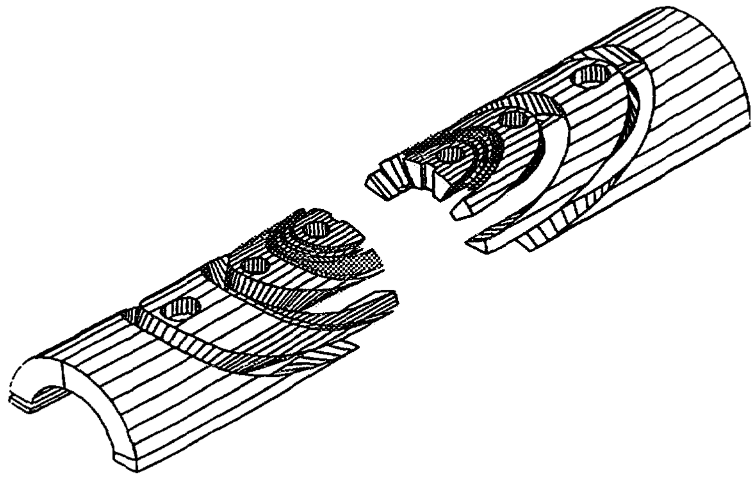


Fig. 14 End spacers for the LHC dipole coil inner shell

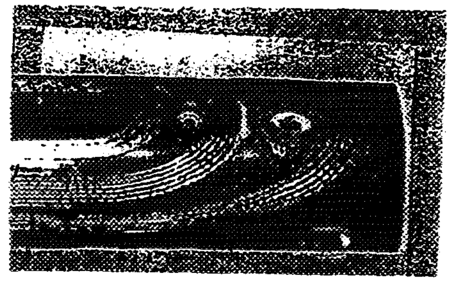


Fig. 15 Coil end (LHC dipole model, inner shell)

3.5 Other points studied in the magnetic design

Effects of remanent field { from the conductor (persistent currents). / from the yoke (or screen).

Stored energy, stray field.

Tolerances on geometry.

Electro-magnetic forces, etc.

4. FIELD QUALITY

Field quality is, of course of great importance to the accelerator designer, as it directly affects beam optics and stability of the beam(s).

4.1 Definition of multipole field components

The field in accelerator magnets can usually be treated as two-dimensional since the beam integrates through the field. It can then be expressed as a power series:

$$B_y + iB_x = B_1 \sum_{1}^{\infty}{}_n (b_n + ia_n) (z/R_r)^{n-1}$$

where B_1 = magnitude of dipole field in the y (vertical) direction;
b_n = normal multipole coefficient;
a_n = skew multipole coefficient
z = $x + iy$;
R_r = reference radius.

n = 1 dipole
n = 2 quadrupole
n = 3 sextupole, etc.

One has to be careful when comparing figures, because the coefficients a_n and b_n are dimensionally dependent. So they must be defined at a given reference radius.

Caution

When consulting literature, attention has to be paid to the fact that in the USA a slightly different nomenclature is used:

$$B_y + iB_x = B_o \sum_0^{\infty} n\,(b_n + ia_n)\,(z/R_r)^n$$

and then $n = 0$ dipole,
$n = 1$ quadrupole,
$n = 2$ sextupole, etc.

4.2 Sources of field errors

Field errors have different origins: conductor placement errors, iron saturation, coil deformations under e.m. forces, persistent and eddy currents.

a) Conductor placement errors

The field errors originating from misalignment of conductors or complete coils can be computed analytically at the design stage from the known possible or imposed manufacturing tolerances. Sensitivity matrices can be computed for displacement of single conductors or conductor blocks, as for example the matrix of Fig. [16] for the LHC dipole, which has a coil aperture diameter of 56 mm [17]. Some examples are given for this magnet in terms of the most sensitive multipole components a_n, b_n defined at the reference radius $R_r = 10$ mm:

- a 0.1-mm increase of the coil azimuthal size (which incidentally would correspond to a compression of about 30 MPa) while current and coil thickness remain unchanged, produces a sextupole and a decapole

$$b_3 = 1.2 \times 10^{-4}, \qquad b_5 = 0.03 \times 10^{-4}$$

- a 0.01-mm inner radius difference between lower and upper poles of the same aperture produces a skew quadrupole

$$a_2 = 0.44 \times 10^{-4}$$

- a 0.1-mm vertical off-centering of the coils with respect to the yoke produces a skew quadrupole

$$a_2 = 0.15 \times 10^{-4} \text{ at } 0.58\text{ T (injec. field), and } 0.09 \times 10^{-4} \text{ at } 8.4\text{ T (nominal field)}$$

- a 1-mm difference in length between lower and upper coil of the same aperture in a 14 m long LHC dipole ($\Delta L/L = 7 \times 10^{-5}$) is equivalent to an "integrated two-dimensional" skew quadrupole

$$a_2 = 0.06 \times 10^{-4}$$

Fluctuations of bending power from magnet to magnet may be also originated:

- a 0.1-mm azimuthal elongation of the coils, with unchanged current and coil thickness, produces a dipole field variation

$$\frac{\Delta B}{B} = -\,1.4 \times 10^{-3}$$

- a 0.1-mm increase of coil inner radius, with unchanged current and coil thickness, produces

$$\frac{\Delta B}{B} = -\ 1.5 \times 10^{-3}$$

- a 0.1-mm increase of yoke aperture (coil-yoke distance) produces

$$\frac{\Delta B}{B} = -\ 1.5 \times 10^{-3}$$

Error matrix for radial displacements of 1 mm of the winding blocks in 1E-4

		Quadrant 1					Quadrant 2					Quadrant 3					Quadrant 4				
Error type	Multipole order	Block 1.1	1.2	1.3	1.4	1.5	Block 2.1	2.2	2.3	2.4	2.5	Block 3.1	3.2	3.3	3.4	3.5	Block 4.1	4.2	4.3	4.4	4.5
b_n	1	-7.89	-6.45	-9.18	-9.20	-3.36	-7.89	-6.45	-9.18	-9.20	-3.36	-7.89	-6.45	-9.18	-9.20	-3.36	-7.89	-6.45	-9.18	-9.20	-3.36
a_n	1	1.48	6.87	1.42	6.28	6.64	-1.48	-6.87	-1.42	-6.28	-6.64	1.48	6.87	1.42	6.28	6.64	-1.48	-6.87	-1.42	-6.28	-6.64
b_n	2	3.68	1.08	5.70	2.60	-2.88	-3.68	-1.08	-5.70	-2.60	2.88	-3.68	-1.08	-5.70	-2.60	2.88	3.68	1.08	5.70	2.60	-2.88
a_n	2	-1.44	-5.18	-1.85	-6.67	-3.88	-1.44	-5.18	-1.85	-6.67	-3.88	1.44	5.18	1.85	6.67	3.88	1.44	5.18	1.85	6.67	3.88
b_n	3	-1.02	0.73	-2.33	0.71	2.09	-1.02	0.73	-2.33	0.71	2.09	-1.02	0.73	-2.33	0.71	2.09	-1.02	0.73	-2.33	0.71	2.09
a_n	3	0.64	1.40	1.20	3.02	-0.37	-0.64	-1.40	-1.20	-3.02	0.37	0.64	1.40	1.20	3.02	-0.37	-0.64	-1.40	-1.20	-3.02	0.37
b_n	4	0.23	-0.36	0.81	-0.87	-0.23	-0.23	0.36	-0.81	0.87	0.23	-0.23	0.36	-0.81	0.87	0.23	0.23	-0.36	0.81	-0.87	-0.23
a_n	4	-0.21	-0.16	-0.60	-0.80	0.79	-0.21	-0.16	-0.60	-0.80	0.79	0.21	0.16	0.60	0.80	-0.79	0.21	0.16	0.60	0.80	-0.79
b_n	5	-0.04	0.09	-0.25	0.42	-0.22	-0.04	0.09	-0.25	0.42	-0.22	-0.04	0.09	-0.25	0.42	-0.22	-0.04	0.09	-0.25	0.42	-0.22
a_n	5	0.06	-0.02	0.26	0.06	-0.20	-0.06	0.02	-0.26	-0.06	0.20	0.06	-0.02	0.26	0.06	-0.20	-0.06	0.02	-0.26	-0.06	0.20
b_n	6	0.01	-0.01	0.07	-0.13	0.10	-0.01	0.01	-0.07	0.13	-0.10	-0.01	0.01	-0.07	0.13	-0.10	0.01	-0.01	0.07	-0.13	0.10
a_n	6	-0.02	0.02	-0.10	0.06	-0.04	-0.02	0.02	-0.10	0.06	-0.04	0.02	-0.02	0.10	-0.06	0.04	0.02	-0.02	0.10	-0.06	0.04
b_n	7	0.00	0.00	-0.02	0.02	0.00	0.00	0.00	-0.02	0.02	0.00	0.00	0.00	-0.02	0.02	0.00	0.00	0.00	-0.02	0.02	0.00
a_n	7	0.00	0.00	0.04	-0.04	0.04	0.00	0.00	-0.04	0.04	-0.04	0.00	0.00	0.04	-0.04	0.04	0.00	0.00	-0.04	0.04	-0.04
b_n	8	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	-0.01
a_n	8	0.00	0.00	-0.01	0.01	-0.01	0.00	0.00	-0.01	0.01	-0.01	0.00	0.00	0.01	-0.01	0.01	0.00	0.00	0.01	-0.01	0.01
b_n	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
a_n	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Error matrix for tangential displacements of 1mm of the winding blocks in 1E-4

		Quadrant 1					Quadrant 2					Quadrant 3					Quadrant 4				
Error type	Multipole order	Block 1.1	1.2	1.3	1.4	1.5	Block 2.1	2.2	2.3	2.4	2.5	Block 3.1	3.2	3.3	3.4	3.5	Block 4.1	4.2	4.3	4.4	4.5
b_n	1	-2.76	-12.30	-2.03	-8.30	-8.66	2.49	12.01	1.70	7.97	8.54	-2.76	-12.30	-2.03	-8.30	-8.66	2.49	12.01	1.70	7.97	8.54
a_n	1	-13.93	-14.84	-11.86	-11.79	-4.22	-13.98	-15.07	-11.91	-12.02	-4.46	-13.93	-14.84	-11.86	-11.79	-4.22	-13.98	-15.07	-11.91	-12.02	-4.46
b_n	2	1.78	6.10	2.07	6.94	3.91	1.61	6.05	1.74	6.79	4.07	-1.78	-6.10	-2.07	-6.94	-3.91	-1.61	-6.05	-1.74	-6.79	-4.07
a_n	2	4.29	1.15	5.82	2.48	-3.07	-4.35	-1.38	-5.93	-2.86	2.85	-4.29	-1.15	-5.82	-2.48	3.07	4.35	1.38	5.93	2.86	-2.85
b_n	3	-0.71	-1.45	-1.30	-2.99	0.46	0.65	1.49	1.10	3.05	-0.28	-0.71	-1.45	-1.30	-2.99	0.46	0.65	1.49	1.10	3.05	-0.28
a_n	3	-1.06	0.81	-2.28	0.83	2.08	-1.10	0.72	-2.38	0.58	2.11	-1.06	0.81	-2.28	0.83	2.08	-1.10	0.72	-2.38	0.58	2.11
b_n	4	0.23	0.15	0.64	0.74	-0.80	0.21	0.17	0.55	0.84	-0.78	-0.23	-0.15	-0.64	-0.74	0.80	-0.21	-0.17	-0.55	-0.84	0.78
a_n	4	0.23	-0.37	0.77	-0.81	-0.19	-0.24	0.38	-0.84	0.82	0.28	-0.23	0.37	-0.77	0.81	0.19	0.24	-0.38	0.84	-0.82	-0.28
b_n	5	-0.06	0.03	-0.28	-0.03	0.19	0.06	-0.02	0.24	0.09	-0.22	-0.06	0.03	-0.28	-0.03	0.19	0.06	-0.02	0.24	0.09	-0.22
a_n	5	-0.04	0.09	-0.23	0.42	-0.24	-0.05	0.09	-0.27	0.41	-0.21	-0.04	0.09	-0.23	0.42	-0.24	-0.05	0.09	-0.27	0.41	-0.21
b_n	6	0.02	-0.02	0.11	-0.07	0.05	0.02	-0.01	0.10	-0.05	0.03	-0.02	0.02	-0.11	0.07	-0.05	-0.02	0.01	-0.10	0.05	-0.03
a_n	6	0.01	-0.01	0.08	-0.12	0.10	-0.01	0.01	-0.08	0.13	-0.10	-0.01	0.01	-0.08	0.12	-0.10	0.01	-0.01	0.08	-0.13	0.10
b_n	7	0.00	0.00	-0.04	0.04	-0.04	0.00	0.00	0.04	-0.04	0.04	0.00	0.00	-0.04	0.04	-0.04	0.00	0.00	0.04	-0.04	0.04
a_n	7	0.00	0.00	-0.01	0.02	0.00	0.00	0.00	-0.02	0.03	-0.01	0.00	0.00	-0.01	0.02	0.00	0.00	0.00	-0.02	0.03	-0.01
b_n	8	0.00	0.00	0.01	-0.01	0.01	0.00	0.00	0.01	-0.02	0.01	0.00	0.00	-0.01	0.01	-0.01	0.00	0.00	-0.01	0.02	-0.01
a_n	8	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	-0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	-0.01
b_n	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
a_n	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

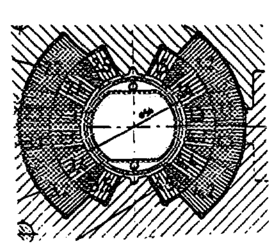


Fig. 16 Sensitivity matrix for radial and azimuthal displacements of conductor blocks in LHC dipole

From these examples one may understand how difficult it is to obtain in superconducting magnets a field quality comparable to that of classical lower field magnets, where the field distribution is determined by the iron-pole profiles which can easily be produced with a hundredth-millimetre precision. In superconducting magnets the coil geometry is the result of assembling stacks of conductors, typically 15 to 20, which are produced with stringent, but not infinitely small, tolerances (in cables the best that can be achieved nowadays is of the order of $\pm$ 0.0025 mm on the thickness) and insulated by wrapping them with tapes which can be industrially produced with a few micrometers tolerance on their thickness.

b) Iron saturation

The field errors originating from the remanence and the variable permeability vary with excitation and depend strongly on the coil-yoke distance. For warm iron magnets (e.g. Tevatron dipoles) these errors can be neglected in practice. For cold iron magnets they have to be carefully evaluated, are in general systematic and affect mainly the first higher multipoles (6-pole and 10-pole in dipoles) and can be compensated either by correction windings in the aperture (e.g. HERA) or by small corrector magnets placed at the magnet ends (e.g. LHC).

c) Coil deformation under the e.m. forces

These errors vary with excitation and can be easily computed after the mechanical analysis of the structure and the determination of the deformed coil configurations, using analytical or other computer programs.

d) Persistent currents in the superconductors

The persistent current errors are a particularity of superconducting magnets. They are due to currents induced in the superconducting filaments by field variations and, contrary to normal conductors where resistance rapidly reduces and after a while eliminates eddy currents, circulate indefinitely as long as the superconductor is kept below its critical temperature. Persistent current errors affect all field multipole components allowed by the symmetry configuration of the magnet, including the fundamental one. Their importance decreases with excitation, but they are particularly disturbing at low field level and especially at injection. They depend on the previous powering of the magnet and vary with time and, therefore, require a careful study of the magnet excitation cycle. Persistent currents are proportional to the effective diameter of the superconducting filaments, so accelerator magnets use filaments as thin as possible, compatibly with economy and quality of production.

This subject is extensively treated in the chapter by A. Devred and in Refs. [3, 18, 19].

e) Eddy currents

Eddy currents occur during field sweep in multistrand conductors both inside the strands, mainly due to coupling between filaments, and between the strands. They distort the magnetic field, and their effects depend on the geometrical and electrical characteristics of strands and cables (matrix and inter-strand resistance, cable aspect ratio, distribution of superconducting filaments, etc.) and, of course, on the field ramp rate.

This subject is also treated in a lecture by A. Devred and in Refs. [20–22].

f) Torsion

In long and slim objects such as superconducting magnets, twist around the longitudinal axis is difficult to prevent and can produce non-negligible field orientation errors. In general a tolerance of about 1 mrad is required in dipole magnets, and, although the beam integrates along its path, local twists should not exceed this value by more than a factor two or three. Excessive twist can be prevented by locking the magnet active part to a structural component of high torsional rigidity, generally a closed cylinder of large diameter, such as the helium vessel, which in many cases is also part of the force containment structure.

4.3 Examples

At the design stage of accelerators, forecasts of field quality can be obtained by applying the considerations of section 4.2. These can be corroborated by existing statistical results, at present from the Tevatron (744 dipoles), from HERA (~ 500 dipoles) and recently from RHIC (~ 100 dipoles measured), and from LHC models and prototypes.

As an example, Table 2 presents the expected error field components in the LHC dipole at injection and nominal operation field. Table 3 shows an estimate of the eddy current errors for given interstrand contact resistance and field ramp rate: note that, *at a given ramp rate, eddy currents produce the same errors at any field level*, so that the relative field errors are highest at injection.

An example of twist in 10-m long magnets is shown in Fig. 17, plotting the field orientation in the two apertures of three LHC dipole prototypes of the first generation. Noticeable is the remarkably good parallelism of the field in the two apertures at all longitudinal positions.

5. FORCES AND MECHANICS

5.1 Some facts

– Electromagnetic forces (e.m.) are in first approximation proportional to B^2

$$\vec{F} = \vec{I} \times \vec{B} = \sim B^2$$

and are extremely high in high-field magnets: e.g. see in Fig. 18 the transverse forces in the LHC dipole.

Table 2
Expected multipole performance, at injection and at 8.40 T
(In units of 10^{-4} relative field error at 10 mm)

	At injection, 0.58 T				At nominal operation, 8.40 T			
	Mean		Random		Mean		Random	
n	norm. b_n	skew a_n	norm. $\sigma[b_n]$	skew $\sigma[a_n]$	norm. b_n	skew a_n	norm. $\sigma[b_n]$	skew $\sigma[a_n]$
2	±0.5	±0.3	0.4	1.0	1.0±0.5	±0.3	0.4	1.0
3	-3.6±0.3*	±0.3	0.5	0.15	0.5±0.3*	±0.3	0.6	0.15
4	±0.2	±0.2	0.1	0.1	±0.2	±o.2	0.1	0.1
5	0.18±0.05*	±0.05	0.08	0.04	0.06±0.05*	±0.05	0.05	0.04
6	-0.004	0.0	0.02	0.01	-0.005	0.0	0.006	0.01
7	-0.026	0.0	0.01	0.01	0.006	0.0	0.009	0.003
8	0.0	0.0	0.005	0.005	0.0	0.0	0.001	0.002
9	0.006	0.0	0.003	0.004	-0.001	0.0	0.001	0.001
10	0.0	0.0	0.002	0.002	0.0	0.0	0.0	0.0
11	0.008	0.0	0.001	0.001	0.008	0.0	0.0	0.0

*Systematic b_3 and b_5 are compensated by correctors at each dipole end. The b_3 values in this table indicate the magnitude of the persistent current effect at injection and of the yoke saturation at 8.4 T, the coil geometry being designed for $b_3 = 0$.

Table 3
Estimation of the additional magnetic errors during the ramp of the field for an interstrand contact resistance of 10 μΩ

	Field at x = 1 cm [mT]		Relative field error at injection	
n	normal	skew	b_n	a_n
1	4.8		9.0	0
2	0.03	0.54	0.056	1.0
3	0.24	0.060	0.45	0.11
4	0.018	0.060	0.034	0.11
5	0.091	0.018	0.17	0.034

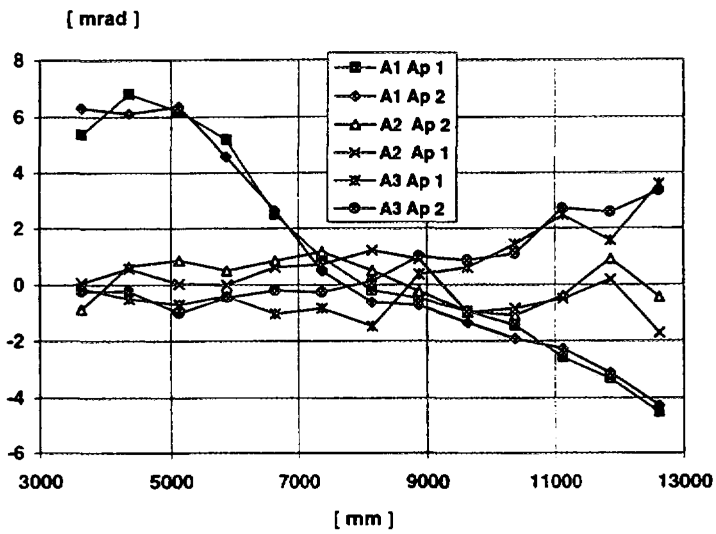


Fig. 17 Field orientation measured in the two apertures of three, 10-m long LHC dipole models (A1, A2, A3)

- Specific heat of materials (NiTi, Cu, etc.) are very low** at 4.2 K (for NbTi, C = ~ 6.5 x 10^{-4} J/g K, for Cu, C = ~ 1 x 10^{-4} J/g K) and is about a factor 10 lower at 2 K.

- A superconductor stays in the s.c. state when temperature, magnetic field and current density are below their critical values.

- The temperature margin, the difference between the conductor working temperature and the critical temperature at the working field and current, is usually very small (1 – 2 K). (E.g. in LHC dipoles ~ 1.2 – 1.4 K.)

- Even extremely small sudden movements of the cable (or even of a strand) or cracking of the insulation generate enough heat to raise local parts of the superconductor above the critical temperature, provoking premature quenching (training).

- It is therefore necessary to limit those sudden movements by a careful design and construction of the force containment structure. Elastic gradual deformations are permitted, to a certain extent.

** ~10^{-4} of the room temperature spec. heat

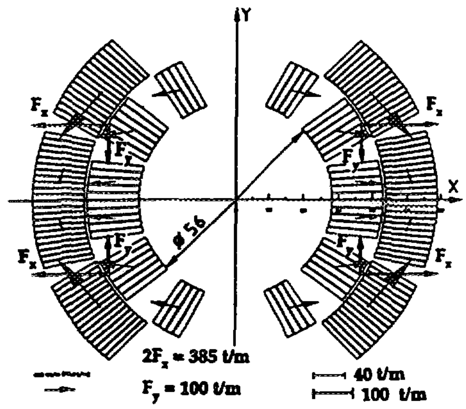


Fig. 18 Transverse electro-magnetic forces in the LHC dipole coils at 9-T field

5.2 "Roman Arch" concept for mechanical stability

Coils are built from different materials: conductors, metallic or insulating spacers, insulations, and glue (epoxy or polyimide resin) to keep everything together, at least during fabrication.

Sometimes the construction is even more loose: one wants to avoid sticking of the insulation to the conductor in order to:

- Let the conductor slide with little friction inside the insulation, to prevent sudden release of heat in case of small movements.

- Let the liquid helium enter into direct contact with the conductor to carry away heat. This is beneficial to stability and essential in case of pulsed magnets or if the magnets have to be operated in helium II (≈2 K), and to remove heat deposited by beam losses (e.g. LHC).

Insulation and glues cannot be relied upon to withstand the high tensile stresses that result from the electromagnetic forces. If nothing is done cracks and micro fissures will develop and the magnet will repetitively quench (training). The solution lies in the application of an adequate pre-compression which will prevent any tensile stress in the coil when e.m. forces are applied. A successful idea was to make the coil behave as a Roman Arch. For cos $n\vartheta$ winding configurations the external structure applies a radial inward compression which, as in Roman arches, is transformed into azimuthal compression inside the coil which counteracts the formation of tensile stresses that would otherwise appear under the action of the electro-magnetic forces (Fig. 19).

A roman arch can be built from materials of poor or even zero tensile properties (e.g. from dry stones), provided they have a good resistance to compression. For the concept to work the coils should not be supported inside.

This concept was first thought of and applied to the design and construction of the ISR low-β quadrupoles (1974) [23], then adopted in practically all s.c. main magnets for accelerators.

In the example of the ISR quadrupoles (Fig. 20), the pre-compression was applied by means of aluminium-alloy shrinking rings through the yoke quadrants and a set of stainless steel spacers. The main advantages of using an aluminium alloy instead of other materials (e.g. stainless steel) for the shrinking rings were the same as those mentioned in section 5.3 for the collars.

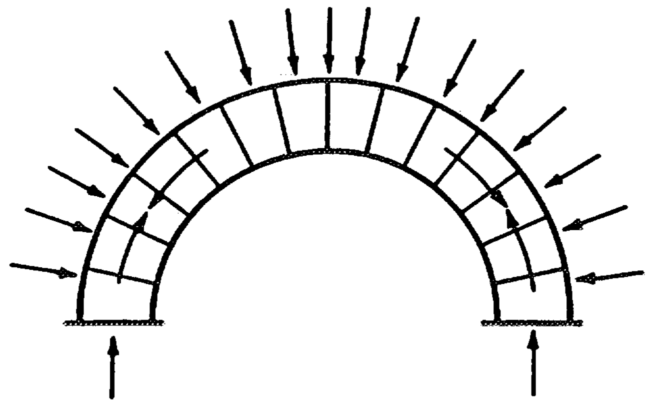


Fig. 19 Roman Arch analogy

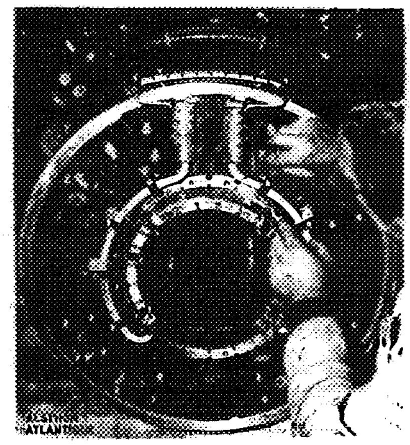


Fig. 20 ISR superconducting quadrupole

5.3 Force-containment structures

The purposes of the force-containment structures are:

- To prevent any sudden movement of the conductors leading to premature quenching and training. For this, the structure must not only be sufficiently strong to hold the e.m. forces without damage, but must also be able to pre-stress the coil in such a way that any tensile stress is prevented at all excitation conditions.

- To limit the elastic deformations of the coils, so that the field quality is maintained at all field levels.

- To guarantee the required structural and dimensional integrity of the magnet.

To achieve this, considerable pre-loads are necessary at operation temperature: e.g. for dipoles operating at 10 T azimuthal average compressive pre-stresses of the order of 80–100 N/mm^2 in the coils are usual.

For magnets working at moderate fields, up to about 5.5 T, coil clamping is in general achieved simply by collars which are adequately compressed when mounted around the coils and locked in position by means of dowels or keys. Different solutions are however adopted concerning the collar material (aluminium alloys, non-magnetic steels, steel-yoke laminations) and their locking system (dowel rods through quasi-circular holes, keys inside grooves).

The main advantage of aluminium alloy collars is that they produce an increase (or at least not a decrease) of the prestress on the coil at cool-down, thanks to their high thermal contraction. In this way the high prestress is applied only when needed and not at room temperature, where creep of insulation or even of copper may occur in the long run. Another advantage is lower cost of material and fabrication. A drawback when compared to stainless or other non-magnetic steels is that more space (wider collars) is required because of lower elastic modulus and tensile/compressive strength. In some cases, however, the coil-to-yoke distance is imposed by the required field quality. In addition, aluminium alloys are more sensitive to stress concentration under fatigue conditions, so that it is not advisable to use grooves and keys to lock them. A problem with present collaring systems is that the coils have to be compressed more than strictly necessary in order to produce the clearances (typically 0.1 mm) which permit the insertion of rods or keys and to compensate for the elongation of the collar "legs" under the reaction of the coils when releasing the press. Concerning the latter, austenitic steels behave better than aluminium alloys owing to their larger elastic modulus (10–20% higher prestress remaining after releasing the press). On the other hand with austenitic-steel collars one is bound to loose an important fraction (~ 30%) of the prestress during cool-down. This can be accepted in magnets for moderate field levels, but is severely penalizing for high-field magnets in which the prestress that must then be applied at room temperature may exceed the acceptable creep limits (magnets should be able to survive long storage time at room temperature).

It is recognized that to achieve efficient designs for fields above 6 T, it is necessary that the rest of the structure (yoke + other components) contributes to the containment of the forces. So, in the HERA dipoles, which were designed for a lower operation field, above 6 T the collars come in contact in the horizontal plane with the yoke, which acts as a stopper to limit the deformation of the coil/collar assembly [24]. This, however, happens after a non-negligible radial outward elastic expansion (~ 0.1 mm) of the assembly at the median plane.

For the LHC 8 to 10 T dipoles (Fig. 3) a new type of mechanical structure was designed [25] in which aluminium-alloy collars are surrounded by a vertically split yoke (with open gap at room temperature) clamped by a stainless steel shrinking cylinder. The collars are clamped around two coils at room temperature with a moderate pressure to avoid risk of room temperature creep. During the cooling process the shrinkage of the collars increases the pre-compression in the coils, while the yoke halves, which are actuated by the outer shrinking cylinder, move horizontally inwards applying additional compressive forces to the collar-coil assembly in the direction just opposite to the main action of the e.m. forces. The gap between the two parts of the yoke closes at a predetermined lower temperature and when cool-down is completed a compressive force is produced at the mating face between the yoke halves. If this force is equal or larger than the horizontal resultant of the e.m. forces, the gap remains closed in all operating conditions and the split iron behaves as a single stiff solid body. Coil displacement and deformation under the electromagnetic forces are, therefore, greatly reduced, as compared to a simple collar clamping system, with beneficial effects on magnet stability and also on field quality.

In the example of the LHC dipoles at 9 T field the horizontal radial elastic expansion of the coil at the median plane is only about 0.05 mm. In the same conditions the maximum compressive stress in the coil inner layer is 81 N/mm^2, to be compared to 126 N/mm^2 computed for a structure of the same dimensions, but without the compressive prestress on the split-yoke mating faces. The price to be paid for this more efficient, but also more complex structure is that its components have to be produced with tighter dimensional accuracy to ensure that they perfectly fit together.

Axial electro-magnetic forces in high-field magnets produce large stresses and strains in the coil. A quick way to determine globally the axial force is to simply take the derivative of the magnetic stored energy with respect to the axial direction dW/dz. In the LHC 15-m long dipole, at 9.6 T, close to the conductor short sample limit, the stored energy per metre length is 670 kJ, the axial force is then 0.67 MN and the axial tensile stress in the coils would be about 100 MPa and the elastic elongation 12 mm. Magnet ends have, therefore, to be adequately supported by the external structure and a detailed stress analysis has to be done for any end design. Part of the axial force is transferred by friction from the coils to the outer structure, in general the shrinking cylinder, via collars and yoke. The rest is taken by end plates and from them transferred to the strongest longitudinal elements, normally the shrinking cylinder. Sharing between the two parts depends on the construction. Measurements have shown that with the LHC dipole construction type only a fraction of 15–20% of the axial electro-magnetic force is taken by the end plates. Figure 21 is a longitudinal section of the magnet showing the special bolts which carry the force from the coil end pieces to the end plates. These bolts are also used to pre-compress slightly the coil axially at assembly.

Design and stress analysis of the force supporting structures are normally first carried out using the classical mechanical engineering methods and then refined with the aid of modern computer packages which can calculate the e.m. force distribution, the behaviour (stress and strains) of the whole magnet and of each of its components under all operational conditions, and during assembly. Many examples of this procedure can be found in the literature [26, 27], a recent treatment for the LHC dipole is reported in Ref. [28].

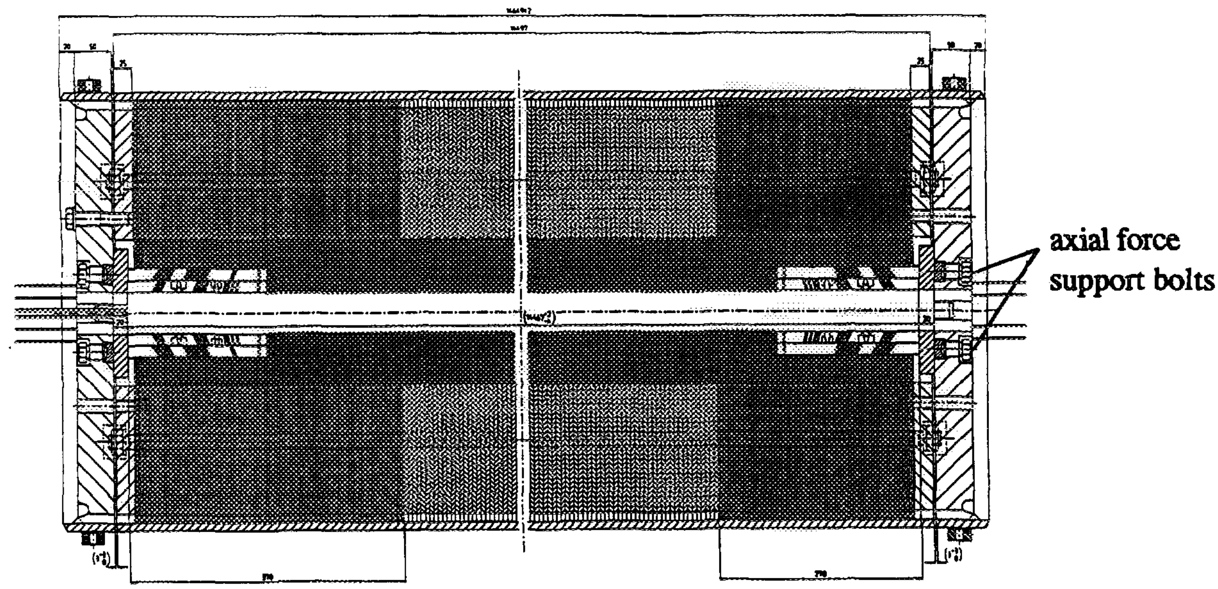


Fig. 21 Longitudinal section of LHC dipole

* * *

REFERENCES

[1] The LHC Study Group, LHC, The large hadron collider accelerator project, CERN/AC/93-03(LHC), 8 Nov. 1993.

[2] M.N. Wilson, Superconducting magnets, Clarendon Press, Oxford 1983.

[3] K.H. Mess, P. Schmüser, Superconducting accelerator magnets, CAS School on Superconductivity in Particle Accelerators, CERN 89-04.

[4] R. Perin, High-field superconducting magnets for particle accelerators, New Techniques for Future Accelerators III, pp. 87-106, Plenum Press, New York, 1990.

[5] M. Linder et al., Construction details and test results from RHIC sextupoles, IEEE Trans. on Magnetics, Vol. 30, No. 4, pp. 1730-1733, July 1994.

[6] P.M. McIntyre, R.M. Scanlan, W. Shen, Ultra-high-field magnets for future hadron collider, Proc. ASC-94, Boston, Oct. 1994.

[7] R.F. Holsinger, Ch. Iselin, CERN POISSON package, CERN computer Library.

[8] R. Perin, S. van der Meer, The program MARE for the computation of two-dimensional static magnetic field, CERN 67-7, March 1967

[9] TOSCA, Vector Fields Ltd, Oxford, U.K.

[10] R.A. Beth, Proc. of 1968 Summer Study on Superconducting Devices, Report BNL 50155, Vol. 3, p. 843, 1968.

[11] R. Perin, Calculation of magnetic field in a cylindrical geometry produced by sector or layer windings, CERN/ISR-MA Internal Note, Nov. 1973.

[12] S. Russenschuck, A computer program for the design of superconducting accelerator magnets, CERN AT/95-39 (MA), LHC Note 354, September 1995.

[13] R. Perin, T. Tortschanoff, R. Wolf, Magnetic design of the superconducting quadrupole magnets for the ISR high-luminosity insertion, CERN/ISR-BOM 179-2, 1979.

[14] T. Tortschanoff, Computation of the integrated field in a s.c. quadrupole magnet with constant perimeter end, Proc. Compumag, 1976.

[15] C. Daum, D. Ter Avest, Three-dimensional computation of magnetic fields and Lorentz forces of an LHC magnet using the method of image currents, Proc. MT-11 Conference, Tsukuba, Japan, 1989.

[16] D. Ter Avest, C. Daum, H.H.J. ten Kate, Magnetic fields and Lorentz forces in an LHC dipole magnet; 3-D analysis using the FEM program TOSCA.

[17] S. Russenschuck and R. Wolf, private communication.

[18] M. A. Green, Residual fields in superconducting magnets, Proc. of Int. Conf. on Magnet Technology, p. 339, 1972.

[19] R. Wolf, Remanent fields in multipoles due to superconductor magnetization, CERN Internal Note ISR-MA, April 1974.

[20] A.P. Verweij, R. Wolf, Field errors due to inter-strand coupling currents in the LHC dipole and quadrupole, CERN Internal Note AT-MA/94-97, March 1994.

[21] T. Ogitsu, Influence of cable eddy currents on the magnetic field of superconducting particle accelerator magnets, SSCL-N-848, January 1994.

[22] V.T. Kovatchev, M.J. Neal, D.W. Capone, W.J. Curr, C. Svenson, Interstrand resistance of SSC magnets, Cryogenics 1994, Vol. 34, No. 10.

[23] R. Perin, Mechanical stability of superconducting quadrupole coils, Proc. 5th Int. Conf. on Magnet Technology, pp. 551-556, Roma, 1975.

[24] H. Kaiser, Design of superconducting dipole for HERA, 13th Int. Conf. on High Energy Accelerators, Novosibirsk, 1986, and DESY-HERA 1986-14.

[25] R. Perin, Preliminary views on 8 tesla dipole magnet for the Large Hadron Collider, CERN LHC Note 32, and CERN-TIS-MC/85-2.

[26] D. Leroy, R. Perin, G. de Rijk, W. Thomi, Design of a high-field twin-aperture superconducting dipole model, IEEE Trans. on Magnetics, Vol. 24, pp. 1373-1376, March 1988.

[27] D. Leroy, R. Perin, D. Perini, A. Yamamoto, Structural analysis of the LHC 10 T twin-aperture dipole, Proc. XI Int. Conf. on Magnet Technology, pp. 159-164, Tsukuba, Japan, 1989.

[28] R. Perin, D. Perini, J. Salminen, J. Soini, Finite element structural analysis of LHC bending magnet, Proc. MT14, Tampere, 12-15 June 1995.

INFLUENCE OF EDDY CURRENTS IN SUPERCONDUCTING PARTICLE ACCELERATOR MAGNETS USING RUTHERFORD-TYPE CABLES

A. Devred[1] *and T. Ogitsu*[2]
Superconducting Super Collider Laboratory*, Dallas, USA

Abstract
A model describing the influence of eddy currents in superconducting particle accelerator magnets using flat, two-layer, Rutherford-type cables is developed. The model relies on a two-dimensional field calculation, and considers two types of eddy currents: 1) the intra-strand eddy currents flowing within the cable strands and coupling the superconducting filaments, and 2) the inter-strand eddy currents flowing from strand to strand through the contact resistances at the crossovers between the two cable layers. The model can be used to compute the power and the field distortions produced by the eddy currents and is applicable to most of the existing designs of superconducting particle accelerator magnets.

1. INTRODUCTION

Most of the large hadron synchrotrons use superconducting particle accelerator magnets. The operation of a synchrotron generally includes an acceleration phase during which the magnets are ramped from a low field (injection mode) to a high field (extraction or collider mode). While ramping, the varying field induces eddy currents in various components of the magnets, which result in heat dissipation and field distortions. The heat dissipation raises the magnet temperature and can lead to quenching. The field distortions perturb the beam circulation and can result in beam losses. Of course, the faster the acceleration, the more serious the potential for ramp-related problems.

A recent example of a fast-cycling machine for which superconducting magnets were considered is the High Energy Booster (HEB) of the Superconducting Super Collider (SSC) [1]. In this accelerator complex, which was to be built near the picturesque town of Waxahatchie, 35 miles south of Dallas, Texas, USA, the HEB would have accelerated protons from 0.2 TeV to 2.0 TeV and would have served as an injector to the final stage, called the Collider. In the Collider, also relying on superconducting magnet technology, protons in two counter-rotating rings would have been accelerated to 20 TeV and stored for up to 24 hours while collision occurred at several interaction points. The injection time from the HEB to the Collider was fixed at 1 hour, during which the HEB was to be cycled eight times between -6650 A and +6650 A, at a ramp rate of 62 A/s. At the end of injection, the Collider was to be ramped from the injection current (of the order of 650 A) to 6700 A at a ramp rate of 4 A/s. Hence, the HEB magnets were supposed to be operated in a bipolar mode and at a relatively large ramp rate.

The SSC project was cancelled on October 20, 1993 by decision of the United States Congress, after a production of 20 5-cm aperture, 15-m long superconducting dipole magnet prototypes, aimed at both the HEB and the Collider, had been completed. All of the prototypes

---

[1] Now at Commissariat à l'Energie Atomique, Centre d'Etudes de Saclay, 91191 Gif-sur-Yvette Cédex, France.
[2] Now at KEK, National Laboratory for High Energy Physics, 1-1 Oho, Tsukuba-shi, Ibaraki-ken, 305, Japan.
* Operated by the Universities Research Association, Inc. for the U.S. Department of Energy under Contract No. DE- AC35-89ER40486.

were thoroughly cold-tested, and it was found that the quench performance at 4 A/s [2-6], and the geometric field errors [7-11] were quite satisfactory. It appeared, however, that most of the prototypes exhibited a strong ramp rate sensitivity which was not acceptable for HEB operations [12-15]. The ramp-rate sensitivity of these magnets fell in at least two categories that could be correlated to the manufacturer and production batch of the strands used for the inner-coil cables. The first category, referred to as *type-A*, is characterized by a strong quench-current degradation at high ramp rates, usually accompanied by large distortions of the multipole fields and large energy losses. The second category, referred to as *type-B*, is characterized by a sudden drop of quench current at low ramp rates, followed by a much milder degradation at larger rates. The multipole fields of the type-B magnets show little ramp-rate sensitivity, and the energy losses are smaller than for the type-A magnets. The test data from the type-A magnets suggests the presence of large eddy currents in the magnet coils while ramping, while the type-B behavior may result from anomalies in the transport-current repartition among the strands of the Rutherford-type cables used in these magnets [14,15].

In this paper, we develop models describing the various types of eddy currents in superconducting particle accelerator magnets along with their effects on the magnets' performance. These models were originally developed to explain the type-A behavior observed on the SSC dipole magnet prototypes. Their formulations, however, are general enough to be applied to most of the existing designs of superconducting particle accelerator magnets using flat, two-layer, Rutherford-type cables. Detailed interpretations of the ramp rate sensitivity data from the SSC prototypes are presented elsewhere [15].

2. EDDY-CURRENT SOURCES

Let us start by reviewing the possible sources of eddy currents in SSC-type magnets.

As is the case for most superconducting particle accelerator magnets, the SSC magnet coils are wound with Rutherford-type conductors, which consist of a few tens of strands twisted together and shaped into a flat, two-layer, slightly keystoned cable. The strands themselves consist of thousands of superconducting filaments twisted together and embedded in a matrix of high purity copper. At liquid-helium temperature, the resistivity of high-purity copper becomes very low, eventually resulting in filament coupling [16]. Although the twisting of the filaments helps to reduce this effect, noticeable eddy currents flowing from one filament to another through the copper matrix can still be generated when the strands are subjected to a varying field. These eddy currents are referred to as *intra-strand* eddy currents.

Furthermore, the mid-thickness of the two-layer cable is smaller than twice the strand diameter and the contact surfaces at the crossovers between the strands of the two layers can be relatively large. Also, during magnet assembly, the coils are pre-compressed azimuthally [17]. Large pressures are thus applied perpendicularly to the cables that keep the strands firmly in contact. The large contact surfaces and high pressures eventually result in low contact resistances at the strand crossovers that couple the cable strands. Loops are thus formed where significant eddy currents can take place when the cable is subjected to a varying field. These eddy currents are referred to as *inter-strand* eddy currents.

In addition, eddy currents are generated in all the other conductive components of the magnet, such as copper wedges and iron-yoke laminations, but these can be shown to only provide minor contributions [18].

This brief review thus points towards the cable and the cable strands as the largest sources of eddy currents. We shall now develop numerical models enabling one to compute the intra-strand and inter-strand eddy currents generated within a magnet coil along with the power they dissipate and the effects they produce on the multipole fields. These models, however, require a computation of the transport-current field. We thus shall begin by describing how to compute the transport-current field.

3. COMPUTING THE TRANSPORT-CURRENT FIELD

In the long, almost straight, section of the magnet, the field can be considered as two-dimensional. Let (O,x,y,z) designate a rectangular coordinate system such that the z-axis is parallel to the ideal beam orbit and O is at the magnet center. It is convenient to introduce the complex function, **B**, defined as

$$\mathbf{B}(x+\mathrm{i}y) = B_y(x,y) + \mathrm{i}B_x(x,y) \ , \tag{1a}$$

where B_x and B_y are the x- and y-components of the field. The complex function **B** can be expanded into a Taylor's series of the form

$$\mathbf{B}(x+\mathrm{i}y) = \sum_{n=0}^{+\infty} (B_n + \mathrm{i}A_n)\left(\frac{x+\mathrm{i}y}{R_{\mathrm{ref}}}\right)^n \ . \tag{1b}$$

Here, B_n and A_n are the so-called *normal* and *skew* $2(n+1)$-pole fields and R_{ref} is the reference radius. (For SSC magnets, $R_{\mathrm{ref}} = 1$ cm.)

The symmetries of a cosine-theta distribution of conductors are such that only even normal multipole fields, also called *allowed* multipole fields, can be non-zero. In real magnets, however, manufacturing errors and non-uniformities in the material properties can result in violations of these symmetries and lead to non-zero *unallowed* multipole fields. Typical examples of such violations are: a top/bottom asymmetry, which results in a non-zero skew quadrupole field (A_1), and a left/right asymmetry, which results in a non-zero normal quadrupole field (B_1) [11].

For SSC dipole magnets, the dipole field is expected to be about 10^4 times larger than any other multipole fields. Hence, it is customary to introduce the dimensionless multipole coefficients, a_n and b_n, defined as

$$a_n = 10^4 \frac{A_n}{B_0}, \tag{2a}$$

and

$$b_n = 10^4 \frac{B_n}{B_0} \ . \tag{2b}$$

The complex transport-current field, $\mathbf{B}_t$, produced by the coil assembly of a dipole magnet like those pictured in Fig. 1 can be calculated by dividing each turn of the coil into elementary current-lines parallel to the z-axis.

Let l be an index referring to the coil layer number, k be an index referring to the turn number within a given coil layer, and j be an index referring to the current-line number within a given turn. From Ampere's theorem, the complex field, $\mathbf{B}_t^{l,k,j}(x+\mathrm{i}y)$ generated by a given current-line is

$$\mathbf{B}_t^{l,k,j}(x+\mathrm{i}y) = \frac{\mu_0 I_t^{l,k,j}}{2\pi\left[(x+\mathrm{i}y) - \mathbf{z}_t^{l,k,j}\right]} \ , \tag{3}$$

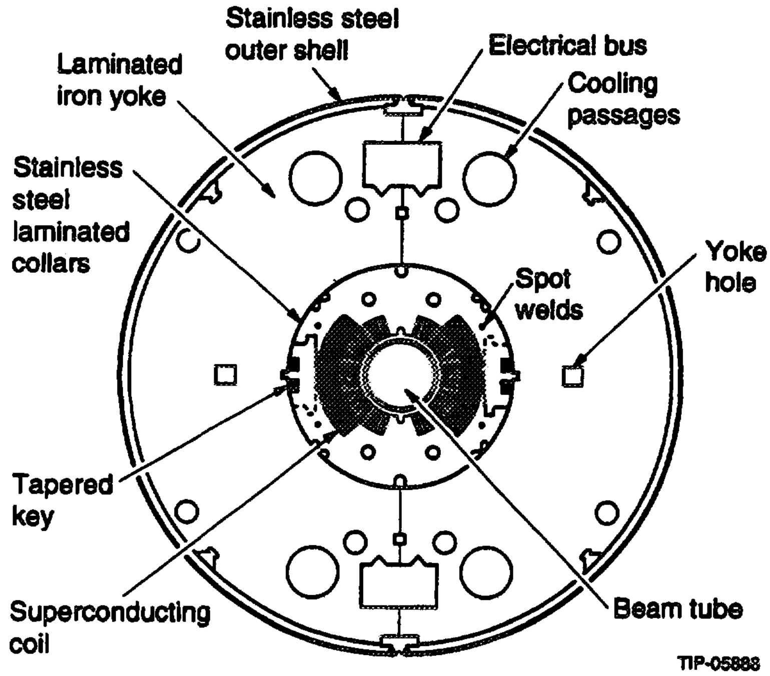


Fig. 1 5-cm aperture SSC/FNAL dipole magnet prototype

where μ_0 is the permeability of vacuum, $I^{l,k,j}$ is the current-line intensity and $\mathbf{z}_t^{l,k,j} = x_t^{l,k,j} + iy_t^{l,k,j}$ is the current-line position in the complex plane. The current-line intensity is related to the transport-current intensity, I_t, by

$$I_t^{l,k,j} = \varepsilon_{l,k} \frac{I_t}{n_{l,k}} , \tag{4}$$

where $n_{l,k}$ is the number of current-lines used to represent turn k of layer l, $\varepsilon_{l,k} = -1$ for a current-line in quadrant 1 or 4 $\left[\mathrm{Re}\left(\mathbf{z}_t^{l,k,j}\right) > 0\right]$, and $\varepsilon_{l,k} = +1$ for a current-line in quadrant 2 or 3 $\left[\mathrm{Re}\left(\mathbf{z}_t^{l,k,j}\right) < 0\right]$.

If the current-line is located inside a circular iron yoke of radius, R_y, the contribution of the iron yoke can be shown to be the same as that of a mirror current-line of intensity, $I_{t,m}^{l,k,j}$, and position, $\mathbf{z}_{t,m}^{l,k,j}$, where [19]

$$I_{t,m}^{l,k,j} = \frac{\mu - 1}{\mu + 1} I_t^{l,k,j}, \tag{5a}$$

and

$$\mathbf{z}_{t,m}^{l,k,j} = \frac{R_y^2}{\left(\mathbf{z}_t^{l,k,j}\right)*} . \tag{5b}$$

Here μ designates the magnetic permeability of the iron yoke and $\left(\mathbf{z}_t^{l,k,j}\right)*$ designates the complex conjugate of $\mathbf{z}_t^{l,k,j}$. Note that the mirror image method is only applicable if the iron-yoke permeability is uniform.

The transport-current field produced by the magnet assembly is obtained by summing the contributions of all the elementary current lines. In the current range where the iron yoke is not saturated, $\mathbf{B}_t$ may thus be written

$$\mathbf{B}_t(x+\mathrm{i}y) = I_t \mathbf{T}_t(x+\mathrm{i}y) \tag{6}$$

where $\mathbf{T}_t$ is the local complex transfer function given by

$$\mathbf{T}_t(x+\mathrm{i}y) = \frac{\mu_0}{2\pi}\sum_{l=1}^{2}\sum_{k=1}^{K_l}\frac{\varepsilon_{l,k}}{n_{l,k}}\sum_{j=1}^{n_{l,k}}\left[\frac{1}{(x+\mathrm{i}y)-\mathbf{z}_t^{l,k,j}} + \frac{\mu-1}{\mu+1}\frac{1}{(x+\mathrm{i}y)-\mathbf{z}_{t,m}^{l,k,j}}\right]. \tag{7}$$

Here, K_l designates the total number of turns in layer l. Equation (7) shows that, in the current range where the iron yoke is not saturated, $\mathbf{T}_t$ is independent of the current and is only determined by coil geometry and by the inner radius and the magnetic permeability of the iron yoke. (In most practical cases, μ can be assumed to be infinite.)

For $|x+\mathrm{i}y| < |\mathbf{z}_t^{l,k,j}|$, the first fraction of Eq. (7) can be expanded as a series

$$\frac{1}{(x+\mathrm{i}y)-\mathbf{z}_t^{l,k,j}} = -\frac{1}{\mathbf{z}_t^{l,k,j}}\frac{1}{1-\dfrac{x+\mathrm{i}y}{\mathbf{z}_t^{l,k,j}}} = -\frac{1}{\mathbf{z}_t^{l,k,j}}\sum_{n=0}^{\infty}\left(\frac{x+\mathrm{i}y}{\mathbf{z}_t^{l,k,j}}\right)^n , \tag{8}$$

and, for $|x+\mathrm{i}y| < |\mathbf{z}_{t,m}^{l,k,j}|$ a similar series expansion can be derived for the second fraction of Eq. (7).

Hence, within the coil aperture, $\mathbf{B}_t$ may be written

$$\mathbf{B}_t(x+\mathrm{i}y) = \sum_{n=0}^{\infty}\left(B_{n,t}+\mathrm{i}A_{n,t}\right)\left(\frac{x+\mathrm{i}y}{R_{\mathrm{ref}}}\right)^n , \tag{9}$$

where R_{ref} is the reference radius introduced in Eq. (1), and $A_{n,t}$ and $B_{n,t}$ are the geometric components of the $(2n+1)$-pole fields given by

$$B_{n,t}+\mathrm{i}A_{n,t} = -\frac{\mu_0 I_t}{2\pi R_{\mathrm{ref}}}\sum_{l=1}^{2}\sum_{k=1}^{K_l}\frac{\varepsilon_{l,k}}{n_{l,k}}\sum_{j=1}^{n_{l,k}}\left[\left(\frac{R_{\mathrm{ref}}}{\mathbf{z}_t^{l,k,j}}\right)^{n+1} + \frac{\mu-1}{\mu+1}\left(\frac{R_{\mathrm{ref}}}{\mathbf{z}_{t,m}^{l,k,j}}\right)^{n+1}\right]. \tag{10}$$

Equation (10) shows that $A_{n,t}$ and $B_{n,t}$ vary linearly as a function of I_t.

In practice, a good computational accuracy can be achieved by taking for $n_{l,k}$ an even number of the order of N_l, where N_l is the number of strands in the layer-l cable, and by representing each turn by two layers of equally spaced current-lines (see Fig. 2). In this paper, we use

$$n_{l,k} = N_l \ , \qquad \text{for } N_l \text{ even,} \tag{11a}$$

and

$$n_{l,k} = N_l - 1 \ , \qquad \text{for } N_l \text{ odd,} \tag{11b}$$

and we take for $\mathbf{z}_t^{l,k,j}$

$$\mathbf{z}_t^{l,k,j} = \left(1 - \frac{2j-1}{n_{l,k}}\right)\left(\frac{3\mathbf{Z}_1^{l,k} + \mathbf{Z}_2^{l,k}}{4}\right) + \frac{2j-1}{n_{l,k}}\left(\frac{\mathbf{Z}_3^{l,k} + 3\mathbf{Z}_4^{l,k}}{4}\right), \quad \text{for } j, 1 \leq j \leq \frac{n_{l,k}}{2}, \tag{12a}$$

and

$$\mathbf{z}_t^{l,k,j} = \left(1 - \frac{2j - n_{l,k} - 1}{n_{l,k}}\right)\left(\frac{\mathbf{Z}_1^{l,k} + 3\mathbf{Z}_2^{l,k}}{4}\right) + \frac{2j - n_{l,k} - 1}{n_{l,k}}\left(\frac{3\mathbf{Z}_3^{l,k} + \mathbf{Z}_4^{l,k}}{4}\right),$$

$$\text{for } j, \frac{n_{l,k}}{2} + 1 \leq j \leq n_{l,k}, \tag{12b}$$

where $\mathbf{Z}_l^{l,k}$ through $\mathbf{Z}_4^{l,k}$ designate the positions in the complex plane of the four corners of turn k of layer l as defined in Fig. 2.

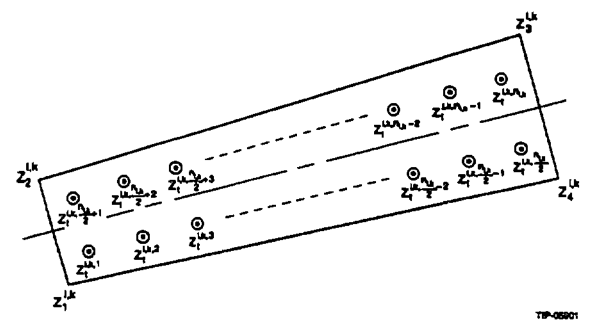


Fig. 2 Current-line model for the calculation of the transport-current field produced by a given turn of a two-layer cosine-theta coil wound with Rutherford-type cables

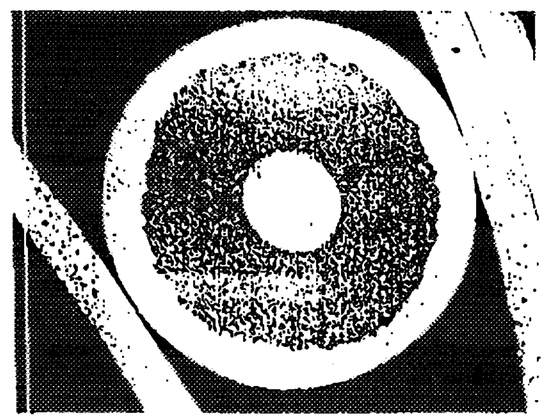


Fig. 3 Micrograph of a SSC inner strand. The strand consists of three concentric regions: 1) a copper inner core, 2) an annular superconducting multifilamentrary composite, and 3) a copper outer sheath

4. MODEL FOR INTRA-STRAND EDDY CURRENTS

4.1 Case of a single strand

Let us first consider a rectilinear and infinite strand, exposed to a uniform, time-dependent, complex external field, $\mathbf{B}_t(t)$, perpendicular to its axis. The shielding currents generated within the strand can be shown to produce a complex magnetic moment, $\mathbf{M}_s$, given by [16]

$$\mathbf{M}_s = M_{sy} + \mathrm{i}M_{sx} = \frac{2\pi R_3^2}{\mu_0}\frac{\mathrm{d}\mathbf{B}_t}{\mathrm{d}t}\tau_s \tag{13}$$

where M_{sx} and M_{sy} are the x- and y-components of the magnetic moment, R_3 is the strand radius, and τ_s is the effective time constant of the shielding currents. The power, P_s, dissipated per strand unit length by the shielding currents is given by

$$P_s = \frac{2\pi R_3^2}{\mu_0}\left|\frac{d\mathbf{B}_t}{dt}\right|^2 \tau_s \quad . \tag{14}$$

As can be seen from the micrograph presented in Fig. 3, the SSC strands consist of three concentric regions: 1) a copper inner core, of outer radius, R_1, and resistivity, ρ_1, 2) an annular multifilamentary composite, of outer radius, R_2, and transverse resistivity, ρ_2, and 3) a copper outer sheath, of resistivity, ρ_3. For such geometry, the effective time constant of the shielding currents can be estimated as [20]

$$\tau_s = \tau_{\text{core}} + \tau_{\text{composite}} + \tau_{\text{sheath}} \ , \tag{15}$$

where τ_{core} corresponds to the shielding currents generated in the inner core

$$\tau_{\text{core}} = \frac{\mu_0}{2\rho_1}\frac{R_1^2}{R_3^2}\left[1+\left(\frac{\pi}{L_s}\right)^2 R_1^2\right]\left(\frac{L_s}{2\pi}\right)^2 \quad , \tag{16a}$$

$\tau_{\text{composite}}$ corresponds to the shielding currents generated in the annular multifilamentary composite

$$\tau_{\text{composite}} = \frac{\mu_0}{2\rho_2}\frac{R_2^2 - R_1^2}{R_3^2}\left[1+\left(\frac{\pi}{L_s}\right)^2 \left(R_1^2 + R_2^2\right)\right]\left(\frac{L_s}{2\pi}\right)^2 \tag{16b}$$

and τ_{sheath} corresponds to the shielding currents generated in the outer sheath

$$\tau_{\text{sheath}} = \frac{\mu_0}{2\rho_3}\frac{R_3^2 - R_2^2}{R_3^2}\left[\frac{R_2^2}{R_2^2 + R_3^2} + \left(\frac{\pi}{L_s}\right)^2 \left(R_2^2 + R_3^2\right)\right]\left(\frac{L_s}{2\pi}\right)^2 \quad . \tag{16c}$$

In Eqs. (16a) through (16c), L_s designates the strand twist pitch.

The resistivity of the inner core and of the outer sheath can be taken equal to the resistivity of bulk copper, ρ_b, which depends on the bulk copper Residual Resistivity Ratio, noted RRR_b (defined as the ratio of the resistivity at 273 K and 10 K to the resistivity at 10 K and 0 T), the local temperature, T, and the local field strength, B_t. For the case of high contact resistance between the superconducting filaments and the copper matrix, the transverse resistivity of the multifilamentary composite can be evaluated as [21]

$$\rho_2 = \frac{1+\lambda}{1-\lambda}\rho_m \quad , \tag{17}$$

where λ is the fraction of superconductor in the composite and ρ_m is the copper matrix resistivity.

For an hexagonal lattice of filaments, we have

$$\lambda = \frac{\pi}{2\sqrt{3}}\left(1+\frac{s}{d}\right)^{-2} \quad , \tag{18}$$

where d is the filament diameter and s is the filament spacing. In the case of SSC strands, d = 6 μm, s = 1 μm, and $\lambda \approx 0.67$.

In the case of SSC strands, the determination of ρ_m is complicated by the fact that, at low temperatures, the interfilament spacing is smaller than the electron mean free path in copper. Hence, at low temperatures, the filaments of the composite act as boundary scatterers, resulting in an enhancement of the copper matrix resistivity with respect to bulk copper. This enhancement can be estimated as [22]

$$\rho_m = \rho_b + \frac{6.56\ 10^{-16}}{s} \tag{19}$$

For $\mathrm{RRR}_b = 200$, we get: ρ_b(10 K, 0 T) = 7.8 10^{-11} Ωm and ρ_m(10 K, 0 T) =7.3 10^{-10} Ωm, which shows that, at low temperatures, the copper matrix can be 10 times more resistive than the inner core or the outer shell.

Furthermore, the RRR values quoted in Tables I(a) through I(c) of Reference [14] as measured on SSC cable short samples or on SSC magnet prototypes during cold testing correspond to

$$\mathrm{RRR} = \frac{\rho_t(273\mathrm{K},0\mathrm{T})}{\rho_t(10\mathrm{K},0\mathrm{T})}, \tag{20}$$

where ρ_t is the longitudinal resistivity of the cable. For SSC-type cables

$$\frac{1}{\rho_t} \approx \left(1 - \frac{1}{\lambda}\frac{1}{1+r_{\mathrm{cus}}}\right)\frac{1}{\rho_b} + \frac{1-\lambda}{\lambda}\frac{1}{1+r_{\mathrm{cus}}}\frac{1}{\rho_m}, \tag{21}$$

where r_{cus} is the strand copper-to-superconductor ratio.

At 273 K, $\rho_m \approx \rho_b$, and Eq. (21) can be written

$$\rho_t(273\mathrm{K},0\mathrm{T}) \approx \frac{1+r_{\mathrm{cus}}}{r_{\mathrm{cus}}}\rho_b(273\mathrm{K},0\mathrm{T})\ . \tag{22a}$$

At 10 K, however, one needs to take into account the enhancement of the copper matrix resistivity due to boundary scattering. Equation (21) then becomes

$$\frac{1}{\rho_t(10\mathrm{K},0\mathrm{T})} \approx \left(1 - \frac{1}{\lambda}\frac{1}{1+r_{\mathrm{cus}}}\right)\frac{1}{\rho_b(10\mathrm{K},0\mathrm{T})} + \frac{1-\lambda}{\lambda}\frac{1}{1+r_{\mathrm{cus}}}\frac{1}{\rho_b(10\mathrm{K},0\mathrm{T}) + \frac{6.56\ 10^{-16}}{s}}\ . \tag{22b}$$

By combining Eqs. (20), (22a), and (22b), and introducing the definition of RRR_b, we obtain

$$\mathrm{RRR} \approx \frac{\lambda(1+r_{\mathrm{cus}})-1}{\lambda r_{\mathrm{cus}}}\mathrm{RRR}_b + \frac{1-\lambda}{\lambda r_{\mathrm{cus}}}\frac{1}{\frac{1}{\mathrm{RRR}_b} + \frac{4.23\ 10^{-8}}{s}} \tag{23}$$

Here we have assumed: ρ_b(273 K, 0 T) ≈ 1.55 10^{-8} Ωm.

Equation (23) allows one to determine RRR_b from the measured RRR of the cable. As an illustration, for RRR = 200 and $r_{\mathrm{cus}} = 1.5$, we get: $\mathrm{RRR}_b \approx 290$, which corresponds to the RRR value of the raw copper thought to have been used by the SSC strand manufacturers [23]. Thus, for the magnets listed in Tables I(a) through I(c) of Reference [15] as having an RRR of the order of 200, we can conclude that coil curing was accompanied by a nearly full annealing of the copper.

Having calculated ρ_1, ρ_2, and ρ_3, we can now go back to the estimation of the effective time constant of the shielding currents. Let x_{core} designate the ratio of the inner core cross-sectional area to the strand cross-sectional area. We have

$$R_1 = \sqrt{x_{\mathrm{core}}}\,R_3 \tag{24}$$

and

$$R_2 = \sqrt{x_{\mathrm{core}} + \frac{1}{\lambda(1+r_{\mathrm{cus}})}}\,R_3 \quad . \tag{25}$$

For SSC inner strands, $R_3 = 0.404$ mm and $x_{\mathrm{core}} \approx 10\%$. With $r_{\mathrm{cus}} = 1.5$, we get: $R_1 \approx 0.128$ mm and $R_2 \approx 0.338$ mm, and with $L_s = 13$ mm, Eqs. (16a) through (16c) become

$$\tau_{\mathrm{core}} = \frac{2.69\,10^{-13}}{\rho_b(\mathrm{RRR}_b, T, B_t)} , \tag{26a}$$

$$\tau_{\mathrm{composite}} = \frac{3.26\,10^{-13}}{\rho_b(\mathrm{RRR}_b, T, B_t) + 6.56\,10^{-10}} , \tag{26b}$$

and

$$\tau_{\mathrm{sheath}} = \frac{3.45\,10^{-13}}{\rho_b(\mathrm{RRR}_b, T, B_t)} . \tag{26c}$$

These time constants can be compared to those derived from ac loss measurements on strand short samples reported in Reference [24]. For RRR = 37 (sample F-1), T = 10 K, and B = 1 T: $\tau_{\mathrm{core}} = 0.8$ ms, $\tau_{\mathrm{composite}} = 0.3$ ms, and $\tau_{\mathrm{sheath}} = 1.0$ ms, which gives: $\tau_s = 2.1$ ms, while the ac loss measurements yielded 3.6 ms. For RRR = 126 (sample F-2), T = 10 K, and B = 1 T: $\tau_{\mathrm{core}} = 2.2$ ms, $\tau_{\mathrm{composite}} = 0.4$ ms, and $\tau_{\mathrm{sheath}} = 2.9$ ms, which gives: $\tau_s = 5.5$ ms, while the ac loss measurements yielded 6.5 ms. The predicted values thus appear to be of the right order of magnitude, but somewhat lower than the values derived from the test data.

4.2 Case of a magnet coil

Having treated the case of a rectilinear and infinite strand, we can now go back to the case of a magnet coil. In the current-line model described in section 3, the effects of intra-strand eddy currents can be calculated by associating with each current-line a complex magnetic moment, $\mathbf{M}_{\mathrm{intra}}^{l,k,j}$, defined as

$$\mathbf{M}_{\mathrm{intra}}^{l,k,j} = -\frac{N_l}{n_{l,k}}\,\frac{2\pi(R_3^l)^2}{\mu_0}\,\frac{d\mathbf{B}_t\left(\mathbf{z}_t^{l,k,j}\right)}{dt}\,\tau_s^{l,k,j} , \tag{27}$$

where R_3^l designates the outer radius of the strands of the layer-l cable, and $t_t^{l,k,j}$ designates the effective time constant of the shielding currents estimated for a strand of the layer-l cable at a field of $\mathbf{B}_t(\mathbf{z}_t^{l,k,j}) = |\mathbf{B}_t(\mathbf{z}_t^{l,k,j})|$. When ramping the transport-current, the time derivative of $\mathbf{B}_t$ is simply given by

$$\frac{d\mathbf{B}_t(\mathbf{z}_t^{l,k,j})}{dt} = \frac{dI_t}{dt}\mathbf{T}_t(\mathbf{z}_t^{l,k,j}). \tag{28}$$

The complex field, $\mathbf{B}_{\text{intra}}^{l,k,j}$, produced by $\mathbf{M}_{\text{intra}}^{l,k,j}$ may be derived by representing the magnetic moment by the current-line doublet shown in Fig. 4. This doublet consists of a current-line of intensity $(-I_d^{l,k,j})$ located at $\mathbf{z}_t^{l,k,j}$, and a current-line of intensity $(+I_d^{l,k,j})$ located at $(\mathbf{z}_t^{l,k,j} + \mathbf{d}_{\text{intra}}^{l,k,j})$, where $\mathbf{d}_{\text{intra}}^{l,k,j}$ is perpendicular to the orientation of the vector magnetic moment. The doublet strength, defined as $|I_d^{l,k,j}\mathbf{d}_{\text{intra}}^{l,k,j}|$, is taken to be equal to $M_{l,k,j} = |\mathbf{M}_{\text{intra}}^{l,k,j}|$. Let $M_{\text{intra},x}^{l,k,j}$ and $M_{\text{intra},y}^{l,k,j}$ designate the x- and y-components of the vector magnetic moment, we thus have

$$I_d^{l,k,j}\mathbf{d}_{\text{intra}}^{l,k,j} = \mathrm{i}\left(M_{\text{intra},x}^{l,k,j} + \mathrm{i}M_{\text{intra},y}^{l,k,j}\right) = -\left(\mathbf{M}_{\text{intra}}^{l,k,j}\right)^* \tag{29}$$

From the expressions derived above for the field produced by a current-line within a circular iron yoke, it follows that, for $|x + \mathrm{i}y| < |\mathbf{z}_t^{l,k,j}|$ and $|x + \mathrm{i}y| < |\mathbf{z}_t^{l,k,j} + \mathbf{d}_{\text{intra}}^{l,k,j}|$, $\mathrm{B}_{\text{intra}}^{l,k,j}$ is given by

$$\mathbf{B}_{\text{intra}}^{l,k,j}(x+\mathrm{i}y) = \sum_{n=0}^{\infty}\left(B_{n,\text{intra}}^{l,k,j} + \mathrm{i}A_{n,\text{intra}}^{l,k,j}\right)\left(\frac{x+\mathrm{i}y}{R_{\text{ref}}}\right)^n \tag{30}$$

where

$$B_{\text{n,intra}}^{\text{l,k,j}} + \mathrm{i}A_{\text{n,intra}}^{\text{l,k,j}} = -\frac{\mu_0 I_d^{l,k,j}}{2\pi}R_{\text{ref}}^n\left\{-\left(\mathbf{z}_t^{l,k,j}\right)^{-(n+1)} + \left(\mathbf{z}_t^{l,k,j} + \mathbf{d}_{\text{intra}}^{l,k,j}\right)^{-(n+1)} + \frac{\mu-1}{\mu+1}\,\frac{-\left[\left(\mathbf{z}_t^{l,k,j}\right)^*\right]^{n+1} + \left[\left(\mathbf{z}_t^{l,k,j} + \mathbf{d}_{\text{intra}}^{l,k,j}\right)^*\right]^{n+1}}{R_y^{2(n+1)}}\right\}. \tag{31}$$

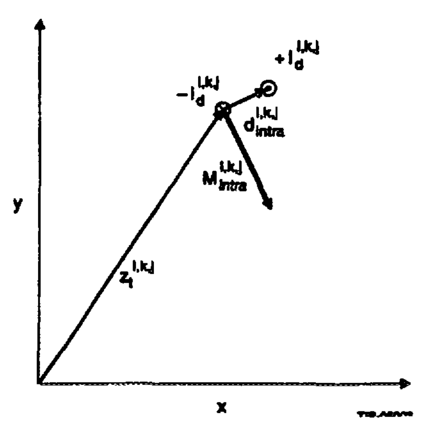


Fig. 4 Representation of a vector magnetic moment by a current line doublet

For $|d^{l,k,j}_{\text{intra}}| << |z^{l,k,j}_t|$, we can write

$$\left(\mathbf{z}_t^{l,k,j}+\mathbf{d}_{\text{intra}}^{l,k,j}\right)^{-(n+1)}=\left(\mathbf{z}_t^{l,k,j}\right)^{-(n+1)}\left(1+\frac{\mathbf{d}_{\text{intra}}^{l,k,j}}{\mathbf{z}_t^{l,k,j}}\right)^{-(n+1)}\approx\left(\mathbf{z}_t^{l,k,j}\right)^{-(n+1)}-(n+1)\,\mathrm{d}_{\text{intra}}^{l,k,j}\left(\mathbf{z}_t^{l,k,j}\right)^{-(n+2)} \tag{32a}$$

Similarly, we have

$$\left[\left(\mathbf{z}_t^{l,k,j}+\mathbf{d}_{\text{intra}}^{l,k,j}\right)^*\right]^{n+1}\approx\left[\left(\mathbf{z}_t^{l,k,j}\right)^*\right]^{n+1}+(n+1)\,\left(\mathbf{d}_{\text{intra}}^{l,k,j}\right)^*\left[\left(\mathbf{z}_t^{l,k,j}\right)^*\right]^n\ . \tag{32b}$$

Combining Eqs. (29), (31), (32a), and (32b) then yields

$$B_{n,\text{intra}}^{l,k,j}+\mathrm{i}A_{n,\text{intra}}^{l,k,j}\approx-\frac{\mu_0(n+1)}{2\pi R_{\text{ref}}^2}\left\{\left(\mathbf{M}_{\text{intra}}^{l,k,j}\right)*\left(\frac{R_{\text{ref}}}{\mathbf{z}_t^{l,k,j}}\right)^{(n+2)}-\frac{\mu-1}{\mu+1}\mathbf{M}_{\text{intra}}^{l,k,j}\frac{R_{\text{ref}}{}^{n+2}\left[\left(\mathbf{z}_t^{l,k,j}\right)^*\right]^n}{R_y^{2(n+1)}}\right\}. \tag{33}$$

It can be seen from Eq. (33) that the dependence of $A^{l,k,j}_{n,\text{intra}}$ and $B^{l,k,j}_{n,\text{intra}}$ on the ramp rate is determined by the dependence of $\mathbf{M}^{l,k,j}_{\text{intra}}$, which were shown to vary linearly as a function of $(\mathrm{d}I_t/\mathrm{d}t)$.

The overall complex field, $\mathbf{B}_{\text{intra}}$, produced by the intra-strand eddy currents generated within the magnet coil is obtained by summing the contributions of all the magnetic moments. Within the coil aperture, $\mathbf{B}_{\text{intra}}$ may be written as

$$\mathbf{B}_{\text{intra}}(x+\mathrm{i}y)=\sum_{n=0}^{\infty}\left(B_{n,\text{intra}}+\mathrm{i}A_{n,\text{intra}}\right)\left(\frac{x+\mathrm{i}y}{R_{\text{ref}}}\right)^n\ , \tag{34}$$

where $A_{\text{n,intra}}$ and $B_{\text{n,intra}}$ are the intra-strand eddy current components of the $(2n+1)$-pole fields given by

$$B_{n,\text{intra}}+\mathrm{i}A_{n,\text{intra}}=\sum_{l=1}^{2}\sum_{k=1}^{K_l}\sum_{j=1}^{n_{l,k}}\mathrm{B}_{n,\text{intra}}^{l,k,j}+\mathrm{i}A_{n,\text{intra}}^{l,k,j}. \tag{35}$$

Similarly, the overall power, P_{intra}, dissipated per coil unit length by the intra-strand eddy currents is given by

$$P_{\text{intra}}=\sum_{l=1}^{2}\sum_{k=1}^{K_l}\sum_{j=1}^{n_{l,k}}P_{\text{intra}}^{l,k,j}, \tag{36}$$

where $P^{l,k,j}_{\text{intra}}$ is the power loss per unit length associated with $\mathbf{M}^{l,k,j}_{\text{intra}}$

$$P_{\text{intra}}^{l,k,j}=\frac{N_l}{n_{l,k}}\frac{2\pi\left(R_3^l\right)^2}{\mu_0}\left|\frac{\mathrm{d}\mathbf{B}_t\left(z_t^{l,k,j}\right)}{\mathrm{d}t}\right|^2t_s^{l,k,j} \tag{37}$$

The above expressions show that $A_{n,\mathrm{intra}}$ and $B_{n,\mathrm{intra}}$ vary linearly as a function of $(\mathrm{d}I_t/\mathrm{d}t)$, while P_{intra} is proportional to $(\mathrm{d}I_t/\mathrm{d}t)^2$.

4.3 Discussion

The model described in the previous section enables one to calculate the field distortions and the power dissipation due to intra-stand eddy currents. Let us first discuss the field distortions.

The distribution of magnetic moments determined above follows the symmetries of the transport-current field. Hence, if we neglect the geometric errors and if the strand properties are uniform, only the allowed multipole fields (*i.e.*, in a dipole magnet, the even normal multipole fields) are affected. For a given magnetic design and given strand geometries, and assuming that the RRR is the same for the two inner and two outer coils, the dependence of $B_{2m,\mathrm{intra}}$ on $(\mathrm{d}I_t/\mathrm{d}t)$ can be expressed as

$$B_{2m,\mathrm{intra}} = B^0_{2m,\mathrm{intra}}(\mathrm{RRR}, I_t)\frac{\mathrm{d}I_t}{\mathrm{d}t}. \tag{38}$$

Here, the dependence of $B^0_{2m,\mathrm{intra}}$ on I_t comes from the fact that, in Eq. (27), $\mathbf{M}^{l,k,j}_{\mathrm{intra}}$ is a function of $\tau_s^{l,k,j}$, which itself is a function of ρ_b, and, thus, depends on the transport-current field at $\mathbf{z}_t^{l,k,j}$.

Table 1 summarizes the values of $B^0_{0,\mathrm{intra}}$, $B^0_{2,\mathrm{intra}}$, and $B^0_{4,\mathrm{intra}}$ obtained for the 5-cm aperture SSC dipole magnet design at 2000 A. In these computations, we assumed: $r_{\mathrm{cus}} = 1.5$ and $x_{\mathrm{core}} = 0.1$ for the inner-layer cable, and $r_{\mathrm{cus}} = 1.8$ and $x_{\mathrm{core}} = 0.1$ for the outer-layer cable. All the other parameters were taken at their design values. The various lines of Table 1 correspond to various RRR values and lists the effective time constants of the shielding currents calculated in average over the inner-layer and outer-layer line-currents. Intra-strand eddy currents appear to have a relatively small effect on the multipole fields (for instance, the values of $B^0_{2,\mathrm{intra}}$, and $B^0_{4,\mathrm{intra}}$ are one order of magnitude smaller than the slopes of multipole fields versus ramp rate, S_{B_2} and S_{B_4}, determined from magnetic measurements performed on SSC dipole magnet prototypes and reported in Table IV of Reference [15]; note also that the signs of $B^0_{2,\mathrm{intra}}$ and $B^0_{4,\mathrm{intra}}$ are opposite to the signs of S_{B_2} and S_{B_4}).

Table 1
Predicted field distortions due to intra-strand eddy currents for 5-cm aperture SSC dipole magnets at 2000 A

RRR	t_s (ms)		$B^0_{0,\mathrm{intra}}$	$B^0_{2,\mathrm{intra}}$	$B^0_{4,\mathrm{intra}}$
	inner[a]	outer[b]	(Gauss/(A/s))	((Gauss/(A/s))	(Gauss/(A/s))
100	4.1	4.9	$-3.2\ 10^{-3}$	$-1.0\ 10^{-3}$	$0.082\ 10^{-3}$
150	5.0	6.6	$-3.9\ 10^{-3}$	$-1.2\ 10^{-3}$	$0.098\ 10^{-3}$
200	5.7	8.0	$-4.5\ 10^{-3}$	$-1.4\ 10^{-3}$	$0.11\ 10^{-3}$

[a] Average over inner-layer line-currents
[b] Average over outer-layer line-currents

Let us now discuss the power dissipation. Similarly to the multipole fields, the dependence of P_{intra} on $(\mathrm{d}I_t/\mathrm{d}t)$ can be expressed as

$$P_{\text{intra}} = P^0_{\text{intra}}(\text{RRR}, I_t)\left(\frac{dI_t}{dt}\right)^2 . \tag{39}$$

Figure 5 presents plots of P^0_{intra} as a function of I_t for various RRR values. The power dissipation appears to decrease as a function of transport current as can be expected from the effects of magneto-resistance.

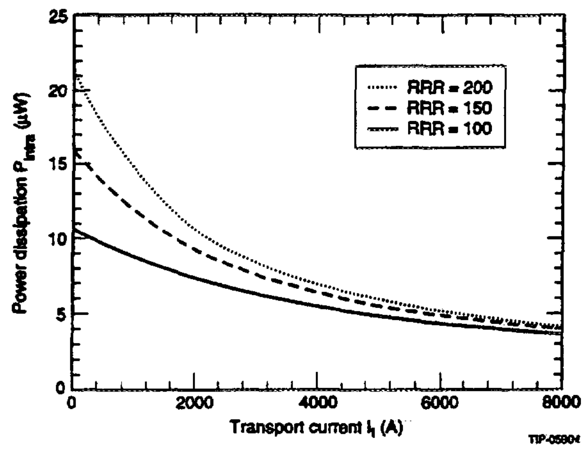


Fig. 5 Computed power dissipated by intra-strand eddy currents per unit length of 5-cm aperture SSC dipole magnets as a function of transport current and coil RRR. (The two inner and two outer coils are assumed to have the same RRR.)

The energy, E_{intra}, dissipated per coil unit length during a current cycle can be estimated as

$$E_{\text{intra}} = \int_{\text{cycle}} P_{\text{intra}}\, dt = E^0_{\text{intra}}(\text{RRR, cycle})\frac{dI_t}{dt} , \tag{40}$$

where

$$E^0_{\text{intra}}(\text{RRR, cycle}) = \int_{\text{cycle}} P^0_{\text{intra}}(\text{RRR}, I_t)|dI_t| . \tag{41}$$

For a 500-5000-500 A cycle, we get

$$E^0_{\text{intra}}(\text{RRR} = 100) = 61.1\ 10^{-3}\text{J/m/(A/s)} , \tag{42a}$$

$$E^0_{\text{intra}}(\text{RRR} = 150) = 76.4\ 10^{-3}\text{J/m/(A/s)} , \tag{42b}$$

and

$$E^0_{\text{intra}}(\text{RRR} = 200) = 87.5\ 10^{-3}\text{J/m/(A/s)} . \tag{42c}$$

For a SSC 5-cm aperture 15-m long magnet, these values scale to numbers in the range 0.9 to 1.3 J/(A/s), which appear to be of the same order of magnitude as the smallest slopes of energy loss versus ramp rate measured on SSC dipole magnet prototypes and reported in Table V of Reference [15]. Given that the time constants we calculated were smaller than those derived from test data on strand short samples, the power, and thus, the energy, we compute may be somewhat underestimated. Hence, we conclude that intra-strand eddy current losses may account for a significant fraction of the small losses measured on some of the SSC dipole magnet prototypes, while they cannot account for the larger losses measured on the other prototypes.

5. MODEL FOR INTER-STRAND EDDY CURRENTS

5.1 Case of a single cable

Let us first consider a single Rutherford-type cable, exposed to a time-dependent external field. As we described earlier, and as is shown in Fig. 6, the Rutherford-type cables used in SSC magnets consist of a few tens of strands, twisted together, and shaped into a flat, two-layer, slightly-keystoned conductor. The strands themselves are straight, except at the cable edges, where they are bent in a hairpin-like manner to ramp from one layer to the other. In such geometry, there are two types of inter-strand contacts: 1) line contacts between adjacent strands, and 2) surface contacts at the crossovers between strands of the two layers.

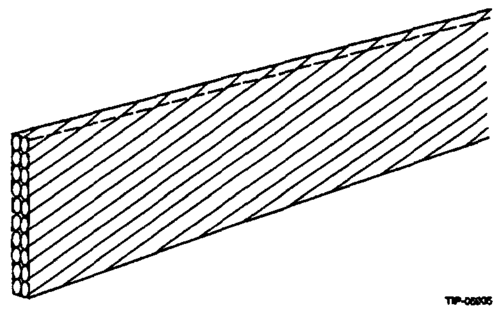


Fig. 6 Rutherford-type cable used in superconducting particle accelerator magnets

Following References [25] through [27], we assume that the crossover contact resistances are much smaller than the line contact resistances, and that the latter can be neglected. Hence, the smallest loops, where inter-strand eddy currents can be generated, are constituted by two adjacent strands of one layer crossing over two adjacent strands of the other layer. These loops are referred to as *elementary* loops, and the cable is represented by the model circuit of Fig. 7.

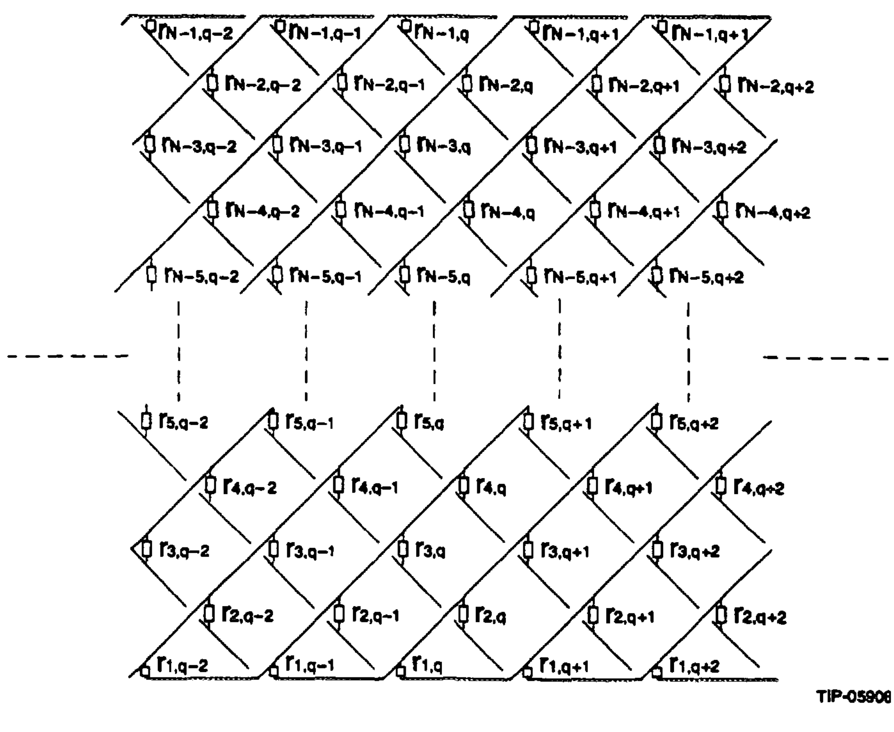


Fig. 7 Model circuit for a N-strand, two-layer Rutherford-type cable exposed to a time dependent magnetic field. The cable parameters and the magnetic field, which may vary across the cable width, are assumed to be uniform along the cable length

Two indexes are required to properly identify the crossover contacts of the model circuit of Fig. 7: one for the rows, p, where $1 \leq p \leq N-1$, and one for the columns, q, where q can assume all relative integer values. A row is defined as a series of crossover contacts on a straight line parallel to the cable axis. A column is defined as a series of crossover contacts on a zigzag line across the cable width. The rows are counted starting from the thin edge of the slightly keystoned cable. The columns are counted from left to right, starting from an arbitrary position along the cable axis. The current circulating in a given cross-over resistance, $r_{p,q}$, is referred to as $i_{p,q}$, and is counted positively when flowing from the bottom to the top layer of the cable (see Fig. 8). The magnetic flux, $\Phi_{p,q}$, through an elementary loop is identified by the indexes of the crossover resistance at its left-hand side corner and is countedpositively when penetrating the cable from the bottom. For a given column of an N-strand cable, there are (N–1) cross-over resistances, cross-over currents, and elementary loops.

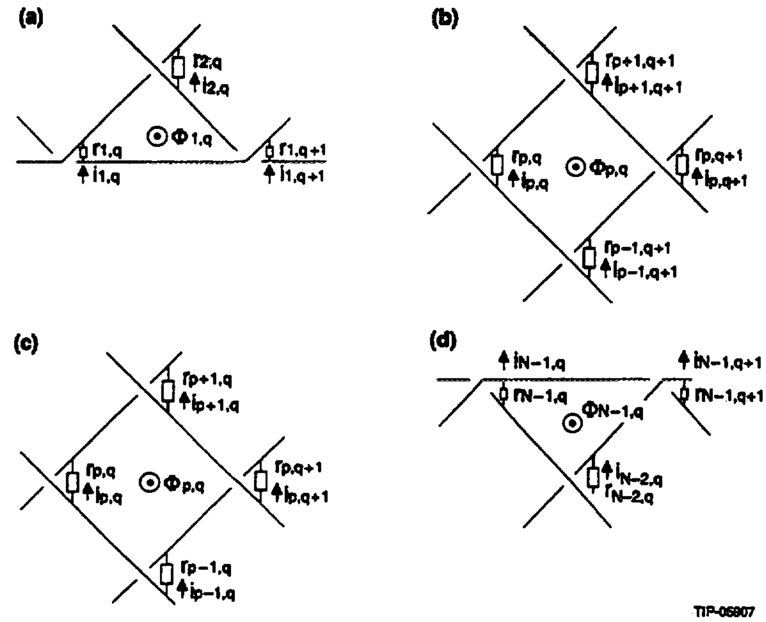


Fig. 8 Elementary loops of the model circuit for a Rutherford-type cable exposed to a time-dependent magnetic field: a) at the thin edge of the slightly keystoned cable, b) in the middle of the cable and for p even, b) in the middle of the cable and for p odd, and d) at the thick edge of the cable and for N even

The crossover currents can be determined by applying Faraday's law to the (N–1) elementary loops of the q-th column. For the loop at the cable thin edge (see Fig. 8(a)), we get

$$r_{1,q} i_{1,q} + r_{1,q+1} i_{1,q+1} - r_{2,q} i_{2,q} = \frac{\mathrm{d}\Phi_{1,q}}{\mathrm{d}t} \tag{43a}$$

For the loops in the middle of the cable, and for p even, $2 \leq p \leq N-2$ (see Fig. 8(b)), we get

$$r_{p,q} i_{p,q} + r_{p,q+1} i_{p,q+1} - r_{p+1,q+1} i_{p+1,q+1} - r_{p-1,q+1} i_{p-1,q+1} = \frac{\mathrm{d}\Phi_{p,q}}{\mathrm{d}t} , \tag{43b}$$

while for p odd, $3 \leq p \leq N-2$ (see Fig. 8(c)), we get

$$r_{p,q} i_{p,q} + r_{p,q+1} i_{p,q+1} - r_{p+1,q} i_{p+1,q} - r_{p-1,q} i_{q-1,q} = \frac{\mathrm{d}\Phi_{p,q}}{\mathrm{d}t} . \tag{43c}$$

For the loop at the thick edge of the cable, and if N is even (see Fig. 8(d)), we get

$$r_{N-1,q}i_{N-1,q} + r_{N-1,q+1}i_{N-1,q+1} - r_{N-2,q}i_{N-2,q} = \frac{\mathrm{d}\Phi_{N-1,q}}{\mathrm{d}t} , \quad (43d)$$

while if N is odd, we get

$$r_{N-1,q}i_{N-1,q} + r_{N-1,q+1}i_{N-1,q+1} - r_{N-2,q+1}i_{N-2,q+1} = \frac{\mathrm{d}\Phi_{N-1,q}}{\mathrm{d}t} \quad (43e)$$

A method to determine the general solutions of the system of Eqs. (43) is developed in Reference [27]. The problem, however, can be simplified considerably by assuming that the crossover resistances, the elementary fluxes, and the crossover currents are uniform along the cable axis, and, for p, $1 \le p \le N-1$, and for all q, satisfy the conditions

$$r_{p,q+1} = r_{p,q} , \quad (44a)$$

$$\Phi_{p,q+1} = \Phi_{p,q} , \quad (44b)$$

and

$$i_{p,q+1} = i_{p,q} . \quad (44c)$$

Then, the dependence of Eqs. (43a) through (43e) on the column number vanishes, and the system can be written

$$2r_1 i_1 - r_2 i_2 = \frac{\mathrm{d}\Phi_1}{\mathrm{d}t} \quad (45a)$$

$$2r_p i_p - r_{p+1}i_{p+1} - r_{p-1}i_{p-1} = \frac{\mathrm{d}\Phi_p}{\mathrm{d}t} , \quad \text{for } p, 2 \le p \le N-2 , \quad (45b)$$

and

$$2r_{N-1}i_{N-1} - r_{N-2}i_{N-2} = \frac{\mathrm{d}\Phi_{N-1}}{\mathrm{d}t} . \quad (45c)$$

(Note that Eqs. (45a) through (45c) apply for both N even and N odd.)

The solutions of the system of Eqs. (45) are

$$i_1 = \frac{1}{Nr_1}\sum_{m=1}^{N-1} (N-m)\frac{\mathrm{d}\Phi_m}{\mathrm{d}t} , \quad (46a)$$

and

$$i_p = p\frac{r_1}{r_p}i_1 - \frac{1}{r_p}\sum_{m=1}^{p-1}(p-m)\frac{\mathrm{d}\Phi_m}{\mathrm{d}t} , \quad \text{for } p, 2 \le p \le N-1 . \quad (46b)$$

In addition, the power, P_c, dissipated per cable unit length by the crossover currents can be derived as

$$P_c = \frac{N}{L_c} \sum_{p=1}^{N-1} r_p \, i_p^2 \quad , \tag{47}$$

where L_C is the cable pitch length.

Given the values of the elementary fluxes and of the crossover resistances, Eqs. (46a), (46b), and (47) enable one to calculate the crossover currents and their power dissipation. The next step is to calculate the eddy currents induced along the cable strands, to which we shall refer as *branch* currents.

Looking again at the model circuit of Fig. 7, let $J_{1,q}$ designate the current circulating in the branch from $r_{1,q}$ to $r_{1,q+1}$ at the thin edge of the cable (see Fig. 9(a)), and let $J_{N,q}$ designate the current circulating in the branch from $r_{N-1,q}$ to $r_{N-1,q+1}$ at the thick edge of the cable (see Fig.9(d)). For the top-layer branches in the middle of the cable, let $J_{p,q}^{\text{top}}$ designate the current circulating in the branch from $r_{p-1,q}$ to $r_{p,q}$, if p is even, $2 \le p \le N-1$, and in the branch from $r_{p-1,q}$ to $r_{p,q+1}$, if p is odd, $3 \le p \le N-1$ (see Figs. 9(b) and 9(c)). For the bottom-layer branches in the middle of the cable, let $J_{p,q}^{\text{bot}}$ designate the current circulating in the branch from $r_{p,q}$ to $r_{p-1,q+1}$, if p is even, $2 \le p \le N-1$, and in the branch from $r_{p,q}$ to $r_{p-1,q}$, if p is odd, $3 \le p \le N-1$. For a given column of a N-strand cable, there are $(N-2)$ top-layer branch currents, and $(N-2)$ bottom-layer branch currents, which, with the two edge branch currents, make a total of $2(N-1)$ branch currents.

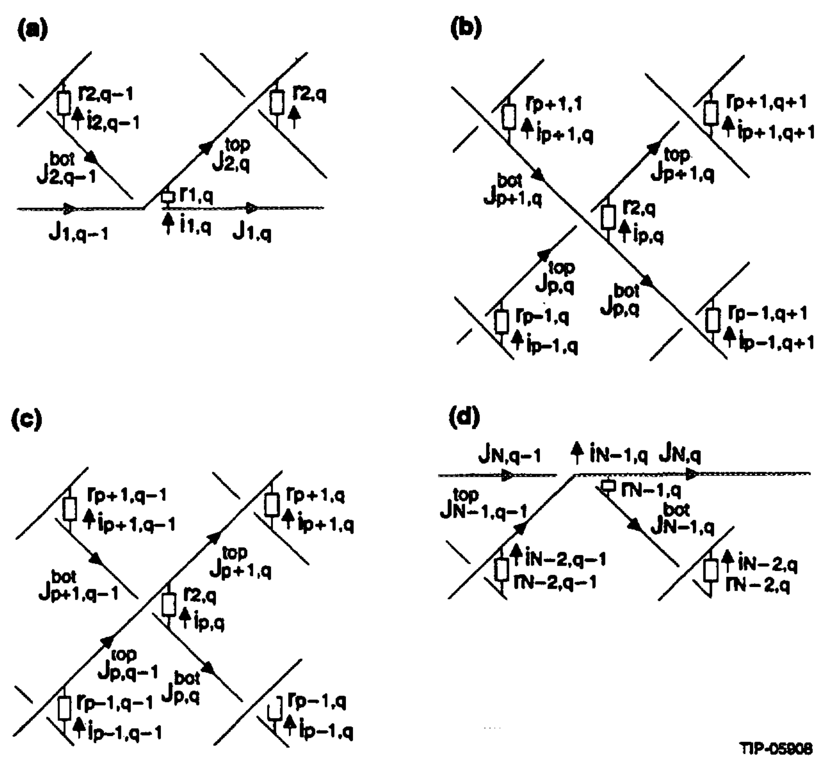


Fig. 9 Nodes of the model circuit for a Rutherford-type cable exposed to a time-dependent magnetic field: a) at the thin edge of the slightly keystoned cable, b) in the middle of the cable and for p even, c) in the middle of the cable and for p odd, and d) at the thick edge of the cable and for N even.

The branch currents can be determined by applying Kirchoff's law at the nodes at both extremities of the $(N-1)$ crossover resistances. For the top and bottom nodes of $r_{1,q}$ (see Fig. 9(a)), we get

$$J_{1,q-1} + i_{1,q} = J^{\text{top}}_{2,q} \,, \tag{48a}$$

and

$$J^{\text{bot}}_{2,q-1} - i_{1,q} = J_{1,q} \,. \tag{48b}$$

For the top and bottom nodes of $r_{p,q}$, and for p even, $2 \le p \le N-2$ (see Fig. 9(b)), we get

$$J^{\text{top}}_{p,q} + i_{p,q} = J^{\text{top}}_{p+1,q} \,, \tag{48c}$$

and

$$J^{\text{bot}}_{p+1,q} - i_{p,q} = J^{\text{bot}}_{p,q} \,, \tag{48d}$$

while for p odd, $3 \le p \le N-2$ (see Fig. 9(c)), we get

$$J^{\text{top}}_{p,q-1} + i_{p,q} = J^{\text{top}}_{p+1,q} \,, \tag{48e}$$

and

$$J^{\text{bot}}_{p+1,q-1} - i_{p,q} = J^{\text{bot}}_{p,q} \,. \tag{48f}$$

For the top and bottom nodes of $r_{N-1,q}$, and if N is even (see Fig. 9(d)), we get

$$J^{\text{top}}_{N-1,q-1} + i_{N-1,q} = J_{N,q} \,, \tag{48g}$$

and

$$J_{N,q-1} - i_{N-1,q} = J^{\text{bot}}_{N-1,q} \,, \tag{48h}$$

while if N is odd, we get

$$J^{\text{top}}_{N-1,q} + i_{N-1,q} = J_{N,q} \,, \tag{48i}$$

and

$$J_{N,q-1} - i_{N-1,q} = J^{\text{bot}}_{N-1,q} \,. \tag{48j}$$

The general solutions of the system of Eqs. (48) can be determined by following a method similar to the method developed in Reference [27] for the crossover currents. Once again, however, the problem can be simplified considerably by assuming that the crossover currents and the branch currents are uniform along the cable axis, and, for p, $1 \le p \le N-1$, and for all q, satisfy Eq. (44c) and the conditions

$$J^{\text{top}}_{p,q+1} = J^{\text{top}}_{p,q} \,, \tag{44d}$$

and

$$J^{\rm bot}_{p,q+1} = J^{\rm bot}_{p,q} \ . \tag{44e}$$

With these assumptions, the dependence of Eqs. (48a) through (48j) on the column number vanishes, and it can be shown readily that

$$J^{\rm bot}_{p} = J^{\rm top}_{p} \ , \quad \text{for} \ \ p, 2 \le p \le N-1 \ . \tag{49}$$

Let J_p designate this common value. It thus appears that, in the middle of the cable, the branch currents can be regarded as flowing along zigzag paths parallel to the cable rows.

Taking Eq. (49) into account, the above system of 2(N–1) unknowns and 2(N–1) equations can be reduced to the following system of N unknowns and (N–1) equations

$$J_p = J_{p-1} + i_p \ , \quad \text{for} \ \ p, 2 \le p \le N \ . \tag{50}$$

The system of Eq. (50) has one more unknown than equation, and thus, is undetermined. An additional equation can be written by expressing that no net current is expected to result from the cable eddy currents. Hence, the sum of the branch currents flowing through a given cross-section of the conductor should be zero

$$\sum_{p=1}^{N} J_p = 0 \ . \tag{51}$$

Combining Eq. (51) and the system of Eq. (50) yields

$$J_1 = -\frac{1}{N}\sum_{m=1}^{N-1}(N-m)\, i_m \ , \tag{52a}$$

and

$$J_p = J_1 + \sum_{m=1}^{p-1} i_m \ , \quad \text{for} \ \ p, 2 \le p \le N \ . \tag{52b}$$

Equations (52a) and (52b) enable one to calculate the branch currents as a function of the crossover currents, which, given the elementary fluxes and the crossover resistances, can be determined using Eqs. (46a) and (46b).

If we further assume that the crossover resistances and the elementary fluxes are all equal, Eqs. (46), (47), and (52) can be written

$$i_p = \frac{p(N-p)}{2r_c}\frac{\mathrm{d}\Phi}{\mathrm{d}t}, \quad \text{for} \ \ p, \ 1 \le p \le N-1 \ , \tag{53}$$

$$J_p = \frac{-N\left(N^2-1\right) + 2p(p-1)(3N-2p+1)}{24r_c}\frac{\mathrm{d}\Phi}{\mathrm{d}t} \ , \quad \text{for} \ \ p, 1 \le p \le N \ , \tag{54}$$

and

$$P_c = \frac{N^2\left(N^4-1\right)}{120L_c r_c}\left(\frac{\mathrm{d}\Phi}{\mathrm{d}t}\right)^2 \ , \tag{55}$$

where r_c and Φ designate the common values of crossover resistances and elementary fluxes. Here we have used the well known identities

$$\sum_{p=1}^{N} p = \frac{N(N+1)}{2} , \tag{56a}$$

$$\sum_{p=1}^{N} p^2 = \frac{N(N+1)(2N+1)}{6} , \tag{56b}$$

$$\sum_{p=1}^{N} p^3 = \frac{N^2(N+1)^2}{4} , \tag{56c}$$

and

$$\sum_{p=1}^{N} p^4 = \frac{N(N+1)(2N+1)\left(3N^2+3N-1\right)}{30} . \tag{56d}$$

Considering that there are $N(N–1)$ elementary loops per cable pitch length, an estimate of Φ is simply given by

$$\Phi \approx \frac{w_c L_c}{N(N-1)} B_{t,\perp} , \tag{57}$$

where w_c is the cable width, and $B_{t,\perp}$ is the supposedly uniform component of the external field perpendicular to the cable.

By combining Eqs. (54) and (57), on can derive the expression of the thin-edge branch current, J_1

$$J_1 \approx -\frac{(N+1) w_c L_c}{24 r_c} \frac{\mathrm{d}B_{t,\perp}}{\mathrm{d}t} , \tag{58}$$

which, with the thick-edge branch current, J_N, can be verified to be the most intense.

By combining Eqs. (55) and (57), one can also derive the power dissipation

$$P_c \approx \frac{\left(N^4-1\right) w_c^2 L_c}{120(N-1)^2 r_c}\left(\frac{\mathrm{d}B_{t,\perp}}{\mathrm{d}t}\right)^2 \approx \frac{N^2 w_c^2 L_c}{120 r_c}\left(\frac{\mathrm{d}B_{t,\perp}}{\mathrm{d}t}\right)^2 , \tag{59}$$

where we recognize an expression similar to that presented in Reference [25]. (Note that, in Eq. (59), L_c corresponds to the full pitch length of the cable, while Reference [25] uses the cable half-pitch length.)

For the SSC inner cable, $N = 30$, w_c = 12.34 mm, and L_c = 86 mm. Assuming a field ramp rate of 0.1 T/s and a crossover resistance of 1 mΩ, we get: $|I_1| \approx$ 137 A, and $P_c \approx$ 1.1 W/m. This simple calculation shows that the branch currents can be very large and that the crossover currents can dissipate sizable power.

5.2 Case of a magnet coil

Having treated the case of a rectilinear and infinite cable, we can now go back to the case of a magnet coil. Assuming that the elementary fluxes, the crossover resistances, the crossover currents, and the branch currents are uniform along the axis of every turn of every layer, the (N_1−1) crossover currents, $i_{1,k,p,} 1 \leq p \leq N_1 - 1$,and N_1 branch currents, $J_{1,k,p,} 1 \leq p \leq N_1 - 1$,of turn k of layer l can be computed as

$$i_{1,k,1} = \frac{1}{N_1 r_{l,k,1}} \sum_{m=1}^{N-1} (N-m) \frac{\mathrm{d}\Phi_{l,k,m}}{\mathrm{d}t} \ , \tag{60a}$$

and

$$i_{l,k,p} = p \frac{r_{l,k,1}}{r_{l,k,p}} i_{l,k,1} - \frac{1}{r_{l,k,p}} \sum_{m=1}^{p-1} (p-m) \frac{\mathrm{d}\Phi_{l,k,m}}{\mathrm{d}t} \ , \quad \text{for} \ \ p, 2 \leq p \leq N_1 - 1 \ , \tag{60b}$$

$$J_{l,k,1} = -\frac{1}{N_1} \sum_{m=1}^{N-1} (N-m) \ i_{l,k,m} \ , \tag{61a}$$

and

$$J_{l,k,p} = J_{l,k,p} + \sum_{m=1}^{p-1} i_{l,k,m} \ , \quad \text{for} \ \ p, 2 \leq p \leq N \ , \tag{61b}$$

where $r_{1,k,p}$ and $\Phi_{1,k,p}, 1 \leq N_1 - 1$ designate the (N_1-1) crossover resistances and elementary fluxes of the given turn.

The position in the complex plane, $\mathbf{z}_{\Phi}^{l,k,p}$ of the center of the p-th elementary loop of turn k of layer l can be estimated to be

$$\mathbf{z}_{\Phi}^{l,k,1} = \left(1 - \frac{1}{2(N_1 - 2)}\right) \left(\frac{\mathbf{Z}_1^{l,k} + \mathbf{Z}_2^{l,k}}{2}\right) + \frac{1}{2(N_1 - 2)} \left(\frac{\mathbf{Z}_3^{l,k} + \mathbf{Z}_4^{l,k}}{2}\right) , \tag{62a}$$

$$\mathbf{z}_{\Phi}^{l,k,p} = \left(1 - \frac{p-1}{N_1 - 2}\right) \left(\frac{\mathbf{Z}_1^{l,k} + \mathbf{Z}_2^{l,k}}{2}\right) + \frac{p-1}{N_1 - 2} \left(\frac{\mathbf{Z}_3^{l,k} + \mathbf{Z}_4^{l,k}}{2}\right), \quad \text{for} \ \ p, 2 \leq p \leq N_1 - 2, \tag{62b}$$

and

$$\mathbf{z}_{\Phi}^{l,k,N_1-1} = \frac{1}{2(N_1 - 2)} \left(\frac{\mathbf{Z}_1^{l,k} + \mathbf{Z}_2^{l,k}}{2}\right) + \left(1 - \frac{1}{2(N_1 - 2)}\right) \left(\frac{\mathbf{Z}_3^{l,k} + \mathbf{Z}_4^{l,k}}{2}\right) , \tag{62c}$$

where $\mathbf{Z}_1^{l,k}$ through $\mathbf{Z}_4^{l,k}$ designate the positions of the four corners of the given turn (see Fig. 2).

Let $E_x^{l,k}$ and $E_y^{l,k}$ designate the x- and y-components of a unit vector parallel to the midplane of turn k of layer l. We have

$$E_x^{l,k} + \mathrm{i}E_y^{l,k} = \frac{Z_3^{l,k} + Z_4^{l,k} - Z_1^{l,k} - Z_2^{l,k}}{\left|Z_3^{l,k} + Z_4^{l,k} - Z_1^{l,k} - Z_2^{l,k}\right|} \ . \tag{63}$$

The components $F_x^{l,k}$ and $F_y^{l,k}$ of a unit vector perpendicular to the midplane of turn k of layer l are given by

$$F_x^{l,k} + \mathrm{i}F_y^{l,k} = \mathrm{i}\left(E_x^{l,k} + \mathrm{i}E_y^{l,k}\right) = -E_y^{l,k} + \mathrm{i}E_x^{l,k} \ . \tag{64}$$

Hence, at position $z_\Phi^{l,k,p}$ the component of the transport-current field, $B_{t,\perp}$ perpendicular to the midplane of the cable may be derived as

$$\begin{aligned} B_{t,\perp}\left(z_\Phi^{l,k,p}\right) &= B_{t,x}F_x^{l,k} + B_{t,y}F_y^{l,k} = \mathrm{Re}\left[B_t\left(z_\Phi^{l,k,p}\right)\left(E_x^{l,k} + \mathrm{i}E_y^{l,k}\right)\right] \\ &= I_t\,\mathrm{Re}\left[\mathbf{T}_t\left(\mathbf{z}_\Phi^{l,k,p}\right)\frac{\mathbf{Z}_3^{l,k} + \mathbf{Z}_4^{l,k} - \mathbf{Z}_1^{l,k} - \mathbf{Z}_2^{l,k}}{\left|\mathbf{Z}_3^{l,k} + \mathbf{Z}_4^{l,k} - \mathbf{Z}_1^{l,k} - \mathbf{Z}_2^{l,k}\right|}\right] \end{aligned} \tag{65}$$

and its time derivative may derived as

$$\frac{\mathrm{d}B_{t,\perp}\left(z_\Phi^{l,k,p}\right)}{\mathrm{d}t} = \frac{\mathrm{d}I_t}{\mathrm{d}t}\mathrm{Re}\left[\mathrm{T}_t\left(\mathbf{z}_\Phi^{l,k,p}\right)\frac{\mathbf{Z}_3^{l,k} + \mathbf{Z}_4^{l,k} - \mathbf{Z}_1^{l,k} - \mathbf{Z}_2^{l,k}}{\left|\mathbf{Z}_3^{l,k} + \mathbf{Z}_4^{l,k} - \mathbf{Z}_1^{l,k} - \mathbf{Z}_2^{l,k}\right|}\right] \ . \tag{66}$$

Considering that there are $N_l(N_l-1)$ elementary loops per pitch length of the layer-l cable, the time derivatives of the elementary fluxes can then be estimated as

$$\frac{\mathrm{d}\Phi_{l,k,p}}{\mathrm{d}t} \approx \frac{w_c^l L_c^l}{2N_l\left(N_l - 1\right)}\frac{\mathrm{d}B_{t,\perp}\left(\mathbf{z}_\Phi^{l,k,p}\right)}{\mathrm{d}t} \ , \quad \text{for } p = 1 \text{ and } p = N_l - 1 \ , \tag{67a}$$

and

$$\frac{\mathrm{d}\Phi_{l,k,p}}{\mathrm{d}t} \approx \frac{w_c^l L_c^l}{2N_l\left(N_l - 1\right)}\frac{\mathrm{d}B_{t,\perp}\left(\mathbf{z}_\Phi^{l,k,p}\right)}{\mathrm{d}t} \ , \quad \text{for } p, 2 \le p \le N_l - 2 \ , \tag{67b}$$

where w_c^l and L_c^l designate the width and the pitch length of the layer-l cable. The difference between Eq. (67a) and Eq. (67b) arises from the fact that, in the model circuit of Fig. 7, the loops at the cable edges are smaller than the loops in the middle of the cable. It can be seen in Eqs. (67a) and (67b) that, as expected, the time derivatives of the elementary fluxes are proportional to $(\mathrm{d}I_t/\mathrm{d}t)$.

Given the values of the crossover resistances, and substituting the above expressions for the time derivatives of the elementary fluxes into Eqs. (60) and (61), it is now possible to calculate the crossover currents and the branch currents in each turn of the coil. As for the time derivatives of the elementary fluxes, these eddy currents are proportional to $(\mathrm{d}I_t/\mathrm{d}t)$.

In the above computation, we assumed that the crossover resistances, the elementary fluxes, and the inter-strand eddy currents were uniform along the cable axis. As a result, in the middle of the cable, the branch currents flow along zigzag paths parallel to the cable rows. Away from the cable, the field produced by a zigzag branch current can be approximated by the field produced by a rectilinear current-line of the same intensity and located at the centerline of the zigzag path. Hence, the field distortions caused by the inter-strand eddy currents can be calculated by associating with each turn of the coil, N_l current-lines, of intensity, $I_e^{l,k,p}$, where

$$I_e^{l,k,p} = J_{l,k,p} \ , \quad \text{for} \ \ p, 1 \le p \le N_l \ , \tag{68}$$

and located at $\mathbf{z}_e^{l,k,p}$, where

$$\mathbf{z}_e^{l,k,l} = \frac{\mathbf{Z}_1^{l,k} + \mathbf{Z}_2^{l,k}}{2} \ , \tag{69a}$$

$$\mathbf{z}_e^{l,k,p} = \left(1 - \frac{2p-3}{2(N_l-2)}\right)\left(\frac{\mathbf{Z}_1^{l,k} + \mathbf{Z}_2^{l,k}}{2}\right) + \frac{2p-3}{2(N_l-2)}\left(\frac{\mathbf{Z}_3^{l,k} + \mathbf{Z}_4^{l,k}}{2}\right), \quad \text{for} \ \ p, 2 \le p \le N_l - 1 \tag{69b}$$

and

$$\mathbf{z}_e^{l,k,N_l} = \frac{\mathbf{Z}_3^{l,k} + \mathbf{Z}_4^{l,k}}{2} \ . \tag{69c}$$

The complex field, $\mathbf{B}_{\text{inter}}^{l,k}$, produced by the branch currents generated in turn k of layer l is obtained by summing the contributions from the N_l current-lines associated with this turn, along with the contributions from the mirror images of these current lines in the iron yoke. Within the coil aperture, $\mathbf{B}_{\text{inter}}^{l,k}$ may be written as

$$\mathbf{B}_{\text{inter}}^{l,k}(x + \mathrm{i}y) = \sum_{n=0}^{\infty} \left(B_{n,\text{inter}}^{l,k} + \mathrm{i}A_{n,\text{inter}}^{l,k}\right)\left(\frac{x + \mathrm{i}y}{R_{\text{ref}}}\right)^n \ , \tag{70}$$

where

$$B_{n,\text{inter}}^{l,k} + \mathrm{i}A_{n,\text{inter}}^{l,k} = -\sum_{p=1}^{N_l} \frac{\mu_0 I_e^{l,k,p}}{2\pi R_{\text{ref}}}\left[\left(\frac{R_{\text{ref}}}{\mathbf{z}_e^{l,k,p}}\right)^{n+1} + \frac{\mu - 1}{\mu + 1}\left(\frac{R_{\text{ref}}}{\mathbf{z}_{e,m}^{l,k,p}}\right)^{n+1}\right] \ . \tag{71}$$

Here, $\mathbf{z}_{e,m}^{l,k,p}$ designates the location of the mirror image in the iron yoke of the current-line located at $\mathbf{z}_e^{l,k,p}$ (see Eq. (5b)). It can be seen from Eq. (71) that the dependence of $A_{n,\text{inter}}^{l,k}$ and $B_{n,\text{inter}}^{l,k}$ on the ramp rate is determined by the dependence of $I_e^{l,k,p}$, which were shown to vary linearly as a function of $(\mathrm{d}I_t/\mathrm{d}t)$.

The overall complex field, $\mathbf{B}_{\text{inter}}$, produced by the inter-strand eddy currents generated within the magnet coil is obtained by summing the contributions from each turn. Within the coil aperture, we simply have

$$\mathbf{B}_{\text{inter}}(x + \mathrm{i}y) = \sum_{n=0}^{\infty} \left(B_{n,\text{inter}} + \mathrm{i}A_{n,\text{inter}}\right)\left(\frac{x + \mathrm{i}y}{r_0}\right)^n \ , \tag{72}$$

where $A_{n,\text{inter}}$ and $B_{n,\text{inter}}$ are the inter-strand eddy current components of the $(2n+1)$-pole fields given by

$$B_{n,\text{inter}} + iA_{n,\text{inter}} = \sum_{l=1}^{2} \sum_{k=1}^{K_l} B_{n,\text{inter}}^{l,k} + iA_{n,\text{inter}}^{l,k} \ . \tag{73}$$

Similarly, the overall power, P_{inter}, dissipated per coil unit length by the intra-strand eddy currents is given by

$$P_{\text{inter}} = \sum_{l=1}^{2} \sum_{k=1}^{K_l} P_{\text{inter}}^{l,k} \ , \tag{74}$$

where $P_{\text{inter}}^{l,k}$ is the power loss per unit length of turn k of layer l

$$P_{\text{inter}}^{l,k} = \frac{N_l}{L_c^l} \sum_{p=1}^{N_l 1} r_{l,k,p} i_p^2 \ . \tag{75}$$

The above expressions show that $A_{n,\text{inter}}$ and $B_{n,\text{inter}}$ vary linearly as a function of (dI_t/dt), while P_{inter} is proportional to $(dI_t/dt)^2$.

5.3 Discussion

The model described in the previous section enables one to calculate the field distortions and the power dissipation due to inter-strand eddy currents. The field distortions and the power dissipation are mainly determined by the values of crossover currents and branch currents, which themselves depend on the values of elementary fluxes and crossover resistances. Hence, for a given magnetic design and given cable geometries, the only variables in the model are the ramp rate and the values of crossover resistances. Let us, for now, assume that the crossover resistance is uniform throughout the coils, and let r_c designate this common value.

Similarly to the intra-strand eddy currents, the distribution of inter-strand eddy currents determined above follows the symmetry of the transport-current field. Hence, if we neglect the geometric errors and if the crossover resistance is uniform, only the allowed multipole fields are affected. In this case, the dependence of $B_{2m,\text{inter}}$ and P_{inter} on (dI_t,dt) and r_c can be expressed as

$$B_{2m,\text{inter}} = \frac{B_{2m,\text{inter}}^0}{r_c} \frac{dI_t}{dt} \ , \tag{76}$$

and

$$P_{\text{inter}} = \frac{P_{\text{inter}}^0}{r_c} \left(\frac{dI_t}{dt} \right)^2 \ . \tag{77}$$

In Eqs. (76) and (77), the inverse proportionality of $B_{2m,\text{inter}}$ and P_{inter} to the crossover resistance comes from the fact that the branch currents and the crossover currents are themselves inversely proportional to r_c.

Table 2 summarizes the values of $B_{0,\text{inter}}^0$, $B_{2,\text{inter}}^0$, $B_{4,\text{inter}}^0$, and P_{inter}^0 obtained for a number of superconducting dipole magnet designs built around the world. Table 2 includes the 5-cm aperture SSC dipole magnet design discussed in this paper, along with the old 4-cm aperture dipole magnet design considered at the beginning of the SSC project [28], and an alternate 5-cm aperture dipole magnet design (referred to as SSC/HEB), developed by SSCL [29]. It also includes the results of computations for the Tevatron dipole magnets [30], the HERA dipole magnets [31], and the dipole magnets for the Relativistic Heavy Ion Collider (RHIC) now under construction at BNL [32]. The values in Table 2 correspond to cases where all crossover contacts are assumed to be conductive. For the SSC designs, the multipole field values are quoted at a reference radius of 1 cm, while for all other designs, the reference radius is 2.5 cm. Table 2 also lists the cable parameters relevant to the computation of inter-strand eddy currents.

Table 2
Predicted field distortions and power dissipation due to inter-strand eddy currents for various superconducting dipole magnet designs

	HERA	RHIC	SSC 4-cm Apert.	SSC 5-cm Apert.	SSC/HEB	Tevatron
Transfer Function (T/kA)[a]	0.935	0.709	1.025	1.048	1.028	0.953
Inner Cable						
Width (mm)	10	9.73	9.296	12.34	12.34	7.8
Twist Pitch (mm)	95	74	79	86	86	57
Number of Strands	24	30	23	30	30	23
Outer Cable						
Width (mm)	10	n/a	9.728	11.68	11.68	7.8
Twist Pitch (mm)	95	n/a	74	94	94	57
Number of Strands	24	n/a	30	36	36	23
Field Distortions[b]						
$B^0_{0,\text{inter}}$ (Gauss.μΩ/(A/s))	+1.5	+0.5	+2.6	+6.5	+6.1	+0.6
$B^0_{2,\text{inter}}$ (Gauss.μΩ/(A/s))	+0.4	+0.2	+0.04	+0.4	+0.4	+0.2
$B^0_{4,\text{inter}}$ (Gauss.μΩ/(A/s))	−0.05	−0.02	−0.05	−0.02	−0.01	−0.01
Power Dissipation						
P^0_{inter} (10^{-3}W.μΩ/m/(A/s)^2)	0.9	0.3	0.7	2.4	2.2	0.4

[a] On magnet axis.
[b] At the reference radius (1 cm for SSC designs, 2.5 cm for all other designs).

In the case of the HERA dipole magnets, the cable strands were coated with a thin layer of 5 wt% silver-95 wt% tin solder called *Stabrite*. The purpose of this coating was to prevent the uncontrolled formation of a copper oxide layer on the strand surfaces, and to make the crossover resistances as uniform as possible along the cable and throughout the magnet coils. Measurements on short samples of HERA cables have shown: r_c = 2.1 ± 0.5 μΩ [33]. Introducing this value into Eq. (76) yields

$$\frac{B_{2,\text{inter}}(R_{\text{ref}} = 2.5\text{ cm})}{dI_t \,/\, dt} \approx 0.2 \text{ Gauss} / (\text{A} / \text{s}) . \tag{78}$$

This predicted ramp rate dependence of the normal sextupole field appears to be in good agreement with the results of magnetic measurements as a function of ramp rate recently carried out at DESY [34]. Such an agreement gives us some confidence that the model we have developed has a sound basis.

In the case of the Tevatron dipole magnets, the coils were wound with so-called *Zebra* cables. In a zebra cable, half of the strands are coated with Stabrite, while the other half are coated with *Ebanol*. Ebanol is a chemical which favors the development of black copper oxide on the strand surfaces. The Stabrite- and Ebanol-coated strands are alternated, yielding a pattern of black and silver stripes resembling a zebra. The purpose of this mixed coating was to reduce the number of crossover contacts which are conductive while still allowing some possibility of current redistribution among the cable strands. If we assume that the Ebanol coating results in a perfect isolation, only the Stabrite/Stabrite crossovers let eddy currents flow, reducing the number of electrical paths to one fourth. Hence, compared to a case where all crossover contacts are assumed to be conductive, the effects of inter-strand eddy currents for a zebra cable are expected to be four times smaller.

In the case of the SSC dipole magnets, the values quoted in Table 2 show that the 5-cm aperture designs are much more sensitive to inter-strand eddy currents than the 4-cm aperture design. This increase in sensitivity can be understood when considering that the 5-cm aperture designs rely on cables that are wider and have a larger number of strands. For instance, Eq. (63) shows that the power dissipation per cable unit length is roughly proportional to the square of the number of strands and to the square of the cable width. Hence, when going from a 23-strand, 9.3-mm wide cable to a 36-strand, 12.3-mm wide cable, the power dissipation in the inner coils can be expected to increase by a factor of 3. This ratio is consistent with the increase in power dissipation per coil unit length seen in Table 2.

As for the intra-strand eddy currents, the energy, E_{inter}, dissipated per coil unit length during a current cycle can be estimated as

$$E_{\mathrm{inter}} = \int_{\mathrm{cycle}} P_{\mathrm{inter}} \mathrm{d}t = \frac{E^0_{\mathrm{inter}}(\mathrm{cycle})}{r_c} \frac{\mathrm{d}I_t}{\mathrm{d}t} , \quad (79)$$

where

$$E^0_{\mathrm{inter}}(\mathrm{cycle}) = \int_{\mathrm{cycle}} P^0_{\mathrm{inter}} |\mathrm{d}I_t| = \frac{P^0_{\mathrm{inter}}}{r_c} \int_{\mathrm{cycle}} |\mathrm{d}I_t| , \quad (80)$$

For the 5-cm aperture SSC design and a 500–5000–500 A cycle, we get

$$E^0_{\mathrm{inter}} = 9000 P^0_{\mathrm{inter}} = 21.6 \ \mathrm{J.\mu\Omega / m / (A / s)} . \quad (81)$$

To complete this discussion of the model with uniform crossover resistance, Figs. 10(a) and 10(b) present three-dimensional plots of $A^{l,k}_{\mathrm{inter}}$ and $B^{l,k}_{2,\mathrm{inter}}$ (z-axis) as a function of turn position in the x-y plane. The computation was done for the 5-cm aperture SSC dipole magnet cross-section, with $r_c = 1$ μΩ and $\mathrm{d}I_t/\mathrm{d}t = 1$ A/s. It can be seen clearly that the amplitude and sign of the contributions to the multipole fields of the inter-strand eddy currents generated in a given coil turn strongly depend on the turn position. In the case of the skew quadrupole field, the upper coil turns all yield a negative contribution, while the lower coil turns all yield a positive contribution. If the crossover resistance is uniform, these contributions cancel out, resulting in a zero $A_{1,\mathrm{inter}}$. For the normal sextupole, the pattern is more complicated, with a change of sign in each coil quadrant, when going from the turns next to the pole to the turns close to the midplane. However, Fig. 10(b) shows also that the contributions of the turns close to the midplane are larger than the contributions of the turns next to the pole, thus resulting in a net, positive $B_{2,\mathrm{inter}}$. Furthermore, unlike the formulae for the intra-strand eddy currents, the signs of $B^0_{2m,\mathrm{inter}}$ and $B^0_{2m,\mathrm{inter}}$ are the same as the signs of the slopes of multipole fields versus ramp rates, S_{B_2} and S_{B_4}, determined from magnetic measurements performed on SSC dipole magnet prototypes and reported in Table IV of Reference [15]. Note also that effects of the same order of magnitude as S_{B_2} and S_{B_4} can be reproduced by considering crossover resistances of the order of, or less than, 10 μΩ. It thus appears that the inter-strand eddy current model we have developed may enable us to simulate behaviors similar to the anomalous behaviors observed on some of the SSC dipole magnet prototypes.

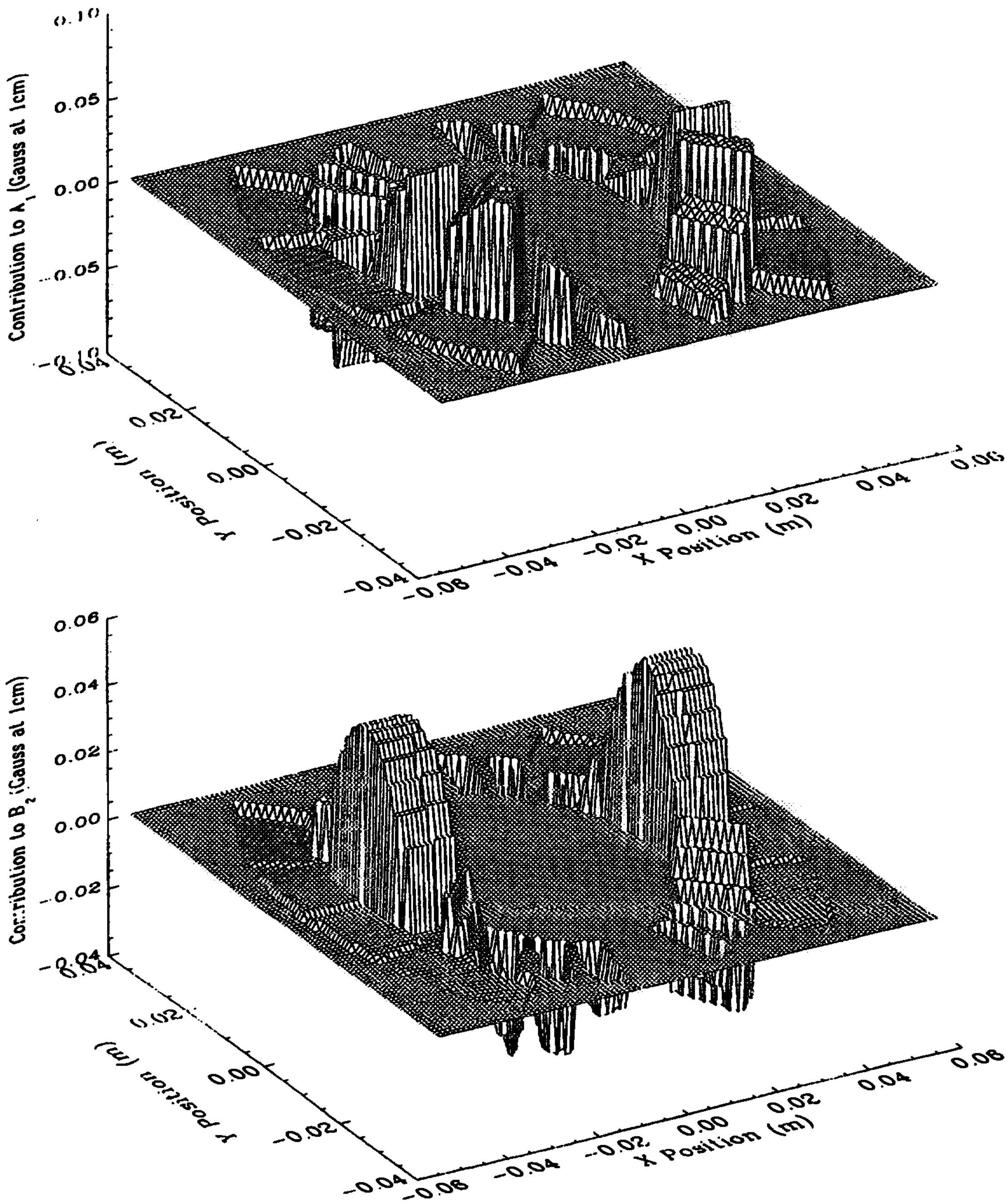


Fig. 10 Contributions to the multipole fields of the inter-strand eddy currents generated in a given coil turn of the 5-cm aperture SSC dipole magnet cross-section as a function of turn position: Top) skew quadruple field, Bottom) normal sextupole field. The multipole fields are in Gauss. The crossover resistance is assumed to be uniform, equal to 1 $\mu\Omega$, and the ramp rate is taken equal to 1 A/s.

6. CONCLUSION

An eddy-current model for superconducting particle accelerator magnets using flat, two-layer, Rutherford-type cables was developed. The model considers intra-strand and inter-strand eddy currents as the major sources of eddy currents. The calculation shows that, for low values of contact resistances at the strand crossovers between the two cable layers, the inter-strand eddy currents can produce large energy losses and field distortions while ramping.

This suggests that inter-strand eddy currents may be the cause of the type-A behavior observed on some of the SSC dipole magnet prototypes. This also indicates that controlling the inter-strand resistance is essential to the good performance of superconducting particle accelerator magnets with fast ramp cycles. Detailed interpretations of the SSC dipole magnet prototypes test data in the light of the eddy current model developed in this paper are presented elsewhere [15].

ACKNOWLEDGEMENTS

The authors wish to acknowledge the contributions of A.A. Akhmetov, A.E. Badea, T. Bush, D.W. Capone, II, M. Coles, R. Coombes, J. DiMarco, G. Ganetis, M. Garber, C. Goodzeit, A.K. Ghosh, R. Gupta, A. Jain, K. Kim, V.T. Kovachev, J. Krzywinski, J. Kuzminski, P.O Mazur, R.G. Mints, G.H Morgan, J. Muratore, W. Nah, D. Orris, J. Ozelis, E.G. Pewitt, P. Radusewicz, W.B. Sampson, R. Scanlan, R.I. Schermer, G. Snitchler, R. Stiening, J.B. Strait, R. Thomas, J.C. Tompkins, M. Wake, P. Wanderer, E. Willen, Y. Yu, Y. Zhao, and H. Zheng. The authors are grateful to Y. Kimura, T. Shintomi, and K. Tsuchiya for their support in the completion of this work. They also wish to thank C. Swenson for many fruitful discussions.

REFERENCES

[1] J.R. Sanford and D.M. Matthews, eds., "Site Specific Conceptual Design of the Superconducting Super Collider," SSCL-SR-1056, (1990).

[2] A. Devred, T. Bush, *et al.*, "Review of SSC Dipole Magnet Mechanics and Quench Performance," *Supercollider 4*, pp. 113–135 (1992).

[3] J. Strait, D. Orris, *et al.*, "Quench Performance of Fermilab/General Dynamics Built Full Length SSC Collider Dipole Magnets," *Supercollider 4*, pp. 365–372 (1992).

[4] M. Anerella, J. Cottingham, *et al.*, "Construction and Test Results from 15-m Long, 50-mm Aperture SSC Collider Dipole Models," *Supercollider 4*, pp. 535–549 (1992).

[5] W. Nah, A. Akhmetov, *et al.*, "Quench Characteristics of 5-cm Aperture, 15-m Long SSC Dipole Magnet Prototypes," *IEEE Trans. Applied Superconductivity*, **3**(1), pp. 658–661 (1993).

[6] J. Kuzminski, T. Bush, *at al.*, "Quench Performance of 50-mm Aperture, 15-m Long SSC Dipole Magnets Built at Fermilab," *Int. J. Mod. Phys. A* (proc. Suppl.) 2B, pp. 588–591 (1993).

[7] P. Wanderer, M. Anerella, *et al.*, "A Summary of SSC Dipole Magnet Field Quality Measurements," *Supercollider 4*, pp. 137–149 (1992).

[8] S. Delchamps, M. Bleadon, *et al.*, "Magnetic Field Measurements of Fermilab/General Dynamics Built Full Scale SSC Collider Dipole Magnets," *Supercollider 4*, pp. 251–258 (1992).

[9] P. Wanderer, M. Anerella, *et al.*, "Magnetic Design and Field Quality Measurements of Full Length 50-mm Aperture SSC Model Dipoles Built at BNL," *Int. J. Mod. Phys. A* (proc. Suppl.) 2B, pp. 641–643 (1993).

[10] J. Strait, R. Bossert, *et al.*, "Magnetic Field Measurements of Full Length 50-mm Aperture SSC Dipole Magnets at Fermilab," *Int. J. Mod. Phys. A* (proc. Suppl.) 2B, pp. 656–658 (1993).

[11] T. Ogitsu and A. Devred, "Influence of Azimuthal Coil Size Variations on Magnetic Field Harmonics of Superconducting Particle Accelerator Magnets," *Rev. Sci. Instrum.*, **65**(6), pp. 1998–2005 (1994).

[12] T. Ogitsu, Y. Zhao, *et al.*, "Influence of Cable Eddy Currents on Magnetic Field Harmonics," *KEK-Proc.* 92-14, pp. 23–27 (1992).

[13] Y. Zhao, A. Akhmetov, *et al.*, "Current Dependence of Harmonic Field Coefficients of 5-cm Aperture, 15-m Long SSC Dipole Magnet Prototypes," *IEEE Trans. Applied Superconductivity*, **3**(1), pp. 674–677 (1993).

[14] A. Devred and T. Ogitsu, "Ramp-Rate Sensitivity of SSC Dipole magnet Prototypes," *Journal of Japanese Cryogenic Engineering Society*, Vol. 29, No. 9, pp. 424–452 (1994).

[15] A. Devred and T. Ogitsu, "Ramp Rate Sensitivity of SSC Magnets," to appear in the Proceedings of the Joint US-CERN–Japan Topical Course on Accelerator Physics and Technology, Maui, Hawaii, USA (Nov. 3–9, 1994).

[16] N.M. Wilson, *Superconducting Magnets*, Clarendon Press, Oxford, p. 174ff (1983).

[17] A. Devred, T. Bush, *et al.*, "About the Mechanics of SSC Dipole Magnet Prototypes," *AIP Conference Proceedings*, **249**(2), pp. 1310–1374 (1992).

[18] R.I. Schermer, private communication.

[19] P. Schmüser, "Superconducting Magnets for Particle Accelerators," *AIP Conference Proceedings*, **249**(2), pp. 1099–1158 (1992).

[20] G.B.J. Mulder and E.M.J. Niessen, "Coupling Losses of Multifilamentary Superconductors Having Several Concentric Regions and Mixed Matrix," *IEEE Trans. Applied Superconductivity*, **3**(1), pp. 142–145 (1993).

[21] W.J. Carr, Jr., "Conductivity, Permeability, and Dielectric Constant in a Multifilament Superconductor," *J. Appl. Phys.*, **46**, pp. 4043–4047 (1975).

[22] W.B. Sampson, M. Garber, and A.K. Ghosh, "Normal State Resistance and Low Temperature Resistance of Superconducting Cables for Accelerator Magnets," *IEEE Trans. Magn.*, **25**(2), pp. 2097–2100 (1989).

[23] C. Swenson, private communication.

[24] Y.Z. Lei, T. Shintomi, *et al.*, "AC Loss Measurements of Rutherford Type Superconducting Cables under Mechanical Stresses," KEK-Proc. 92-14, pp. 23–27 (1992).

[25] G.H. Morgan, "Eddy Currents in Flat Metal Field Superconducting Braids," *J. Appl. Phys.*, **44**, pp. 3319–3322 (1976).

[26] V.E. Sytnikov and G.G. Svalov, "Coupling losses in superconducting transposed conductors located in changing magnetic field," *Cryogenics*, **29**, p. 926–930 (1989).

[27] A.A. Akhmetov, A. Devred, and T. Ogitsu, "Periodicity of Cross-Over Currents in a Rutherford-Type Cable Subjected to a Time-Dependent Magnetic Field," *J. Appl. Phys.*, **75**(6), pp. 3176–3183 (1994).

[28] G.H. Morgan, private communication.

[29] D. Orrell, private communication.

[30] F.T. Cole, M.R. Donaldson, *et al.*, eds., "A Report on the Design of the Fermi National Accelerator Laboratory Superconducting Accelerator," FNAL (1979).

[31] H. Kaiser, "Design of Superconducting Dipole for HERA," *Proceedings of the XIII-th International Conference on High Energy Physics Accelerators*, Novosibirsk, USSR, August 7–11, pp. 49–52 (1986).

[32] P.A. Thompson, R.C. Gupta, S.A. Kahn, H. Hahn, G.H. Morgan, P.J. Wanderer, and E. Willen, "Revised Cross Section for RHIC Arc Dipole Magnets," *Proc. of 1991 IEEE Part. Acc. Conf.*, pp. 2245–2247 (1991).

[33] A.K. Ghosh, private communication.

[34] H. Brück, private communication.

IMPACT OF PERSISTENT CURRENTS ON ACCELERATOR PERFORMANCE

B.J. Holzer
DESY, Hamburg, Germany

Abstract
Persistent currents are superconducting eddy currents circulating inside the filaments of a s.c. magnet cable. They are induced by external field changes and due to their contribution to the multipole components of the magnets they have a strong influence on the accelerator. In this paper the impact of persistent currents (p.c.) on the performance of a large superconducting accelerator is discussed. Based on the experience at the proton ring of the HERA collider, the impact on magnet quality and beam parameters, as well as on the practical handling of the machine is studied. Detailed measurements concerning stability and reproducibility of the machine are presented at static operation and during beam acceleration. The correction system which is used to keep the impact of p.c. within the required tolerances is presented.

1 INTRODUCTION

Superconducting magnets have played an essential role in the design and construction of the big hadron accelerators such as the TEVATRON p-p̄ collider at Fermilab and the proton-lepton collider HERA at DESY. Even more important is the impact of this magnet technology for the new machines such as RHIC and LHC that are in their design and construction phase at present.

The state of the art in the fields of magnet design using superconductors, quench safety and the physics of superconductivity itself has been presented in great detail in the contributions to this school. And we had the opportunity to look closer to the wonderful possibilities that are offered by nature if we take advantage of superconductivity in a particle accelerator.

The keyword here will be *"persistent currents"*. Unfortunately superconducting magnets are — as conventional magnets too — not perfect. They suffer from imperfections and non-reproducibilities and it will be our task now to have a closer look at the influence of superconducting eddy currents on the performance of the machine and on the beam parameters. ... and maybe we'll find that the wonderful world of superconductivity in particle accelerators is not as perfect as it could be.

2 PERSISTENT CURRENTS

Persistent currents are long-lasting eddy currents induced by external changes of the magnetic field and that circulate inside the superconducting filaments of a magnet cable. Figure 1 shows a so-called Rutherford-type cable as used in the HERA main dipole magnets [1].

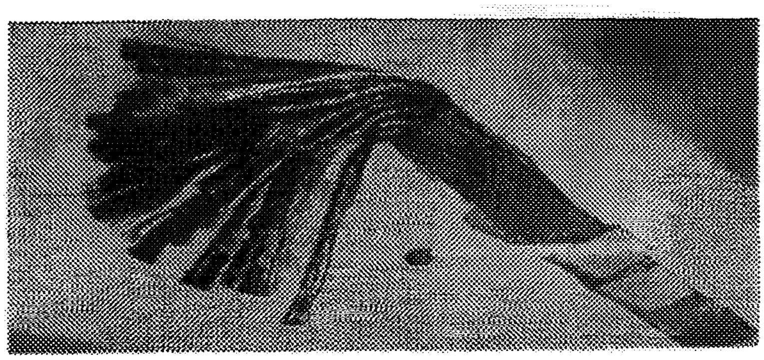


Figure 1: A superconducting cable as used in HERA consists of 24 wires each containing 1230 filaments. At the end of the cable the copper matrice surrounding the niobium-titanium filaments is removed to show the filaments.

It consists of 24 wires each containing 1230 superconducting filaments. The typical diameter of such a niobium-titanium alloy fiber is about 14 μm.

A. Devred's chapter gives a theoretical introduction to the problem of persistent currents and models their effects. Assuming that the induced persistent magnetization currents circulate inside the filaments at a constant current density, the model can be applied to predict the multipole components of superconducting magnets with great accuracy. Figure 2 shows a schematic view of the persistent currents inside a filament according to the Bean-model [2].

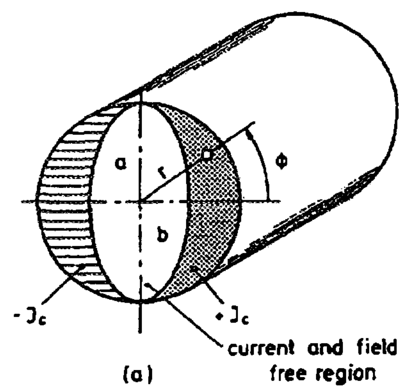


Figure 2: Model of the persistent currents inside a filament. Induced by a change of the external magnetic field, currents are running back and forth in the filament at a critical current density J_c .

Based mainly on the experience and measurements at HERA we will point out the impact of persistent currents on the performance of an accelerator and the possibilities to counteract their effects.

2.1 Historical Remark

The very first observation of the influence of persistent currents was made at Fermilab in 1987 when the new TEVATRON machine came into operation. During the initial coasting beam tests large changes in the beam behavior over the course of several hours were observed. Variations in initial tunes and chromaticity occurred which depended on the recent history of the machine [3] and the corresponding value of the sextupole component [4] is plotted in Fig. 3. The values correspond to 150 GeV beam energy, the nominal injection energy of the accelerator. A decrease of up to a factor of two in a time period of 6 hours is observed, indicating that persistent currents may lead to machine parameters varying with time. As will be shown later a similar effect occurs in HERA at low energies.

3 MULTIPOLE CONTRIBUTIONS OF PERSISTENT CURRENTS

The persistent currents inside the filaments of a superconductor severely affect — at least at low magnetic fields — the quality of the magnet: They contribute to the multipole components of the magnetic field.

With r denoting the radial, and θ the azimuthal direction, the azimuthal field component B_θ of a magnet can be expanded in a series of normal and skew components b_n and a_n:

$$B_\theta(r,\theta) = B_{\mathrm{main}} * \Sigma(\frac{r}{r_0})^{n-1} * \{b_n * \cos(n\theta) + a_n * \sin(n\theta)\} \quad (1)$$

where r_0 denotes the reference radius (25 mm for the HERA magnets). It corresponds to about $\frac{2}{3}$ of the inner coil radius and approximates the free-bore radius of the beam-pipe. B_{main} is the "main field" of the magnet, i.e. the dipole field in a main bending magnet and in the case of a quadrupole lens $B_{\mathrm{main}} = r_0 * \mathrm{g}$ with g being the gradient.

In figure 4 such a multipole composition is presented for the main bending and focusing magnets of the HERA proton ring [5] with the multipole order n up to 16. The measurements were obtained using a

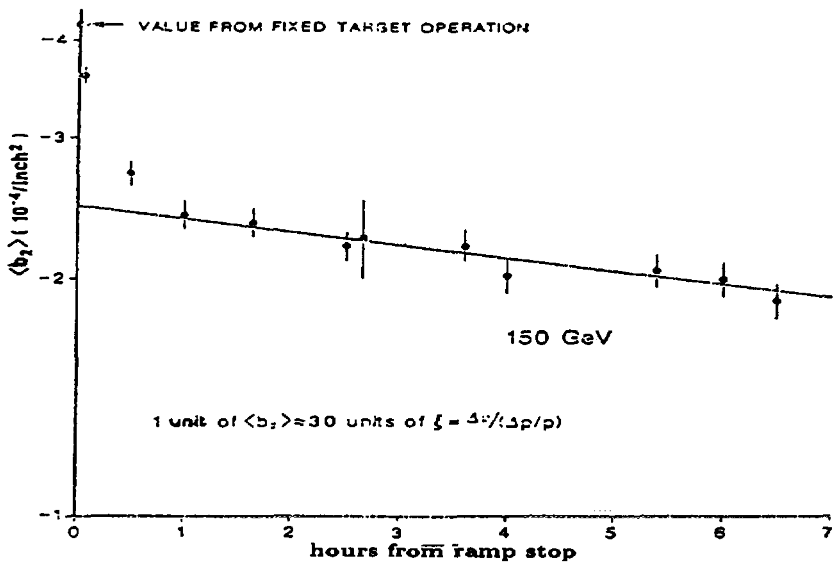


Figure 3: Variation with time of the sextupole component of the TEVATRON magnets at the injection energy of 150 GeV. This plot shows the first evidence of persistent current effects in an accelerator.

system of rotating coils [6] at a transport current through the magnets of $I_0 = 5000$ A which corresponds to an energy of about 800 GeV and a magnetic induction of roughly 4.7 Tesla. On the left hand side of the figure the multipole components of the dipole magnet are presented. The right hand side shows those of the main quadrupole lenses. The average values are plotted with error bars indicating the rms spread of the magnet sample. The data are split with respect to the two production lines (i.e. German/Italian in the case of the bending magnets and German/French for the quadrupole lenses). The corresponding skew components of the two HERA main magnet types are summarized in figure 5 in the same way.

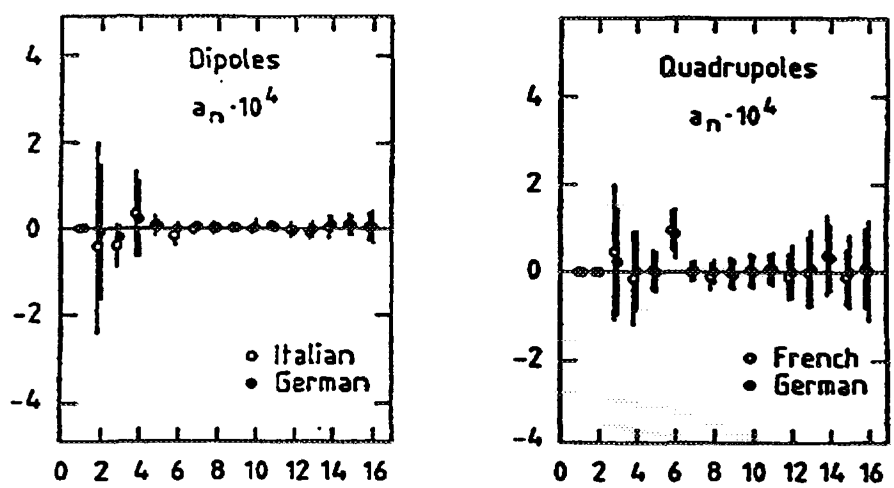


Figure 4: Normal multipole coefficients b_n up to order 16 of the HERA main dipole and quadrupole magnets measured at a magnetic field of about 5 Tesla.

While the contribution of persistent-currents to the multipole composition of the magnetic fields can be neglected completely at the high fields of about 5 Tesla where the measurements shown in figures 4 and 5 have been taken, this is generally not true for lower fields i.e. close to the injection energy of the machine). The plots of figures 4 and 5 therefore give us an information on the true multipole composition of our magnets without persistent current contributions. As can be seen from the data the multipole coefficients of the HERA proton magnets are in the order of 10^{-4} both for the skew and the normal components — and let's assume that these are good magnets.

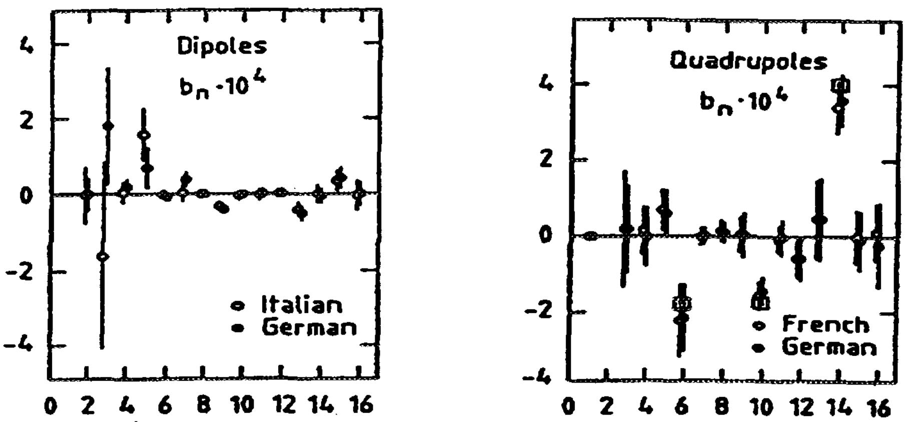


Figure 5: Skew coefficients a_n up to order 16 of the HERA main dipole and quadrupole magnets measured at a magnetic field of about 5 Tesla.

At low energies the situation looks quite different: The field quality of the superconducting magnets is seriously affected by persistent currents inside the filaments of the magnet cable. Multipoles of all orders allowed by the coil geometry can be created, i.e. in the case of a dipole magnet $n = 1, 3, 5, \ldots$ and for a quadrupole $n = 2, 6, 10, \ldots$

For transport currents between 100 and 1000 A corresponding to magnetic fields in the bending magnets of B = 0.09 T to B = 0.9 T, the persistent-current contribution to the dipole field in the bending magnet and to the quadrupole component in the main quadrupole lenses of HERA are shown in figures 6 and 7 respectively. The star in the plot refers to the injection field of B_{inj}=0.2267 Tesla at a transport current of $I = 244$ A [5].

Each dot represents a measurement of the persistent-current dipole contribution to the main dipole field (or quadrupole contribution to the main gradient respectively) for the given transport current through the coil.

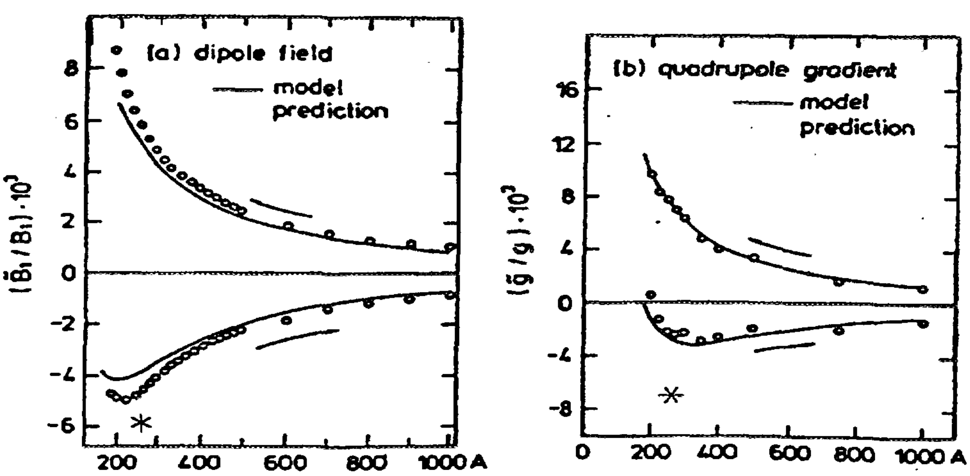


Figure 6: Contribution of the persistent-current fields to the main dipole field and to the quadrupole gradient as a function of the transport current measured at low fields in the dipole and quadrupole magnets of the HERA proton ring. The range of the plotted transport currents corresponds to beam energies of 16 to 164 GeV. The injection energy of 40 GeV is marked by a star.

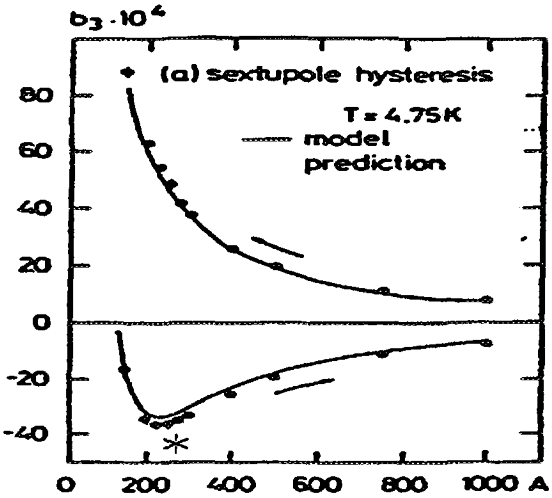


Figure 7: Persistent current contribution to the sextupole component of the main bending magnet in HERA. The value of the normal sextupole coefficient b_3 is shown as a function of the transport current (dots) and compared to model calculations (solid curve). The star refers to the value at injection energy.

Three characteristics of the graphs should be pointed out:

- The persistent current contributions are large compared to the normal multipole imperfections of the magnetic field plotted in figures 4 and 5.

  At 40 GeV injection energy of the HERA p-ring the transport current amounts to 244 A which is close to the value where the maximum deviation of the persistent current field error in figure 6 occurs.

  The magnetic field at injection in the HERA machine is about $B_0 = 0.22668$ T and it has to be reproduced to a level of $\frac{\Delta B}{B} \approx 0.1$ Gauss to guarantee a good energy matching between the preaccelerator PETRA and HERA during the transfer of the proton bunch trains. As can be deduced from the plot, the persistent-current effect surpasses this required accuracy by more than a factor of 100; resulting in a field contribution of approximately $\Delta B_{pc} \approx 11$ Gauss, far too much to be neglected.

- The persistent-current contribution in both the dipole field and the quadrupole field of figure 6 shows a strong dependence on the main field of the corresponding magnets. Depending on the change of the transport current through the coils, even the sign of the field contribution changes; for the ramp-up direction of the magnet a negative persistent-current component is detected, for the down ramp a positive one leading to a hysteresis like behavior of the magnetic field as a function of transport current and the history of the magnet. (The ramp direction during the measurement is indicated in the plots by arrows.)

  At higher fields the effect diminishes and can be neglected completely in HERA at the flat top energy of 820 GeV. (See the multipole components given at luminosity energy in figures 4 and 5.)

  At lower fields however — especially at the 40 GeV injection energy — the ambiguity of the strength and sign of the persistent current contribution has to be corrected accurately. During the measurements this has been achieved by a quench of the magnet and a well-defined procedure to apply the required value of the transport current for each measurement.

  Presently for routine operation of the accelerator, a special magnet cycle procedure is performed preceding each injection of HERA and the persistent-current contributions are corrected according to the results of online field measurements.

- In the two plots of figure 6 model calculations also have been included [7]. They are in excellent agreement with the measurements on the hysteresis, the diminishing effect at higher fields, and also the absolute strength of the p.c. contributions.

In addition to the dipole field distortion due to persistent currents, the sextupole contribution in the main bending magnets is of great importance for the machine performance. In figure 7 the normal sextupole coefficient b_3 caused by persistent currents is shown. Again, as in figure 6, it is plotted as a function of the transport current through the magnet. The arrows in the figure indicate the direction of the change of the transport current during the measurement: the "up—ramp direction" and "down—ramp direction" respectively.

As in the case of the persistent-current dipole and quadrupole components, the typical hysteresis is observed and the effect is decreasing for higher fields. The result of the model description is included in the plots of figures 6 and 7 (solid curve in the plots). The influence of this distortion on the chromaticity ξ of the machine is remarkable:

The natural chromaticities of the HERA proton ring at injection energy amount to $\xi_x = -44$ and $\xi_z = -47$ in the horizontal and vertical direction. The b_3 component due to the influence of the persistent-current fields is, according to figure 7, about $b_3 \approx 32 * 10^{-4}$ at the injection current of 244 A. The resulting distortion of the chromaticity ξ_{pc} surpasses the natural one by roughly a factor of 5:

	natural	b_3 (dipoles)
ξ_x	-44	-275
ξ_z	-47	+245

Table 1: Natural chromaticity compared to the persistent-current sextupole contribution b_3.

For completeness, the decapole contribution in the dipole and the 12-pole and 20-pole contribution in the quadrupole magnets are summarized in figures 8 and 9.

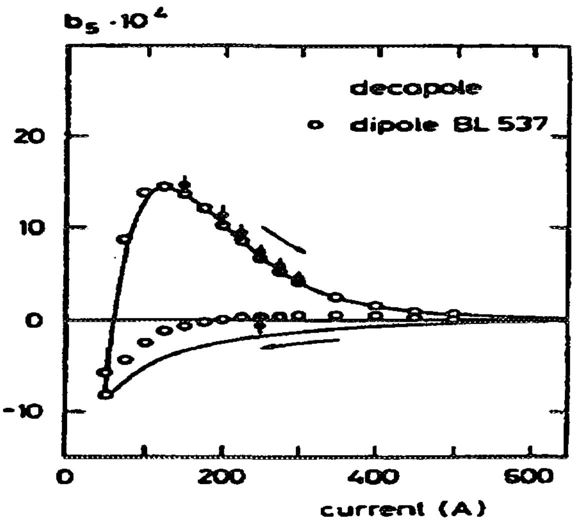


Figure 8: Persistent-current contribution to the decapole component of the main bending magnet in HERA. The results of the measurements are represented by dots in the plot. The solid curve describes the prediction of model calculations.

The persistent-current effects discussed above have a considerable influence on the parameters of the accelerator: The size of the p.c. contributions to the energy and chromaticity are large compared to the nominal values of the machine. Therefore two questions have to be discussed now:

Are the persistent-current effects stable?
Are they reproducible?

Unfortunately the answer is a twofold "No".

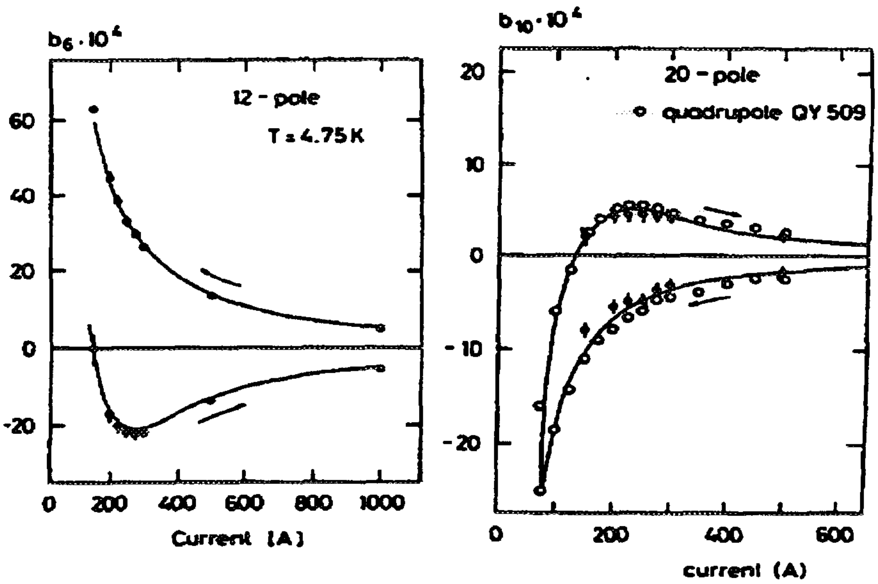


Figure 9: Persistent-current contribution to the 12-pole and 20-pole component of the main quadrupole magnets in HERA. The results of the measurements are represented by dots in the plot. The solid curve describes the prediction of model calculations.

The influence of p.c. on the machine performance depend on the history of the magnets, i.e. on the detailed excitation curve the magnets have been passed through, the effects decaying as a function of time. During the construction phase of HERA this behavior was studied in detail and we will summarize here the most important results. Preceding to the measurements each magnet was prepared by a well-defined initialization cycle consisting of

- a quench of the magnet
- application of the maximum transport current (e.g. $I_{\mathrm{max}} = 6000$ A)
- putting the magnet to the minimum current (e.g. $I_{\mathrm{min}} = 50$ A)
- adjusting the required current for the measurement (e.g. $I_{\mathrm{injection}} = 250$ A)

At the current of $I = 250$ A which is close to the actual injection current used today ($I = 244$ A) the persistent-current effects were observed as a function of time. Figure 10 shows the dipole field contribution measured in Gauss in a HERA bending magnet. The values are plotted as a function of time after the last step of the above procedure had been finished. It is evident that the p.c. contribution is not constant.

The two curves in the plot correspond to two different excitation cycles the magnet had passed through. The open circles indicate a maximum current in the preceding cycle of $I_{\mathrm{max}} = 6500$ A whereas the dots were measured after a cycle of only $I_{\mathrm{max}} = 2000$ A [8]. The horizontal axis is plotted on a logarithmic scale and both curves show a logarithmic reduction of the persistent-current effects — which can be explained by flux creeping in the super conducting wire [9]. The rate of the decay as deduced from the plot however is strongly dependent on the magnet history.

(Latest results from theoretical investigations indicate that the decay of persistent-current fields is related to at least three effects: flux creeping of the inner filament persistent currents, decay of inter-strand eddy currents floating in loops between the different strands of a super conducting cable and redistribution of currents among cable strands [10].)

As the persistent currents are induced by a change of external field (the main-dipole field in this case) they are counteracting and thus reducing the dipole field of the accelerator magnets. The decay

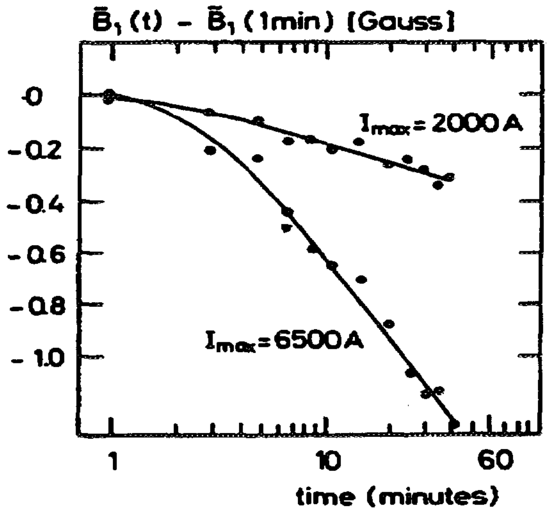


Figure 10: Dipole contribution of persistent currents at injection energy measured as a function of time. For two different values of the transport current in the preceding magnet cycle the value of the p.c. dipole field at the time of the measurement is compared to its value 1 minute after the end of the cycle. It is plotted versus a logarithmic time scale.

of the persistent currents shown in the last plot therefore predicts an increase of the measured magnetic dipole field during the injection phase of the accelerator which is indeed observed.

The same behavior is found in the sextupole contribution shown in figure 11. Again we have plotted the values as a function of a logarithmic-scaled time axis to point out the decay of the p.c. due to flux creeping. The numbers given in the plot indicate the maximum transport current of the magnet cycle preceding the measurement. As in the case of the dipole contribution, the decay rate of the persistent-current sextupole field at injection energy is stronger if higher values of the transport currents have been applied in the recent magnet history.

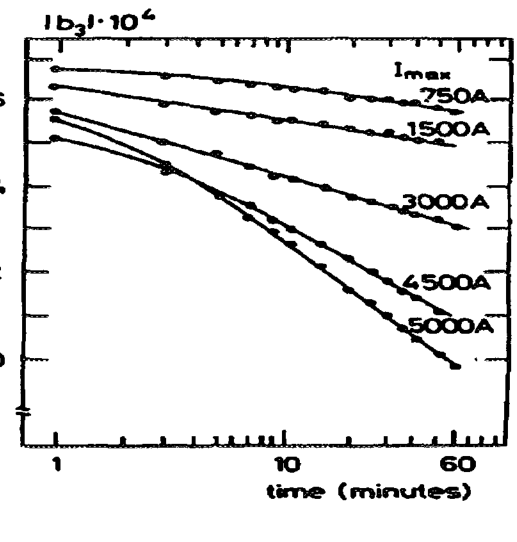


Figure 11: Sextupole contribution of persistent currents at injection energy measured as a function of time. For different values of the transport current in the preceding magnet cycle the value of the p.c. sextupole field at the time of the measurement is compared to its value 1 minute after the end of the cycle. The values of the transport current reached in the preceding cycle are indicated in the figure.

The influence of temperature changes on the magnetization currents was studied at a HERA dipole magnet in the temperature range of 4.4 K to 6.4 K. In figure 12 the sextupole contribution is shown as a function of the transport current through the magnet — and as a function of the coil temperature. At injection energy which requires a transport current of 244 A the sextupole contribution is varying by about 15 % for the temperature variation given in the plot. The influence on the accelerator performance however is relatively small in this case as the temperature around the ring can be kept constant at a level of about Δ T ≈ 0.2 K and the corresponding variation of $b_3 = 5 * 10^{-5}$ at injection energy is relatively small.

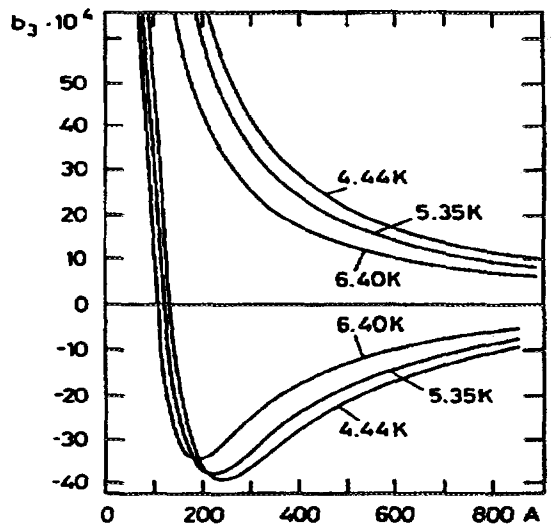


Figure 12: Sextupole contribution of persistent currents measured at different magnet currents as a function of the helium temperature.

4 PERSISTENT-CURRENT EFFECTS DURING ACCELERATION

As we have seen in the preceding section the decay of the persistent currents in the superconducting magnets of the HERA proton accelerator play an essential role at low energies. Especially during the injection period where the accelerator is kept at an energy of 40 GeV for 30 to 50 minutes the continuously decaying contributions have to be taken into account.

During acceleration however the situation is changing drastically: In the first steps of the acceleration the persistent-current fields are re-induced by the changing magnetic field of the coil. After a short time (some seconds, depending on the recent history of the magnet) the persistent-current effects have reached their full strength again and we have to deal with the original hysteresis curve. The situation is shown in figure 13. In the upper part of the figure — for clarity — we have plotted the hysteresis-like behavior of the 6-pole component already shown in figure 7. At the injection energy of 40 GeV measurements were performed during the 30 minutes injection period. The results are shown in the lower part of the plot where a better resolution was chosen: The measurement points show the decay of the p.c. 6-pole contribution at 40 GeV. The dots refer to a preceding magnet cycle where the usual maximum current of $I_{max} = 6000$ A was applied. The open circles describe the behavior after cycling the magnets a second time at a lower current of $I = 2000$ A.

In both cases the decay of the persistent-currents is seen, and also the re-approach to the hysteresis curve during acceleration. But a strong dependence is found on the history of the magnets. After the second cycle of $I_{max} = 2000$ A the decay turns out to be about half as strong and the re-induction of the p.c. due to the changing field on the ramp is much faster.

The strong contribution of persistent-current effects to the sextupole component of our magnets was already pointed out in figure 7 and table 1 for HERA at injection energy. Considering now the re-induction of the effects during the ramp of the accelerator as shown in the last figure we expect a nonlinear deviation of the chromaticity of the machine from its ideal behavior during the acceleration process.

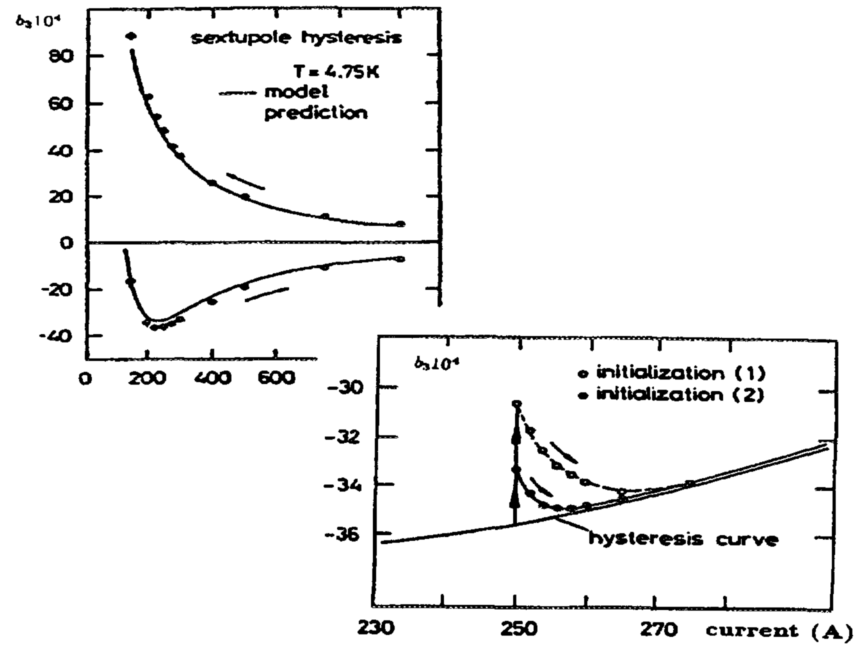


Figure 13: Persistent-current contributions during acceleration: The upper-left part of the figure shows the well-known sextupole contribution presented in figure 7. At the 244 A injection value of the transport current the decay of the p.c. b_3 component is plotted in the lower-right part of the figure where a better scale of the plot was chosen. The dots correspond to the decay after a maximum current of 2000 A in the preceding cycle. The open circles show the behavior after $I_{max} = 6500$ A. In both cases the persistent currents are re-induced to their full strength during acceleration and the measurement points approach the old hysteresis curve again.

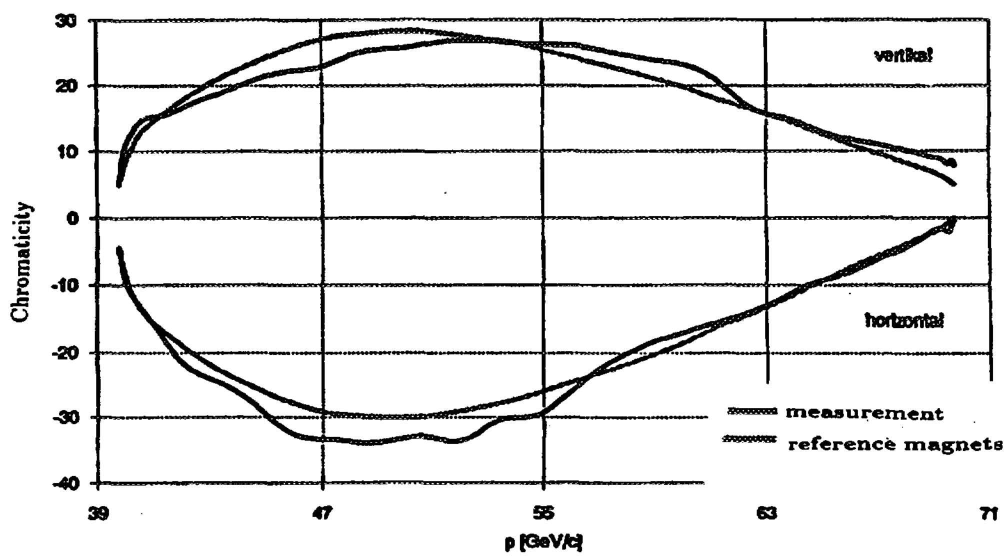


Figure 14: Chromaticity measurement during acceleration between 40 GeV and 70 GeV. The solid line shows the chromaticity measured during the first step of the HERA acceleration procedure. The dashed line corresponds to the chromaticity values predicted by the measured values of the changing sextupole component in the HERA reference magnets. Both curves are in excellent agreement prooving the reliability of the reference magnet system which is the basis for the correction system discussed in the next section.

In figure 14 the deviation of the chromaticity ξ from its nominal value (i.e. close to +1) is plotted as a function of the energy of the beam between the injection energy and 70 GeV [11]. The solid line shows a measurement of the horizontal and vertical chromaticity ξ_x and ξ_z of the beam. Directly after the start of the ramp both chromaticities are deviating strongly in opposite directions due to the nonlinearity of the persistent-current 6-pole component. The dotted line in figure 14 shows the horizontal and vertical chromaticities predicted by an online measurement of the sextupole component in the HERA reference magnets (see below).

Based on the b_3 component measured with a rotating coil in these magnets the contribution of the changing sextupole field is calculated and plotted on the same scale in figure 14. The predicted and the measured chromaticities are in excellent agreement thus forming the basis for the HERA online correction system that will be discussed in section 7.

HERA is not the only machine that suffers from persistent-current effects. Figure 15 shows the sextupole contribution in the main dipole magnets of the TEVATRON accelerator during a complete cycle of the machine. The hysteresis-like behavior, the decreasing influence at higher energies and the decay of the p.c. effects at the injection energy of 150 GeV, correspond to the plots of figures 7 and 13 for HERA [12].

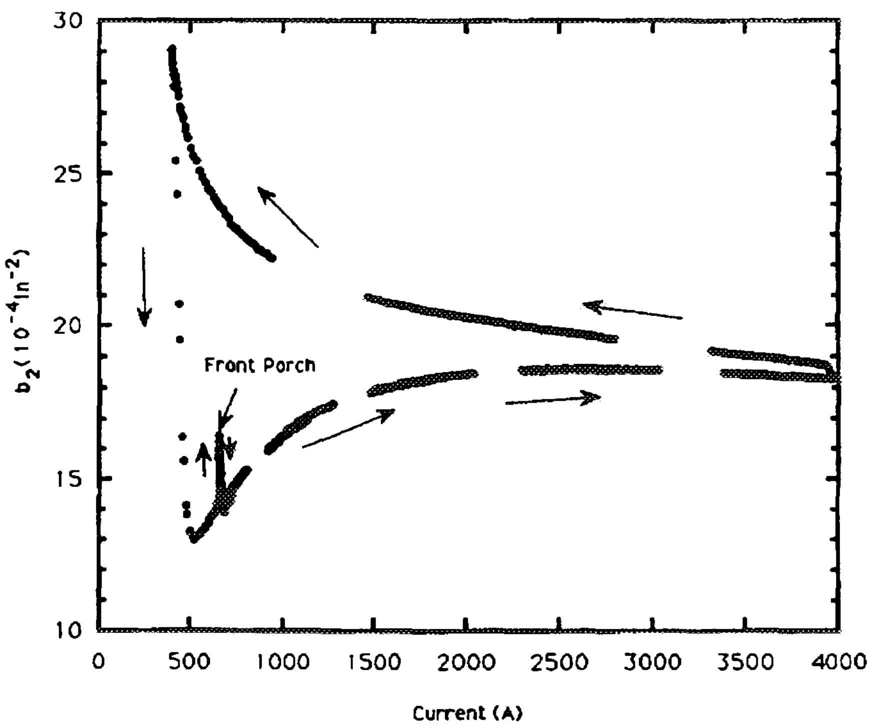


Figure 15: Sextupole contribution in the TEVATRON main magnets. As in HERA these magnets show a strong hysteresis-like behavior as a function of the transport current due to the persistent-current fields. The arrows indicate the change of the transport current and the direction of the decaying contribution during the injection period at 150 GeV.

5 THE HERA MULTIPOLE CORRECTION SCHEME

As we have seen in the last sections the influence of the persistent currents at injection and the early steps of the acceleration is too strong to be neglected. The HERA proton machine therefore is equipped with different multipole correction coils to achieve the required flexibility in the machine handling and to correct the measured field distortions of the superconducting magnets.

A concept of distributed correction coils was chosen to obtain an extended system of correction magnets [13]. Therefore the superconducting main dipole magnets were equipped with 6 m long superconducting quadrupole and sextupole correction coils which are mounted in two layers on the beam pipe.

Decapole correctors of 2.9 m length cover the remaining part in half of the dipoles and 12-pole windings are located in the main quadrupole magnets. In figure 16 the geometry of the quadrupole and sextupole correction windings located inside the dipole magnet is shown.

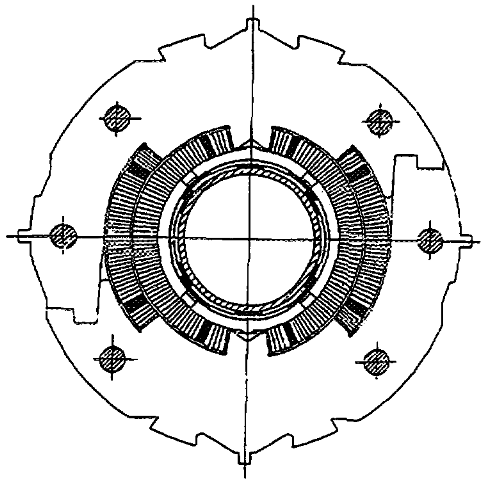


Figure 16: Cross section of a HERA dipole magnet showing the main dipole coil and the quadrupole and sextupole correction windings on the beam-pipe.

The system of extended correctors allows to correct the p.c. multipoles at the location of their origin i.e. inside the main magnets. But there is a clear disadvantage of this system: A mutual influence of the main coil and the correction windings is expected (and indeed found by measurements) as the windings of different geometries interact with each other in the mutual changing magnetic fields.

In the correction coils that are nested inside the main dipole magnets the persistent-current effects are generated in two ways [14]: Firstly they act as passive superconductors. Their filaments are magnetized if the main dipole filed is changed. Secondly, whenever one of the beam-pipe correction coils is excited, its field changes the magnetization current in the other layer and in the main coil.

A detailed analysis of these effects can be found in [14] and [15]. Here we will just summarize the most important results and discuss the impact on the handling of the machine.

There are three major effects:

1. *Change of the transfer function of the correction coil windings*

   During the measurements of the HERA dipole magnets the correction windings inside these magnets were excited to currents of ± 20 A while the main field was kept at about $B_0 = 0.23$ T which corresponds roughly to the injection field used today. The quadrupole and sextupole fields of the correctors measured under these conditions turned out to be about 10% lower than computed theoretically ... and measured in the normal conducting state of the magnets. In systematic measurements this effect on the transfer function of the correction magnets was studied. The transfer function is defined by

   $$B_i = F_i * I$$

   the index i denoting the specified coil.

   In both cases the transfer function turns out to be lower in the superconducting state than in the normal conducting state:

   $$\frac{F(4.5\mathrm{K})}{F(15\mathrm{K})} \approx 0.9$$

   There is clear evidence that this effect is caused by persistent currents in the main coil counteracting the fields of the correctors: If the field of the main magnet is cycled ($B_0 = 0.23\mathrm{T} \rightarrow 0.9\mathrm{T} \rightarrow$

0.05T → 0.23T) while keeping the transport current in the correction coils at a fixed value, the effect disappears and the transfer functions F_1 and F_2 reach their design values. The persistent-current patterns in the filaments of the main magnet are "overwritten" by the large field sweep, and the attenuation of the corrector field disappears.

2. *Passive magnetization of correction coil windings*

Even in the absence of any transport current the correction coil windings affect the multipole content of the superconducting main dipoles due to the magnetization currents that are induced by the main field. It turns out that the windings of the quadrupole correction coil generate a persistent-current sextupole field as the lowest multipole order whereas the 6-pole windings produce a decapole.

In figures 17 and 18 both multipoles are plotted as a function of the transport current in the main dipole coil. In both cases the circles in the plot correspond to the data measured in a single magnet and the error bars indicate the average of data taken in a sample of 315 dipole magnets. The solid curve shows the result from model calculations.

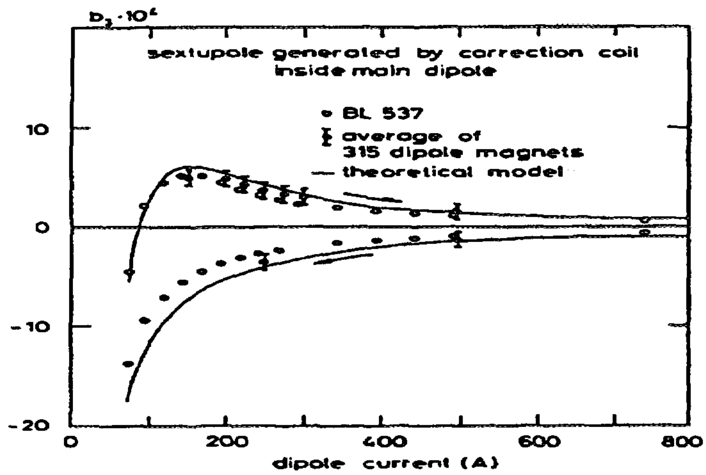


Figure 17: Effect of the passive magnetization of the quadrupole correction-coil windings in the changing external dipole field: A sextupole component is generated. The circles show measurements on a single magnet, the error bars represent the average of 315 magnets and the curve corresponds to model predictions.

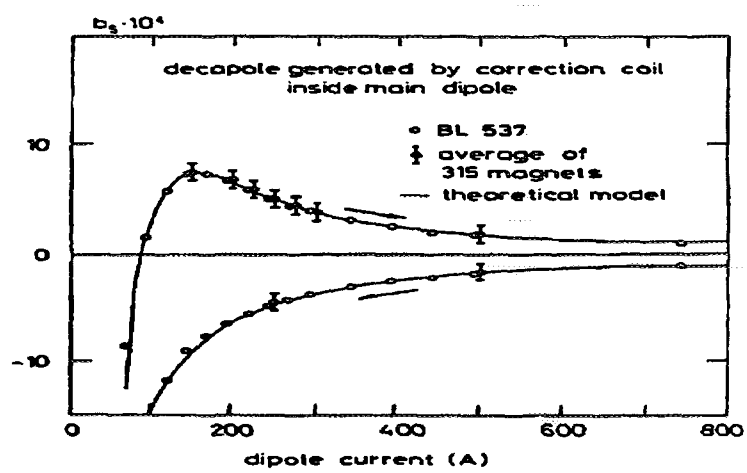


Figure 18: Effect of the passive magnetization of the sextupole correction-coil windings in the changing external dipole field: A 10-pole component is generated.

If compared to the persistent-current created multipoles that are generated in the main magnet without correction windings (see figures 7 and 8) the well-known hysteresis-like behavior shows up again. In the case of the 6-pole contribution however the effect of the correction coils is about one order of magnitude smaller; and it turns out to be opposite in sign thus counteracting the original effect and lowering its impact on the beam.

3. *Higher multipoles generated by the correction coils*

   Beside the two passive effects of the correction coils discussed above there is an active influence of the correction coil field on the multipole content of the main dipole magnet. A change of the transport current through the correction coils will induce magnetization currents in the main dipole magnet and distort the multipole components of its field.

   As usual the effect is most pronounced at low energies i.e. close to the injection energy of HERA. In figure 19 the results are summarized.

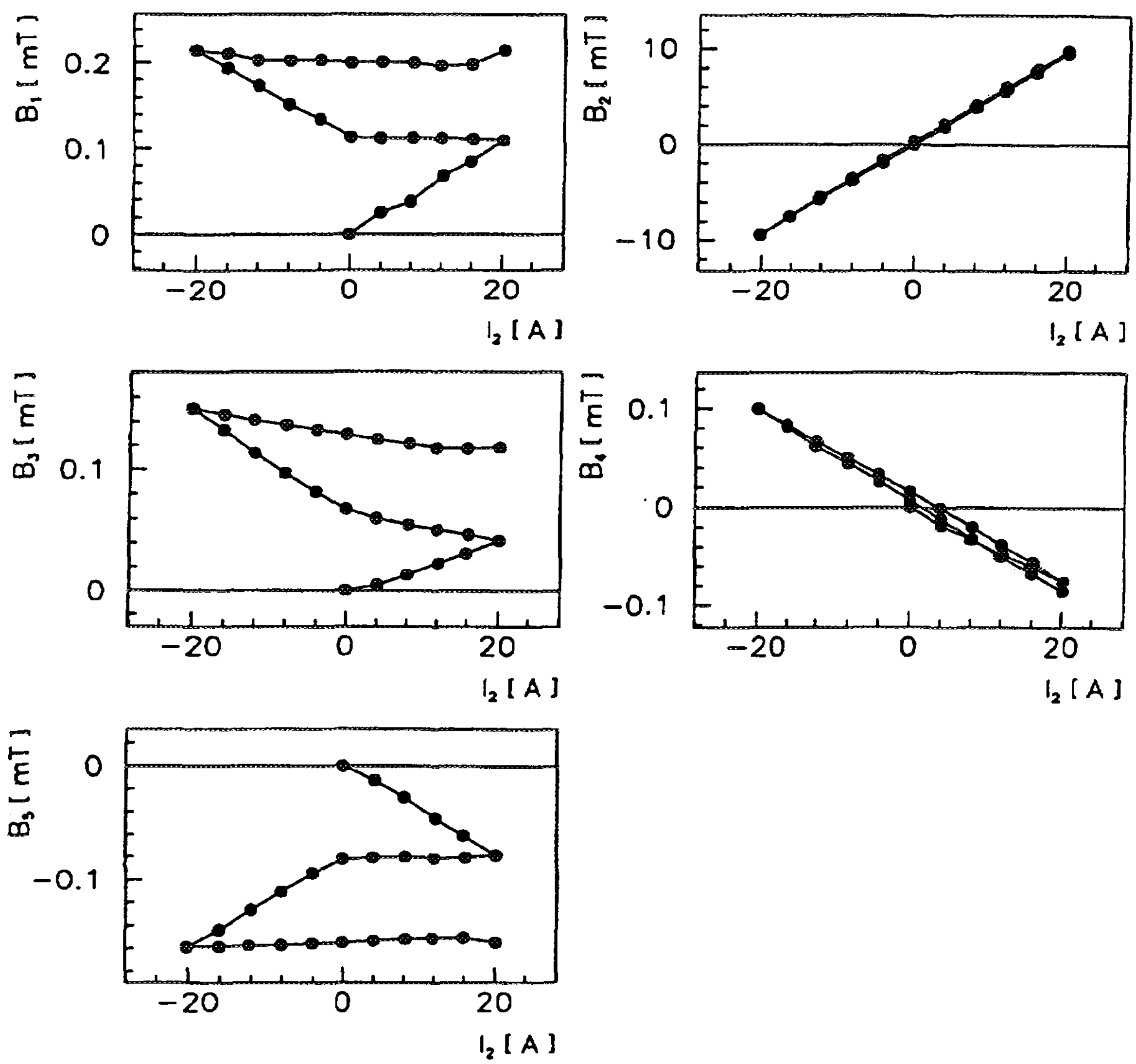


Figure 19: Multipoles generated by changing the 4-pole correction coil in the main dipole magnet at injection energy. The correction-coil current is changed between $I = \pm 20$ A leading to a zig-zag like increase of multipole fields.

At a temperature of 4.75 K the field in the dipole magnet was set to its injection value and the correction coil was powered between +20 A and -20 A repetitively. The resulting multipoles were measured. They are plotted up to multipole order 5 in the figure. If the *correction coil current* is increased from 0 A to 20 A multipoles are induced as expected, leading to the first branches in the pattern of the plots. If the current, however, is returned to zero the multipoles do not vanish. Instead, a further zig-zag like increase is observed until they reach a saturation value were dependence on the correction coil current is no longer obtained.

For completeness the same effect is shown on the down-ramp branch of the main coil in figure 20. Except for a change of the sign the situation looks qualitatively the same.

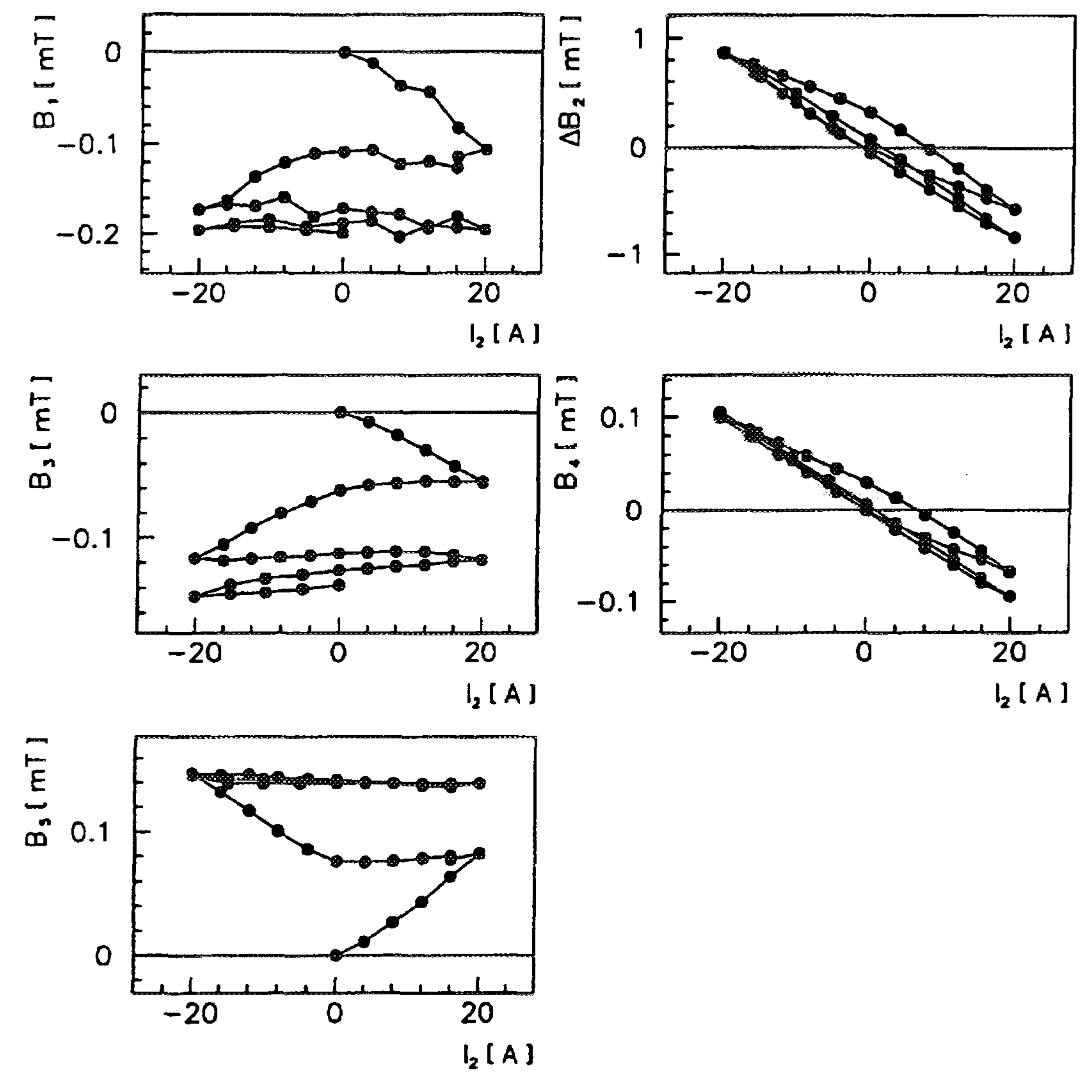


Figure 20: Multipoles generated by changing the 4-pole correction coil in the main dipole magnet at the down-ramp branch of the dipole field. The applied transport current corresponds, as in the previous figure, to the injection field.

It is a justified question now whether these exotic and surprising effects can be explained by persistent currents. Theoretical calculations have been done based on the Bean-model. For the given geometry and parameters of the HERA correction and main-coil windings the effect of the interacting coils and fields was calculated and found to be in perfect agreement with the measurement [14]. We will not go into these details here. Instead, the persistent-current nature of the problem can be proved in another way: In figure 21 the measurements of the preceding plots are repeated at a higher temperature of 25 K where the magnets are in their normal-conducting state. Again the transport current in the quadrupole correction coil was excited within the limited range of +10 A and −10 A due to the higher temperature. Except for the quadrupole field following the current change in the 4-pole correction coil no multipole excitation is observed within the measurement errors.

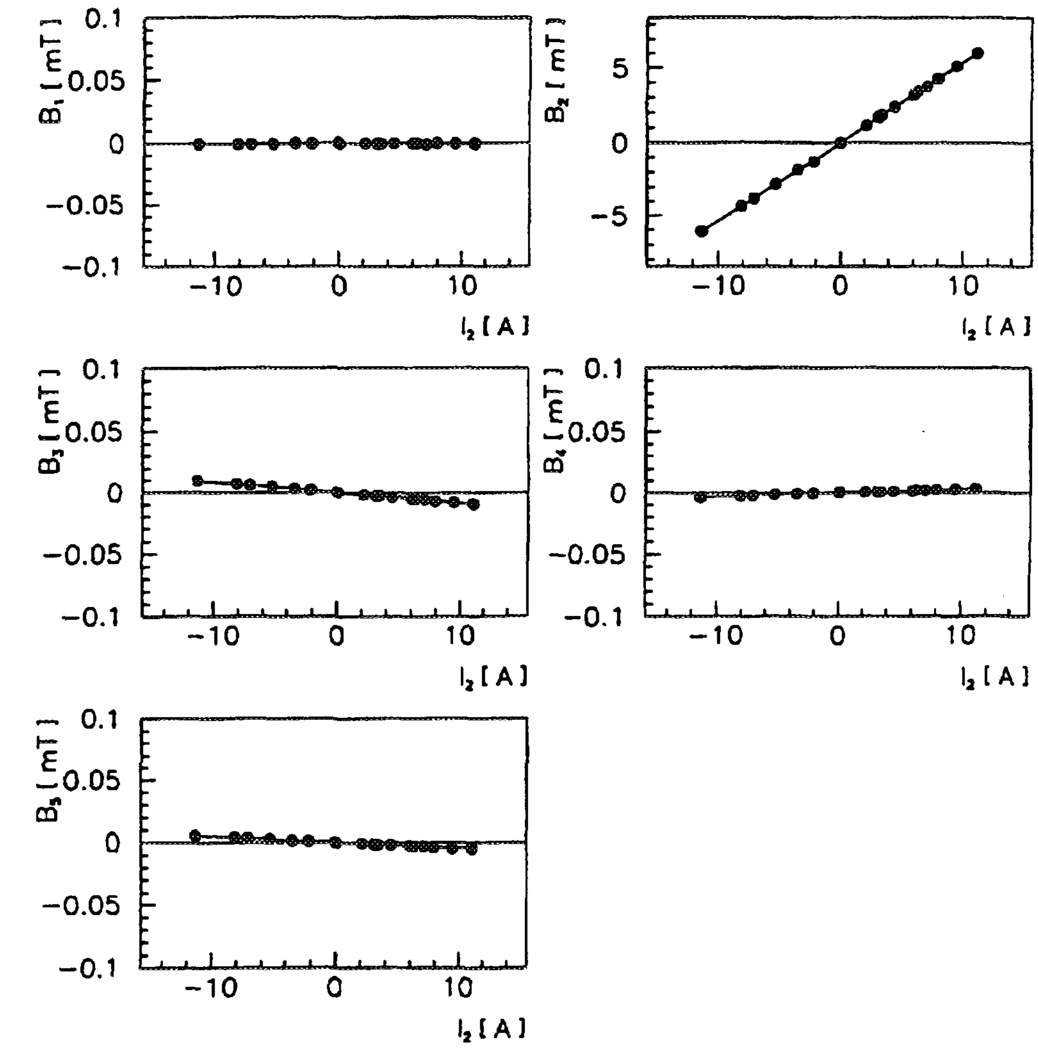


Figure 21: Measured multipole components at a temperature of 25 K. In the normal-conducting state no multipole components are induced by changing the correction coil current which indicates the persistent-current nature of the problem shown in figures 19 and 20.

6 IMPACT OF PERSISTENT CURRENTS ON ACCELERATOR PERFORMANCE

Before we study the impact of persistent currents on the performance of our accelerator it is worth summarizing briefly the effects that we have presented in the preceding sections.

- Persistent-currents affect the machine parameters most seriously at low energies, that is in general at injection and during the early steps of the acceleration process, where the influence on the beam is large. As was demonstrated in the case of the chromaticity and the energy, it can exceed the required tolerances by big factors. Therefore at least the most dominant contributions have to be corrected.

- Persistent currents are not persistent. They decay as a function of time, the rates depending on the history of the magnets. As a consequence a detailed knowledge of the recent "history" of the magnets in the accelerator is required as well as the time that has passed since the last magnet cycle was performed. Based on these data, corrections can be calculated and applied to multipole windings in the machine. An even better method is a measurement of the persistent-current effects online and to correct them accordingly.
- Persistent currents can be re-induced after their decay by changes in the external field, which is the case during acceleration for example.
- Persistent-current effects show a hysteresis-like behavior. The multipole contributions depend strongly on the way the transport current had been changed before. A dedicated procedure is required therefore to cycle the magnets and to end up with a well-known situation at injection energy after each change of the transport current.
- The extended correction coils that are nested in the HERA main magnets have to be measured under real conditions. Their transfer function is influenced by persistent currents in the main coil leading to lower fields for a given correction current.
  A mutual influence of these coils and the main magnet coil is detected which affects the multipole components of the main magnet. Changes of the transport current in the main coil generate multipoles due to persistent currents in the correction windings and vice versa.

Due to the effects summarized above a well-defined procedure has been established at HERA to prepare the machine for injection. This "magnet cycle" consists of four steps:

1. The main magnets (dipole and quadrupole magnets) are set to a field value of about 5 Tesla which is close to the maximum field under normal conditions. $B_0 = B_{\max}$
2. The normal-conducting magnets in the accelerator are cycled, that means set to their maximum fields, set back to zero current and then to their nominal values. At the same time the superconducting magnets including the correction coil windings are set to their nominal injection values.
3. The main magnets are set to their minimum values
4. The main magnets are set to their injection values.

This procedure overwrites all existing persistent-current patterns in the main magnets and the eventual existing multipole contributions due to current changes in the correction coils. The remaining effects are the well known hysteresis curve of the persistent-current multipoles in the main magnet field and their decay.

7 CORRECTION OF THE PERSISTENT-CURRENT EFFECTS

Following the procedure described above the machine is well prepared for injection and the only remaining problems we have to deal with are to correct the time dependence of the persistent-current contributions to the machine parameters such as energy, tune and chromaticity.

At HERA these parameters are measured online. The machine is equipped with two reference magnets — one for each magnet production line. They are powered in series with the main magnet chain in the accelerator tunnel and they represent the behavior of the whole magnet ensemble (see figure 22).

The reference magnets are equipped with NMR and Hall probes to measure the magnetic dipole field, and with rotating coils for the 6-pole contribution.

A change of the injection field for example due to the decay of the persistent-current fields is detected by the NMR's and corrected by a corresponding increase of the horizontal orbit correction coils in the ring. During the acceleration procedure a stationary pick-up coil is used to measure the dipole field in the reference magnets and to lock all other elements in the machine to the ramping dipole field — however complicated the hysteresis curve might look.

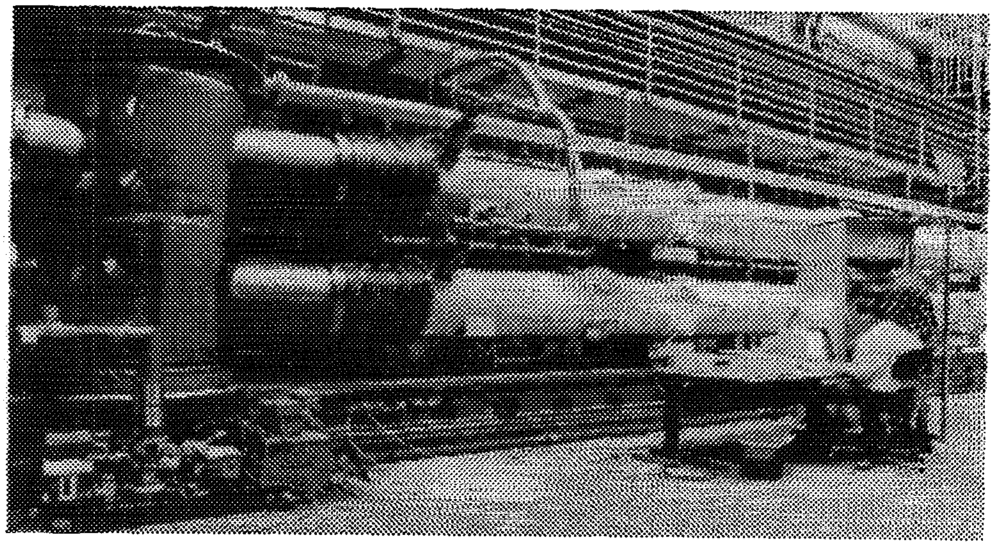


Figure 22: The HERA reference magnet system: Two dipole magnets representing the magnet ensemble in the accelerator are equipped with various measuring devices. During injection and acceleration the field measurements are used to correct the unstable persistent-current contributions and to establish accurate knowledge of the fields at any time.

The chromaticity of the machine is measured by the rotating 6-pole coils in the reference magnets. Changes of the sextupole contribution due to persistent-current decay, hysteresis or re-induction are thus detected. Correction currents are calculated and applied to the machine during injection and during the acceleration up to an energy of 150 GeV where persistent-current effects no longer play an essential role in the magnets. The power of this online measurement system is demonstrated in figure 23 where the measured chromaticity of the ring is plotted at injection energy as a function of time.

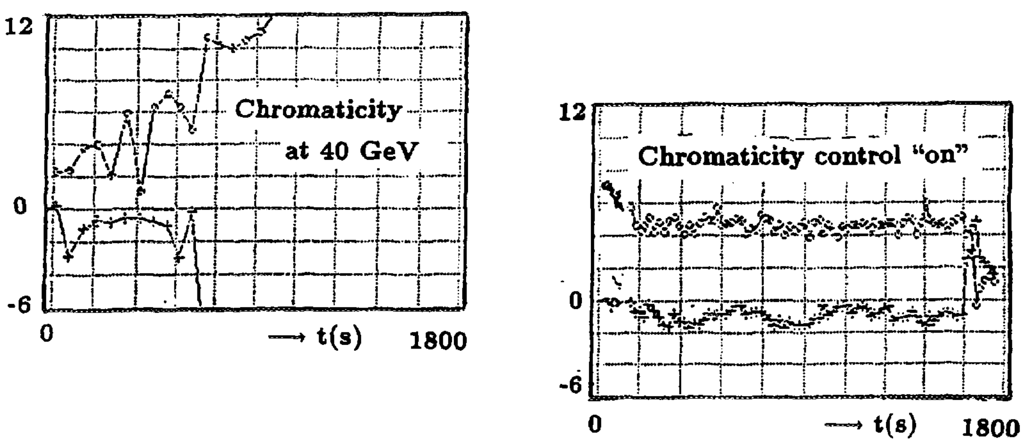


Figure 23: Chromaticity measurement at HERA during injection energy: Due to persistent-current decay the chromaticity is drifting in both planes in opposite directions as indicated in the left part of the plot. The right part shows the same situation if the changing 6-pole contribution is measured in the reference magnets. Corrections can be applied and the chromaticity is kept constant during the complete injection phase.

In the first part of the figure the decaying persistent currents lead to a fast change of the horizontal and vertical chromaticities in opposite directions. Within a couple of minutes the chromaticities have reached intolerable values. In the second part, the same situation is shown but with the correction system switched on. It counteracts the decaying persistent-current sextupole fields, and the chromaticity in both planes is kept close to the desired values.

References

[1] H.H.J. ten Kate, "Practical Superconductors", in CERN Accelerator School on Superconductivity in Particle Accelerators, Haus Rissen, Hamburg, Germany, June 1988, CERN 89-04 (1989).

[2] C.P. Bean, Phys. Rev. Lett. vol. 8, pp. 250, 1962

[3] D.A. Herrup, et al, "Time Variations of Fields in Superconducting Magnets and Their Defects on Accelerators", Fermi National Accelerator Laboratory, TM-1543, Aug. 1988

[4] D.A. Finnley et al, "Time Dependent Chromaticity Changes in the TEVATRON", IEEE Particle Accelerator Conference, Washington, USA, March, 1987, pp. 151-3

[5] P. Schmueser, "Field Quality Issues in Superconducting Magnets", IEEE Particle Accelerator Conference, San Francisco, USA, May, 1991

[6] P. Schmueser, R. Meinke, Y. Zhao, "Methods of Harmonic Measurements in the Superconducting HERA Magnets and Analysis of Systematic Errors", DESY-Hera-91-13, Sep. 1991

[7] H. Brueck et al, Z. Phys. C44, 385, 1989

[8] H. Brueck et al, "Time Dependence of Persistent Current Field Distortions in the Superconducting HERA Magnets", European Particle Accelerator Conference, Nice, France, June 1990.

[9] Y. B. Kim, C.F. Hempstead, A.R. Strnad, Phys. Rev. Letters 9, pp 306, 1963

[10] A. Devred, priv. com.

[11] O. Meincke, "Measurement and Control of Q-Values and Chromaticity in the HERA Proton Ring" (in german), DESY-Hera-92-08, Apr. 1992

[12] D.A. Herrup, "Compensation of Time Dependent Persistent Current Effects in the TEVATRON", Fermilab-FN-602, Apr. 1993

[13] C. Daum, J. Geerinck, P. Schmueser, R. Heller, H. Moeller, H. Muto, S. Schollmeier, P. Bracke, DESY-Hera-89-09

[14] M. Pekeler et al, "Coupled Persistent Current Effects in the Hera Dipoles and Beam Pipe Correction Coils", DESY-Hera-92-06, Apr. 1992

[15] H. Brueck et al,"Persistent Current Effects in the Superconducting Hera Magnets and Correction Coils", European Particle Accelerator Conference, Nice, France, June 1990.

NEXT PAGE(S) left BLANK

QUENCH PROTECTION

K.H. Mess
DESY, Hamburg, Germany

Abstract
Quench protection starts with the design of the magnets. Hence tools to estimate the stability of a magnet against temperature perturbations are discussed. Stable magnets quench reluctantly, of course. The low expansion rate may in turn have an influence on the hot-spot temperature. Methods to measure and estimate quench expansion are presented and the problems with quench detection and energy extraction are discussed using existing solutions as examples.

1. TRANSITION TO THE NORMAL CONDUCTING STATE

A strand of a superconducting cable consists of matrix material, for example, copper or aluminium, which surrounds the hard (type II) superconductor, for example, NbTi. Let the strand have the perimeter P and the cross sectional area A, and let it be completely immersed in liquid helium at the bath temperature T_B to keep it superconducting. An external magnetic field B is applied perpendicular to the wire. The wire will remain in the superconducting state if the current density in the superconductor $J = I/(1-f)A$, the external field B, and the temperature T are all below the "critical surface", where f is the matrix fraction of the wire. Figure 1 shows this "surface" that describes the phase transition. The critical surface of a piece of superconductor depends on the chemical consistency and the treatment during production. Approximate properties of NbTi-copper composite conductors, useful for calculations, can be found in the literature [1], [2], [3]. Superconductivity vanishes for any set of two parameters, if the third parameter increases beyond the critical value. Some authors [3] distinguish between transitions by simultaneous current and field increase (along arrow C, conductor limited quenches) and by local temperature fluctuations (along arrow E, energy deposited quenches). The distinction helps to identify the reason for quenches and to find possible cures. Moreover, the current density is normally constant throughout the coil or large fractions of it, while the temperature and the field may vary. In a magnet the field B correlates with the current I. Thus magnets operate on a load line, as the line C in Fig. (1) indicates for $T = T_B = 4.2$K.

If the superconducting compound wire has for some reasons a local temperature above the bath temperature, as indicated in Fig. 2, the thermal energy varies along the strand because of the heat flow. If the temperature is high enough Joule heat generation will take place. Wire motions in the magnetic field, cracking epoxy, eddy current heating and other distributed sources may add further heat. The heat conduction through the surface of the strand to the coolant will cool the wire. All energy variations will result in a local change of the thermal energy. The equation for the heat balance in a piece of conductor of length dx can be written:

$$C_{avg}(T)\frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(k_{avg}(T)\frac{\partial T}{\partial x}\right) + Q(T) + g_{dist}(t,x) - W(T). \qquad (1)$$

Heat conduction, Joule heating, additional distributed heat sources, and the energy absorbed by the coolant determine the change in temperature. C_{avg} is the volumetric specific heat capacity, averaged over the wire. The thermal conductivity depends on the direction of heat flow. In the axial direction the heat conductivities add up according to their relative area like $k_{avg}(T) = fk_m(T) + (1-f)k_s(T)$, where the subscripts m and s stand for matrix and

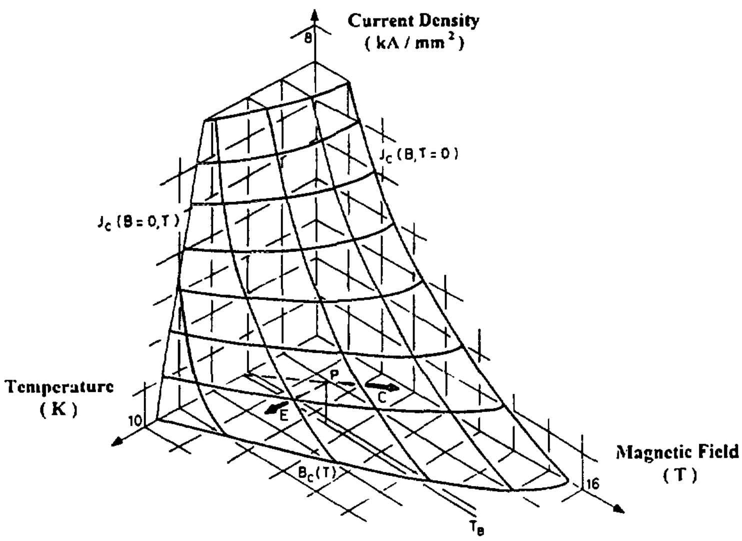


Fig. 1 The phase transition surface of NbTi

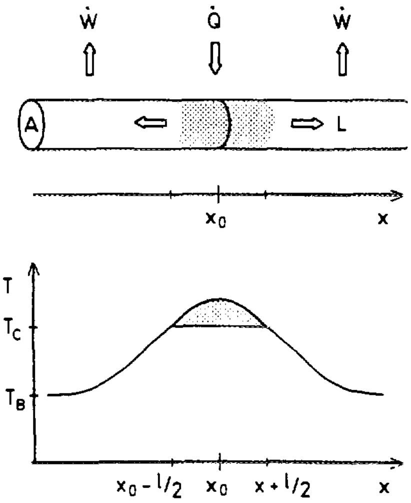


Fig. 2 Sketch of the heat balance in a quenching cable and of a possible temperature profile

superconductor respectively. In the radial direction the expression for the heat conductivity must take the distribution of filaments in the matrix and a possible cladding with solder or varnish into account, which renders the transverse heat conductivity ill defined. It is customary to hide the general lack of knowledge by simply scaling the transverse heat conductivity from the axial heat conductivity by applying some factor. Copper and aluminium are by far the better conductors and also transport the heat much better than NbTi. It is therefore a fair approximation to ignore the contribution to the heat conductivity by the superconductor altogether if material with a small resistivity ρ_m at low temperatures is selected as matrix.

The cooling term in Eq. (1) depends on the temperature difference times the heat transfer coefficient h, which varies with temperature and cooling conditions.

The Joule heating term needs detailed explanation. If the temperature is low enough the current density in the matrix $J_m = I_m / f \cdot A$ will be zero and the current will entirely flow in the superconductor with a density $J = I / ((1-f) \cdot A)$. For convenience the average current density is called $J_{avg} = I / A$.

Close to the critical surface some of the flux in the superconductor starts moving. The changing total flux through the superconductor induces an electrical field, which shows up as if there were a resistance. The resistances of the flux-flow and of the matrix act in parallel. The increase in matrix current will be such that the voltage due to the surplus current beyond the critical current in the superconductor, which feels the flux flow resistivity ρ_f, equals the voltage in the matrix:

$$V_m = J_m \rho_m = (J_s - J_c)\rho_f = (J - J_m - J_c)\rho_f \ . \qquad (2)$$

As the flux flow resistance is quite large for small surplus currents approximately all current above the critical current is transferred to the matrix. The so called current-sharing model states that the superconductor carries as much current as possible up to the critical current density J_c. Any additional current flows through the matrix.

As the critical current density depends almost linearly on the temperature for a given field, the current-sharing (or heat generation) temperature is:

$$T_g = T_B + \left(T_c(B) - T_B\right)\left(1 - \frac{J}{J_c(B,T_B)}\right). \quad (3)$$

Below the current-sharing temperature the Joule heating is zero because the superconductor carries the current. Above the critical temperature the superconductor is practically free of current and the average resistivity is $\rho_{avg} = \rho_m / f$ because the resistance of NbTi exceeds the resistance of good copper by a factor 2000.

Figure 3 illustrates the current-sharing. The thick line indicates the current density in the superconductor; the dotted line shows the current density in the matrix.

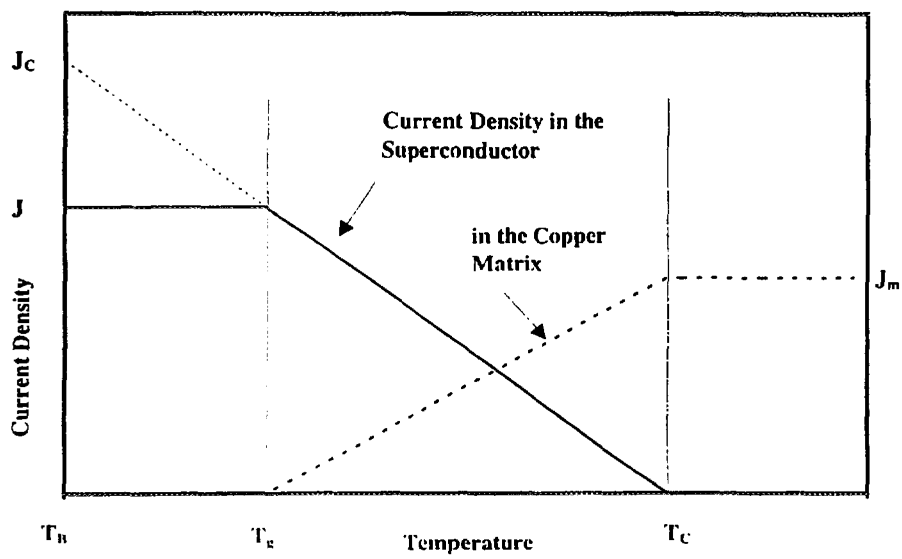


Fig. 3 Distribution of the current density between superconductor and matrix in the current sharing region

The Joule heating equals the total transport current density times the electrical field:

$$Q(T) = J_{avg} \cdot E = J_{avg} \frac{dU}{dx} = \frac{I}{A}\frac{dU}{dx}. \quad (4)$$

The differential voltage dU depends on the temperature difference over a distance dx along the wire and equals the current through the matrix times the differential resistance dR_m. Using the linear approximations for the matrix current density one calculates the Joule heating term as:

$$Q(T)=\begin{cases} 0 & \Leftarrow T \le T_g \\ \dfrac{\rho_m(T)J_{avg}^2}{f}\dfrac{T-T_g}{T_c-T_g} & \Leftarrow T_g \le T \le T_c \\ \dfrac{\rho_m(T)J_{avg}^2}{f} & \Leftarrow T \ge T_c \end{cases} \qquad (5)$$

Usually knowledge about the distributed heat sources g_{dist} is quite limited. All magnets have welds or solder joints somewhere in the cable. A good solder joint has a resistance of $10^{-9}\,\Omega$. Hence a typical current of 5000 A creates a steady heat load of 5 μW. This can already present a stability problem under some circumstances.

Accelerator magnets, fusion reactor magnets or detector magnets for high energy experiments will always be exposed to some level of radiation. A bunch consisting of 10^{11} protons of 1 TeV deposits an enormous energy density of 10 J/cm^3 in a typical copper-stabilised NbTi coil.

Other heat sources are related to the Lorentz forces in the magnet. They deform the coil. This may result in movements or bending of conductors and hence in friction [4]. For magnets under SSC design conditions [3] a 10 μm movement over a conductor length of 500 μm is sufficient to release 10 μJ.

2. STABILITY

Quench protection means first of all protection against quenches. This is best achieved by constructing a magnet that hardly quenches under normal operating conditions. What can reasonably be expected and which are the sensitive parameters?

Little can be done against occasional temperature disturbances in the conductor. The cooling and the heat capacity of the cable determine the stability against such disturbances. Its extension in space and time and the shape of the temperature distribution will determine whether the zone of disturbance grows or decreases. The minimal zone that cannot be cooled away is often called the "minimum propagating zone" or MPZ [5].

Equation (1) describes the heat balance along the conductor. The most interesting questions are: What are the conditions that cause the expansion of a normal zone, and what is the expansion rate? What are the conditions for a normal zone that does not expand? If a current-sharing or normal conducting zone is very short, the heat can be conducted away. Alternatively if the heat produced is small enough, it can be cooled away on the helium wetted surface. Thus there are two extreme cases. In the first case heat sources can be assumed to be uniformly distributed along the wire. That is equivalent to the absence of longitudinal heat conduction. All cooling takes place locally through the surface to the helium. In the second case, a very short hot spot expands by heat conduction along the wire in both directions. The heat flow into the helium is neglected because the surface is small or well insulated.

2.1 Cooling of an isothermal wire

If a long region is normal conducting or current-sharing, the heat conduction along the wire is insignificant and the helium absorbs all heat Q locally. Under these conditions the critical current density is reached at $T = T_g$. In steady state the power generation must be less or equal to the cooling. The ratio of Joule heating to surface cooling is called the Steckly parameter α_{St}.

$$\alpha_{St} = \frac{\rho_m J_C^2 (1-f)^2 A}{fPh(T_C - T_B)} \tag{6}$$

If the cooling always dominates over heat generation (for $\alpha_{St} < 1$) the wire is cryogenic stable. A coil with $\alpha_{St} < 1$ can not quench provided liquid helium is available. Large coils like solenoids in the experiments at storage rings or for tomography, or in toroids to store magnetic energy are in this very favourable situation. But accelerator or fusion reactor magnets are usually far away from this ideal state. In the HERA dipoles, for instance, $\alpha_{St} = 22.3$; so the coils are definitely not cryo stable. Full cryogenic stabilisation is uneconomical for accelerator magnets. One has therefore to accept that such magnets may and will quench.

In reality, the normal zone does not extend to infinity nor can the cooling through the surface be ignored. Taking both heat-conduction and cooling into account simultaneously one arrives at the Maddock [6] criterion, also called "equal area theorem". It states that heat generation and heat removal should cancel each other globally but not necessarily locally. The conductor will be stable, provided the surplus heat at the hot spot can be conducted to a cooler place, where surplus of cooling is available. The difference of the integrals over cooling and heating along the wire from the superconducting region up to the hottest spot is what counts. If the difference vanishes, the conductor stays stable. As the temperature increases monotonously, the integration over space can be replaced by integration over temperature. The theorem states:

$$\int_{T_B}^{T_H} \left[\frac{Ph}{A}(T - T_B) - \frac{A}{P} Q(T) \right] k(T) dT = 0, \tag{7}$$

where T_H is the maximum temperature of the normal zone.

2.2 Cooling by heat conduction along the wire

To find the smallest normal conducting region that does not propagate, and to estimate the energy needed to create such a minimum propagating zone let us rewrite Eq. (1) in three dimensions and study the expansion of a small normal conducting zone. The aim is a solution for the stationary case without external cooling (adiabatic limit). Additional cooling would allow a larger zone to be stable. It is therefore possible to ignore the cooling, if one is only interested in the minimal zone.

The thermal conductivity in the longitudinal direction, k_x, is well known, while the radial heat conduction must be estimated. It turns out to be sensible to set $\alpha^2 = k_r / k_x$. α is typically 2-3%. Equation (1) can be rewritten in spherical co-ordinates introducing a scaled radial co-ordinate $R = r/\alpha$.

$$\frac{d^2T}{dR^2} + \frac{2}{R}\frac{dT}{dR} + \frac{(1-\varepsilon)Q}{k_x} = 0. \tag{8}$$

The factor $(1-\varepsilon)$ takes care of the fact that a fraction $\varepsilon \approx 0.1$ of the coil (if Rutherford cable is assumed) consists of voids filled with helium.

The solution of Eq. (8) consists of three parts. Close to the origin, in the region with Joule heating, (that is, inside a "sphere" of radius R_g) the temperature is:

$$T = T_g + A\frac{R_g}{R\cdot\pi}\sin\left(\frac{\pi\cdot R}{R_g}\right), \tag{9}$$

with
$$R_g = \sqrt{\frac{k_x(T_C - T_g)}{(1-\varepsilon)Q(T_g)}}\cdot\pi = \frac{\sqrt{f}}{J_{avg}}\sqrt{\frac{k_m(T_C - T_g)}{(1-\varepsilon)\cdot\rho_m}}\cdot\pi. \tag{10}$$

If the temperature exceeds the critical temperature Eq. (8) overestimates the heat production. Thus the Eqs. (9) and (10) describe an upper limit only. In real space the minimum propagating zone is an ellipsoid, oriented along the quenching wire, with a total length of $2\cdot R_g$ and a transverse diameter of $2\cdot R_g\sqrt{\frac{k_r}{k_x}} = 2\alpha R_C$.

Outside the heat producing zone, $(R \geq R_g)$, the heat generating term in Eq. (8) vanishes and the solution is

$$T_{out} = T_g(1-(1-\frac{R_g}{R})A).$$

At the cold side the temperature reaches the bath temperature in principle at infinity. In reality the zone cannot extend that far. Firstly it would take a very long time to reach a steady state and secondly an infinite energy would be required. Far away, at a distance $R_{cold} = x\cdot R_g$, the longitudinal heat flow is insignificant compared to the cooling by the helium through the insulation. At this point and beyond, Eq. (8) is invalid. The longitudinal temperature gradient does not need to be continuous any more. The cold boundary determines the constant A and hence the maximum temperature inside the minimum propagating zone. If the temperature reaches the temperature of the bath at $R_{cold} = x\cdot R_g$, the temperature profile can be written as

$$T = \begin{cases} T_g + \dfrac{x(T_g - T_B)R_g}{(x-1)\pi R}\sin\left(\dfrac{\pi R}{R_g}\right) & \Leftarrow R \leq R_g \\ T_g - \dfrac{(R-R_g)x(T_g - T_B)}{R(x-1)} & \Leftarrow x\cdot R_g \geq R \geq R_g \\ T_B & \Leftarrow R \geq x\cdot R_g \end{cases} \tag{11}$$

Figure 4 shows as an example four temperature profiles for minimum propagating zones in the DD0019 prototype magnet for SSC [3] at its nominal operating point (at that time). The bath temperature is 4.35K in this case and the field about 6.9T. The four curves correspond to x = 1.25, 1.5, 1.75 and 2 respectively. The heavily drawn curve describes the most probable profile (x = 1.5), as explained below. The horizontal scale gives the longitudinal size of the disturbance.

The energy needed to create a minimum propagating zone depends on the actual temperature profile and hence on the boundary where the bath temperature is reached again. Because the temperature depends only on the radius the energy is described by the double integral

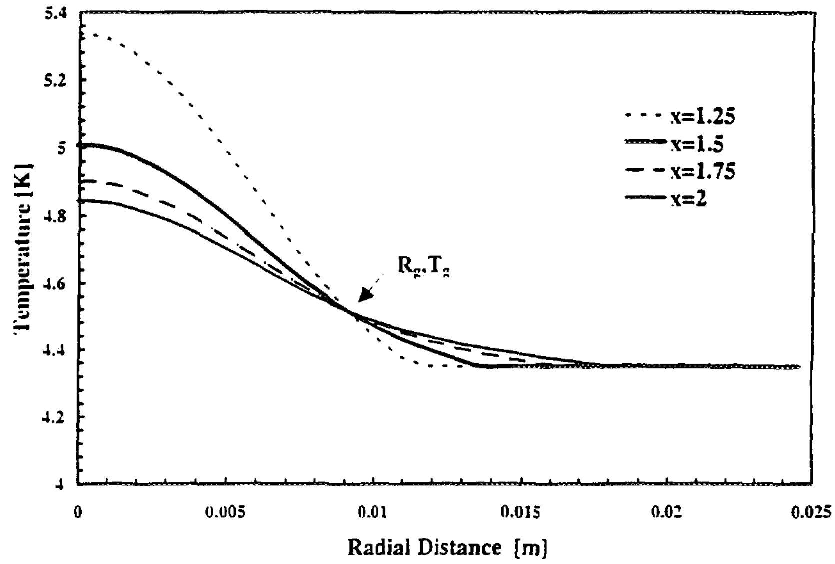


Fig. 4 Temperature profiles of minimum propagating zones. The bath temperature is reached at the radial distance $r = xR_g$.

$$E = \int_0^{xR_g} \int_{T_B}^{T(r)} C_{avg}(\vartheta) d\vartheta \; 4\pi (r\alpha)^2 dr. \tag{12}$$

The factor α^2 takes into account that the lateral dimensions have to be rescaled to physical dimensions. Analytic calculations of this integral are quite complicated. Even approximations lead to integrals that have to be solved numerically. It turns out that (12) has a shallow minimum between $1.4 < x < 2.5$. The energy needed to create the temperature distribution varies here by less than 20% as a function of x; the exact minimum depends on other parameters, however. The quoted minimum energy to produce a minimum propagating zone is a good measure of the relative stability of cable configurations or operating points.

Figure 5 shows the minimum propagating energy for a typical HERA dipole magnet (at 4.5 K) and for the SSC prototype magnet DD0019 (at 4.35 K) as a function of the magnetic field, assuming α = 3%. The bath temperatures correspond to their nominal values respectively. The arrows indicate the peak fields in the coils under nominal conditions. Obviously the SSC magnet reaches a much higher field. At the nominal operating point the minimum propagating energy has dropped below 10 μ J, in fortuitous agreement with measurements [3] and other calculations [7]. The HERA design is considerably more conservative. The minimum propagating energy at nominal conditions is 0.15 mJ and the corresponding volume is $7.1 \cdot 10^{-9}$ m^3.

It is often assumed that a lower bath temperature would provide for a more stable operation. Figure 6 shows the minimum propagating energy for the HERA coil as a function of the bath temperature. The gain in safety is quite moderate when lowering the bath temperature because the heat capacity is so small. However, below the lambda point the improved heat removal changes the situation completely.

The size of the minimum propagating zone depends on the geometric mean of heat conductivity and the electrical conductivity properly scaled by the relative fraction of copper.

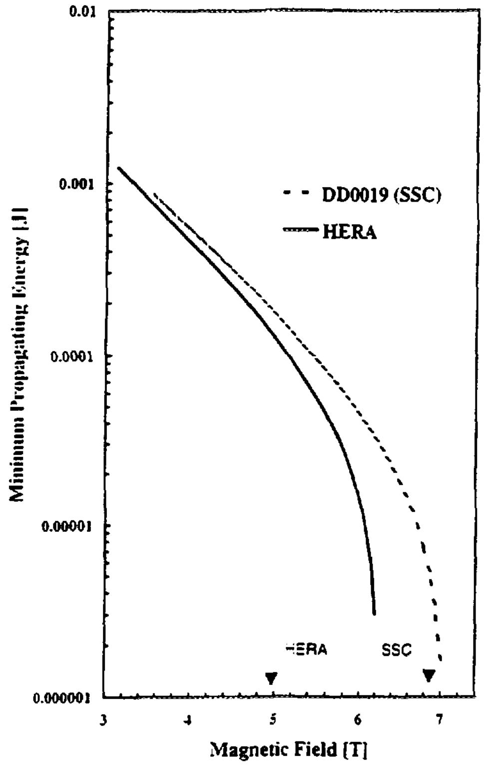


Fig. 5 Minimum propagating energy as a function of the magnetic field for a typical HERA dipole (T_B = 4.5 K) and the early SSC prototype magnet DD0019 (T_B = 4.35 K). The arrows indicate the respective peak field values under nominal operating conditions.

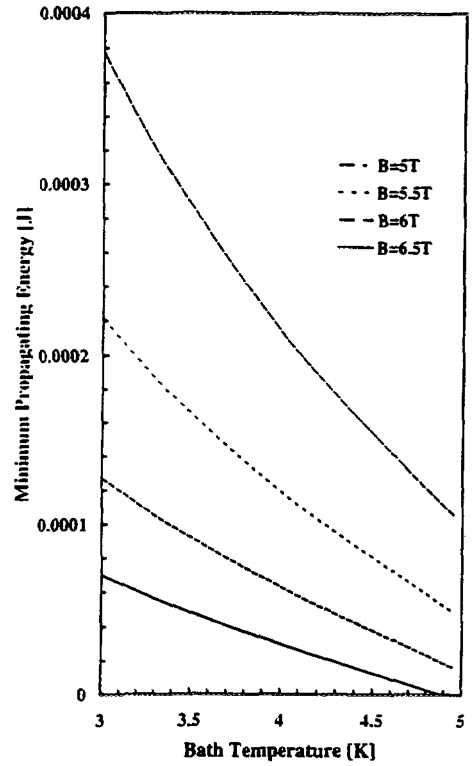


Fig. 6 Minimum propagating energy for a typical HERA dipole as a function of the bath temperature for four values of the highest magnetic field in the coil.

Removal of matrix material decreases the size of the minimum propagating zone if the amount of superconductor is unchanged. It also decreases the minimum propagating energy, because less material needs to be heated above the current-sharing temperature.

3. QUENCH EXPANSION

One would like the largest possible minimum propagating zone for magnet protection, because a large zone also means a "large" energy to trigger the quench. One may hope that such a kind of disturbance is rare and hence the likelihood of quenches is small. However once a quench has started, one would like the normal conducting zone to expand quickly. Often the magnet has to absorb the stored magnetic energy or at least a sizeable fraction of it. A large conducting volume results in a low energy density and hence a low maximum temperature.

Once a normal zone has started to grow it will continue to do so as long as the current density and the magnetic field are high enough. The low heat conduction of the insulation and latent heat of the helium content in the cable impede the transverse expansion. Therefore the normal zone will expand dominantly along the cable.

Quench velocities in a piece of cable in the field of an extra magnet are easy to measure. Voltage taps distributed along the wire are sufficient to detect the arrival of the normal zone by measuring the resistive voltage over the normal zone. Figure 7 shows a typical example.

The voltage rises linearly when the normal zone expands between two voltage taps. Thereafter the voltage rises slowly because the resistance is almost constant below some 25 K.

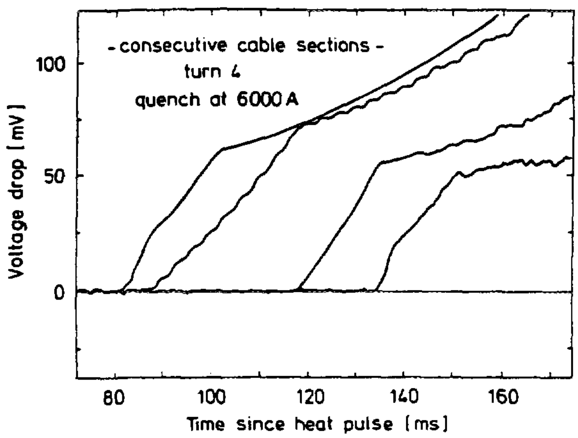


Fig. 7 Measurement of the voltage drop over consecutive cable sections during a quench as a function of time

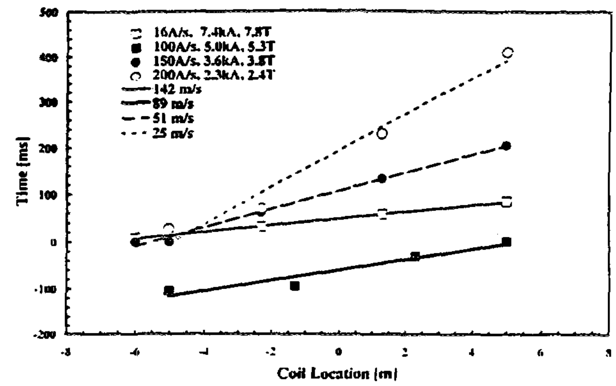


Fig. 8 Measurements of the time of detection of a local field distortion in "quench antenna" coils as a function of the coil position. The four sets of data correspond to quenches occurring at different ramp rates and hence currents and magnetic fields in the SSC dipole prototype DCA 312. The lines are fits to the data [14].

Measurements at real magnets are more difficult. Basically four methods exist to measure the quench velocity in a magnet. One can try to attach voltage taps to the coil during assembly. This does not work too well because the attached wires are always a hindrance during the coil compression under curing. At best a few connections in the magnet heads, where access is somewhat easier, are possible.

Another method was used to observe quenches in a short HERA prototype magnet [8]. Devices that carried needles like a porcupine were inserted into the coil aperture. The needles could be expanded and pierced into the inner coil layer by turning a key from one of the magnet ends. Several porcupines could be inserted. However, it is both difficult and dangerous to insert such a device because insufficient alignment results in shorts between the coil windings.

A developing quench is rather violent and produces ultrasonic noise. It was observed that the flux redistribution indeed makes noise [9] which can be picked up by microphones. In particular, if the magnetic field changes rapidly, this method of monitoring has some advantages because the acoustic emissions are not electromagnetic and thus immune against electromagnetic noise. Acoustic measurements with several microphones can also reveal the origin and propagation of a quench. The drawback is, however, that quite often acoustic signals are emitted that do not correspond to a quench or vice versa [10].

The most elegant way to observe, measure, and trace back quenches has recently been invented by Krzywinski [11] and since been used at CERN [12] and SSCL [13], [14].

A moving wire or the current redistribution between strands at the front of an expanding normal zone creates field distortions. They induce signals in pick-up coils, properly located inside the free aperture of the quenching magnet. Such antennae can be made insensitive to changes of the main dipole field by quadrupolar or sextupolar coil arrangements. This works very much in the same way as in the measurement of the magnetic multipole components. The twin aperture magnets for LHC offer the favourable possibility to subtract signals of corresponding radial coils in the two apertures. Radial coils are easier to produce to the same dimensions and the results are easier to interpret.

To detect quenches and observe the quench propagation it is sufficient to insert four radial coils, rotated by $\pi/2$, preferentially all mounted on the same shaft and covering the length of the magnet under investigation. To cancel dipole contributions either a corresponding coil in the second aperture or the properly weighted average of the other three coils can be used. Alternatively, sets of coils can be made that measure quadrupolar and sextupolar field components, both regular and "skew", i.e., shifted by $\pi/4$ and $\pi/6$ respectively. For quadrupole magnets, of course, sextupolar and octupolar pick-ups are necessary.

The starting point (radius and azimuth), the direction and the change of the magnetic strength characterise the transverse motion of a magnetic moment. Four different coils are sufficient to measure this. At SSCL this technique allowed to locate the quench origin at the inner edge of a particular winding turn.

The longitudinal position of the quench origin can be deduced from the development of the signal with time. To achieve this at least two, preferentially many, sets of pick-up coils are stacked along the length of the magnet. Depending on the circumstances, many identical sets of coils can be mounted on a common shaft or the sets are positioned individually, employing the technique of the moles, as for harmonic measurements. If more than one set of coils detects the quench and if the quench velocity is reasonably constant the axial position of the quench origin can be determined to better than 1 cm.

To measure the velocity of the quench front two methods can be used. In the simplest case one uses the pick-up coils as voltage tap replacements. The distance between the pick-up coils is known and times of the arrival of the quench front can be detected. Figure 8 shows an example from SSCL [14]. The aim of the experiment was to find the reason for the ramp rate dependence of the apparent critical current. Hence the various ramp rates given in the index correspond in reality to different current densities and fields as indicated. Note that the measured velocities are as large as 100 ms^{-1}.

In the second approach, the magnetic flux in the pick-up coil is calculated, by integrating the induced voltage. In Fig. 9 this is done for four consecutive pick-up coils in a one-meter long LHC model magnet [12]. The time intervals at which the propagating normal zone passes by can easily be measured and compared to the slope. Again about 100 ms^{-1} are observed. One may even argue that the velocity increases slightly. In Fig. 10 [12] the signals in three pick-up coils at the same longitudinal position are shown for a somewhat longer time. The second bump (or dip, respectively) corresponds to the current redistribution in one of the adjacent turns. The quench obviously needs 14 ms to propagate azimuthally by one turn. Unfortunately, a complete and consistent set of quench velocity measurements in a large magnet has not so far been published.

The quench antenna method is also useful to study other phenomena. It has been observed that sharp signals are accompanied by mechanical oscillations. The damping of the oscillations depends on the absence or presence of the flux iron. Obviously the potential of the method has not yet been fully exploited.

Measurements on cables showed that the normal zone expands at a constant velocity, except in the very beginning, where the manner in which the quench has been initiated has some influence. In real magnets, however, the magnetic energy has to be dumped and the current has to be decreased in order to protect the coil from melting. The rapidly changing

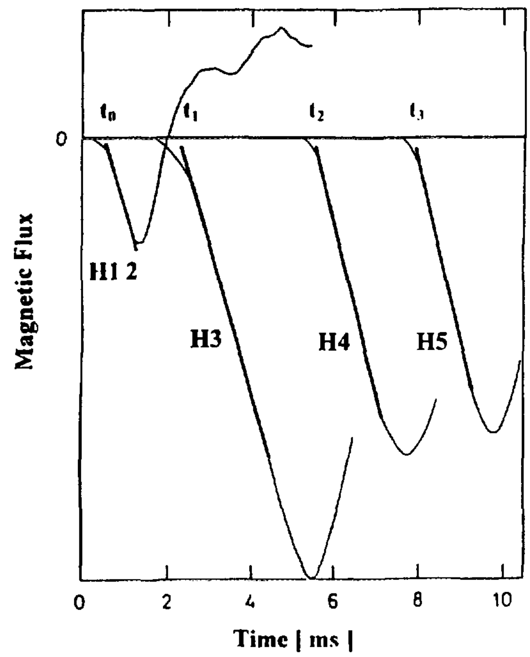


Fig. 9 The development of magnetic flux in a series of "quench antenna" coils as a function of time shows the longitudinal propagation of the quench in a LHC model magnet [12].

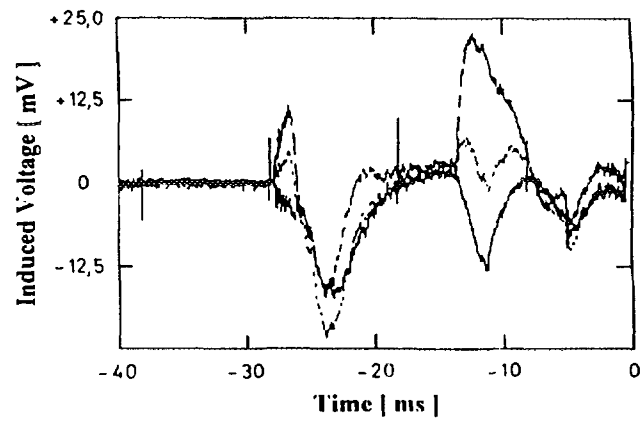


Fig. 10 The superimposed voltages of three "quench antenna" coils of the same longitudinal section show signals when the quench front of one cable passes by and when, after 14 ms, the normal zone has expanded to the adjacent winding.

magnetic field causes movements of the coil and eddy current heating. (This may even happen at normal ramp rates, as was demonstrated by the measurements of Fig. 8.) The helium may be blasted along the wire, preheating the wire ahead of the front. All these effects are difficult to take into account. Needless to say, approximate explicit formulas can only be calculated under the assumption that all material parameters depend on temperature in very simple ways. Often some averaged values have to be assumed.

In the calculations for the quench velocity the cable is imagined as a homogenous piece of metal with direction dependent heat conductivity. The longitudinal heat conductivity is high. The radial conductivity is low and not well known. The superconducting material, the "dirty" copper in the centres of the strands, the poor contact between the strands, and finally the helium in the voids contribute to the lateral conductivity in a complicated manner. Nonetheless, one can schematically write the radial quench velocity inside the cable as

$$v_r = \sqrt{\frac{k_r}{k_x}} v_x . \tag{13}$$

It is tempting to apply the same technique as used in deriving the minimum propagating zone and the temperature profile. This leads however to integrals that do not have a closed form representation. One has therefore to start the analysis at a time at which the quench is already well in progress. At such a late time the radial expansion has come to a temporary halt at the surface of the wire or cable, because the heat conductivity in the helium-filled insulation is very low and the latent heat comparably large. The quench propagation happens mainly along the wire. Therefore a surface-cooled rod resembles the quenching cable, neglecting all transverse expansion for the moment. In such a picture the radial temperature gradient vanishes everywhere but on the cooled surface. Thus heat conduction is longitudinal only and the one-dimensional description is quite adequate.

In the simplest approximation one can ignore the current sharing, all temperature dependencies and the cooling. The result is the velocity in the adiabatic limit. It applies for well insulated and relatively heavily-stabilised cable only. However, it gives the order of magnitude for current densities close to the critical density, provided later mentioned complications can be ignored.

In the adiabatic limit, with constant heat capacity and constant heat conduction, and without current-sharing, the quench velocity can be calculated to [2]:

$$v_{ad} = \frac{J_{avg}}{C_{avg}} \sqrt{\frac{\rho_{avg} * k_{avg}}{(T_S - T_B)}}, \tag{14}$$

where J_{avg} and ρ_{avg} are averaged over the conductor and T_S is the temperature at which all Joule heating is assumed to take place in the absence of current sharing. Using the Wiedemann-Franz-Lorentz law, which relates electrical and thermal conductivity, Eq. (14) can be written as $\left(L_0 = 2.45 \cdot 10^{-8}\,\mathrm{W\Omega K^{-2}}\right)$

$$v_{ad} = \frac{J_{avg}}{C_{avg}} \sqrt{\frac{L_0 T_S}{T_S - T_B}}. \tag{15}$$

Heat capacity and resistivity are not constants and current-sharing and cooling can usually not be ignored. In particular cooling can decrease the normal zone again. That would formally correspond to a negative velocity, which Eq. (14) cannot describe. Various approximations have been tried. However, theoretical formulae are mainly useful for interpolating measurements because it is only on the basis of a measurement that one can decide which approximations may be adequate.

M. Wilson [2] takes both steady-state contributions and transient terms into account. Using the relative current density ($i = J/J_C$) the steady-state contribution is formulated in terms of the ratio of cooling on the surface and the heating:

$$y = \frac{hP(T_S - T_B)}{A J_{avg}^2 \rho_{avg}} = \frac{(T_S - T_B)}{(T_C - T_B)\alpha_{St} i^2}. \tag{16}$$

The transient term z compares the latent heat with the heat capacity:

$$z = \frac{Q_L}{C_{acg}(T_S - T_B)}. \tag{17}$$

Combining all effects Wilson finds for the quench velocity:

$$v = v_{ad} \frac{(1 - 2y)}{\sqrt{yz^2 + z - y + 1}}. \tag{18}$$

Without transient effects, Eq. (18) can be rewritten in terms of the relative current density i and the Steckly parameter α_{St}

$$v = v_{ad}\frac{1-2y}{\sqrt{1-y}} = \sqrt{\frac{hkP}{AC^2}}\frac{\frac{(T_C - T_B)}{(T_S - T_B)}\alpha_{St}i^2 - 2}{\sqrt{\frac{(T_C - T_B)}{(T_S - T_B)}\alpha_{St}i^2 - 1}}. \tag{19}$$

With the special assumption $T_S = (T_t + T_C)/2$ Eq. (19) becomes

$$v = \sqrt{\frac{hkP}{AC^2}}\frac{\alpha_{St}i^2 - 2 + i}{\sqrt{1 - i/2}\sqrt{\alpha_{St}i^2 - 1 + i/2}}. \tag{20}$$

Cherry and Gittelman [15] assumed that all Joule heating appears suddenly at $T_S = T_C$, which simplifies the formula (19) to

$$v = \sqrt{\frac{hkP}{AC^2}}\frac{\alpha_{St}i^2 - 2}{\sqrt{\alpha_{St}i^2 - 1}}. \tag{21}$$

More elaborate formulae can be found in the literature that modify Eq. (20) slightly to fit data somewhat better. None of them describes the high quench velocities in long SSC and LHC dipoles mentioned earlier.

Many factors and mechanisms contribute to the expansion of the normal zone in a magnet but are omitted in the calculations. Incidentally, large and rapid changes of the magnetic field and the current result in a further velocity increase [16], as do thermo-hydraulic effects.

Thus, the quench performance of a new design can at best be roughly estimated and modelled. Measurements on a sizeable piece of cable under realistic conditions are strongly advisable to gain insight into which quench expansion mechanism dominates. The quoted formulae may then be used to interpolate between measurements.

4. HEATING OF THE COIL AFTER A QUENCH

4.1 Hot-spot temperature

The quenching region is at all temperatures ranging from the bath temperature up to a maximum temperature which is found at the point where the quench was initiated. The hottest spot is also the spot most in danger and hence one has to be concerned mostly with this peak temperature. As a simplification local adiabaticity is assumed, because it is always a conservative assumption. Furthermore a quench lasts only about one second which is an order of magnitude less than large scale heat exchange in a cryostat. Under this assumption the locally produced heat results in a local temperature rise:

$$J_{avg}^2(t)\rho_{avg}(T)dt = C_{avg}(T)dT. \tag{22}$$

All quantities are averaged over the winding cross section, including insulation and helium in the voids. Equation (22) reads after rearrangement and integration:

$$\int_0^{\infty} J_{avg}^2(t)dt = \int_{T_B}^{T_H} \frac{C_{avg}(T)}{\rho_{avg}(T)}dT = J_0^2\tau = F(T_H) \tag{23}$$

The rearrangement implies that $\rho(T)$ does not depend explicitly on the time. This is not exactly true. Hence an effective resistivity should be used as approximation.

For a given coil the function $F(T_H)$ can be used to estimate the maximum temperature T_H. $F(T_H)$ depends only on known material constants and can be calculated. The integral over the squared current density is easily determined in the case of a single magnet. In fact, for simplicity only the integral over the square of the current is often calculated and quoted in units of $10^6 A^2S$, called MIITS. Here the current density squared is used in order to be applicable to all similarly built cables.

A direct measurement of the temperature is cumbersome and requires many temperature sensors in the coil. Alternatively, one can determine the average temperature between two voltage-taps from the resistance, i.e., from the voltage drop. Figure 11 shows such a set of measurements [8] plotted against the integral over the squared current density. To relate that to currents, one has to know that the cable area was $1.32 \cdot 10^{-5}$ m^{-2}. The fully drawn curve shows $T_H(F)$, which is the inverse of the function $F(T_H)$ calculated for a HERA type coil. The equivalent curve for a pure copper coil $(\rho_{4.2K} = 10^{-10}\,\Omega\text{m})$ is drawn dashed. The case of pure NbTi cannot be illustrated in the same figure. The curve would essentially coincide with the vertical axis because the large resistivity results in a dramatic heating. Despite the fact, that both C and ρ are complicated functions of the temperature, the function $T_H(F)$ can usually be approximated fairly well as a parabola with some offset.

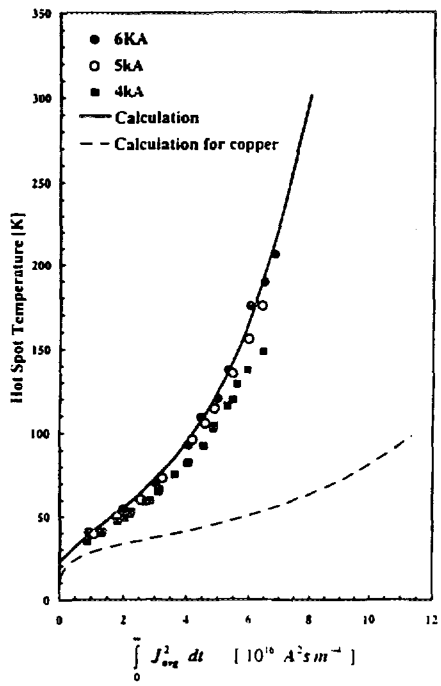


Fig. 11 Hot-spot temperature in a short test dipole for HERA as a function of $\int J^2 dt$: full curve calculated, dashed curve for pure copper $(\rho_{4.2K} = 10^{-10}\,\Omega\text{m})$.

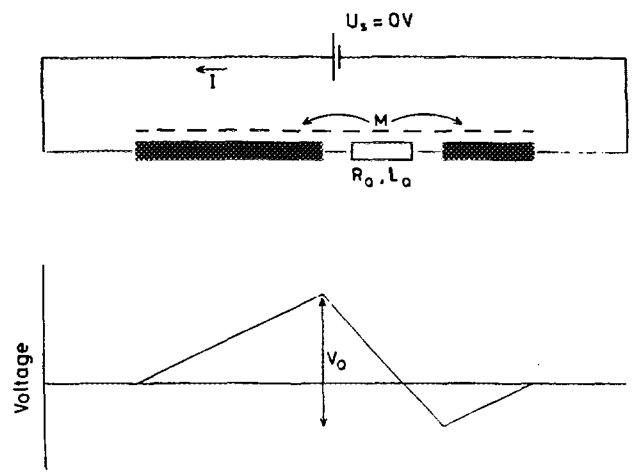


Fig. 12 Equivalent electrical circuit of a quenching magnet and the voltage distribution

The hot spot temperature increases roughly as the square of the current density. This is so because the characteristic time for a quench is roughly inversely proportional to the quench velocity. This in turn is roughly proportional to the current density. In heavily stabilised cable, as is used in the interconnection region between two magnets, the copper properties dominate completely. In these regions a quench can last until the adiabaticity fails after some tens of seconds. The coil, however, reaches at the hot spot a temperature of 100 K in 0.3 seconds.

Good engineering practice would call for 100 K as the upper limit, because the thermal expansion and the mechanical stresses in the coil and support structure start to increase above this temperature. Common practice is however, to go far beyond this point in order to save coil volume, conductor, and hence cost.

One could believe that a current decay time of less than a second may cause high voltages across the terminals of the quenching magnet. Referring to Fig. 12 the voltage can be written as $V_Q = I\, R_Q - \left(L_Q + M\right)\frac{dI}{dt}$ and $L\frac{dI}{dt} = I\, R_Q$, because the power supply is assumed to be an ideal current source. L is the inductance of the coil, M the mutual inductance between the normal conducting and the superconducting parts of the magnet. R_Q denotes the resistance and L_Q the inductance of the normal conducting part. Combining the two equations yields

$$V_Q = I\, R_Q\left(1 - \frac{L_Q + M}{L}\right) \approx I\, R_Q\left(1 - \frac{M}{L}\right) \tag{24}$$

which is always less than or equal to $V_Q^{max} = I\, R_Q$.

4.2 Numerical calculations

Equation (1) cannot be integrated analytically even if all material properties are assumed to be constant, although most of them depend in reality on the temperature and on the magnetic field. Also approximate solutions fail if external actions, like firing of heaters or bypass thyristors (diodes) to control the hot spot temperature, are to be taken into account. A series of numerical programs has been written to simulate quenches on the computer. The first program published was "QUENCH" [17]. Starting with a current I_0 at the time t_0 the quench velocity is calculated using the calculated magnetic field and the corresponding critical temperature to determine the material properties. Assuming a constant expansion speed the volume at the time t_1 is calculated. Now the average temperature in the normal conducting volume V_1 and the current decay, if applicable, is determined. In the next step the material properties are recalculated using the new temperature, a new normal conducting layer is added, and the temperatures in the inner layers are updated. Each layer keeps its own record of temperature history. Magnetic field and temperature distribution determine the total coil resistance and hence the resistive voltage drop at any time interval. From the external protection resistors or diodes, the inductance, and the coil resistance the coil current can be calculated which is then used for the next time interval. At this point also inductive couplings and other complications can be taken into account. For the transverse expansion either of the two approximations, mentioned above, can be used. The simple Euler algorithm with variable material constants converges acceptably provided the time steps are small enough. In a similar fashion K. Koepke predicted successfully the behaviour of the TEVATRON magnets [18]. These programs as well as the adopted version of QUENCH for HERA [19] or QUENCH-M [20] are bulky FORTRAN programs specially tuned for a particular magnet type and protection circuit. The ideas lead Pissanetzky and Latypor [21] to a modern version, applicable to magnets with a single or with multiple coils, with or without iron, operating in the persistent mode or from external power. The method assumes again that a quench starts at an arbitrary point of the coil and propagates in three dimensions. Multiple independent fronts can coexist. Local magnetic fields and inductive couplings of the coils are calculated by the finite element method. All properties, like fields and temperatures, are obtained by solving the corresponding equations at each point in space and time. For the LHC magnets Hagedorn and Rodriguez-Mateos designed a different generally applicable simulation package. The general simulation tool, called QUABER [22], is based on a professional tool, called SABER (trademark of Analogy Inc.). Bottura and Zienkiewicz [23,24] developed a finite element program for magnets with "cable-in-conduit", that are magnets with forced helium flow. All programs mentioned above and many variants of them are able to describe the "typical" quench fairly well. However the result of the model calculation depends essentially on the assumptions. It is advisable to investigate as many different input assumptions as reasonable to determine the uncertainty in the calculations.

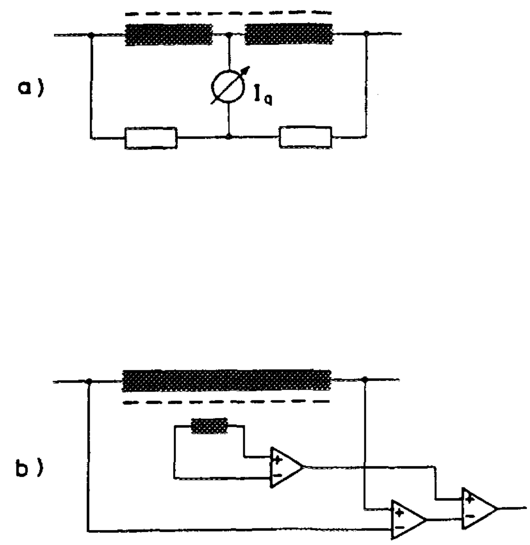
5. QUENCH DETECTION AND EXTERNAL SAFETY CIRCUITS

The minimum response necessary to prevent conductor burnout when a quench occurs is the disconnection of the power supply. This requires of course the detection of the quench. Several signals can be used to sense a quench. Acoustic emissions precede a quench and follow it. This signal, however, is not very specific because noise emission is not always accompanied by a quench. A resistive voltage $U_q = R_q I$ builds up when a normal zone grows and expands. The rising resistance leads also to a change in current which in turn induces an inductive voltage. In the same fashion an inductive voltage arises when the current changes for other reasons or whenever the coil is magnetically coupled to a coil with changing current [25]. Somehow the inductive voltages have to be cancelled. In the simple case of a single coil, as indicated in Fig. 13a, a single bridge circuit will be a reliable solution. For this purpose a centre tap on the magnet coil is needed and the bridge has to be balanced to better than 0.5%. Once set properly, which may be tedious, the bridges can stay unchanged for years as experience shows. Of course, this method can never detect a quench that develops in both half coils identically. Some additional measures exclude this rather exotic case. In a large system, for example, the bridge method can be repeated for groups of magnets.

Instead of subtracting the inductive voltage by a bridge directly, one can also measure it by some additional device and subtract it from the coil voltage electronically. Figure 13b indicates as an example the measurement with an additional field coil. Alternatively the average voltage of a large set of identical magnets in series can serve as a measurement of the inductive voltage. In either case, problems may arise with the dynamic range, the initial adjustment and eventual drifts with temperature.

Magnetically coupled coils, like stacked correction coils or the windings in fusion reactor magnets, present a particular problem. Current changes in any of the coupled coils induce voltages in all other coils too. Hence the subtraction method has to be expanded to all combinations of all coils. This procedure is somewhat akin to the Gaussian elimination in matrix diagonalisation [25].

The magnet coils float during a quench at an unpredictable potential with respect to ground. The measurement technique has to take care not to destroy the magnets by more than one unintended short to ground. Thus it is advisable to add a resistor in series with the potential tap as close to the coil as practicable in order to limit the possible current to ground and hence the damage. If one applies the bridge detection method high-valued series resistors will decrease the sensitivity. In this case a protection with high voltage fuses is possible. In fact the simplest high voltage fuse for this purpose is a piece of wire-wrap wire, as used in electronics, at a sufficient distance from conducting material. Note that the fuses have to be tested continuously as parts of the quench detection circuit.



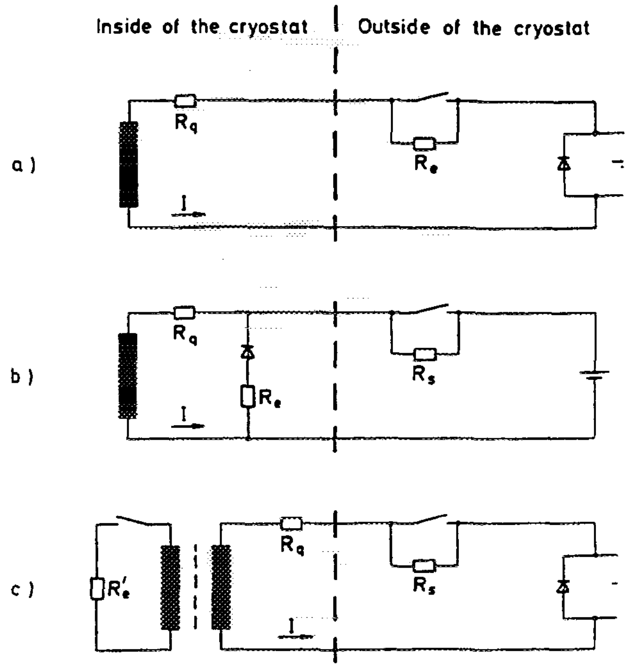


Fig. 13 Quench detection by a) measuring the current through a bridge or b) comparing the total voltage with the inductive voltage.

Fig. 14 Possible energy extraction schemes for a single magnet: a) external resistor, b) internal diode and resistor plus external resistor, c) magnetically-coupled resistor.

After the detection of the quench the stored magnetic energy has somehow to be dumped. Rephrasing the power supply to pump the energy back into the mains is possible but usually too slow. Basically all ideas about how to "protect" superconducting coils have been well known for many years [26]. For a single magnet, one can switch off the power supply and dissipate a large fraction of the stored energy by one of the circuits sketched in Figs. 14a, b and c. In circuit (a) the current continues to flow through a diode ("free wheel diode") and a load resistor R that determines basically the maximum voltage $U_{max} = I_0 R$, the exponential decay time $\tau = L/R$, and hence the hot spot temperature. In circuit (b) an extra diode has been added. It could also be installed inside the cryostat as a so called "cold diode". The current will commute partially into the diode branch once the diode knee voltage is reached. This is a few tens of a volt at room temperature and a few volts at 4.2 K. The current will commute independent of the position of the current switch and the quench detection. In the worst case the diode will melt, but not the magnet. Circuit (c) contains an inductively-coupled resistor R'. This can be the support cylinder of a large solenoid or some other structural element. The heat produced in R', because of the change in the primary current and the magnetic coupling, can be used to accelerate the quench propagation. This solution, termed "back quench", is preferred for slowly ramped magnets but it is not directly applicable for fast ramping superconducting devices.

To summarise, a number of reliable methods have been developed to protect a single magnet after a quench. The quench signal has to be detected and discriminated from noise signals. The power supply has to be switched off without interrupting the magnet current. The stored energy has to be dissipated in suitable devices. If necessary, the quench can be spread artificially by activating heaters in or at the windings.

6. PROTECTION OF A STRING OF MAGNETS

An accelerator consists of large number of magnets in series. A fusion reactor magnet, even worse, consists of a large number of magnetically-coupled coils. The protection of such a string or group of coils is a challenge. For example, the inductance in the HERA ring adds up to $L = 26.5$H. At 5.5 T, 470 MJ are stored in the ring; an energy sufficient to melt 780 kg of copper.

Unfortunately, a simple switch, as in the case of one magnet, cannot work. A magnet is barely able to absorb its own stored energy. Simple switching-off dumps almost all stored energy into the quenching magnet and destroys it. On the other hand, energy extraction with external resistors would require an enormous resistance and hence a voltage of more than 300 kV.

The recipe is therefore:

- detect the quench,
- isolate the quenching magnet,
- spread the energy,
- subdivide the inductance (if possible).

The principle is best discussed by explaining a few examples. Figure 15a shows the quench detection system of the TEVATRON, the first large superconducting accelerator. It is based on the measurement of voltage differences. Average voltage differences are calculated, including the inductive voltages during ramps, and compared with the measured values. A significant discrepancy indicates a quench. The large values for the resistors, chosen for safety reasons, together with the cable capacity introduce a sizeable signal distortion that has to be corrected.

The system developed for HERA (Fig. 15b) is based on bridge circuits for each magnet. Additional bridges over a number of magnets increase the redundancy. A radiation-resistant magnetic isolation amplifier that is insensitive to noise pickup amplifies the bridge current. A similar but more versatile version has been proposed for UNK [27]. For the LHC also a bridge circuit is in discussion. However, isolation amplifiers with semiconductors are presently favoured.

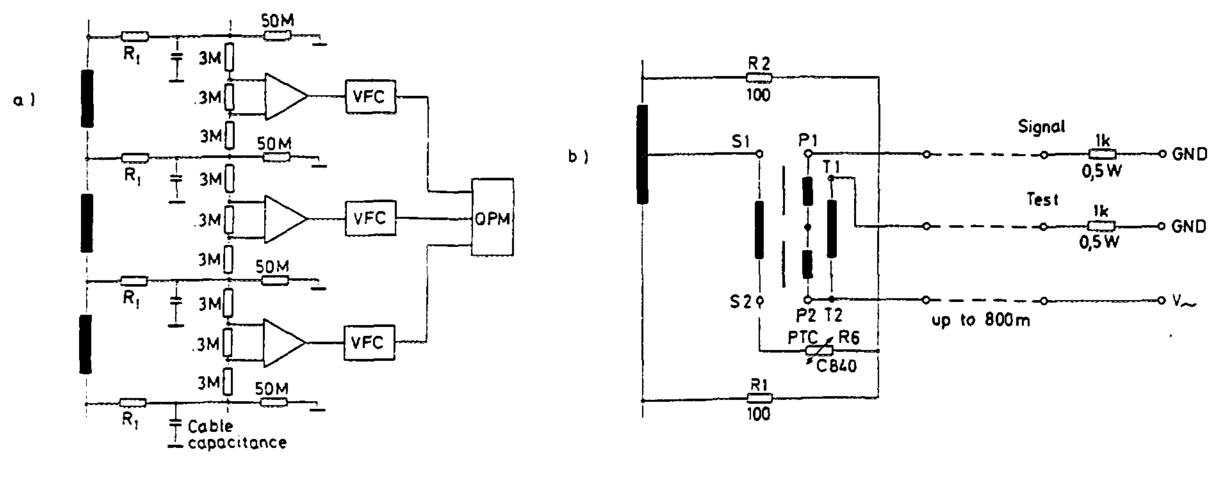


Fig. 15 Quench detection circuits used in large systems, a) TEVATRON, b) HERA

Next the energy of the unquenched magnets has to be kept away from the quenching magnet. Basically, guiding the main current around the magnet achieves this effect. Figure 16 shows an equivalent circuit diagram. The total inductance L of the magnet string is much larger than the inductance L_l of a single magnet. Hence the main current I decays with a much larger time constant than the current I_Q in the quenching magnet. The differential equation for I_Q is (neglecting diode voltage drops)

$$L_1 \frac{dI_Q}{dt} + I_Q R_Q(t) = (I - I_Q) R_b . \qquad (25)$$

Since $R_Q(t)$ grows with time an analytic solution is not available. But once the whole coil has become normal one arrives at a steady state solution

$$I_Q = I \frac{R_b}{R_b + R_Q} \approx I \frac{R_b}{R_Q} . \qquad (26)$$

To minimise the current remaining in the coil, the resistor R_b in the bypass line should be made as small as possible.

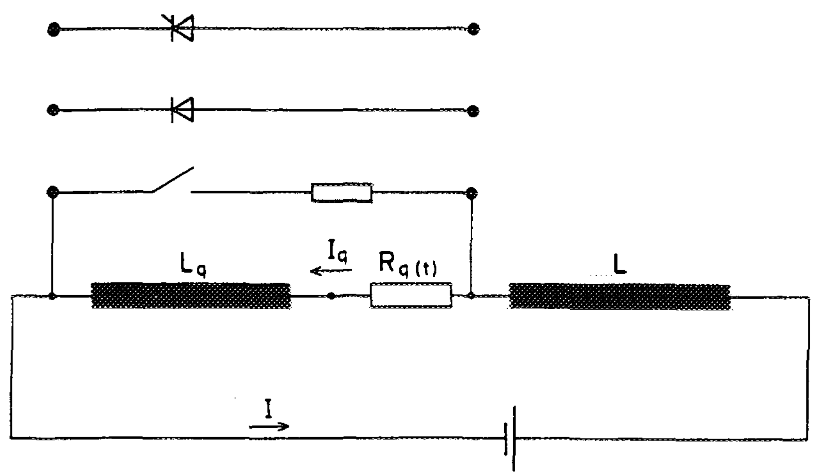


Fig. 16 Equivalent circuit of the bypass of a quenching magnet

Two basic solutions exist. Thyristors act as fast switches at fast ramping machines like the FNAL TEVATRON, SSC or UNK. Figure 17 shows part of the electrical circuit for the TEVATRON. In the TEVATRON magnets the return bus is integral part of the coil and has to be protected. Therefore half the magnets are fed by one bus with half a winding powered by the return bus. The interleaving other half of the magnets is connected in the opposite way. Also shown are the heaters that are needed to distribute the stored energy in a magnet group evenly. The energy of the rest of the ring is bypassed by thyristors. They have to be mounted outside the cryostat and therefore current feed-throughs are needed. These require a very careful design since their electrical resistance (which is the main contribution to R_b) should be small. Their thermal resistance, on the other hand, should be large to avoid a heat load on the liquid helium system. During a quench the safety current leads heat up considerably which brings the connection points to the superconductor in danger of quenching also. In addition a fast recooling time is also an important design criterion. The development of high temperature superconductor current feed-throughs may alter the situation considerably.

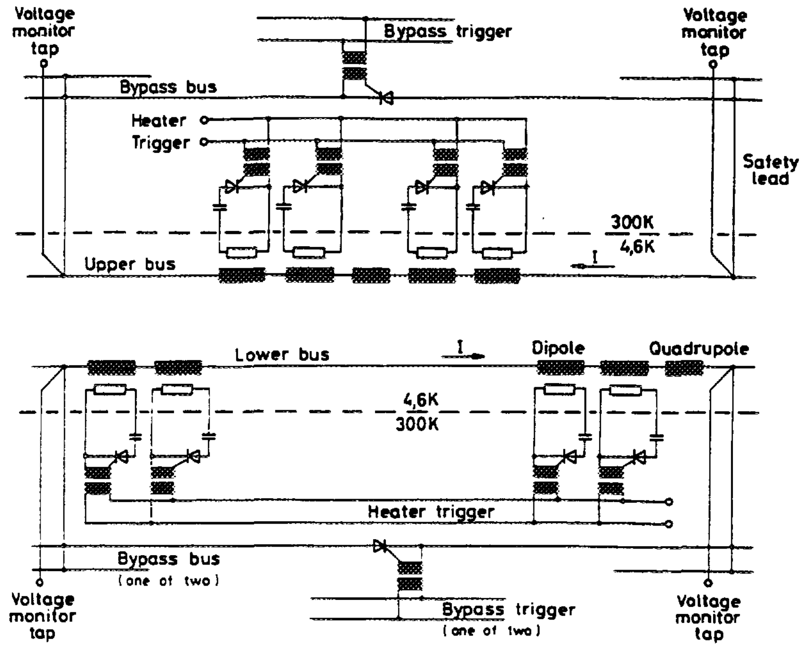


Fig. 17 The magnet connections in the TEVATRON. Dipoles and quadrupoles share two busses. Groups of magnets are protected against the total stored energy by bypass thyristors.

Diodes can replace the thyristors in storage rings that have a low ramp rate and hence small inductive voltages during the normal operation. Cottingham [28] first proposed for ISABELLE to mount diodes inside the liquid helium cryostat. This solution has several advantages. The bypass resistance is normally much smaller. Each magnet can have its own bypass diode. There is no heat load on the cryogenic system due to the safety leads. Finally, the cryostats are easier to build and cheaper if the current feed-throughs are missing. This concept has since successfully been adopted or is proposed for HERA, RHIC and LHC.

The bypass diode has to be selected carefully. Firstly, the diode should have a low dynamic resistance and this should not change as a result of aging or neutron bombardment. Secondly, the proper backward voltage has to be selected carefully. High voltage diodes have a thick p-n junction that is susceptible to radiation damage and which has a large resistance. On the other hand, if magnets in a string quench, not all will quench simultaneously. The inductive voltage that develops during the ramp down will concentrate on the still superconducting magnets and may exceed the backward voltage of the diode. Note that the backward voltage depends on the operating temperature. The extraordinary high current in the LHC magnets and the expected level of radiation pose severe constraints on the cold diodes. Nevertheless promising solutions have been found [29].

Relatively conservative magnets, as the HERA and RHIC magnets, do not in principle need artificially quench spreading, in particular if every magnet has its own bypass. Quench heaters need some energy storage, some firing electronics, and feed-throughs into the cryostat. The heater band has to be in close thermal contact with the coil, because the heat must reach the coil as fast as possible. In fact, the best place is in-between the two coil layers [30]. This is of course hazardous. Good heat conduction means little electrical insulation and hence the risk of shorts to the coil. In summary, quench heaters are costly and a potential danger themselves. However, for safety against quenches in the coil heads, even the HERA dipole magnets are equipped with heater strips, very much as in the TEVATRON magnets.

For the LHC magnets to avoid excessive energy densities artificial quench spreading is essential.

Finally, it is necessary to subdivide the machine into as many independent current circuits as feasible. This can be achieved in two ways. At HERA, all magnets are fed by one power supply. This results in good tracking of bending power and focal strength. In total 10 mechanical switches break the circuit in case of a quench into nine pieces separated by resistors, in fact, just steel pipes. The resistors are matched to the inductance such that the centres of the resistors and of the nine magnet strings are virtually at ground potential. Hence the ring virtually breaks up into nine independent sub-circuits. If a switch fails to open, the symmetry is broken; therefore the installation of an additional equalising line is necessary. The solution for RHIC, shown in Fig. 18, is similar. One twin power supply feeds one of the two rings. The subdivision follows of course the geometry of the tunnel and the switches consist of thyristors.

Very large rings contain so much energy that the virtual subdivision is not safe enough. Figure 19 shows one of the proposed solutions for the LHC [31]. The accelerator is divided into 16 independent units each of which stores more than HERA. An equal number of circuit breakers and dump resistors are required. The fact that each half octant can be switched off independently is clearly an advantage. A further division is unfortunately excluded from the geography of the LEP tunnel. On the contrary, cost optimisation may lead to a solution employing only one power supply per octant.

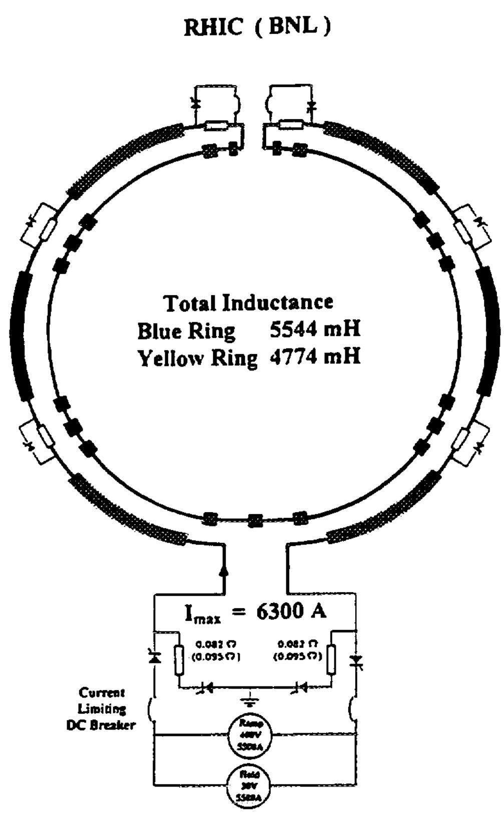


Fig. 18 The proposed connection scheme for one of the RHIC magnet rings. The second ring is almost identically to the one shown.

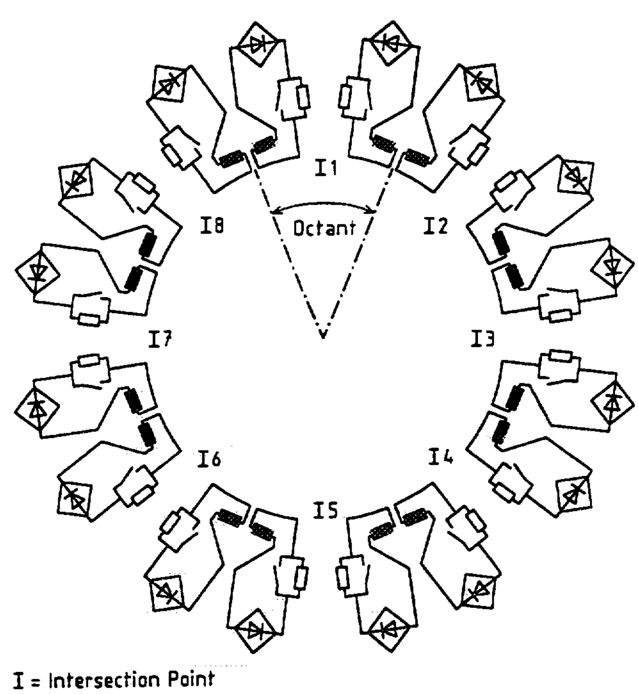


Fig. 19 The proposed subdivision of the LHC magnet string. The number of power supplies may be changed.

7. SUMMARY

Effective quench protection cannot be added afterwards. It is an integral part of the magnet and system design from the very beginning. The costs for cumbersome and complicated protection should be weighed against the cost of more stabilisation copper. Moreover aspects of reliability play also an important role. Quenches can and will always happen. Hence the quench protection has to be reliable and fail safe. Heaters should be added if necessary, and the overall layout of the power circuit has to be planned and simulated carefully. Magnetic coupling can add to the problems but it can also be turned into an advantage by spreading quenches over large volumes. A low energy density in a quenching magnet has to be tolerated and quench protection is basically the combination of measures to keep the energy density low in case of a quench.

* * *

REFERENCES

[1] M.S. Lubell, IEEE Trans. Magn., Vol 19, 1983, p754ff

[2] M. Wilson, Superconducting Magnets, Oxford Science Publication, Clarendon Press, Oxford

[3] A. Devred, Quench Origins, The physics of particle accelerators, Vol 2, M. Month et al (ed), 1992, American Institute of Physics

[4] H. Maeda et al, Cryogenics, June 1982, p287ff

[5] S. Wipf, Stability and Degradation of Superconducting Current Carrying Devices, Los Alamos 7275, 1978

[6] Maddock et al., Cryogenics 9 (1969), p261ff

[7] K.-Y. Ng, Minimum Propagating Zone of the SSC Superconducting Dipole Cable, SSC-180, 1988

[8] D. Bonmann et al, Investigations on Heater Induced Quenches in a Superconducting Test Dipole Coil for the HERA Proton Accelerator, DESY-HERA 87-13 (1987)

[9] H. Nomuraet al, Cryogenics, 1980, p283ff

[10] J. Chikaba et al, Cryogenics 1990, Vol 30, p649ff

[11] D. Leroy et al, IEEE Trans. Appl Sup Vol 3, pp 781ff,1993

[12] A. Siemko et al, Quench Localization in the Superconducting Model Magnets for the LHC by means of Pick-up Coils, to be published in IEEE Trans. Mag.

[13] T. Ogitsu et al, IEEE Trans. Mag Vol 30, No 4, p2773ff, 1993

[14] T. Ogitsu, Influence of Cable Eddy Currents on the Magnetic Field of Superconducting Particle Accelerator Magnets, SSCL-N-848, 1994 and Thesis (Institute of Applied Physics, University Tsukuba, Japan, 1994)

[15] W.H. Cherry, G.I. Gittelman, Solid State Electronics 1 [1960], p287ff

[16] Buznikov et al, Cryogenics, Vol 34, 1994, p761ff

[17] M.N. Wilson, Computer Simulation of the Quenching of a Superconducting Magnet, Rutherford High Energy Laboratory, RHEL/M 151, 1968

[18] K. Koepke, TMAX program, Fermilab, Batavia, Illinois, USA

[19] U. Otterpohl, Untersuchungen zum Quenchverhalten supraleitender Magnete, DESY-HERA 84/05 and Diploma Thesis, University Hamburg

[20] T. Tominaka et al., IEEE Trans. Mag Vol 28, 1992, p727ff

[21] Pissanetzky et al., Full Featured Implementation of Quench Simulation in Superconducting Magnets, Cryogenics, Vol 34,

[22] D. Hagedorn et al, Modelling of the Quenching Process in Complex Superconducting Magnet Systems, 12th International Conference on Magnet Technology, Leningrad 1991, LHC Note 159

[23] L. Bottura et al, Cryogenics, Vol 32, 1992, p 659ff

[24] L. Bottura et al, Cryogenics, Vol 32, 1992, p 719ff

[25] M.A. Hilal et al., IEEE Trans. Appl. Superconductivity, Vol 4, N03, 1994, p109ff

[26] P.F. Smith, The Review of Scientific Instr., 1963, Vol 34, p368ff

[27] I.M. Bolotin et al, The Quench Detector on Magnetic Modulator or the UNK Quench Protection System, Supercollider 4, J. Nonte ed., Plenum Press, NY, 1992

[28] J.G. Cottingham, Magnet Fault Protection, BNL-16816 (1971)

[29] D. Hagedorn, W. Nägele, Quench Protection Diodes for the Large Hadron Collider LHC at CERN, 1991 Cryogenic Engineering Conference, Huntsville, Alabama, LHC Note 148

[30] G. Ganetis, A. Stevens, Results of Quench Protection Experiment on DM1-031, SSC technical Note No 12, Brookhaven National Laboratory

[31] L. Coull et al, LHC Magnet Quench Protection System, 13th Intern. Conference on Magnet Technology (MT13), Victoria, Canada, 1993, LHC Note 251

BIBLIOGRAPHY

M. Wilson, Superconducting Magnets, Oxford Science Publication, Clarendon Press, Oxford.

NEXT PAGE(S) left BLANK

SUPERCONDUCTING CAVITIES — BASICS

W. Weingarten
CERN, Geneva, Switzerland

Abstract
The basic properties and design criteria of superconducting cavities for electron accelerators are discussed with special emphasis on the following topics: technical motivation for the use of superconducting cavities; design and production; surface resistance; the critical field of superconductors; anomalous losses; materials other than niobium; technological achievements for accelerating cavities.

1. INTRODUCTION

High-energy physics has prospered for the last few decades thanks largely to progress in particle accelerator techniques. The quest was always for larger centre-of-mass energies and larger luminosities. The most recent accomplishment is the operation at centre-of-mass energies between 91 GeV and 2 TeV of LEP at CERN, SLC at SLAC, HERA at DESY and the Tevatron at Fermilab.

An increasingly demanding problem is how to reduce investment and operation costs of such large accelerators. For this reason the benefits of superconductivity are exploited for magnets (at Tevatron and HERA) and RF cavities (at LEP and HERA). To combine larger energies with affordable costs is different for proton and electron accelerators. Circular accelerators offer a cost-effective way to obtain high energies because the particles traverse the accelerating sections many times. The energy limit is determined by the emission of synchrotron radiation, which increases with the fourth power of γ, the ratio of the particle energy over its rest mass. Accelerators for protons, a relatively massive particle, are still far away from this limit. Therefore future proton accelerators such as the LHC at CERN will be circular machines. On the other hand, circular accelerators for electrons, relatively light particles, emit large amounts of synchrotron radiation. An estimation shows [1] that future accelerators for electrons will be linear colliders like the SLC at Stanford, the pioneering one. The reason is that the investment costs are proportional to the length (for linear colliders) or the circumference (for circular colliders). For a beam energy E and an accelerating gradient g, they are proportional to E/g and to $E^2/\sqrt{g}$, respectively. Hence, above a certain energy a linear collider is cheaper. Since particles will traverse the accelerating section only once, it is essential to have high accelerating gradients.

The other demanding need is to keep the mains power as low as possible for a given luminosity and particle energy; in other words, the efficiency of transforming mains power to beam power should be a maximum. In principle, radio-frequency (RF) super-conductivity is supposed to provide a technical solution for both constraints, that is a high accelerating gradient and high mains-to-beam power conversion efficiency.

2. TECHNICAL MOTIVATION FOR THE USE OF SUPERCONDUCTING CAVITIES

2.1 The mains-to-beam-power conversion efficiency

Electrical power from the mains to the beam is transferred via the following chain: high-voltage power supply — klystron — waveguide — RF cavity — beam. For the time being the largest power loss has to be tolerated between the RF cavity and the beam. In order to establish the accelerating voltage V in the cavity, RF currents are induced at its surface generating the heat P_c. A shunt impedance $R = V^2/(2P_c)$ is defined to account for these RF losses. R should be as large as possible. RF accelerating cavities manufactured from high-conductivity copper have shunt impedances (per unit length) of typically 10–15 MΩ/m. With the beam current I_b (d.c. component), the beam power $P_b = VI_b$ (for the particle riding on the crest of the RF wave), the conversion efficiency is defined as

$$\eta = \frac{P_b}{P_b + P_c} = \frac{1}{1+\frac{P_c}{P_b}} = \frac{1}{1+\frac{V}{2RI_b}} . \tag{1}$$

For the LEP RF cavity (copper) with R = 43 MΩ, I_b = 6 mA, V = 3 MV we get $\eta \approx 15\%$.

As can be inferred from Eq. (1), normal-conducting (nc) cavities made from copper offer high conversion efficiencies for high beam currents and low accelerating voltages. The shunt impedance R of an RF cavity is related to its Q value by $R = (R/Q)\ Q$. (R/Q) is independent of the RF frequency and is determined by the geometry of the cavity. The Q value for its part is inversely proportional to the surface resistance R_s, $Q = G\ R_s^{-1}$, with G another geometrical constant (of typically 250 Ω). Hence, in order to have large shunt impedances, we are faced with the development of RF cavities with surfaces of low surface resistance.

2.2 The cryogenic efficiency

It is the large 'improvement factor' of the shunt impedance (~ 10^5) compared with conventional copper cavities that makes superconducting (sc) cavities so attractive. However, the dissipated power $Q_2 = P_c$ has to be removed at cryogenic temperatures, in common practice at the boiling temperature under normal pressure of liquid helium, T_2 = 4.2 K (Fig. 1). Hence, the entropy current $S = Q_2/T_2$ which, according to the second law of thermodynamics, in the ideal (reversible) case is Q_1/T_1 at room temperature $T_1 \approx 300$ K.

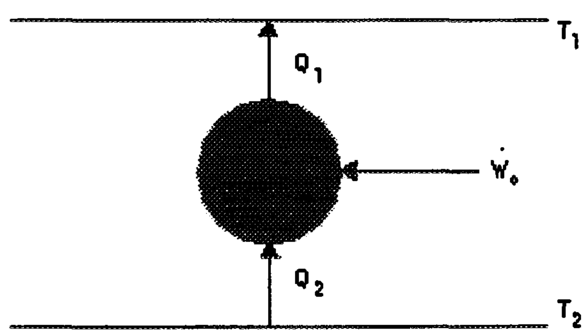


Fig. 1 Schematic power flow in a refrigerator

According to the first law of thermodynamics, the power

$$\dot{W}_c = Q_1 - Q_2 = Q_2(Q_1/Q_2 - 1) = Q_2(T_1/T_2 - 1) \tag{2}$$

has to flow into the compressor. The Carnot efficiency η_c for a refrigerator is defined as

$$\eta_c = Q_2 / \dot{W}_c = T_2 / (T_1 - T_2) . \tag{3}$$

For T_1 = 300 K and T_2 = 4.2 K, η_c = 1/70. The 'thermodynamic efficiency'

$$\eta_{td} = \dot{W}_c / \dot{W} \tag{4}$$

is the ratio of the power W_c needed to operate the compressor in the ideal case, to the 'real' power $\dot{W}$. The total cryogenic efficiency is

$$\eta_{cr} = Q_2 / \dot{W} = (Q_2 / \dot{W}_c)(\dot{W}_c / \dot{W}) = \eta_c \eta_{td} \,. \tag{5}$$

With $\eta_{td} \approx 0.3$ for large units the total cryogenic efficiency is $\eta_{cr} = 4.5 \times 10^{-3}$. Unavoidably, in an sc accelerator some power P_{cr} flows into the liquid He, even in the absence of RF (standby heat load of cryostat). The efficiency η for a sc accelerator of RF-to-beam power conversion is then

$$\eta = \left[1 + (P_c + P_{cr}) / (P_b \eta_{cr})\right]^{-1} . \tag{6}$$

As an example, for the sc cavity and cryostat for LEP with $P_c = 50$ W, $P_b = 50$ kW and $P_{cr} = 25$ W, we obtain a total efficiency of $\eta = 0.75$, which is larger by a factor of 5 than for a conventional RF system (Table 1).

Table 1

RF-to-beam power conversion efficiency η for nc and sc cavities for the e^+e^- collider LEP

	nc	sc	
Beam current	6	6	mA
Accelerating voltage	3	8.5	MV
Field gradient	1.4	5	MV/m
Shunt impedance R	43	3150 [1]	MΩ
P_c	105000	52 @ 4.2 K	W
P_b	18000	51000	W
P_{cr}	0	25	W
η_{cr}	100	0.45	%
η_{RF}	15	75	%

[1] including the total cryogenic efficiency η_{cr}

2.3 Lower impedance of sc cavities

Normal-conducting cavities are optimized for a high shunt impedance. The dissipated RF power to be removed gives the performance limit, and issues of low surface electric and magnetic fields are of minor importance. Therefore these cavities are designed as shown schematically in Fig. 2 (top).

Superconducting cavities, on the contrary, are designed for low surface electric fields and other features (e.g. to avoid electron multipacting or to efficiently extract the power which the beam transfers into the higher-order modes). Issues of high shunt impedance are of minor importance, because it is sufficiently large thanks to the sc material. A typical value is 1 GΩ/m (including the total cryogenic efficiency η_{cr}). Hence larger iris openings and a rounded shape are used [Fig. 2 (bottom)].

As sc cavities have a larger field gradient (about a factor 4) and a larger iris diameter, they present more than a factor 10 lower impedance (for the same accelerating voltage) to the beam compared to nc cavities.

The rise time of beam instabilities is inversely proportional to the product of impedance and current. The beam remains stable if the rise time is longer than the damping time of synchrotron and betatron oscillations. There is a minimum tolerable rise time, and, consequently, the maximum current is inversely proportional to the impedance. Hence the advantage of sc cavities over nc cavities.

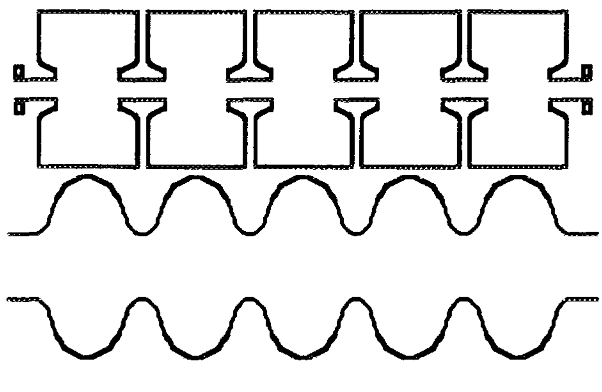


Fig. 2 Cavity shape (schematic): the nc cavity (top) represents a larger impedance to the beam due to the smaller beam holes and nose cones compared with the sc one (bottom).

3. DESIGN AND PRODUCTION

3.1 The choice of gradient and cost limits

The maximum gradient is not necessarily that for which the construction cost is a minimum. This also depends on the type of machine (accelerator tunnel layout, luminosity, beam energy etc.). However the larger the accelerator the more important become economical constraints.

If the total RF voltage, V, is fixed, the total costs are given by three terms. The first one is proportional to the total length L of the cavities (or cryostats), the second one is proportional to the total RF power P_{RF} to be transmitted to the beam (independent of L), and the third one to the cooling power (the power P_d dissipated in the RF cavities), hence inversely proportional to the shunt impedance or total length L. The cost minimum is located where the first and third term are equal. The following example is based on numbers, which are close to CERN's experience [2]. Figure 3 displays these three terms and their dependence on the accelerating gradient E_a (which is for constant $V = E_a \cdot L$ inversely proportional to L). It shows that a gradient of about 10 MV/m is an optimum choice for this particular application.

However, having in mind even larger applications, such as the sc linear collider TESLA under design, at higher frequency (1.3 GHz) and higher energy (500 GeV), costs will increase by at least an order of magnitude compared to the largest existing accelerators. There are two major contributions (which depend on E_a), linear investment and cooling power costs. The latter may be reduced (technical Q-values being in the 10^9 range) by pulsing the accelerator. The former has to be reduced by novel and more economic industrial production methods. At CERN, coating the copper substrate cavity with thin sc layers of Nb [3], which can be hydroformed [4] may be a step in this direction.

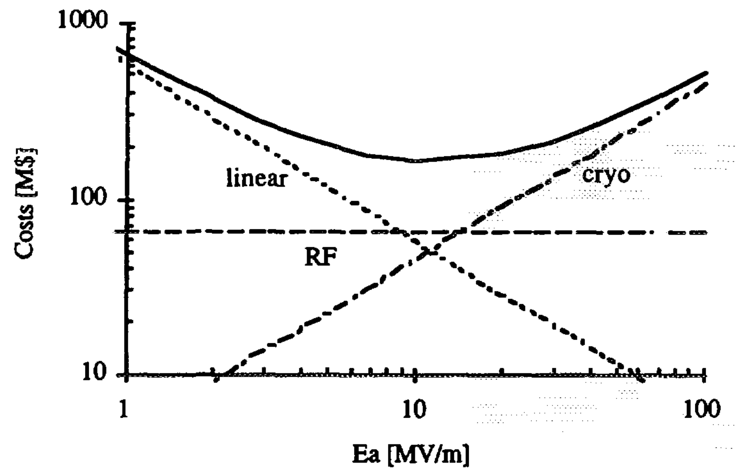


Fig. 3 Cost estimate for an sc RF system (f = 352 MHz, V = 2 GV, 12 MW RF power, 5 W/m standby heat load, $Q_0 = 4\cdot 10^9$).

3.2 The choice of the operating frequency

Without loss of generality, R per length L (which is called r) is required to be maximized:

$$\begin{aligned} \frac{R}{L} = r &= \frac{V^2}{2P_c L} = \frac{E_a^2 L^2}{2\omega U L}\frac{\omega U}{P_c} \\ &= \frac{E_a^2}{2\omega U/L}\frac{\omega U}{P_c} = \frac{r}{Q}Q = \frac{r}{Q}\frac{G}{R_s}, \\ &= \frac{r}{Q}\frac{G}{R_s^{BCS} + R_{s0}} \end{aligned} \tag{7}$$

where we have made use of the definition of $r/Q = E_a^2/(2\omega U/L)$, $Q = \omega U/P_c$, $V = E_a L$, $R_s Q = G$, the geometry factor. U is the stored energy, E_a the accelerating gradient, R_s^{BCS} the BCS part of the surface resistance, and R_{s0} its residual part (cf. sections 4.1 and 6.1). Now we shall look at the dependence on the frequency f of the different factors. From Fig. 8 (see below) we learn that for niobium the condition $R_s^{BCS} \gg R_{s0}$ is valid for frequencies larger than 3 GHz, and the condition $R_s^{BCS} \ll R_{s0}$ is valid for frequencies smaller than 300 MHz:

$$\underbrace{(r/Q)}_{\propto f}\frac{G}{\underbrace{R_s^{BCS}}_{\propto f^2} + R_{res}} \propto \begin{cases} f^{-1}, \; R_s^{BCS} \gg R_{res}, \; f > 3\ \text{GHz} \\ f, \; R_s^{BCS} \ll R_{res}, \; f < 300\ \text{MHz} \end{cases} \tag{8}$$

To maximize r, we are forced to prefer lower frequencies for $f \geq 3$ GHz, and higher frequencies for $f < 300$ MHz. Therefore, neither very low nor very high frequencies are useful, and $\beta = 1$ accelerating structures are operated between 300 MHz and 3 GHz, approximately.

3.3 Design criteria

The most critical part of an sc cavity is the surface. This comes from the fact that the current flows with nearly no losses. Any contamination on the surface will give rise to extra heat and will drastically increase the losses. Due to the low thermal capacity at low temperatures even a minute heat source may provoke a large increase in temperature. Superconductivity may be destroyed. Therefore the critical steps of the assembly are done in clean rooms of class 100 and lower (Fig. 4 [5]).

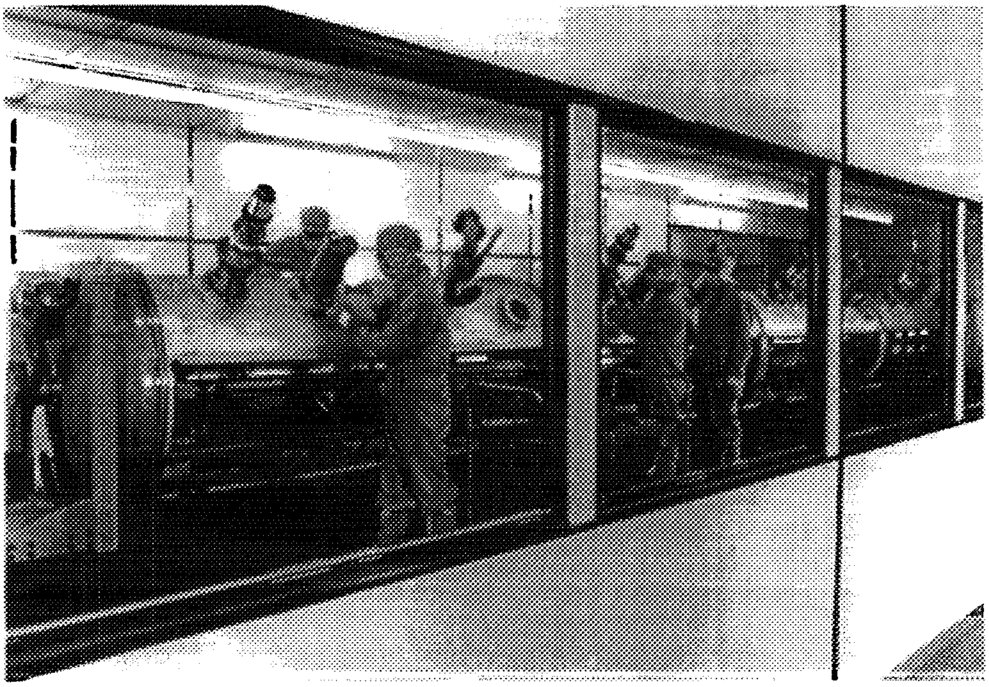
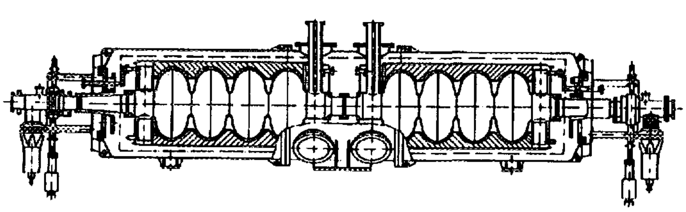


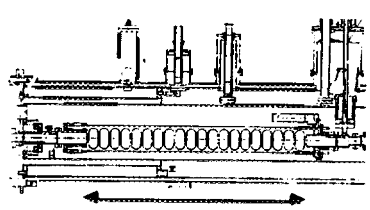
Fig. 4 Four-cavity module for LEP being equipped with HOM and power couplers in a class 100 clean room

Local heat sources may also be created by electron impact (multipacting or field emission). The former can be suppressed by the choice of the geometry ("spherical" cavities [6]). The latter may be reduced to tolerable levels by rinsing the surface with ultrapure agents (mostly water and methanol) and avoiding any dust precipitation during assembly. Other considerations are: economy of fabrication (spinning and deep drawing of sheet metal), the reduction of any heat leak (coaxial couplers), the efficient removal and dissipation at room temperature of higher-order mode (HOM) induced power, no or few mobile mechanical parts at cryogenic temperatures, and attachment of sensors at critical parts to probe the local vacuum, temperature and e^--current.

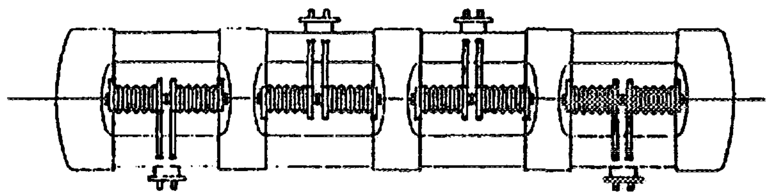
Nowadays the design for $\beta = 1$ cavities and cryostats is largely converging as shown in Figs. 5 and 6.



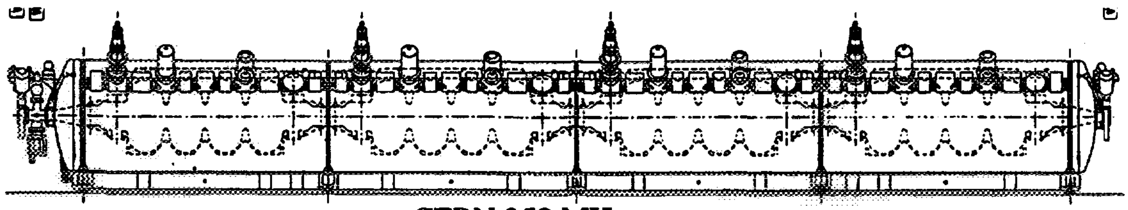
DESY 500 MHz



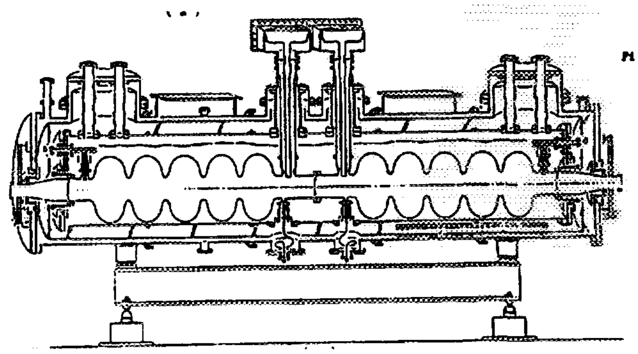
Darmstadt 3 Mhz



CEBAF 1500 MHz



CERN 352 MHz



KEK 508 MHz

Fig. 5 Cryostat and cavity design for $\beta = 1$ accelerating structures at different frequencies (approximately to scale)

The following is a summary of the design criteria and state of the art:

- Access must be possible to the cryostat's vital parts without either the vacuum having to be broken or the cryostat having to be dismounted from the beam line
- All ancillary equipment to be cleaned to the same standards as those of the cavity
- Rinsing to be done with clean and dust-free fluids
- Assembly to be done under dust-free conditions
- Spherical shape to avoid e^- multipacting with slightly inclined iris to facilitate the drainage of rinsing fluids
- Metal sheet forming
- Minimizing of mechanical vibrations
- Compact coaxial antenna type couplers for RF power input and HOM power damping
- RF windows — if possible — between beam and insulation tank vacuum to minimize consequences of break

- No holes in the cavity proper: couplers attached to the beam tube
- Number of cells limited: careful computer-aided designing of cavity geometry imperative
- Vacuum, e^- and temperature sensors available at critical regions of power coupler
- Frequency tuners without moving parts.

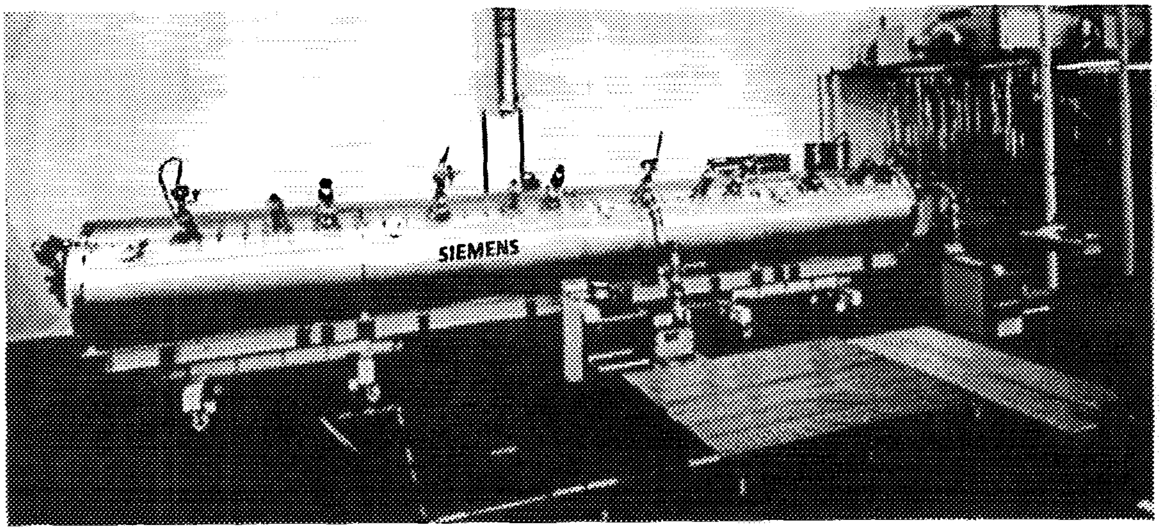


Fig. 6 Fully equipped four-cavity module for LEP.

4. SURFACE RESISTANCE

4.1 The surface impedance of superconductors

Imagine a plane wave propagating in the x-direction towards an sc metal sheet in the y-z plane. The metal is supposed to contain sc and nc electrons (two-fluid model). Only the nc component interacts with its environment; the sc component carries current without interaction. The RF magnetic field penetrates this sheet to within the penetration depth λ_L. According to the Maxwell equation curl $E = -\partial B/\partial t$, the RF magnetic field is accompanied by an electric field in the y plane $E_y = j\omega\lambda_L B_z = j\omega\lambda_L\mu_0 H_z = j\omega\lambda_L\mu_0 H_{z0}\exp(-x/\lambda_L)$. This interacts with the nc electrons (still present at non-zero temperatures) and gives rise to a power dissipation per square metre

$$P_c = (\sigma_n/2)\int_0^\infty E_y^2(x)dx = (1/4)\lambda_L\sigma_n E_{y0}^2 = (1/4)\omega^2\mu_0^2\lambda_L^3\sigma_n H_{z0}^2 \,. \tag{9}$$

where σ_n is the conductivity of the nc electrons.

By definition, $P_c = (1/2)\, R_s\, H_z{}^2$, and we obtain

$$R_s = (1/2)\omega^2\mu_0^2\lambda_L^3\sigma_n \,. \tag{10}$$

λ_L is the characteristic penetration depth of a static or microwave magnetic field in a superconductor such as niobium (London penetration depth $\lambda_L = [m/(n_s e^2\mu_0)]^{1/2}$). λ_L depends on the density n_s and mass m of the sc electrons. σ_n is proportional to the density of the nc electrons, which decreases with the temperature according to a Boltzmann law, with the energy gap Δ as the characteristic energy. The precise theory will result in an expression for the so-called BCS surface resistance (from the inventors of the microscopic theory of superconductivity, Bardeen, Cooper and Schrieffer). It can be approximated for $T < T_c/2$ (T_c is the critical temperature of the superconductor) as

$$R_s = (A/T)\omega^2 \exp(-\Delta / kT),\ 2\Delta \approx 3.5kT_c. \qquad (11)$$

The ω^2 dependence of the surface resistance is a consequence of the frequency-independent penetration depth [7]. Measurements of the temperature and frequency dependence of the BCS surface resistance are shown in Figs. 7 and 8 [8,9].

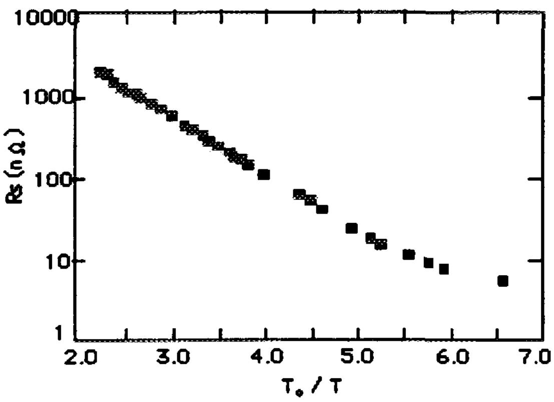


Fig. 7 Surface resistance of a 3-GHz niobium-sheet cavity (*RRR* = 40) vs. T_c/T. The residual resistance is 4 nΩ.

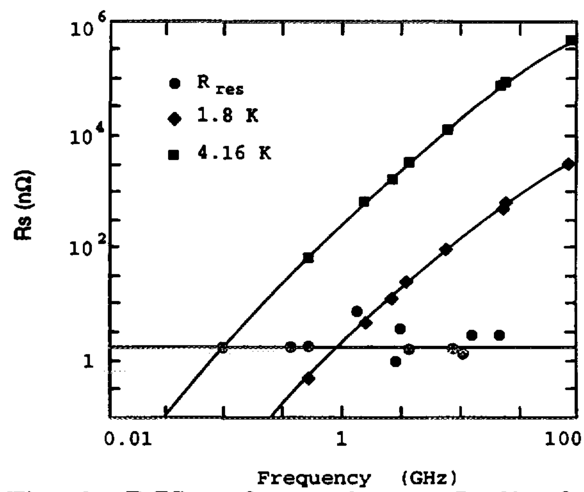


Fig. 8 BCS surface resistance R_s^{BCS} of niobium vs. frequency at 4.2 K (extrapolated to 1.8 K). Lowest residual surface resistance data are indicated at the bottom.

The surface impedance for the sc and the nc case is derived more formally in Table 2. For the former, the constituent equations are the London equations (3rd row), for the latter the constituent equation is Ohm's law. These go together with the Maxwell equations (2nd row), which are always valid. The surface impedance is a complex number; in the nc case the real part (surface resistance) is equal to the imaginary part (surface reactance), whereas in the sc case they are different. The surface resistance as indicated in Eq. 10 is reproduced, and the surface reactance is proportional to the penetration depth λ_L.

For the sake of completeness it should be mentioned that the surface resistance R_s depends through λ on the mean free path of the electrons ℓ [10] and hence the *RRR* value (the residual resistivity ratio *RRR* is the ratio of the resistivity at room temperature and low temperature in the nc state). For niobium, there is a minimum surface resistance at about *RRR* = 7.

5. THE CRITICAL FIELD OF SUPERCONDUCTORS

In d.c. the external magnetic field starts penetrating the sc material at the critical field B_c (type I) or B_{c1} (type II superconductors). For RF this is not true for reasons which are not completely understood. It is believed that the creation of nc islands in the sc metal needs a period of time which is larger than one RF period. The limiting field in RF superconductivity is therefore larger than both B_c and B_{c1}, and is called the critical superheating field B_{sh} [11].

Table 2

Comparison of superconductor (two-fluid model) with normal conductor

<table>
<tr><th>Superconducting metal</th><th>Normal-conducting metal</th></tr>
<tr><td colspan="2">$\mathrm{curl}\,\boldsymbol{H} = \boldsymbol{j} + \varepsilon_0 (\mathrm{d}/\mathrm{d}t)\boldsymbol{E}$
$\mathrm{curl}\,\boldsymbol{E} = -\mu_0 (\mathrm{d}/\mathrm{d}t)\boldsymbol{H}$
$\mathrm{div}\,\boldsymbol{H} = 0$
$\mathrm{div}\,\boldsymbol{E} = 0$</td></tr>
<tr><td>$(\mathrm{d}/\mathrm{d}t)\,\boldsymbol{j} = \boldsymbol{E}/(\mu_0 \lambda_L^2) + \sigma_n (\mathrm{d}/\mathrm{d}t)\boldsymbol{E}$
$\mathrm{curl}\,\boldsymbol{j} = -\boldsymbol{H}/\lambda_L^2 - \sigma_n \mu_0 (\mathrm{d}/\mathrm{d}t)\boldsymbol{H}$</td><td>$\boldsymbol{j} = \sigma_n \boldsymbol{E}$</td></tr>
<tr><td colspan="2">$\Delta\,\boldsymbol{E} = -K^2 \boldsymbol{E}$
$\Delta\,\boldsymbol{H} = -K^2 \boldsymbol{H}$</td></tr>
<tr><td>$K^2 = -\,\lambda_L^{-2}\,(1 + \mathrm{i}\sigma_n \mu_0 \omega \lambda_L^2 - \varepsilon_0 \mu_0 \omega^2 \lambda_L^2)$
$Z_{ssc} = (1/2)\omega^2 \mu_0^2 \lambda_L^3 \sigma_n - \omega\mu_0\lambda_L$,
$\varepsilon_0 \mu_0 \omega^2 \lambda_L^2 \ll \sigma_n \mu_0 \omega \lambda_L^2 \ll 1$.</td><td>$K^2 = \varepsilon_0 \mu_0 \omega^2 (1 - \mathrm{i}\sigma_n/(\omega\varepsilon_0))$
$Z_{snc} = (\omega\mu_0\delta/2)(1 - \mathrm{i}) =$
$\sqrt{[\omega\mu_0/(2\sigma_n)]}(1 - \mathrm{i}),\ \omega\varepsilon_0/\sigma_n \ll 1$.</td></tr>
</table>

Effective electron mass m

Electron mean free path l

Conductivity of the nc electrons $\sigma_n = l n_n e^2/(m v_F)$

London penetration depth $\lambda_L = \sqrt{[m/(n_s e^2 \mu_0)]}$

Surface impedance $Z_s = (E_z/H_y)|_{x=0} = \mu_0\omega/K$

Skin depth $\delta = \sqrt{[2/(\mu_0 \sigma_n \omega)]}$

Fermi velocity v_F

Density of nc electrons n_n

Density of sc electrons n_s

The relevance of B_{sh} can be understood as follows [12]. The entry of a vortex line into the bulk of a superconductor is governed by two forces (Fig. 9). The first one is due to the fact that the shielding currents of a line only have a component tangential to the surface (the normal component vanishes), $j_\perp = 0$. This boundary condition can be easily assured by the superposition of two lines, the original one and its mirror image with respect to the metal surface. The force between these two lines decays exponentially with twice the distance x_L between the line and the metal surface ($\sim\exp(-2x_L/\lambda)$). One can convince oneself of the attractive nature of this force and of its independence of the RF magnetic field amplitude B.

The second force is created by the line and the RF currents $\bar{j}$, which are proportional to B. This force is repulsive and decays exponentially with a characteristic length x_L [$\sim\exp(-x_L/\lambda)$]. The superposition of these two forces results in three distinct cases (Fig. 10):

(a) For B smaller than the lower critical magnetic field B_{c1}, $B < B_{c1}$, the attractive interaction between line and image line is dominant, thus preventing the entry of the line into the bulk.

(b) For $B_{c1}, < B < B_{sh}$, the attractive interaction beats the repulsive one for small x_L in such a way that a surface barrier still prevents the line from entering into the bulk.

(c) For large B, $B > B_{sh}$, this surface barrier has disappeared and the line enters the bulk.

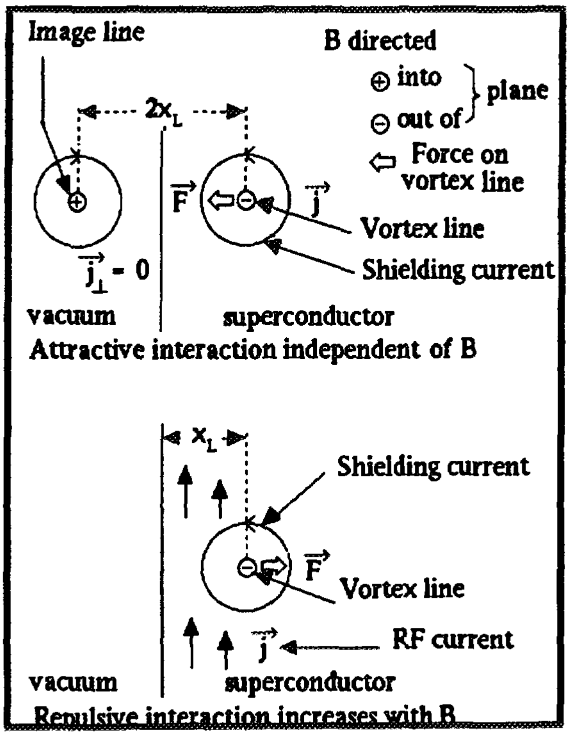


Fig. 9 Forces acting on a vortex line entering a superconductor: an attractive force between the line and its image (independent of the RF magnetic field amplitude B) opposes the entry of the vortex line into the bulk (upper part), a repulsive force between the line and the RF currents j (proportional to B) favours the entry of the vortex line into the bulk (lower part).

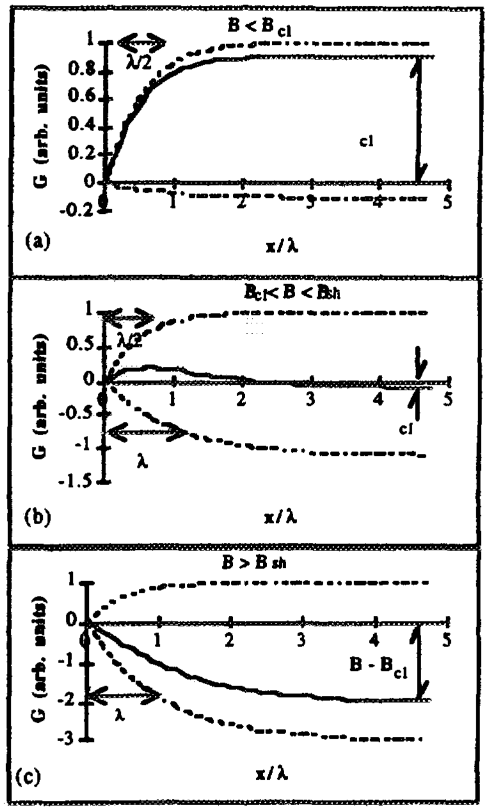
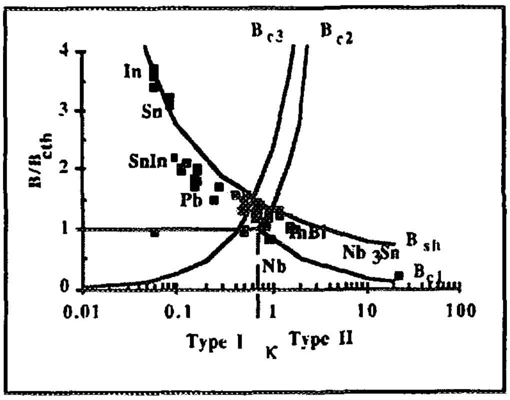


Fig. 10 The Gibbs potential G vs the distance from the metal surface x_L/λ of the vortex line. G is a superposition (full line) of an attractive and a repulsive interaction (dotted lines). These two forces have different characteristic lengths of exponential decay (indicated by the arrows). Depending on the RF magnetic field amplitude B, $G > 0$, the line cannot enter (a), G has a barrier for small x_L, the line still cannot enter (b), and $G < 0$, the line can enter (c).

A series of experiments on small samples of Sn-In and In-Bi alloys indicates that the limiting field in RF is indeed B_{sh} for type I and type II superconductors [13]. This seems to be confirmed by Fig. 11 where maximum magnetic fields experimentally obtained under RF conditions (normalized to B_{cth}) are compared with various critical fields (B_{c1}, B_{c2}, B_{c3}, B_{sh}) as a function of the Ginzburg-Landau parameter κ. All data are compatible with the assertion that the limiting field in RF is indeed B_{sh}.

If this is confirmed, then very high magnetic fields can, in principle, be hoped for in sc accelerating cavities. With typically 4 mT RF surface magnetic field per MV/m accelerating gradient, the ultimate limit for niobium is then 60 MV/m, and for Nb_3Sn 100 MV/m accelerating gradient.



Material	κ	B_{cth0} [mT]	B_{exp}/B_{cth}	Refs.
Sn	0.086	30.9	3.11	13
			3.24	13
In	0.06	29.3	3.41	13
			3.61	13
			3.72	13
			0.953	15
Pb	0.5	80.4	0.98	15
			1.38	14
			1.46	14
Nb	0.9	~ 200	0.815	16,17
			1.17	18
			1.96	18
Nb_3Sn	22.8	~ 535	0.199	19
SnIn	0.1- 1.0		2.2 - 0.85	13
InBi	0.16-1.74		1.84- 1.01	13

B_{cth} thermodynamic critical field
B_{exp} experimentally obtained maximum field in RF
The index '0' refers to the temperature $T = 0$ K.

Fig. 11 Critical fields of different nature (1st and 2nd critical field B_{c1}, B_{c2}, surface critical field B_{c3}, and superheating critical field B_{sh}, normalized to the thermodynamic critical field B_{cth}) vs. the Ginzburg-Landau parameter κ. Maximum magnetic fields obtained under RF are indicated as dots.

6. ANOMALOUS LOSSES

6.1 Phenomenology

As seen above, there should be no reason for accelerating fields of 60 MV/m in sc niobium cavities not to be obtained. By lowering the temperature of the liquid He bath sufficiently, the RF losses should be negligibly small. At 350 MHz and 2 K, as an illustration, the sc LEP accelerating cavities should have a Q value of 8×10^{11} and dissipate only 30 W at 60 MV/m accelerating gradient. Nevertheless, experimentally obtained best results in these cavities are $Q = 3 \times 10^9$ with about 200 W dissipated power at 10 MV/m accelerating gradient. Unfortunately, several phenomena, in general not directly linked to RF superconductivity, make this goal hard to achieve. Some typical performance limits are also visible in Fig. 12.

Diagnostic tools have been developed that allow the study of performance limits. Firstly, the response of the cavity to an RF pulse can now be analysed [20]. Secondly, from the Q vs. E curve, information on the status of the cavity can be drawn. Thirdly, a most powerful diagnostics device is the recording of the surface temperature of the cavity at high field by temperature sensors ("temperature mapping" [21,22], Fig. 13).

Performance limits go together with a specific mechanism of RF loss, which we will call "anomalous". These mechanisms are commonly observed, but they are atypical of a surface as described in the framework of the BCS theory, being free from impurities and contamination.

Residual loss is independent of the temperature. The physical origin is not well known in most cases; what could be identified is loss from dielectric materials and trapped magnetic flux.

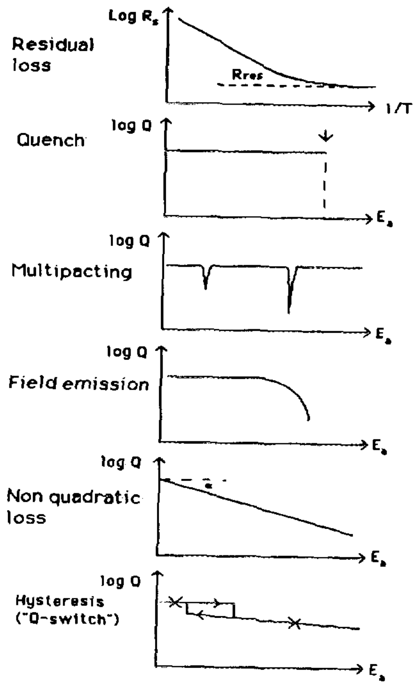


Fig. 12 $R_s\ (T_c/T)$ and $Q(E_a)$ curves that show typical performance limits

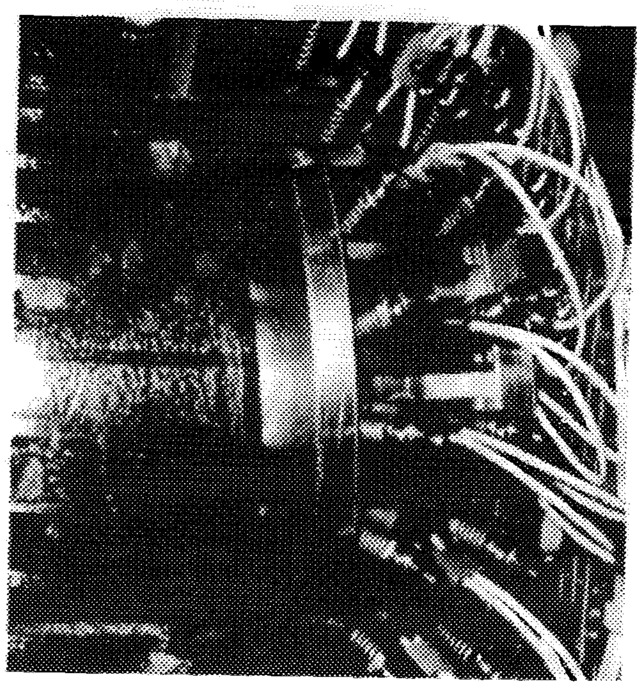


Fig. 13 Rotating-arm temperature sensors on 1.5-Ghz niobium-coated copper cavity

A *"quench"* is often observed in niobium cavities made of industrial-grade niobium without post-purification. A quench is caused by nc inclusions and other imperfections (welding beads, residues from the cleaning treatment). These inclusions are heated in the RF magnetic field and drive the sc metal near by into the nc state. Then an instability shows up: the stored energy of the cavity is dumped in this nc region, by this action the RF field in the cavity is abruptly decreased. At low RF field, the RF dissipation is again low enough for the niobium to go back into the sc state. As the RF coupling has been adjusted for this situation, the cavity again fills with RF energy until the cycle starts as described above ('self-pulsing breakdown').

Electron multipacting (from *multi*ple im*pact*) is an electron multiplication phenomenon which can happen if the secondary electron emission coefficient is larger than 1. As the electrons are resonant with the RF field amplitude, multipacting is visible at distinct field levels. If the electron trajectories are restricted to a point of the surface (by electron kinematics in the RF field), the currents can be sufficiently high to dissipate enough power to induce a breakdown in spite of the low impact energy (about 50–1500 eV). Properly cleaned niobium cavities (low δ) of rounded shape have been shown to be virtually free of multipacting [6].

Field emission is another loss due to electron impact. The current is generated by the large electric surface field at point-like emitting sites (e.g. dust particles). Contrary to electron multipacting, the current is relatively small (~ nA), but the impact energy is large (~ MeV). The current increases exponentially with the RF surface field amplitude, so that the gradient is limited even when the RF power is increased. Here also, breakdown may occur. In addition, heat and large amounts of X-radiation are generated by bremsstrahlung of the electrons when stopped at the cavity wall.

Non-quadratic loss manifests itself by a non-horizontal *Q vs. E* curve. It is often observed in cavities coated with niobium or other materials (NbN, NbTiN, YBaCuO). It is hypothesized that the nc surface increases with the RF field amplitude (presumably at "weak links").

Hysteretical Q vs. E curves are observed for sc material in loose contact with the substrate (welding beads in niobium sheet-metal cavities or peeled-off niobium flakes in niobium sputter-coated cavities). When the RF field amplitude is increased, these irregularities switch to the nc state at a certain RF magnetic field amplitude, upon which the Q-value abruptly goes down by a small amount, which corresponds to the tiny surface of the flakes. When the RF field amplitude is decreased, they remain in the nc state until the RF dissipation becomes so small that the cooling is sufficient to put them back into the sc sate.

6.2 Thermal breakdown

Whereas in X-band cavities of relatively small surface exposed to the RF field (some square centimetres), high surface magnetic fields were obtained more than two decades ago, this is not the case in larger cavities for accelerator application at lower frequencies. It has long been suspected that small nc surface defects with much higher RF losses, relatively widely dispersed, were the origin. If so, by statistical arguments, the probability of obtaining high surface fields was much larger in small single-cell cavities of higher frequency. Furthermore, the number of tests performed with small cavities at high frequency was much larger, which made a test result with a very high surface field more probable.

In some experiments at CERN, "hot spots" of enhanced RF loss causing breakdown were detected and localized by temperature mapping, cut out and inspected under a scanning electron microscope [23]. It turned out that, for accelerating fields less than ~ 8 MV/m (on commercial grade niobium in 1980), a surface defect which had clearly induced the thermal breakdown could be attributed to every "hot spot". They were welding "beads", welding holes, chemical residues and nc inclusions of a diameter of typically 100 µm.

Why do these defects cause a breakdown?

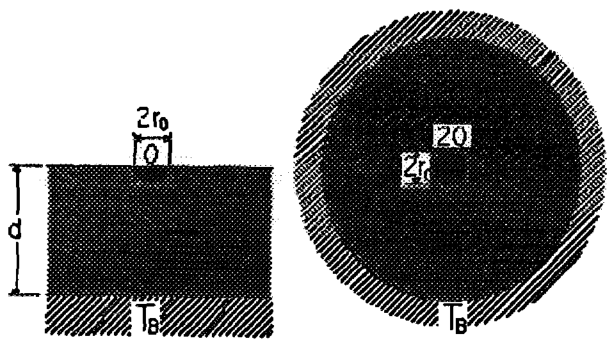


Fig. 14 Modelling the thermal breakdown (quench) caused by an nc defect, producing heat at a rate Q.

Let us examine a defect, represented by a half-sphere with radius r_0 of nc metal exposed to the RF field. It is embedded in an sc metal of wall thickness d and thermal conductivity (4.2 K) λ [Fig. 14 (left)], cooled by liquid He at temperature T_B. The defect represents a heat source of strength Q [W]. Under certain assumptions ($r_0 << d$, no temperature drop across the metal–liquid interface), the problem is equivalent to a circular symmetric one with a heat source of twice the strength Q, for which the heat conduction equation

$$\Delta T \approx 2(Q/\lambda)\delta(r) \tag{12}$$

is solved by

$$T \approx Q/(2\pi r) + T_B . \tag{13}$$

The heat source strength is

$$Q \approx (1/2) R_s^n H^2 \pi r_0^2 , \tag{14}$$

where $R_s{}^n$ is the surface resistance of the nc metal and H the local magnetic surface field. Hence

$$T \approx (1/4) R_s^n H^2 r_0^2 / (\lambda r) + T_B . \tag{15}$$

Breakdown will occur when the temperature at the nc–sc interface ($r = r_0$) approaches T_c. This gives a condition for the maximum RF field H_{max} [24,25],

$$H_{max} \approx \left[4(T_c - T_B)\lambda / \left(R_s^n r_0\right)\right]^{1/2} . \tag{16}$$

With typical values λ = 10 W/(mK), T_B = 1.8 K, $R_s{}^n$ = 12 mΩ, $r_0 = 10^{-4}$ m, we obtain $B_{max} = \mu_0 H_{max}$ = 20 mT, which corresponds to 5 MV/m accelerating gradient, in qualitative agreement with the experimental results obtained a decade ago with reactor-grade niobium. Computer simulations have shown [26] that for a defect radius r_0 > 20 μm, $\lambda \le 75$ W/(mK), Eq. (16) is correct to within 20%.

It becomes immediately clear that, to increase the maximum surface field H_{max}, the number and size of the surface defects have to be reduced, and, on the other hand, sc metal of increased thermal conductivity must be made available ('thermal stabilization' of defects). It was found [25] that the thermal conductivity of niobium is most severely affected by interstitial impurities such as oxygen, nitrogen and carbon. These impurities could be considerably reduced by repeated furnace treatments and/or solid-state gettering by wrapping a foil of metal with a large affinity to these impurities around the cavity in a furnace at 1200–1400 °C. Yttrium and titanium were used as getter material [28,29].

Another way has been chosen which is to coat high-thermal-conductivity material such as OFHC copper with a thin sc film of lead [28] or niobium [3] of ~ 1–2 μm, which is thick enough by far to carry the supercurrent. As an example, LEP-type cavities sputter-coated with niobium exhibit maximum accelerating fields and Q values comparable with niobium sheet metal or even better. Although occasional problems occur due to poor adhesion of the niobium film to the copper substrate at contamination at the copper/niobium interface, these cavities are not limited by a thermal breakdown. In addition, they show almost no extra RF loss due to trapped magnetic flux up to at least more than twice the Earth's magnetic field, significantly alleviating the task of obtaining a low residual surface resistance.

6.3 Electron field emission

Since the thermal stabilization of defects has been achieved, thermal breakdown of sc cavities has become less frequent, so the maximum fields obtained are significantly increased. Other breakdown phenomena, typical of higher electric fields, such as electron emission, become more important.

Field emission of electrons from metallic surfaces is determined by the work function and, according to the Fowler–Nordheim law, should occur only at very large electric fields (~ GV/m). However, from temperature mapping (cf. Section 6.1) of sc cavities it became clear that field emission originates from isolated sites at surface fields which are lower by a factor β, that is around 10 MV/m (Fig. 15 [21, 29, 30]). The fit parameter β is called the 'field enhancement factor' and amounts to the order of several hundred. The physical explanation for this phenomenological factor is not understood. DC field emission experiments have shown that small (< 1 μm) isolated sites on niobium surfaces emit electrons [31]. There is no evidence to believe that field emission observed in sc RF cavities is of a different nature.

Field emission electron loading in RF cavities can be understood on a statistical basis [30]. The emitting sites are randomly distributed on the surface of the cavity with an average spatial density depending on the preparation and treatment. The intrinsic features of the site (described by their β-value) are also statistically distributed: large β-values are less

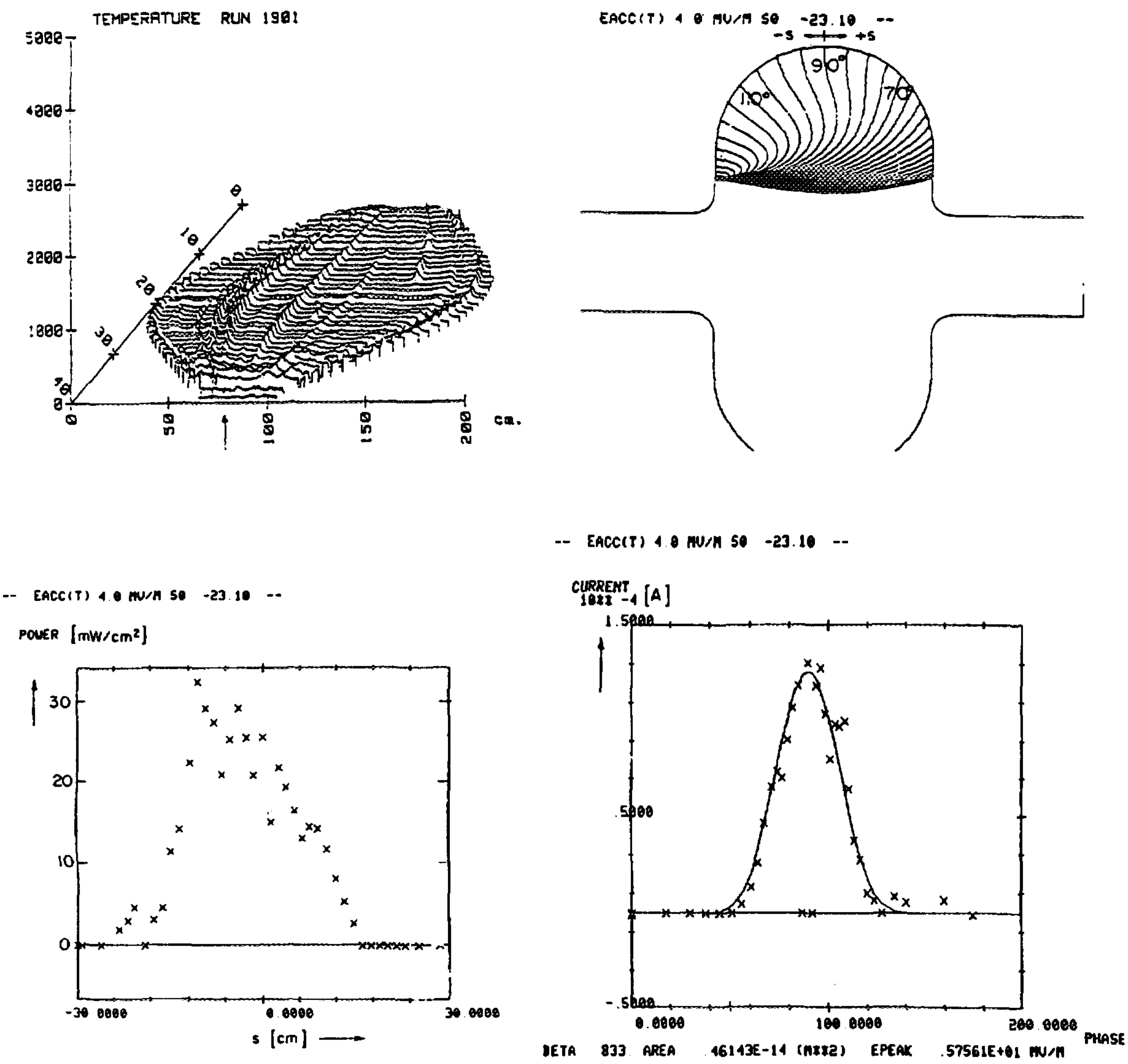


Fig. 15 Diagnosing field emission in a 500-MHz monocell cavity (top left: temperature map showing lines of temperature signals from electron impact; right: simulating electron kinematics in the RF field; bottom left: deposited power density along line of electron impact; right: the instantaneous electron current within one RF half period (180°) with the fit parameter β = 833 and the emitting area $A = 0.46 \times 10^{-14}$ m^2.

probable than small ones. Large β-values mean that electron emission will start at low surface electric fields. To find an emitter with a large β-value is more probable in a cavity with a large surface (low frequency) than in a cavity with a smaller one (high frequency). Hence it is more difficult to obtain large surface electric fields in a low-frequency cavity.

In accordance with the preceding ideas very large surface electric fields (145 MV/m) have been obtained in a niobium cavity in which the electric fields were concentrated into a small area [32]. This observation indicates a lower limit for the maximum electric surface field which could be obtained in a perfect niobium cavity.

Considerable progress in reducing the density of emission sites has been made in recent years by heat treatment of cavities above 1400°C [31], by operating the cavity in pulsed mode with very high RF power [33], and by improving the rinsing methods (rinsing under high pressure) [34].

7. MATERIALS OTHER THAN NIOBIUM

It is evident that superconductors with a larger pairing energy Δ (higher T_c), should allow a smaller surface resistance at the same temperature, provided A (Eq. 11) is about the same. This is the reason why both the conventional and new high-T_c superconductors such as Nb_3Sn and YBaCuO are also being investigated for RF applications. The surface resistance of commercially available niobium at 4.2 K is given by

$$R_s^{BCS}/n\Omega = 105\ (f/\mathrm{GHz})^2 \exp(-18/(T/\mathrm{K}))/(T/\mathrm{K}).$$

For Nb_3Sn-coated cavities ($T_c = 18.1$ K) we have

$$R_s^{BCS}/n\Omega = 105\ (f/\mathrm{GHz})^2 \exp\ (-40/(T/\mathrm{K}))/(T/\mathrm{K}).$$

For comparison, the surface resistance for copper at room temperature is

$$R_s/m\Omega = 7.8\ (f/\mathrm{Ghz})^{1/2},$$

10^5 times larger than for niobium at 4.2 K and 350 MHz. Consequently, the shunt impedance is larger by the same factor for sc cavities.

As was already mentioned, materials with a high-enough Cooper pairing energy 2Δ would allow a considerable reduction of the BCS surface resistance. According to the BCS theory, Δ is related to the critical temperature $\Delta = 3.5\ kT_c$. Therefore, superconductors with a higher critical temperature than niobium are being intensively studied. Among them are the well-known 'old' high-T_c superconductors, such as the B1 and A15 compounds NbN and Nb_3Sn. The potential of attaining a low surface resistance in a Nb_3Sn cavity is large, provided that the residual surface resistance can be made sufficiently low (Fig. 16 [35]). The critical temperatures are $T_c = 17$ K for NbN and 18.5 K for Nb_3Sn. Since their discovery in 1986 [36], the "new" high-T_c superconductors such as $YBa_2Cu_3O_{7-\delta}$ ($T_c = 93$ K) [37] have also been studied for RF applications.

Thin coatings of these materials on RF cavities are produced either by thermal diffusion out of the vapour phase (Nb_3Sn, NbN), by sputter deposition (NbN, NbTiN), by evaporation (NbN), or by electrophoretical deposition in a high-static magnetic field from an organic suspension of powdery sc material and sintering ($YBa_2Cu_3O_{7-\delta}$) [38]. The best results on RF cavities (made completely from high-T_c materials) are shown in Table 3.

Table 3

Accelerating cavity results for 'old' and 'new' high-T_c superconductors

Material/ Substrate	Fabrication method [a]	Frequency [GHz]	R_{BCS} [nΩ]	R_{s0} [nΩ]	T [K]	E_{amax} [MV/m]	Ref.
Nb_3Sn/Nb	DF	3.0	~16.0	94.0	4.2	9.9	39
	DF	0.5	~0.5	11.0	4.2	4.9	40
NbN/Nb	DF	4.5	~500	690	4.2	2.3	41
NbTiN/Cu	SpF	0.5	9.0	18.0	4.2	5.4	42
$YBa_2Cu_3O_{7-\delta}$	SiB	7.0	42×10^6	1.8×10^6	77	–	43

(a) D = diffusion, Sp = sputtering, Si = sintering of bulk material, F = film and B = bulk.

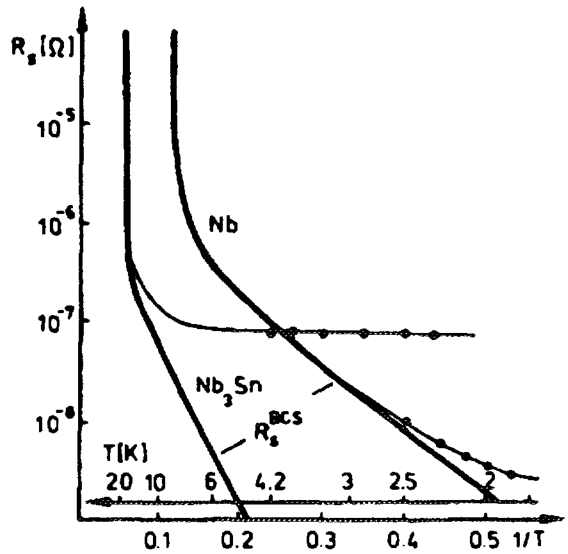


Fig. 16 The surface resistance of niobium in comparison with the A15 superconductor Nb_3Sn

8. TECHNOLOGICAL ACHIEVEMENTS FOR ACCELERATING CAVITIES

What matters in the end, from the standpoint of the accelerator builder at least, are the characteristics of the accelerating structure in the accelerator itself. It is here where the achievements of the technology of RF superconductivity should pay off and have to be assessed. Table 4 gives the performance of RF structures being permanently operated in electron accelerators.

In several laboratories accelerating gradients between 5 and 10 MV/m were obtained in structures of 1 m length and more. There is increasing evidence that structures, after being installed in the accelerator, do not degrade. At CERN, a LEP cavity has been operated in the SPS for more than 10000 h without any irreversible degradation. However, examples demonstrating the opposite can also be given, after the structures have been tested in the laboratory, equipped with additional components and moved to the accelerator.

The TRISTAN e^+e^- storage ring at KEK has been operated for several years with 32 500-MHz cavities to increase the energy of its beams of 10 mA each. The largest system of sc cavities actually in operation is at CEBAF, with 338 five-cell 1500-MHz cavities delivering 800 MV RF voltage.

The working principle of acceleration of electron beams with sc cavities has been demonstrated: in several laboratories the installation of larger numbers is envisaged. DESY has installed 16 cavities in HERA, and CERN will upgrade LEP to be able to create W^+W^- pairs in 1996 with 192 sc cavities, to complement the existing copper cavities. The first four modules (consisting of four four-cell cavities each) are installed and commissioned (Fig. 17).

Table 5

RF superconductivity based accelerators or accelerators with sc cavities permanently installed

Laboratory	Year of first operation	Frequency [GHz]	Total obtained voltage [MV]	Max. obtained gradient [MV/m]	Refs.
HEPL	1977	1.3	50	2.2 (a)	46
University of Illinois (MUSL-2)	1977	1.3	13	2.3(a)	47
CERN (SPS)	1988	0.35	16	7.1	48
Darmstadt (S-DALINAC)	1989	3.0	80	10.1	49
KEK (TRISTAN)	1989	0.5	200	>9.0	50
CERN (LEP)	1990	0.35	75	5	51
DESY (HERA)	1991	0.5	76	4	52
CEBAF	1994	1.5	800	8	53

(a) Structure design not multipactor-free

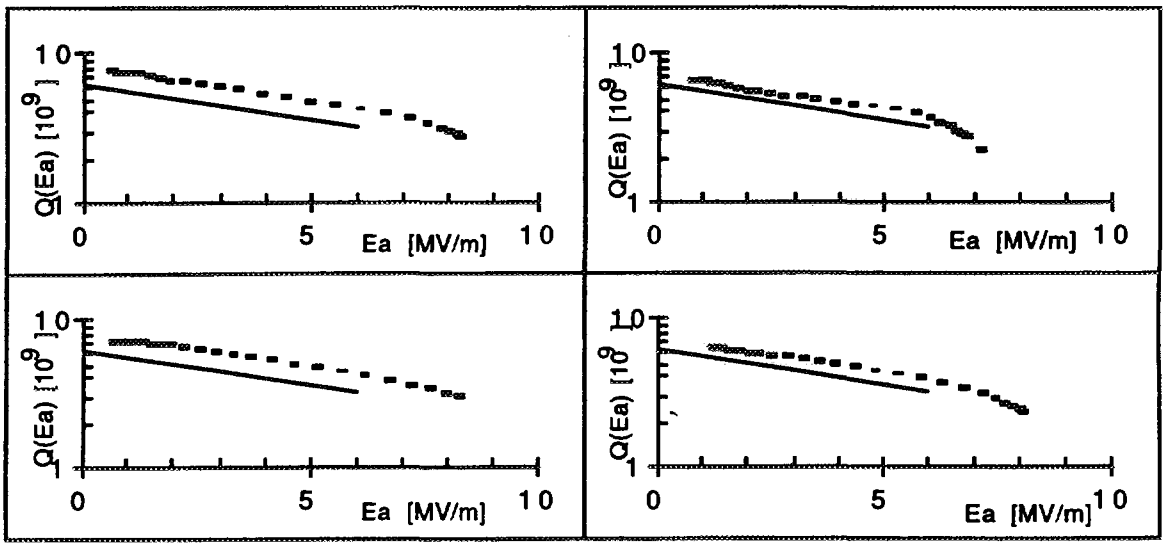


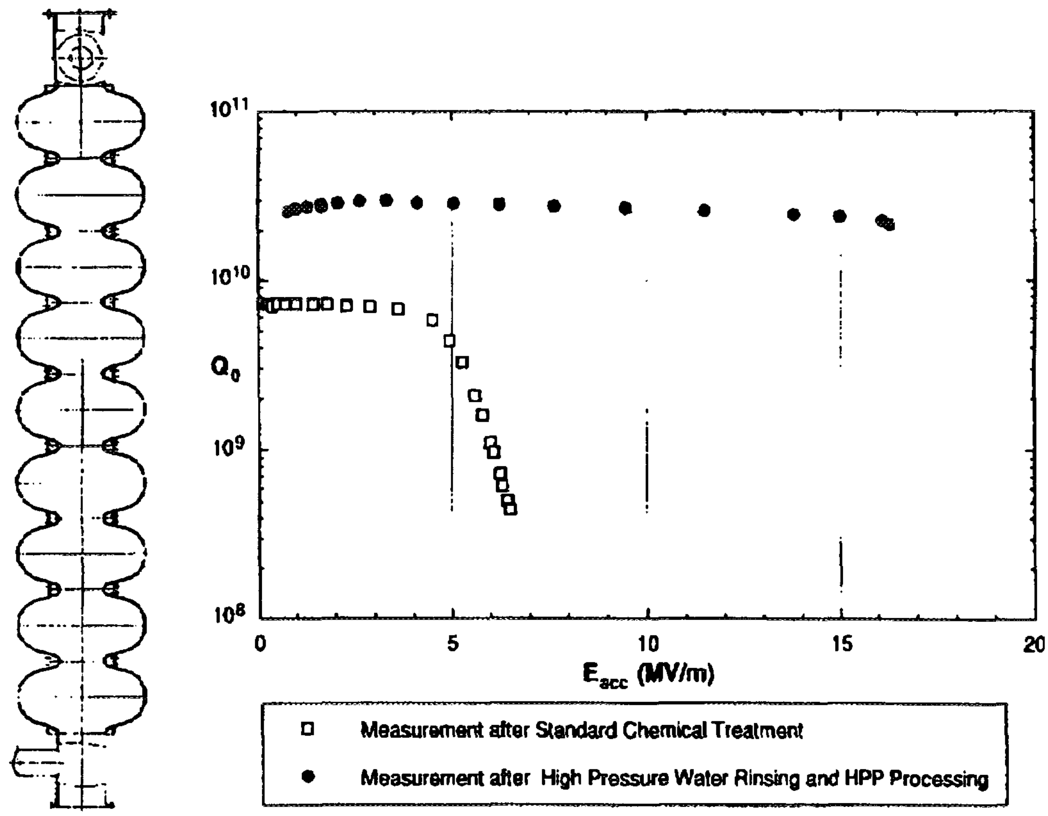
Fig. 17 Typical Q vs. E_a curves of individual four-cell cavities for LEP in horizontal test after assembly into a module (no couplers). The straight line gives the required performance.

9. CONCLUSION AND OUTLOOK

The big advance of the last decade in RF superconductivity for electron accelerators was the mastering of industrial fabrication and the putting into operation of larger numbers of structures with gradients between 5 and 10 MV/m and for beam currents up to 30 mA. Existing and new accelerators are already based completely, or to a large extent, on this technology. For the future, one technological incentive is the application of RF sc structures for high-current accelerators (particle "factories" [52]), and another is its application for high gradient accelerators and TeV linear colliders [53], for example, the TESLA project (Fig. 18). For CERN's CLIC project, a big fraction of the accelerator complex is based on sc RF technology [54]. Superconducting RF technology is still far from its theoretically possible limits, in particular after the discovery of high-T_c superconductors some years ago, leaving room for considerable improvement of performance in the future.

ACKNOWLEDGEMENTS

I would like to thank my colleagues from CERN for many clarifying discussions.



(Courtesy CERN Courier 74, vol. 8, p. 22)

Fig. 18 After high pressure water rinsing and high peak power (HPP) radio-frequency treatment at DESY's TESLA Test Facility, promising accelerating fields have been produced by prototype superconducting cavities (1.3 GHz) for a proposed TeV Superconducting Linear Accelerator (TESLA).

* * *

REFERENCES

[1] B. Richter, IEEE Trans. Nucl. Sci. NS-26 (1979) 4261.

[2] W. Weingarten, Proc. XVth Int. Conf. High Energy Acc., Hamburg (Germany) July 20-24, 1992, Ed. J. Rossbach, Singapore, 1993, Int. J. Mod. Phys. A (Proc. Suppl.) 2B (1993) 678.

[3] C. Benvenuti et al., Appl. Phys. Lett. 45 (1984) 583, C. Benvenuti et al., IEEE Trans. Magn. MAG-21 (1985) 153.

[4] C. Hauviller, Proc. Particle Accelerator Conference, Chicago, 1989 (IEEE, New York, 1989), p. 213.

[5] T. Tajima et al., Proc. 6th Workshop RF Supercond., CEBAF, Newport News VA (USA), Oct. 4-8, 1993, Ed. R. Sundelin, 944.

[6] U. Klein and D. Proch, Proc. Conf. Future Possibilities for Electron Accelerators, Charlottesville, Va., 1979, Eds. J.S. McCarthy and R.R. Whitney (Univ. Virginia, Charlottesville, 1979), N1–N17, V. Lagomarsino et al., IEEE Trans. Mag. MAG-15 (1979) 25.

[7] H. Piel, Proc. CERN Accelerator School, Hamburg, 1988, Ed. S. Turner, CERN 89–04 (1989), p. 149;
H. Lengeler, ibid., p. 197.

[8] H. Lengeler et al., IEEE Trans. Magn. MAG-21 (1985) 1014.

[9] U. Klein, Thesis, Wuppertal Univ., WUB–DI 81–2 (1981); G. Müller, Thesis, Wuppertal Univ., WUB–DI–83–1 (1983).

[10] J. Halbritter, Zeitschrift für Physik 238 (1970) 466.

[11] J. Matricon and D. St. James, Phys. Lett. 24A (1967) 241.

[12] P.G. de Gennes, Superconductivity of Metals and Alloys, W.A. Benjamin, New York 1966.

[13] T. Yogi, G.J. Dick and J.E. Mercereau, Phys. Rev. Lett. 39 (1977) 826.

[14] T. Yogi, Thesis, California Institute of Technology, Pasadena (1977).

[15] Y. Bruynseraede et al., Physica 54 (1971) 137.

[16] K. Schnitzke et al., Phys. Lett. A45 (1973) 241.

[17] H. Padamsee et al., IEEE Trans. Magn. MAG-15 (1979) 602.

[18] T. Hayes et al., Proc. Int. Conf. High Energy Particle Accelerators 1995, Dallas, TX, USA (in press).

[19] B. Hillenbrand et al., IEEE Trans. Magn. MAG-13 (1977) 491.

[20] H. Piel, Proc. Workshop RF Supercond., Karlsruhe 1980, Ed. M. Kuntze, KfK Report 3019 (1980) 85.

[21] H. Piel and R. Romijn, CERN EF/RF/80–3; Ph. Bernard et al., Nucl. Instrum. Methods 190 (1981) 257.

[22] Ph. Bernard et al., Proc. 6th Workshop RF Supercond., CEBAF, Newport News VA (USA), Oct. 4-8, 1993, Ed. R. Sundelin, 739.

[23. H. Padamsee, J. Tückmantel and W. Weingarten, IEEE Trans. Magn. MAG-19 (1983) 1308.

[24] H. Padamsee, IEEE Trans. Magn. MAG-19 (1983) 1322.

[25] H. Padamsee, Proc. 2nd Workshop on RF Supercond., CERN, 1984, Ed. H. Lengeler (CERN, Geneva, 1984), p. 339.

[26] H. Padamsee, IEEE Trans. Magn. MAG-21 (1985) 1007.

[27] P. Kneisel, J. Less Common Metals, 139 (1988) 179.

[28] J.R. Delayen, Proc. 3rd Workshop RF Supercond., Argonne IL (USA), Sep. 14–18, 1987, Ed. K.W. Shepard, ANL-PHY-88-1, 469.

[29. J. Tückmantel and W. Weingarten, CERN/EF/RF 82-6; W. Weingarten, Proc. 2nd Workshop on RF Supercond., Geneva, July 1984, Ed. H. Lengeler, 551.

[30] W. Weingarten, Proc. XIIIth Int. Symp. Discharges and Electrical Insulation in Vacuum, Paris (France), June 27-30, 1988, Les Ulis 1988, Ed. J.M. Buzzi and A. Septier, 480.

[31] Ph. Niedermann et al., J. Appl. Phys. 59 (1986) 892; J. Vac Sci. Technol. A8 (1990) 594.

[32] D. Moffat et al., Proc. 4th Workshop on RF Supercond., Tsukuba, 1989, Ed. Y. Kojima (KEK, Tsukuba), p. 445.

[33] C. Crawford et al., Part. Acc. 49 (1995) 1.

[34] D. Bloess, private communication; Ph. Bernard et al., Proc. 5th Workshop on RF Supercond., Hamburg, 1991, Ed. D. Proch [Part. Accel. 40 (1992)], p. 487.

[35] G. Arnolds-Mayer and N. Hilleret, Adv. in Cryogenic Eng. Mat. 28 (1982) 611.

[36] J.G. Bednorz and K.H. Müller, Z. Phys. B64 (1986) 189.

[37] M.K. Wu et al., Phys. Rev. Lett. 58 (1987) 908.

[38] G. Müller, Proc. 2nd Workshop on RF Supercond., CERN, 1984, Ed. H. Lengeler (CERN, Geneva 1984), p. 267.

[39] D. Dasbach et al., IEEE Trans. Magn., MAG-25 (1989) 1862; M. Peiniger et al., Proc. 3rd Workshop on RF Supercond., Argonne, 1987, Ed. K.W. Shepard (report ANL–PHY–88–1, Argonne, 1988), p. 503.

[40] G. Arnolds-Mayer and E. Chiaveri, Proc. 3rd Workshop on RF Supercond., Argonne, 1987, Ed. K.W. Shepard (report ANL–PHY–88–1, Argonne, 1987), p. 491.

[41] P. Fabbricatore et al., IEEE Trans. Magn. MAG-25 (1989) 1867.

[42] C. Benvenuti, Proc. 5th Workshop on RF Supercond., Hamburg, 1991, Ed. D. Proch [Part. Accel. 40 (1992)], p. 189.

[43] E. Minehara, N. Nagai and M. Takeuchi, Proc. 4th Workshop on RF Supercond., Tsukuba, 1989, Ed. Y. Kojima (KEK, Tsukuba), p. 695.

[44] J.R. Calarco et al., IEEE Trans. Nucl. Sci. NS-24 (1977) 1091.

[45] P. Axel et al., IEEE Trans. Nucl. Sci. NS-24 (1977) 1133.

[46] D. Boussard et al., in Proc. Part. Acc. Conf., Chicago, 1989, Eds. F. Bennett and J. Kopta, (IEEE, New York, 1989), p. 1783.

[47] J. Auerhammer et al., Proc. 6th Workshop on RF Supercond., Newport News, Va., 1993, Ed. R. Sundelin (CEBAF, Newport News, Va.), p. 1203.

[48] T. Furuya, Proc. 6th Workshop on RF Supercond., Newport News, Va 1993, Ed. R. Sundelin (CEBAF, Newport News, Va.), p. 23.

[49] G. Cavallari et al., Proc. 5th Workshop on RF Supercond., Hamburg, 1991, Ed. D. Proch [Part. Accel. 40 (1992)], p. 23.

[50] D. Proch et al., Proc. 6th Workshop on RF Supercond., Newport News, Va., 1993, Ed. R. Sundelin (CEBAF, Newport News, Va.), p 77.

[51] A. Hutton, Proc. 1994 European Particle Accelerator Conference, London, 27 June–1 July 1994, Ed. V. Suller and Ch. Petit-Jean-Genaz, (World Scientific, Singapore, 1994), p. 15.

[52] H. Padamsee et al., Ibid, p. 2048.

[53] R. Brinkmann, Ibid, p. 363.

[54] W. Schnell, CERN-LEP-RF/86-06 and Proc. Symposium on Advanced Accelerator Concepts, Madison, 1986, AIP Conf. Proc. 156, p. 12 and CERN-LEP-RF/86-06, 1986 (CLIC Note 13) Feb. 1986.

NEXT PAGE(S)
left BLANK

MATERIALS FOR SUPERCONDUCTING CAVITIES

B. Bonin
Commissariat à l'Energie Atomique, DSM/DAPNIA, France

Abstract
The ideal material for superconducting cavities should exhibit a high critical temperature, a high critical field, and, above all, a low surface resistance. Unfortunately, these requirements can be conflicting and a compromise has to be found. To date, most superconducting cavities for accelerators are made of niobium. We shall discuss here the reasons for this choice. Thin films of other materials such as NbN, Nb_3Sn, or even YBCO compounds can also be envisaged and are presently investigated in various laboratories. We show here that their success will depend critically on the crystalline perfection of these films.

1. INTRODUCTION

The advantages of superconducting cavities over normal conducting cavities are well known. These advantages can be exploited in many different ways since they permit continuous operation of the accelerator, improve the energetic conversion to the beam, relax the constraints on cavity design and minimize the cavity impedance seen by the beam. According to the accelerator characteristics, e.g. duty cycle, intensity, beam dynamics properties, a different priority is given to each of these advantages. Nevertheless, it should be remembered that all the advantages of superconducting cavities stem from one single property of the superconducting material, namely its very low *surface resistance*. We shall discuss first the requirements imposed on the superconductor by this criterion of minimal surface resistance.

2. CRITERIA OF CHOICE DERIVED FROM THE SURFACE RESISTANCE

The power dissipation per unit area of a superconductor in RF regime is related to the surface resistance of the material R_s via:

$$P = \frac{1}{2}.R_S.H^2, \quad (1)$$

where H is the RF magnetic field amplitude. A well known expression for the surface resistance of a superconductor can be derived from the two-fluid model [1]:

$$R_S = \frac{A}{T}.\sigma_n.\omega^2.\lambda^3.e^{-B.T_c/T} + R_{\text{residual}}, \quad (2)$$

where A and B are two constants which depend only weakly on the material, ω is the RF pulsation, σ_n the normal state conductivity of the material, λ its effective penetration depth, and T_c the superconductor critical temperature. The first term, improperly called R_{BCS}, corresponds to the dissipation due to normal electrons, and the second, R_{residual}, is associated with imperfections in the superconductor structure and behavior.

In many practical cases, the BCS term takes a non negligible fraction of the total surface resistance. As can be seen from Eq. (2), R_{BCS} depends strongly on the superconductor penetration depth and critical temperature. It is thus crucial to maximize T_c (requirement 1) and to minimize λ (requirement 2).

In Eq. (2), the penetration depth λ is an effective value, related to the London superconductor penetration depth λ_L by $\lambda = \lambda_L \cdot \sqrt{\frac{\xi_0}{\xi}}$, where ξ_0 and ξ are the coherence lengths in the pure and real material respectively. In the real material, the coherence length is given by

$$\xi^{-1} = \xi_0^{-1} + l^{-1},$$

where l is the electron mean free path. Two extreme cases can then be envisaged:

For clean superconductors, i.e. those with a large electron mean free path, $l >> \xi_0$, thus $\xi \approx \xi_0$, and Eq. (2) gives $R_{BCS}^{clean} \propto l$. For very clean material, Eq. (2) is no longer valid, and more sophisticated calculations based on BCS theory [2] predict a roughly constant surface resistance, independent of l, and thus, of the material purity.

For dirty superconductors, i.e. those with a small electron mean free path, $l << \xi_0$, thus $\xi \approx l$, hence $R_{BCS}^{clean} \propto l^{-1/2}$. The surface resistance thus diverges for very dirty superconductors (Fig. 1).

Between the clean and dirty limits, R_{BCS} takes a minimum value when the electron mean free path becomes comparable to the coherence length.

Pure metals and pure intermetallic compounds with a well defined stoechiometric composition, like Nb_3Sn, are usually clean superconductors, except if they have many crystalline defects; on the other hand, alloys enter the category of dirty superconductors, due to their very small electron mean free path. Consequently, they display large BCS surface resistance. For the same reason, i.e. a small electron mean free path, alloys also have a poor thermal conductivity at cryogenic temperatures, thus hampering the thermal stability of a cavity. For these two reasons, alloys are not suitable materials for superconducting cavities.

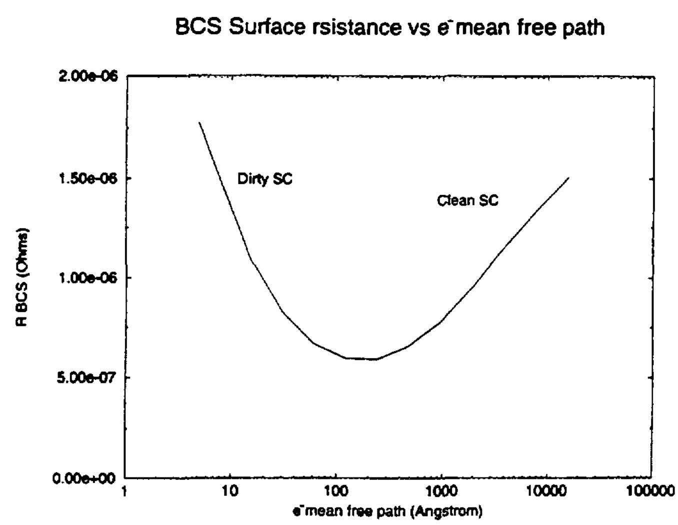


Fig. 1 The BCS surface resistance of niobium at 1.5 GHz and 4.2 K, as a function of the electron mean free path.

Usually, the optimum working conditions of the superconducting cavity are met when $R_{BCS} \approx R_{residual}$. It is therefore very important to minimize this last term. Many causes contribute to the residual surface resistance [3]. Some of these are extrinsic (trapped flux), and can be avoided. Some causes are intrinsic and are due to structural imperfections of the material, like inhomogeneities, grain boundaries or surface serration. The superconducting wave function or order parameter is sensitive to defects larger or of the same size as the coherence length ξ [4]. Materials with a large ξ–value will thus "forgive" large defects without an appreciable increase of the residual surface resistance. This is clearly a desirable feature for cavity applications where square meters of superconducting surface are exposed to the RF field, and are presumably difficult to prepare completely "defect-free".

Putting requirements 2 (small penetration depth) and 3 (large coherence length) together, we get the description of a type I superconductor. These are universally known as low T_c superconductors, and this is clearly in contradiction with requirement 1. The BCS theory [2] gives a relationship between the coherence length and the critical temperature:

$\xi_0 = 0.18\dfrac{\hbar . v_F}{k.T_c}$, where v_F is the Fermi velocity. The inverse relationship between the coherence length and the critical temperature indicates that the contradiction between requirements 1 and 3 is very deep indeed. Therefore, the ideal superconductor for RF applications does not exist, and subsequent choices clearly result from a compromise.

Lead, as an archetype of a type I superconductor, has been used for low frequency cavities, and has yielded a very low residual surface resistance. It is cheap, and easily available in a pure form. Unfortunately, at frequencies higher than a few hundred MHz, the BCS surface resistance becomes prohibitive, due to the low critical temperature of this material. Moreover, it has poor mechanical characteristics and oxidizes easily, with a subsequent degradation of the properties of the superconducting surface. For these reasons, lead tends to be progressively replaced by niobium, and is now confined to low frequency applications.

Type II superconductors can have a large T_c and a reasonably small penetration depth, so that their BCS surface resistance can be small, even at rather high cryogenic temperature. But their coherence length ξ is small, so type II superconductors tend to display rather high residual surface resistance, unless they are prepared "defect-free" (Fig. 2). Table 1 summarizes the characteristics of various superconductors.

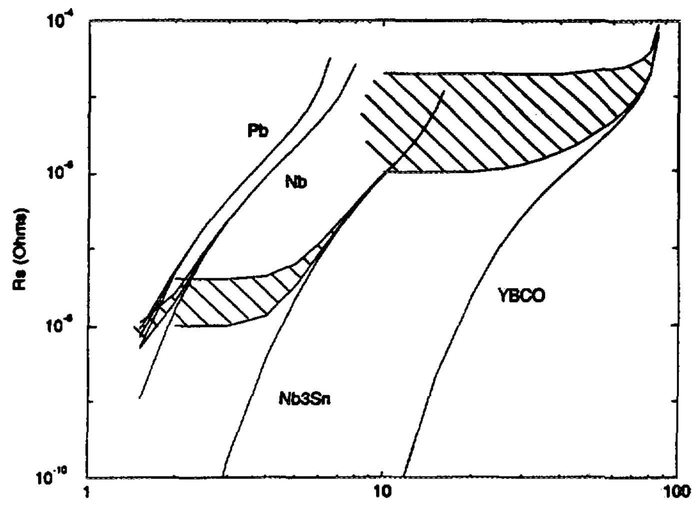


Fig. 2 The surface resistance vs temperature for a few typical superconductors at 1.5 GHz. In this diagram, the full lines show the BCS contribution, and the hatched areas represent the total surface resistance for "state-of-the-art" materials.

Table 1
Characteristics of various superconductors

Material	T_c (K)	λ (nm)	ξ_0 (nm)
Pb	7.2	39	83-92
TYPE I ↑			
TYPE II ↓			
Nb	9.2	32-44	30-60
$Nb_{0.6}Ti_{0.4}$	9.8	250-320	4
NbN	15-17	200-350	3-5
Nb_3Sn	18	110-170	3-6
YBCO	94	140	0.2-1.5

3. CRITICAL FIELD

High magnetic fields are present in accelerating cavities. In many cases, the amplitude of the RF magnetic fields approaches the order of magnitude of the superconductor critical fields. A superconducting material with a high critical field is thus desirable for RF applications (requirement 4). Since the nucleation time of a vortex is usually large compared with the RF period, the relevant critical field of an RF superconductor is not the usual critical fields H_{c1} or H_{c2}, but is believed to be the superheating field H_{sh} [5]. H_{sh} is related to the thermodynamic critical field H_{th} via:

$$H_{sh} = 0.89 \cdot \frac{H_{th}}{\sqrt{\kappa}} \qquad \kappa << 1$$

$$H_{sh} \approx H_{th} \qquad \kappa >> 1$$

It should be noted that for type II superconductors, H_{sh} can be significantly larger than H_{c1}.

Theoretical arguments in favor of the hypothesis that the limiting RF field is H_{sh} are supported by the fact that experimentally, the superheating limit has been approached, but never broken [6]. Table 2 gives the thermodynamic field, superheating field and maximum attained RF field for various superconductors.

4. NIOBIUM

In view of the above criteria, Nb appears as a serious candidate for superconducting cavities. It has the highest T_c and H_{sh} of all pure metals. Being a soft type II superconductor, it occupies a position of compromise between the four requirements mentioned above.

Niobium homogeneity and purity are important issues for RF applications because it determines the thermal stability of the cavity. It was quickly realized that a frequent gradient limitation in superconducting cavities is due to thermal instabilities triggered by microscopic hot spots, for example normal conducting inclusions. This led researchers to investigate in detail the thermal behavior of niobium cavities, in relation to the material characteristics [7].

Table 2

Thermodynamic, superheating and maximum obtained RF field of some superconductors

Material	H_{th} (A/m)	H_{sh} (A/m)	H_{RF} max (A/m)
Pb	6.4 10^4	8.4 10^4	6.4 10^4
Nb	1.6 10^5	1.9 10^5	1.6 10^5
Nb_3Sn	2.5 10^5	3.2 10^5	8.0 10^4
YBCO	≈6 10^5 -10^6	≈6 10^5 - 8 10^5	8.0 10^4

Niobium thermal conductivity

Two parameters are relevant for the description of the thermal behavior of the cavity, namely the niobium thermal conductivity and the Kapitza resistance at the niobium-helium interface. In the case of a hot spot, most of the thermal gradient is located in the niobium sheet, and the thermal properties of the interface play a minor role [7]. For a good thermal stability, a niobium cavity must thus be made from a material with high thermal conductivity. At cryogenic temperatures, the main heat carriers in niobium are electrons, and their mean free path is limited primarily by collisions with impurity atoms [8].

The electron mean free path is usually given in terms of a quantity (residual resistivity ratio), defined as:

$$\mathrm{RRR} = \frac{\rho_{300K}}{\rho_{0K}}$$

where ρ_{300K} is the room temperature resistivity (this term is constant, ρ_{300K} = 1.45 10^{-7} $\Omega^{-1}.\mathrm{m}^{-1}$) and ρ_{0K} the normal state resistivity of niobium at zero temperature.

The approximate relationship giving the electron mean free path in niobium as a function of RRR is:

$$l(T = 0\mathrm{K})(\mathrm{Angstrom}) \approx 27.\mathrm{RRR}.$$

The thermal conductivity of niobium (Fig. 3) has been measured as a function of purity and past metallurgical history. A useful rule of thumb is:

$$\lambda(T = 4.2\mathrm{K})(\mathrm{W.m^{-1}.K^{-1}}) \approx \mathrm{RRR}/4.$$

With usual values of RRR (a few hundreds), this relationship gives a thermal conductivity significantly smaller for niobium than for OFHC copper.

In the superconducting state, paired electrons in niobium decouple from the lattice and no longer participate in heat conduction. Heat is then carried by the small fraction of unpaired electrons. The poor thermal conductivity of niobium is thus intrinsically due to its superconducting nature. A Nb layer of 1 μm thickness would be sufficient for superconducting

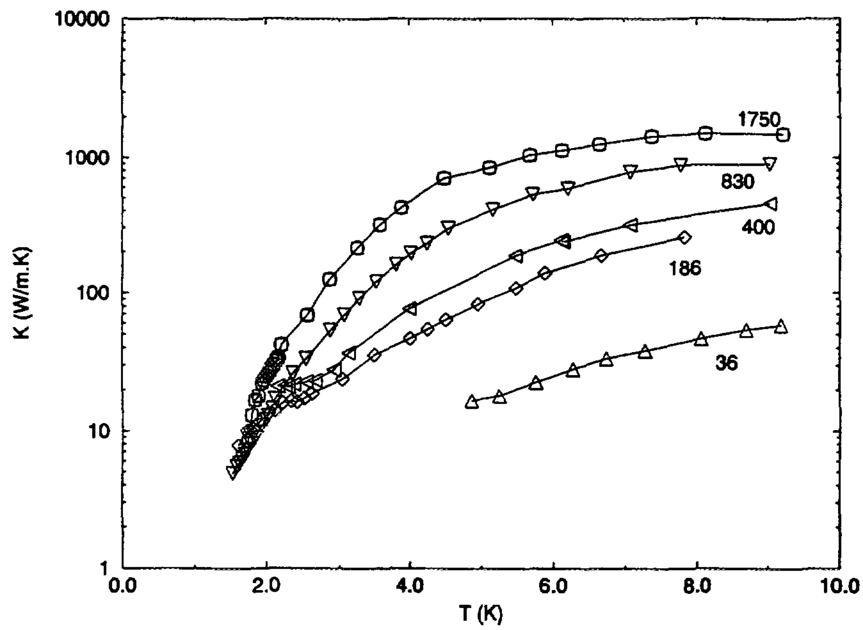


Fig. 3 The thermal conductivity of niobium as a function of temperature, for various RRR values.

cavities. Expensive as it is, the rest of the material serves merely as a substrate, despite its poor mechanical and thermal properties!

The above arguments show how important it is to make superconducting cavities with very pure material. One of the most significant advances in cavity performance is due to a recent effort from niobium suppliers to produce niobium of improved purity.

Niobium production

Niobium production from the ore to the raw ingot involves a complicated path, with chemical treatments and high temperature steps, which have been described in [9]. The purified ingot is then rolled into sheet form. Recrystallization annealings are usually performed by the supplier, between the rolling steps or before the sheet delivery, in order to warrant a uniform and well controlled grain size (typically 50 µm, for good sheet deformation capabilities). The cavity can then be formed.

The main impurities which contribute to the RRR degradation are C, N and O [9]. For state-of-the-art material (RRR 300), these are present at a concentration of about 100 at. ppm; Tantalum is also present in large amounts (1000 at. ppm). This element is difficult to separate from niobium because both have very similar chemical properties.

Niobium purity is improved by electron beam melting (2400 °C) in high vacuum. Light impurities (C, N, O, H) are vaporized during the process; four to six passes are generally considered to be necessary to reach a RRR level of the order of 300 [9].

Niobium post-purification

Post-purification of niobium can be achieved by a heat treatment associated with solid state gettering [10]. The material, either in sheet or in finished cavity form, is heated under high vacuum at temperatures ranging between 1000 °C and 1400 °C. This temperature is sufficient to permit diffusion of interstitial impurities C, N, O in the bulk material. The material is thus homogenized, thanks to the dissolution of the inclusions. At the same time, a liner of titanium placed close to the material to be purified sublimates. Its vapor deposits at the niobium surface and acts as a getter for the interstitial impurities which diffuse here. After a few hours of heat treatment, the bulk niobium is purified to a large extent. This technique is now capable of improving the niobium RRR by a factor of 10 (from 30 to 300 for "reactor grade" material, and from 200 to 2000 for "state-of-the-art" material). Niobium thermal conductivity can be considerably improved by this treatment, with a subsequent improvement of the cavity quench threshold. In spite of its drawbacks (cost, severe degradation of its mechanical properties), niobium post-purification is presently an interesting route to very high cavity performance.

5. THIN FILMS

Superconducting cavities internally coated with a thin superconducting film can also be envisaged. For a complete screening of the RF field by the superconductor, a minimum thickness of 10 times the penetration depth λ should be deposited. This corresponds to a thickness ranging between 0.5 and 2 µm and can be achieved by many deposition techniques. Considerable advantages can be expected from such "thin film" cavities:

- a cheap substrate can be used, with a subsequent cavity cost saving;
- a good heat conductor can be chosen for the substrate, giving a good cavity thermal stability.
  These two criteria can be met by using copper as the substrate.
- materials unavailable in bulk form can be deposited in thin films, with potentially interesting superconducting properties.

Nb/Cu thin films

A number of superconducting materials have been investigated for RF applications. So far, the most successful is niobium, sputter-deposited on a copper substrate. The deposition method, pioneered at CERN [11], is DC magnetron sputtering. Here, the copper cavity plays the combined roles of substrate and vacuum vessel, and the niobium target (cathode) is introduced at the centre of the vessel. The method is well suited to large cavities, working at low frequency, and at 4.2 K. In this case, it provides an appreciable cost saving in supply material. In addition, the RRR of the deposited material is around 30, close to the optimum value which minimizes the BCS contribution to the surface resistance. The archetype of this kind of application is LEP2 cavities. After a difficult start, this technology has now been transferred to industry, which provides cavities meeting the LEP2 specifications with a good success rate.

Nb(Ti)N thin films

Deposition of niobium-nitride films has also been attempted in various laboratories[12]. Here, the hope is to exploit the high critical temperature of this intermetallic compound (17 K) to lower the BCS contribution to the surface resistance, and to permit cavity operation at temperatures higher than 2 K (the goal is 4.2 K). The deposition method used so far is reactive sputtering, with a pure metallic target of Nb or NbTi alloy, and introduction of nitrogen in the sputtering gas.

So far, full success cannot be claimed. However, encouraging results have been obtained on samples, with low surface resistance at low field levels (400 mΩ at 4 GHz), and a BCS contribution much lower than for niobium (three times less at 4.2 K). Nitride accelerating cavities with competitive characteristics have not yet been produced.

Nb_3Sn layers

Thin films of Nb_3Sn have had more success (accelerating gradients as high as 15 MV/m, cavity Q-value higher than 10^{10} at low fields) [13], but the fabrication method (start from a pure niobium cavity, evaporate tin on its surface, and heat up until Nb_3Sn is formed to a few µm thickness) is probably expensive and does not lend itself to an easy industrialization. It might be interesting, however, in cases where cost is not a very important criterion, or to upgrade the performance of existing cavities.

YBCO

Theoretically, very low surface resistance could be expected from high T_c superconductors like YBCO. So far, the experimental results fall short of these expectations [14]. Sintered ceramics display a very large residual resistance. Moreover, R_{res} increases dramatically with increasing RF field. Thick coatings prepared by laser ablation, sputtering or other deposition techniques also display more or less the same behavior, whose cause can be found in the granular nature of the material. Epitaxially grown thin films have much smaller surface resistance, and their surface resistance is independent of the amplitude of the RF field. Their superconducting characteristics could make them attractive for accelerator applications. Unfortunately, they require a monocrystalline substrate, and so far have only been produced on flat samples of very restricted area.

Granular superconductivity

Most of the high T_c superconductor films and all the sputtered films have in common a very unfortunate feature: they display an increase of surface resistance with increasing RF field. This behavior limits severely the accelerating gradient obtainable from "thin film" or "high T_c " cavities. The disease seems to be worse for materials with higher T_c, or, more precisely, with smaller coherence length. The cause of this misfortune can be traced back to the granular nature of this kind of superconducting materials. The RF dissipation of granular superconductors has been investigated by many authors [15]. Here, the superconductor is modelled as an array of superconducting grains separated by weak links which, in the present case, can be structure defects like grain boundaries. The granularity is more pronounced if the grains are smaller, the boundaries more resistive or the coherence length smaller. If the current induced by the RF field through the weak link stays below its critical current, the weak link behaves as a resistively shunted inductance; in the opposite case, the weak link can be modelled as a pure resistor and the power dissipation of the array rises abruptly. These models provide guidelines to produce better, less granular superconducting films. They indicate that a weak link is a defect as large or larger than the coherence length. Since ξ_0 is small for high T_c materials, the challenge is to make perfect films, with a minimum number of structure defects.

6. CONCLUSION

In view of the arguments discussed above, one can understand why superconductors for magnet applications and superconductors for cavities are so different. The only requirement they have in common is a high critical temperature. Whereas superconductors for magnets need a high H_{c2}, and thus a small coherence length, materials for cavities need a low surface resistance and a large superheating field H_{sh}. In order to ensure an effective flux pinning, the former needs to be a dirty superconductor, the latter a clean one. Table 3 summarizes the requirements put on superconducting materials for RF and DC applications.

Overall, niobium seems to be an excellent compromise between the conflicting requirements imposed on a superconductor suitable for RF cavities, and its supremacy will probably last for a long time. In view of their superior critical fields and temperatures, compounds like NbN, Nb_3Sn or even YBCO are attractive substitutes to niobium, but it will be necessary to produce them in a very pure and perfect way in order to avoid problems related to the low coherence length of these materials.

Table 3
Main requirements of SC materials for RF and DC applications

Regime	H_c	H_{c2}	j_c	R_s	T_c	ξ	λ	Material
DC	-	large	large	-	high	small (pinning is needed)	large (bulk current)	NbTi Nb_3Sn
RF	large	-	-	small	high (R_s depends on T/T_c)	large (to give little sensitivity to structure defects)	small $R_s \propto \lambda^3$	Nb Nb(Ti)N ? Nb_3Sn? YBCO?
	Primary requirements					Consequences		

* * * * *

REFERENCES

[1] H. Piel, "Superconductivity in Particle Accelerators", CAS, Hamburg, 1988, ed. S. Turner, CERN 89-04 (1989).

J.H. Hinken, "Superconducting electronics", Springer Verlag

[2] C.C. Mattis and J. Bardeen, Phys. Rev. 111 412 (1958)

[3] J. Pierce, "Superconducting microwave resonators" in "Methods of Experimental Physics" vol. 2, ed. R. Coleman, Acad. Press, N.Y. (1974)

M. Tigner and H. Padamsee in "Physics of high energy particle accelerators", SLAC summer school (1982), ed. M. Month, AIP Conf. Proc. 105, 801 (1983)

J. Halbritter, J. Appl. Phys. 42 82 (1971)

P. Kneisel, O. Stolz, J. Halbritter, Proc. of the Appl. Superconductivity Conf. IEEE, New York (1972) p 657

[4] P. Schmüser, these proceedings

[5] T. Yogi and M. Mercereau, Phys. Rev. Lett. 39 826 (1977)

J. Halbritter, 1st Workshop on RF Superconductivity, Karlsruhe, KfK 3019 (1980)

[6] W. Weingarten, Particle World 1 93 (1990)

G. Müller, Proc. 3rd Workshop on RF Superconductivity, Argonne (1987) ANL-PHY-88-1 p 331

[7] H. Padamsee, 1st Workshop on RF Superconductivity, Karlsruhe, KfK 3019 (1980)

H. Padamsee, IEEE T. Mag. 19 1322 (1983)

[8] K.K. Schulze, Journal of Metals, May 1981, p33

[9] M. Hörmann, J. of Less Common Metals 139 1 (1988)

[10] T. Haas, Surf. Sci. 5 345 (1966)

E. Fromm, Vacuum 19 191 (1969)

H. Farrell, Surf. Sci. 38 31 (1973)

[11] G. Arnolds-Mayer and W. Weingarten, IEEE T. Mag. 23 1620 (1987)

C. Benvenuti et al., IEEE T. Mag. 21 153 (1985)

[12] C. Benvenuti et al., 5th Workshop on RF Superconductivity, DESY, Hamburg, ed. D. Proch, pp 189 and 518 (1991)

R. di Leo, A. Nigro, G. Nobile and R. Vaglio, J. Low Temp. Phys. 78 41 (1990)

P. Bosland et al, 6th Workshop on RF Superconductivity, CEBAF, USA, ed. R. Sundelin, (1993) p. 1028

[13] J.J. Hanak et al. RCA Rev. 25 342 (1964)

M. Peiniger and H. Piel, IEEE T. NS 32 3612 (1985)

R. Röth and P. Kneisel, 7th Workshop on RF Superconductivity, Gif sur Yvette, ed. B. Bonin (1995)

[14] G. Müller, this CERN course.

[15] L. Cohen, I.R. Gray, A. Porch and J.R. Waldram, J. Phys. F: Metals Physics, 17 L179 (1987)

A.M. Portis, IEEE T. Mag. 25 2390 (1989)

O.G. Vendyk et al., Zh. Tekh. Fiz. 59 107 (1989)

B. Bonin and H. Safa, Superc. Sci. Technol. 4 257 (1991)

C. Attanasio, L. Maritato and R. Vaglio, Phys. Rev. B 43 6128 (1991)

QUEST FOR HIGH GRADIENTS

D. Proch
Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany

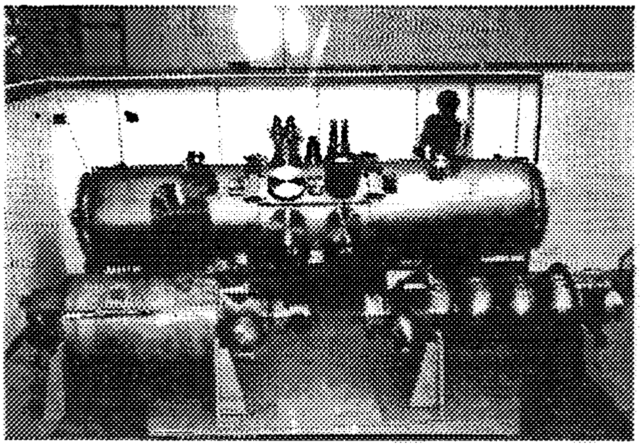
Abstract
In this chapter the method of reaching high accelerating gradients in superconducting cavities is presented. In the first part fundamentals of cavity design and measurement techniques are described. In the main part field limitation by multipacting, quench and field emission are discussed. Finally the state of the art and recent measurements with single and multicell structures are outlined.

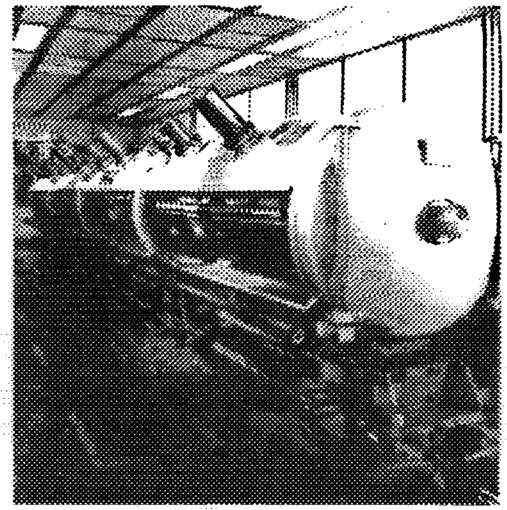
1. INTRODUCTION

There is an increasing number of accelerator projects with superconducting cavities as accelerating structures: storage rings (CERN [1], DESY [2], KEK [3], see Fig. 1), linacs for electrons or positrons (Frascati [4], CEBAF [5], HEPL [6], Darmstadt [7]), linacs for heavy ions (Argonne [8], Legnaro [9], Jaeri [10]), development programmes for high current application (Cornell [11], KEK [12]) and for a superconducting linear collider (TESLA [13]). The advantage of superconducting over normal conducting cavities is that accelerating gradients can be established with less AC power (including the cryogenic plant) and that higher gradients are possible for continuos wave operation. Another advantage of the low power demand is that the cavity geometry does not need to be strictly optimised for minimum power demand. Therefore a shape of the cavity can be chosen which produces less beam-cavity interaction at frequencies higher than the accelerating mode (so-called higher order modes, HOM). Operating systems work at gradients far below the theoretical limit (at less than 20 %) whereas the best values of single-cell measurements come near to this limit. In this chapter I will describe the quest for high gradients by following the historical development of this research area. Finally the state of the art is presented. I will not describe fundamentals of superconductivity (see Schmüser, this course) and basics of RF superconductivity (see Weingarten, this course). I will also concentrate on solid niobium as the cavity material and on those structures which have a phase velocity near, or equal to, the speed of light.

2. SOME CAVITY PROPERTIES

Figure 2 shows the typical shape of a superconducting accelerating structure. The rounded shape is chosen to suppress multipacting (see next sections). The electromagnetic fields inside the cavity are determined by the conditions of Maxwell's equation. For a simple cavity shape (pill box, cylinder) the analytic solution is given in standard text books (for example in [14]). More complex structures can be analysed by finite element programmes, for example URMEL [15] or in 3D by MAFIA [16]). The electric field lines inside the resonator are closed by surface currents in the uppermost metal boundary. The ratio of the accelerating electric field on the axis, the surface current in the metal and the magnetic field at the surface is determined by the cavity shape (and not by the conductivity!). In Fig. 3 typical numbers for the maximum surface electric field (at the iris) and the maximum magnetic field (at the equator) are given for an accelerating field of 1 MV/m. These numbers can be optimised by appropriate shaping of the cavity. The RF loss is proportional to the square of the surface current and inverse proportional to the conductivity. In case of a normal conducting cavity the favourite material is Cu and its conductivity is given. Therefore careful shaping of the cavity is needed to reduce the RF losses.





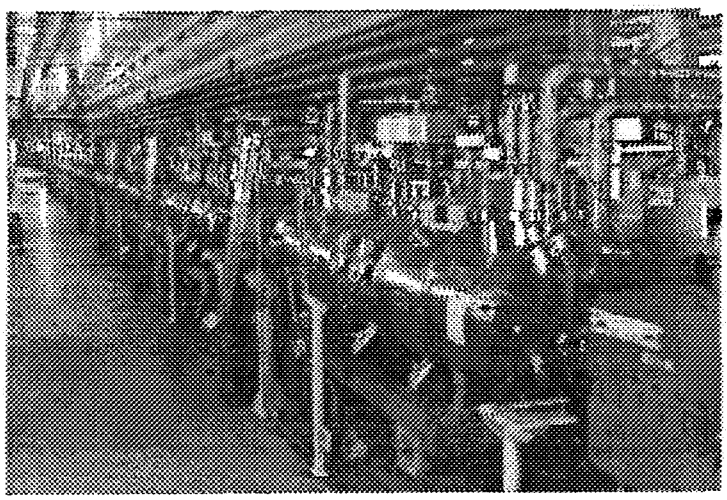


Fig. 1 Examples of superconducting cavity installations. Top left: 500-MHz DESY module with two 4-cell cavities per cryostat, top right: 356-MHz CERN module with four 4-cell cavities per cryostat, bottom: 500-MHz installation at KEK in the TRISTAN tunnel.

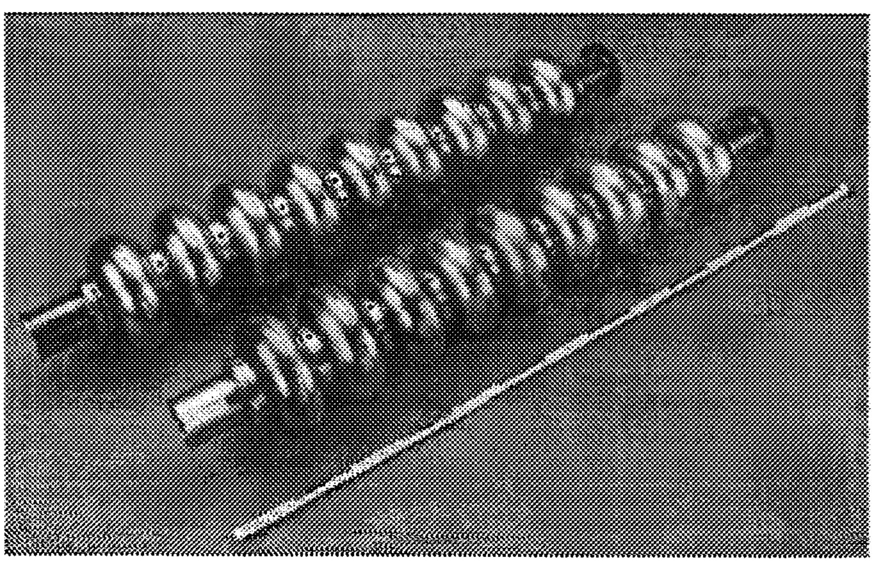
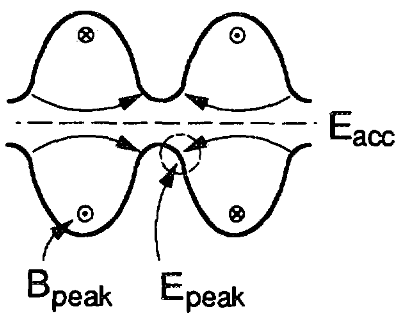


Fig. 2 Nine-cell 1.3-GHz cavities for TESLA. The rounded shape of a typical superconducting cavity design is chosen to suppress multipacting.



$$E_{peak} \simeq 2 \cdot E_{acc}$$

$$\frac{B_{peak}}{E_{acc}} \simeq \frac{4 \text{ mT}}{\text{MV/m}}$$

$$\text{Nb:} \quad B_c \simeq 200 \text{ mT}$$

$$\Longrightarrow \text{max } E_{acc} \simeq 50 \text{ MV/m}$$

Fig. 3 Cross section of a typical superconducting cavity design. The number for maximum surface electric and magnetic field is given for 1 MV/m accelerating gradient.

For a superconducting cavity the resistivity is enhanced by several orders of magnitude as compared to Cu and is very small (but not zero as for dc current, see Weingarten's chapter). Therefore the shape of a superconducting cavity is optimised for other properties: low excitation of higher frequencies by the beam, low surface magnetic field to enlarge the limit by superconducting- to normal-conducting phase transition, low surface electric field to suppress field emission or appropriate shape to avoid multipacting.

The efficiency of a cavity to produce an accelerating field at a given RF power is defined by the shunt impedance R. This is equivalent to Ohm's law where the resistance is the proportionality factor between the square of the voltage and the loss. Therefore Eq. (1) is called the "Ohm's law for the accelerator":

$$U^2 = R \cdot P \tag{1}$$

where:

U is the accelerating voltage per cavity,
R is the shunt impedance per cavity,
P is the RF power loss per cavity.

At a given power loss P the accelerating voltage can be enlarged by better shaping of the cavity and by using lower-loss wall material. Therefore it is plausible that the shunt impedance can be divided into two factors: R/Q which only depends on the cavity shape and Q which mainly depends on the conductivity of the cavity wall.

$$R = \frac{R}{Q} \cdot Q \tag{2}$$

where:

R/Q is the "cavity shape factor",
Q is the quality factor of the cavity resonator.

Thus Eq. (1) becomes

$$U^2 = \frac{R}{Q} \cdot Q \cdot P \tag{3}$$

The R/Q value can be calculated by several cavity codes or can be measured in comparison with a simple cavity which can be evaluated analytically. Superconducting cavities have a typical value of R/Q = 100 Ω per cell. As mentioned above, with normal conducting cavities the shunt impedance and thus the R/Q value must be optimised to reduce the RF power. Here typical values of R/Q are around 200 Ω per cell.

The quality factor Q is proportional to the ratio of the stored energy and the power loss

$$Q = \omega \cdot \frac{W}{P} \tag{4}$$

where:

$\omega = 2\pi f$, angular frequency,

W is the stored energy,

P is the RF power loss.

Typical values for Q are 40000 for copper cavities at 500 MHz and 10^{10} for niobium cavities at 1.3 GHz, 1.8 K. With these numbers, one can easily calculate (one cell has a length of 0.5 wavelength) that the RF power loss per metre at 1 MV/m is 37.5 kW for normal conducting and 0.1 W for superconducting cavities. As a rule of thumb the AC power demand of a refrigerator is 1 kW for 1 W cooling power at 1.8 K. Therefore the total power demand for a superconducting cavity is 100 W at this gradient.

2.1 How to measure the accelerating gradient of a superconducting cavity

The accelerating gradient can be measured according to Eq. (3) by knowledge of the geometric factor R/Q and measurement of the quality factor Q and the RF power loss P. The R/Q value can be obtained from programmes like URMEL [15], the power loss P is determined by microwave measurements of the cavity absorbed power. The quality factor Q can be measured by the bandwidth of the resonator according to Eq. (5).

$$Q = \frac{f}{\Delta f} \tag{5}$$

where:

Q is the quality factor,

f is the resonant frequency,

Δf is the bandwidth (full width, half power).

The resonant bandwidth of a superconducting cavity is very small. At a frequency of 1.3 GHz and quality factor of 10^{10}, the bandwidth is just 0.13 Hz and in order to measure the bandwidth, the generator has to be more stable than this. On the other hand the superconducting resonator will vibrate due to noise coming from the cryogenics or the surrounding environment. Therefore the resonance curve itself will move and a fixed, highly-stable oscillator will be unable to excite the resonance. This is why the generator has to be stabilised by the cavity itself (another way is to close a self-excited loop of an amplifier and cavity circuit). Figure 4 shows a resonance curve and the change of the phase of the cavity field when passing through the resonance. This phase information can be used to lock the generator frequency on the centre of the resonance curve. Figure 5 shows the principal microwave circuit of such a stabilisation technique. The phase difference of the forward RF power and the cavity field is measured. The variable phase shifter is adjusted in such a way that the signal is zero at the centre of the resonance curve. Above (below) this frequency, the phase difference is positive (negative). This phase difference signal is converted to an error voltage and finally controls the generator frequency via a frequency tuning circuit.

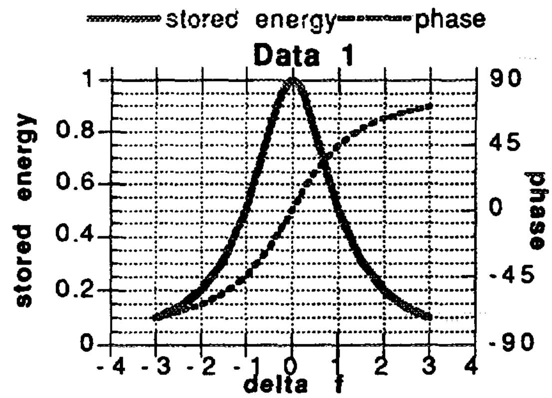


Fig. 4 Amplitude and phase response of a resonator.

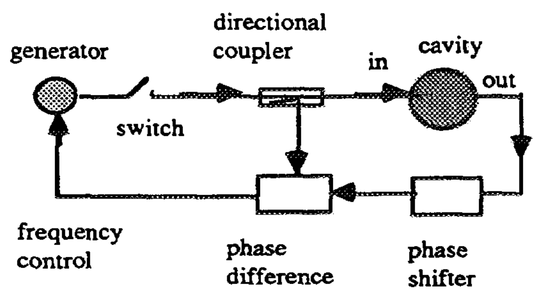


Fig. 5 Principal circuit of a frequency stabilisation system. The phase response of the resonator is used to stabilise the generator to the centre frequency of the resonator.

The quality factor can be determined by measuring the bandwidth according to Eq. (5). But at high Q values this is an inconvenient method. Here it is easier to measure the decay time of the cavity field after switching off the resonator.

$$U(t) = U_0 e^{-\frac{t}{\tau}} \tag{6}$$

where:

$U(t)$ is the cavity field after switching off the generator,
t is the time since switching off the resonator,
U_0 is the cavity field at equilibrium condition,
τ is the time constant of the exponential decay of the cavity field.

The quality factor Q is given by:

$$Q = \omega \cdot \frac{\tau}{2} \tag{7}$$

In Eq. (7) the time constant is $\tau/2$ because the quality factor Q is related to the stored energy (which is proportional to the square of the cavity field).

With the above example ($Q = 10^{10}$, $f = 1.3$ GHz) the decay time constant is 2.4 sec and can be measured easily. The drive power is switched by the p-i-n modulator in Fig. 5.

The accelerating voltage U now can be calculated with Eq. (3) and (6) because R/Q, Q and P are known. The on-time of the pulsed RF power has to be long enough for complete filling of the cavity.

$$U(t) = U \cdot (1 - e^{-\frac{t}{\tau}}) \tag{8}$$

where:

$U(t)$ is the time dependent filling of the cavity field,

t is the time since switching on the resonator

U is the value of the cavity field after complete filling.

Figure 6 shows the microwave signals of a measurement set up for superconducting cavities according to the arrangement of Fig. 5. In this example, the RF drive is switched on for 20 sec. The cavity voltage increases (Eq. (8)) and for 10 sec a flat top of U is reached. After switching off the generator, the cavity field decays (Eq. (6)) and the decay constant is measured to determine the quality factor.

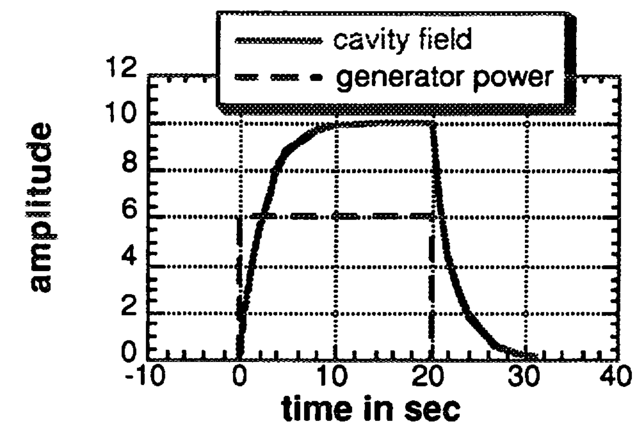


Fig. 6 Oscilloscope trace of the cavity voltage for a pulsed generator power. The example is given for the following parameters: $f_0 = 1.3$ GHz, $Q_0 = 10^{10}$, generator power is 70 W, 9-cell cavity with R/Q-value of 1000 Ω (TESLA case).

3. FIELD LIMITATION BY MULTIPACTING

The first superconducting cavities were made from lead (in the late sixties). Then niobium was chosen as material because of its superior properties ($T_c = 9.25$ K, $B_c = 190$ mT as compared to $T_c = 7.2$ K and $H_c = 804$ G for lead). Figure 7 shows a cross section of such an early Nb resonator. The two cups were machined from a solid Nb cylinder and welded at the equator. This fabrication technique as well as the shape (long cylindrical part at the equator, sharp corner at the transition to the iris) is typical for the design of normal conducting resonators. The cavity was measured in the pulse mode technique with the frequency stabilisation circuit as described before. The measurements, however, showed dramatic deviations as compared to the expected curve (see Fig. 8). The cavity started filling, but at a certain field value, no further increase could be gained. The excessive forward RF power was absorbed by the cavity without any increase of the field level. The transition at the limiting field was sharp. Sometimes a low level barrier could be overcome, but the same limiting phenomenon started at a somewhat higher field level again.

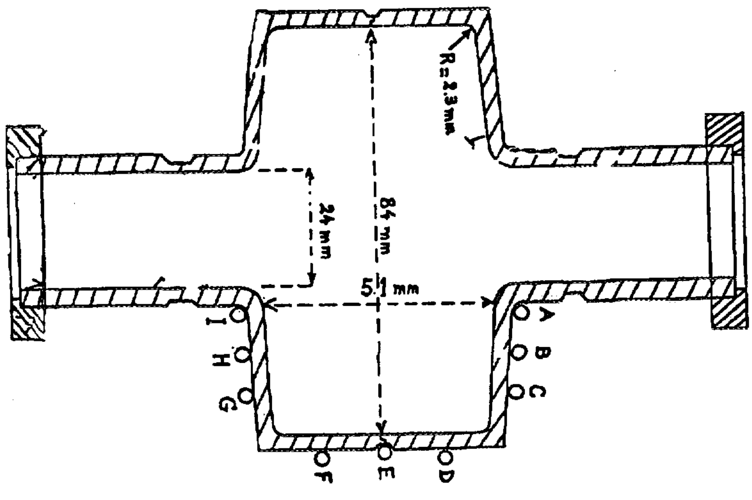


Fig. 7 Cross section of an early superconducting resonator. A cylindrical part at the large diameter and a sharp corner at the transition to the iris region is typical for this design.

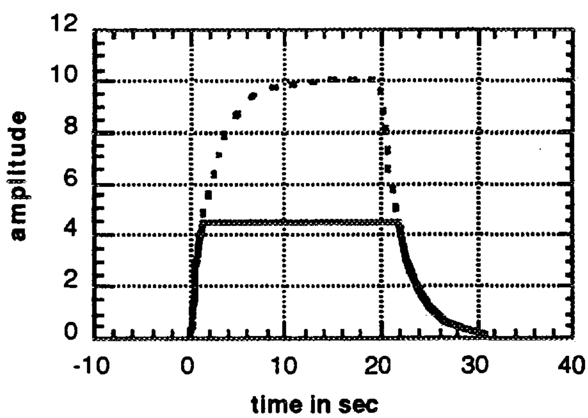


Fig. 8 Oscilloscope trace of the cavity field. The multipacting phenomenon (see text) limits the cavity field with a sharp transition at low field levels.

The analysis of the experimental observation concluded that at a certain field level all additional RF power was absorbed by an unknown mechanism. This kind of resonant absorption suggested that multiple electron impacting, so-called "multipacting" might be involved in the limiting mechanism. An electron avalanche is initiated by the multipacting effect at certain resonant conditions of the electromagnetic field. The multipacting current withdraws energy from the electromagnetic field. The magnitude of the multipacting current is stabilised to such a value that the cavity field is limited at the critical value.

A simple example of multipacting happens in the electric field between parallel plates (see Fig. 9). A first electron is accelerated by the electric field and crosses the gap to the other plate during one half RF cycle. If the time of flight is $N \cdot T/2$; $N = 1, 3, 5 \ldots$; T = period time; the electron could cross the gap backwards again (resonant conditions). If the impact energy is in the range between 50 eV and 500 eV (depending somewhat on the type of metal and surface condition), more than one primary electron will be created (see Fig. 10). Thus an electron avalanche will grow and absorb any RF power which exceeds the necessary value to establish the field value at the resonant condition.

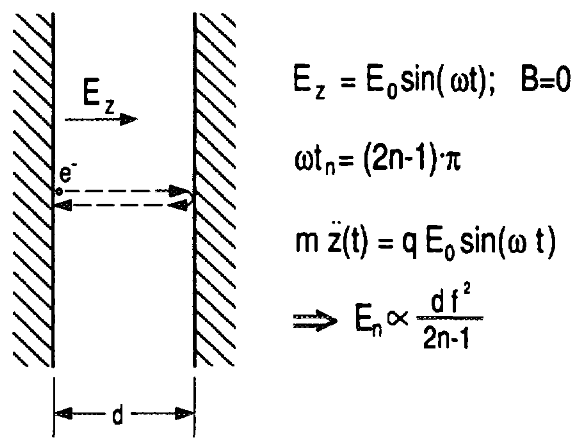


Fig. 9 Parallel-plate geometry as example of two-side multipacting in a time dependent electric field

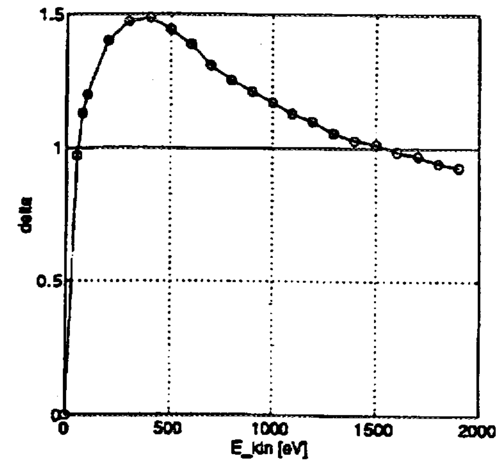


Fig. 10 Production rate (delta) of secondary electrons after impacting a metal surface by one primary electron (= secondary yield) as function of the impact energy

There was the suspicion that such a two-point multipacting between both end plates of the resonator might be the reason for the observed limitation. The counter argument is that the magnetic field component of the resonator field (azimutal field around the axis of rotation) will bend the electrons to the outer wall and prevent a closed orbit.

Experiments were undertaken to find the location of enhanced losses. Temperature sensors were placed at the outer cavity wall. The temperature sensors were specially selected carbon resistors with a high gradient of resistivity vs. temperature at the operating temperature between 4.1 K and 1.8 K. Heat pulses were detected at the outer cylindrical wall next to the iris transition. To simulate a possible multipacting resonance, tracking codes were developed. The trajectory of an electron which was started at the suspicious surface of the cavity area was calculated under the influence of the electromagnetic field. Resonant electron trajectories were found. Figure 11 gives an example of these trajectories. The characteristics of this multipacting are as follows:

- Electrons start and return to the same surface ("one-point multipacting").
- The resonant trajectories are closed orbits with the shape of a circle (first order, N=1) of an eight-like shape (second order, N=2), double-eight-like shape (third order, N=3) and so on.
- The closed orbit is created by the bending force of the magnetic field (cyclotron resonance). Only a small electric field is needed to accelerate the electron away from the surface.
- The trajectory circle is very small compared to the cavity dimensions (only about 1 % or less). Earlier simulations did not find these resonances since the cavity cross section was scanned with a mesh size in the dangerous area that was too large to resolve the trajectories.

The most successful remedy against this one-point multipacting was the design of a cavity shape which avoids the conditions for multipacting: no constant magnetic field along the surface and strong electric field to eject electrons far away from the dangerous surface. These conditions are fulfilled by a spherical (or more generally by an elliptical) cross section of the cavity [17]. Figure 12 shows such a favourable cavity shape. There is no cylindrical part with constant magnetic and low electric field. There is a zero crossing of the electric field at the equator (from conditions of symmetry) and the field strength of the electric field increases strongly due to the fast transition to the iris geometry. It was observed that another type of multipacting is created at this geometry: two side trajectories crossing the equatorial plane. Fortunately these trajectories proved to be unstable so that in the experiment this limitation could be overcome easily.

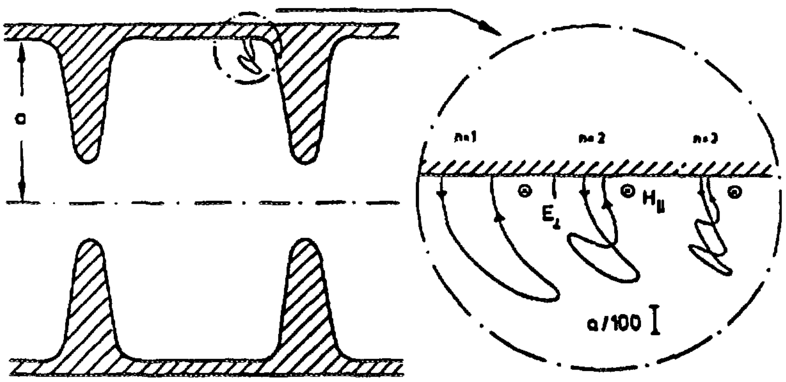
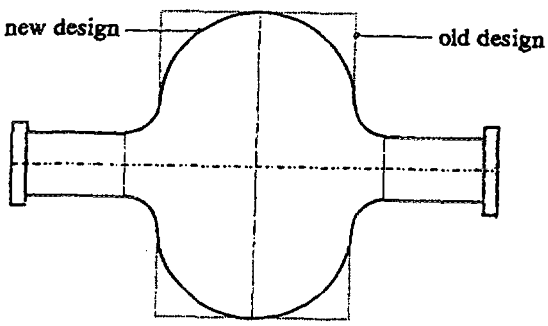


Fig. 11 Trajectories of "one-point" multipacting in resonators with a long cylindrical part at the large diameter



multipactor-free cavity shape

Fig. 12 New cavity shape to suppress multipacting. The spherical (or generally speaking the elliptical) cross section of the cavity avoids resonant conditions for multipacting. The "old" cavity shape with sharp corners and a cylindrical outer contour is shown in dashed lines. This shape produces multipacting resonances.

4. FIELD LIMITATION BY QUENCH

After the elimination of multipacting, higher gradients of superconducting cavities were expected. Figure 13 shows that a new type of limitation was observed. The cavity experienced a regular filling behaviour. At some level, however, the stored energy was absorbed during a time period much shorter than the natural filling time constant. The field in the cavity went to nearly zero (less than 1 % of the value just before the breakdown) and remained at this condition for some time although the generator offered forward power continuously. After some time the cavity started filling, reached the critical value and broke down again.

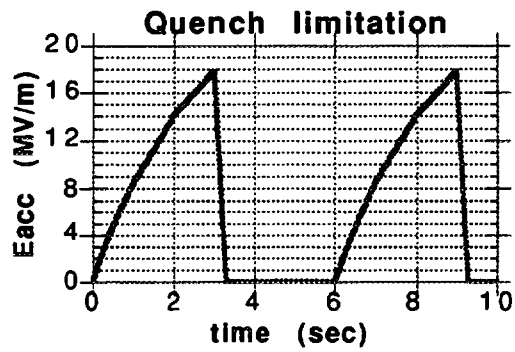


Fig. 13 Typical oscilloscope trace of the cavity field with a breakdown due to a quench

The diagnostic system to detect locations of enhanced losses in a cavity had been further developed. Instead of installing many fixed temperature sensors to the outer wall of the cavity (which means assembly effort to place the thermometers in good contact with the cavity surface and to dismantle all after the test), a system of rotating sensors has been developed. A frame with about 200 resistors per 1 metre cavity length is installed around the cavity. The temperature of the sensors and thus the temperature of the outer cavity wall is measured for the condition of high field and no field in the cavity at one angular position. Then the resistor frame is rotated (by about 3°) and a new measurement is carried out. In this way a temperature map of the whole cavity surface is created. This diagnostic tool gives valuable information about the location, spatial distribution and (if calibrated) about the absolute value of the dissipated power. Figure 14 gives an example of the mechanical arrangement, Fig. 15 shows a typical temperature map indicating the presence of one hot spot. After localising, the hot spot cavities were inspected from the inside to identify the reason of the breakdown. In many cases "defects" were detected: cracks in a weld, foreign particles sitting on the surface, "drying spots" from the last water rinsing etc. These findings concluded the following breakdown mechanism (see Fig. 16):

- A normal-conducting particle is sitting at the surface and produces enhanced RF losses.

- The enhanced heat is transferred through the niobium to the outside cooling helium. Therefore the temperature of the inner niobium surface next to the normal-conducting defect is raised.

- With increasing RF field, the heat production of the normal-conducting spot is increased and the surface temperature rises further.

- At a critical RF field the niobium surface next to the hot spot reaches the transition temperature of niobium (or strictly speaking: the reduced critical field at the enhanced temperature is reached).

- The niobium underneath the hot spot will switch to normal conductivity and further increase the heat flux. Finally the normal-conducting area will grow and dissipate the stored cavity energy in a short time. This breakdown process is named a "quench".

There are two principal ways to reduce the quench limitation:

1 Any kind of contamination has to be avoided by quality control of the niobium, extremely careful cleaning of the cavity surface and final handling of the cavity under best dust-free conditions. Also auxiliary equipment (pumping lines, coupler, etc.) has to be treated in the same way to avoid contaminating the cavity surface.

Scratches of the niobium or "loose" niobium particles (welding balls) have to be avoided, too.

2 If there are remaining defects on the surface, the other way is to improve the thermal conductivity of the niobium itself. In this way the temperature increase due to enhanced heat flux is reduced. Model calculations concluded [18] that the quench field (due to a normal-conducting spot) increases with the square root of the thermal conductivity of the niobium. This is valid as long as the Kapitza resistance (= thermal resistance of the Nb-to-liquid-helium interface) is small compared to the thermal resistance of the Nb itself.

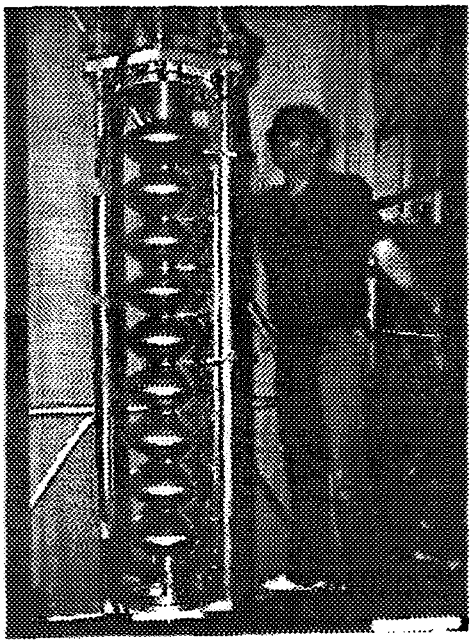


Fig. 14 A rotating temperature-mapping system assembled around a 9-cell 1.0-GHz niobium cavity. Carbon resistors are used to detect heat pulses from "bad spots" inside the cavity.

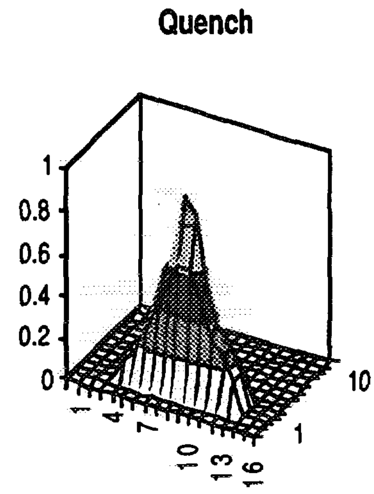


Fig. 15 Typical temperature map showing a single heat spot. In this plot the azimuthal and meridial position of the temperature sensor is transferred in rectangular x, y coordinates. The vertical amplitude corresponds to the measured increase of temperature.

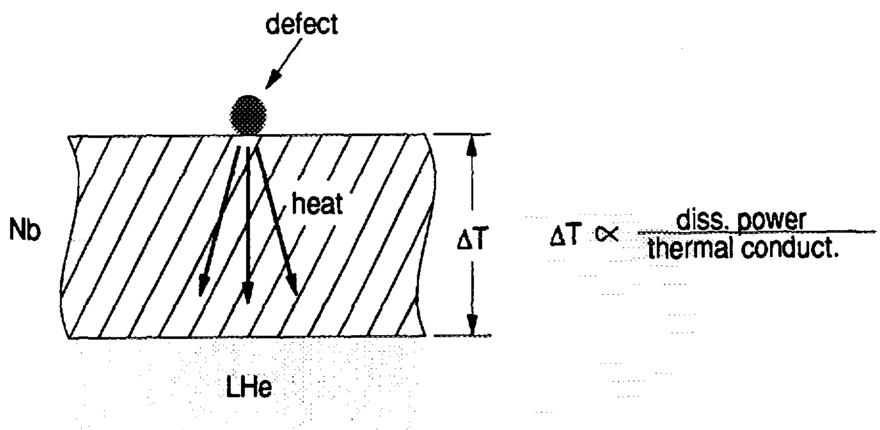


Fig. 16 Model of a breakdown due to excessive heating by a local defect ("quench").

Great progress has been made in clean treatment and improved thermal conductivity of the niobium during the last decade. Nevertheless, more effort is needed to transfer the excellent performances of single-cell cavities to real accelerating structures with all their auxiliary equipment (couplers, HOM dampers, tuners, etc.).

5. LIMITATION BY FIELD EMISSION OF ELECTRONS

If a cavity does not show multipacting or quench behaviour, a third type of phenomena usually limits the available gradient. Figure 17 demonstrates such a limitation. At first the cavity follows the expected filling curve. But at higher field levels, the rate of increase of the electric field slows down. An equilibrium level is reached which is lower than the expected value. If the generator power is raised further, the cavity field value is only slightly enlarged. This behaviour can be more clearly analysed if the standard Q vs. cavity voltage diagram is plotted (see Fig. 18). At low field values, the cavity quality factor stays constant (as expected). At high fields, the quality factor drops down dramatically for further increased cavity voltage. Under these conditions, γ-ray activity is noted outside the cryostat. A temperature map shows the locations of additional heat dissipation (see Fig. 19). Heat is produced along a line. If the temperature map is orientated like the projection of a globe (the beam axis penetrates the north and south pole) then the line of loss always follows a meridian.

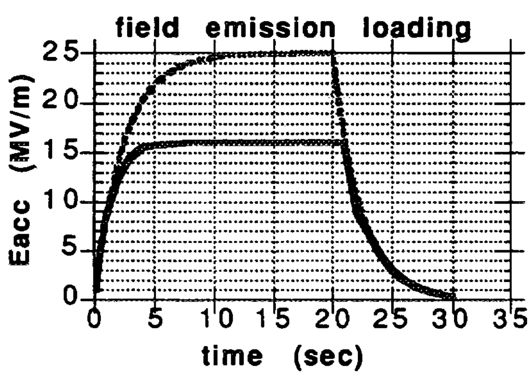


Fig. 17 Typical oscilloscope trace of the cavity field with field emission loading. The field follows the expected filling curve up to some critical value. Above this level the increase of the cavity field is considerably reduced and an equilibrium state is reached which is lower than the calculated value.

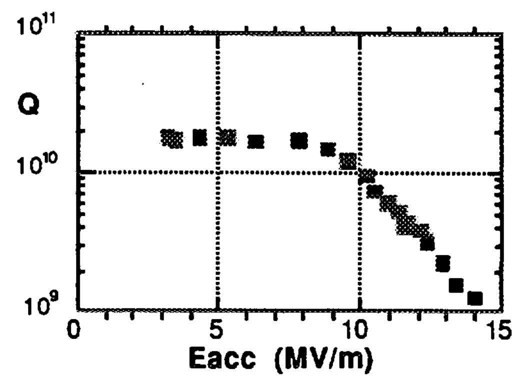


Fig. 18 Measured quality factor of a cavity versus accelerating gradient. Above 10 MV/m field, emitted current is present and is accelerated by the RF field. This results in additional losses so that the quality factor decreases.

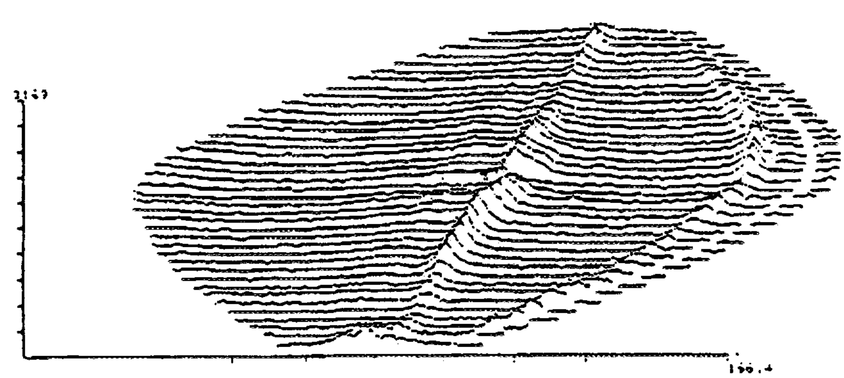


Fig. 19 Temperature map of a cavity loaded by field emission current. Two emitters produce two lines of additional losses. The lines follow a meridian.

The explanation of this limitation is that electrons are produced by field emission at the location of high surface electric fields, i.e. at the area of the iris. Then electrons are accelerated by the electric fields of the cavity. Because of the rotational symmetry of the cavity, there is no azimutal electric field compound (in the accelerating mode TM_{010}). Therefore the electron will hit the cavity surface along a meridian line and produce heat by impact.

Field emission of metal surfaces is a tunnel effect of electrons through the surface barrier. The extracted current grows exponentially (Fowler-Nordheim, see Bonin's chapter) with the surface field. That explains why the quality factor of the cavity rapidly decreases (i.e. the additional losses increase quickly) beyond the onset of field emission.

Field emission studies with DC and RF fields conclude that particles on the surface are the most likely reason for field emission. Here the microscopic electric field is enhanced by geometric effects. Although particles could be identified as the source of field emission, the detailed mechanism of why some emit but many do not is not understood. A detailed description of RF field emission is given in B. Bonin's chapter.

6. STATE OF THE ART

6.1 Results of large scale production

Niobium sheets for cavity production can be obtained from industry in high quality. A thermal conductivity values of 75 W/mK at 4.2 K is a standard quality. Usually the value of RRR (residual resistance ratio: ratio of electric resistance at room temperature and at 4.2 K in the normal-conducting state) is quoted because this quantity is easier to be measured. RRR and thermal conductivity values are related to each other (Wiedemann-Franz law) and the above value corresponds to RRR = 300. This is an improvement of a factor ten compared to the first cavities for storage ring application. Progress was made in handling, welding and dust free assembly of cavities, too.

At CEBAF, 338 Nb cavities (5-cell, 1.5-GHz) were fabricated during 1991–1994 by industry and chemically treated and assembled by the laboratory. The Nb sheet material has a RRR value of around 250. This is a large data base of cavity performance and is discussed here as a representative of the state of the art of superconducting cavity production. The cavities are assembled in pairs and are measured in a vertical cryostat. After this acceptance test, four pairs are grouped in a (horizontal) cryostat for installation in the accelerator tunnel. The vertical pair measurements gave the following results:

- Figure 20 shows the three typical qualities of performance: limitation by field emission (here at 10 MV/m) with the typical fast Q degradation by field emission; limitation by a quench with a nice flat Q vs. E_{acc} behaviour until quench break down at 13 MV/m; finally one of the best cavities which could be operated up to 19 MV/m (limited by available RF power)
- Figure 21 shows a "scatter plot" of the maximum gradient and Q values reached in all cavities during acceptance tests. It can be seen that the average value is 9.8 MV/m, but that there is a large spread of performance from 3 MV/m up to 18 MV/m.
- The cavities are limited by quench or field emission to about the same amount.

All cavities have been installed in the tunnel and commissioning of the linac started at the end of 1994. The design gradient is 5 MV/m. It will be interesting to see whether the cavity performance is preserved from vertical pair test to operation in the linac.

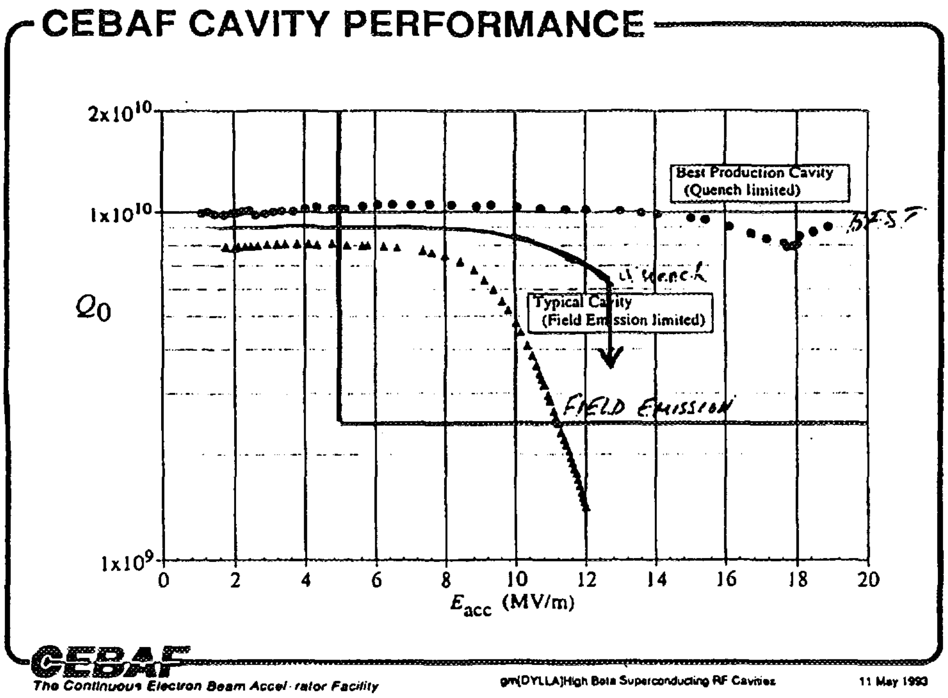


Fig. 20 Measurement of 5-cell (1.5-GHz) cavities at CEBAF. The three cavities are representatives of the typical performance: limitation by quench, field emission or (best cavity) only by available RF power.

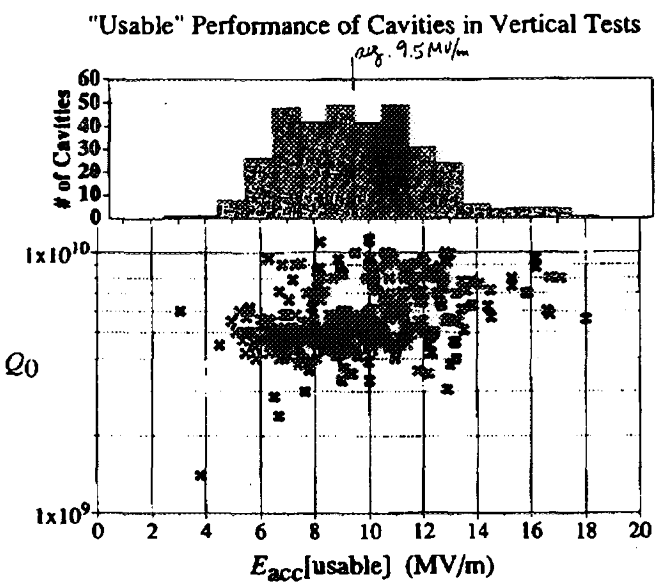


Fig. 21 Performance of all 338 CEBAF 5-cell cavities (1.5-GHz) during their acceptance test (pair measurement in vertical cryostats).

6.2 Automized chemistry and clean-room assembly

It has been mentioned that dust particles are the most likely source of field emission. Therefore absolutely clean conditions are essential to prepare a high-gradient cavity. The standard procedure of cavity fabrication is to make cups from niobium sheets (by deep drawing or spinning) and to weld those cups at the equator and iris region by electron-beam welding. After fabrication the cavity surface is cleaned by removing about 100 μm of the niobium material at the inner surface. This is done by etching the niobium with a solution of H_3PO_4, HNO_3 and HF. Is is important to use the best grade of acid to avoid contamination of the clean niobium surface by residues in the acid. After the acid treatment the niobium surface has to be rinsed with high quality clean water. The best result can be expected if a closed circuit of acid treatment, water rinsing and subsequent drying is installed. Such a system has been developed at Saclay and is being used for the TESLA Test Facility (TTF) at DESY (TESLA = TeV range Electron Superconducting Linear Accelerator). After this treatment couplers, higher-order mode dampers, pick-up probes etc. have to be attached to the cavity. This must be done under dust-free conditions, too. Therefore clean rooms are installed at all laboratories which handle superconducting cavities. The quality of the clean room is class 10 (10 particles of μm size per cubic foot and minute) and corresponds to the requirements in semiconductor industry.

6.3 High-pressure water rinse

Sapphire samples have been used to measure the cleanliness of the standard niobium surface treatment. These samples accompanied the niobium during each step of treatment. Afterwards the size and number of particles on the surface were measured with a laser reflectometer. The result was that hundreds of particles (per square inch) of μm size are still present on the surface. Spraying the surface with high-pressure water (about 100 bar) is a common method in the semiconductor industry to further clean the surface. Recently the same method has been used to clean cavities. At the TTF (DESY) a 2-m long rod is inserted in the vertical cavity after chemical etching and rinsing is done with ultra-pure water. The head of the hollow rod has a spraying nozzle and is slowly rotated and moved along the cavity axis to reach the whole inner cavity surface for cleaning. This high-pressure water cleaning is now the standard cleaning procedure for the TTF cavities. Figure 22 shows the improvement of the cavity performance after such a high-pressure water rinse.

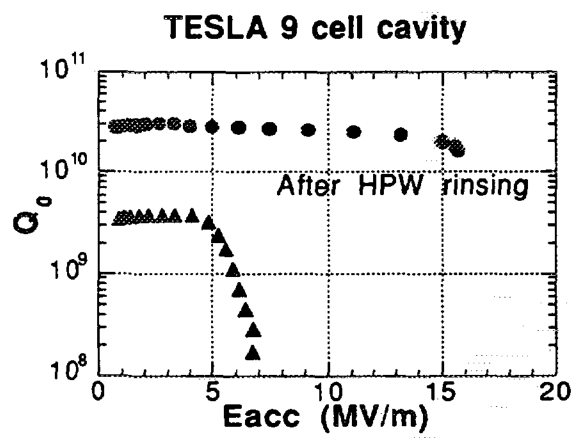


Fig. 22 Best improvement of cavity performance after high pressure water cleaning.

6.4 Post purification by Ti heat treatment

A high thermal conductivity of the niobium is essential to avoid limitation by quench. During welding of the niobium cavities using an electron-beam welding machine, the hot niobium can absorb oxygen, carbon or nitrogen due to insufficient vacuum conditions. This will result in a reduced thermal conductivity at the welded region. Unfortunately, the weld at the equator is at the location of highest surface current. Therefore a post purification of the niobium cavity after fabrication could "repair" the thermal conductivity at those deteriorated regions and improve the thermal conductivity elsewhere. The procedure is to heat the cavity and some titanium (placed inside and outside the cavity) at 1400 °C for about 4 hours. The evaporated Ti covers the niobium surface. The Ti film will getter dissolved gases (O, C, N) from the bulk niobium because of its higher affinity compared to niobium. The mobility of the dissolved gases in niobium at 1400 °C is high enough to acheive a sufficient cleaning of the bulk niobium by this solid-state diffusion process within 4 hours. After the heat treatment the cavity has to be cleaned of Ti by etching away about 100 μm of Nb surface. Typically the thermal conductivity can be improved by a factor of 2–3 by this procedure. It has to be mentioned, however, that the niobium is softened by the heat treatment (the yield strength $\sigma_{0.2}$ reduces typically from 40 N/mm^2 to 10 N/mm^2). This reduced mechanical strength has to be compensated by careful handling and the use of appropriate fixtures during assembly of the cavities.

6.5 High power processing

Investigation of RF field emission leads to the conclusion that small particles are the origin of the emitted current. Therefore very careful cleaning is needed to suppress this limitation. Nevertheless, some "residual" dust will remain on the surface or may contaminate the cavity during the final assembly procedure. Therefore an "in situ" cleaning procedure, which can be applied after assembly or during operation of the superconducting cavity is of great help. Such a method was developed at Cornell recently [19] and has been successfully applied at several laboratories. The cavity is operated with short (approx. 50 to 500 μsec) RF pulses of high power (about 1 MW for a 9-cell TESLA cavity, 1.3 GHz). The cavity is heavily overcoupled to the feeding line in order to reduce the filling time (some 100 μsec, TESLA cavity) as compared to a matched superconducting resonator (about 1 sec, TESLA cavity). Therefore the cavity field reaches a high value in a short time. A field emitter will heat up, accordingly, and reach a high temperature. The process is a transient behaviour so that a thermal equilibrium state is not yet reached. Many photos of such an event suggest that the emitting particle is "exploded" due to excessive heating. Figure 23 shows such an exploded emitter and its characteristic "star burst" signature.

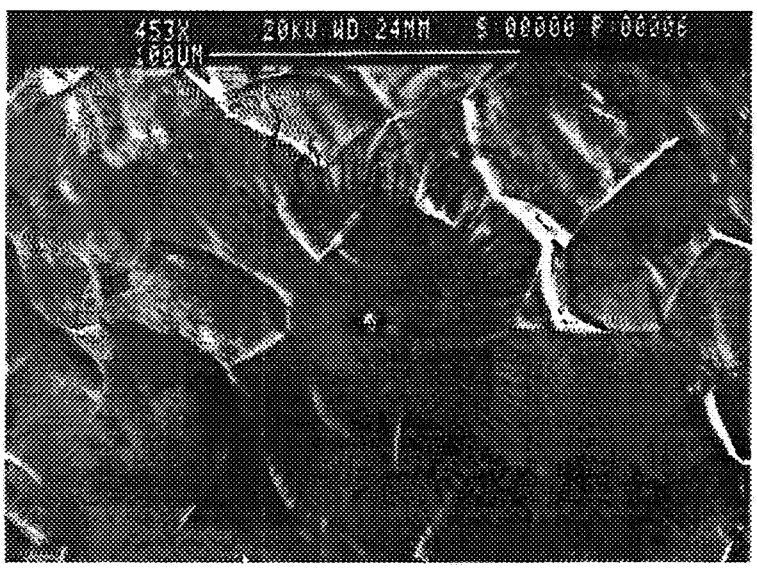


Fig. 23 "Star burst" picture of a field emitter after processing with high peak RF power. It is assumed that the field emitting particle "explodes" during the processing.

During high power processing, many processing events are observed by a short breakdown in the stored cavity field. Figure 24 shows a typical Q vs. E_{acc} curve before and after high power processing. It can be seen that the onset of field emission is considerably enlarged. There is a concern that the left over material of an exploded field emitter might raise the RF loss of the cavity or produce new field emitters. Indeed, it is observed that the Q of a cavity is slightly reduced after high power processing. But this is the only known way to cure a cavity from field emission in situ, i.e. without dismantling, cleaning and repeated cool down.

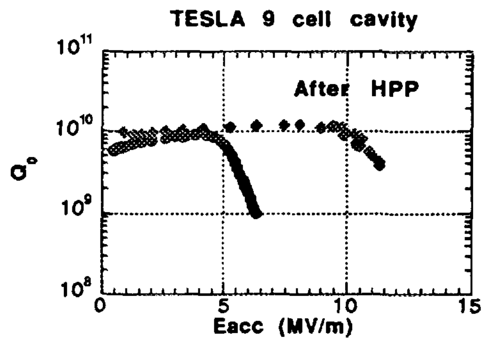


Fig. 24 Performance of a cavity before and after processing with high peak RF power.

6.6 Best results of single and multicell cavities

It is a general finding that high gradients can be reached more easily with single-cell cavities than with multicell cavities. This is explained by a statistical model for the defect probability. In the case of field emission this is justified by the observation that many small particles (i.e. possible field emitters) can still be found on a thoroughly cleaned surface. In the case of a quench defect, the statistical argument seems to be plausible, too. Therefore single-cell cavities are good test vehicles to investigate new treatment methods, or to search for the next high gradient limits. Considerable effort in quality control, cleaning and handling techniques is needed, however, to transfer the results from single cell cavities to multicell structures.

Single-cell activities at KEK [20] and CEBAF [21] are presented here as an example. At KEK niobium material with moderate thermal conductivity (RRR = 200) from Japanese companies is used. The cavities are electropolished by 200 μm. This gives smoother surfaces but the injected hydrogen (during the electropolish) must be removed by a moderate heat treatment at around 750 °C. Finally the cavities are slightly (20 μm) chemically etched and cleaned by high pressure water. Figure 25 shows that gradients around 30 MV/m with low field emission could be reached. All tests were finally limited by thermal quench.

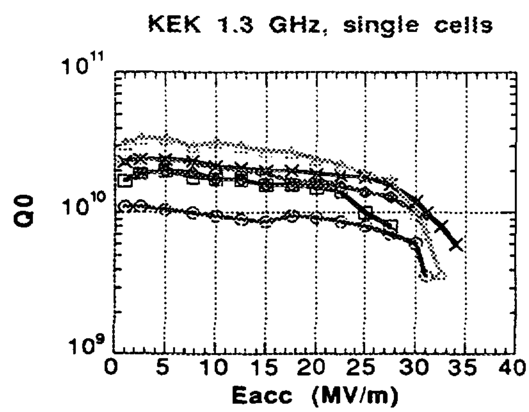


Fig. 25 Recent single-cell measurements (1.3 GHz) at KEK. All fields were limited by quenches.

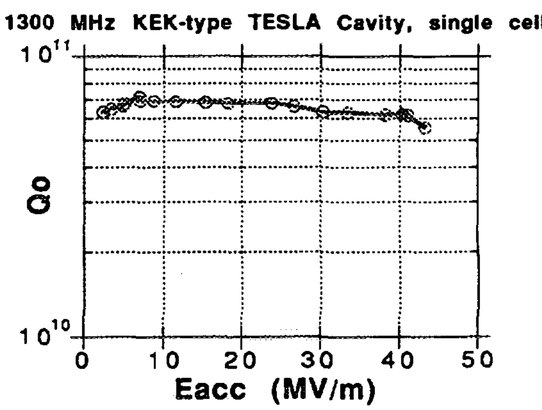


Fig. 26 Best single-cell measurements (1.5 GHz) at CEBAF (T = 1.5 K). The field was limited by quench.

At CEBAF [5], single-cell niobium cavities are post purified by the 1400 °C Ti treatment, followed by chemical etching (150 μm) and high-pressure water rinsing. The best measurement is shown in Fig. 26. The measured gradient of 43 MV/m comes close to the expected theoretical limit of 50 MV/m (with the corresponding surface magnetic field equal to the thermodynamic critical field). It should be noted that this high field was measured at the rather low temperature of 1.5 K. Many other tests with single cells reached gradients around 30 MV/m with only slight field emission.

At Cornell the technique of high power processing has been developed [19]. Several 5-cell cavities (TESLA shape; the furnace for the Ti treatment accepts no larger structures than 5 cells) have been processed. Figure 27 shows the improvement after high power processing. As a rule of thumb, the onset of field emission could be raised by a factor of 2.

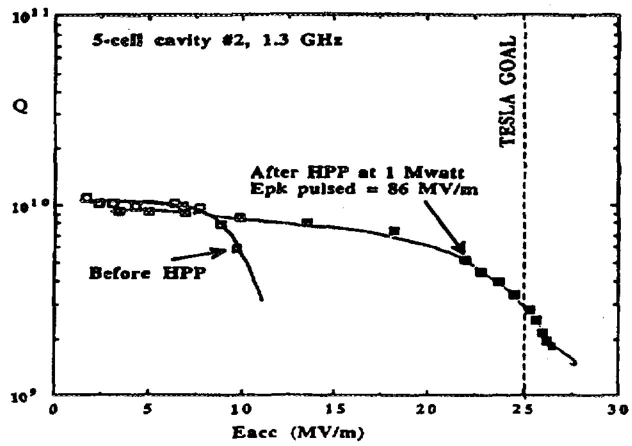


Fig. 27 Improvement of cavity performance by high power processing (HPP). Measurement of 5-cell, 1.3-GHz cavities at Cornell.

At DESY the infrastructure for processing of TESLA cavities was completed in 1994 and several 9-cell TESLA cavities (1.3 GHz) have already been measured. The standard procedure is:

- chemical etching of 100 μm to clean the niobium surface,
- Ti treatment at 1400 °C to improve the thermal conductivity,
- chemical etching of 100 μm to remove the Ti layer,
- high pressure water (100 bar) spray to clean the surface,
- high power processing to do in situ cleaning.

Figure 28 shows the measured Q vs. E_{acc} curve of the first series production cavity under continuous wave operation. The specified values for the test linac are $Q = 3 \times 10^9$ at E_{acc} = 15 MV/m. In TESLA the superconducting cavities will be pulsed with 800 μsec flat top at 25 MV/m, rep. rate 10 Hz. These conditions could be established with the first series production cavity attached to a high power input coupler (see Fig. 29).

ACKNOWLEDGEMENTS

Exchange of information and stimulating discussions with my colleagues from CEBAF, CERN, DESY, Cornell, KEK and Saclay are gratefully acknowledged.

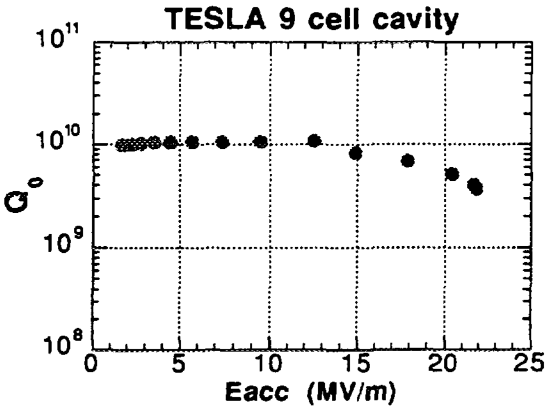


Fig. 28 Measurement of the first 9-cell TESLA cavity (1.3 GHz) under continuous wave operation

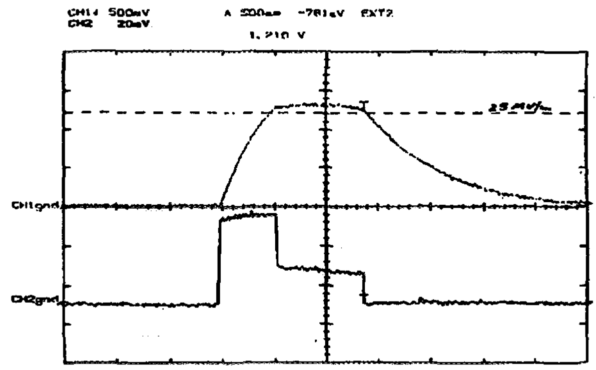


Fig. 29 Pulsed measurement of the first 9-cell TESLA cavity. The beam loading during the 800 μsec flat top was simulated by shaping the klystron power (upper curve E_{acc}, lower curve klystron power).

REFERENCES

[1] G. Cavallari et al, Acceptance Tests of Superconducting Cavities and Modules for LEP from Industry, Proc. EPAC 3, London (1994), p. 2042 - 2044.

[2] B. Dwersteg et al, Operating Experience with Superconducting Cavities in HERA, Proc. EPAC 3, London (1994), p. 2039 - 2041.

[3] S. Noguchi et al, Recent Status of the TRISTAN Superconducting RF System, Proc. EPAC 3, London (1994), p. 1891 - 1893.

[4] M. Castellano et al, Early Results of Beam Acceleration in the SC Linac LISA, Proc. EPAC 1, London (1994), pp. 731.

[5] A. Hutton for the CEBAF Commissioning Team, Commissioning at CEBAF, Proc. EPAC 1, London (1994), p. 15 - 19.

[6] M. S. McAshen et al., Appl. Phys. Lett. 22 (1973) 605 and J.R. Calarco et al, IEEE Trans. Nucl. Sci. NS-24 (1977) 346.

[7] S. Döbert et al., Status of the S-DALINAC and Experimental Developments, Proc. EPAC 1, London (1994), 719 - 721.

[8] R.C. Pardo et al, Operational Status of the Uranium Beam Upgrade of the ATLAS Accelerator, Proc. PAC 93, 3, (1993) p.1694 - 1699.

[9] G. Fortuna et al., Nucl. Instr. Methods A328 (1993) 236 and L. Badan et al, Contribution to the EPAC, London (1994).

[10] T. Ishii, M. Shibata, S. Takeuchi, Nucl. Instr. Methods A328 (1993) 231.

[11] H. Padamsee et al, Development and Tests of a Superconducting Cavity for High Current Electron Storage Rings, Proc. EPAC 3, London (1994), p. 2048 - 2051.

[12] T. Takashaki et al, 9th Symposium on Accelerator Science and Technology, Tsukuba, Japan, (1993) pp 327.

[13] M. Leenen for the TESLA Collaboration, The Infrastructure for the TESLA Test Facility (TTF) - A Status Report, Proc. EPAC 3, London (1994), p. 2060 - 2062.

[14] S. Ramo, J.R. Whinnery, T. van Duzer, Fields and Waves in Communication Electronics, Wiley Internat. Edition (1965).

[15] U. Laustroer, U. van Rienen, T. Weiland, URMEL and URMEL-T User Guide, DESY M-87-03 (1987) and T. Weiland, TBCI and URMEL - New Computer Codes for Wake Field and Cavity Mode Calculations, IEEE Trans. on Nuclear Science, vol NS 30 (1983) p. 2489 - 2491.

[16] M. Bartsch et al, "MAFIA Release 3. The New Version of the General Purpose Electromagnetic Design code Family," Proc. of EPAC Conf., Vol 1, pp 249-25 (1990).

[17] U. Klein, D. Proch, Proc. Conf. of Future Possibilities for Electron Accel., ed. JS McCarthy, RR Whitney. Charlottesville: Univ. Virginia (1979) p. N1-17.

[18] H. Padamsee, IEEE Trans. Mag. 19, (1983) 1322.

[19] C. Crawford et al, Achieving the TESLA Gradient of 25 MV/m in Multicell Structures at 1.3 GHz, Proc. EPAC 1, London (1994), p. 37 - 39.

[20] S. Noguchi, private communication.

[21] P. Kneisel, private communication.

FIELD EMISSION IN RF CAVITIES

B. Bonin
Commissariat à l'Energie Atomique, DSM/DAPNIA, France

Abstract
Electron field emission limits the accelerating gradient in superconducting cavities. The present paper shows how and why it is an important problem. The phenomenology of field emission is then described, both in DC and RF regimes. In a third part, the merits of a few plausible "remedies" to field emission are discussed.

1. WHY IS IT A PROBLEM ?

When exposed to an intense, properly oriented electric field, a conducting surface can emit electrons. In the case of an accelerating cavity, these electrons are accelerated by the RF field in the cavity. In normal conducting cavities, "field emission" must be avoided because it is a precursor to vacuum breakdown and is likely to cause dark current. Superconducting cavities are even more sensitive to field emission: in a superconducting cavity, even the small additional dissipation of RF power due to the electron loading of the cavity may correspond to a significant and undesirable degradation of the cavity Q-value, and an increase of the cavity cryogenic consumption (Fig. 1).

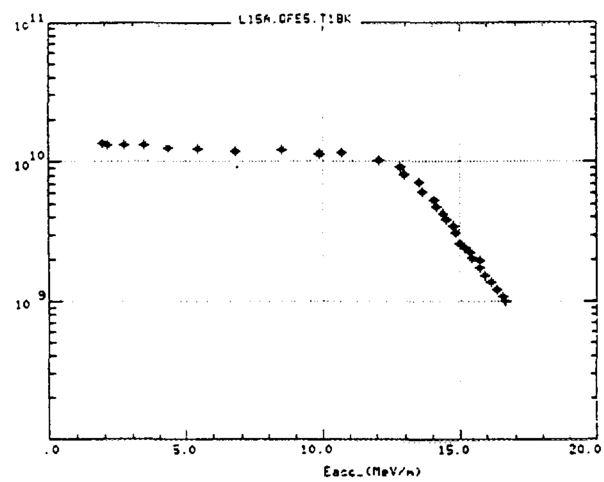


Fig. 1 Q-value vs accelerating gradient for a typical superconducting cavity. In this example, Q-degradation beyond E_{acc} = 12 MV/m is due to electron loading.

Furthermore, the emitted electrons follow complicated trajectories inside the cavity (Fig. 2). If the cavity is excited on the TM010 mode, the electron trajectories remain in the same meridian as their starting point. The electrons ultimately land on the cavity wall with an energy roughly proportional to the accelerating gradient (typically a few hundreds of keV for an accelerating gradient of 10 MV/m in a 1.5-GHz single-cell cavity). The corresponding energy deposit is usually rather local. It creates a hot zone on the superconducting surface, which can be identified in the temperature maps of the cavity as a ridge parallel to a cavity meridian (Fig. 3). This energy deposit is liable to hamper the thermal stability of the cavity, thereby limiting the accelerating gradient. The heating is also accompanied by copious X-ray emission originating from the electron landing point.

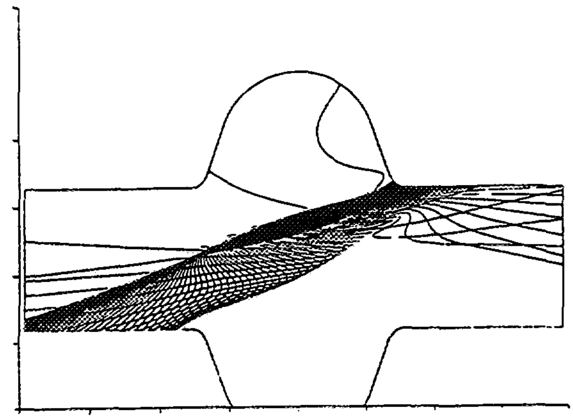


Fig. 2 Trajectories of electrons emitted inside the cavity

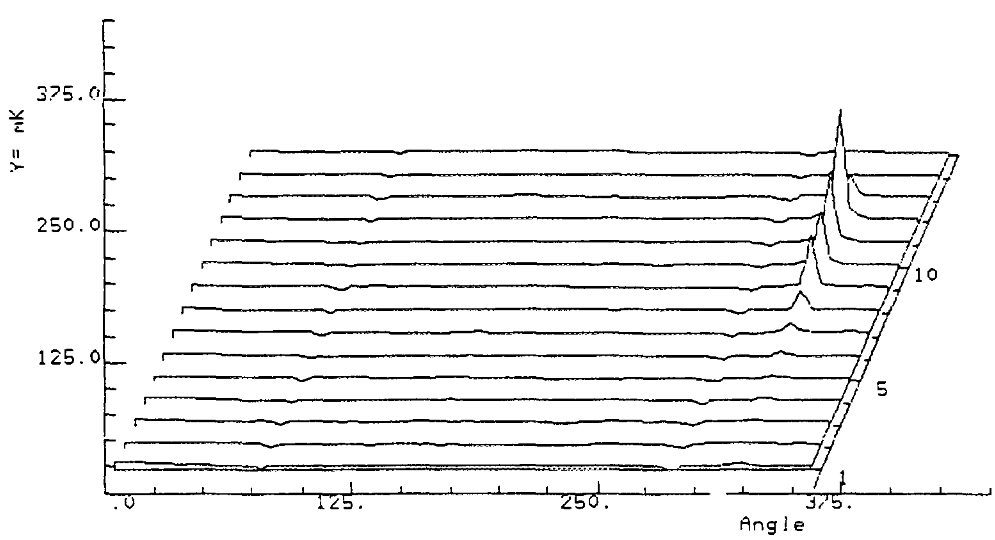


Fig. 3 Temperature map of a superconducting cavity plagued by field emission

The order of magnitude of the field emission current able to limit superconducting cavity performance is around a few µA. The field emission "disease" is very common indeed: for example, 60% of the CEBAF superconducting cavities are limited in gradient by field emission. This alone would justify the research efforts from many laboratories, aiming at a better understanding of the phenomenon, and at finding ways to avoid or to cure it. However, field emission is not always considered as a harmful phenomenon. It is exploited in vacuum electronics to produce cold electron sources. Here, the efforts go in the opposite direction, aiming at finding ways to enhance and to stabilize the emission. The author found it especially useful to compare the results from both communities, ie those who promote and those who fight against field emission.

2. PHENOMENOLOGY OF FIELD EMISSION

Fowler-Nordheim theory

In a metal, electrons are usually prevented from escaping by a potential barrier separating the Fermi level in the metal and the vacuum level. When an electric field is applied, the vacuum level takes a slope, and the barrier becomes triangular (Fig. 4). The width of the barrier decreases with increasing field. When it becomes thin enough, the probability for electrons to tunnel through the barrier becomes non negligible, and a field emission current arises. The height of the barrier is known as the metal work function, and amounts to 4–5 eV, according to the nature of the metal. Given this height, Fowler and Nordheim [1] calculated the tunnel probability through the barrier and derived the following relationship between the current density j and the the electric field E, for a perfect metal with a Fermi-Dirac electron density at 0 K:

$$j = \frac{A.E^2}{\Phi} \exp(-B\frac{\Phi^{3/2}}{E}) \tag{1}$$

If j is expressed in $A.m^{-2}$, E in $V.m^{-1}$, and Φ in eV, the constants A and B are respectively:

$$A = 1.54.10^{-6}$$
$$B = 6.83.10^{9}.$$

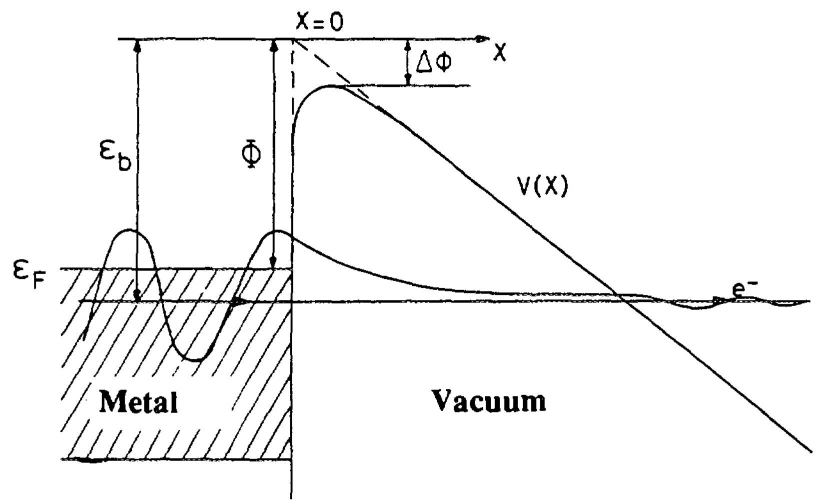


Fig. 4 Energy diagram of a metal-vacuum interface. The electrons fill the energy levels of the conduction band, up to the Fermi level. If an electric field is applied to the surface, the potential barrier seen by the electrons becomes triangular.

This law predicts a negligible current for fields below a few GV/m, and a very steep rise of the current above this limit. For very clean metal surfaces of restricted area (the tip of a sharp needle), the F-N law has been verified in a large domain of current densities, despite some difficulties to evaluate properly the magnitude of the local electric field in the metal vicinity. For current densities larger than $j \approx 10^{12}$ A/m^2, the agreement between theory and experiment apparently breaks down [2], probably because space charge phenomena arise, and because tip breaking due to the Joule effect tends to modifiy the surface morphology.

If the field emitted current density is maintained well below 10^{12} A/m^2, emission appears to be stable in time if the tip is maintained in ultra high vacuum. In vacua less than 10^{-8} mbar, species adsorbed on the tip can modify the shape of the potential barrier seen by the electrons, and the current density fluctuates [3].

Enhanced field emission (EFE)

For large area electrodes with a less well characterized surface, the phenomenology appears to be entirely different [4]. A significant current can arise for electric fields as low as a few MV/m. Many studies have shown that the emission is localized on microscopic emitting sites [5]. Each site seems to obey a phenomenological F-N law, with effective parameters: there is still a linear relationship between Log(I/E^2) and $1/E$ (Fig. 5), but if one assumes an unchanged value for the work function Φ of the emitting site, the local electric field seems to be enhanced by a factor β of the order of 100, and the effective emission is restricted to small, but very variable areas S [6]. The current I emitted by one site is then given by:

$$I = S.\frac{A.\beta^2.E^2}{\Phi}\exp(-B\frac{\Phi^{3/2}}{\beta.E}) \qquad (2)$$

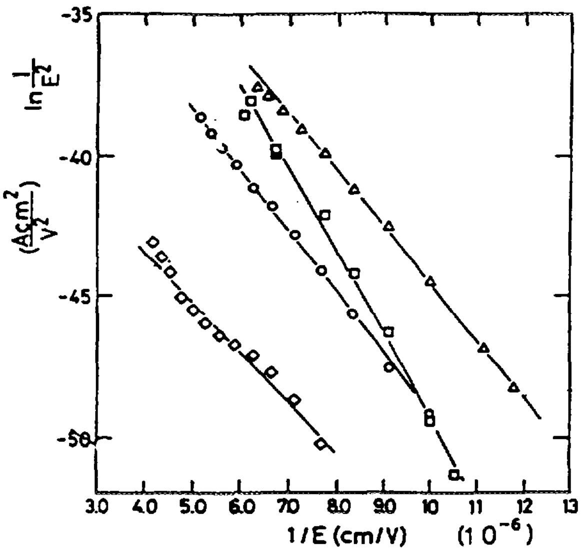


Fig. 5 Field emission characteristics of four emitting sites (from P. Niedermann, Ref. [7])

Recent investigations have correlated these emitting sites with surface defects, identifiable by electron microscopy [7]. (More precisely, not all the surface defects behave as emitters, but all emitters could be identified as previously existing surface defects.)

It was proved that particulate contaminants were powerful field emitters, with somewhat unstable field emission characteristics. Irregularly shaped conducting particles seem to emit much more than smooth ones [8]. Scratches, or other geometrical defects of the surface also behave as field emitters, with a more stable emission.

On a given surface, there is a considerable scatter in emitter characteristics. This results in a widely spread hierarchy of emitters. Consequently, even on surfaces of very large area, there is only one or a few emitter(s) whose current dominates over all the other ones.

Microscopic investigation of emitting sites reveals thermal effects associated with field emission. After exposure to electric field and emission of 1-100 μA, emission of some sites is often observed to drop.abruptly. This change appears to be associated with the sudden melting of the emitter (Fig. 6). Frequently, this melting takes place at the sharpest protrusion on the surface, thus suggesting that the previous active emitter was located there. Craters of molten material are also frequently observed [9], and seem to be correlated with a micro breakdown, and with the death of an emitting site. New emitters are often formed on the crater rim.

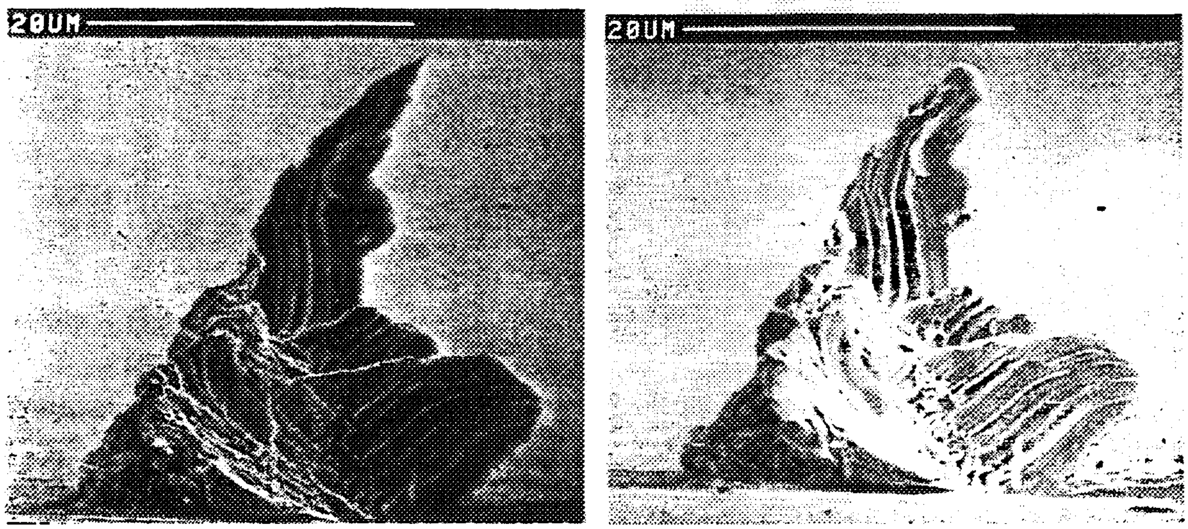


Fig. 6 Microphotograph of an emitting site: a) before emission; b) after emission. Note the apex melting.

Enhanced field emission in RF regime

Enhanced field emission has been extensively studied in DC regime, but has received much less attention in RF regime. So far, we have no indication that EFE should obey a different mechanism in DC and RF regimes. The typical tunneling time through a barrier of height h can be estimated via the uncertainty principle:

$$\tau \approx \frac{h}{\Phi} \approx 2.10^{-16} \text{ s.}$$

This time is short compared with the RF period, it is thus legitimate to assume that the field emission process is instantaneous and occurs in a "quasi static" electric field. With this approximation, the average current emitted in RF regime can be calculated, using Eq. (2):

$$\begin{aligned} < I > &\approx S.\frac{C.\beta^{2.5}.E_{peak}{}^{2.5}}{\Phi}\exp(-B\frac{\Phi^{3/2}}{\beta.E_{peak}}) \\ C &= \frac{2.\sqrt{2}}{3.\pi}.\frac{A}{\sqrt{B.\Phi^{3/2}}}. \end{aligned} \tag{3}$$

Note that the law obtained analytically is very similar to the F-N law, except the 2.5 exponent in the prefactor. This law is in excellent agreement with experiment [10], with effective parameters β and S quite similar to the ones.extracted from DC measurements. This certainly adds credit to the conjecture that EFE obeys the same basic mechanism in DC and RF regimes.

The geometrical model

Many explanations of EFE have been proposed [11]. In this course, presentation of the simplest one is probably sufficient. Many geometrical defects such as scratches or conducting particles have been shown to emit mainly because there is a geometrical enhancement of the electric field at the apex of a protrusion of microscopic or even nanometric scale [8].

The degree of generality of this mechanism is still under discussion. It does not account for the role of adsorbates, which are noted for their capability to modify field emission, not for the odd energy spectra which were observed by Latham on some emitters [12]. However, there is no doubt that at least a large fraction of the emitting sites obey this simple mechanism.

Field emission from superconductors

Is field emission the same for normal metals and for superconductors? The gap brought by superconductivity in the electronic density of states is too narrow to have a significant consequence on the field-emitted current. Experimentally, no difference of field emission from a niobium tip was noted at the superconducting transition [13].

3. HOW TO AVOID FIELD EMISSION

Instabilities

EFE is a rather unstable phenomenon: emitters are known to switch "on" and "off", and the field-emitted current often fluctuates. It is interesting to analyze these instabilities, because there is some hope to use them to destroy emitters.

Firstly, one can expect mechanical instabilities of the emitters. Because of the large field enhancement at the emitter apex, the electrostatic pressure

$$p = \varepsilon_0 . E^2_{microscopic}$$

gets close to the yield stress of usual metals. Necking or even breaking of the apex can thus occur, and the subsequent modification of the surface geometry results in changes in its field emission characteristics [14]. If the emitting site is a particle sitting on the surface, the same electromagnetic pressure can overcome adherence forces. Larger particles are loosened at lower electric fields. Note that the force exerted by the substrate on the particle is always repulsive, independent of the sign of the electric field, so that this "cleaning" mechanism is effective in RF as well as in DC regime [15]

The second kind of instability is of thermal origin. Several mechanisms can affect the emitter temperature [16]. Joule heating due to the field emitted current through the very narrow emitter apex does not make a significant contribution to the elevation of the apex temperature, unless the current density exceeds 10^{12} A/m^2. In RF regime, Joule heating due to the current induced through the emitter can contribute significantly to increasing the temperature of the emitter. Ion bombardment can also heat up the emitter. The process may be initiated by ionization of species desorbing from the surface. Ions accelerated by the field can deposit their kinetic energy on the metal surface at the end of their flight. This heating then promotes further desorption and ion bombardment. Further heating may cause vaporization of the metal itself, accompanied by a considerable enhancement of the vapor and ion density in the vicinity of the emitting site. This process is believed to lead to explosive phenomena and vacuum breakdown [17]. It leads generally to the destruction of the original emitting site, and to the formation of craters (Fig. 7). All these heating mechanisms are more effective if emitters are particulate contaminants in poor thermal contact with the substrate.

The emitters are thus mechanically and thermally fragile. This has been a matter of concern for the designers of electron sources based on field emission, but serves also as a guideline to find effective processing recipes to minimize field emission in RF cavities.

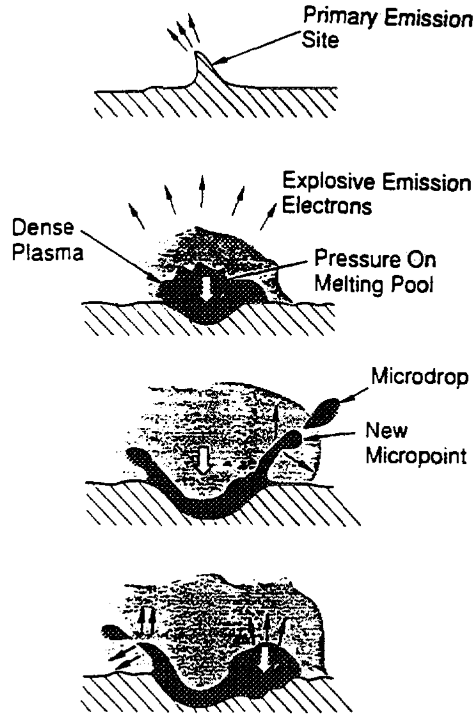


Fig. 7 Formation of a microbreakdown (from Ref. [17])

Helium processing

It can be envisaged to modify the emitters by ion bombardment. This method is used with some success in superconducting cavities under the generic name of "Helium processing". Helium gas at low pressure (10^{-5} mbar) is introduced in the cavity and RF field is applied. A reduction of the field emission in the cavity is often observed [18], probably due to the bombardment of the emitting site by the helium ions. This bombardment may cause a gas desorption on the emitter, or a destruction of the emitter by sputtering. In spite of its effectiveness, helium processing has severe drawbacks, related to the risk of deposition of sputtered contaminants on the cavity surface. Degradation of the cavity Q-value has often been observed after helium processing.

RF processing

It has long been recognized that both field emission current and breakdown probability decrease by exposing a virgin surface to an intense electric field. This kind of processing is currently applied in DC regime to upgrade the breakdown threshold of metal surfaces in vacuum, and in RF regime on normal conducting cavities. In most of these cases, this processing is made "in situ", i.e, the surface is not reexposed to atmosphere after the treatment. When it is reexposed, the benefits of the processing are partly lost.

High-field processing has also been applied to superconducting cavities. Here, specific problems arise. Despite the long cavity filling time, the surface must be exposed to intense electric fields in a short time, if cavity quench is to be avoided. This short pulse time, combined with a severe electron loading during the pulse, requires large RF peak powers and an adjustable RF coupling.

In a remarkable series of experiments, the Cornell laboratory has established high-peak-power processing (HPP) as a very effective curative treatment against field emission [19] (Fig. 8). The parameter which determines the degree of reduction of field emission is the maximum electric field reached during the HPP stage. For example, by using processing fields as high as 60 MV/m, it is possible to eliminate field emission for CW fields of the order of 30–40 MV/m. The effectiveness of HPP probably lies in its capacity to heat the emitters selectively, initiating breakdowns which ultimately burn the emitters. Craters are formed during the treatment [9]. It has been shown that these craters themselves may eventually emit, but these can be processed in turn in the same way. The efficiency of the treatment saturates when the emission from the new craters equals the emission of the formerly destroyed emitters.

Despite impressive laboratory results, the usefulness of HPP might be reduced if the processed surface has to be re-exposed to air, because of the risk of ulterior particulate contamination. If HPP is done in situ, i.e without any re-exposition to air, it remains to be seen to what extent this kind of processing is really useful on real accelerators, with limited RF power and fixed couplings.

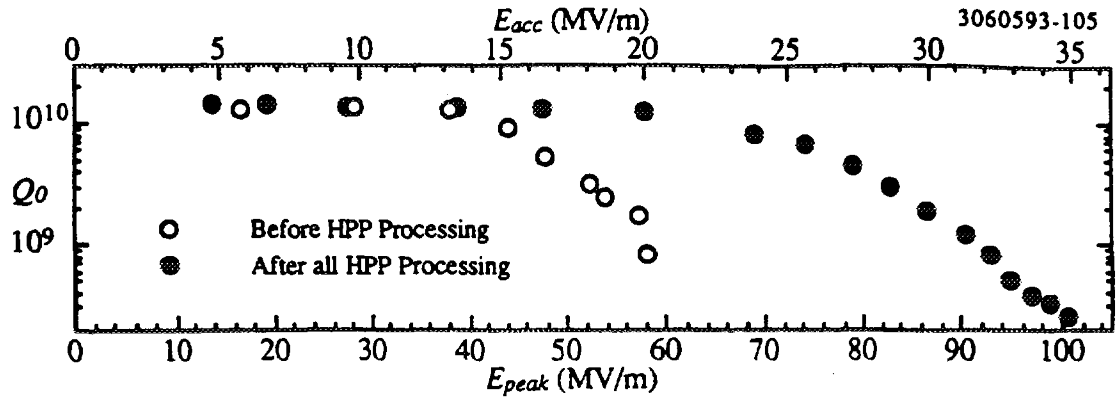


Fig. 8 High-peak-power processing can cure field emission in superconducting cavities (from Ref. [19]).

4. CONCLUSION

Clearly, the first cause of field emission in superconducting cavities is particulate contamination. Cleanliness is an indispensable prerequisite in avoiding enhanced field emission from extended surfaces. Effective removal of micron-sized particles is a rather difficult task, because the mechanical forces applied to the particle by the cleaning process decrease faster than the adherence forces when the particle size decreases. Advanced cleaning techniques like high-pressure rinsing [20] have been used to clean superconducting cavities, with a statistically significant reduction of field emission. The effectiveness of this treatment may be limited by eventual contaminations occurring later in the cavity history, but this kind of cleaning is probably one of the most precious preventive treatments available against enhanced field emission.

Systematic use of both preventive and curative remedies (i.e. ultra-clean treatments and high-voltage processing techniques) holds the promise of greatly improving the superconducting cavity performance level and reproducibility.

REFERENCES

[1] R.H. Fowler and L. Nordheim, Proc. Roy. Soc. London A 119, 173 (1928)

[2] W.P. Dyke and J.K. Trolan, Phys. Rev. 89, 799 (1953)

[3] C.B. Duke and M.E. Alferieff, J. Chem. Phys. 46, 923 (1967)

J.W. Gadzuk, Phys. Rev. B1, 2110 (1970)

R. Morin et al. Surf. Sci. 155, 187 (1985)

[4] R.J. Noer, Appl. Phys. A 28, 1 (1982)

[5] B.M. Cox and D.E.J. Wort, Vacuum 20, 453 (1972)

C.S. Walters, M.W. Fox, R.V. Latham, J. Phys. D Appl. Phys. 12, 969 (1979)

B.M. Cox, J. Phys. D Appl. Phys.8, 2065 (1975)

[6] H. Padamsee, "High Voltage Vacuum Insulation in RF Superconducting Cavities", p 431 in "High Voltage Vacuum Insulation" edited by R.V. Latham, Acad. Press (1995)

M. Jimenez et al., J. Phys. D: Applied Phys. 27, 1038 (1994)

Ph. Niedermann, PhD thesis, University of Geneva, (1986)

[7] Ph. Niedermann, PhD thesis, University of Geneva, (1986)

P. Niedermann et al., J. Appl. Phys. 59, 892 (1986)

M. Jimenez et al., J. Phys. D: Applied Phys.26, 1503 (1993)

B. Bonin, Vacuum 46, 907 (1995)

[8] M. Jimenez et al., J. Phys. D : Applied Phys. 27, 1038 (1994)

[9] D. Moffat et al., Particle Accelerators 40, 85 (1992)

[10] S. Noguchi, Nuclear Instruments and Methods 179, 205 (1981)

C. Lyneis, Proc. 1st Workshop on RF Superconductivity, Karlsruhe, Germany, KFK report 3019, ed. M. Kuntze (1980)

W. Weingarten, Proc. 2nd Workshop on RF Superconductivity, Geneva, Switzerland, ed. H. Lengeler, CERN (1985)

[11] N.S. Xu, "The Physical origin of prebreakdown electron pin-holes", p115 in "High Voltage Vacuum Insulation" edited by R. V. Latham, Acad. Press (1995), and refs. therein

[12] N.K. Allen and R.V. Latham, J. Phys. D: Appl. Phys. 11, 657 (1978)

[13] A. Léger, J. de Physique 29, 646 (1968)

[14] S.I. Shkuratov et al., Le Vide, les Couches Minces 272, 209 (1994)

[15] R.V. Latham, "Microparticle phenomena" p 229 in "High Voltage Vacuum Insulation", edited by R.V. Latham, Acad. Press (1995)

J. Tan, H. Safa, B. Bonin, Proc. 7th Workshop on RF Superconductivity, Gif sur Yvette, France, ed. B. Bonin (1995)

[16] L.W Swanson et al., Phys. Rev. 151, 327 (1966)

N.M. Miskovsky et al., J. Vac. Sci Technol. B11, 366 (1993)

[17] G.A. Mesyats et al., IEEE Trans. Elec. Insul. EI 18, 218 (1983)

E.A. Litvinov et al., IEEE Trans. Elec. Insul. EI 20, 683 (1985)

G.N. Fursey et al., IEEE Trans. Elec. Insul. EI 20, 659 (1983)

J.W. Wang, Ph. D., SLAC report 339 (1989)

[18] C. Athwal and W. Weingarten, CERN/EF/RF 84-7 (1984)

[19] J. Graber, PhD thesis, Cornell University, (1993)

J. Graber et al, 5th Workshop on RF Superconductivity, DESY, Hamburg, Germany, DESY report M 92 01 ed. D. Proch (1992)

J. Kirchgessner et al. Part. Acc. Conf. PAC 1993, Washington CLNS 93/1209

J. Graber, ibid., CLNS 93/1215

[20] D. Bloess, private communication

P. Kneisel et al, Proc. 6th Workshop on RF Superconductivity, CEBAF, USA, ed. R. Sundelin (1994)

C.Z. Antoine et al., ibid., p 456

COUPLERS FOR CAVITIES

E. Haebel
CERN, Geneva, Switzerland

Abstract
The first part of this chapter is on higher-order-mode couplers. To introduce the subject a coupling approach is examined which makes use of the beam tubes as wave guides with internal vacuum compatible RF loads. Then it is shown how coupling can be considerably enhanced if couplers are themselves regarded as resonators. Examples of couplers used in the LEP and HERA storage rings serve to illustrate this approach. In this first part we regard the cavity as a multi-frequency generator furnishing, via the coupler, power to a room temperature load. In the second part the direction of power flow is reversed. Now, at a single frequency, the cavity acts as a low-temperature load to which RF power has to be delivered and which can be considerably higher than the higher-order-mode power.

1. INTRODUCTION

Cavities are the 'motors' of accelerators, passing energy to the charged particle beams. But charges passing through a cavity are 'active devices' and can either receive energy from, or furnish energy to, a cavity. In an accelerator the first process is the required one though the second inevitably takes place and may hamper the efforts to produce a high quality beam. Couplers are needed to replenish the cavity energy which had been lost to the beam. Such main or power couplers may have to handle very large amounts of RF power, up to the MW level.

Couplers are also needed to mitigate the unwanted interactions between particles and cavities. They then are called *higher-order-mode* (HOM) couplers or dampers. In fact, to be more precise, couplers are RF devices which allow energy to be exchanged with the *modes of oscillation* of a cavity. So, at the beginning of this coupler chapter a short discussion of modes is needed.

2. MODES AND THEIR CLASSIFICATION

The ideal cavity is a closed volume, completely surrounded by a metallic boundary as indicated in Fig. 1. For infinite conductivity of the metal walls a closed volume can store electromagnetic field energy U for infinite time. Storage is in the form of *free oscillations* at eigenfrequencies ω_n, the spectrum of which depends on the size and shape of the volume. At each ω_n, field energy U_n changes periodically between its two possible forms, electric and magnetic, and the patterns of the corresponding fields E_n and H_n are characteristic of each oscillation mode. Computer codes are available (e.g. MAFIA) to calculate the eigenfrequencies and fields of such modes but for simple boundary shapes analytical solutions also exist. Two examples are shown in Fig. 2.

For accelerator cavities it is usual to classify the cavity modes into two groups: the first contains only a single member, the mode used for particle acceleration, the second group contains all the others. Since, by cavity design, the accelerating mode usually has the lowest frequency, one calls it the *fundamental mode* (fm) and all the others the *higher-order modes*. Another name for the latter is *parasitic modes*, indicating that they are deleterious and unwanted. I will also talk about longitudinal and dipole modes. Cavities for accelerators have axial symmetry. Then in a cylindrical coordinate system ϕ, r and z, longitudinal modes have fields with no ϕ dependence whereas for dipole modes fields vary with cos ϕ.

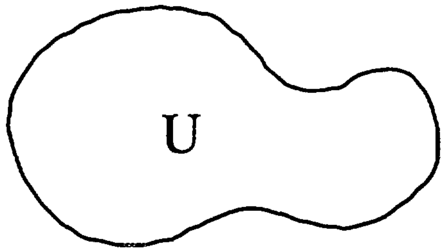


Fig. 1 Example of an ideal cavity

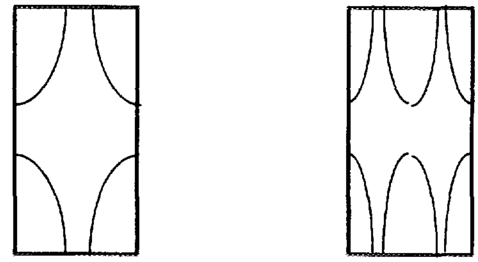


Fig. 2 The TM011- and TM012-mode E-field patterns of a 'pillbox' cavity

3. BEAM TUBES AS HOM DAMPERS

To integrate a cavity into an accelerator one has to attach it to the beam tubes These have little influence on modes with eigenfrequencies lower than the cut off frequency of wave propagation in the beam tubes. Cavity modes with higher frequencies are often said to be *propagating*. It is assumed that, by exciting waveguide travelling-wave modes, such cavity modes will quickly loose their stored energy i.e. become very effectively damped.

The interesting question is now: Can *all* HOM of a cavity be made propagating? What beam tube diameter is required to achieve this goal? To get a first idea we do some back of the envelope calculations. They become simple if the cavity has the shape of a pillbox with radius a and length l. For this simple shape the fundamental TM_{010} resonance is at the cut off frequency of the TM_{01} waveguide mode in a tube of radius a. This mode has the propagation constant:

$$k^2 = \left(2\pi/\Lambda\right)^2 = \left(\omega/c\right)^2 - \left(2.41/a\right)^2$$

At the cut off the propagation constant is zero and hence at the TM_{010} resonance frequency f_0:

$$\omega_0/c = \left(2.41/a\right)$$

Note that f_0 is independent of the pillbox length 1. Cavities for ultra-relativistic particles have lengths around $\lambda_0/2$ where λ_0 is the free space wavelength at f_0. The HOM next to the fundamental then is the TE_{111} dipole mode. At its resonance the cavity length is one half of the TE_{11} mode wavelength. Working again with the propagation constant we can therefore determine its frequency from

$$\left(2\pi/\Lambda\right)^2 = \left(2\pi/2l\right)^2 = \left(\omega_1/c\right)^2 - \left(1.841/a\right)^2$$

The beam tube radius b, which at ω_1 just allows propagation of the TE_{11} mode can then be calculated since

$$\omega_1 / c = 1.841 / b \ .$$

Writing finally $l = p\ \lambda_0/2$ and rearranging we obtain for the ratio a/b of cavity and beam tube radius:

$$(a/b)^2 = 1 + (1/p)^2 (2.41/1.84)^2$$

To have, according to this formula, propagation of all HOM, the beam tubes must be much wider than conventionally used and a reasonable value of $a/b = 2$ is only obtained if $p = 0.75$ i.e. for a cavity *shorter than* $\lambda_0/2$. This prevents us using such a HOM damping technique for multicell π-mode cavities. Even for a single cell the prognostic from this formula is too optimistic. A check with a code like URMEL reveals [1] that attaching wide beam tubes to the cell *lowers* its TE_{111} frequency, so the mode 'refuses' to propagate *however wide the beam tubes are made!*

3.1 Confined and trapped modes

Figure 3 shows E-field patterns of a wide-beam-tube design [2] as studied at CERN for possible use in LHC. The FM frequency is 400 MHz and the beam tube diameter 30 cm, about one half of the cavity diameter. The TE_{11} and TM_{01} waveguide cut off frequencies are 586 and 764 MHz respectively and, as we see in comparing frequencies, in addition to the FM also the TE_{111} and TM_{110} dipole modes remain *confined* to the cavity. The TM_{011} mode (first harmonic of the fundamental) just reaches the region of propagation. But even a frequency above cut off is no guarantee of sufficient mode damping. At about three times the fundamental mode frequency (at 1232 MHz) we see a HOM which excites only a very small field in the beam tubes. For a cell length of $\lambda_0/2$ the excitation would be even weaker. We have here the example of a *trapped mode* which, though *nominally propagating*, may remain weakly damped [3]. Trapped modes become a real headache when designing multicell cavities [4]. They also have smaller cell-to-cell coupling than other modes so, with increasing cell numbers, field distributions become very sensitive to perturbations.

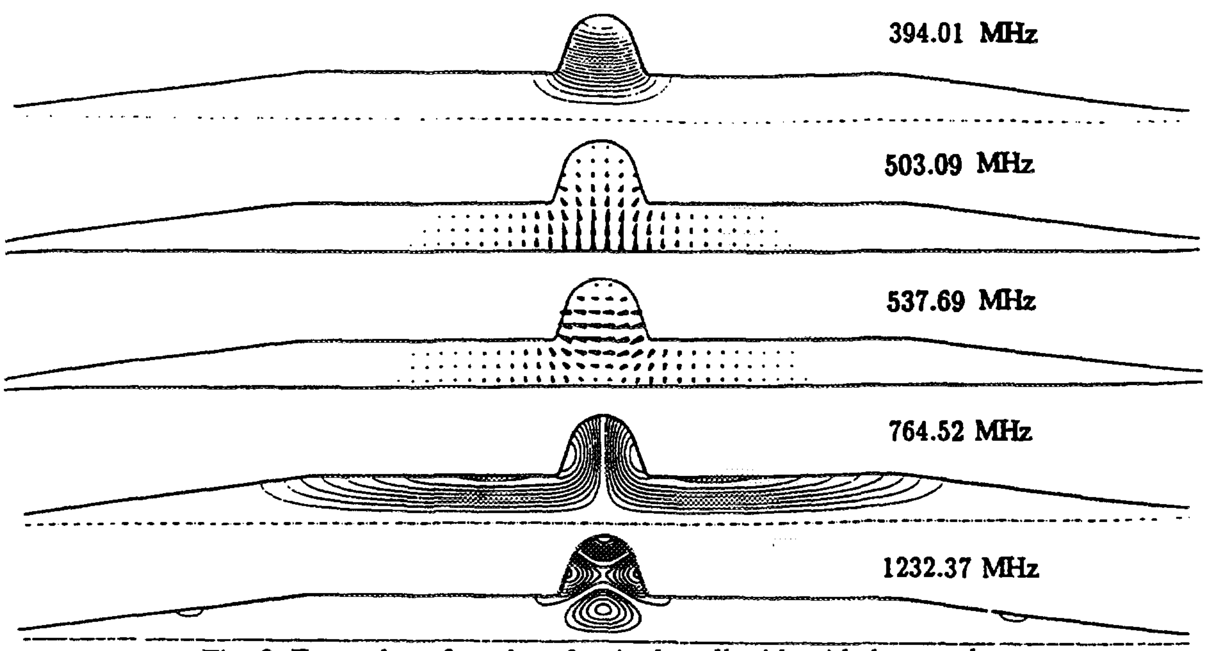


Fig. 3 Examples of modes of a single cell with wide beam tubes

3.2 Deconfining the TE_{111} mode

To escape from the dilemma of the first two dipole modes not propagating, two solutions were proposed and developed. The first, as pursued at KEK in Japan, simply widens the beam tube further, but only behind a coupling aperture to the cell[1] as sketched in Fig. 4. However, not only the TE_{11} cut off is lowered but also, unnecessarily, the TM_{01} cut off, so at the RF absorbing material the FM amplitude might become too high. Therefore, at Cornell the beam tube is widened azimuthally only in sectors, which creates the geometry of a ridged waveguide (see Fig. 4) and lowers particularly the waveguide's dipole mode cut off. These techniques are used for one of the beam tubes, the other tube (see Fig. 5) remaining a simple one. The use of different beam tube sections on the two sides of the single cell helps to avoid trapped modes,

[1] This concept was already applied by the Wuppertal team in their designs of 3-GHz structures.

measurements on a copper model at Cornell demonstrating efficient damping of all significant modes to external Q values smaller than 50.

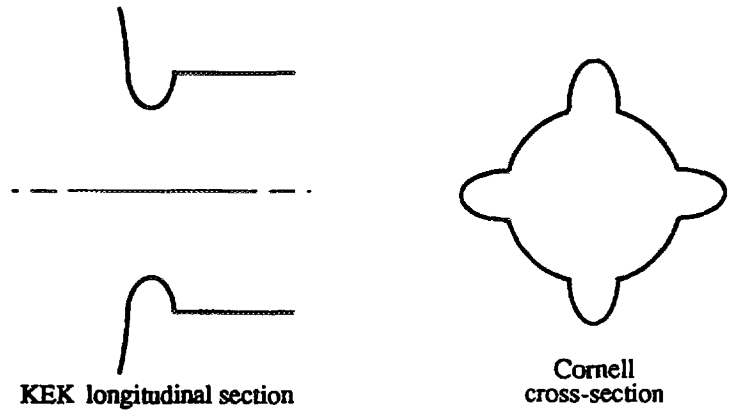


Fig. 4 Beam tube forms at KEK and Cornell

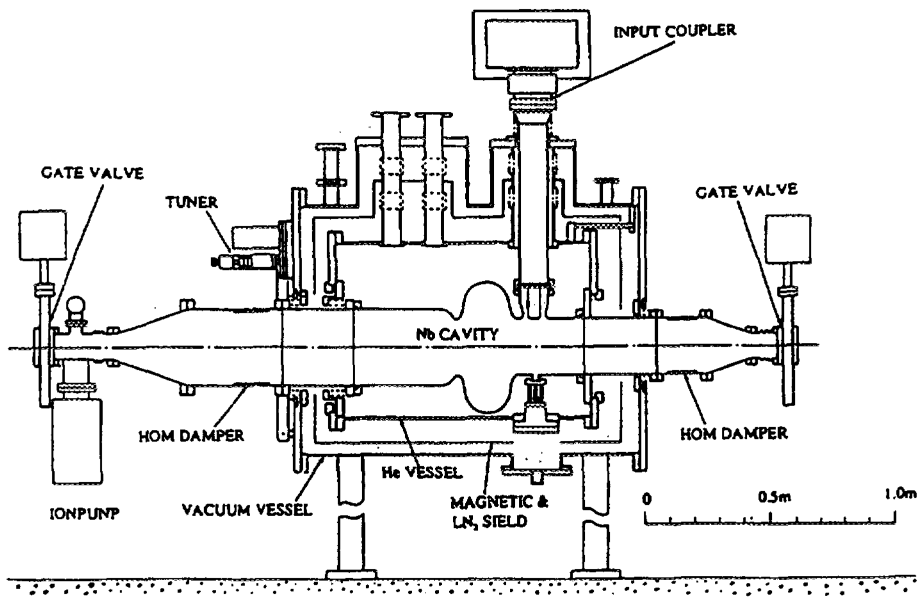


Fig. 5 KEK B-factory accelerating module

3.3 The beam tube RF load

Beam tube HOM couplers have to be terminated into a RF load. If high HOM powers are expected *this load must be at room temperature.* It also must consume little space within the beam tube aperture and should allow cooling to extract the developed heat. The solution which imposes itself is to clad a ring portion of the inner beam tube surface with a uhv-compatible absorbing material. Since, adjacent to a metal surface, only the magnetic field has a component parallel to the absorbing surface layer the material must have *magnetic* RF losses. *Ferrites* are such materials and special uhv compatible ferrites were available from earlier work at CERN [5] to dampen parasitic resonances in the beam pickups of the Antiproton Accumulator. The problem was now to attach such ferrites with a good mechanical and thermal bond to a metal substrate. A technique using soft soldering [6] has been developed at Cornell while KEK [7,8] has succeeded in brazing ferrites. The main obstacle to bonding of ferrites to a metal is their difference of thermal expansion. It is important to keep the dimensions of ferrite tiles small and to use a ductile metal like copper as substrate. Brazing then can be done in a vacuum oven using a standard Ag-Cu brazing alloy while pressing both partners together with 4 kg/cm^2 force.

3.4 General design implications

The concept of using the beam tubes for HOM damping implies an accelerating-module design where each single-cell cavity is housed in a separate cryostat. As illustrated by the KEK design for a B-factory, the large beam tube sections protrude out of the cryostat to allow ferrite loads at room temperature. The large diameter beam tubes are wide open ports for heat to reach the liquid helium vessel by radiation and conduction. *Obviously the simplicity of the HOM damping scheme has to be paid for by enhanced refrigeration costs.*

3.5 Ferrite beam tube loads for LEP 2

At CERN we[2] prepare ferrite absorbers for mounting them, if required, into the 10-cm diameter beam tube sections between the 4-cavity modules. In LEP 2, once all the copper cavities have been replaced by superconducting ones, the transversal beam impedance will be reduced, so higher bunch charges can be accelerated. In addition, when the particle energy is increased, bunches may become shorter than they are now. As a consequence the HOM power deposition within a module will increase, particularly at high frequencies, where the performance of the HOM couplers is not well known. Figure 6 shows the power spectrum predictions [9] according to the ABCI program. Since this power is deposited into longitudinal modes and since above 2.2 GHz the TM_{01} waveguide mode starts to propagate in a 10-cm diameter tube we can reasonably hope to intercept *the high frequency tail* of the spectrum (about 1 kW for 2 x 4 bunches of 1 mA) at room temperature inbetween the modules.

The ferrite tiles had to be integrated into the pumping boxes between the modules (Fig. 7). The basic idea for their manufacture was first to braze small (3 x 15 x 18 mm^3) ferrite tiles onto flat copper strips which then, by electron beam welding, were joined to form a 10-cm diameter absorber tube. In between the strips 2-mm wide slots allow for pumping. The tube is then integrated into the cylindrical pumping box. Water cooling loops at both ends of the absorber tube remove heat.

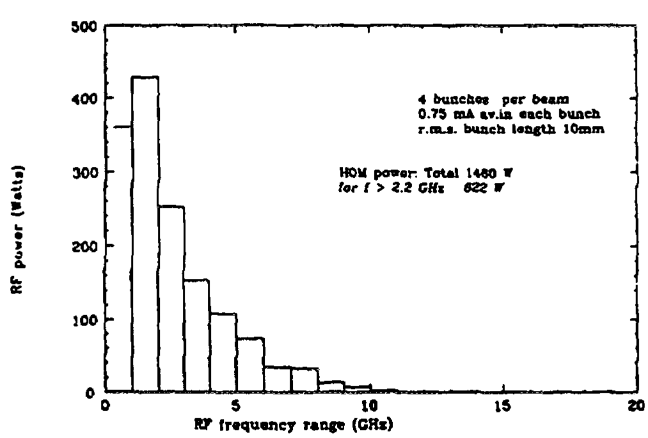


Fig. 6 LEP2 module HOM power from ABCI calculations for $\sigma_s = 10$ mm

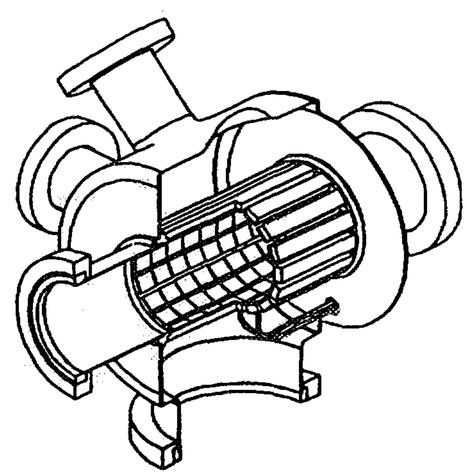


Fig. 7 Pumping box with integrated HOM load

[2] F. Caspers, E. Haebel, N. Hilleret, V. Rödel, B. Trincat, R Valbuena

3.6 Concluding remarks on waveguides as HOM couplers and definition of external Q

Apart from the technically difficult vacuum RF load, (but which is only needed if the beam tubes themselves are to be used as waveguides), this HOM coupler concept is of great simplicity. We cut an opening into the wall of the cavity and, through this aperture, let RF power radiate into a wave guide which carries the power as a travelling wave to a termination at room temperature. For a mode with angular frequency ω_n , stored energy U_n and radiating the power P_n we may, since $P_n \propto U_n$, characterize the coupling by defining an *external* Q :

$$Q_{\text{ex},n} = \frac{\omega_n U_n}{P_n} \tag{1}$$

Below cut off a guide cannot carry travelling waves. Thus, by a proper choice of the transverse guide dimensions, coupling to the FM is easily suppressed. But guide and cavity dimensions become comparable and integration into a cryostat while keeping heat leaks at an acceptable level is difficult[3]. So *coupling to transmission lines* has become a more widespread technique. We then need a *filter* to suppress FM coupling but we can adapt the line cross-section to the HOM power and minimize heat leaks.

4. COUPLING TO TRANSMISSION LINES AND THE EQUIVALENT-GENERATOR APPROACH

Amongst the different forms of transmission lines the coaxial one is best suited for our purpose. For the moment we will assume that the line is terminated by a matched load, $R=Z_w$, where Z_w is the *wave impedance*, so again energy transport is by a travelling wave.

Let us now examine the interface between cavity and line in more detail. We have two basic choices. As depicted below we can either leave the inner conductor end 'open', forming a *probe*, or 'shorted', forming a *loop*. Seen from the cavity the transmission line terminates this probe or loop by a resistor $R = Z_w$.

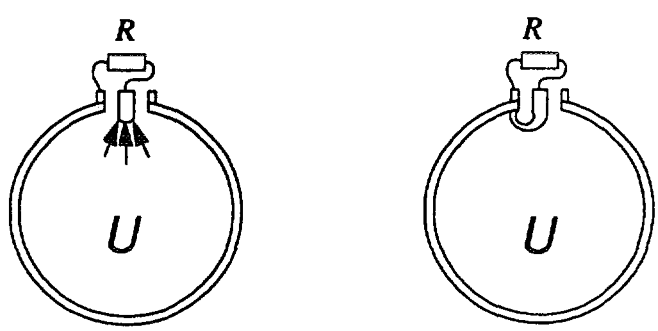


Fig. 8 Two possibilities of coupling, *probe* or *loop*

4.1 Non-resonant coupling

For a given loop or probe the question is now: Which value of R gives the lowest Q_{ex}?

---

[3] At CEBAF HOM couplers in waveguide technique are used but with loads in the liquid He bath, but this is only possible because the HOM power is very small.

To answer the question let a mode at ω be excited by some auxiliary device to oscillate with *constant stored energy U*. We then may regard the probe or the loop as the output port of a RF generator with *emf* V_0 and *internal impedance* Z_i. We can determine V_0 and Z_i by 'measuring' *open circuit voltage* V_0 and *short circuit current* I_0 and have for the loop:

$$I_0 = \frac{\Phi_m}{L_s} = \frac{\mu_0}{L_s} \iint \vec{H} \cdot d\vec{s}$$

$$V_0 = j\omega\Phi_m = j\omega\mu_0 \iint \vec{H} \cdot d\vec{s} \tag{2}$$

L_s is the self-inductance of the loop. H is the mode's magnetic field and we integrate over the loop surface. Evidently $Z_i = j\omega L_s$ and according to *Thevenin's theorem* the coupling port can be described by the *loop inductance* in series with the *induced voltage* as in Fig. 9

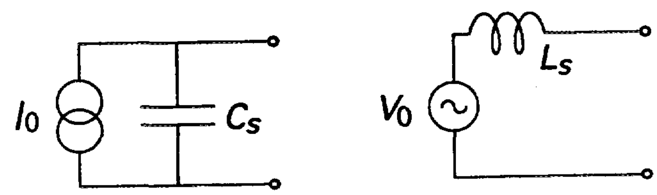


Fig. 9 Equivalent generator circuits for probe and loop coupling

A port equipped with a probe has an equivalent circuit dual to that of the loop. C_s is the fringe field capacitance of the probe tip. I_0 is a short circuit current and corresponds to the displacement current of that part of the cavity E-field which ends on the probe surface.

$$I_0 = j\omega\Phi_e = j\omega\varepsilon_0 \iint \vec{E} \cdot d\vec{s} \tag{3}$$

Which power is now extracted from the mode? We find for the *loop* (Y is the *admittance* 'seen' by the voltage source V_0):

$$P = \frac{1}{2} V_0^2 \operatorname{Re}(Y) = \frac{1}{2} V_0^2 \frac{1}{Z_w} \frac{Z_w^2}{Z_w^2 + (\omega L_s)^2} \tag{4}$$

and for the probe with $Y_w = 1/Z_w$ (Z is the *impedance* 'seen' by the current source I_0)

$$P = \frac{1}{2} I_0^2 \operatorname{Re}(Z) = \frac{1}{2} I_0^2 \frac{1}{Y_w} \frac{Y_w^2}{Y_w^2 + (\omega C_s)^2} \tag{5}$$

The presence of L_s (or C_s) diminishes the power flow on the transmission line and hence the damping of the mode.

4.1.1 The limit of obtainable damping

Evidently here we have a dilemma. If, to obtain more damping, we increase the surface of a coupling loop, then also L_s will increase and take away at least part of the potential benefit of the higher induced voltage. And similarly, for a probe a bigger surface will increase C_s.

Let us study this important phenomenon in more detail. We use a geometry of transmission line and loop (see Fig. 10) which hardly would be used in a real construction project but has the advantage that the loop's self-inductance can be expressed by a simple formula. We use a strip line and as loop a *solenoid with only one turn* so (r is the radius of the solenoid and l its length):

$$L_s \approx \mu_0 \pi \frac{r^2}{l}$$

Figure 11 sketches how such a loop could couple to the TM_{010} mode of a pillbox cavity. The coupler is in the symmetry plane of the cavity and the loop at the equator where the magnetic field B is maximal and to first order constant over the loop area. Then:

$$\frac{1}{P} = \frac{2Z_w}{\left(\omega r^2 \pi B\right)^2} \; \frac{Z_w^2 + \left(\omega \mu_0 \pi r^2 / l\right)^2}{Z_w^2}$$

and it follows
$$Q_{ex} = \frac{\omega U}{P} = 2\frac{\omega U}{B^2}\left\{\frac{Z_w}{\pi^2 \omega^2 r^4} + \frac{\mu_0^2}{l^2 Z_w}\right\} \tag{6}$$

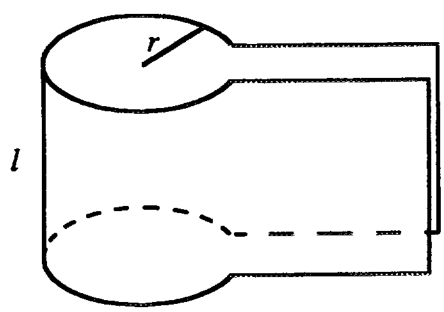


Fig. 10 Stripline with loop of solenoidal geometry

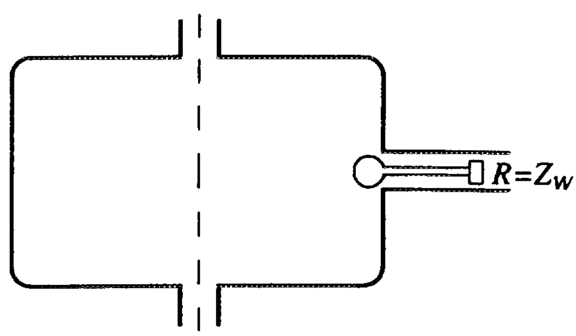


Fig. 11 Pillbox cavity with the coupling loop of Fig. 11 at the equator

The factor $\omega U/B^2$ describes the cavity and loop position, the two additive terms within the curly brackets the coupler itself. The salient point is now that one of the two terms is independent of the loop radius. So it constitutes a lower limit to the obtainable *external Q* (an upper limit of damping). It does not pay to increase the loop radius much beyond the value leading to

$$\frac{Z_w}{\pi^2 \omega^2 r^4} = \frac{\mu_0^2}{l^2 Z_w}$$

But this expression can be rearranged to give

$$\omega L_s = Z_w \tag{7}$$

and in this form is of general validity. Having chosen a wave impedance Z_w for the transmission line and a conductor for the loop, Eq. (7) limits the useful loop size. On the other hand, as the second term in (6) shows, if the loop area linked to flux remains fixed, increasing the conductor cross-section will lead to improved damping.

An example:

To close this section let us now calculate the *lower external Q limit* for the fundamental mode if we mount this loop coupler on the equator of the 4-cell LEP cavity. For an accelerating voltage of V = 1.7 MV (1 MV/m of accelerating gradient) this cavity has at the equator a magnetic induction of B = 4 mTesla and in the π-mode a (R/Q) of 230 Ω. We now can evaluate $\omega U = V^2/(2R/Q)$ and obtain, for a line impedance of 50 Ω and a loop length of 8 cm, from the second term of Eq. (6):

$$(Q_{ex})_{min} = 3900$$

For the reasonable loop diameter of 4.2 cm which produces $\omega L_s = 50\ \Omega$, the external Q will be twice the minimal one and increasing the loop diameter by a factor $\sqrt{2}$ to 6 cm we will approach the minimal value to within 25%:

$$Q_{ex} = 5000$$

However, much lower values of Q_{ex} can be obtained from the same loop at the price of some more sophistication.

4.2 Resonant coupling

In the simple non-resonant approach Q_{ex} has a lower limit, since part of the induced voltage is lost as voltage drop across the internal impedance. But this impedance is not resistive as in ordinary generators. For a coupler it is a pure reactance and, in contrast to a resistance, a reactance can most easily be compensated by an opposite one, at least at certain frequencies.

For a loop coupler the simplest measure is to connect a capacitor in series with the loop [10]. If we do so for the 6-cm diameter loop we need a capacitive reactance of 100Ω at 352 MHz. From $1/(\omega C) = 100\ \Omega$ we obtain C = 4.55 pF. Such a capacity is small enough to be easily realised in practice. The result of compensation is often spectacular. In our example, adding the capacitor would divide the external Q by 5!

$$Q_{ex} = 1000$$

And lower values are within reach as the following discussion will show.

4.2.1 The damping limit for resonant coupling

With the compensation condenser added we have the equivalent circuit of Fig. 12. Evidently now, at the mode frequency, we can increase the extracted power, and hence the damping, if we reduce R (e.g. by connecting a $\lambda/4$ *transformer* between condenser and transmission line) and, in reducing R to zero, it appears as if in the limit Q_{ex} = 0 could be obtained. On the other hand, with R = 0 there are no losses in the system, and Q_{ex} must be infinite! To remove this contradiction we have to realise that, in adding the compensation condenser, we transformed the coupler *itself* into a *resonator* with its own quality factor $Q_{co} = \omega L_s/R$ and tuned to the mode frequency. Figure 13 depicts the situation in representing the mode by a parallel LC-resonator which couples via the mutual inductance $M = k\sqrt{L}\sqrt{L_s}$ to the loop. k is the coupling factor and it follows from the definition of M that the flux of cavity field through the loop is

$$\Phi = M\, I = k\sqrt{L}\sqrt{L_s}\, I$$

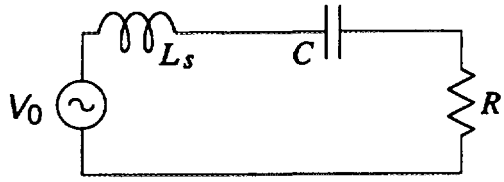


Fig. 12 Capacitive series compensation of loop inductance

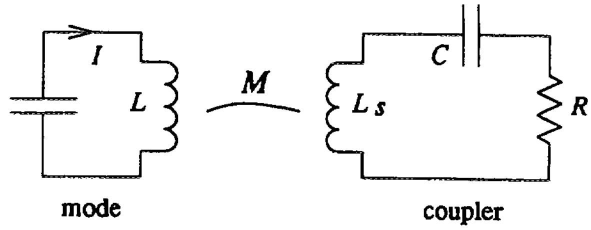


Fig. 13 More realistic representation of coupling

We will now express the previous formulae in a coupled-resonator notation and this will help to see the limit of their range of validity. Using $0.5\, L\, I^2 = U$ we get

$$\Phi^2 = k^2 L_s \cdot LI^2 = 2k^2 L_s U \tag{8}$$

With $P = 0.5 V_0^2/R$ we have now

$$P = \frac{1}{2}\frac{(\omega\Phi)^2}{R} = \frac{1}{2}\frac{(\omega\Phi)^2}{\omega L_s} Q_{co}$$

and substituting from (8) gives

$$P = k^2 \omega U Q_{co}$$

If now $\omega U/P = Q_{ex}$ then

$$Q_{ex} = \left(\frac{1}{k}\right)\cdot\frac{1}{kQ_{co}} \tag{9}$$

Equation (9) does not give anything new. It still predicts infinite damping for infinite coupler Q. But it is written in the language of coupled-resonator theory which says that for $R = 0$ the presence of the coupler resonator will simply split the cavity mode into two, with frequency difference $\Delta f = k\, f$.

On the other hand, modes of a coupled system are sensitive to perturbations. Damping is one of the possible reasons for perturbations and splitting disappears when $kQ_{co} < 1$. Substituting this into (9) will give a reasonable estimate of the minimal external cavity Q that can be produced with this resonant coupler technique.

$$(Q_{ex})_{min} \approx \frac{1}{k} = \frac{\sqrt{2UL_s}}{\Phi} \tag{10}$$

In our LEP cavity example, with the loop of 6-cm diameter we have $L_s \approx 46$ nH and, for a stored energy of 2.9 Ws, a magnetic flux of $\approx 11.3\ 10^{-6}\ V_s$ through the loop. Using (8) we calculate a coupling factor of $\approx 2\%$ and

$$(Q_{ex})_{min} \approx 50$$

The coupler Q must then be 50 too and this requires a resistive termination of 2Ω. In designing couplers for HOM damping one hardly would go to such an extreme because the coupler bandwidth becomes too small and about one coupler per mode would be needed!

4.2.2 Fundamental mode filter requirements

As mentioned earlier coupler constructions based on transmission line techniques have to use a filter to suppress coupling to the fm. This appears to complicate the design but the filter can always be integrated into a reactance compensation scheme. For instance, if in Fig. 12 an inductance $L = L_s$ is connected in parallel to C, then a stop filter for the FM at ω_0 is formed (see Fig. 14). But at higher frequencies the filter reactance becomes capacitive to compensate the loop reactance at $\sqrt{2}\omega_0$. More general for this circuit, if the HOM frequency at which we want compensation is at ω_c then L and C are given by:

$$1/C = \omega_0^2 L = \left(\omega_c^2 - \omega_0^2\right) L_s \tag{11}$$

Voltage and current in the filter may reach high values. To illustrate this point we do the 'tentative design' of a LEP cavity HOM coupler compensated at 500 MHz and using a loop (as sketched in Fig. 11) with 3-cm diameter and 4-cm length. Then at 352 MHz $\omega L_s = 50\ \Omega$ and, since $500/352 \approx \sqrt{2}$, $L = L_s$. At an accelerating gradient of 6 MV/m the equatorial magnetic cavity field is 24 mT and induces the voltage $\omega\Phi = 37.5$ kV, which drops across the filter condenser C and drives a current of $I = 750$ A through L. If C has a gap of 3 mm, (a bigger gap could lead to multipacting), a surface field of 12.5 MV/m results and if the conductor carrying I has a diameter of 3 cm a surface magnetic field of ≈ 10 mT will be produced, *values comparable to those in the cavity! Construction from superconducting materials using cavity assembly and surface preparation techniques is mandatory.*

Two of the four HOM couplers of the CERN 5-cell 500-MHz prototype sc cavity, built for beam tests in PETRA at DESY, were of this type [11]. An outline of their geometry is given below in Fig. 15. With a coupler $Q \approx 2$ they produced *external* Q values ≈ 40000 and were operated up to accelerating gradients of 4MV/m (limit due to a cavity quench).

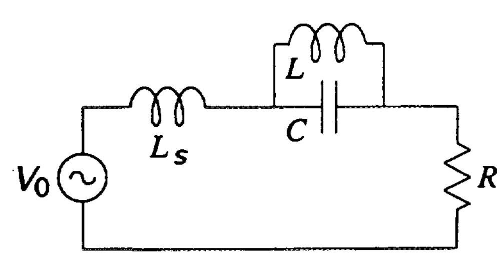


Fig. 14 Resonant HOM coupler with fundamental mode stop filter

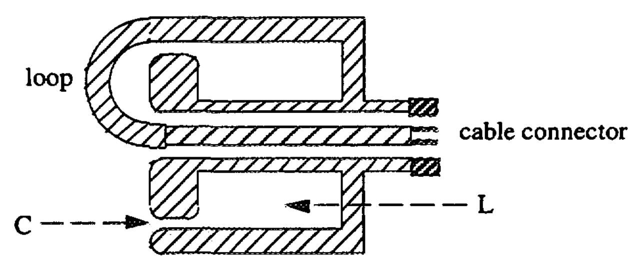


Fig. 15 Outline of resonant HOM loop coupler with FM filter

However, this approach was not further pursued for two reasons. Making an opening at the cavity equator causes local field enhancement just where the FM B-field has its maximum, and putting the coupler at the maximum in turn causes very high field values in the coupler's filter. Looking for a coupler position with smaller FM field and above all a more favourable ratio of HOM to FM fields, a *position on the beam tubes* of the cavity was identified as optimal. The TM_{01} mode being in cut off, damping in the beam tube then acts as a prefilter for the FM .

4.2.3 Couplers on the beam tube with several resonances

Another feature of the last generation of HOM coupler designs is that they are tailored to be resonant at *several* frequencies. Figure 16 illustrates for the LEP cavity the distribution of

HOM with significant R/Q values. As we see, high R/Q HOM come in three clusters around 480 MHz, 650 MHz and 1.1 GHz, so at these frequencies should be resonances of the HOM coupler. A heuristic approach leading to a coupler with the required three resonances is the following. First realize that 650 and 1100 MHz are nearly harmonic and that a transmission line resonator of length $\lambda/2$ at 650 MHz with shorts at both ends would have a second resonance at 1300 MHz.

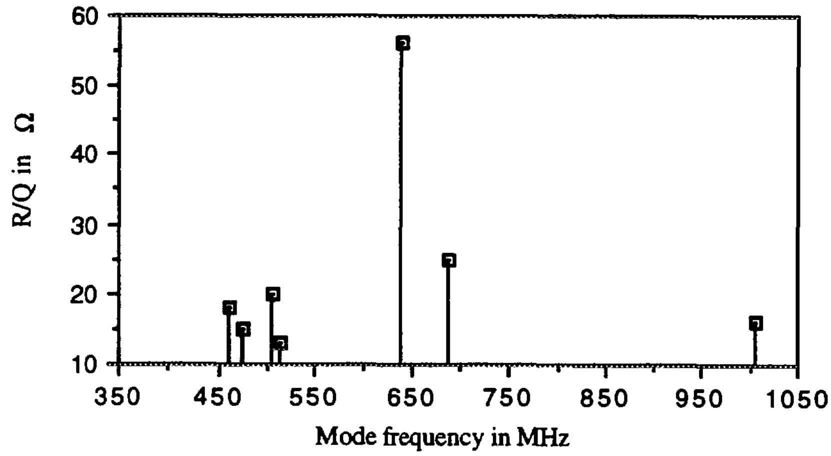


Fig. 16 The high R/Q HOM of the LEP cavity

Replacing the shorts by small inductances L_1 & L_2, (and L_1 would be the coupling loop), we can move these resonances nearer to each other. A third lower frequency resonance finally is obtained by connecting the termination R (see Fig. 17) via a capacitor C to L_2. L_2, C and R then form a second resonator, which shares with the line resonator the common element L_2. In other words: the two resonators are coupled, allowing to split the original 650 MHz resonance into two.

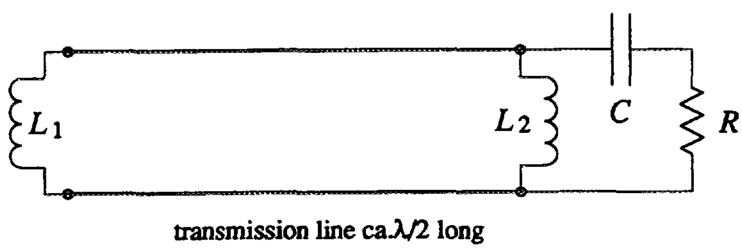


Fig. 17 Circuit of a transmission line coupler with three resonances

The HOM couplers [12] used at DESY in HERA and more recently in the TESLA [13,14] project, as well as at CERN on the LEP Nb-Cu cavities all use this approach. As shown in Fig. 18 the HERA coupler has a small antenna added to the loop so that coupling to electric cavity fields is enhanced. The FM filter is in the back of the coupler, parallel to L_2, forcing a voltage zero across C. In CERN (see Fig. 19) the filter has been moved to the front in returning L_1 via a capacitor to ground. In this way coupling to FM E-fields is eliminated. Also, to avoid magnetic coupling, the loop is perpendicular to the cavity axis. HOM coupling is electric for the longitudinal modes and predominantly magnetic (to their B_z in the beam tube) for the dipole modes. This modification made it possible to have a *demounting flange*, which has to carry only the HOM power. In fact, for a successful fabrication of sc cavities by sputter-coating [15], demountability of couplers is a precondition.

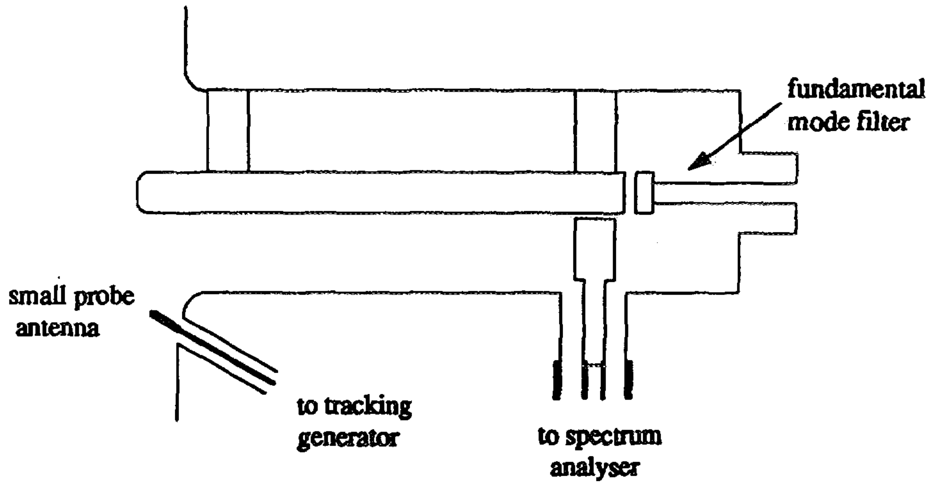


Fig. 18 Scheme of the HERA HOM coupler

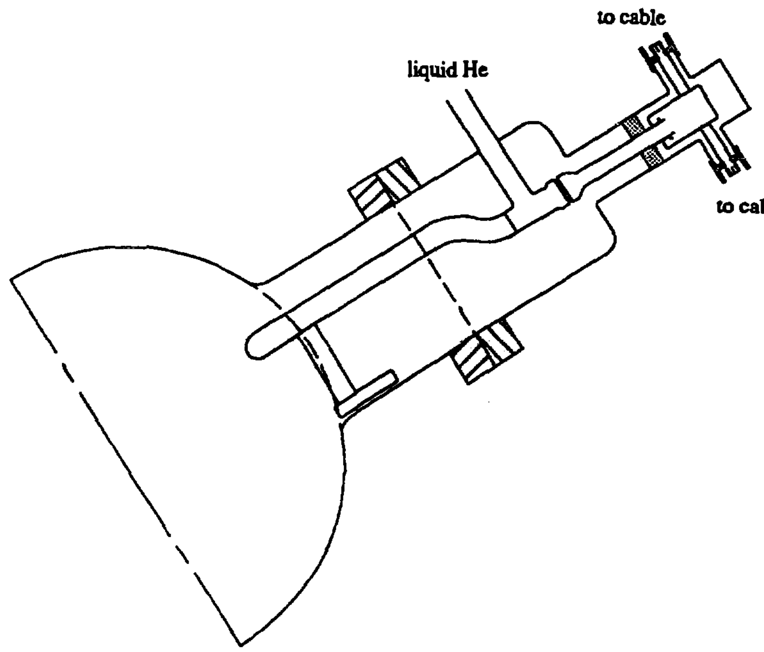


Fig. 19 Demountable coupler for sputter-coated LEP cavities

At SACLAY and CERN it has also been found that letting a loop protrude into the beam tube is another way to get sufficient coupling to the *E-fields* of the longitudinal (TM_{0nm}) modes. This is an interesting variant since it is simpler to fabricate and has been chosen for the LEP cavities. A schematic drawing together with a table of external Q_s is given in Appendix A.

4.3 Design aids

The equivalent generator approach combined with a circuit model of the coupler lends itself readily to making meaningful estimations of mode damping. One ingredient is a calculation (by a network analysis code) of the real part of the coupler's input impedance for probes (or admittance for loops) *including* into the network the coupler's C_s (or L_s respectively).

The second ingredient is an estimation of the probe's short-circuit current I_0 (or the loop's open-circuit voltage V_0) with the help of a cavity code which gives the stored energy U and the fields E and H at the coupler's location. We assume that these fields are not too much perturbed by the presence of the coupler, and estimating the integrals of (2) and (3) we have all the needed information to calculate Q_{ex}.

During the conceptual design phase of a HOM coupler it is most useful to plot the calculated real part of the input impedance (or admittance) against frequency to identify the position of maxima (resonances) and their relative height. After building hardware models such 'sensitivity' curves can be verified in measuring the transfer function between a small loop or probe (which replace the cavity field) and the coupler's load as illustrated in Fig. 18.

More recently a code (HFSS) has become available which, with the exclusion of the interaction region between coupler and cavity fields, allows one to calculate the s-parameters of a complete hardware model. Using it greatly facilitates the translation from the circuit model to the final RF structure.

5. THE HIGHER ORDER MODE POWER

Obviously a good knowledge of the HOM power is needed if one has to decide on the size of the couplers and of cables, connectors and vacuum feedthroughs which serve to connect the couplers to room temperature loads.

5.1 Beam with sinusoidally varying linear charge density

To calculate the RF power (Ref. [16] contains a more general treatment of the subject) from a particular mode with frequency ω_m first imagine a beam with sinusoidally varying linear charge density which passes an AC current I with frequency ω through a reference plane of the cavity exciting oscillating fields of the same frequency. If ω is in the vicinity of ω_m this mode's field will predominate with an amplitude depending on ω as typical for a resonance, i.e. reaching at $\omega = \omega_m$ a maximum proportional to I. We may measure the accelerating effect of the field by test charges (which have the same speed $v \approx c$ aas the paticles of the beam) and call the *factor of proportionality* between the found *accelerating voltage* V and I (at $\omega = \omega_m$) the *effective shunt impedance R.*

$$V = R\,I \tag{12}$$

R in turn is proportional to the loaded Q of the mode and we call the factor of proportionality *(R/Q)*.

$$R = \left(\frac{R}{Q}\right) Q \tag{13}$$

The method used to define V implies for the power P_m lost by the beam at resonance to the mode's field:

$$P_m = \frac{1}{2} I\,V$$

using (12) it follows that

$$P_m = \frac{1}{2}\frac{V^2}{R}$$

But conservation of energy demands that P_m is equal to the power dissipated in the cavity walls and in the HOM coupler load

$$P_{\text{diss}} = P_m = \frac{1}{2}\frac{V^2}{R} \tag{14}$$

Further by definition

$$Q = \frac{\omega U}{P_{\text{diss}}} \tag{15}$$

where U is the stored energy of the mode. Combining (11), (12) and (13) gives:

$$\left(\frac{R}{Q}\right) = \frac{1}{2}\frac{V^2}{\omega U} \tag{16}$$

Using a cavity code (SUPERFISH, URMEL, MAFIA...) *(R/Q)* can be calculated from Eq. 16 *for all non-propagating modes and P_m could be determined for given I and Q.* But it is more important to realise that all these equations map on the circuit equations of a parallel *LC*-resonator (see Fig. 20) to which the beam current is connected in the form of a current source and which has $1/(\omega_m C) = (R/Q)$.

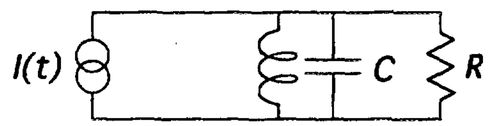


Fig. 20 Circuit model of beam-mode interaction

Trusting the equivalence between circuit model and mode we can now determine the response of the mode *to other beam current forms.*

5.2 Response to a single bunch

For a point charge q passing the reference plane at $t = 0$ the beam current has the form of a single δ-current pulse.

$$I(t) = q\,\delta(t)$$

Such a pulse charges the condenser discontinuously to the voltage q/C which, if the condenser had no charge before, is the initial amplitude of a subsequent *free oscillation*:

$$V(t) = q\,h(t) \quad \textit{with} \qquad h(t) = \frac{u(t)}{C}e^{\gamma t}$$

$u(t)$ is the unit step function.

With the filling time T_f:

$$\gamma = -\frac{1}{T_f} + j\omega_m = -\frac{\omega_m}{2Q} + j\omega_m$$

Further, for a ***linear*** system such as the circuit here and *any* beam current $I(t)$

$$V(t) = \int_{-\infty}^{\infty} I(\tau) h(t-\tau) d\tau$$

and for our $h(t)$: $$V(t) = \frac{e^{\gamma t}}{C} \int_{-\infty}^{\infty} I(\tau)\, u(t-\tau)\, e^{-\gamma\tau} d\tau \qquad (17)$$

We now examine a *single* beam-current pulse $I(\tau)$ and $V(t)$ at times t when the current has returned to zero. During the pulse then $u(t - \tau) = 1$. Also, in accelerators the beam current pulses are very short compared to $2Q/\omega_m$. So it makes no difference if we neglect the attenuation and replace the integral in (17) by

$$\int_{-\infty}^{\infty} I(\tau) e^{-j\omega_m \tau} d\tau$$

This is the Fourier transform of the current pulse $I(t)$ at ω_m. A single gaussian bunch of charge q with

$$I(t) = \frac{q}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2}\frac{t^2}{\sigma^2}}$$

has the Fourier transform

$$I(\omega) = q e^{-\frac{(\omega\sigma)^2}{2}}$$

and ***after the passage*** of a single gaussian bunch

$$V(t) = V_b e^{\gamma t} \qquad (18)$$

with $$V_b = \omega_m \left(\frac{R}{Q}\right) e^{-\frac{(\omega_m \sigma)^2}{2}} q \qquad (19)$$

5.3 The HOM voltage due to a bunched beam

We may now go a step further and try to determine the response to a *beam* made up of bunched charges q which, passing through the reference plane, constitute a *pulsed current with period* T_b. In a steady state the excited oscillations must have the same period T_b. On the other hand, in between bunch passages the mode oscillations must be *free* oscillations i.e. representable by a *phasor* V turning with angular speed ω_m.

Formally, we may turn this phasor back or forward to the moment of the last or next bunch passage and in this way construct phasors V_+ and V_-. Further, to have the required overall periodicity with T_b, at the moment of a bunch passage, the accelerating voltage induced by the bunch must update V_- to become V_+. Central to the derivation is now that, due to the *linearity* of the circuit, this updating voltage must be V_b. Writing complex quantities with a tilde:

$$\tilde{V}_+ = \tilde{V}_- + V_b$$

but also $$\tilde{V}_- = \tilde{V}_+ e^{\gamma T_b}$$

$$\tilde{V}_+ = V_b \frac{1}{1 - e^{\gamma T_b}}$$

and for the time interval of free oscillations:

$$\tilde{V}(t) = V_b \frac{e^{\gamma t}}{1 - e^{\gamma T_b}} \qquad for \qquad 0 < t < T_b \tag{20}$$

5.4 The HOM power from a bunched beam

$\tilde{V}(t)$ drops across the shunt impedance R, causing in average the dissipation

$$P_{\text{mode}} = \frac{1}{2} \frac{1}{T_b} \int_0^{T_b} \frac{\left|\tilde{V}(t)\right|^2}{R} dt$$

P_{mode} is furnished by the bunched beam. Substituting (20)

$$P_{\text{mode}} = \frac{V_b^2}{2RT_b\left|1 - \exp(\gamma T_b)\right|^2} \int_0^{T_b} e^{-\frac{2t}{T_f}} dt$$

using $\tau = \dfrac{T_b}{T_f} = \dfrac{\omega_m T_b}{2Q}$ *and* $\delta = \omega_m T_b$ *(modulo* 2π*)*

we have $\left|1 - \exp(\gamma T_b)\right|^2 = 1 - 2\cos\delta \exp(-\tau) + \exp(-2\tau)$

And evaluating the integral we obtain with $R = Q(R/Q)$ the result:

$$P_{\text{mode}} = \frac{\frac{1}{2}V_b^2}{\left(R/Q\right)} \frac{1}{\omega_m T_b} \mathrm{F}(\tau,\delta) \tag{21}$$

where $$\mathrm{F}(\tau,\delta) = \frac{(1 - \exp(-\tau))(1 + \exp(-\tau))}{1 - 2\cos\delta \exp(-\tau) + \exp(-2\tau)} \tag{22}$$

For gaussian bunches of charge q and substituting from (19)

$$P_{\text{mode}} = \frac{1}{T_b} k_m q^2 \mathrm{F}(\tau,\delta) \tag{23}$$

k_m is called the loss parameter [16] of the mode at ω_m.

$$k_m = \frac{1}{2}\omega_m \left(\frac{R}{Q}\right) e^{-(\omega_m \sigma)^2} \tag{24}$$

The pulsed current of a bunched beam corresponds to a spectrum of discrete lines at distance $1/T_b$ which has the Fourier transform of a single pulse as envelope curve. $\delta = 0$ (modulo 2π) means that ω_m coincides with one of the spectral frequencies. At $\delta = \pi$ the mode is just in between two spectral lines. Figure 21 shows a plot of $F(\tau,\delta)$ for these two cases. Note that $F(\tau,\pi) = 1/F(\tau,0)$. F is plotted against $2/\tau = 4Q/(\omega_m T_b)$.

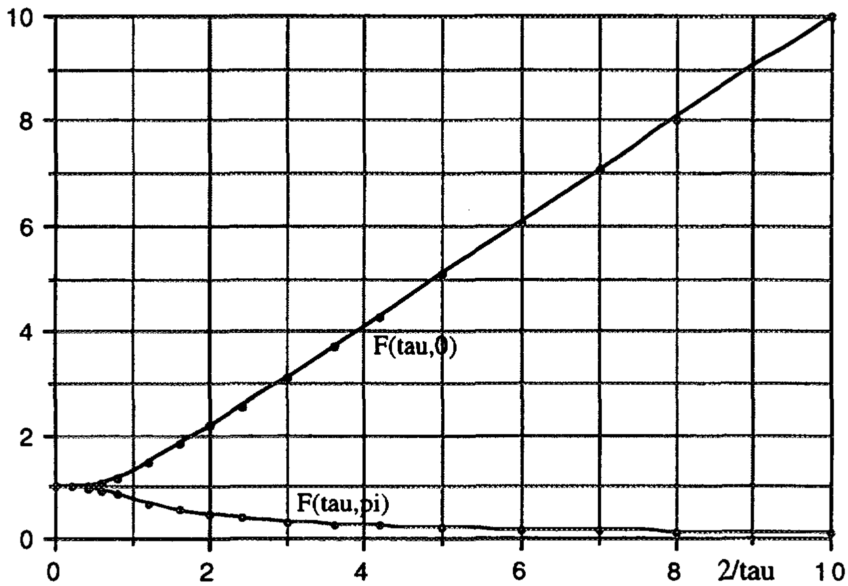


Fig. 21 F(τ,0) and F(τ,π) plotted against $2/\tau = 4Q/(\omega_m T_b)$

5.4.1 The strong damping limit

Whatever the δ, for increasing damping (increasing τ) F approaches one, and P_{mode} becomes

$$P_{\text{mode}} \approx \frac{1}{T_b} k_m q^2 \tag{25}$$

In the limit each bunch of the beam sees an "empty" cavity and we conclude that the bunches now loose an energy $k_m q^2$ to each mode. For strong damping of all HOM the expression for the total HOM power thus must have the simple form

$$P_{\text{HOM}} \approx \frac{1}{T_b} k q^2 \tag{26}$$

with

$$k = \sum k_m \tag{27}$$

If the beam is kept on the cavity axis where transversal modes have no accelerating field, only longitudinal modes contribute to k. Codes like TBCI and ABCI allow one to calculate the k_m and their sum, even including the frequency range where modes propagate.

5.4.2 The weak damping limit

If $\tau = 0.5\omega_m T_b/Q$ is small enough to approximate exp(-τ) by 1-τ then we can simplify setting:

$$F(\tau,0) \approx \frac{2}{\tau} = \frac{4Q}{\omega_m T_b}$$

Formula (21) now takes the form

$$P_{\text{mode}} \approx \frac{1}{2}\left(\frac{2q}{T_b} e^{-\frac{1}{2}(\omega_m \sigma)^2}\right)^2 \left(\frac{R}{Q}\right) Q$$

The expression in brackets is the intensity of the beam current's spectral line at ω_m. Calling it the beam's RF current at ω_m we have:

$$P_{\text{mode}} \approx \frac{1}{2} I_{\text{rf}}^2 \left(\frac{R}{Q}\right) Q = \frac{1}{2} I_{\text{rf}}^2 R \qquad (28)$$

We arrived at a formula rigourously valid for a *single* harmonic current source and conclude that used for a bunched beam this approximation is better the smaller the mode's loaded bandwidth compared to $1/T_b$.

5.5 Choices of damping

As we have seen, for a bunched beam we can never reduce the HOM power to zero. The lower limit of power given by (26) is the higher the smaller the number of bunches circulating in a machine for a given average DC beam current $I_0 = q/T_b$. In fact, by substitution of I_0 into (26):

$$P_{\text{HOM}} = k\, I_0^2\, T_b \qquad (29)$$

Small numbers of bunches also produce a dense spectrum of beam lines, so the scatter of HOM frequencies due to dimensional tolerances of cavity production may become comparable to $1/T_b$. It is then impossible to avoid resonances between modes and beam lines and filling times T_f should be made equal or smaller than T_b at least for longitudinal modes with significant *(R/Q)* values.

On the other hand, if high numbers of equally spaced bunches circulate in a machine it may be possible to detune high *(R/Q)* modes from beam lines and weak damping gives an advantage, provided beam stability requirements allow high external Q values. In all these cases the resulting *loaded* Q will be many orders of magnitude smaller than the Q_0 of modes in a sc cavity. So HOM dissipation in the cavity walls will be negligible compared to the dissipation due to the generator driven fundamental mode.

6. THE FUNDAMENTAL MODE POWER COUPLER

6.1 Matching

Power couplers must have specified properties only near to a single frequency. In this respect they differ considerably from HOM couplers, which in general have to cover broader frequency ranges. Also, at least for the applications developed so far, they must couple only loosely to the fm. If again we measure coupling by an external Q, values smaller than 1E+5 will hardly be needed and reactance compensation techniques to enhance coupling are not required, rather *'matching'* is required.

The concept of *matching* comprises measures which, by an appropriate choice of impedances, optimize the generation and the transport of power. A simple high power RF system has three building blocks: the power amplifier, a line (transmission line or wave guide) and the load. We then want to minimize power losses on the line. This implies that power is transported by a pure traveling wave and so the line must see a load equal to its wave impedance. In most cases, to realise this condition, we need to connect special impedance transforming *matching devices* between the line and the load proper and to produce a *'load match'* is one of the functions of the power coupler.

Another matching task is at the output port of the power amplifier which, for best efficiency, must work on a tightly specified resistive termination. If lines of standard wave impedance Z_0 are used, the needed transformers form part of the amplifier circuit. The system is now in a state where it works with highest efficiency, but more matching can be done to improve its behaviour in the presence of load matching *errors*. The problem is that klystron and tetrode power amplifiers have an internal or 'source' resistance, which may be an order of magnitude higher than their optimal load resistance. Hence, looking back to the amplifier, the line is not matched. This has the consequence that reflections at the load are back reflected from the amplifier adding to the forward power which thus depends on the load. Matching at the amplifier side of the line is called *'source matching'*. At microwave frequencies a convenient way of source matching is to connect a *circulator* between amplifier and line. For a fixed drive power to the amplifier the forward power is then independent of load matching errors.

6.2 Matching the external Q

The load to which a power coupler shall match is a special one: The FM of a sc cavity, which exchanges energy with the bunches of a particle beam. If we represent the mode by a parallel LC resonator (with $\omega_c L = 1/(\omega_c C) = (R/Q)$) then the beam is a current source connected to its terminals. In the discussion here, which focuses on matching, it is sufficient to regard the coupler as an impedance transformer for the wave impedance Z_0 of the line, so that looking from the resonator to the line we see a transformed wave impedance Z (see Fig. 22).

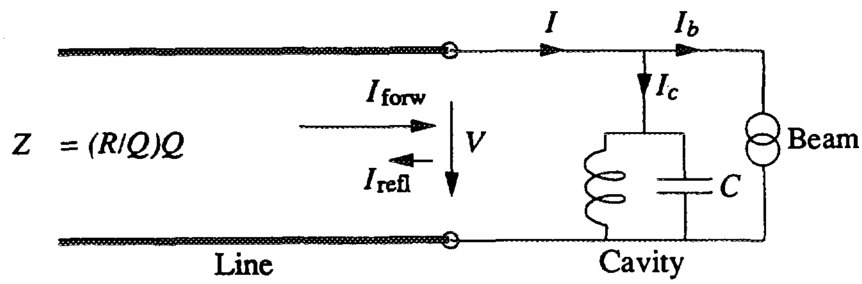


Fig. 22 Line with attached equivalent circuit of loss-free cavity and beam

If this line is source matched then, *seen from the resonator*, its input impedance is also Z, producing as *external Q*

$$Q = \omega_c C\, Z = (R/Q)^{-1}\, Z \qquad (30)$$

We will now assume that the resulting bandwidth $\Delta\omega_c = \omega_c/Q$ is small compared to the bunch frequency. With the resonator tuned near to one particular line of the beam current's spectrum the voltage due to all other lines may then be neglected i.e. the current source in Fig. 22 can be thought of as emitting a *sinusoidal* RF current of frequency ω_g, identical to that of the RF power generator.

In the following derivations we will write complex variables (phasors) with a tilde, their modules (amplitudes) and other real quantities without.

The circuit equations are here:

$$\tilde{V}_{\text{forw}} + \tilde{V}_{\text{refl}} = \tilde{V}$$

and
$$\tilde{I}_{\text{forw}} - \tilde{I}_{\text{refl}} = \tilde{I} = \tilde{I}_c + \tilde{I}_b \tag{31}$$

$$2\tilde{I}_{\text{forw}} = (\tilde{I}_{\text{forw}} + \tilde{I}_{\text{refl}}) + \tilde{I}_c + \tilde{I}_b$$

but $I_{\text{forw}} + I_{\text{refl}} = (V_{\text{forw}} + V_{\text{refl}}) / Z = V / Z$ and $\tilde{I}_c = \tilde{V} / \tilde{Z}_c$

$$2\tilde{I}_{\text{forw}} = \tilde{V}\left(\frac{1}{Z} + \frac{1}{\tilde{Z}_c}\right) + \tilde{I}_b \tag{32}$$

where
$$\frac{1}{\tilde{Z}_c} = j\left(\frac{\omega_g}{\omega_c} - \frac{\omega_c}{\omega_g}\right)\left(\frac{R}{Q}\right)^{-1} \tag{33}$$

Equation (32) is a relation between phasors in a complex current plane. If now we wish to represent (32) by a *phasor diagram*, we are free to define the time zero and thus to decide which of the phasors shall be collinear with the real axis.

- Accelerator physicists often declare the RF beam current as real. In fact it appears natural to regard bunch charges as real quantities and hence also the resulting beam current.

- Circuit analysis usually takes the generator signal, here the forward current, as reference and regards it as real.

- For people who develop sc cavities, the accelerating voltage is a main concern. Regarding it as the reference has advantages. The cavity current I_c (since we neglect cavity losses) is now an imaginary quantity for any tuning and the angle of the RF beam current phasor gives directly the synchronous phase Φ *if we use the convention to count it from the crest of the cavity voltage.* But the forward current now is in general complex and the expressions which relate forward and beam power are under general conditions, rather complicated.

In the following we need to find the optimal conditions for power transfer to the beam. We regard an *'accelerating station'* i.e. a *group of cavities* which, from a power divider, all receive the *same forward power*. Also, the distance of cavities on the beam axis has been chosen to make for all of them the beam current phasor equal in amplitude and phase which, *referenced to the forward current*, is called *'station phase'* φ. Under these conditions it is most convenient to regard the forward current as a real quantity, as in the *phasor diagram* of Fig. 23.

Since Z is real but $\tilde{Z}_c$ imaginary, the phasors $\tilde{V} / Z$ and $\tilde{V} / \tilde{Z}_c$ are at a right angle and meet on a circle which has $2I_{\text{forw}} - \tilde{I}_b$ as diameter. Figure 23 represents a non matched situation: both cavity detuning and coupling are incorrect. To match we first retouch the detuning so that $\tilde{V}$ becomes collinear with I_{forw} (the line 'sees' a real impedance). The result is shown in Fig. 24:

$I_c = V/Z_c$ now compensates the imaginary component I_{bi} of the beam current and if we substitute Z_c from (33) we get the detuning condition

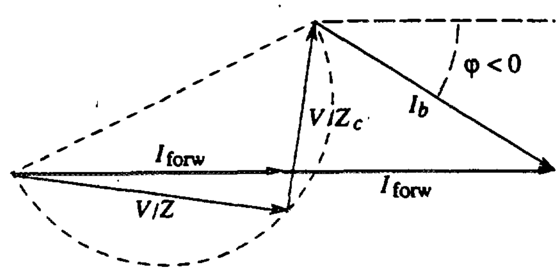


Fig. 23 Equation (32) in a phasor diagram representation

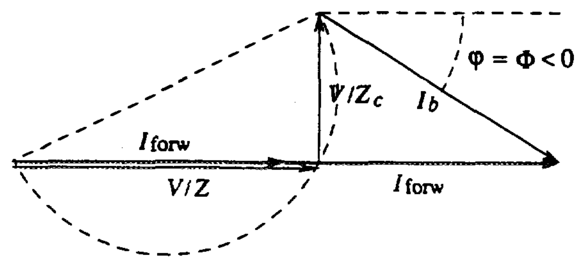


Fig. 24 Phasor sum for a correct cavity detuning

$$I_c = \frac{V}{Z_c} = V\left(\frac{\omega_g}{\omega_c} - \frac{\omega_c}{\omega_g}\right)\left(\frac{R}{Q}\right)^{-1} = -I_{bi} = -I_b \sin\Phi \tag{34a}$$

or

$$\frac{(\omega_g - \omega_c)}{0.5(\omega_c/Q)} \approx -Q\left(\frac{R}{Q}\right)\frac{I_b}{V}\sin\Phi \tag{34b}$$

ω_c/Q is equal to the loaded 3dB bandwidth $\Delta\omega_c$ if the line which feeds power to the cavity is source matched. When (34) is satisfied, station phase φ and synchronous phase Φ *(measured from the crest of the cavity voltage)* become equal.

If the detuning condition is met, Eq. (31) reads:

$$I_{\text{forw}} - \tilde{I}_{\text{refl}} = I_{br} = I_b \cos\Phi$$

and to obtain zero reflection we have in addition to satisfy $I_{\text{forw}} = I_{br}$. But for zero reflection we have also $V = V_{\text{forw}}$. Thus (since $Z_m = (R/Q)Q$):

$$V = V_{\text{forw}} = I_{\text{forw}} Z_m = I_{\text{br}}\left(R/Q\right)Q_m$$

Solving for Q_m and using $2P_b = V I_{br}$ we obtain the condition in a more practical form:

$$Q_m = \frac{1}{2}\frac{V^2}{P_b}\left(R/Q\right)^{-1} \tag{35}$$

P_b is the power transferred to the beam. Note that, for a given cavity voltage amplitude, Q_m depends only on P_b and not on the particular combination of beam current and synchronous phase angle which produces P_b. The phasor diagram for the matched case is shown in Fig. 25 below:

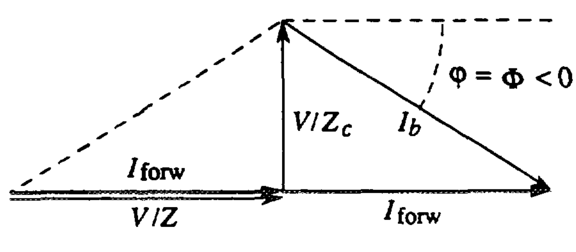


Fig. 25 Phasor diagram at the match point

Example:

The cavities in LEP2 work at 10.2 MV (6 MV/m) and transfer 120 kW to the beam. With $(R/Q) = 230\ \Omega$ we calculate from (35) $Q_m = 1.9E6$. The corresponding 3 dB *bandwidth* is 187 Hz. The RF beam current is 28 mA and the synchronous phase -32.8°. Measured in units of half a bandwidth the required cavity detuning is then 0.64 .

6.2 Matching error analysis

In the following we will use a more formal way of analysis to discuss also the consequences of *matching errors*. Introducing normalized values of detuning, beam current and cavity voltage:

$$d = Q\left(\frac{\omega_g}{\omega_c} - \frac{\omega_c}{\omega_g}\right) \approx \frac{(\omega_g - \omega_c)}{0.5\Delta\omega_c} \qquad \tilde{i}_b = \frac{I_b}{I_{\text{forw}}} = i_{br} + \mathrm{j}\, i_{bi} \quad \text{and} \quad \tilde{v} = \frac{\tilde{V}}{V_{\text{forw}}}$$

and solving (32) and (33) for $\tilde{V}$ we get:

$$\tilde{v} = \frac{2 - \tilde{i}_b}{1 + \mathrm{j}d} = \frac{(2 - i_{br}) - \mathrm{j}\, i_{bi}}{1 + \mathrm{j}d} \tag{36}$$

and calculate from $\tilde{V}$ the reflection coefficient $\tilde{\rho}$.

$$\tilde{\rho} = \frac{\tilde{V}_{\text{refl}}}{V_{\text{forw}}} = \frac{\tilde{V} - V_{\text{forw}}}{V_{\text{forw}}} = \tilde{v} - 1 = \frac{(1 - i_{br}) - \mathrm{j}(d + i_{bi})}{1 + \mathrm{j}d} \tag{37}$$

Let the cavity detuning at which $\tilde{V}$ and $\tilde{\rho}$ become real be d_0. At d_0 the expressions in the denominator and nominator of (36) must have equal phases:

$$\frac{d_0}{1} = -\frac{i_{bi}}{2 - i_{br}}$$

and if we substitute this result back:

$$\mathrm{v} = (2 - i_{br}) \tag{38}$$

and

$$\rho = 1 - i_{br} \tag{39}$$

Evidently complete matching is obtained if, in addition to $d = d_0$, also $i_{br} = 1$.

Finally, to evaluate also for the general non-matched case the power P_b transfered to the beam we determine $1-|\rho|^2$. From (37):

$$|\rho|^2 = \frac{(1 - i_{br})^2 + (d + i_{bi})^2}{1 + d^2} \tag{40}$$

$$\boxed{\frac{P_b}{P} = 1 - |\rho|^2 = \frac{i_{br}(2 - i_{br}) - i_{bi}(2d + i_{bi})}{1 + d^2}} \tag{41}$$

P is the forward power needed to transfer P_b to the beam.

6.2.1 Coupling errors

Let us now compare two cavities of the accelerating station. One, the reference cavity, shall be matched i.e. have (P_m is the forward power and I_{brm} the real component of the RF beam current phasor):

$$I_{\text{brm}} = \left(I_{\text{forw}}\right)_m = \sqrt{2P_m} / \sqrt{Z_m} \quad \Rightarrow \quad \sqrt{2P_m} = I_{\text{brm}}\sqrt{Z_m}$$

$$\text{and} \quad V_m = \left(V_{\text{forw}}\right)_m = \sqrt{2P_m}\sqrt{Z_m} \quad \Rightarrow \quad \sqrt{2P_m} = \text{V}_{\text{m}} / \sqrt{Z_m}$$

The second cavity shall be correctly detuned ($d = d_0$) and sees the same RF beam current. But with $Z \neq Z_m$ it will receive from its line a different value of forward current and of forward voltage:

$$I_{\text{forw}} = \frac{\sqrt{2P_m}}{\sqrt{Z}} = I_{\text{brm}}\sqrt{\frac{Z_{\text{m}}}{Z}} \quad \Rightarrow \quad \text{i}_{\text{br}} = \frac{I_{\text{brm}}}{I_{\text{forw}}} = \sqrt{\frac{Z}{Z_m}} = \sqrt{\frac{Q}{Q_m}}$$

$$\text{and} \qquad V_{\text{forw}} = \sqrt{2P_m}\sqrt{Z} = V_m\sqrt{\frac{Z}{Z_m}} = V_m\sqrt{\frac{Q}{Q_m}}$$

Substituting into (39) and (38) we find for the second non-matched cavity:

$$\rho = 1 - \sqrt{\frac{Q}{Q_m}} \tag{42}$$

and for the ratio of cavity voltages:

$$\boxed{\frac{V}{V_m} = \sqrt{\frac{Q}{Q_m}}\left(2 - \sqrt{\frac{Q}{Q_m}}\right)} \tag{43}$$

From $P_b = (1 - |\rho|^2)\, P_m$ we verify that (43) also describes the ratio of powers transferred to the beam. P_b does not vary strongly with Q. For $Q/Q_m = \sqrt{2}$ we find $P_b/P_m = 0.964$.

A more serious consequence of a coupling error is increased voltage in the coupler's transmission line. V_{refl} adds to V_{forw} to form a standing wave pattern with voltage maxima of $V_{\max} = (1+|\rho|)\, V_{\text{forw}}$. Here we must differentiate between two cases: overcoupling, $Q < Q_m$, and undercoupling, $Q > Q_m$:

$$1 + |\rho| = \begin{cases} 2 - \sqrt{Q/Q_m} & \text{when} \quad Q \le Q_m \\ \sqrt{Q/Q_m} & \text{when} \quad Q > Q_m \end{cases} \tag{44}$$

Present day couplers work at field levels far below those found in cavities. So, why is increased voltage a problem? The answer is that, in contrast to sc cavities, power couplers may be plagued by multipacting discharges within their operating power range and that, in the presence of mismatch, the forward power at which a given multipacting resonance is met, becomes reduced by $(1 + |\rho|)^2$.

Example:
If, with the important mismatch of $Q/Q_m = \sqrt{2}$, one sends 120 kW to the cavity one loses only the moderate reflected power of 4.3 kW. But multipacting levels up to an equivalent power of 170 kW are now within the operating range of the coupler!

6.2.2 Beam current changes

We now change the real component of the RF beam current from I_{brm} to I_{br} and simultaneously, by a servo loop, the forward power from P_m to P *so that the voltage of the reference cavity stays constant.* Equation (38) then gives the condition for the changed forward current:

$$V = V_{forw}(2 - i_{br}) = I_{forw} Z_m\left(2 - \frac{I_{br}}{I_{forw}}\right) \equiv V_m = I_{brm} Z_m$$

or

$$2I_{forw} = I_{br} + I_{brm} \tag{45}$$

and remembering that $I_{brm} = (I_{forw})_m$ we have

$$I_{forw} = \frac{1}{2} I_{brm}\left(1 + \frac{I_{br}}{I_{brm}}\right) = \frac{1}{2}(I_{forw})_m\left(1 + \frac{I_{br}}{I_{brm}}\right) \tag{46}$$

it follows that

$$\boxed{\frac{P}{P_m} = \frac{1}{4}\left(1 + \frac{I_{br}}{I_{brm}}\right)^2} \tag{47}$$

There is now a standing wave also on the transmission line to the reference cavity. To know its maximal voltage we again calculate $(1+|\rho|)V_{forw}$. From (39) and (46), after some algebra (see Appendix B)

$$V_{max} = (1 + |\rho|)\, V_{forw} = \begin{cases} (V_{forw})_m & \text{when } I_{br} \le I_{brm} \\ \dfrac{I_{br}}{I_{brm}}\,(V_{forw})_m & \text{when } I_{br} > I_{brm} \end{cases} \tag{48}$$

Evidently, to avoid increased voltage, coupling has to be correct for the *highest* expected beam current.

6.2.3 Beam current changes and coupling errors combined

Finally, applying P to the second cavity (which has $Q \neq Q_m$) we have the forward current (more details in appendix C):

$$I_{forw} = \frac{\sqrt{2P}}{\sqrt{Z}} = \sqrt{\frac{P_m}{2Z}}\left(1 + \frac{I_{br}}{I_{brm}}\right) = \frac{1}{2} I_{brm}\sqrt{\frac{Z_m}{Z}}\left(1 + \frac{I_{br}}{I_{brm}}\right) \tag{49}$$

and

$$V_{forw} = I_{forw} Z \frac{Z_m}{Z_m} = \frac{1}{2} V_m \sqrt{\frac{Z}{Z_m}}\left(1 + \frac{I_{br}}{I_{brm}}\right) \tag{50}$$

Substituting (49) and (50) into (38) we obtain the generalised form of (43):

$$\boxed{\frac{V}{V_m} = \sqrt{\frac{Q}{Q_m}}\left\{\left(1 + \frac{I_{br}}{I_{brm}}\right) - \frac{I_{br}}{I_{brm}}\sqrt{\frac{Q}{Q_m}}\right\}} \tag{51}$$

Note, that as a consequence of equation (51), for $I_{br} < I_{brm}$ and undercoupling ($Q > Q_m$) the cavity field becomes *higher than the nominal one.* This is illustrated in Fig. 26 below, taking the LEP cavity with its nominal gradient of 6 MV/m as example. The used range of Q/Q_m corresponds to what is found in practice for the present fabrication methods of cavities and couplers.

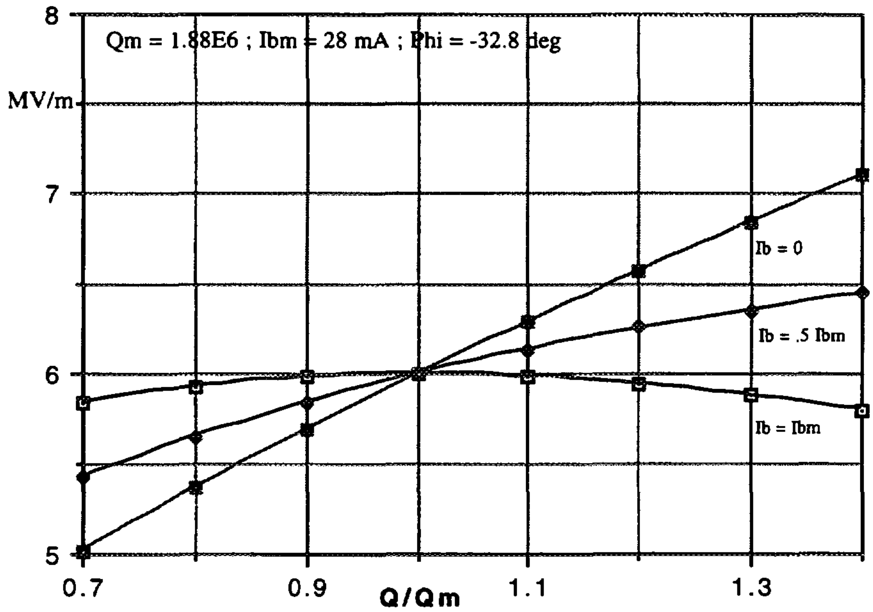


Fig. 26 Coupling dependence of the cavity field for three different beam currents

6.2.4 Tuning errors

All the cavities in an accelerating station have independent tuning loops so, in steady state conditions, tuning errors are insignificant. The situation is different for fast perturbations. Tuner constructions available for sc cavities, since they change the length of the cavity, are rather limited in speed. Even the fastest designs, using magnetostrictive rods, cannot tune out vibrations or beam current changes with frequencies above 10 Hz. It is then necessary to *change the phase* and to *increase the amplitude* of the forward current *to keep the cavity voltage-phasor constant* which, for a calculation of the now required additional forward power, forms the most convenient reference of the phasor diagram in Fig. 27.

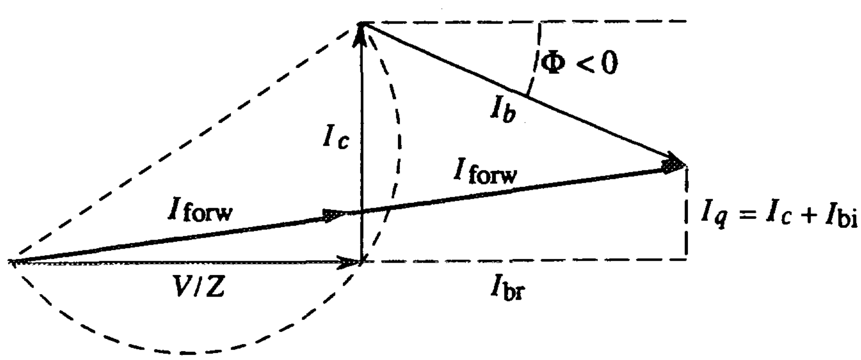


Fig. 27 Same diagram as in Fig. 25, but here *V/Z* is the reference

Note that here, since we decompose in the direction of V, the values of I_{br} and I_{bi} are different from those in Fig. 25! With $I_q = I_c + I_{bi}$ we now read from the diagram (I_q is a measure of the deviation from correct tuning):

$$(2I_{\text{forw}})^2 = \left(\frac{V}{Z}+I_{\text{br}}\right)^2 + (I_c+I_{\text{bi}})^2 = \left(\frac{V}{Z}+I_{\text{br}}\right)^2 + I_q{}^2$$

it follows $$P = \frac{1}{2}(I_{\text{forw}})^2 Z = \frac{Z}{8}\left\{\left(\frac{V}{Z}+I_{\text{br}}\right)^2 + I_q{}^2\right\} \qquad (52)$$

Evidently, for given cavity voltage, RF beam current and synchronous phase (i.e. beam power) the required forward power P depends on Z (coupling) and I_q (tuning) and to *minimize P* we have to satisfy the two equations:

$$\frac{\partial}{\partial I_q}P(Z,I_q)=0 \quad \text{(i)} \qquad \text{and} \qquad \frac{\partial}{\partial Z}P(Z,I_q)=0 \quad \text{(ii)}$$

or $$I_q = I_c + I_{\text{bi}} = 0 \quad \text{(iii)} \qquad \text{and} \qquad I_q^2 + I_{\text{br}}^2 - \left(\frac{V}{Z}\right)^2 = 0 \quad \text{(iv)}$$

Satisfying (iii) and (iv) we have a match and the minimum minimorum of the forward power. But with a tuning error, $I_q \neq 0$, we still can obtain a *relative minimum* of the forward power by satisfying only (iv). We find the optimal coupling under such conditions to be *tighter than the matching one:*

$$Z_{\text{opt}} = \left(\frac{R}{Q}\right)Q_{\text{opt}} = \sqrt{\frac{V^2}{I_{\text{br}}^2 + I_q^2}} \leq \frac{V}{I_{\text{br}}} = Z_m \qquad (53)$$

We now can imagine two scenarios. In the first we had not foreseen a tuning error and matched at a nominal beam current I_{brm} i.e. we had chosen $Z = V/I_{\text{brm}}$. Substituting this value of Z into Eq. (52) we get:

$$P = \frac{V}{8}\left(I_{\text{brm}} + 2I_{\text{br}} + \frac{I_{\text{br}}^2}{I_{\text{brm}}} + \frac{I_q^2}{I_{\text{brm}}}\right)$$

and normalising by the beam power in matched conditions, $P_{\text{bm}} = 0.5VI_{\text{brm}}$, we obtain a more general form of Eq. (47):

$$\frac{P}{P_{\text{bm}}} = \frac{1}{4}\left\{\left(1+\frac{I_{\text{br}}}{I_{\text{brm}}}\right)^2 + \left(\frac{I_q}{I_{\text{brm}}}\right)^2\right\} \qquad (54)$$

I_q is related to the deviation df from correct detuning. To first order $I_q/I_{\text{brm}} = 2df/\Delta f$ where Δf is the loaded 3 dB bandwidth.

Example:
Recently it has been realised [17] that operated at 6 MV/m the LEP cavity suffers from a ponderomotive mechanical oscillation [18]. These oscillations may be suppressed by operating the cavity on its resonance frequency i.e. with $\omega_c = \omega_g$. But then $I_c = 0$ i.e. $I_q = I_{bi}$ and at the nominal beam current I_{bm}:

$$\frac{P}{P_{\text{bm}}} = \frac{1}{4}\left\{(1+1)^2 + \left(\frac{I_{\text{bim}}}{I_{\text{brm}}}\right)^2\right\} = 1 + \frac{1}{4}\text{tg}^2\phi \qquad (55)$$

Now for $P_{bm} = 120$ kW and $\Phi = -32.8°$ $P = 132.5$ kW and $(1 + |\rho|)^2 P = 225$ kW.

In the second scenario we know I_q in advance and choose $Z = Z_{opt}$. For *zero detuning* we have $Z_{opt} = V/I_b$ and the phasor diagram forms an isosceles triangle. Below the LEP example (Φ = -32.8°) is illustrated.

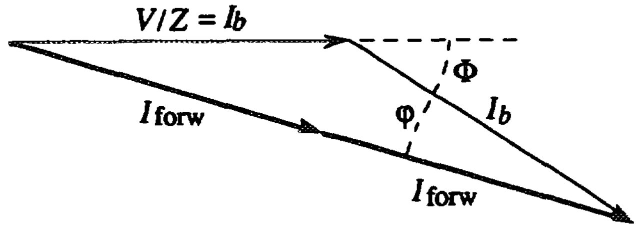


Fig. 28 Phasor sum for zero detuning and optimal coupling

The synchronous phase is now two times the station phase. Substituting Z_{opt} into (52) we find:

$$P = \frac{1}{4} V I_b (1 + \cos\phi) \tag{56}$$

6.2.5 Tuning and coupling errors combined

Finally we will try to visualize what consequences coupling errors have if the cavities of an accelerating station are kept at resonance ($\omega_c = \omega_g$). Here right from the start we had to use a direct numerical approach: First for a reference cavity with $Q = Q_m$ = 1.88E6 and Φ = -32.8° the station phase φ was determined by solving (36) numerically. Remembering that φ and P are the same for all cavities in a station we then can proceed to determine Φ and P_b for cavities with different *external* Q_s in using (36) and (41). The results are shown in Fig. 29. In contrast to the case of correct tuning the synchronous phase (the phase angle *difference* between cavity voltage and RF beam current) now depends on cavity coupling and rises with increasing *external* Q. Undercoupling thus leads to a sharper increase of power reflection and at the upper limit of the external Q range the equivalent travelling wave power $(1+|\rho|)^2 P$ rises to 280 kW.

6.3 Hardware considerations

Power couplers for sc cavities are normal conducting (nc) devices. Hence their designs could take over many features of earlier constructions for nc cavities. Such couplers may be subdivided into three functional units: first the coupling element proper, probe, loop or coupling iris. Second a ceramic window which seals off the cavity vacuum while letting through the RF power. Third a transition piece to the standard wave guide of the power distribution system.

Couplers for sc cavities have one element more to bridge the gap between room and liquid He temperature while keeping, compared to the cavity dissipation, the heat flux into the He bath small. This thermal transition is a length of guide or line where special care has been taken to minimize both the metal cross-section and thermal conductivity by using copper-plated stainless steel. In addition, cooling with cryogenic gases (He or N_2) is employed to intercept heat at discrete points or continuously in a counter current-flow fashion, very similar to the current leads for sc magnets.

For powers around 100 kW and frequencies around 400 MHz wave guides have unnecessarily large cross-sections. So the coaxial line geometry is the preferred ending, to avoid heat conduction by the inner conductor, in an open circuit as the probe which couples to the electric cavity field within the cut-off tube near to one of the cavity end cells. Thus from the inner conductor only heat radiation can reach the He bath and its operation even at room-temperature becomes feasible if no cold window is employed.

Use of such a second cold window (at ≈ 70 K) was the rule in earlier coupler constructions [19,20]. It was felt that only in this way the sensitive sc surfaces of the cavity could be protected if, under the action of the RF fields, gas molecules would be desorbed from 6warm parts of the coupler and especially from its first room-temperature window.

With the argument that the thermal transition piece should act as an efficient baffle, later [21] only the warm window was kept[4] allowing the use of window constructions with a proven performance from copper-cavity work. Figure 30 shows the LEP power coupler as an example of this approach. Window and 'door-knob' transition to a waveguide are taken over from the coupler for the LEP copper cavities.

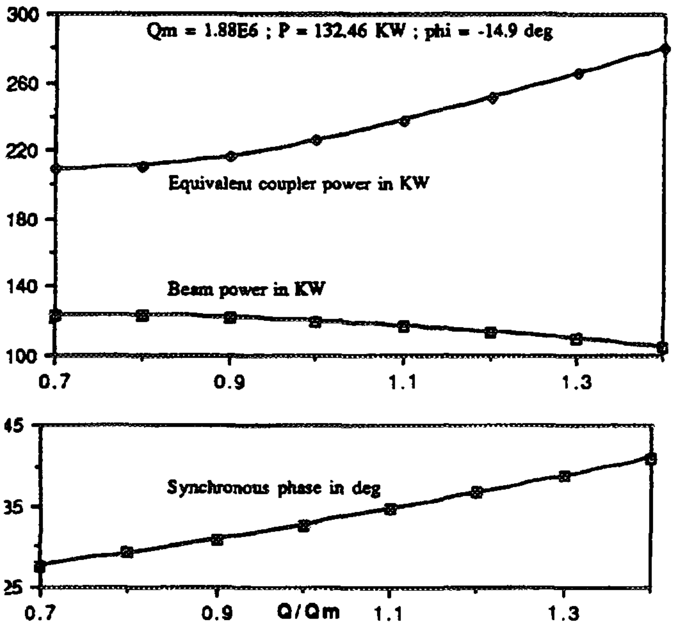


Fig. 29 Influence of coupling errors for LEP cavities operated at resonance

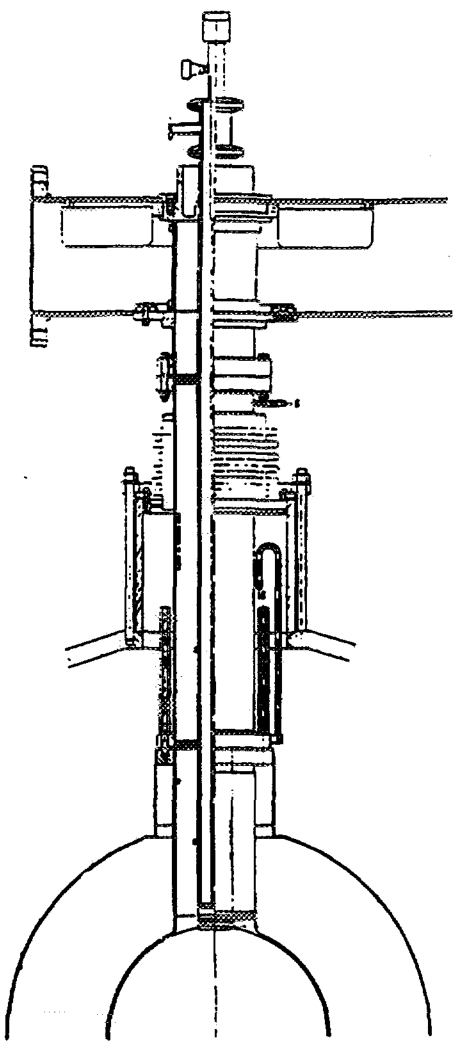


Fig. 30 A LEP power coupler with 50 Ω coaxial line

6.4 Conditioning

Applying for the first time higher powers to vacuum RF devices is often a very delicate operation. In fact, if the power is raised abruptly one risks provoking violent discharges which may destroy the device. Two phenomena combine to create this danger. The first is the

[4] The developments for the 1.5 GHz TESLA very-high-gradient sc linac, to be operated at 1.8 K, have reintroduced a second 70 K window as heat shield and to seal the cavity at an early stage of manufacture against dust particle intrusion.

presence of adsorbed layers of gas (predominantly H_2O) molecules on RF boundary surfaces. The second is the existence of electron orbits which, starting and ending on surfaces, become self replicating at certain field levels. If now at such a kinematic "multipactor" resonance an initial electron has an impact energy high enough to release more than one secondary then an avalanche occurs with a proportional increase of gas desorption. If the process is not controlled the local gas pressure may rise to the level where ordinary gas discharges set in.

But gas desorption is not only a danger. It is also the effect which makes *"conditioning"* possible. In fact, the secondary emission coefficient of a copper surface (and similarly for other metals) decreases when H_2O layers are removed [22], so multipactor discharges have the tendency to eliminate themselves. To make a discharge disappear one only has to "tickle" it long enough, taking care that pressure bursts do not rise into the unsafe region above 10^{-7} Torr.

6.5 Deconditioning

Couplers for copper cavities are conditioned together with the cavity. Couplers for sc cavities need more elaborate procedures. Their operation is often plagued by the *reappearance* of multipactor discharges *(deconditioning)*. This is caused by the baffle action of the warm-cold transition [23].

It has been shown that multipactor orbits in coaxial [24] lines with wave impedances bigger than 50 Ω start and end on the outer conductor alone which, acting as a cold trap, is unfortunately also the zone to where molecules liberated in the warm window area are cryopumped. Multipactoring recreated in this way is very difficult to condition. Probably molecules desorbed by electron impact are only partially pumped away, the remainder is re-adsorbed at cold surfaces nearby.

To alleviate the problem of deconditioning, power couplers for sc cavities are, after a thorough bake out at 200°C, *preconditioned* on a separate room-temperature test stand and only then transferred to the cavity. Since, during this last operation, contact with humid air is unavoidable LEP couplers are subsequently baked a second time *in situ.* The window is heated to 200°C for 24 hours. Only then are the cavities cooled and RF power applied for the final conditioning.

As a further safeguard against the reappearance of multipacting the LEP coupler has been fitted with condensers [25] which isolate the central conductor of the coaxial line for DC voltages but let the RF power pass. So, if required, a bias voltage can be applied [26]. It has been found that a bias of 2.5 kV suppresses all multipactor resonances up to a travelling-wave power of at least 200 kW.

Finally, since multipactor resonances belong to certain *E*-field levels at the *outer* conductor, power flow before multipacting sets in will be the higher the *smaller the central conductor's diameter*. With respect to multipacting characteristics the conventional choice of a 50 Ω wave impedance is not optimal. In consequence the LEP power coupler's antenna diameter recently has been reduced from 45 mm to 30 mm which corresponds to $Z_0 = 75\ \Omega$.

REFERENCES

[1] V. Rödel, E. Haebel, The Effect of the Beam Tube Radius on Higher Order Modes in a Pill-Box RF Cavity, SL/RFS/Note 93-17.

[2] V. Rödel, L. Verolino, Geometry of a Superconducting 400 MHz Accelerating Cell for the LHC, SL/RFS/Note 91-11.

[3] W. Hartung, In Search of Trapped Modes in the Single Cell Cavity Prototype for CESR-B, Proc. of the 1993 PAC, Washington, p.898.

[4] E. Haebel and A. Mosnier, Large or Small Iris Aperture in SC Multicell Cavities? Proc. of the 5th Workshop on RF-Superconductivity, DESY, 1991, p. 823.

[5] F. Caspers, Experience with UHV Compatible Microvave Absorbing Materials at CERN, PS 93-10 (AR).

[6] D. Moffat et al., Design and Fabrication of a Ferrite-lined HOM Load for CESR-B, Proc. of the 1993 PAC, Washington, p. 977.

[7] T. Tajima et al., Bonding of a Microwave-Absorbing Ferrite with Copper for the HOM Damper of the KEK B-Factory SC Cavities, KEK Preprint 93-152, November 1993.

[8] T. Tajima et al., Development of HOM Absorber for the KEK B-Factory SC Cavities, Proc. of the 6th Workshop on RF-Superconductivity, CEBAF, 1993, p.962.

[9] E. Plawski, Influence of the Bunch Length on the HOM Power Deposition in the SC LEP Cavities, CERN LEP2 Note 94-19.

[10] W. Schminke, SPS/ARF/WS/gs/81-68,CERN 1981.

[11] E. Haebel, Fundamental and HOM couplers on SC Cavities for Electron Storage Rings, Proc. of the 1983 PAC, Santa Fe, New Mexico, p. 3345.

[12] E. Haebel and J. Sekutowicz, Higher Order Mode Coupler Studies at DESY, DESY-Report M-86-06, July 1986.

[13] J. Sekutowicz, Higher Order Mode Coupler for TESLA, Proc. of the 6th Workshop on RF-Superconductivity, CEBAF, 1993, p. 426.

[14] S. Chel et al., Thermal Tests of HOM Couplers for Superconducting Cavities, Proc. of the 1994 EPAC, p. 2007 .

[15] Ph. Bernard et al., Demountable E/H-Field HOM Couplers for the Niobium Sputtered LEP Cavity, Proc. of the 5th Workshop on RF-Superconductivity, DESY, 1991, p. 956.

[16] P.B. Wilson, High Energy Electron Linacs: Application to Storage Ring RF Systems and Linear Colliders, SLAC-PUB-2884, Feb. 1982.

[17] D. Boussard et al., Electroacoustic Instabilities in the LEP2 Superconducting Cavities, Proc. of the 7th Workshop on RF-Superconductivity, Gif sur Yvette, 1995.

[18] M.M. Karliner et al., Instability in the Walls of a Cavity due to Ponderomotive Forces of the Electromagn. Field, Soviet Physics-Technical Physics, Vol. 11, No 11, May 67.

[19] A. Citron et al. and H. Lengeler, The Karlsruhe-CERN Superconducting RF Seperator, Nucl. Instr. Meth. 164 (1979) p. 31.

[20] L. Szeczi and R. Lehm, Theor. und Techn. Beschr. eines 100 KW Koppelsystems für einen supraleitenden Beschleunigungsresonator, Primärbericht 9/81, KFK, Karlsruhe.

[21] E. Haebel, Beam tube Couplers for the Superconducting LEP Cavity,, Proc. of the 2nd Workshop on RF-Superconductivity, CERN, 1984, p. 299.

[22] J. Barnard et al., Secondary Electron Emission from Various Technical Materials and Condensed Gases, Proc. of the 7th Workshop on RF-Superconductivity, Gif sur Yvette, 1995.

[23] E. Haebel et al., Gas Condensation on Cold SuRFaces, a Source of Multipacting Discharges in the LEP Power Coupler, Proc. of the 7th Workshop on RF-Superconductivity, Gif sur Yvette, 1995.

[24] E. Somersalo et al., Analysis of Multipacting in Coaxial Lines, Proc. of the 1995 PAC, Dallas, Texas.

[25] H.P. Kindermann et al., Status of the RF Power Couplers for Superconducting Cavities at CERN, to be Published in Proc. of the 1996 EPAC, Sitges (Barcelona).

[26] J. Tückmantel et al., Improvements to Power Couplers for the LEP2 Superconducting Cavities, Proc. of the 1995 PAC, Dallas, Texas.

APPENDIX A

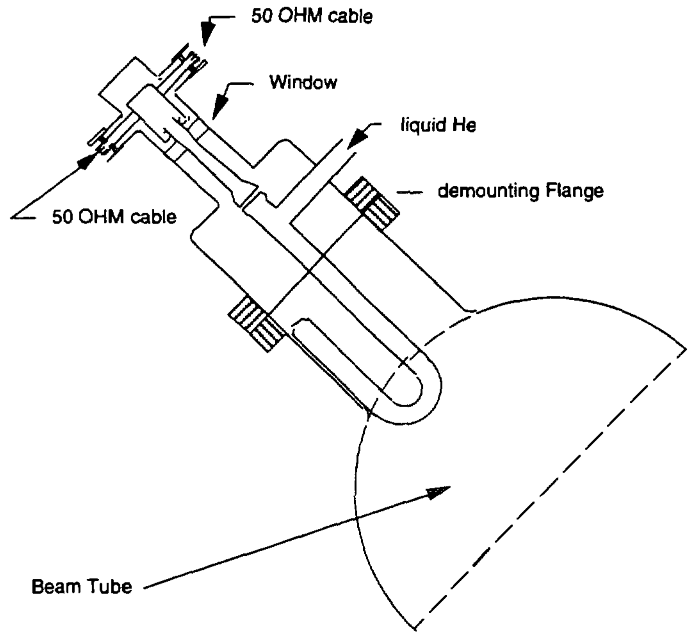


Fig. 1 Geometry of the HOM coupler used in LEP

Table 1
Damping of significant HOM with two couplers per LEP cavity

f/MHz	461	476	506	513	639	688	1006
Mode	TE_{111}	TE_{111}	TM_{110}	TM_{011}	TM_{111}	TM_{111}	TM_{012}
$(R/Q)/\Omega$	18	15	20	13	56	25	16
Q_{ex}	17 000	14 000	5600	5700	7000	1000	2000

APPENDIX B

From

$$V = I_{forw}\, Z_m \left(2 - \frac{I_{br}}{I_{forw}}\right) \equiv V_m = I_{brm}\, Z_m \qquad \text{(B1)}$$

we obtain

$$2I_{forw} = I_{brm} + I_{br} \qquad \text{(B2)}$$

and using Eq. (39) from the main body of this lecture we have $$\rho = 1 - \frac{I_{\text{br}}}{I_{\text{forw}}} = \frac{2I_{\text{forw}} - 2I_{\text{br}}}{2I_{\text{forw}}} = \frac{I_{\text{brm}} - I_{\text{br}}}{I_{\text{brm}} + I_{\text{br}}} \tag{B3}$$

and remembering that $I_{\text{brm}} = (I_{\text{forw}})_{\text{m}}$ we get

$$1 + |\rho| = \begin{cases} \dfrac{2I_{\text{brm}}}{I_{\text{brm}} + I_{\text{br}}} = \dfrac{(I_{\text{forw}})_m}{I_{\text{forw}}} = \dfrac{(V_{\text{forw}})_m}{V_{\text{forw}}} & \text{when } I_{\text{br}} \le I_{\text{brm}} \\ \dfrac{2I_{\text{br}}}{I_{\text{brm}} + I_{\text{br}}} = \dfrac{I_{\text{br}}}{I_{\text{brm}}} \dfrac{(V_{\text{forw}})_m}{V_{\text{forw}}} & \text{when } I_{\text{br}} > I_{\text{brm}} \end{cases} \tag{B4}$$

and $$V_{\text{max}} = (1 + |\rho|)\, V_{\text{forw}} = \begin{cases} (V_{\text{forw}})_{\text{m}} & \text{when } I_{\text{br}} \le I_{\text{brm}} \\ \dfrac{I_{\text{br}}}{I_{\text{brm}}} (V_{\text{forw}})_{\text{m}} & \text{when } I_{\text{br}} > I_{\text{brm}} \end{cases} \tag{B5}$$

APPENDIX C

From the main body of the lecture:

$$V = V_{\text{forw}}\left(2 - \frac{I_{\text{br}}}{I_{\text{forw}}}\right) \tag{C1}$$

and $$\sqrt{\frac{P_m}{2Z}} = \frac{\sqrt{2P_m}}{2\sqrt{Z}} = \frac{(I_{\text{forw}})_m \sqrt{Z_m}}{2\sqrt{Z}} = \frac{1}{2} I_{\text{brm}} \sqrt{\frac{Z_m}{Z}} \tag{C2}$$

and $$I_{\text{forw}} = \frac{\sqrt{2P}}{\sqrt{Z}} = \sqrt{\frac{P_m}{2Z}}\left(1 + \frac{I_{\text{br}}}{I_{\text{brm}}}\right) \tag{C3}$$

it follows that $$I_{\text{forw}} = \frac{1}{2} I_{\text{brm}} \sqrt{\frac{Z_m}{Z}}\left(1 + \frac{I_{\text{br}}}{I_{\text{brm}}}\right) \tag{C4}$$

and $$V_{\text{forw}} = I_{\text{forw}} Z \frac{Z_m}{Z_m} = \frac{1}{2} V_m \sqrt{\frac{Z}{Z_m}}\left(1 + \frac{I_{\text{br}}}{I_{\text{brm}}}\right) \tag{C5}$$

substituting (C4) and (C5) into (C1):

$$\frac{\text{V}}{V_m} = \sqrt{\frac{Z}{Z_m}}\left[\left(1 + \frac{I_{\text{br}}}{I_{\text{brm}}}\right) - \frac{I_{\text{br}}}{I_{\text{brm}}} \sqrt{\frac{Z}{Z_m}}\right] \tag{C6}$$

CRYOGENICS

Jörg Schmid
CERN, Geneva, Switzerland

1 INTRODUCTION

The subject of this Accelerator School is *Superconductivity in Particle Accelerators.* Superconductors are characterised by a high degree of order in their electronic system: it is due to this high degree of order that superconducting electrons can travel as a perfectly ordered assembly through the crystal lattice without transferring energy to it. Using the terminology of thermodynamics, we say that the assembly of superconducting electrons is in a state of zero entropy.

Cryogenics is the art of creating states of low, even zero entropy. Cryoplants are *entropy pumps* that extract entropy from the system to be cooled and discharge it to an environment at higher (usually ambient) temperature, just as vacuum pumps extract gas from the system to be evacuated and discharge it to an environment at higher (usually ambient) pressure. Both cryoplants and vacuum pumps require an external supply of energy.

Cryoplant processes are closely related to heat engine processes; a heat engine can be considered as an *entropy mill,* in which entropy flows from a heat source to a heat sink (which is usually at ambient temperature). Such a process, if carefully designed, can provide external mechanical or electrical energy ('work').

Both cryoplants and heat engines are governed by the Second Law of thermodynamics, which states that the entropy output is always greater than (or — in the never reached ideal, 'reversible' case — equal to) the entropy input:

$$S_{out} \geq S_{in} ,$$

i.e. entropy always increases. There is no such thing as an entropy sink where entropy disappears without reappearing somewhere else. This is generally true, and neither a cryoplant as a whole nor any of its components is an exception.

The trick for cryoplants is to design a process that reduces *local* entropy, while allowing the *overall* entropy to increase, which is all the Second Law requires. How this package deal can be arranged will be shown in the following.

People used to purely mechanical or electrical systems tend to avoid talking about entropy. Indeed, many aspects of cryogenics can be described without reference to it. But the loss in deeper understanding is hardly justified by the momentary convenience of not having to learn how to use a professional tool, which is after all quite easily handled.

I believe there are two main reasons for the reluctance of people to deal with entropy:

- *People are not used to the terminology.* When arguing about energy, mass or temperature, we think we know what we are talking about. Only when challenged to *explain* what *is* energy, mass or temperature, do we realise how superficial our true understanding is of these (and many other) currently used physical concepts. We use them because we are used to them, not because we understand them. I am inclined to think that the roots of the entropy concept are simpler and easier to grasp than those of energy, mass and temperature, and I will try to illustrate this at the end of my talk.
- Different from energy and mass, *entropy is not a conserved quantity.* Any real thermodynamic process produces entropy. People facing the question of where the entropy comes from look in vain for a mysterious external source. In fact, entropy comes from the

lack of unlimited funds and time, from sloppy engineering, and from unreasonable operation. The task of the cryogenic engineer is to keep control of entropy proliferation.

2 *UN PEU D'HISTOIRE* (FIGURE 1)

Modern thermodynamics begins with a booklet of some 60 pages, entitled *Réflexions sur la puissance motrice du feu,* by Sadi Carnot, published in 1824. Carnot, born in 1796, was a young French officer and son of a hero of the French Revolution who was convinced that France's defeat in 1814/1815 was due not to military, but to industrial inferiority — particularly with respect to England. He considered that he could best serve his country by working on the emerging technology of steam engines. Carnot tried to treat heat engines as analogous with water mills. In a water mill, work is produced by a flow of water passing from a high to a low level; in a heat engine a flow of heat passes from a high to a low *temperature level.* The work produced by the water mill is proportional to the quantity of water and to the level difference. Carnot aimed at showing that in a heat engine the work produced is proportional to the quantity of heat and the temperature difference. Analysing a gas process, and using ideal gas relations (Boyle's law discovered 1662, Gay-Lussac's law discovered 1802, and Poisson's law, brand new, discovered 1822), he found, however, that the factor of proportionality itself depends on temperature. In modern terminology, we would write

$$\text{Work Output } W = \text{Heat Input } Q_{in} * f(T_{in}) * (T_{in} - T_{out}),$$

with

$$f(T_{in}) \leq 1/T_{in},$$

the equal sign referring to a reversible process. The factor $(T_{in} - T_{out})/T_{in}$ is called Carnot efficiency.

Carnot himself did not identify $f(T_{in})$ [1]. In fact, his fundamental assumption — the analogy between heat (*'le calorique'*) being the driving fluid of the heat engine as water is the driving fluid of the water mill — was not quite correct; whilst water input and output are equal, energy conservation requires a difference between heat input and output, the two being related to the work output W of the process by

$$\boxed{W = Q_{in} - Q_{out}\ .} \qquad (1)$$

However, in 1824 the law of conservation of energy was yet to be discovered; it took another 20 years until the *First Law of thermodynamics* was established[2] on the basis of systematic experiments by James Prescott Joule and William Thomson (who later became Lord Kelvin[3]). Little is known about Carnot's later work. He died in 1832, aged 36, from cholera and, as was the case in epidemics, was buried with all his papers and belongings. From

---

1 He writes: 'Nous sommes hors d'état de déterminer rigoureusement, avec les seules données expérimentales que nous possédons, la loi suivant laquelle varie la puissance motrice de la chaleur dans les différents degrés de l'échelle thermométrique...'.

2 The first publication (1842) of convincing arguments and a formulation of the general principle of energy conservation, with heat included as a manifestation of energy, are due to Robert (von) Mayer, a medical doctor with a strong interest for general science who lived in a small town in Germany. This outsider's publication was, however, largely ignored by the contemporary scientific community.

3 Joule (born 1818) and Thomson (born 1824) were not only brilliant scientists, but also strongly engaged in practical and technical work. Joule, the son of a wealthy brewer in Manchester, profited greatly in his experimental work from the availability of the excellent brewery workshop. Thomson, who did most of his scientific work at the University of Glasgow, made his fortune at a young age with an invention related to transatlantic telegraphy, for which he was knighted by Queen Victoria in 1892.

remaining fragments, it appears that he worked on experiments which would have demonstrated the equivalence of heat and work and that he had quite clear ideas about the impossibility of what was later to be called a *perpetuum mobile of the second kind.*

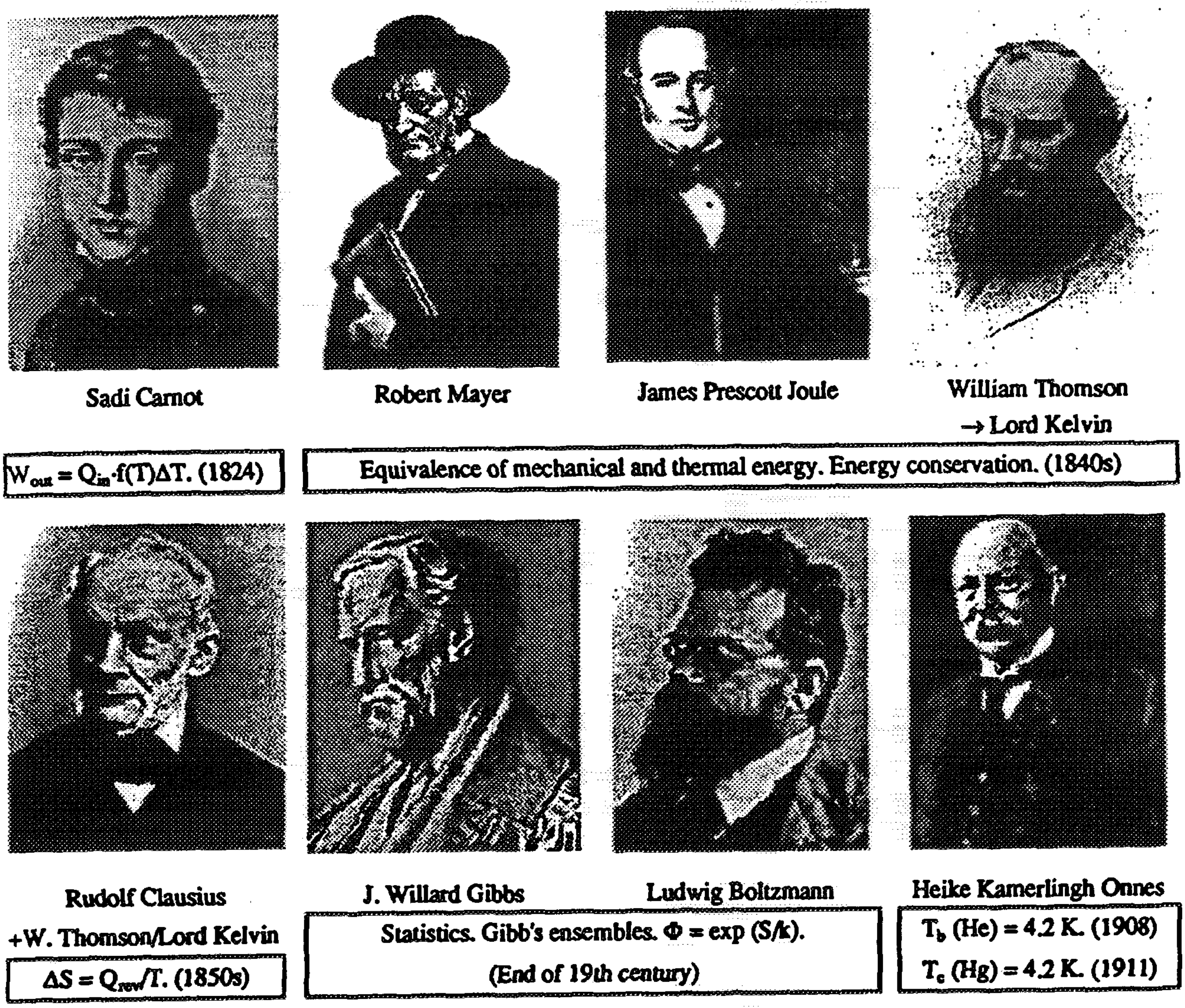


Fig. 1 The pioneers of thermodynamics

The *Second Law of thermodynamics* (formulated in 1850 by Rudolf Clausius[4]) can be illustrated by the statement that, for any continuous process driven by a flow of energy and involving heat input and output,

$$\boxed{\frac{Q_{out}}{T_{out}} \geq \frac{Q_{in}}{T_{in}}.} \qquad (2)$$

The equal sign stands for a *reversible process*, reversible meaning that input and output are interchangeable. This *reversible* process indeed matches Carnot's picture, the mill now being

---

4 Clausius (born 1822) is the systematic thinker sometimes referred to as the person who made thermodynamics a science. In 1850, when he published his paper with formulations of both the First and the Second Law, he was named professor of physics at the Artillery and Engineering School at Berlin. Later, he became a professor at Zürich, Würzburg and Bonn.

driven not by a flow of heat, but by a flow of a quantity Q/T, which Clausius called *entropy* S and which in the *reversible* case remains unchanged while passing through the engine. In the real, *irreversible* case, however, entropy increases — S_{out} is always greater than S_{in}. The process is said to have internal sources of entropy. The water mill analogy becomes somewhat artificial, but the entropy concept remains useful, the point of interest now being the nature of the entropy sources.

Clausius and Thomson proved that entropy is a *function of state*. This means that any substance in a given thermodynamic state (defined, say, by temperature and pressure) has a *measurable* content of entropy. Entropy can thus be *tabulated*, and entropy change associated with a change of thermodynamic state can be read from such tables as a difference between initial and final state. Table 1 is an example, showing data for helium.

Atomistic interpretations of the laws of thermodynamics were worked out in the last quarter of the 19th century by J. Willard Gibbs, Ludwig Boltzmann and others[5], who developed appropriate statistical methods. The atomistic picture gives a plausible interpretation of the fact that entropy is a function of state and that it inevitably increases in all real processes. Many scientists and philosophers, however, were suspicious about any speculation concerning the atomic nature of matter, and it took quite some time until the new theories were generally accepted.

Cryogenic techniques leading successively to the liquefaction of oxygen, nitrogen and hydrogen were developed in the late 19th century. Helium was first liquefied in 1908 by Heike Kamerlingh Onnes. Superconductivity was discovered in 1911 by Kamerlingh Onnes and his assistant Gilles Holst. The first signs of superfluidity were observed immediately, but the main phenomena were discovered only in the 1930s; the term 'superfluid' was coined by Pjotr Kapitza in 1938.

3 ENTROPY MILLS AND ENTROPY PUMPS

The picture of thermal engines developed by the generation of Joule, Thomson and Clausius is illustrated in Fig. 2. This picture constitutes the basis for the evaluation of the performance of cryoplants (and heat engines in general), and we will refer to it in the following.

The upper row of Fig. 2 refers to *heat engines*, the lower row to *heat pumps*. The left column refers to *ideal cycles*, the right column to *real cycles* with internal entropy production.

For the purpose of the drawing (arrow widths), we measure temperatures in units of the lower temperature. At the lower temperature level, the temperature is thus '1 temperature unit', and entropy ($\dot{Q}/T$) and heat ($\dot{Q}$) arrows have the same width. The *upper temperature is assumed to be three times higher than the lower temperature* ('3 temperature units'). At the upper temperature level, heat arrows are therefore three times larger than entropy arrows. Work arrows (= arrows of non-thermal energy $\dot{W}$ = arrows of energy that does not carry entropy) use, of course, the same scale as heat arrows. In all four diagrams it can be seen that energy is conserved.

---

5 Gibbs was born in 1839 in Connecticut and after five years of study in Europe became professor at Yale in 1871, where he spent the rest of his life far from philosophical quarrels about the significance of modern theories. Boltzmann, born in 1844 in Vienna, was professor at several Austrian and German universities, in particular at Vienna, centre of the positivistic school formed around Ernst Mach, who strongly criticised any attempt to introduce atomistic speculations into science. Boltzmann suffered deeply from the conflict and committed suicide in 1906. Mention must also be made of James Clerk Maxwell, born 1831, most famous for his work on electrodynamics. Maxwell, Gibbs and Boltzmann, although working without direct connection, highly respected each other's work.

Table 1

Helium

enthalpies and entropies for saturated liquid and vapour ⇒

and for gas

or subcooled liquid or supercritical fluid

⇓

saturated...		LIQUID		VAPOUR	
p [bar]	T [K]	h [J/g]	s [J/g K]	h [J/g]	s [J/g K]
1.0	4.208	9.94	3.560	30.74	8.492
1.2	4.407	11.07	3.782	30.61	8.208
1.4	4.584	12.20	3.997	30.28	7.933
1.6	4.744	13.38	4.213	29.72	7.650
1.8	4.891	14.67	4.442	28.87	7.339
2.0	5.026	16.17	4.707	27.48	6.952
2.2	5.151	18.90	5.199	18.90	5.199

	1 bar		2 bar		5 bar		10 bar		20 bar	
T [K]	h [J/g]	s [J/g K]	h [J/g]	s [J/g K]	h [J/g]	s [J/g K]	h [J/g]	s [J/g K]	h [J/g]	s [J/g K]
4	8.95	3.32	9.30	3.22	10.66	3.02	13.21	2.81	18.51	2.56
4.2	9.90	3.55	10.14	3.42	11.35	3.18	13.81	2.95	19.01	2.68
4.4	32.38	8.87	11.07	3.64	12.09	3.35	14.43	3.10	19.54	2.81
4.6	33.92	9.22	12.16	3.88	12.88	3.53	15.09	3.25	20.09	2.93
4.8	35.36	9.52	13.53	4.17	13.73	3.71	15.78	3.39	20.67	3.05
5	36.74	9.80	15.68	4.61	14.66	3.90	16.51	3.54	21.26	3.17
6	43.05	10.96	38.84	9.06	21.42	5.12	20.75	4.31	24.61	3.78
7	48.90	11.86	45.85	10.14	33.93	7.05	26.41	5.18	28.59	4.39
8	54.54	12.61	52.11	10.98	43.78	8.37	33.93	6.18	33.29	5.02
10	65.52	13.84	63.80	12.28	58.47	10.01	50.42	8.02	44.88	6.31
12	76.29	14.82	74.99	13.30	71.09	11.17	65.06	9.36	58.35	7.53
14	86.95	15.64	85.93	14.15	82.92	12.08	78.28	10.38	72.05	8.59
16	97.54	16.35	96.72	14.87	94.34	12.84	90.68	11.21	85.34	9.48
18	108.08	16.97	107.43	15.50	105.52	13.50	102.59	11.91	98.14	10.23
20	118.60	17.52	118.07	16.06	116.53	14.08	114.18	12.52	110.52	10.88
30	170.90	19.64	170.74	18.19	170.29	16.26	169.61	14.77	168.62	13.25
40	223.01	21.14	223.03	19.70	223.09	17.78	223.23	16.32	223.66	14.83
60	327.04	23.25	327.22	21.81	327.75	19.90	328.66	18.46	330.50	17.00
80	430.97	24.75	431.22	23.31	431.96	21.40	433.21	19.96	435.70	18.51
100	534.86	25.91	535.15	24.47	536.00	22.56	537.42	21.12	540.25	19.68
120	638.74	26.85	639.05	25.41	639.96	23.51	641.48	22.07	644.50	20.63
140	742.62	27.65	742.93	26.21	743.88	24.31	745.45	22.87	748.58	21.43
160	846.48	28.35	846.81	26.91	847.77	25.01	849.37	23.57	852.57	22.13
180	950.35	28.96	950.67	27.52	951.65	25.62	953.27	24.18	956.51	22.74
200	1054.20	29.51	1054.50	28.07	1055.50	26.16	1057.10	24.72	1060.40	23.29
220	1158.10	30.00	1158.40	28.56	1159.40	26.66	1161.00	25.22	1164.30	23.78
240	1261.90	30.45	1262.30	29.01	1263.20	27.11	1264.90	25.67	1268.10	24.23
260	1365.80	30.87	1366.10	29.43	1367.10	27.53	1368.70	26.09	1372.00	24.65
280	1469.70	31.25	1470.00	29.81	1471.00	27.91	1472.60	26.47	1475.80	25.03
300	1573.50	31.61	1573.80	30.17	1574.80	28.27	1576.40	26.83	1579.70	25.39
350	1833.20	32.41	1833.50	30.97	1834.40	29.07	1836.00	27.63	1839.20	26.19
400	2092.80	33.11	2093.10	31.67	2094.10	29.76	2095.60	28.32	2098.80	26.89
450	2352.50	33.72	2352.80	32.28	2353.70	30.37	2355.30	28.94	2358.40	27.50
500	2612.10	34.26	2612.40	32.83	2613.30	30.92	2614.90	29.48	2617.90	28.04
550	2871.80	34.76	2872.10	33.32	2873.00	31.42	2874.50	29.98	2877.50	28.54
600	3131.40	35.21	3131.70	33.77	3132.60	31.87	3134.10	30.43	3137.10	28.99

An inlet/outlet temperature ratio of 3 would be respectable for a heat engine (900 K/300 K), but can be considered at best as a borderline case for a cryoplant (300 K/100 K). For a liquid-helium cryoplant this ratio would be $300/4.2 \approx 70$.

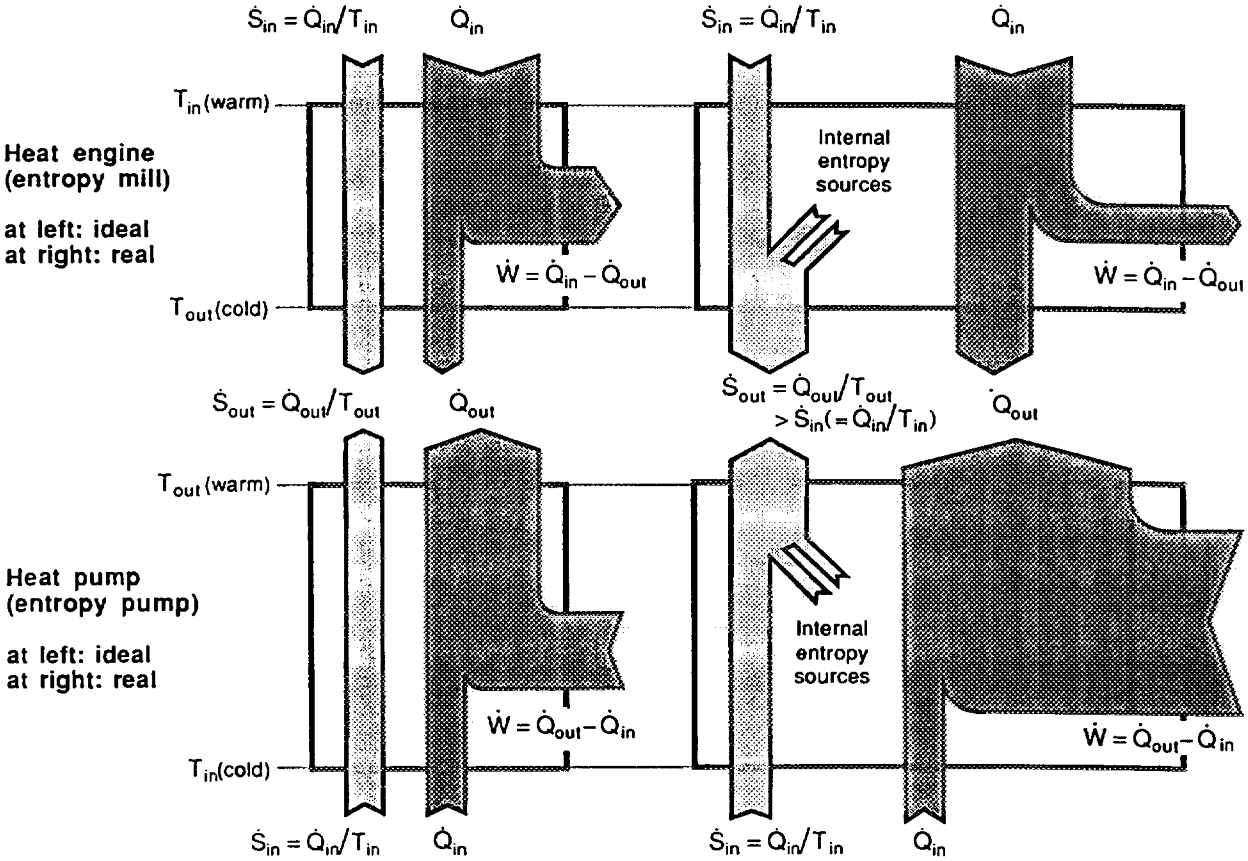


Fig. 2 Entropy flow $\dot{S}$ and heat flow $\dot{Q}$ in heat engines and heat pumps

For the *ideal cycles* at left in Fig. 2, input and output entropies are equal. For the heat engine, heat output is 1 unit, heat input is 3 units, and the difference (2 units) is available as work. For the heat pump, heat input is 1 unit, heat output is 3 units, whilst the difference (2 units) must be made up by work.

For the *real cycles* at right, internal entropy sources add an amount of entropy equal to the input entropy, with the result that the output entropy is twice the input entropy. For the heat engine, the heat output increases from 1 to 2 units; since the heat input is unchanged at 3 units, only 1 unit instead of 2 is available as work. For the heat pump, the heat output increases from 3 to 6 units, and the difference between these and the unchanged heat input of 1 unit (5 units) must be made up by work.

In Fig. 2 it is tacitly assumed that the heat input $\dot{Q}$ occurs at a well-defined temperature, T_{in}. The procedure can be generalised to heat input in a *range* of temperatures — for example, heat input between 30 K and 75 K when cooling a thermal shield, or heat input between saturated liquid conditions at 4.5 K and 300 K when cooling current leads to a superconducting magnet. In such cases, the entropy input for a given heat input must be taken from tables.

Note that the world record for efficiency of large helium cryoplants is presently held by the LEP2 cryoplants at CERN (which will be discussed in Section 7). In these plants, the incoming entropy is not only doubled, but tripled by internal entropy sources. In other, smaller or older plants (enormous progress has been made in the last ten years), the entropy increase due to internal entropy sources may be by a factor of 5 to 10 or more. A paper size of about 1.5 × 1.5 m^2 would be needed in order to show the heat and work arrows to correct scale (with the width of the entropy input arrow unchanged).

Table 2 gives a scheme for evaluating the performance of cryoplants in line with the principles of Fig. 2. The table will be referred to in the discussion of cryoplant cycles. It contains some technical terms about which details will be given in later sections.

Table 2

Scheme to evaluate cryoplant performance according to the principles of Fig. 2

1	Define the duty loads (heat load $\dot{Q}_{in}$ and entropy load $\dot{S}_{in}$) of the cryoplant. Examples: – A heat load of 1 W/4.5 K constitutes an entropy load of 1/4.5 W/K ≈ .222 W/K, – a liquefier load of 1 g/s, provided as saturated liquid at 1.4 bar and recovered as gas at 1 bar, 300 K constitutes, according to Table 1 a heat load (Δh) of 1'573.5–12.2 = 1571.3 W/g and an entropy load (Δs) of 31.61–3.997 = 27.67 W/g K, – etc. Include parasitic loads. When calculating the entropy load, decide whether you will count the entropy production during heat transfer from the load to the cycle as part of the load or as part of the cycle.
2	Identify internal entropy sources. Entropy sources regularly encountered will be discussed in Sections 5 and 6 (heat transfer from the compressor coolers to the environment in Section 5.1, entropy production associated with compression, expansion, counterflow heat exchange, internal heat and gas leaks and unsteady operation conditions in Sections 6.1 to 6.5).
3	Identify enthalpy extraction $\dot{H}_{exp}$ by expansion devices (enthalpy drop in piston expanders and turbines). Decide whether the extracted enthalpy will be fed back into the cycle as work or whether it will be dissipated as heat. In virtually any practical cryoplant, it will be dissipated at ambient temperature; if this is the case, treat the resulting entropy, obtained by dividing the extracted enthalpy by the ambient temperature, as an internal entropy source.
4	Adding entropies from the duty loads and from the internal entropy sources gives the output entropy $\dot{S}_{out}$
5	Dividing the output entropy by the ambient temperature gives the output heat $\dot{Q}_{out}$.
6	The required power (work) input is $\dot{W} = \dot{Q}_{out} - \dot{Q}_{in}$.

4 PROPERTIES OF THE IDEAL GAS AND REAL GASES. THE TEMPERATURE-ENTROPY DIAGRAM

4.1 Terminology

The quantities of interest for us are pressure p, temperature T, volume V, internal energy U, enthalpy H (about which more will be said in Section 4.2) and entropy S. They are all functions of state, which means that their numerical value can be taken from a table or a diagram. Many variables exist for which this is not the case, the most important ones in thermodynamics being heat and work. Both are forms of energy, and, knowing the state of something, we can say how much energy it contains, i.e. how much energy was necessary to bring it to its present state from a state of reference (say $T = 0\,K$, $p = 1$ bar). However, we cannot say how much of it was provided in the form of heat and how much in the form of work.

In the following, we will generally use upper case for absolute values and lower case for specific values per unit mass, i.e. v, u, h and s for specific volume, internal energy, enthalpy and entropy per — for example — gram: $v = V/m$, ..., where m is the mass of the sample.

Exceptions to this are capitals used for C_p and C_v, the specific heats *per mole* [J/mol K] at constant pressure and volume, respectively. Using molar specific heats is justified because they are closely related to the universal gas constant $R \approx 8.31$ J/mol K, also a molar quantity. Individual specific heats and the individual gas constant of a substance are obtained by dividing the molar values by the molecular mass M [g/mol]: $c_p = C_p/M$ etc. Such quotients will often appear in a technical context.

Hybrid units within the metric system are still frequently used in cryogenics — for example, 'Joule/g' for specific energies (the Joule in the numerator is based upon the kg, $1\,J = 1\,kg\,m^2\,s^2$, rather than upon the gram used in the denominator). Consistency of units must be checked, and conversion factors (usually powers of 10) may be necessary to obtain the result in the desired unit.

4.2 Enthalpy

When describing thermodynamic processes, we have the choice of the variables best adapted to the problem. Two sets are particularly convenient:

T, V, U and S	for illustration of physical context,
T, p, H and S	for quantitative treatment of technical systems.

Internal energy is the true thermal energy of the molecules. It is *constant in a closed system* in thermal equilibrium that does not exchange energy with its environment (for example, because its volume is kept constant and it is thermally insulated), but internal energy *may vary in an open system* connected to external energy sources or sinks. In open systems it is convenient to define functions of energy that are constant under the actual conditions. For our purposes the most important function is enthalpy, introduced for dealing with compressible systems, which, when travelling in an ambience of varying pressure, change volume and therefore exchange energy (work) with their environment. Their internal energy is not constant, but the energy passed to the environment can be treated as potential energy pV of the travelling sample, and, in the absence of other external influences, the sum of the two is again constant. This sum is called

$$\text{Enthalpy} \quad H = U + p * V. \tag{3}$$

In fact, introducing enthalpy (or other functions of energy) is a trick of bookkeeping. It depends on the situation whether I focus on the money in my pocket (when shopping), or on my bank account (when making standing orders), or on the sum of both (when paying taxes).

4.3 Entropy — more

The form in which entropy has been introduced was matched to a particularly simple example — the water mill/pump analogue for a heat engine, in which the parts of heat and work in energy transactions are clearly defined and hidden transformations of work into heat are excluded. Not all situations encountered in practice are so clear. For an unambiguous definition of entropy it is necessary to refer to internal energy rather than to heat (which, as mentioned, is not a function of state). The following equation is suitable for comparison of entropies when *no changes of volume* are involved:

$$\Delta S = \int_{\text{reference state}}^{\text{actual state}} \frac{dU}{T}$$

integration *at constant volume* (the formula does not account for entropy change due to volume change).

Textbooks often introduce Q_{rev} as the heat that would have been necessary to reach the state in question by a reversible approach, starting from an agreed reference state. By an irreversible approach, *less* heat would be required and the difference would be made up by work W:

$$Q_{rev} = Q_{irr} + W ,$$

$$Q_{irr} < Q_{rev} .$$

Heat that eventually lead to a measured increase of internal energy was not necessarily introduced as heat, but (in the case of an irreversible process) partly or totally as work. When concluding from heat input on internal energy change we must be sure that no complementary work input occurred.

The complete formulae relating entropy to traditional quantities and offering a basis for measurement (which is essentially a measurement of specific heat) are

$$s(T,v) = s(T_0,v_0) + \int_{T_0}^{T} \left(\frac{\partial s}{\partial T}\right)_v dT + \int_{p_0}^{p} \left(\frac{\partial s}{\partial v}\right)_T dp$$

with

$$\left(\frac{\partial s}{\partial T}\right)_v = \frac{m c_v}{T} \quad \left(\frac{\partial s}{\partial v}\right)_T = \left(\frac{\partial p}{\partial T}\right)_v .$$

4.4 The temperature–entropy diagram of the ideal gas

Figure 3 shows the isochors and isobars (lines of constant volume and constant pressure) of an ideal monatomic gas of molecular mass 4 g/mol ('ideal helium') in a logarithmic temperature-entropy diagram. The intention of this section is to illustrate the physical content of such a diagram and the quantitative relations underlying its structure.

In a T–s diagram, isotherms (lines of constant temperature) are obviously straight horizontal lines.

An ideal gas is characterised by the fact that its internal energy is proportional to temperature and independent of volume and pressure. For this reason, the isenergs (lines of constant internal energy) are parallels to the isotherms.

The equation of state of the ideal gas,

$$p * v = (R/M) * T , \qquad (4)$$

states that $p * v$, the term corresponding to the potential energy of the gas, depends on temperature only, hence isenthalps (lines of constant enthalpy) are horizontal lines as well. A few isenthalps are indicated in the diagram, whilst the isenergs have been omitted to avoid graphic overloading.

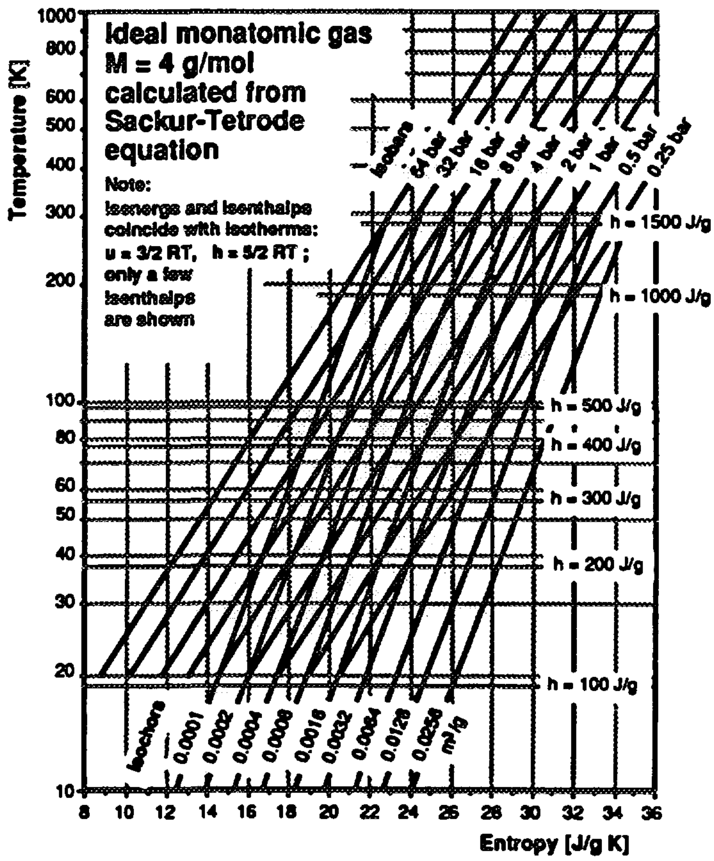


Fig. 3 Temperature-entropy diagram of 'ideal helium'

The universal gas constant in (4),

$$R \approx 8.31 \text{ J/mol K},$$

is related to Boltzmann's constant,

$$k \approx 1.38 * 10^{-23} \text{ J/K},$$

and Avogadro's number,

$$N_A \approx 6.02 * 10^{23} \text{ mol}^{-1},$$

by

$$R = N_A * k.$$

Specific heat at constant volume, $c_v = C_v/M$, is introduced as the proportionality factor between internal energy and temperature and, similarly, specific heat at constant pressure, $c_p = C_p/M$, as the proportionality factor between enthalpy and temperature:

$$u = c_v T = (C_v/M) * T, \quad h = c_p T = (C_p/M) * T; \qquad (5\text{ a/b})$$

with $h = u + p*v = u + (R/M)*T$ according to (3) and (4) following

$$c_p = c_v + R/M, \quad C_p = C_v + R. \qquad (6\text{a/b})$$

From atomistic considerations it follows that for ideal *monatomic* gases $C_v/R = 3/2$, from which in turn follows with (6) $C_p = 5/2$ and $C_p/C_v = c_p/c_v = 5/3$ for the monatomic gas.

Injection of a small amount of heat $c_v\,dT$ (at constant volume) or $c_p\,dT$ (at constant pressure) into a gas sample should lead, according to Clausius' assumption, to an increase ds of entropy:

$$\text{isochoric case:}\quad ds=\frac{c_v dT}{T}=c_v\,d\ln T \;\Rightarrow\; \left(\frac{\partial s}{\partial \ln T}\right)_v=c_v\;,$$

$$\text{isobaric case:}\quad ds=\frac{c_p dT}{T}=c_p\,d\ln T \;\Rightarrow\; \left(\frac{\partial s}{\partial \ln T}\right)_p=c_p\;. \qquad \text{(7a/b)}$$

From this follows that in a log T-*versus*-s diagram, isochors are straight lines with a slope of $1/c_v$ and isobars are straight lines with a slope of $1/c_p$. This is the reason for choosing a logarithmic scale for the ordinate of the temperature–entropy diagram.

Poisson's equation in its three variations [the second and third following from the first one with the ideal-gas equation (4) and the specific-heat relation (6b)] requires that *for states of equal entropy*

$$\left(\frac{p_2}{p_1}\right)=\left(\frac{v_2}{v_1}\right)^{-C_p/C_v},\quad \left(\frac{T_2}{T_1}\right)=\left(\frac{v_2}{v_1}\right)^{-R/C_v},\quad \left(\frac{T_2}{T_1}\right)=\left(\frac{p_2}{p_1}\right)^{R/C_p}. \qquad \text{(8a/b/c)}$$

In differential form (after logarithmic differentiation) we obtain

$$C_v\,d\ln p+C_p\,d\ln v=0,\qquad C_v\,d\ln T+d\ln v+0,\qquad C_p\,d\ln T-R\,d\ln p=0$$

for isentropic states, i.e. for ds = 0.

Generalising, i.e. inserting ds for 0 in these equations, we find the following equations, in which Poisson's equation is included:

$$ds=C_v\,d\ln p+C_p\,d\ln v,\qquad ds=C_v\,d\ln T+R\,d\ln v,\qquad ds=C_p\,d\ln T-R\,d\ln p\;.$$

By integration we obtain the following three formulae:

$$\begin{aligned} s&=C_v\ln p+C_p\ln v+\text{const}\,,\\ s&=C_v\ln T+R\ln v+\text{const}\,,\\ s&=C_p\ln T-R\ln p+\text{const}\,.\end{aligned} \qquad \text{(9a/b/c)}$$

The second and the third of these equations represent the isochors and isobars in the T–s diagram. Obviously, exponential progression on volume and pressure (for example, steps 1, 2, 4, 8, ...) leads to equidistant isochors and isobars in our logarithmic representation.

In most cases, entropy differences rather than absolute values are required for practical calculation purposes, therefore the undefined integration constants in (9) are not a very serious shortcoming. Absolute values of entropy can, however be measured and, in certain cases, also calculated; the formula for the ideal gas, used for the calculation of the Fig. 3 diagram, was derived between 1911 and 1913 by O. Sackur and H. Tetrode on the basis of early quantum-statistical considerations.

Note that the quoted physical relations, except those of (7), were essentially known to Carnot.

4.5 The temperature–entropy diagram of real gases. Helium

Real gases show characteristic departures from ideal gas relations. Particularly important are those for isenthalpic changes of state, which occur when a compressible fluid flows, without heat input or output, in a conduit where it may undergo pressure variations.

A temperature-entropy diagram for a real gas (in this case helium) is sketched in Fig. 4. It can be seen that at ambient temperature and moderate pressure the isenthalps are in good approximation, as postulated for the ideal gas, horizontal straight lines, and that the isobars and isochors in this logarithmic representation are sloping straight lines. However, at low temperature or high pressure or both, departures from ideal gas behaviour are seen. Thus, near state A {p = 200 bar, h = 1000 J/g} we notice that the 1000 J/g isenthalp rises from left to right; an isenthalpic expansion from 200 bar to 60 bar would lead to a temperature rise from 177 K to 186 K. Near state B {p = 10 bar, h = 50 J/g}, however, we notice that the 50 J/g isenthalp drops from left to right; an isenthalpic expansion from 10 bar to 1 bar would lead to a temperature decrease from 10 K to 7.2 K. Finally, an isenthalpic expansion from state C {p = 4 bar, h = 16 J/g} to p = 1 bar would lead to a state below the bell-shaped curve ('saturation curve') that separates the one-phase from the two-phase domain of the fluid.

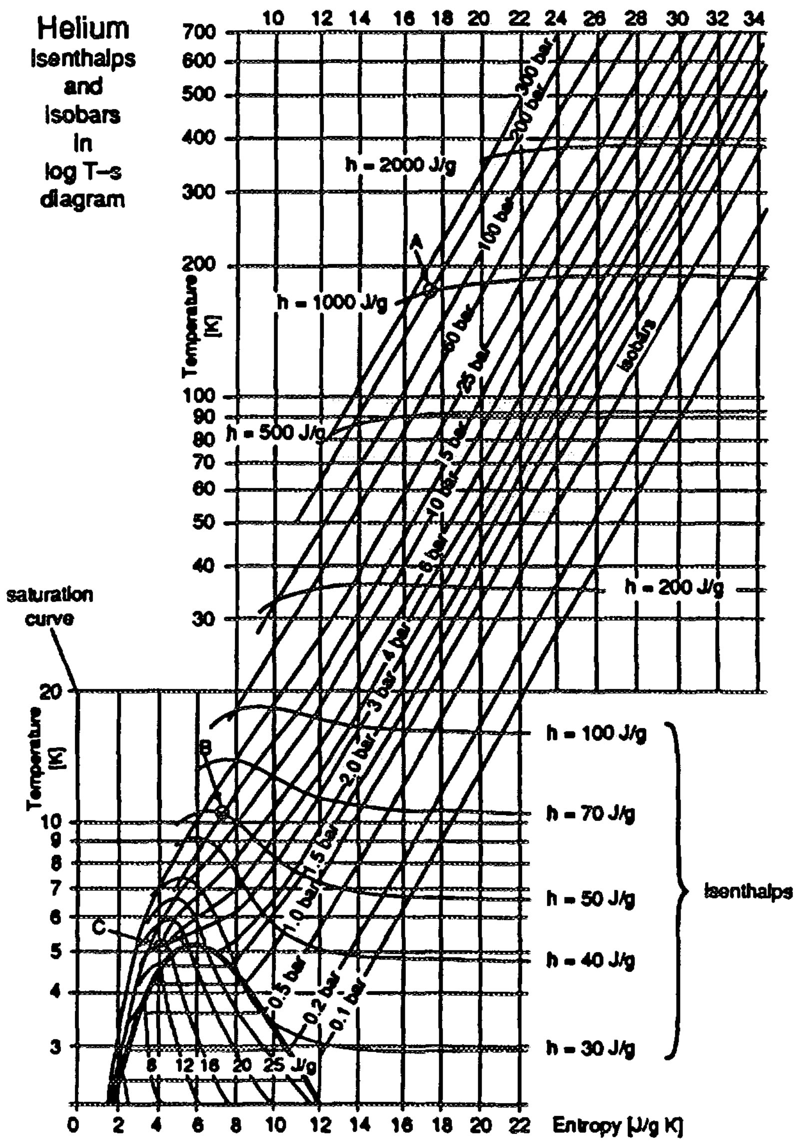


Fig. 4 Temperature-entropy diagram of real helium

Careful measurements of these phenomena have been taken since 1852 by Joule and Thomson, who studied the departure of real gases from ideal gas laws. The temperature increase or decrease after isenthalpic pressure changes is called the *Joule-Thomson effect*. The observed phenomena were explained in 1876 by Johannes van der Waals[6] on an atomistic basis. Van der Waals assumed that real molecules, different from ideal-gas particles, interact by small but finite attractive forces and have a small but finite volume. Both facts lead to a volume-dependence of enthalpy in states of high density. Molecular attraction lets internal energy u and enthalpy h = u + pv *decrease* with decreasing distance between molecules (in the T–s diagram: when going from right to left at constant temperature); the energy reduction is independent of temperature and becomes negligible when internal energy is high at high temperature. Finite molecular volume has no effect on internal energy, but influences the term pv in enthalpy. At high temperature, it lets enthalpy *increase* with decreasing intermolecular distances. The behaviour of helium at state A in Fig. 4 is thus explained by the finite molecular volume, the behaviour at state B by the intermolecular attraction, which, in the region below state C, finally leads to liquefaction.

5 CRYOPLANTS

5.1 Evaporative refrigeration at temperatures close to ambient

As a particularly simple example, let us discuss the following cycle, suitable for refrigeration at close-to-ambient temperatures and therefore preferred in domestic appliances.

The cycle uses a refrigerant fluid which at ambient conditions is gaseous but readily liquefies when compressed. A high molar heat of evaporation is desirable. Typical examples are shown in Table 3.

Table 3

Refrigerants for evaporative cooling close to ambient temperature

	Boiling temperature at 1 bar	Saturation pressure at 0 °C	Latent heat of evaporation	Critical temperature	Critical pressure
Refrigerant R40 = CH_3Cl	–24 °C	5.0 bar	17.6 kJ/mol	143 °C	67 bar
Refrigerant R12 = CCl_2F_2	–30 °C	5.8 bar	20.2 kJ/mol	112 °C	41 bar
Refrigerant R22 = $CHClF_2$	–41 °C	9.4 bar	20.2 kJ/mol	96 °C	49 bar
Propane = C_3H_8	–42 °C	8.3 bar	18.8 kJ/mol	97 °C	42 bar
Isobutane = $(CH_3)_3CH$	–11 °C	3.0 bar	21.3 kJ/mol	135 °C	37 bar
Ammonia = NH_3	–33 °C	8.7 bar	23.3 kJ/mol	132 °C	115 bar
Sulphur dioxide = SO_2	–10 °C	3.2 bar	24.9 kJ/mol	158 °C	79 bar

Figure 5 shows the refrigerator components (a compressor with its aftercooler-heat exchanger, a throttling device and an evaporator equipped with external fins for heat exchange with the load) and the T–s diagram of the cycle.

6 Van der Waals, born 1837, Nobel laureate 1910, was a self-taught scientist who for several years worked in public schools as a teacher before being nominated professor of physics at the newly created University of Amsterdam in 1877. He had already attracted great attention with his doctoral thesis, entitled 'On the continuity of the liquid and gaseous state', presented in 1873.

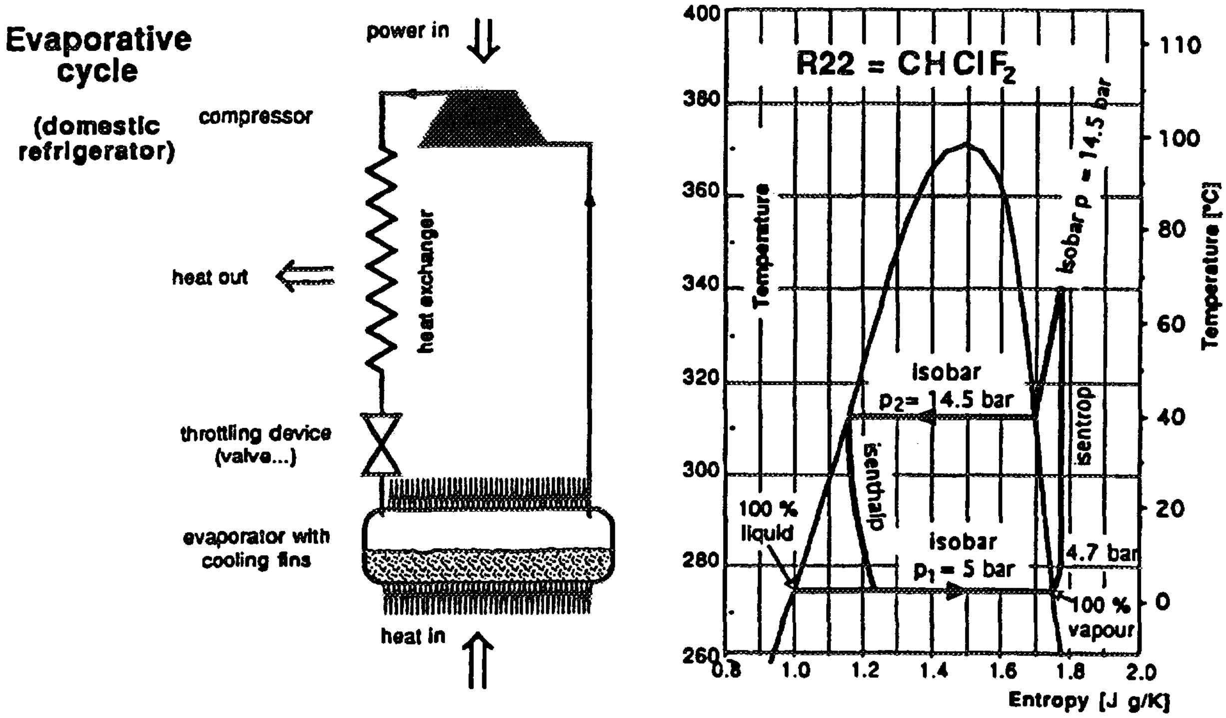


Fig. 5 Flow scheme and temperature-entropy diagram of the evaporative refrigeration cycle (reverse Hirn cycle)

The cycle operates between pressures p_1 (compressor suction) and p_2 (compressor discharge), with p_1 the vapour pressure of the refrigerant at the *desired cooling temperature* T_1, and p_2 being chosen somewhat higher than the vapour pressure of the refrigerant at *ambient temperature* T_a. It is convenient to use a refrigerant for which p_1 is somewhat (but not much) higher than 1 bar (atmospheric pressure).

Under compressor suction conditions (p_1, $T_1 \approx T_a$) the refrigerant is gaseous. By adiabatic (approximately isentropic)[7] compression from p_1 to p_2, the refrigerant temperature is raised from T_1 to T_2, well above T_a; the refrigerant is still gaseous. After cooling in the aftercooler from T_2 to T_a at constant pressure (i.e. along the p_2 isobar), associated with heat transfer to the environment, the refrigerant is completely liquefied. The liquid passes through the throttling device (valve, pipe constriction etc.), where its pressure drops from p_2 to p_1. The throttling, an isenthalpic process, leads to phase separation into a low-enthalpy, low-entropy liquid phase and a high-enthalpy, high-entropy vapour phase in such proportions, x_{liq} and $x_{gas} = 1 - x_{liq}$ that the initial enthalpy h is conserved:

$$x_{liq}\, h_{liq} + (1-x_{liq})\, h_{gas} = h\ .$$

The available refrigeration capacity for extraction of entropy or heat is

$$\dot{S}_{in} = \dot{m}_c \cdot x_{liq} \cdot \left(s_{gas} - s_{liq}\right); \quad \dot{Q}_{in} = \dot{m}_c \cdot x_{liq} \cdot \left(h_{gas} - h_{liq}\right) = T_1 \cdot \dot{S}_{in}\ , \qquad (10\ a/b)$$

where $\dot{m}_c$ is the compressor throughput.

---

7 *Adiabatic* means that the system neither gives nor takes *thermal* energy. *Isentropic* means that the entropy does not change during the process considered. Since thermal energy input or output always involves a change of entropy, adiabasy is a necessary condition for isentropy, but not a sufficient one: a change of entropy (always towards increasing entropy) is possible even without thermal energy transfer by internal dissipation of non-thermal energy ('work').

Note that this is the entropy *received* by the cycle; the entropy *delivered* by the load, which must be at a higher temperature than the gas in order to ensure the desired heat transfer, is smaller.

The entropy discharge to the environment is greater than $\dot{S}_{in}$. Obviously, some entropy is produced by the isenthalpic throttling process (isenthalpic expansion left–right towards increased entropy). Also, rather than being exactly isentropic (i.e. vertical), the compression line will slope from left to right towards increased entropy. Numeric values of entropy production can be determined from tables on the basis of measured input and output data (for example, pressure and temperature).

A major entropy source (often dominant) is the aftercooler. By heat transfer to the environment, the enthalpy of the process gas (inlet and outlet temperatures T^+ and T^-, enthalpies h^+ and h^- and entropies s^+ and s^- assumed) is reduced by an amount $\dot{H}=\dot{m}_c*(h^+-h^-)$. The entropy of the process gas is reduced by $\dot{S}_{HX}=\dot{m}_c*(s^+-s^-)$. The entropy of the environment is, however, increased by the much larger amount $\dot{S}_{env}=\dot{H}/T_a$, where T_a is the ambient temperature. That $\dot{S}_{env}>\dot{S}_{HX}$ follows from the fact that at any longitudinal element of the heat exchanger the heat exchanger temperature T is greater than the ambient temperature T_a. The entropy production is

$$\dot{S}_{env}-\dot{S}_{HX}=\frac{\dot{H}}{T_a}-\dot{m}_c\left(s^+-s^-\right)=\int_{T^-}^{T^+}\frac{\dot{m}c_p dT}{T_a}-\int_{T^-}^{T^+}\frac{\dot{m}c_p dT}{T}=\dot{m}c_p\int_{T^-}^{T^+}\left(\frac{1}{T_a}-\frac{1}{T}\right)dT>0.$$

With this information about internal entropy sources, the performance of the plant can be evaluated using Table 2 and Fig. 2.

5.2 Large-scale liquefaction of oxygen, nitrogen, air etc: The Linde Cycle

The cycle introduced by Carl von Linde[8] in 1895 aimed at large-scale air liquefaction. Air, oxygen and nitrogen belong to what were then called 'permanent' gases, because they will not liquefy under any pressure at ambient temperature: their critical temperature[9] is below ambient. The process described in the preceding section thus cannot be used. Before 1877, when Georges Cailletet in Chatillon-sur-Seine (France) and Raoul Pictet in Geneva succeeded in liquefying minute quantities of oxygen, it was doubtful whether 'permanent' gases could be liquefied at all.

Cailletet used a single-shot non-cyclic process. Pictet used a 'cascade process', precooling his oxygen cycle by a CO_2 cycle (critical temperature 32 °C), in turn precooled by an SO_2 cycle (critical temperature 158 °C). By pumping on the CO_2 cycle, he reached a temperature of about −140 °C ≈ 130 K, sufficiently below the critical temperature of oxygen, 155 K, to allow a small quantity, compressed to 320 bar, to be liquefied. Cascade processes continued to be used for a long time afterwards. Historically, they were used for the first liquefaction of hydrogen by James Dewar in 1898 and of helium by Heike Kamerlingh Onnes in 1908. Kamerlingh Onnes' apparatus is shown in Fig. 9 as an example of a cryoplant made before the age of stainless steel and aluminium. Cascade cycles are, however, no longer of practical interest in large modern laboratories.

---

8 Linde belonged to a new generation of engineers who systematically used the new thermodynamics for optimised plant design on the basis of theoretical considerations. The economic production of oxygen from air by cryogenic separation (1901), an achievement of great importance for steel manufacture, had been an early target of Linde's systematically pursued work.

9 Critical temperature = temperature of the highest point of the saturation curve in the T–s diagram.

Linde ensured the precooling of his final expansion stage by inserting a counterflow heat exchanger (see Fig. 6) between the compressor and the expansion device. He made use of the fact that the *inversion temperature*[10] for oxygen, nitrogen and air is above ambient temperature, leading to a Joule-Thomson effect that, even at ambient temperature, ensures an appreciable temperature decrease when high-pressure (200–300 bar) gas is expanded. In the counterflow heat exchanger, the low-pressure gas, cooled after expansion in the 'Joule-Thomson valve' and returning to the compressor suction, precools the high-pressure gas prior to expansion. Because of the gradually decreasing temperature upstream of the Joule-Thomson valve, the cooling effect is enhanced until eventually the expansion leads to liquefaction.

Linde cycle: Carl Von Linde 1985. First industrial liquefaction of air

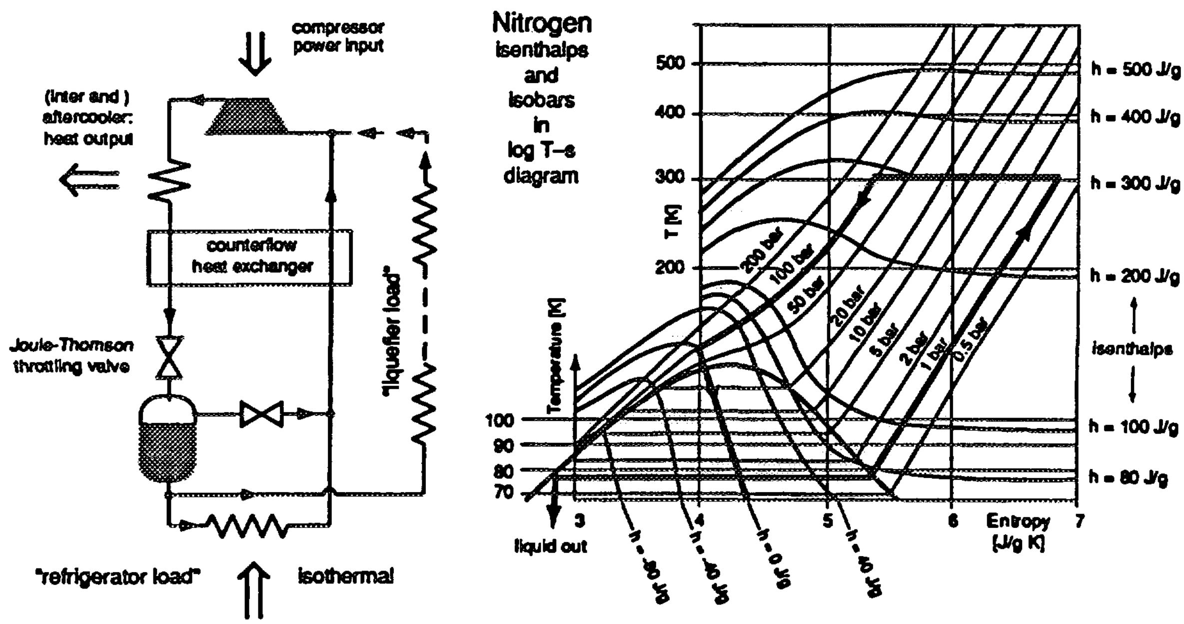


Fig. 6 Flow scheme and temperature-entropy diagram of the Linde cycle, using counterflow heat exchangers and making use of the Joule–Thomson effect

The phenomena of counterflow heat exchangers will be discussed in more detail in Section 6.3. Two typical technical designs are shown in Fig. 7.

The total entropy production of the heat exchanger follows from the specific entropy values at its terminals and from the known mass flow values $\dot{m}$:

$$\dot{S}_{\text{heat exchanger}} = \sum_i \dot{m}_i s_i \ ,$$

the summation extending over all i terminals (2 inlets and 2 outlets in case of a two-passage heat exchanger), with inlets being counted positive and outlets negative.

With this entropy source included, the performance of the plant can be evaluated by means of Table 2 and Fig. 2.

---

10 Inversion temperature = the temperature that, at a given pressure, separates the regions where the Joule-Thomson effect leads to heating and to cooling (somewhere between A and B in Fig. 4).

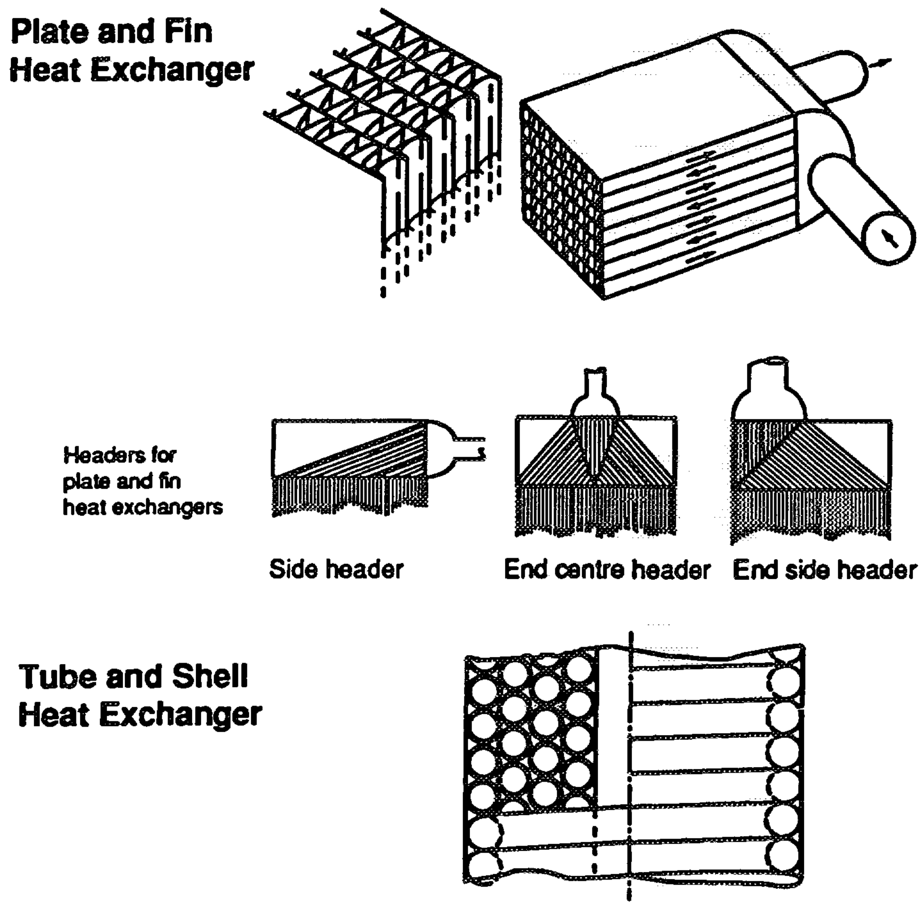


Fig. 7 Technical forms of simple counterflow heat exchangers

5.3 Work-extracting expanders: The Brayton Cycle

The cooling effect of the previously described cryoplant cycles is due to intermolecular forces within the refrigerant. If these forces are strong, they lead to direct liquefaction at ambient temperature when the gas is compressed (case C in Fig. 4). If they are weaker, they usually still lead to a perceptible cooling effect under isenthalpic expansion at ambient temperature (Joule-Thomson effect, case B in Fig. 4). In a few cases, however — namely neon, hydrogen and helium, the intermolecular forces are so weak that at ambient temperature liquefaction is not only impossible, but isenthalpic expansion leads to heating rather than cooling (case A in Fig. 4).

Cooling of such gases, and, generally, the cooling of the ideal gas, in which no intermolecular forces exist, requires an *expansion engine* to extract molecular energy to the outside. In an expansion engine the gas molecules discharge part of their kinetic energy to a receding piston or a turbine wheel, which then transmits it to the outside. The resulting cooling effect is the inverse of the heating effect in a compressor.

A particularly simple cycle consisting of a compressor, an expansion engine, a counterflow heat exchanger between the two, and the heat exchangers linking the cycle to the load and to the environment, is shown in Fig. 8. The T–s diagram shows the two adiabatic (ideally isentropic) phases of compression and expansion with input/output of external work, and the two isobaric phases of internal counterflow heat exchange. For the ideal cycle (full lines), heat is transferred from the load to the cycle at 1 bar between 70 K and about 120 K, and from the cycle to the environment at 4 bar between about 500 and 300 K. It is obvious from the drawing that the entropy transfer is equal in the two cases (horizontal distance between the 1-bar and 4-bar isobars equal at 300 K and at 120 K). For the real cycle (dashed lines), both the expansion and compression lines are diverted from vertical towards increasing entropies, which increases the entropy output and reduces the entropy input to the cycle. To simplify the argument, we have tacitly assumed that no entropy is produced in the counterflow heat exchanger, i.e. that a temperature difference between inlet and outlet exists neither at the cold nor at the warm end of the heat exchanger.

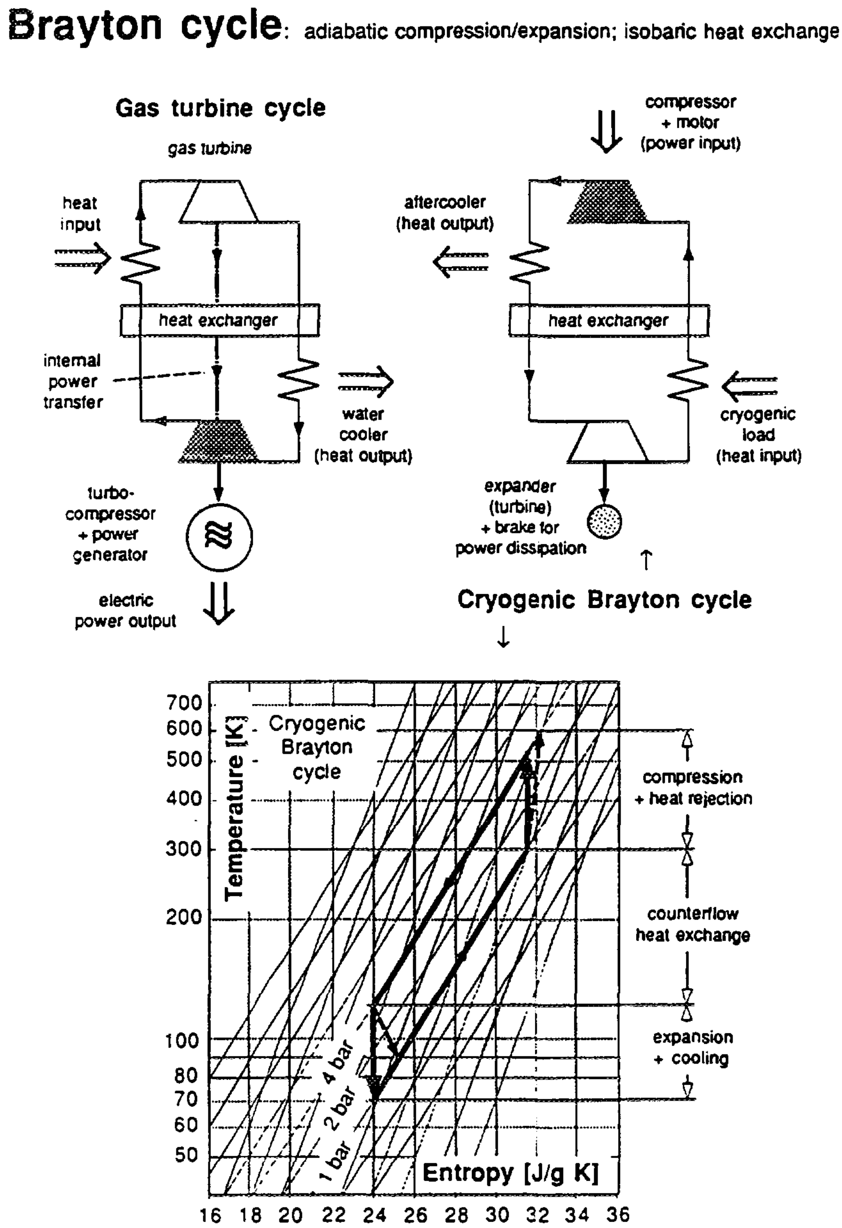


Fig. 8 Flow schemes and temperature-entropy diagram of the Brayton cycle. Top left: gas turbine cycle (Brayton cycle operating as heat engine). Top right: cryogenic Brayton cycle (Brayton cycle operating as heat pump). Bottom: Temperature-entropy diagram of Brayton cycle with ideal gas.

The principle of this cycle, generally referred to as the Brayton cycle, is widely used and by no means limited to helium, hydrogen and neon. George B. Brayton, an American engineer, worked on internal combustion engines in the late 1800s.

A few words should be said about expanders and expander cycles (in particularly turbine cycles).

A *piston expander* can be considered as an inverted reciprocating compressor. Molecules lose part of their momentum when reflected from the receding piston, thereby transferring energy to the piston, which then transmits it to the outside.

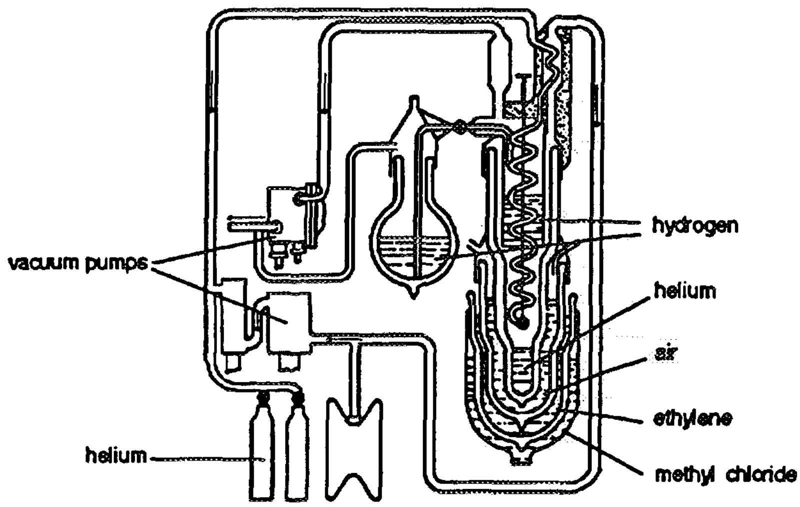


Fig. 9 Cascade liquefier used by Kamerlingh Onnes for the first helium liquefaction in 1908. From J.A. van Lammeren, Technik der tiefen Temperaturen, Springer, Berlin 1941, quoted by R. Plank, Hdb.d.Kältetechnik, Springer, Berlin, Vol. 1 (1954).

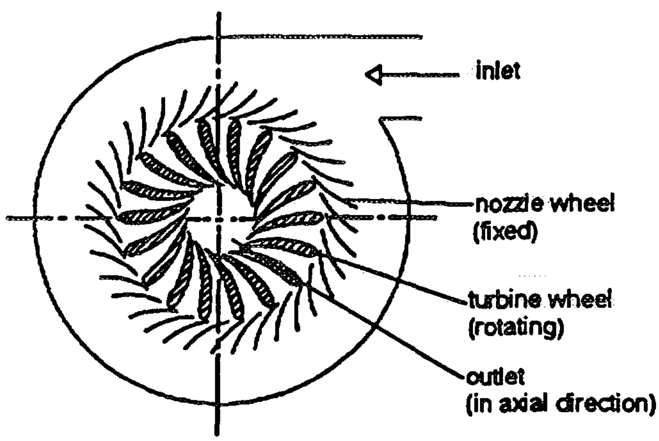


Fig. 10 Radial inflow - axial outflow gas turbine

A *turbine expander* consists of two wheels: the fixed nozzle ring and the rotating turbine wheel (see Fig. 10). In the nozzle wheel, the gas undergoes adiabatic expansion, during which its temperature is reduced according to equation (8c). The corresponding enthalpy reduction [equation (5b)] is balanced by an increase in kinetic energy of the gas[11], which is transferred to the turbine wheel and from there to an external device — for example, an electric generator or a device for producing useful mechanical work such as the compressor of a gas turbine cycle (see

---

[11] The reduction of enthalpy (temperature) of a gas and the corresponding increase of kinetic energy can be observed when the valve of a compressed-gas cylinder is opened and gas is discharged in a jet: the valve and the top of the gas cylinder cool down rapidly. The often heard, but incorrect explanation is that this is the Joule-Thomson effect — during the expansion, which is by no means isenthalpic, internal energy is transformed into kinetic energy. The effect corresponds to the expansion in the nozzle wheel of the turbine.

below), or both. In cryogenics, where energies involved are small and turbine design has to respect stringent requirements because of the low operating temperature (in particular excellent thermal insulation), the complication associated with energy recovery is economically not justified. The power extracted from a cryogenic turbine is therefore usually dissipated as heat in an appropriate external brake circuit at ambient temperature. Thus the expander produces the following entropies:

– on the process gas side (at low temperature) $\dot{m} * (S_{expander\ outlet} - S_{expander\ inlet})$

– on the brake side (at ambient temperature T_a) $\dot{m} * (h_{expander\ inlet} - h_{expander\ outlet}) / T_a$.

Both entropies must be treated as internal entropy sources when evaluating the performance of the plant by calculating the required power input according to Table 2.

It is plausible that the work required by the compressor is greater than the work delivered by the expander, ideally by a factor corresponding to the ratio of compressor temperature to expander temperature[12], because this gives the ratio of the gas volumina handled by the compressor and the expander. (It is volume that determines the energetic performance of a volumetric machine such as a compressor or an expander.)

If we assume that the work produced by the expander is entirely recovered, we find for the so-called Coefficient Of Performance (COP) of the ideal cycle the expression

$$\frac{\text{Work required by compressor} - \text{Work recovered from expander}}{\text{Work delivered by expander}} = \frac{T_{compressor} - T_{expander}}{T_{expander}},$$

the heat-pump analogue to the Carnot efficiency, usually given as 'watt per watt'. (The term 'coefficient of performance' is misleading: a *low* COP is desirable.)

The cryogenic Brayton cycle is the reverse of a *gas turbine cycle* of the same name, sometimes plainly referred to as *the* gas turbine cycle. Here (top left of Fig. 8) a turbine is driven by hot gas; the heat is injected either by internal combustion or via a heat exchanger upstream of the turbine. We consider the second case (the first case would lead to an open circuit without parallel in cryogenics). The mentioned heat exchanger corresponds to the water cooler downstream of the cryogenic compressor. The residual heat of the expanded gas is partly used to preheat, by counterflow heat exchange, the high-pressure gas feeding the turbine. The remaining heat is transferred to the environment by an air- or water-cooled heat exchanger. The gas, now at ambient temperature, is then recompressed to the turbine inlet pressure by a turbocompressor, preheated by the low-pressure gas in counterflow heat exchange as described, exposed to the external heat source and, having reached its final high temperature, is reinjected into the turbine. An essential point about the gas turbine cycle is that the compressor can be mechanically coupled to the turbine, which thus directly provides the driving power for the compressor.

5.4 Regenerators: The Stirling Cycle

Considering that a compressor is in principle an inverse expander, and vice versa, we notice a remarkable symmetry in the Brayton cycle (Fig. 11, top left). The Stirling[13] cycle profits

---

12 If temperatures are taken at corresponding points of the compressor and the expander — for example, inlet or outlet — it is easily shown that the argument is valid even though compression and expansion take place over a temperature range rather than at a fixed temperature.

13 Robert Stirling, a minister of the Church of Scotland, in 1816 invented and patented a hot-air engine using this cycle. Its potential to run as a refrigerater has been known since 1834 but was not much exploited. However, small Stirling-cycle hot-air engines driving a dynamo to power mobile radio sets that were developed and built by Philips in the early 1940s, turned out to be quite efficient as heat pumps, and Philips decided to develop them for cryogenic use. Since the 1960s, powerful cryogenerators for cooling at liquid-nitrogen temperature (up to 20 kW) and, in a two-stage version, at liquid-hydrogen temperatures, were brought to the market.

from this symmetry by using the same reciprocating engine alternatingly (half-stroke by half-stroke) as a compressor and an expander and by replacing the counterflow heat exchanger by a single-passage heat exchange/heat storage device called a ***regenerator***, through which the refrigerant shuttles between a warm compression space and a cold expansion space.

The principle of the Stirling engine is illustrated, in simplified form, in the lower part of Fig. 11. An essential component is the displacer, used to move the gas back and forth between the water-cooled warm compression space (at the bottom) and the cold expansion space equipped with an external heat exchanger connected to the load (at the top). The four phases of the cycle are explained in Table 4.

Stirling cycle: adiabatic compression/expansion isochoric regenerative heat exchange

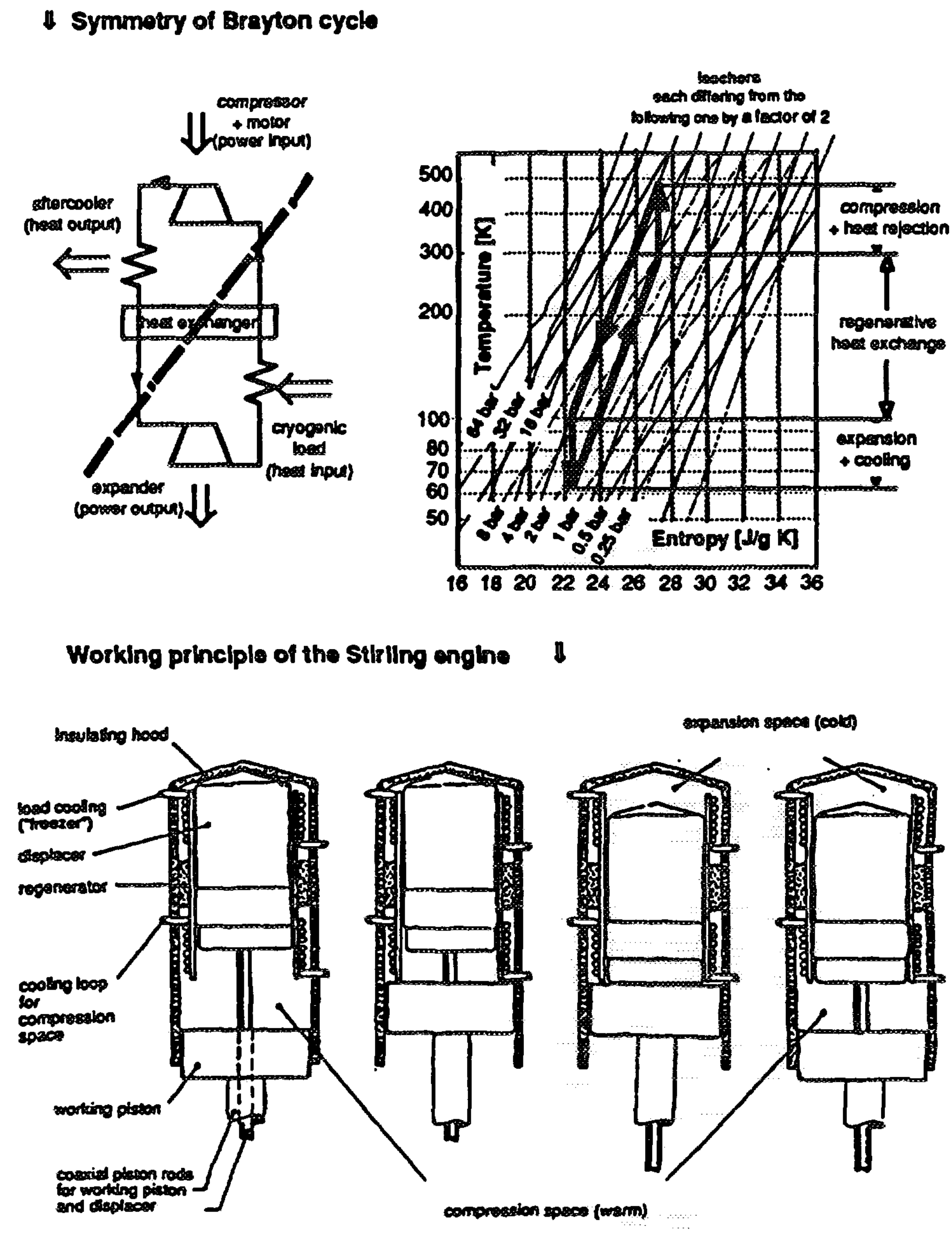


Fig. 11 The Stirling cycle. Top left: relation to the Brayton cycle. Top right: Approximation of temperature-entropy diagram of Stirling cycle. Bottom: Mechanical concept and motion of the working piston and the displacer. Note their phase difference of 90°.

Table 4

Operational phases of the Stirling cycle

Phase	Working piston position	Displacer position	Gas mainly located in
1	down	up	compression space
1→2	*compression*		
2	up	up	compression space
2→3	*gas displacement warm→cold*		
3	up	down	expansion space
2→4	*expansion*		
4	down	down	expansion space
4→1	*gas displacement cold→warm*		
1	down	up	compression space
...	...		

While moving between the warm and the cold space, the gas passes through the regenerator, a structure of large heat exchange surface and high heat capacity, consisting, for example, of stacked stampings of fine copper or phosphor bronze gauze, or of stratified lead beads. By contact with the gas, the regenerator is periodically heated at its warm end and cooled at its cold end, and a regular temperature gradient is established both in the regenerator and in the gas, exactly as in a counterflow heat exchanger. However, whilst the temperature distribution in the conterflow heat exchanger is established by continuous heat transfer from the warm to the cold gas, the regenerator is subject to alternating heat transfer gas→regenerator (when the gas moves warm→cold) and regenerator→gas (when the gas moves cold→warm).

At upper right of Fig. 11 is the T–s diagram of the ideal cycle, which is adiabatic-isochoric, since the regenerative cooling–heating takes place at constant volume. (The Brayton cycle is adiabatic-isobaric.) Equality of entropy input and output can be shown using the same arguments as for the Brayton cycle (horizontal distance of isochors equal at 300 K and at 100 K). The diagram can only give a rough idea of the process in a Stirling engine. Since there is no continuous flow, gas elements in different locations are subject to different cycles. Moreover, a real engine, rather than producing abrupt up–down motions, will run piston and displacer in a sinusoidal mode with a 90° phase shift, and the distinction between compression/expansion and displacement phases is problematic.

Since the pressure drop across the displacer is small (just the flow resistance of the gas), the displacer motion requires little energy, and the leakage between compression and expansion space bypassing the regenerator is kept small. The T–s diagram shows that the pressure amplitude everywhere in the entire gas space (on either side of the displacer) is remarkably high — in spite of the small volume change produced by the piston (a factor of two is assumed in the diagram). The reason for this large amplitude is the isochoric pressure change when the gas passes the regenerator in alternating directions. In the compression phase the piston acts under higher pressure than in the expansion phase. The work required for compression is therefore greater than the work recovered during expansion. The ratio is the ratio of the temperatures, because they determine the pressures. Our arguments used to find the Carnot efficiency of the Brayton cycle (we noticed the difference of volumetric throughput of compressor and expander) can be repeated, *mutatis mutandis*, for the Stirling cycle: we have to replace volume by pressure. Again we find the Carnot efficiency corresponding to the parallelogram-shaped cycle in the log T–s diagram.

Piston and displacer are connected to the same crank drive, which not only supplies the work for compression, but also recovers the work from expansion. This feature accounts for the remarkable efficiency of the process, in particular when used for cooling at relatively high temperature (for example, liquid nitrogen temperature).

CERN uses two-stage Stirling machines for the cooling of liquid-hydrogen targets[14].

5.5 Combination of Brayton and Linde Cycle: The Claude Cycle

Most large, modern cryoplants in particle physics laboratories can be regarded as a combination of a Linde cycle, used for liquefaction, and several 'degenerated' Brayton cycles used for precooling the Linde cycle. Such a compound cycle for air liquefaction was first developed by Georges Claude[15] in 1902; the company *L'Air Liquide* was created on the basis of his patents. The principle of the Claude cycle is shown in Fig. 12. At left, the Brayton and the Linde cycles are shown separately; at right is shown how Claude, by matching the compressor pressures and interconnecting the gas cycles, arrived at an economic integration of the two. Note that the heat exchanger parallel to the expander has unbalanced mass flow: the high-pressure mass flow warm→cold is smaller than the low-pressure mass flow cold→warm, since part of the high-pressure stream is deviated to the expander. This fact, which has important consequences for the heat exchanger operation, will be further analysed in Section 6.3.

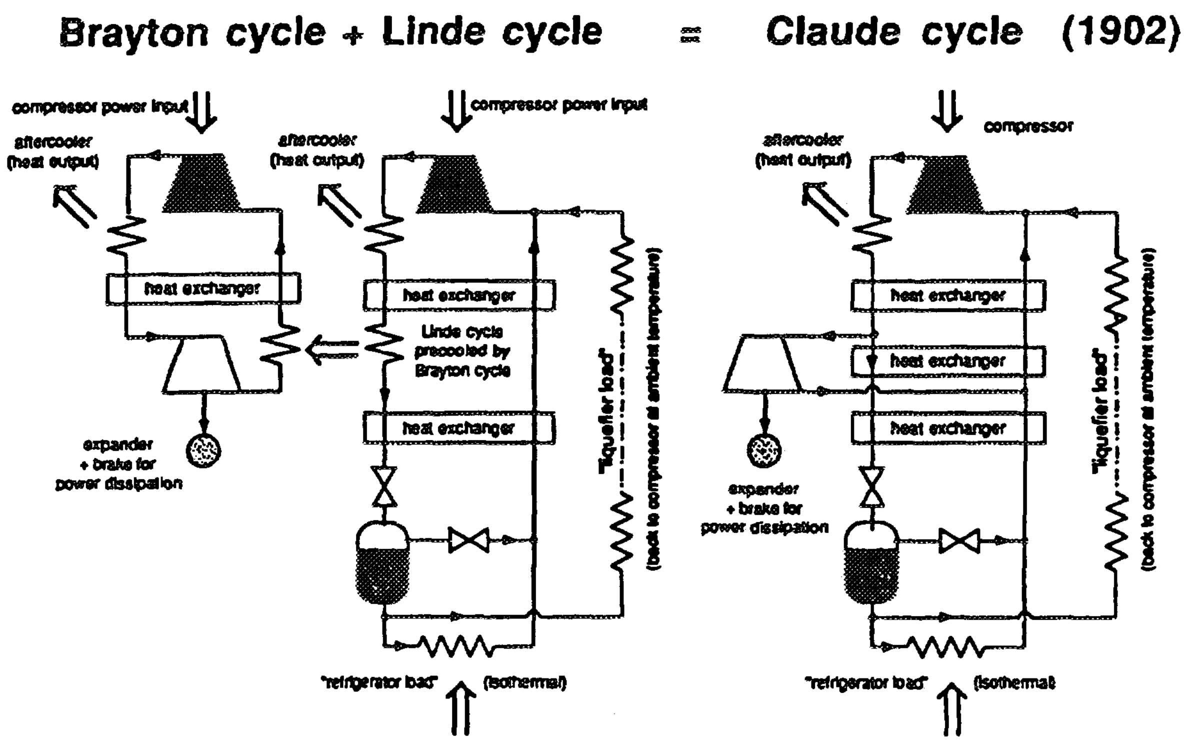


Fig. 12 The Claude cycle as a combination of Brayton cycle and Linde cycle

---

14 Related cycles using regenerators rather than counterflow heat exchangers, but operating between two fixed pressures provided by an external compressor rather than at variable pressure under isochoric conditions are the Solvay cycle and the Gifford–McMahon cycle. Several multistage Gifford–McMahon cryocoolers are used at CERN for cooling at $T \geq 20$ K.

15 Claude had begun his career with the development, in 1897 of the method, still used today, to dissolve acetylene in acetone for storage and transport in order to avoid the risk of autoignition and explosion of this metastable compound. Among many other inventions in various fields, the use of gas discharge tubes as light sources (1910) is particularly remarkable.

Claude used a reciprocating piston expander. Modern cryoplants of the size of interest in particle physics laboratories (cooling capacity above 100 W/4.5 K) generally use turbines[16], mainly because the maintenance requirements of turbines are modest and because turbines of remarkable efficiency have been developed. Piston expanders have their advantages for small plants: they can handle high pressure differences, and their speed can be adapted to wide ranges of gas flow, allowing the plant output to be matched to load requirements without loss of efficiency. Turbines lose a good deal of their efficiency if operated at non-optimal speed, and efficient part-load operation can only be obtained at the price of a considerable sophistication of plant design.

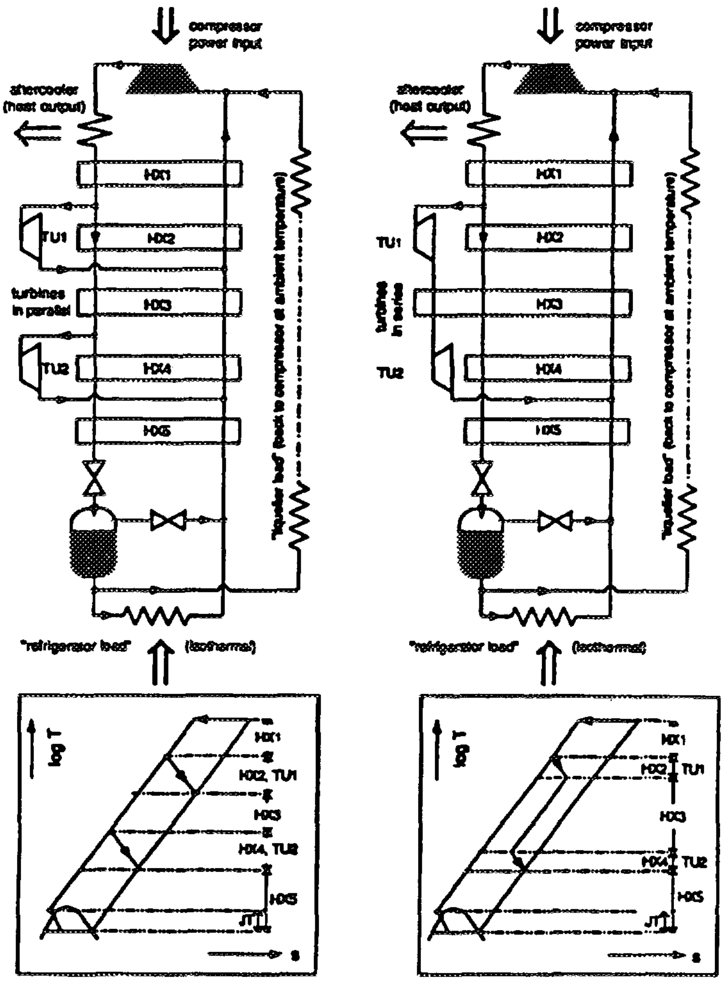


Fig. 13 Two versions of a 2-stage Claude cycle: At left: Brayton stages in parallel. At right: Brayton stages in series

It is possible (and in general necessary for plants designed for cooling below liquid-nitrogen temperatures) to have turbines at several temperatures. Figure 13 shows two two-stage cycles, the turbines being run either in parallel or in series. The T–s diagrams are given at the bottom of the page. Operating turbines in series has the advantage of resulting in smaller pressure and enthalpy differences for handling by each turbine, and, therefore, in smaller gas

---

16 Turbines were proposed for cryogenic use as early as 1898 by Lord Rayleigh, but tremendous technical problems, to the solution of which Pjotr Kapitza (1939) made decisive contributions, delayed their practical introduction. Turbines for hydrogen and helium became available around 1960; the refrigerator of the CERN 2 m bubble chamber was one of the first hydrogen cryoplants using turbines.

velocities. There are other reasons for preferring the serial scheme in many cases. On the other hand, the three-passage heat exchanger involves some complications. The relation of the parallel scheme to the original Claude scheme is obvious, whilst the serial scheme may require some contemplation. The essential point is that in both cases the turbines, bypassing the high-pressure passage of HX2 and HX4, interact with the heat exchange process by reducing the high-pressure flow to be cooled with respect to the cooling low-pressure flow.

I would define the characteristic features of a generalised Claude cycle as follows:

In a string of counterflow heat exchangers, part of the gas is diverted from the high-pressure stream to one or several energy-extracting expanders and then reinjected into the cycle at a lower temperature. As a consequence, the mass flow balance in the counterflow heat exchanger partly bypassed on its high-pressure side by the expander is shifted in favour of the cooling (low-pressure) stream. This results in a reduction of the cold-end temperature difference between the counterflow paths and therefore of the entropy production of the plant.

More about this will be said in Section 6.3.

5.6 Carnot's Cycle

Let us finally discuss the ideal cycle studied by Carnot in 1824 to determine general efficiency limits for thermal engines. Figure 14 illustrates a Carnot cycle in both heat engine and heat pump mode (cycle progressing clockwise and anticlockwise, respectively). Although Carnot engines have never been built for technical purposes, it is useful to discuss even this ideal process with realistic gas data; we again use those of Fig. 3. We assume ambient temperature at the beginning of the cycle and a starting pressure high enough to avoid subatmospheric pressures in all phases.

As Fig. 14 shows, both cycles use the isentropic change of state for switching between heat-sink and heat-source temperature. The required energy for switching from low to high temperature (corresponding to the distance between the horizontal isenthalps in a T–s diagram) is quantitatively recovered when switching back to low temperature in a later phase.

During the isothermal phases, the internal energy of the gas, proportional to its temperature, is constant. The energy (work) put into the system during compression at ambient temperature is therefore exactly compensated by the energy (heat) discharged to the environment. Similarly, the entire energy (work) extracted from the system during expansion at heat-source temperature is exactly compensated by the energy (heat) the system receives from the heat source. Input and output entropies are equal, provided no temperature differences occur during heat transfer load→cycle and cycle→environment.

That the entropy intake at heat source temperature is equal to the entropy output at ambient temperature is as obvious in this diagram as in the diagrams shown previously for the ideal Brayton and Stirling cycles (the parallelogram has now become a rectangle); it is not a specific feature of the Carnot cycle.

The simplicity and transparency of Carnot's fundamental *gedankenexperiment* make it an ideal tool for theoretical considerations. In particular, the assumption of heat input and output at two well-defined temperatures simplifies arguments.

For practical purposes, however, the Carnot cycle has severe drawbacks:

- The condition of isothermality requires careful control of the compression and expansion process in the isothermal phases of the cycle in order to keep temperature gradients negligible and avoid entropy production. Such control is all the more difficult to ensure as the gas pressure varies strongly during the isothermal phases.
- The beauty of avoiding heat exchangers (which are a horror for the theorist because of the complex phenomena they exploit) is stained by the fact that most so-called isentropic

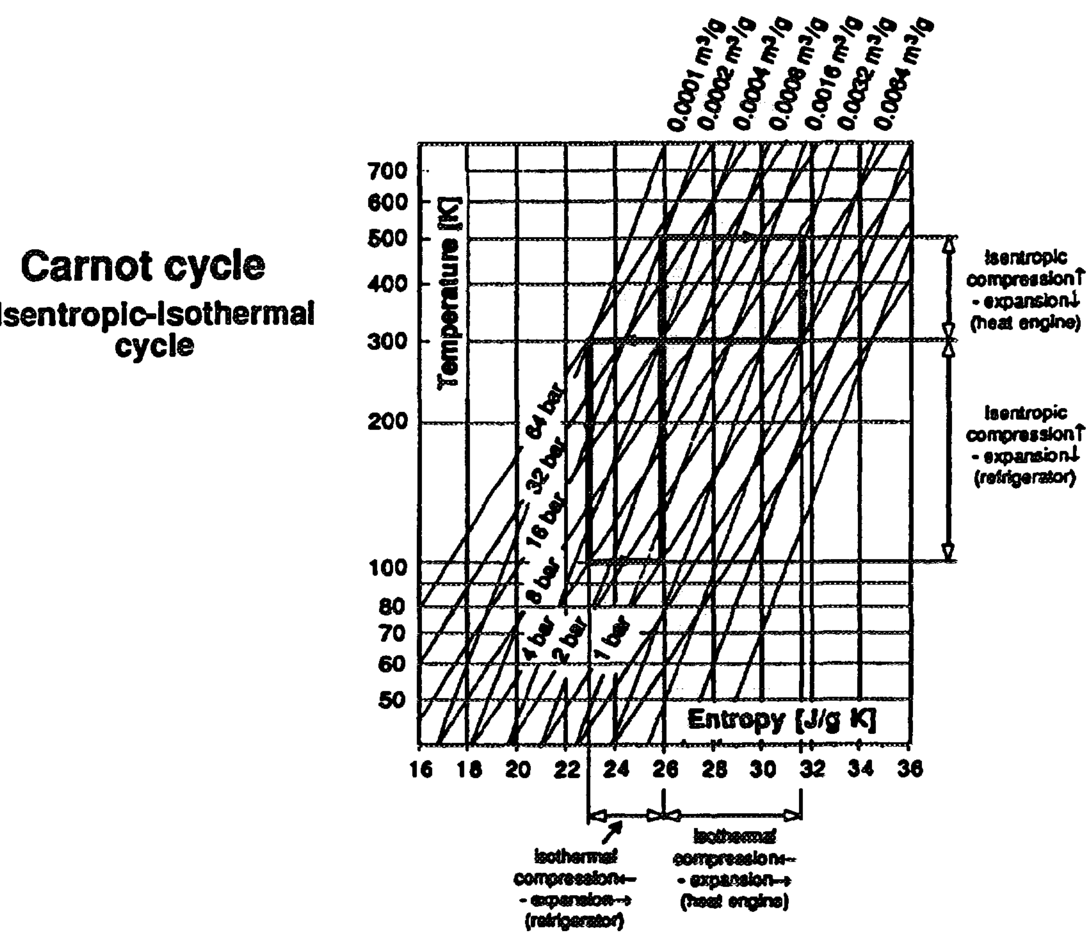


Fig. 14 Carnot cycle operating as heat engine between 500 and 300 K and as refrigerator between 100 K and 300 K

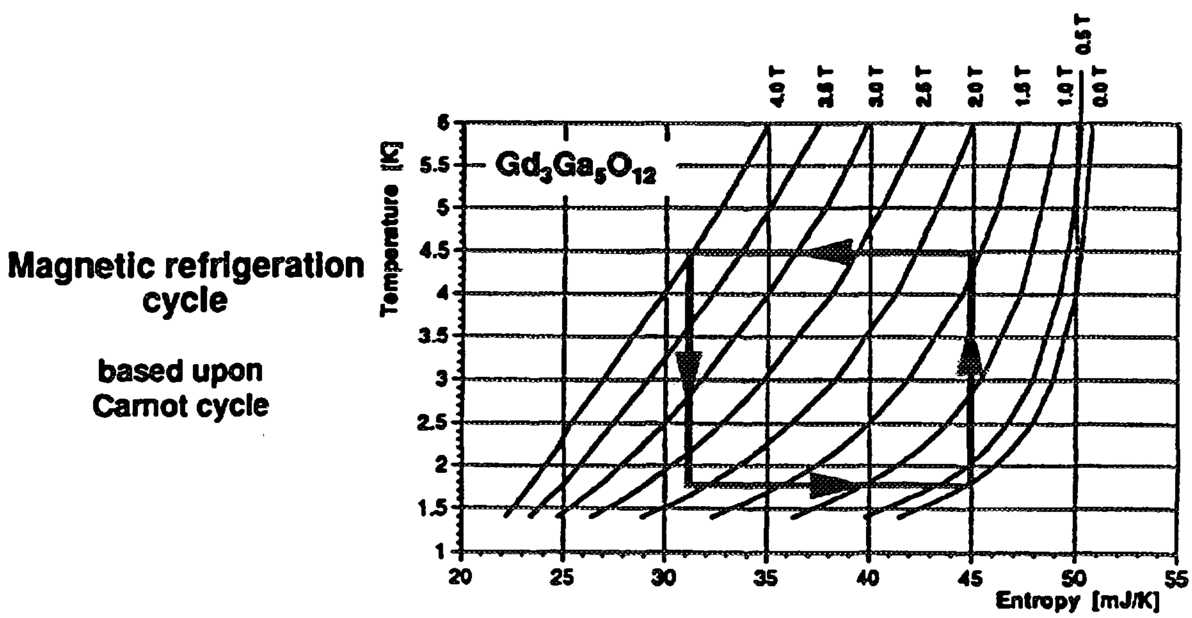


Fig. 15 Application of Carnot cycle for cooling by adiabatic demagnetisation of $Gd_3Ga_5O_{12}$ (GGG = Gadolinium Gallium Garnet)

processes involve non-negligible residual entropy production. Such entropy production would occur in the compressor and in the expander.

- Very high pressure ratios[17] would be required to span temperature differences of practical interest.
- After all, a real Carnot cycle would not even be free from internal heat exchange, since not only the gas, but also its massive containment must follow the temperature cycle, and this again leads to entropy production.

In total, it turns out that adiabatic switching between temperature levels would produce more entropy than a well-designed counterflow heat exchanger.

In spite of all this, the Carnot cycle has found practical application in refrigerators based upon adiabatic demagnetisation, which hold some promise for the pumping of heat at low temperatures and low temperature ratios, for example from 1.8 K to 4.5 K. Magnetic refrigeration cycles use the fact that in a certain temperature range[18] the entropy of suitable paramagnetic crystals depends on external magnetic fields, similar to the entropy dependence on compression of a gas. A T–s diagram for the crystal gadolinium gallium garnet ($Gd_3\,Ga_5\,O_{12}$, 'GGG'), with indication of a possible Carnot cycle, is shown in Fig. 15. The desired Carnot cycle would have the phases:

(α) Magnetisation of the crystal in thermal contact with liquid helium at 4.5 K, accompanied by heat transfer crystal→helium under isothermal conditions (4.5 K),

(β) Thermal isolation of the crystal; demagnetisation, accompanied by a temperature drop in the crystal from 4.5 K to 1.8 K,

(γ) Establishment of thermal contact between the crystal and the load; continued demagnetisation accompanied by heat transfer load→crystal under isothermal conditions (1.8 K),

(δ) Thermal isolation of the crystal; magnetisation accompanied by a temperature increase in the crystal from 1.8 K to 4.5 K.

The GGG diagram shows that the theoretical entropy throughput would be about 14 mJ/g K per magnetic cycle 0 T→4 T→0 T. In reality, much less can be expected, since the ideal Carnot cycle is implemented in very rough approximation only. In particular, phases (β) and (δ) are not clearly separated from (α) and (γ), which results in entropy production due to undesired internal heat transfer.

5.7 Summary of cryogenic cycles. Nomenclature

The *cryogenic* cycles discussed (disregarding those cycles working close to ambient temperature) have in common:

- a compression phase,
- an expansion phase,
- between the two, a temperature-staging device (in most cases a counterflow heat exchanger) ensuring that compression and expansion take place at different temperatures levels,
- heat exchangers for heat (entropy) transfer from the load to the cycle and from the cycle to the environment.

---

17 Or the availability of highly efficient *cold compressors* (suitable for gas compression at low temperature) for multistage configuration of Carnot cycles...

18 Suitable for magnetic cooling are temperatures close to the characteristic temperature for spontaneous ordering of magnetic ion dipoles. The phenomena are somewhat comparable to the behaviour of a real gas near condensation temperature. The interaction of the ion dipoles with the crystal lattice must be small.

The purpose of *compression* is to bring the refrigerant gas into a low-entropy state, preparing it for taking over entropy from the load. Compression is combined with entropy discharge to the environment via heat exchangers that are more or less closely combined with the compressor. To promote heat (entropy) transfer, a certain overtemperature with respect to ambient temperature is necessary.

The purpose of *expansion* is to bring the low-entropy refrigerant to a temperature somewhat lower than the load temperature, thus promoting heat (entropy) transfer from the load to the refrigerant. We distinguish between isenthalpic expansion in throttle valves, where cooling results from work against Van der Waal forces, and adiabatic (ideally: isentropic) expansion in a piston expander or turbine, where cooling results from energy (work) transfer to the outside. All these devices produce entropy, which can be directly read from the T–s diagram (expansion line dropping from left to right, in the sense of increasing entropy). Gas passing a Joule–Thomson throttling valve usually (but not always) enters the liquid domain under the saturation curve of the T–s diagram. Gas passing a turbine or a piston expander may also do so; a special design is then required, and we speak of *wet expanders*.

The purpose of the *temperature staging device* is to prepare the refrigerant for compression and expansion at optimum conditions. If a counterflow heat exchanger is used, this is done by internal heat and entropy transfer between the refrigerant in its low- and high-pressure (high and low entropy) states. In the case of the Carnot cycle, it is done by isentropic compression/expansion. In simple, two-path counterflow heat exchangers with equal mass flow in both channels, the entropy production appears as an increased horizontal distance between the points representing the terminals at the warm end with respect to the cold end. For more complex situations (multipath heat exchangers, different mass flows in channels etc.), this simple graphic illustration is, however, not possible, and (in all cases, simple and complex) entropy production is better evaluated numerically from entropy tables.

Usually, the heat exchanger connecting the cycle to the load is fed by the refrigerant in saturated liquid form[19], and heat transfer from the load leads to progressive evaporation of this liquid. This process is isothermal, characterised by a horizontal line in the T–s diagram, and efficient heat transfer to an isothermal load can be ensured by relatively simple means.

The heat exchanger connecting the cycle to the environment is usually fed by hot gas coming from the compressor, because the efficient isothermal compressor, sufficiently cooled during the entire compression process, has not yet been invented. Heat transfer to the environment thus involves high temperature differences and leads to substantial entropy production (usually the dominant component of the plant's entropy budget).

The entropy production associated with heat (entropy) transfer from the load to the cycle and from the cycle to the environment does not appear in the usual simple T–s diagram. Representation would require separate diagrams (sometimes given as inserts) for load and cooling water conditions. However, since the transferred heat as well as the temperatures of the load and the cooling water are usually known, the entropies taken from the load and received by the cooling water are readily calculated from transferred enthalpies, and compared to the corresponding values received (by the load heat exchanger) and discharged (by the compressor heat exchangers).

---

19 Refrigerators are therefore often incorrectly called liquefiers. In principle, the term *liquefier* is reserved for the operation mode of a cryoplant in which the liquid is extracted from the plant and supplied to an external user who will return the evaporated gas at ambient temperature. Contrastingly, in a *refrigerator* the refrigerant is returned to the plant at or close to saturated vapour conditions; a refrigerator is designed for essentially isothermal cooling. All intermediate cases are possible with gas recovery at a temperature well above saturation, but well below ambient; such plants are loosely called cryoplants or cryogenerators.

Many closely related cycles figure under different names. There are national preferences. Sometimes, due to technical shortcomings, a given cycle can be considered as a crude approximation of several ideal cycles. In general, it is better to give a brief description of the cycle than to rely upon the clarity of a name.

To the best of my knowledge:

- the Bell–Coleman cycle is identical to the Brayton cycle, which is sometimes plainly called the *gas turbine cycle*,
- the *ideal* gas turbine cycle is sometimes called the Joule cycle,
- heat engine cycles corresponding to the reverse of the evaporative refrigeration cycle described in Section 5.1, are called (Clausius–)Rankine cycle and Hirn cycle, the evaporative refrigeration cycle itself is sometimes called reverse Hirn cycle,
- the Ericson cycle, which is similar to the Brayton cycle, but differs from it by aiming at isothermal rather than isentropic compression/expansion, is sometimes also called the Ackeret–Keller cycle,
- the Linde cycle is sometimes referred to as Joule–Thomson cycle or Hampson cycle (George Hampson patented, a few days before Linde, a cycle with counterflow heat exchangers, but which was not specifically designed to work on the basis of the Joule–Thomson effect),
- the Stirling cycle is sometimes referred to as the Philips cycle.
- the Gifford–McMahon cycle and the Solvay cycle are similar to the Stirling cycle in using regenerators and displacers, but differ from it in using external compressors and operating between two constant pressure levels (isothermal-isobaric cycles comparable to the Ericson cycle).

As for the name 'Claude cycle', I use it in a very broad sense for the combination of one or several expander cycles with a liquefaction cycle making use of the Joule–Thomson effect, such as the Linde cycle and its variations. A definition is given at the end of Section 5.5.

6 ENTROPY PRODUCTION IN CRYOPLANTS

There are countless sources of entropy. We can only discuss the most important ones: entropy production associated with compressors, turbines and heat exchangers and entropy production due to the mixing of cold and warm fluids and to unsteady operation conditions.

6.1 Entropy production associated with compression

The largest single source of spurious entropy in cryoplants is usually associated with the compressor. Three types of compressors are currently used for helium service:

- Reciprocating (piston) compressors,
- Screw compressors, and
- (at very low temperatures only[20]) Turbocompressors.

A *reciprocating compressor* performs compression in a close-to-isentropic process. For helium with $C_p/R = 5/2$, a pressure ratio of 3 according to equation (8 c) leads to an isentropic temperature rise from 300 K to 465 K ≈ 190 °C — close to the tolerable limits of compressor

---

20 The requirement of circumferential velocities comparable to sound velocity is not favourable for turbocompressors for compressing light gases such as helium at ambient temperature: achievable rotational speeds are too low for good efficiency.

components. For this reason, compression to higher pressures is undertaken in a multi-stage process — for example, from 1 bar to 20 bar in stages 1 bar → 3 bar → 8 bar → 20 bar, with recooling to ambient temperature between stages ('intercooling')[21]. Multistage compression may greatly reduce entropy production. Whilst entropy production in a close-to-isentropic process is *a priori* low, the energy invested in the heating of the gas must somehow be discharged, either to a work-producing secondary cycle such as a gas turbine — a theoretical solution generally ruled out for many practical reasons — or by transfer to the environment (usually a water circuit) in the form of heat. Generalising equation (8c) for compression from the initial pressure p_0 to the final pressure p_n in n stages of equal compression ratio $(p_n/p_0)^{1/n}$, we find that the required heat transfer is

$$\Delta h = n C_p * \Delta T(\text{stage}) = n C_p * T_0 * \left[(p_n/p_0)^{R/nC_p} - 1\right]$$

and the corresponding entropy transfer $\Delta S_n = \Delta h/T_0$. Compared to the entropy transfer necessary during isothermal compression,

$$\Delta s_T = R \ln (p_n/p_0),$$

following from (9c), we have the relative increase of entropy production

$$\frac{\Delta s_n}{\Delta s_T} = \frac{nC_p}{R} * \frac{(p_n/p_0)^{R/nC_p} - 1}{\ln(p_n/p_0)} .$$

For $p_n = 20$ bar, $p_0 = 1$ bar, $C_p/R = 2.5$ we find (for example)

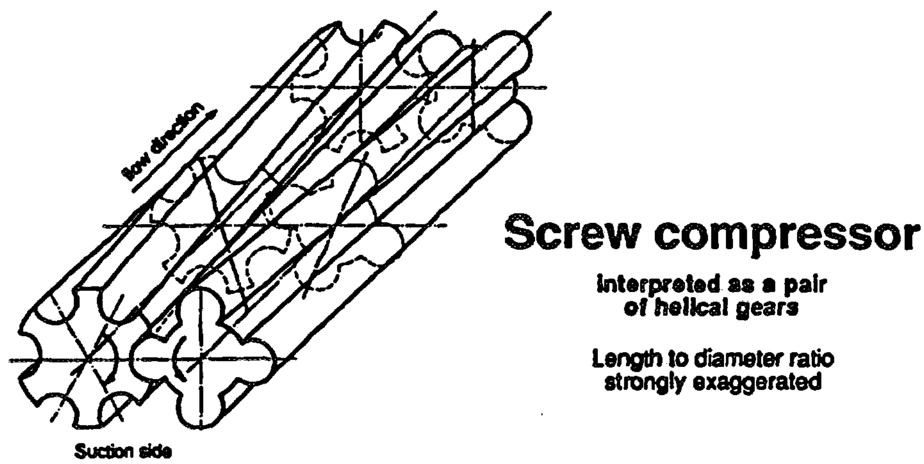
Number of stages n	1	2	3	4	-	∞
$\Delta s_n/\Delta s_T$	1.93	1.37	1.23	1.17	-	1.00

Multistage compression is thus obviously advisable for energetic reasons, but high investment costs and maintenance requirements usually limit the number of stages to two or three.

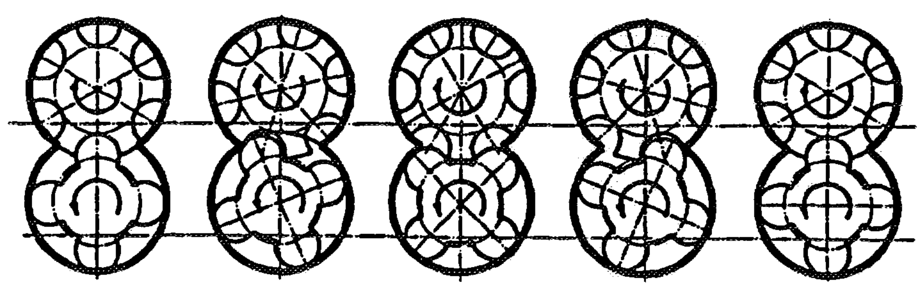
Screw compressors involve lower compression temperatures. The advent of the screw compressor in cryogenics is of recent date, since milling machines for economic rotor production became available only in the 1960s and 1970s. It may be useful to recall the working principles of a screw compressor. A screw compressor (Fig. 16) consists of a pair of helical gears ('screws', 'rotors') of a tooth profile specially designed to compress the fluid trapped between mating threads and to displace it in axial direction. The rotors are installed in a close-fitting casing with apertures for suction and discharge at opposite ends. The upper part of the figure gives an intuitive view of the progressive trapping of fluid between mating teeth. The length of the rotors is strongly exaggerated; in reality rotors are less than two diameters long and the pitch is comparable to the length. A series of cuts through the engaging pair is given in the lower part of the figure, which can be read either as a movie sequence of gear positions at a given location or as a picture of simultaneous gear positions at various axial locations. Helium is usually compressed as a mixture with (uncompressible) oil; quantitative separation after compression is required in a carefully designed coalescer/purifier. The oil not only reduces the otherwise

21 C_p/R and C_v/R values are particularly low for monatomic gases such as helium (3/2 and 5/2, respectively), and therefore the heating effect is particularly strong. For polyatomic gases, C_p/R and C_v/R values are higher — for example, 5/2 and 7/2 for diatomic ideal gases. Hydrogen and air compressors therefore require a smaller number of stages for the same final pressure, and it is for this reason usually not possible to use a hydrogen or air compressor for helium without major modifications.

prohibitively high leakage across rotor gaps, but also improves cooling; in fact it is the high heat capacity of the oil that limits the helium temperature to less than 100 °C even at compression ratios much higher than those of reciprocating compressors. The reduced temperature and of course the reduced leakage (which, however, remains considerable) are features improving efficiency, but other mechanisms — for example, internal friction in the oil-gas mixture — result in new losses. For the time being, the efficiency of screw compressors is no better than that of reciprocating compressors. Their main advantages are low maintenance requirements, vibration-free operation and therefore modest requirements on foundations, and relatively low cost.



Below: Cut across screw compressor seen from suction side
⇓ can be read either as a cinematic sequence of gear positions at a given location
or as a picture of simultaneous gear positions at various axial locations of the cuts
see table below



Location on gear in units of pitch of male screw	⇑ Position at time ⇓	⇑ Position at time ⇓	⇑ Position at time ⇓	⇑ Position at time ⇓	⇑ Position at time ⇓
0	0	1/16 rev	2/16 rev	3/16 rev	0, 4/16 rev...
1/16	1/16	2/16 rev, 6/16 rev...	3/16 rev, 7/16 rev...	0, 4/16 rev...	1/16 rev, 5/16 rev...
2/16	2/16 rev	3/16 rev, 1/16 rev...	0, 4/16 rev...	1/16 rev, 5/16 rev...	2/16 rev, 6/16 rev...
3/16	3/16 rev	0, 4/16 rev...	1/16 rev, 5/16 rev...	2/16 rev, 6/16 rev...	3/16 rev, 7/16 rev...
4/16	0, 4/16 rev...	1/16 rev, 5/16 rev...	2/16 rev, 6/16 rev...	3/16 rev, 7/16 rev...	0, 4/16 rev...
-	-	-	-	-	-

Fig. 16 Working principle of a screw compressor

In recent years cryoplants for cooling temperatures below 4.2 K have increasingly made use of *cold compressors* designed to bring helium evaporated at subatmospheric pressure to a pressure more favourable for efficient operation of heat exchangers (heat exchangers for very-low-pressure gas are bulky, expensive and may be major producers of entropy). Considerable development work is going on. Turbocompressors are generally used at very low temperatures. They benefit from the fact that the required rotor speeds are moderate, as molecular speed and sound velocity decrease with temperature. The compression process is adiabatic, leading to a considerable temperature increase. Required compression ratios may be very high (10 mbar : 1 bar = 1 : 100), and many stages may be necessary. When cold compressors are connected in series, pressure and temperature variations at the outlet of the first compressor complicate operation of the second, since efficient operation of turbomachines depends crucially upon inlet conditions kept to the design point within narrow tolerances. A method for fixing inlet conditions is the recooling of 'compressed' low-pressure helium to a well-defined temperature in counterflow heat exchange with gas from other branches of the process. The layout of the recooling loop can strongly influence its entropy production.

Generally, it must be stressed that a compressor/aftercooler assembly is to be judged by its *entropy* production and *entropy* rejection rather than by its heat production and heat rejection, since it is the final *entropy* transfer to the environment at ambient temperature that largely determines the power requirements of the plant. The entropy eventually discharged can be traced back to the many individual entropy sources, of which it constitutes the arithmetic sum, whilst a similar procedure for heat is not possible, since most of the heat eventually discharged has its origin in non-thermal power provided by the compressor. Praising low-heat production (and corresponding low-power consumption) as a specific advantage of cold compressors is an incorrect appreciation of the situation.

6.2 Entropy production associated with expansion

Expansion takes place not only in sophisticated work-extracting expansion engines such as piston expanders and turbines, but also in simple devices such as valves, and generally in any part of the cycle where flowing compressible fluids undergo a pressure drop. In the first case, the goal is *isentropic* expansion, and entropy production means deviation from that goal. In the second case, we limit our discussion to expansion without external heat transfer; the change of state is then *isenthalpic*.

Isentropic expansion in piston expanders or turbines occurs along a vertical line in the T-s diagram (see Fig. 4). Adiabatic, but not exactly isentropic, expansion occurs along an inclined line more or less close to vertical. Suppose the inlet state is given by pressure p and temperature T; the associated enthalpy and entropy values h and s can be found in tables. The ideal outlet state is determined by the given outlet pressure p' and the isentropy condition that the outlet entropy is $s'_{ideal} = s$. T'_{ideal} and h'_{ideal} again are easily determined. To determine the real outlet conditions, one parameter, for example, temperature T'_{real}, must be measured. h'_{real} and s'_{real} are again easily determined. If the mass flow is $\dot{m}$, the entropy production is $\dot{m}\,(s'_{real} - s)$. The work extracted is equal to the enthalpy decrease $\dot{m}\,(h - h'_{real})$. The value $\eta = (h - h'_{real})/(h - h'_{ideal})$ is generally referred to as the efficiency of the expansion.

Entropy production in turbines (we skip reciprocating expanders in view of their limited practical interest) may be due to turbulent dissipation of kinetic energy in the turbine wheels, but also due to heat inleak by conduction or radiation, or to inleak of warm helium from the gas bearings carrying the rotor of modern cryogenic turbines.

Entropy production during isenthalpic expansion is as easily determined as the entropy difference between outlet and inlet state: $\Delta s = s(h,p) - s(h,p')$.

For isenthalpic expansion $p \rightarrow p'$ of ideal gases, for which isenthalps and isotherms coincide, the entropy production follows, with $T' = T$ and $\Delta p = p - p'$, from (9c),

$$\Delta s_h = \Delta s_T = \frac{R}{M}\ln\frac{p}{p'} = \frac{R}{M}\ln\left(1+\frac{p-p'}{p'}\right) = \frac{R}{M}\ln\left(1+\frac{\Delta p}{p'}\right) \approx \frac{R}{M}\left[\frac{\Delta p}{p'}+\frac{1}{2}\left(\frac{\Delta p}{p'}\right)^2+\ldots\right].$$

At high pressures, the entropy production due to pressure drop in normal flow channels may be negligible because $\Delta p/p'$ is small. At low pressure, however, and in particular at subatmospheric pressure, pressure drop in pipework and heat exchangers may be a major source of entropy.

6.3 Entropy production associated with counterflow heat exchange

This section is devoted to conterflow heat exchange. Simple heat exchange has already been treated in Sections 5.1 and 6.1 as a major problem in the context of compressor heat rejection.

Balanced and unbalanced heat exchangers

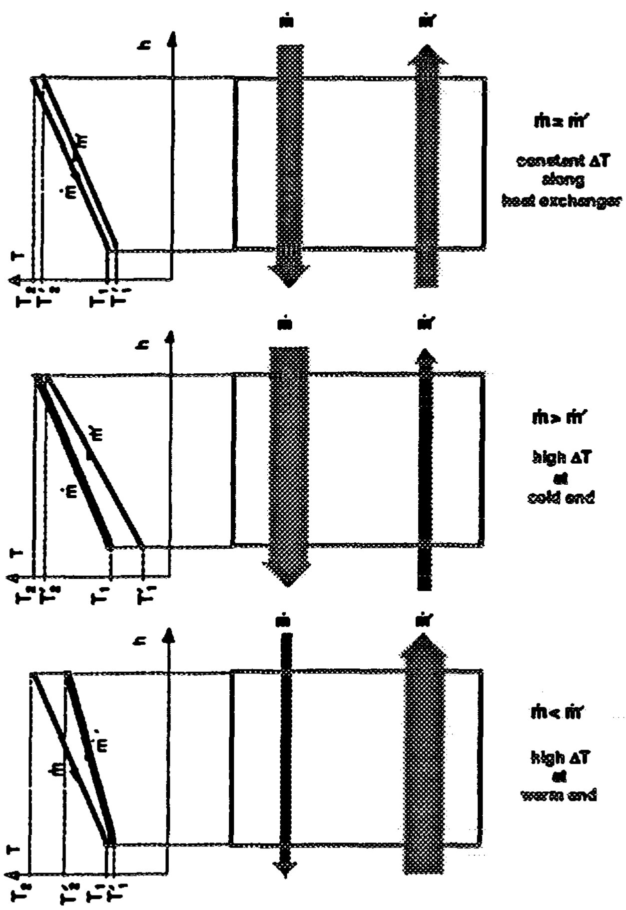


Fig. 17 Temperature distribution on balanced and unbalanced counterflow heat exchangers

We consider (top picture of Fig. 17) a two-passage counterflow heat exchanger, in which heat is transferred from a stream $\dot{m}$ (cool*ed* stream, flow direction $T_2 \rightarrow T_1$, i.e. warm→cold, $T_2 > T_1$) to a stream $\dot{m}'$(cool*ing* stream, flow direction $T_1' \rightarrow T_2'$, i.e. cold→warm). We assume that there is no other heat exchange with the exterior. As a result of the internal heat exchange, the temperature of stream $\dot{m}'$ increases in flow direction, while $\dot{m}$ decreases in flow direction. The First Law requires that, at any longitudinal element of the heat exchanger, the heat (enthalpy

$d\dot{H}$) *given* by stream $\dot{m}$ equals the enthalpy $d\dot{H}'$ *received* by stream $\dot{m}'$; the difference of enthalpies between the two streams is therefore the same at any two corresponding locations. Applying this conclusion to the four terminals, we find $\dot{H}_2 - \dot{H}_2' = \dot{H}_1 - \dot{H}_1'$, hence $\dot{H}_2 - \dot{H}_1 + \dot{H}_1' - \dot{H}_2' = 0$, or, summing over all terminals, with the convention to count as positive the flows entering the heat exchanger ($\dot{H}_2, \dot{H}_1'$) and as negative the flows leaving the heat exchanger ($\dot{H}_2', \dot{H}_1$):

$$\sum \dot{H} = 0 \,. \tag{11}$$

This equation is valid for any number of passages, i.e. also for multistream heat exchangers.

Let us consider the simplest case, namely that of two equal mass flows having the same constant specific heat: $\dot{m} = \dot{m}'$ and $c_p = c_p'$. We then have because of $\dot{H}_2 - \dot{H}_2' = \dot{m} c_p (T_2 - T_2') = \dot{H}_1 - \dot{H}_1' = \dot{m} c_p (T_1 - T_1')$

$$T_2 - T_2' = T_1 - T_1' \quad \text{if} \quad \dot{m} c_p = \dot{m} c_p = \dot{m}' c_p' \,, \tag{12a}$$

i.e. a constant temperature difference between corresponding locations at any place of the heat exchanger.

With similar reasoning we find that

$$T_2 - T_2' < T_1 - T_1' \quad \text{if} \quad \dot{m} c_p > \dot{m}' c_p' \,, \tag{12b}$$

i.e. the temperature difference increases from warm (T_2) to cold (T_1), and

$$T_2 - T_2' > T_1 - T_1' \quad \text{if} \quad \dot{m} c_p < \dot{m}' c_p' \,, \tag{12c}$$

i.e. the temperature difference decreases from warm to cold.

Summing up, we state that the *temperature difference* between corresponding locations *increases in the flow direction of the dominating stream* (the stream with the larger $\dot{m} c_p$). This is illustrated in Fig. 17.

In cryoplants we frequently encounter case (12b);. $\dot{m}'$ may be lower than $\dot{m}$ because some liquefied or cold gas has been withdrawn at the cold end of the plant to be returned at ambient temperature only (liquefier mode). In addition, the specific heat of real gases is somewhat pressure-dependent and often higher at high pressure than at low pressure. For this reason, temperature differences in heat exchangers in the absence of expansion engines tend to increase from warm to cold. When discussing the Claude cycle we noted the fact that the introduction of the expansion engine in parallel to a heat exchanger creates an imbalance of mass flow in the heat exchanger in favour of the low-pressure stream, as the expander takes gas from the high-pressure stream warm→cold, bypassing the heat exchanger, and eventually restitutes it to the low-pressure stream cold→warm. It thus creates a situation corresponding to (12c), leading to decreasing temperature differences in the flow direction of the high-pressure stream. This temperature decrease can be designed so as to compensate earlier temperature increases and to keep temperature differences under control. An illustration is given in Fig. 18.

The entropy production in a counterflow heat exchanger is

$$\dot{S} = \int_{h_{out}}^{h_{in}} \left(\frac{\dot{m}'}{T'} - \frac{\dot{m}}{T} \right) dh \,; \qquad \dot{m}' \text{ not necessarily equal to } \dot{m} \,; \quad T' < T \,.$$

The usefulness of thermodynamic computer programs for avoiding tedious integration is demonstrated by Table 5 prepared with HEPAK. The table refers to a 'balanced' heat exchanger handling equal mass flow of helium at pressures 20 bar and 1 bar. The table was calculated for a given temperature difference of 10 K at the warm end of the heat exchanger and for finite enthalpy transfer elements $\Delta\dot{h}=100$ J/g (steps 1580, 1480, 1380... J/g in the high-pressure channel, and 1522, 1422, 1322... J/g in the low-pressure channel). It is obvious that the entropy flow cold→warm shown in the last column increases continuously from 25.78 to 31.44 J/g, due to the entropy produced during heat transfer. The total entropy production can be calculated by summation over all terminals, using the convention to count as positive the flows entering the heat exchanger and negative the flows leaving it:

$$\Delta\dot{S}=\sum\dot{S}=\sum\dot{m}\,s>0 \tag{14}$$

(s from tabulated values or computer programs).

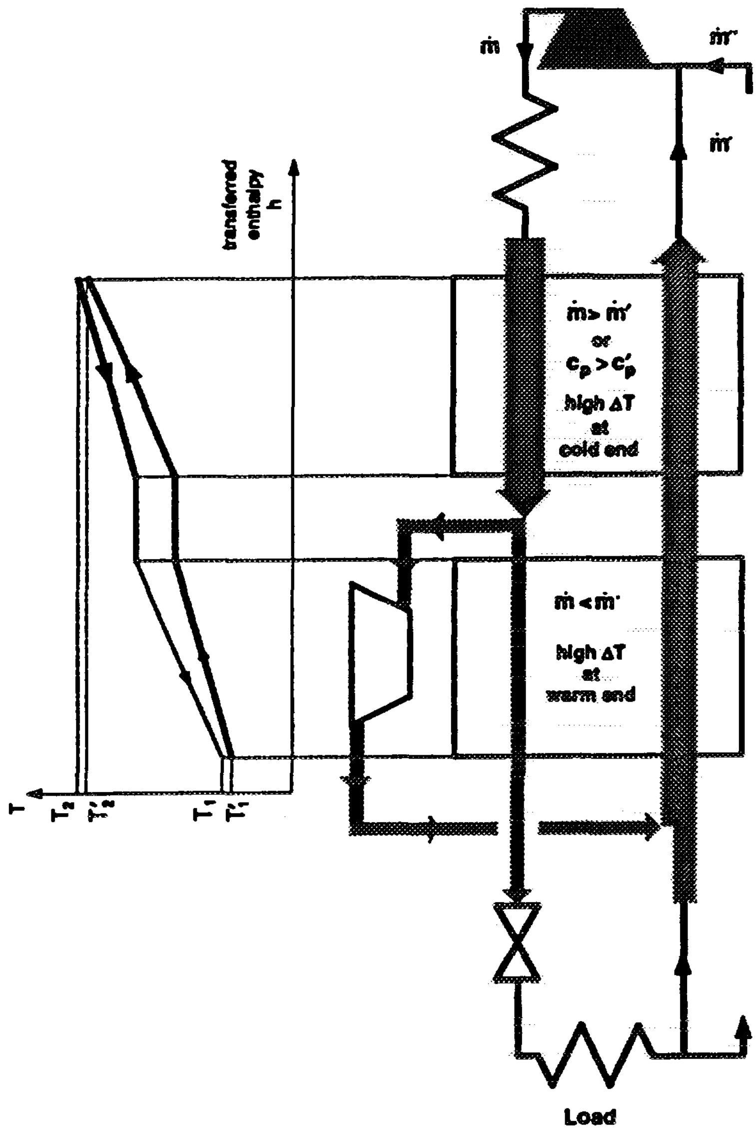


Fig. 18 Temperature distribution on the two heat exchangers of a Claude stage

This formula allows entropy production to be calculated for heat exchangers of any number of passages. It is the most important formula for the evaluation of entropy production in a cryoplant of any complexity.

The reason for keeping temperature differences small by introduction of expansion engines in Claude stages is of course the intention to minimise entropy production. As, according to (13), entropy production depends on

$$\Delta\frac{1}{T}=-\frac{\Delta T}{T^2},$$

it is most important to achieve small values for ΔT at the lowest temperatures of the plant.

Table 5

Entropy production on heat exchanger 300 K... 100K

pressure p	temperature T	enthalpy h	entropy s	pressure p	temperature T	enthalpy h	entropy s	entropy flow upward Δs = s(1 bar)–s(20 bar)
bar	K	J/g	J/g K	bar	K	J/g	J/g K	J/g K
20	300.07	1580	25.39	1	290.08	1522	31.44	6.05
20	280.81	1480	25.05	1	270.82	1422	31.08	6.03
20	261.54	1380	24.68	1	251.57	1322	30.70	6.02
20	242.29	1280	24.28	1	232.31	1222	30.28	6.00
20	223.03	1180	23.85	1	213.05	1122	29.83	5.98
20	203.77	1080	23.38	1	193.80	1022	29.34	5.96
20	184.52	980	22.87	1	174.54	922	28.80	5.93
20	165.28	880	22.30	1	155.29	822	28.19	5.90
20	146.04	780	21.65	1	136.03	722	27.50	5.85
20	126.82	680	20.92	1	116.78	622	26.71	5.79
20	107.62	580	20.06	1	97.52	522	25.78	5.71

We have tacitly assumed in the preceding that, in any cross-section, the heat exchanger is characterised by only two temperatures — one for the cooled stream (T) and one for the cooling stream (T). That this is so is by no means granted by nature but depends on the skill of the designer and the manufacturer of the plant. In particular, flow distribution in a heat exchanger as shown in Fig. 7 can be such that, although the flow, averaged over the entire heat exchanger cross-section, is balanced, certain parts are preferred by the cooled stream and others by the cooling stream, which leads to local unbalance of flow and may result in unexpected temperature differences and major losses of cooling capacity. The reason for this may be an inadequate header design (see Fig. 7), obstructions in passages or, particularly in horizontal heat exchangers at very low temperatures, stratification due to gravity.

6.4 Entropy production due to heat leaks or gas leaks

The consequences of heat inleaks to cold parts for the performance of a cryoplant are often underestimated. Since $\dot{S}=\dot{Q}/T$, the entropy production due to a given heat inleak increases with 1/T. An often encountered case of bad practice is heat introduction by cables of 'small' cross-section handling electrical signals. Such cables can constitute a heat load of quite unexpected importance because of the very high thermal conductivity of good electrical conductors at cryogenic temperatures. Bridging of large temperature differences must therefore be avoided, and incoming heat should be intercepted at the highest possible temperature level, in order to minimise entropy production.

Often overlooked are gas leaks between different temperature levels via leaking valves or faulty components. The requirements on sealing properties (geometric matching) are higher at low than at high temperatures because of the increased gas density, and they are more difficult to meet because of the lack of elasticity of cold sealing materials. Mixing of warm and cold gas is a typical example of an entropy producing process: while enthalpies are additive, resulting in a mixing enthalpy

$$h' = \frac{\dot{m}_1 h_1 + \dot{m}_2 h_2}{\dot{m}_1 + \dot{m}_2},$$

a similar relation does not hold for entropy. The mixing entropy is given by the mixing enthalpy and other boundary conditions (e.g. pressure, if mixing occurred at constant pressure) and can be determined from tables. Always,

$$s' = s(p, h') > \frac{\dot{m}_1 s_1 + \dot{m}_2 s_2}{\dot{m}_1 + \dot{m}_2}.$$

Internal gas leaks, but also other heat leaks, can have further serious consequences by unbalancing critical heat exchangers, shifting working points for delicate components (for example, the Joule–Thomson valve) or creating an unexpected pressure drop at constrictions in the process piping.

6.5 Entropy production due to unsteady operating conditions

Whilst entropy production in compressors, turbines and heat exchangers falls under the responsibility of the cryoplant designer or manufacturer, who may also be responsible for some heat and gas leaks, operating conditions are usually imposed by the operator or user. It is therefore particularly important that the user be aware of the consequences of his requirements for the performance of the plant. This awareness is often missing. In particular, users often require the cryoplant to be able to cope with rapid changes of load conditions ('rapid' in cryogenics means 'within a few minutes...'), not realising that any control action on a thermodynamic system results in temporarily increased temperature differences in heat exchangers and therefore in increased entropy production, thus jeopardising efficiency.

Ensuring steady operation conditions is therefore mandatory for efficient plant operation. It turns out that primitive and apparently brutal methods are often more efficient than sophisticated controls designed by specialists of non-thermal systems. A barbaric method that is often successful consists of adding complementary heat (provided by an electric heater) to the load, to keep constant the sum of the original load and the stabilising complement. The entropy produced by the heater is then smaller than that which would have been produced by an imperfect control system.

In the worst case, unsteady operating conditions include the breakdown of utility supply (electricity or water) or off-design operation of the load (for example, quenches of superconducting magnets). Economic considerations generally exclude a cryoplant design capable of coping with such situations. Asking for increased cooling capacity for all eventualities or for maximal redundancy of components or for ultimate investments in controls is usually not economic, being less efficient than an investment in reliable conventional utilities and restricting operating requirements to those matched to the realistic potential of the cryoplant.

7 HOW TO DEAL WITH COMPLEX CRYOGENIC SYSTEMS

Figure 19 shows the flow scheme and the associated T–s diagram of a large cryoplant at CERN. The plant handles a total entropy load of about 12 kW/4.5 K and uses seven turbines. Two plants of this type were provided by Linde/Sulzer and two similar ones, designed to the same specification, by L'Air Liquide. The specification was drawn up to meet the needs of LEP2, but with the possibility of a later upgrade by about 50% for use with LHC. In view of the upgrade, two low-pressure paths are provided: one, handling the return gas from the liquid-helium vessel and passing by points ...39,...,31; and the other handling the return gas of the upper turbines and passing by points 29,...,21. At present the two paths are at identical temperature and pressure conditions; for the upgrade, however, they can be separated and the turbine path operated at higher pressure and mass flow without affecting the pressure in the liquid-helium vessel. The upgrade will also require the installation of an additional turbine between TU3 and TU5; this turbine will bear the now missing number 4.

We call the cycle a multistage Claude cycle because its central Linde cycle is precooled by seven turbines, bypassing the cooled (high-pressure) stream of heat exchangers HX 2, 4, 6, 8, 10 and 12, and shifting the mass flow balance there in favour of the cooling stream. We introduced a corresponding definition of the generalised Claude cycle in Section 5.5.

The purpose of this section is to show how to use Table 2 and Fig. 2 to analyse a complex installation. The plant shown in Fig. 19 serves as an example; details are neglected. Since entropy considerations play an important role in what follows, numeric values of entropies at component terminals are explicitly given in the T–s diagram. Entropies are given in conventional units [J/g K or W/g K (for entropy flow)] and in units based upon the temperature level of the main heat load of the plant, 4.5 K (a practice introduced with Fig. 2), namely J/4.5 K or W/4.5 K.

Step 1: Definition of duty loads

The *plant specification* defines the following simultaneous loads:

Designation	Quantity	Supply conditions	Recovery conditions
Main load	10 kW	saturated or subcooled liquid at 4.5 K	saturated vapour at 4.5 K
Liquefier load	13 g/s	saturated or subcooled liquid at 4.5 K	gas at 300 K, ~ 1 bar
Heat intercept load	6.7 kW	gas, T < 75 K, any pressure	gas, T < 75 K, any pressure

The total *heat load* is

$$\dot{Q}_{in} = 10\,kW + 13\,g/s * (1'573.5 - 11.6)\,J/g + 6.7\,kW = 37.0\,kW\,,$$

taking the enthalpy values for helium gas at 1 bar and 300K (1'573.5 J/g) and at 4.5 K (11.6 J/g) from HEPAK.

The corresponding *entropy loads* $\dot{S}_{in}$ follow from the data of the T–s diagram:

Designation	Entropy load
Main load	510.5 g/s * (8.250–3.741) J/g K = 2'302 W/K = 10.358 kW/4.5 K, with respect to the specified 10 kW, a slight reserve is included
Liquefier load	13 g/s * (31.477–3.781) J/g K = 360 W/K = 1.620 kW/4.5 K
Heat intercept load	28.1 g/s * (19.568–13.301) J/g K = 176.1 W/K = 0.792 kW/4.5 K
Total $\dot{S}_{in}$ according to flow sheet	(2'302 + 360.5 + 176.1) W/K = 2.84 kW/K ≈ 12.8 W/4.5 K

Parasitic radiative and conductive heat loads are neglected.

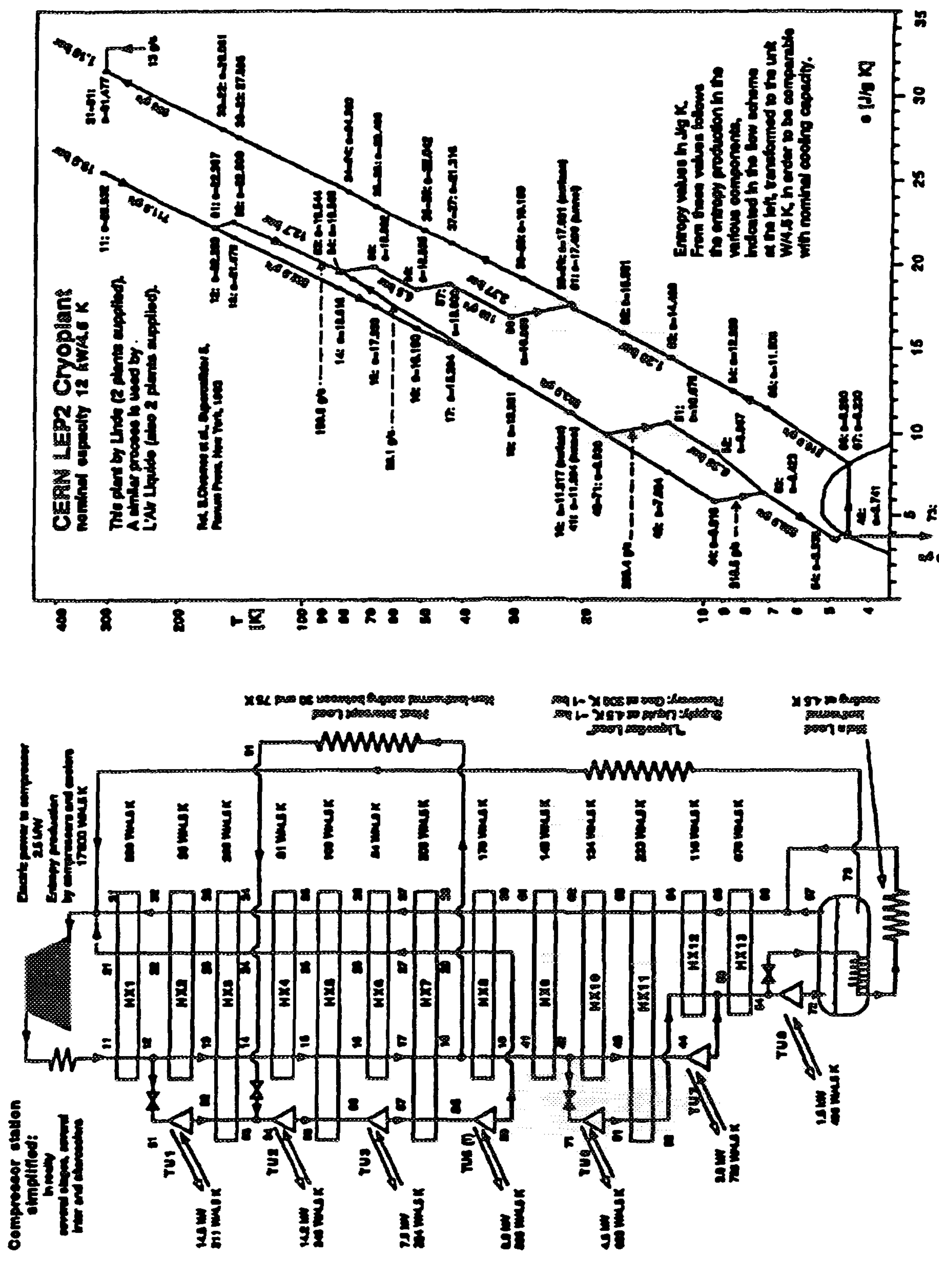


Fig. 19 Flow scheme and temperature-entropy diagram of a large cryoplant at CERN

Step 2: Identification of internal entropy sources

We might individually calculate the entropy load of all components — for example, that of heat exchanger HX11:

$$\dot{S}_{HX11} = \sum_i \dot{m}_i s_i$$
$$= \dot{m}_{51}s_{51} - \dot{m}_{52}s_{52} + \dot{m}_{43}s_{43} - \dot{m}_{44}s_{44} + \dot{m}_{63}s_{63} - \dot{m}_{64}s_{64}$$
$$= 205.4*(10.675 - 8.867) + 318.5*(7.684 - 5.916) + 510.9*(12.539 - 14.469)[g/s*W/gK]$$
$$= -51.6\,W/K = -232\,W/4.5K$$

(negative since coming out of the heat exchanger). Such calculations are extremely useful for the individual evaluation of components. They are easily done on a spreadsheat, the usefulness of which, and of programs like HEPAK, is gratefully acknowledged, remembering the time when such calculations had to be done on slide rules and mechanical calculators with graphical data from diagrams on A0 format. The results for our present example (including *internal* entropy production by turbines, i.e. entropy production at process temperature) are given in the flowsheet of Fig. 19.

Step 3: Identification of enthalpy extraction by expansion devices

The power extracted by turbines is indicated in the flow sheet of Fig. 19; the total is

$$\sum \dot{H}_{turbines} = 14.5 + 14.2 + 7.8 + 8.9 + 4.5 + 3.0 + 1.5 = 54.4\,kW.$$

This energy will be dissipated at 300 K (entropy production 54.4/300 = 0.2 kW/K = 0.8 kW/4.5 K).

Step 4: Output entropy, not including entropy production by compressor coolers

The output entropy can be found by summing up the entropies found in the preceding steps or, more easily, directly from the T–s diagram of Fig. 19:

$$\dot{S}_{out(excl.compressor)} = \dot{m}_{11}*(s_{11} - s_{21}) + 0.2\,kW/K = 711.9*(25.532 - 31.477) - 200[W/K]$$
$$= -4.43\,kW/K = -19.9\,kW/4.5K.$$

Comparing this with the result of Step 1, we notice an entropy increase by 66%.

Step 5: Output heat

The heat output corresponding to $\dot{S}_{out(excl.compressor)}$ is 4.43 kW/K*300 K = 1.33 MW.
The electric power received by the compressor is 2.5 MW.

The difference is due to entropy production in the compressor and the equipment associated with it (in particular heat exchangers). Expressed in units of entropy, it corresponds to (2.5–1.33) MW/300 K = 3.9 kW/K = 17.6 kW/4.5 K.

Conclusion:

The duty entropy load input of the cryoplant is 12 kW/4.5 K. Internal entropy sources from heat exchangers, turbines etc. increase it by a factor of 1.66. Entropy production in the compressor and associated equipment increase it by another factor of 1.88. In other words: the entropy discharged to the environment is about 3.1 times that received by the load. 0.66 times the load entropy is produced by heat exchangers and turbines, 1.47 times the load entropy is produced by the compressor.

Note that this is one of the most efficient cryoplants in the world.

8 WHAT IS ENTROPY?

Entropy is *the* fundamental quantity characterising thermodynamic systems.

Let us recall the main difference between mechanics and thermodynamics:

- In mechanics we deal with strongly constrained systems. The number of components and of degrees of freedom is small. Their parameters, if not already known, can in principle be determined, and their control is the task of the mechanical engineer.
- In thermodynamics we deal with very loosely constrained systems. The number of components and of degrees of freedom is enormous; we would not dream of determining or controlling instantaneous parameters of individual components (which, for simplicity, in the following we will call particles). Nevertheless, the few constraints that are known (for example, given number of particles, given total energy, restriction to a given volume, etc.) allow some important predictions to be made about the behaviour of the system.

Rather than dealing with *the real* state of the system, as we would do in mechanics, in thermodynamics we deal with *all possible* states compatible with the constraints. We call them *microstates*. The most important thing we know about them is that their number exists and that it is finite, although very large. (That the number is finite follows from our limited measurement accuracy, which at any rate does not go beyond Heisenberg's uncertainty principle: it would not be reasonable to count two states as different if we cannot, even in principle, discriminate between them.) We call this number Φ.

To calculate the value of Φ is *not* our objective — it can be done for some ideal systems and is a task for theoreticians, far beyond the scope of this seminar. But the simple fact that this number exists has important consequences in combination with the — plausible and experimentally confirmed — Fundamental Assumption of thermodynamics:

> The Fundamental Assumption of thermodynamics is that the probability of being *real* (and not only possible) is the same for all microstates.

Based upon this assumption, we may compare two macroscopic states (we will call them *macrostates*), each loosely constrained and comprising a large number of microstates — say, Φ_1 and Φ_2. This does not mean very much as long as the two macrostates have nothing to do with each other.

Suppose, however, that they underlie the same constraints. It then follows from the Fundamental Assumption that the one of the two macrostates containing more microstates than the other has a greater probability of being real, the probability being proportional to Φ. We may go further and decompose the state of a real system into a set of macrostates differing from each other by one property that interests us — for example, the distribution of energy over the degrees of freedom of the system — described by a suitable function. We may scan the real system according to an appropriate scheme. According to the Fundamental Assumption, the macrostate containing the largest number of microstates has the highest probability of being real. We conclude that the real distribution of energy is the one that corresponds to this macrostate.

Let us discuss the behaviour of two systems that, to begin with, underlie two different constraints — for example, total energy, volume and particle number N_1, V_1, U_1 and N_2, V_2, U_2; the respective microstate numbers are Φ_1 and Φ_2.

As long we keep the two systems separate, the number describing the pair is

$$\Phi_{separate}(N_1, V_1, U_1; N_2, V_2, U_2) = \Phi_1(N_1, V_1, U_1) * \Phi_2(N_2, V_2, U_2),$$

the product resulting from the fact that each possible state of system 1 can be combined with each possible state of system 2.

If we admit interaction between the two — for example, heat exchange without particle exchange and without change of individual volumes V_1 and V_2 — we get

$$\Phi_{heat\,change}(N_1, V_1, U_1; N_2, V_2, U_2) = \Phi_1(N_1, V_1, U_1^*) * \Phi_2(N_2, V_2, U_2 + U_1 - U_1^*),$$

which leaves us with the problem of determining the energy U_1^* of system 1 after heat transfer.

The law of conservation of energy admits for U_1^* any value between 0 and $U_1 + U_2$. The Fundamental Assumption requires U_1^* to take the value for which $\Phi = \Phi_1 * \Phi_2$ becomes a maximum. This condition is equivalent to the condition

$$(\partial \ln \Phi_1/\partial U)_{N,V} = (\partial \ln \Phi_2/\partial U)_{N,V}. \qquad (15)$$

If this were not so — for example, if $\partial \ln \Phi_1/\partial U$ were greater than $\partial \ln \Phi_2/\partial U$ — energy would flow from system 2 to system 1 and the sum $\ln \Phi_1 + \ln \Phi_2$ — i.e. the product $\Phi_1 * \Phi_2$ — would increase, which would mean that the system had not yet reached equilibrium. (We introduced the logarithm in order to find the maximum of the product $\Phi_1 * \Phi_2$; differentiation of Φ_1 and Φ_2 would have lead to the maximum of $\Phi_1 + \Phi_2$.)

A quantity that becomes a maximum in thermal equilibrium between two closed systems of constant volume that exchange neither heat nor work nor particles with their environment must be closely related to entropy, which has exactly this property. In fact, entropy as defined by Clausius is related to Φ by Boltzmann's law

$$\boxed{\Phi = \exp(S/k)} \quad \text{or} \quad \boxed{S = k \ln \Phi} \qquad (16)$$

where k is Boltzmann's constant, introduced in Section 4.4. From experience we know that the quantity that becomes equal for two systems in thermal contact is temperature, and in fact we can identify $\partial \ln \Phi/\partial U$ with $1/T$:

$$\boxed{\left(\frac{\partial S}{\partial U}\right)_{N,V} = k\left(\frac{\partial \ln \Phi}{\partial U}\right)_{N,V} = \frac{1}{T}.} \qquad (17)$$

This is one of the most important equations in thermal physics. Referring to the question of this section's headline, we might thus answer that entropy is the abscissa of the log T–s diagram, the ordinate being $-\ln(\partial s/\partial u)$.

Equation (17) leads to the procedure for measuring entropy at constant volume, introduced in Section 4.3: measuring the change of temperature for a controlled change of internal energy ΔU (produced, for example, by an electric heater), we have

$$\Delta S \approx \Delta U/T = m\,c_p\,\Delta T/T,$$

a procedure that can be repeated stepwise in order to determine the temperature dependence of entropy at constant volume over an extended range. Volume and pressure dependence can be measured by similar methods.

Also, the other essentially thermodynamic quantities can be interpreted with reference to entropy — for example, pressure:

$$p = -\left(\frac{\partial U}{\partial V}\right)_S = T\left(\frac{\partial S}{\partial V}\right)_U = -\left(\frac{\partial U}{\partial V}\right)_T + T\left(\frac{\partial S}{\partial V}\right)_T. \qquad (18)$$

A numeric example for equation (16) may be instructive: 1 g of helium at 300 K, 1 bar has an entropy of 31.44 J/K. With $k = 1.38 * 10^{23}$ J/K we find

$$\Phi = \exp(S/k) = \exp(31.44/1.38 * 10^{-23}) \approx 10^{2.28 * 10^{24}},$$

i.e. a number of $2.28 * 10^{24}$ digits. Printed at 12 characters/inch, this number has a length of $4.82 * 10^{21}$ m ≈ half a million light-years, i.e. about five times the diameter of our galaxy. In comparison, the superconducting electrons in a superconductor are in a fully defined state described by a single macroscopic wave function, which contains all the information we can possibly have about the system. Its entropy is zero, Φ has 1 digit. Appreciate the difference!

'Boltzmann's constant' k, one of the fundamental constants of Nature, was introduced under that name by Planck in homage of Boltzmann, and the formulation $S = k \ln W$, engraved on Boltzmann's tomb, is due to Planck. (W stands for *Wahrscheinlichkeit*, probability, corresponding to Φ.) Boltzmann himself spoke about proportionality. In deep depression, attacked and despised by the positivistic school as a guru of unscientific atomistic metaphysics, he committed suicide in 1906. The triumph of the atom after the overwhelming discoveries at the end of the 19th and the beginning of the 20th century, however, brilliantly confirmed Boltzmann's conclusions and the power of his methods. But the positivistic warnings survived. They, too, were justified. The atom and its constituents turned out not to be the expected classical particles, and the new physics, quantum mechanics, eventually resigned from describing them in pictures taken from our familiar world, limiting itself to *statistical* statements about *probable* results of experiments.

Superconducting electrons and superfluid helium atoms are a particularly remarkable species of quantum particles. But that is another story...

9 REFERENCES

Cryogenics can be seen under many aspects, and many good books are available on the subject. This is not the place for a detailed review. Anyhow, most people confronted with cryogenics are not looking for a textbook they would read right through but for specific information providing insight into a feature that is the object of their current interest. In this respect, it is certainly good advice to scan the proceedings of recent conferences on cryogenic engineering. Two such conferences are held alternatingly every two years: the International Cryogenic Engineering Conference (ICEC) and the American Cryogenic Engineering Conference (CEC). The CEC Proceedings are published under the title 'Advances in Cryogenic Engineering'. The ICEC Proceedings were published in varying form and by several editors in the past; the last conferences were the object of a special edition of the journal *Cryogenics*. This journal is also is a good source of information.

My personal recommendation is:

Charles Kittel and Herbert Kroemer, Thermal Physics,
Second edition 1980, Freeman, San Francisco: 473 pages,

a fascinating introduction to thermodynamics, of great originality and in a very unpretentious form. It has strongly influenced my thinking about cryogenic problems, although cryogenics is only a small part of the book.

A very helpful and time-saving tool for access to numeric data about helium properties is HEPAK, a program written by V. Arp, R.D. McCarty and B.A. Hands and published by Cryodata, P.O.Box 558, Niwot, Colorado, USA 80544.

NEXT PAGE(S) left BLANK

SUPERFLUIDITY

W F Vinen
School of Physics & Space Research, University of Birmingham, Birmingham, UK

Abstract
The paper contains a brief and elementary introduction to the phenomenon of superfluidity in liquid helium-4. The experimentally observed properties of superfluid ^4He are described, and outline explanations are given in terms of the existence of a Bose condensate and a special form of excitation spectrum. The quantization of superfluid circulation is described, and the role played by free quantized vortex lines is explained.

1. INTRODUCTION

A fluid composed of electrically neutral particles that can, in certain circumstances, flow without friction is called a superfluid. There is a close analogy to superconductivity, which relates to frictionless flow of the electrons in certain metals at a low temperature. Two superfluids are known in terrestrial nature: liquid ^4He at a temperature below about 2 K; and liquid ^3He at a temperature below about 2 mK. The neutron fluid in a neutron star may also be a superfluid.

All superfluids owe their properties to some type of Bose condensation. In liquid ^4He, in which the atoms are themselves bosons, the Bose condensation is more or less straightforward; in liquid ^3He, in which the atoms are fermions, there is, loosely speaking, a Bose condensation of Cooper pairs, as in the electron fluid in a superconductor. However, superfluid ^3He is quite complicated, even in comparison with a conventional superconductor, because the Cooper pairs are formed in states with non-zero angular momentum and non-zero total nuclear spin. In this paper we shall concentrate on the simplest superfluid, formed from liquid ^4He. Liquid ^4He, in either its normal phase or its superfluid phase, is used as a coolant in practical applications: most commonly in superconducting magnets and SQUID systems.

It is not practicable to include in this paper a complete set of references. Much of the material is covered in reference [1], although sometimes from a slightly different point of view. A few other references are included, especially where the material is not covered in reference [1].

2. THE THERMODYNAMIC PROPERTIES OF LIQUID ^4He

The phase diagram is shown in Fig. 1. There is no conventional triple point, and no solid phase exists at pressures below about 25 atm. There are two liquid phases, helium I and helium II, separated by the "λ-line". Helium I is a conventional (normal) liquid; helium II is a superfluid.

The specific heat along the vapour pressure line is shown in Fig. 2. There is a sharp peak at the λ-line, of a type found in other systems at a temperature that marks the onset of some type of ordering process: for example, at the Curie point of a ferromagnetic.

3. THE ELEMENTARY SUPERFLUID PROPERTIES OF HELIUM II

Superfluid helium will flow without friction even through a very narrow channel, provided that the flow velocity is less than some critical value; it will flow easily even in anadsorbed film. The critical velocity depends on channel width (increasing with decreasing

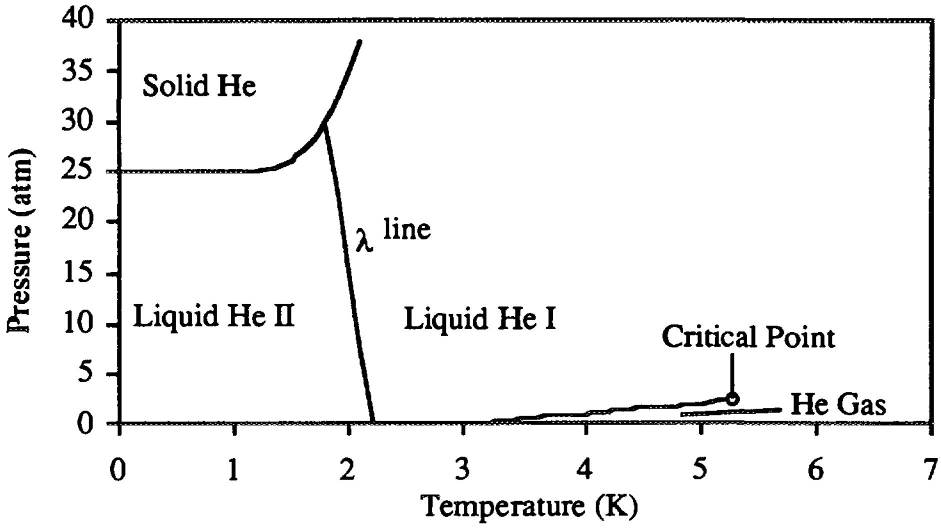


Fig. 1 The phase diagram for ^4He

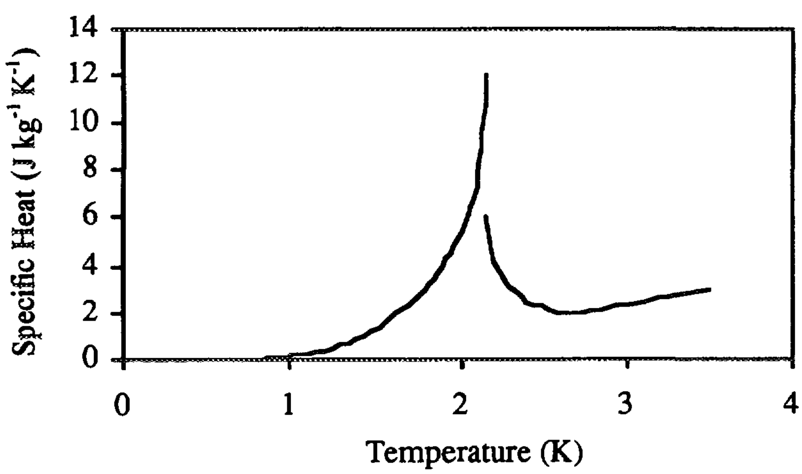


Fig. 2 The heat capacity of ^4He

width), and is typically a few cm per second. This frictionless flow suggests that the helium can behave as though it has no viscosity. However, an observation of the damping in the liquid of a disc oscillating at small amplitude in its own plane leads to a conventional viscosity equal in order of magnitude to that of Helium I. The behaviour appears to be somewhat irrational.

The thermal conductivity of helium I is very low. However, that of superfluid helium is high and unconventional: the heat flow is generally proportional to neither the cross-sectional area nor the temperature gradient. The heat conductivity is very high indeed at low heat currents, but the non-linearity means that it can be much reduced at high heat currents. Detailed investigation shows that heat conduction in superfluid helium obeys a wave equation ($\nabla^2 T = \left(1/c_2^2\right)\partial^2 T/\partial t^2$) rather than the usual diffusion equation ($K\nabla^2 T = \rho C \partial T/\partial t$); thus temperature fluctuations do not diffuse, but instead they propagate as a wave (called *second sound*).

When superfluid helium flows through a narrow channel, it emerges at a lower temperature: the *mechanocaloric effect*. Closely related to this effect is a large thermomechanical effect; a temperature gradient along a narrow channel containing superfluid helium leads to a large pressure gradient.

4. HELIUM AS A QUANTUM LIQUID: BOSE CONDENSATION

The fact that helium does not form a solid even at $T=0$, except at high pressure, is clearly indicative of quantum effects. Each atom in the liquid is confined by its neighbours; this confinement leads to a large zero-point energy, which increases with increasing liquid density; the zero-point energy therefore tends to keep the atoms apart to an extent that prevents solidification, the attractive forces between the helium atoms being quite weak.

The third law of thermodynamics requires that the entropy of the helium, even in its liquid state, must tend to zero as the temperature tends to zero. Superfluid helium at $T=0$ must therefore be an *ordered liquid*. We have noted that the peak in the heat capacity at the λ-line is characteristic of an order-disorder transition. It must surely follow that superfluidity is associated with this ordering process.

What type of ordering is possible in a liquid? To obtain a useful hint let us look at the predicted properties of a hypothetical ideal gas at very low temperatures (hypothetical because all real gases liquefy at a low temperature). The properties depend on the statistics of the particles: ^4He atoms are bosons, so we look at the properties of an ideal Bose gas. What we find predicted is *Bose condensation*. Below a certain critical temperature a finite fraction of the particles occupy the lowest quantum state in the vessel containing the gas, this fraction increasing to unity at $T \to 0$. The ordering at $T=0$ takes the form of putting all the particles in the same quantum state.

Can a similar effect occur in a liquid? And if so, is it responsible for superfluidity?

5. THE TWO-FLUID MODEL

Before we answer these questions we shall introduce a phenomenological model that describes the observed properties of superfluid helium, even when they appear to be somewhat irrational.

According to the two-fluid model superfluid helium consists of two interpenetrating fluids: a *normal component*, density ρ_n, behaving like a conventional viscous fluid and carrying all the entropy of the system; and a *superfluid component*, density ρ_s, behaving like an ideal classical inviscid liquid and carrying no entropy. In the simplest cases the two fluids can move relative to one another without frictional interaction. The proportion of superfluid, ρ_s/ρ, decreases from unity at $T=0$ to zero at the λ-transition (Fig. 3).

In terms of the two-fluid model heat flow in superfluid helium takes place by counterflow of the two fluids; as long as the relative velocity of flow is less than a critical value, which we shall discuss later, the only dissipative process opposing this counterflow arises from the viscosity of the normal fluid and is very weak. The apparently conflicting values of the viscosity obtained from flow through a narrow channel and from the damping of an oscillating disc are easily understood: flow through the channel involves only the superfluid component; the oscillating disc is damped by the normal fluid. Two types of wave motion are possible in the system: one in which the two fluids oscillate in phase (ordinary or *first sound*); and one in which they oscillate in antiphase (*second sound* or temperature waves). A measurement of the speed of second sound provides a good way of obtaining from experiment the normal and superfluid densities as functions of temperature (Fig. 3). They can also be obtained from the *Andronikashvili experiment*, in which a measurement is made of the period of oscillation of a pile of closely spaced discs suspended by a torsion fibre in the

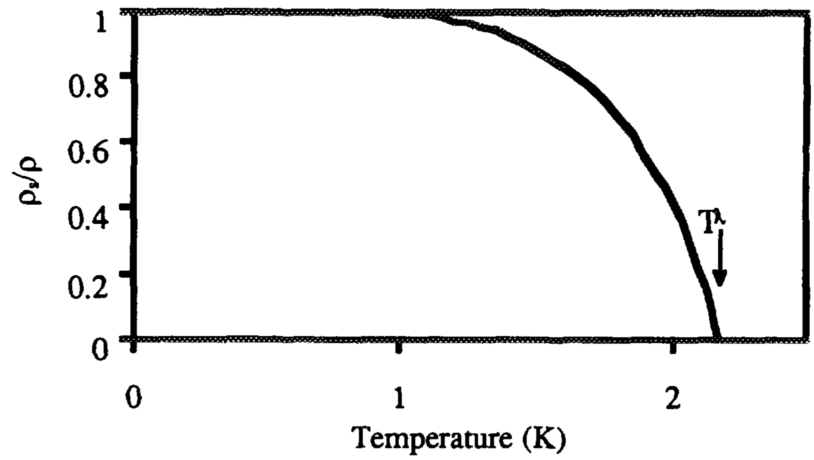


Fig. 3 The density of the superfluid component plotted against temperature

liquid. Only the normal fluid is dragged with the discs, so only the normal fluid contributes to the moment of inertia of the disc system. Explanations of the thermomechanical and mechanocaloric effects are fairly obvious and are left to the reader to consider.

When the two-fluid model was first proposed it was thought that it might be associated with Bose condensation, the normal fluid component being identified with the excited atoms and the superfluid component with the condensed atoms. We know now that this view is an oversimplification, as we shall see in the next section.

6. BOSE CONDENSATION IN LIQUID HELIUM

That a form of Bose condensation does indeed occur in liquid helium below the λ-transition was probably first demonstrated theoretically by Penrose & Onsager [2] (perhaps also by Bogolubov [3]), and confirmation has been provided by much subsequent theoretical study. Direct experimental evidence has been hard to find, but reasonably convincing evidence has now come from experiments on deep inelastic neutron scattering, which measures in principle the momentum distribution of the bare helium atoms.

However, Bose condensation in a liquid differs from that in the ideal gas in that the fraction of atoms in the lowest momentum state (which can be formally defined in terms of the single-particle density matrix) is much smaller and, in liquid helium at low pressures, reaches a value of only about 0.1 even at T=0. This means that this *condensate fraction* cannot be identified with the superfluid fraction ρ_s/ρ. We shall return to the question of what determines the superfluid fraction a little later.

6.1 Bose condensation and superfluidity

Let us imagine that the superfluid phase of liquid helium is contained in a long tube, and that we suddenly set the whole liquid into motion along the tube with velocity **V** by means of an impulsive pressure gradient. Since the whole liquid is set in motion the condensate must also be moving with the same velocity, so that the wave function of the state into which condensation has occurred takes the form

$$\Psi = \Psi_0 \exp(i\mathbf{k}\cdot\mathbf{r}), \tag{1}$$

where $\hbar \mathbf{k} = m_4 \mathbf{V}$ and m_4 is the mass of a helium atom.

The condensate contains a macroscopic number of atoms in a single quantum state (this is still true even when the condensate *fraction* is relatively small). We can therefore regard Ψ, not as a single particle wave function, but rather as a *coherent particle wave* (the *condensate wave function*), analogous to the coherent electromagnetic wave produced by a laser.

As the liquid helium moves along the tube the helium atoms interact with the walls of the tube; scattering of atoms will occur, and some atoms will be scattered out of the condensate. This scattering out of the condensate will reduce the *amplitude* of condensate wave function, but it may not change the phase. Let us suppose that the phase is indeed maintained. The scattering will lead to equilibration of the helium with the walls, but the equilibration takes place in such a way that the phase of the condensate wave function is constrained to remain unchanged. This is a case of *broken symmetry*, and can be compared with the situation in a ferromagnetic material in which the overall magnetization is constrained to point in a certain direction, although the individual atomic magnetic moments are free to point in, for example, either of two opposing directions. Given the broken symmetry the equilibrium state of the helium may involve a non-zero mass current density, which is likely to be of the form

$$\mathbf{J}_s = \rho_s \mathbf{V}, \tag{2}$$

at low velocities. We have therefore described a persistent (frictionless) flow of the *superfluid component*, density ρ_s, moving with velocity $\mathbf{v}_s = \mathbf{V}$.

We emphasize that the equilibrium associated with the broken symmetry can be only metastable; the state of the helium with no supercurrent density has a lower energy. The metastable state must have a finite lifetime. The fact that superflow in helium can take place with no observable friction shows that this lifetime can be very long. Later we shall discuss what determines this lifetime, a discussion that will make clearer why the phase of the condensate wave function tends to be constrained to remain unchanged.

The broken symmetry represents an ordering in the liquid helium, associated with the Bose condensation. The condensate wave function Ψ can be regarded as the *order parameter* in the superfluid phase. It is analogous to the Ginsburg-Landau order parameter in superconductivity.

Although we have developed a satisfying description of superfluidity, it tells us nothing about the magnitude of the superfluid density ρ_s. Indeed we have so far presented no argument for supposing that ρ_s is non-vanishing at temperatures below the Bose condensation temperature.

7. THE NORMAL FLUID FRACTION: PHONONS AND ROTONS

Curiously enough some understanding of what determines the densities ρ_s and ρ_n was provided by Landau [4, 5] before the concepts of Section 6 had been developed. Landau introduced the idea that the low-lying excited states of helium (those that can be thermally excited at low temperatures) can be described in terms of a low-density gas of weakly interacting excitations. Nowadays we are familiar with this type of idea in connection with crystalline solids, where the excited states can be described in terms of quantized normal modes (lattice vibrations), which we call *phonons*. Landau suggested that the lowest-lying excitations in liquid helium are also phonons (longitudinal only), but that there are other excitations at somewhat higher energy, which he called *rotons*. We now know that the phonons and rotons are not clearly distinguishable, and that they form a continuous spectrum as shown in Fig. 4.

Direct experimental evidence for this form of the excitation spectrum has been provided by inelastic neutron scattering.

The thermally excited states of helium can be described in terms of a weakly interacting gas of phonons and rotons only as long as the density of these excitations is small. This means that the temperature must not be too large; in practice less than about 1.8 K. At higher temperatures the situation becomes more complicated.

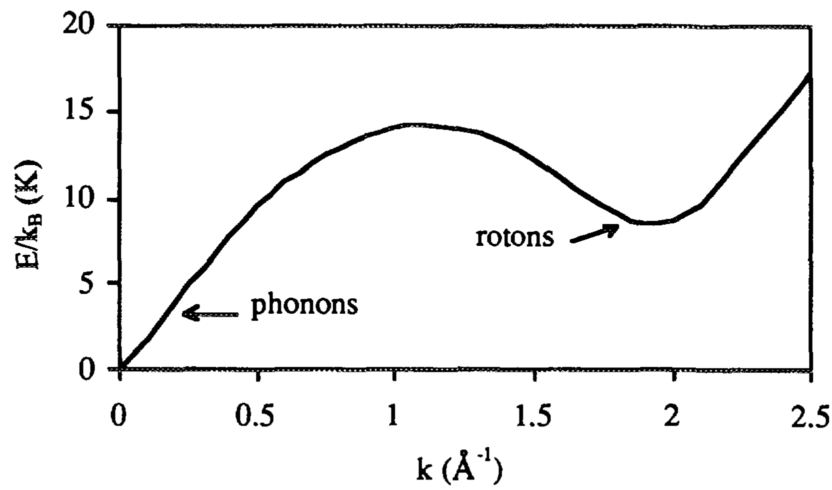


Fig. 4 The excitation spectrum in superfluid ^{4}He

We shall now argue that, at the low temperatures, the normal fluid can be identified with the gas of excitations, and that we can use this picture to calculate the normal and superfluid fractions.

Let us think again of the liquid helium as having been suddenly set in motion along a tube with velocity $\mathbf{V}$. The equilibration process (subject to the constraint of an unchanged phase of the order parameter), which we described in general terms earlier, can now be seen to be the process in which the excitation gas comes into equilibrium with the walls of the channel, the background fluid continuing to move with the velocity $\mathbf{V}$. Since the excitations are effectively an ideal gas, the equilibration is easily described; one has simply to calculate the distribution function for the excitations when they are in equilibrium with the channel walls, in the presence of the background velocity $\mathbf{V}$. Having obtained the excitation distribution function, we can easily use it to calculate the momentum density, $\mathbf{P}_{ext}$, associated with the excitations. To obtain the new total momentum density in the liquid we must add to $\mathbf{P}_{ext}$ the original momentum density, $\rho\mathbf{V}$, giving

$$\mathbf{P}_s = \mathbf{P}_{ext} + \rho\mathbf{V}. \tag{3}$$

For a sufficiently low temperature we find that $\mathbf{P}_s > 0$, $\mathbf{P}_s$ being proportional to $\mathbf{V}$ for small $\mathbf{V}$, and we identify this residual momentum density with that of the moving superfluid

$$\mathbf{P}_s = \rho_s\mathbf{V}. \tag{4}$$

Hence we have shown that ρ_s is non-zero, and we have a method for calculating its value at low temperatures. The results are in very good agreement with experiment, when use is made of the detailed form of the excitation spectrum obtained from inelastic neutron scattering.

7.1 Superfluidity and the form of the excitation spectrum

It must be emphasized that ρ_s is found to be non-zero only if the excitation spectrum has the appropriate form. The existence of only phonons at the lowest energies is a *sufficient* condition. Otherwise the liquid is not a superfluid, even though there may be a Bose condensate. The ideal Bose gas provides an example of a system for which the excitation spectrum does *not* have the right form and it is not a superfluid.

It seems therefore that superfluidity requires both a condensate and the appropriate form of excitation spectrum. Particle interactions are required in order to produce the required form of excitation spectrum.

Space does not allow us to discuss exactly why the excitation spectrum has the form observed. Much theoretical work has been devoted to this question; perhaps the most elegant discussion was provided by Feynman [6].

8. MACROSCOPIC QUANTUM EFFECTS

Superfluidity is clearly a quantum phenomenon. But most properties of condensed matter depend ultimately on quantum effects. What is unique to a superfluid is the appearance of quantum effects on a macroscopic scale, as we now describe. These quantum effects are related to the *quantization of superfluid circulation,* which arises from the long-range phase coherence in the condensate wave function.

8.1 Rotation of the superfluid, and the quantization of superfluid circulation

As is clear from Eq. (1), the velocity, $\mathbf{v}_s$, of the superfluid component is linked to the phase, S, of the condensate wave function. In general

$$\mathbf{v}_s = \frac{\hbar}{m_4}\nabla S . \tag{5}$$

It follows that $\mathrm{curl}\,\mathbf{v}_s = 0$, which suggests that the superfluid cannot rotate. (We note in passing that this irrotation condition is the analogue of the London equation in superconductivity.)

The irrotation condition forbids rotation only in a simply-connected volume. In a multiply-connected volume it allows a non-zero circulation round any circuit that cannot be continuously deformed into a loop of zero size while remaining in the liquid. The circulation in the superfluid component is defined as

$$\kappa = \oint \mathbf{v}_s \cdot d\mathbf{r} , \tag{6}$$

where the integral is taken round the circuit concerned (Fig. 5). However, in a superfluid the *circulation cannot take any value. The condensate wave function must presumably be single-valued,* and it is easy to see that this leads to the condition

$$\kappa = n\frac{h}{m_4}, \tag{7}$$

where n is an integer. The quantum of circulation, h/m_4, is macroscopically large (roughly 10^{-7} m^2s^{-1}). The condition (7) has been verified experimentally by measuring the circulation round a stretched wire running through the liquid; the Magnus (lift) force that arises from the circulation modifies the modes of transverse vibration of the wire in a way that can be easily observed and measured. The quantization of circulation in a superfluid is analogous to the quantization of flux in a superconductor.

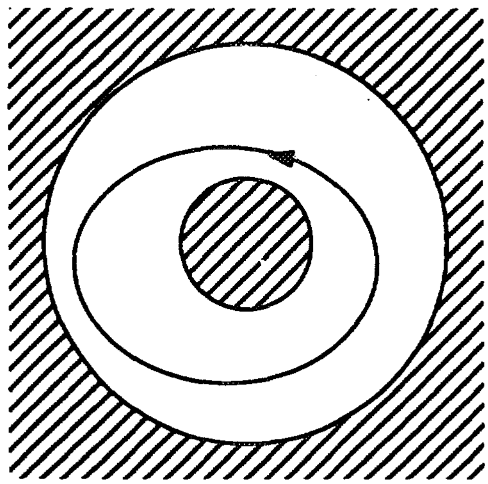


Fig. 5 A path round a torus along which the circulation does not necessarily vanish

8.2 Free quantized vortex lines

So far we have implicitly assumed that a volume of helium can be multiply-connected only if it is contained in a vessel with suitable geometry. However, a volume of liquid can also become multiply-connected if a small hole runs through it. We can establish a single quantum of superfluid circulation round this hole, and we have created a free quantized vortex line. Left to itself the hole acquires a size comparable with the interatomic spacing, so the circulation exists round what is virtually a line through the liquid. This type of quantized vortex line is the analogue of a flux line in a type II superconductor.

If the superfluid phase of liquid helium is situated in a rotating vessel of reasonable size, it is observed that the whole liquid appears to rotate with the vessel; the observed shape of the meniscus of the liquid indicates this quite clearly. The irrotation condition implies that the superfluid cannot be rotating like an ordinary classical liquid. What happens is that the superfluid becomes filled with an array of quantized vortex lines, all parallel to the axis of rotation (Fig. 6). For an angular velocity of 1 rad s^{-1}, the spacing between the lines is about 0.2 mm, and the spatially averaged flow velocity in the superfluid is then closely similar to that associated with solid-body rotation. The discrete vortex structure will give rise to dimples on the free liquid surface, but they turn out to be too small to see.

The presence of the free vortex lines changes some of the properties of the superfluid phase. The excitations that constitute the normal fluid are scattered by the cores of the vortices, and therefore any relative motion of the superfluid and the normal fluid results in a frictional force between them: a force of *mutual friction*. An observation and study of this force in uniformly rotating liquid helium provided the first experimental evidence for the existence of quantized vortex lines and indeed for the quantization condition on the circulation.

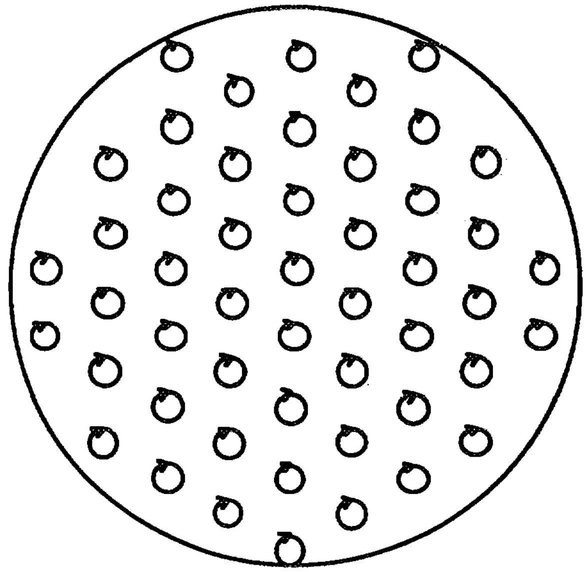


Fig. 6 Vortices in uniformly rotating superfluid helium

8.3 The breakdown of ideal frictionless superflow

As we mentioned earlier, frictionless superflow is observed only at velocities that are smaller than some critical value, which is typically of order 1 cm s^{-1}. Flow at supercritical velocities is found to be accompanied by a mutual friction between the two fluids, similar to that found in the uniformly rotating liquid. Detailed study showed that this "supercritical mutual friction" was due to the generation within the superfluid component of an irregular array of quantized vortices. The process is the quantum analogue of the production of turbulence in a classical fluid. An understanding of this quantum turbulence is necessary for a full understanding of heat conduction in superfluid helium, and it may therefore be important if superfluid helium is to be used as a cooling agent.

In connection with the use of helium as a cooling agent it should be added that, when heat flows from a solid body into liquid helium, there is a large thermal resistance (of order 2 K W^{-1} cm^2) associated with the boundary (the *Kapitza resistance).* Heat flow through the boundary involves the transfer of energy between the excitations in the liquid and those in the solid, and the Kapitza resistance arises from a kind of acoustic mismatch across the boundary.

8.4 The metastability of superflow

We emphasized in Section 7 that ideal frictionless superflow can be only metastable, even at low velocities. If, for example, superflow is established round a toroidal channel, the supercurrent cannot persist for ever; it must ultimately decay, although the time associated with this decay may be very long.

We have seen in the preceding section that the generation of quantized vortices gives rise to a frictional force on the superfluid, and it must be the generation of such lines that causes the ultimate decay of a supercurrent. Let us look at the fundamental nature of this process in more detail.

Consider superflow round the toroidal channel shown in Fig. 7. The velocity of flow, $\mathbf{v}_s$, must be such that an integral number, N, of quanta of circulation are associated with the hole in the centre of the toroid. Decay of the supercurrent will involve the loss of one or more of these quanta. Such a decay can occur only if a free vortex is created in the helium and then moves across the channel. Such a process changes the phase of the order parameter within the helium, and it is often referred to as phase slip. In practice the phase cannot change in any other way.

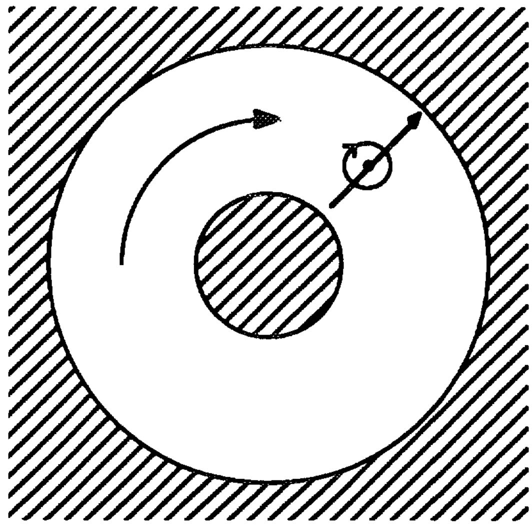


Fig. 7 Decay of a persistent supercurrent by vortex motion

It is easily seen that the creation and movement of the vortex across the channel is opposed by a potential barrier. This barrier has its origin in an attraction that exists between a vortex and a neighbouring wall (or equivalently between a vortex and its image in the wall). *The metastability of the superflow depends on the existence of this barrier*; the presence of a barrier of this type is characteristic of all metastable states. The lifetime of the state is determined by the rate at which the barrier can be surmounted.

The barrier height for the vortex motion in various geometries has been calculated. Its height depends on the velocity v_s; for small velocities, of order 1 cm s^{-1}, and for temperatures not too close to the λ-line, the barrier turns out to be so high that the lifetime ought to be astronomical. In reality the lifetime is often much shorter; frictionless superflow seems to break down much more easily than is predicted. As is now widely accepted, this easy breakdown is associated with the almost inevitable presence in any real volume of helium of a few free vortices that are probably pinned to protuberances on the walls of the containing vessel and are remnants of a large concentration of vortices that is known to be created as the helium is cooled through the λ-transition [7]. These remanent vortices act to nucleate new vortices, in much the same way as remanent dislocations act to generate new dislocations in a solid crystal and hence reduce its intrinsic strength. In the past few years experimental conditions have been found in which the effect of these remanent vortices has been minimized or eliminated, and the vortex nucleation barrier has then been shown to agree with that predicted from the image forces. The barrier can still be overcome at sufficiently high flow rates, and recent experimental and theoretical studies (see, for example, references [1, 8, 9]) have demonstrated the existence of processes involving either thermal excitation over the barrier or quantum tunnelling through it, depending on the temperature.

9. SUMMARY AND CONCLUSIONS

In this paper we have tried to describe some of the remarkable phenomena associated with superfluidity in liquid ^4He, and to give an indication of the way these phenomena can be understood in terms of effects associated primarily with a form of Bose condensation in the liquid. One might wonder why such a wealth of strange phenomena has not led to more practical applications.

ACKNOWLEDGEMENT

I am grateful to Philip Elliott for help with the preparation of this paper.

REFERENCES

[1] D.R. Tilley & J. Tilley, Superfluidity and Superconductivity, 3rd Edition (Institute of Physics Publishing, Bristol & Philadelphia, 1990)

[2] O. Penrose & L. Onsager, Phys.Rev. **104**, 576 (1956)

[3] N.N. Bogolubov, J.Phys. USSR **11**, 23 (1947)

[4] L.D. Landau, J.Phys. USSR **5**, 71 (1941)

[5] L.D. Landau, J.Phys. USSR **11**, 91 (1947)

[6] R.P. Feynman, Progress in Low Temperature Physics vol.1, p.17 (North-Holland Publishing Company, Amsterdam, 1955)

[7] P.C. Hendry, N.S. Lawson, R.A.M. Lee, P.V.E. McClintock & C.D.H. Williams, Nature **238**, 315 (1994)

[8] G.G. Ihas, O. Avenal, R. Aarts, R. Salmelin & E. Varoquaux, Phys.Rev.Lett. **69**, 327 (1992)

[9] J.C. Davis, J. Steinhauer, K. Schwab, Yu.M. Mukarsky, A. Amar, Y. Sasaki & R.E. Packard, Phys.Rev.Lett. **69**, 323 (1992)

NEXT PAGE(S)
left BLANK

OPERATING AT 1.8 K: THE TECHNOLOGY OF SUPERFLUID HELIUM

Ph. Lebrun
CERN, Geneva, Switzerland

Abstract
The technical properties of helium II ("superfluid" helium) are presented from the user point of view. Its applications to the cooling of superconducting devices, particularly in accelerators and colliders, are discussed in terms of heat transfer capability and limitations in conductive and convective modes. Large-capacity refrigeration techniques below 2 K are reviewed from the point of view of thermodynamic cycles as well as process machinery. Examples drawn from existing or planned projects illustrate the presentation.

1. INTRODUCTION

Once a curiosity of nature and still today an arduous research topic in condensed-matter physics, superfluid helium*) has recently become a technical coolant for advanced superconducting devices, to the point that it is now implemented in industrial-size cryogenic systems, routinely operated with high reliability. Two main reasons dictate the use of superfluid helium as a coolant in superconducting devices, namely the lower temperature of operation, and the enhanced heat transfer properties at the solid-liquid interface and in the bulk liquid.

The lower temperature of operation is exploited in high-field magnets [1, 2], to compensate for the monotonously decreasing shape of the superconducting transition frontier in the current density-versus-magnetic field plane, shown in Fig. 1 for superconducting materials of technical interest. In this fashion, the current-carrying capacity of the industrial Nb-Ti superconducting alloys can be boosted at fields in excess of 8 T, thus opening the way for their use in high-field magnet systems for condensed-matter physics [3, 4, 5], magnetic confinement fusion [6, 7] and circular particle accelerators and colliders [8, 9]. In the case of high-frequency superconducting devices, such as acceleration cavities [10], the main drive for superfluid helium cooling is the exponential dependence of the BCS losses on the ratio of operating-to-critical temperature. Accelerators based on this technology, such as medium-energy, high-intensity machines [11, 12] and future high-energy lepton colliders [13, 14] operate in the temperature range which minimizes capital costs and overall energy consumption. This issue is schematized in Fig. 2.

The technical heat transfer characteristics of superfluid helium basically derive from its peculiar transport properties, such as high heat capacity, low viscosity and good effective thermal conductivity [15, 16]. In order to be fully exploited, however, both in steady-state and transient regimes, e.g. for power heat transport over macroscopic distances as well as intimate stabilization of superconductors, they require elaborate thermohydraulic design of the cooling circuits, conductor, insulation and coil assemblies. This often conflicts with other technical or economic requirements of the projects and acceptable trade-offs have to be found.

---

*) Strictly speaking, we are referring to the second liquid phase of helium, called He II, which exhibits the unusual bulk properties associated with superfluidity and is therefore also called "superfluid". This is not to be confused with the entropy-less component of the phenomenological two-fluid model accounting for the behaviour of He II, for which some authors prefer to keep the qualificative "superfluid".

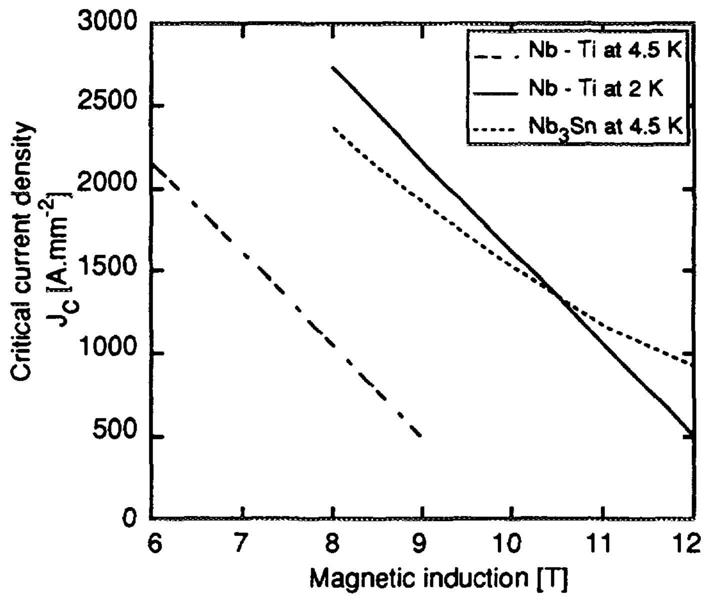


Fig. 1 Critical current density of technical superconductors

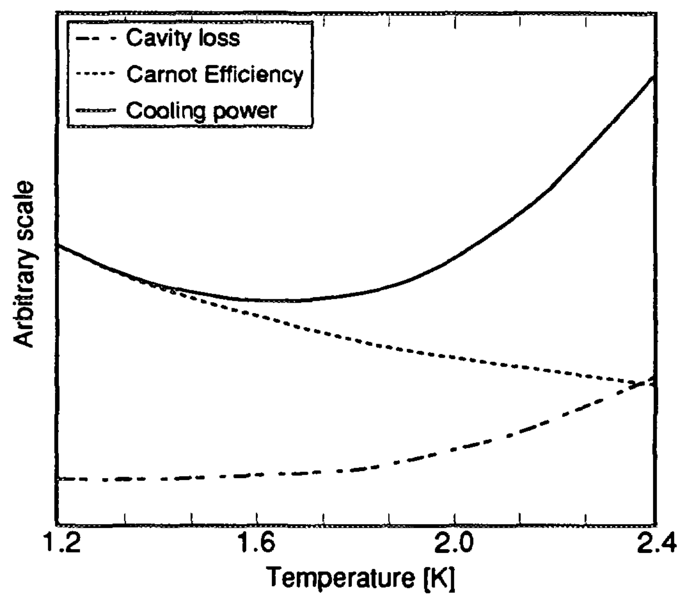


Fig. 2 Optimal operating temperature of RF superconducting cavities

2. PRESSURIZED VERSUS SATURATED SUPERFLUID HELIUM

A look at the phase diagram of helium (Fig. 3) clearly shows the working domains of saturated helium II, reached by gradually lowering the pressure down to below 5 kPa along the saturation line, and pressurized helium II, obtained by subcooling liquid at any pressure above saturation, and in particular at atmospheric pressure (100 kPa).

Although requiring one more level of heat transfer and additional process equipment — in particular a pressurized-to-saturated helium II heat exchanger — implementation of pressurized helium II for cooling devices brings several important technical advantages [17]. Avoiding low-pressure operation in large and complex cryogenic systems clearly limits the risk of air inleaks, and resulting contamination of the process helium. Moreover, in the case of electrical devices, the low dielectric strength exhibited by low-pressure helium vapour [18], in the vicinity of the minimum of the Paschen curve [19], brings the additional risk of electrical breakdown at fairly low voltage. Operating in pressurized helium II avoids this kind of problem.

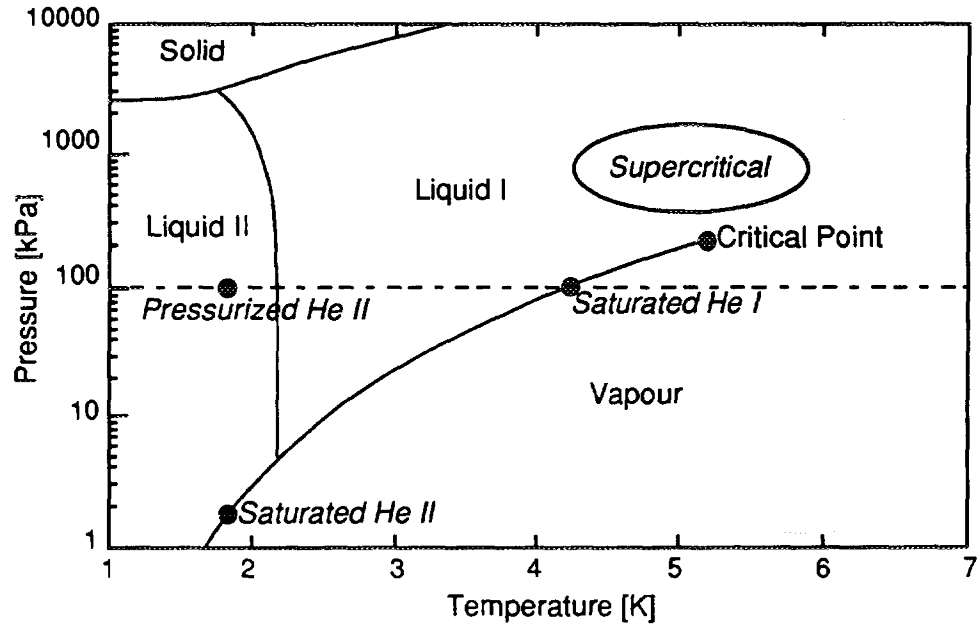


Fig. 3 Phase diagram of helium

However, the most interesting and specific aspect of pressurized helium II in the operation of superconducting devices, stems from its capacity for cryogenic stabilization. As a subcooled (monophase) liquid with high thermal conductivity, pressurized helium II can absorb in its bulk a deposition of heat, up to the temperature at which the lambda-line is crossed, and local boiling then starts due to the low thermal conductivity of helium I. Quasi-saturated helium II, which is in fact slightly subcooled due to the hydrostatic head below the surface of the liquid bath, may only absorb heat deposition up to the point at which the saturation line is crossed, and change of phase occurs. The enthalpy difference from the working point to the transition line is usually much smaller in the latter case. The argument, developed in reference [20], typically yields an order of magnitude better performance in favour of pressurized helium II.

3. CONDUCTION COOLING

In the following we shall only consider conductive heat transport in helium II at heat fluxes of technical interest (typically above 1 $kW.m^{-2}$). For most practical geometries, this means working in the "turbulent" regime with full mutual friction between the components of the two-fluid model [21]. In this regime, helium II exhibits a large, finite and non-linear bulk heat conductivity, the value of which depends both on temperature and heat flux. While the general patterns of this behaviour can be predicted by the Gorter-Mellink theory*), practical data useful for engineering design has been established in a number of experiments [22-27].

Consider conduction in one-dimension, e.g. in a tubular conduit of length L, the ends of which are maintained at temperatures T_c and T_w. The steady-state heat flux q' is given by:

*) C.J. Gorter and J.H. Mellink introduced in 1949 the idea of an interaction producing mutual friction between the components of the two-fluid model, to account for the observed transport properties of helium II.

$$q'^n L = X(T_c) - X(T_w) \quad (1)$$

where the best experimental fit for n is 3.4, and $X(T)$ is a tabulated function of temperature, physically analog to a conductivity integral [22]. A plot of this function in Fig. 4 reveals that the apparent thermal conductivity of helium II goes through a maximum at around 1.9 K.

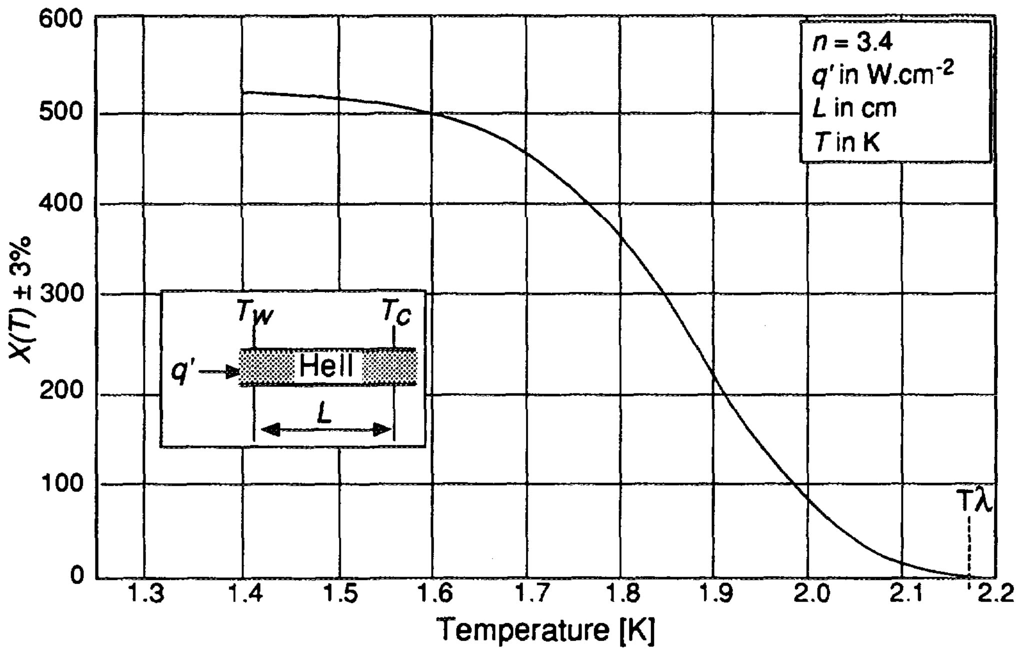


Fig. 4 Thermal conductivity integral of pressurized superfluid helium [22]

As an example, the heat flux transported by conduction between 1.9 and 1.8 K in a 1-m long static column of helium II is about 1.2 $W.cm^{-2}$, i.e. three orders of magnitude higher than that conducted along a bar of OFHC copper of the same geometry! The non-linearity with respect to heat flux also results in a much weaker dependence of conduction upon length, or thermal gradient. While the heat flux conducted in a solid is directly proportional to the thermal gradient applied, doubling the conduction length in a column of helium II only reduces the heat flux by some 20 %.

The variation of $X(T)$ also implies that, for each value of the cold boundary temperature T_c, there exists a maximum possible heat flux at which T_w reaches the lambda point, and the helium column ceases to be superfluid. Values of this limiting heat flux, which also weakly depends on L, range from a fraction to a few units of $W.cm^{-2}$, for practical cases of interest. This clearly brings an intrinsic limitation in the applicability of helium II conduction for quasi-isothermal cooling of long strings of superconducting devices in an accelerator. Transporting tens of watts over tens of meter distances would then require several hundred mK temperature difference and a large cross-section of helium, which is both impractical and thermodynamically costly. For a more precise estimate, consider a uniformly heated tubular conduit of length L, operating between temperatures T_c and T_w, and apply the helium II steady-state conduction equation to this fin-type geometry. After integration:

$$q'^n_{total} L = (n+1) [X(T_c) - X(T_w)] \quad (2)$$

where q'_{total} is the total heat flux flowing through the section at temperature T_c, near the heat sink. As an example, cooling a 50-m long cryomagnet string, with a uniform linear thermal load of 1 $W.m^{-1}$, by conduction between 1.9 K (temperature of the warmest magnet) and 1.8 K (temperature at the heat sink), would require a helium II cross-section of 90 cm^2, i.e. a 10.7-cm diameter conduit. In view of such constraints, the conduction-cooling scheme originally proposed for the LHC project [28] was later abandoned. It is however currently used for cryogenic testing of single LHC magnets [29].

The high thermal conduction in helium II can however be exploited to ensure quasi-isothermality of helium enclosures of limited spatial extension, such as the helium bath of a superconducting magnet under test. Knowledge of temperature changes at any point in the bath then permits to assess enthalpy changes of the system, and thus to perform calorimetric measurements. This technique proves very convenient for measuring minute heat inleaks [30] or energy dissipation [31] produced by ramping losses or resistive transitions in superconducting magnets.

4. FORCED CONVECTION OF PRESSURIZED SUPERFLUID HELIUM

To overcome the limited conduction of helium II in long strings of cryogenic devices, the obvious issue is to create a forced circulation of the fluid in a cooling loop, thus relying on convective heat transfer. One can then benefit from an additional control parameter, the net velocity imparted to the bulk fluid. In the following we shall only discuss convection in channel diameters of technical interest, i.e. typically greater than 1 cm. The flow induced by a pressure gradient across an hydraulic impedance is then essentially determined by the viscosity of the bulk fluid. Assuming that internal convection between the components of the two-fluid model is independent of the net velocity, reduces the problem to the behaviour of a flowing monophase liquid with high, non-linear thermal conductivity. The steady-state heat transport Q' between two points 1 and 2 of the cooling loop is then given by the difference in enthalpy of the fluid flowing with a mass flow-rate m':

$$Q' = m' (H_2 - H_1) \tag{3}$$

An estimate of the potential advantage of forced convection over conduction can be made, using the same geometry and temperature boundary conditions as described in Section 3 above. Consider helium II pressurized at 100 kPa, flowing in a heated pipe of length 1 m and cross section 1 cm^2, and assume its temperature increases from 1.8 K at pipe inlet, to 1.9 K at outlet. It is easy to show that for flow velocities above 0.2 m.s^{-1}, convective heat transport exceeds conduction.

The above calculation however neglects pressure drop along the flow. A look at the pressure-enthalpy diagram of helium (Fig. 5) reveals a positive Joule-Thomson effect [32]: the enthalpy of the fluid increases both with increasing temperature and pressure, so that an isenthalpic expansion results in a temperature increase. For example, pressurized helium II flowing across a pressure gradient of 50 kPa will warm up from 1.8 K to 1.9 K, in the absence of any applied heat load. The magnitude of this effect requires precise knowledge of the thermohydraulic behaviour of helium II, in order to validate its implementation in long cooling loops [33].

Following early exploratory work [34, 35], several experimental programs have investigated heated flow of pressurized helium II in pipes and piping components [36, 37], culminating with the 230-m long test loop in Grenoble [38, 39] which gave access to high Reynolds numbers and extended geometries characteristic of accelerator string cooling loops. In parallel to that work, mathematical models were developed for calculating combined conductive and convective heat transport processes in complex circuits [40, 41], and validated on experimental results. Pressure drop and heat transfer — both steady-state and transient — in flowing pressurized helium II may now be safely predicted for engineering purposes, using well-established laws and formulae.

The implementation of forced-flow cooling requires cryogenic pumps operating with pressurized helium II. Although most of the experimental work has been performed using positive displacement, i.e. bellows- or piston-pumps originally developed for helium I [42], the thermomechanical effect, specific of the superfluid, may also be used for driving cooling loops by means of fountain-effect pumps [43–46]. In spite of their low thermodynamic efficiency [47], a drawback of limited relevance for using them as circulators which have to produce low pumping work, fountain-effect pumps are light, self-priming and have no moving parts, assets of long-term reliability e.g. for embarked applications in space [48]. At higher heat loads, they have been considered [49] and tested [50] for forced-flow cooling of superconducting magnets: the overall efficiency of the process may then be improved by configuring the cooling loop so as to make use of the heat load of the magnet proper to drive the thermomechanical effect in the pump [51].

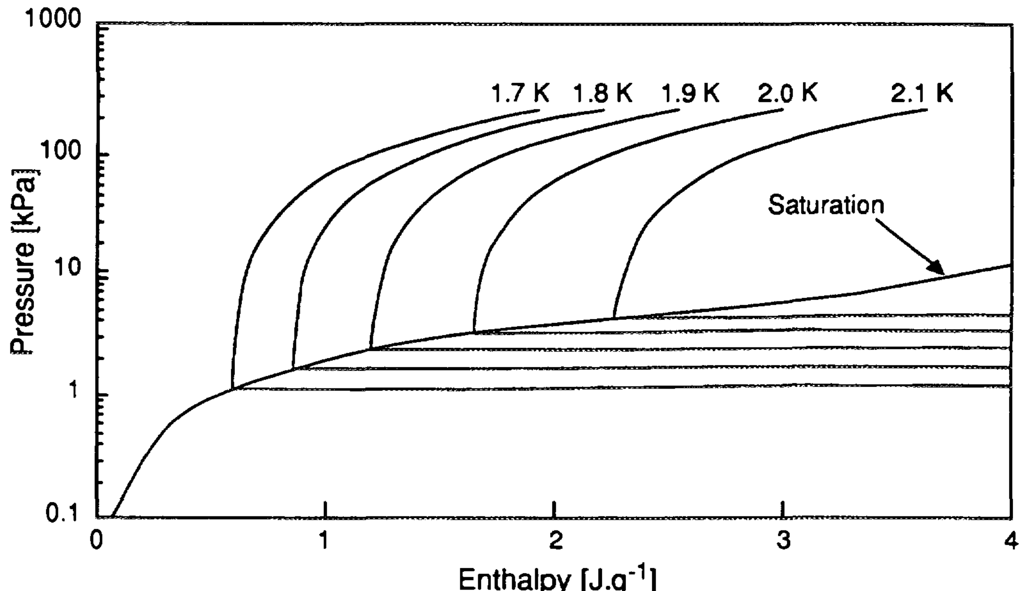


Fig. 5 Pressure-enthalpy diagram of superfluid helium

5. TWO-PHASE FLOW OF SATURATED SUPERFLUID HELIUM

The conductive and convective cooling systems described above, both transport heat deposited or generated in the load, over some distance through pressurized helium II, up to a lumped pressurized-to-saturated helium II heat exchanger acting as quasi-isothermal heat sink. This is achieved at the cost of a non-negligible — and thermodynamically costly — temperature difference, thus requiring to operate the heat sink several hundred mK below the temperature of the load.

A more efficient alternative is to distribute the quasi-isothermal heat sink along the length of the accelerator string. In this fashion the conduction distance — and hence the temperature drop — in pressurized helium II is kept to a minimum, typically the transverse dimension of the device cryostat. This leads to the cooling scheme proposed for the LHC at CERN, and schematized in Fig. 6: the superconducting magnets operate in static baths of pressurized helium II at around atmospheric pressure, in which the heat load is transported by conduction to the quasi-isothermal linear heat sink constituted by a corrugated copper heat exchanger tube, threading its way along the magnet string, and in which flowing two-phase saturated helium II gradually absorbs the heat as it vapourizes [9].

Although potentially attractive in view of its efficiency in maintaining long strings of magnets at quasi-uniform temperature, this cooling scheme departs from the well-established wisdom of avoiding long-distance flow of two-phase fluids at saturation, particularly in horizontal or slightly inclined channels. Moreover, no experimental data was available for flowing saturated helium II, and very little for other cryogenic fluids in this configuration.

Following first exploratory tests [52] which demonstrated the validity of the concept on a reduced geometry, a full-scale thermohydraulic loop [53] permitted to establish the stability of horizontal and downward-sloping helium II flows, to observe partial (but sufficient) wetting of the inner surface of the heat exchanger tube by the liquid phase, thanks to flow stratification, and to address process control issues and develop strategies for controlling uniformity of temperature at strongly varying applied heat loads, in spite of the low velocity of the liquid phase. As long as complete dryout does not occur, an overall thermal conductance of about 100 $W.m^{-1}.K^{-1}$ can be reproducibly observed across a ND40 heat exchanger tube, made of industrial-grade deoxidized phosphorus copper.

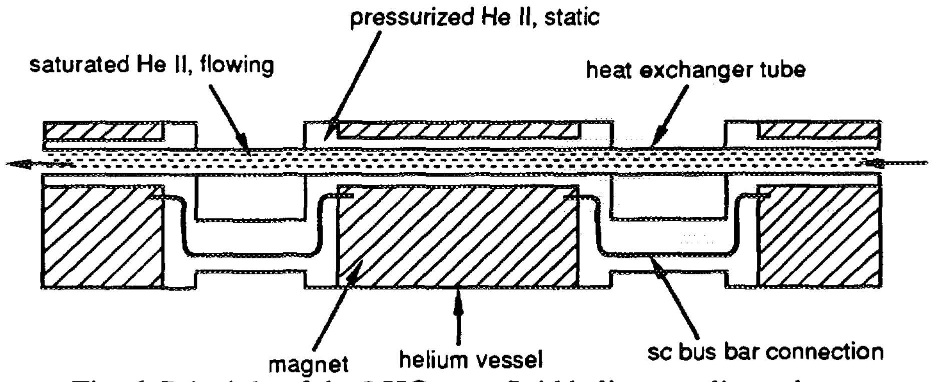


Fig. 6 Principle of the LHC superfluid helium cooling scheme

Once the wetting of the inner surface of the tube is guaranteed, the heat transfer from the pressurized to the saturated helium II is controlled by three thermal impedances in series: solid conduction across the tube wall, and Kapitza resistance at the inner and outer interfaces between tube wall and liquid. While the former can be adjusted, within technological limits, by choosing tube material and wall thickness, the latter, which finds its origin in the refraction of phonons at the liquid-solid interfaces and is thus strongly temperature-dependent, largely dominates below 2 K [54]. The use of high-purity, cryogenic-grade copper such as OFHC is thus not required, and pressurized-to-saturated helium II heat exchangers can be made of DLP grade - "plumber's" copper.

The final validation of the two-phase helium II flow cooling scheme for LHC has been performed successfully on a 50-m long test string, equipped with full-scale prototype cryomagnets, operated and powered in nominal conditions [55]. At varying heat loads exceeding 1 $W.m^{-1}$, all magnets in the string were maintained in a narrow range of temperature, a few tens of mK above the saturation temperature of the flowing helium II. Thermal buffering provided by the pressurized helium II baths contributed to limit temperature excursions, at the cost of introducing strong non-linearities and time delays in the system, which must be coped with by elaborate, robust process control [56]. In addition to that applied work, more fundamental experimental studies are presently conducted on a specially-instrumented, 90-m long test loop at CEN-Grenoble, comprehensively equipped with diagnostics and a transparent section for visual observation and interpretation of the flow patterns.

6. REFRIGERATION CYCLES AND EQUIPMENT

The properties of helium at saturation (Fig. 3) impose maintaining an absolute pressure below 1.6 kPa on the heat sink of a 1.8 K cryogenic system. Bringing the saturated vapour up to atmospheric pressure thus requires compression with a pressure ratio exceeding 80, i.e. four times that of refrigeration cycles for "normal" helium at 4.2 K. For low-power refrigeration, e.g. in small laboratory cryostats, this is achieved by means of standard Roots or rotary-vane vacuum pumps, handling the very-low pressure gaseous helium escaping from the bath after it has been warmed up to ambient temperature. This technology may be pushed to higher flow-rates using liquid-ring pumps, adapted for processing helium by improving the tightness of their casing and operating them with the same oil as that of the main compressors of the 4.2 K cycle [57], or oil-lubricated screw compressors operating at low suction pressure. In any case, compression at ambient temperature is hampered by the low density of the gaseous helium, which results in large volume flow-rates and thus requires large machinery, as well as in costly, inefficient heat exchangers for recovering enthalpy of the very-low pressure stream.

The alternative process is to perform compression of the vapour at low temperature, i.e. at its highest density. The pumps and recovery heat exchangers get smaller in size and less expensive, but the work of compression is then injected in the cycle at low temperature, so that the inevitable irreversibilities have a higher thermodynamic weight. Moreover, the pumping machinery which handles cold helium must be non-lubricated and non-contaminating, which seriously limits the choice of technology. Hydrodynamic compressors, of the centrifugal or axial-centrifugal type, have been used in large-capacity systems [58]. Their limited pressure ratio however imposes arranging them in multistage configurations [59, 60], thus narrowing the operational range of the system, in particular for startup or off-design modes, unless they are used in combination with more compliant volumetric machines, or in a scheme which permits independent adjustment of flow-rate or wheel inlet conditions [61, 62].

The practical ranges of application of these different pumping techniques appear in Fig. 7, setting a *de facto* limit for warm compression above 20'000 $m^3.h^{-1}$, or typically 300 W at 1.8 K.

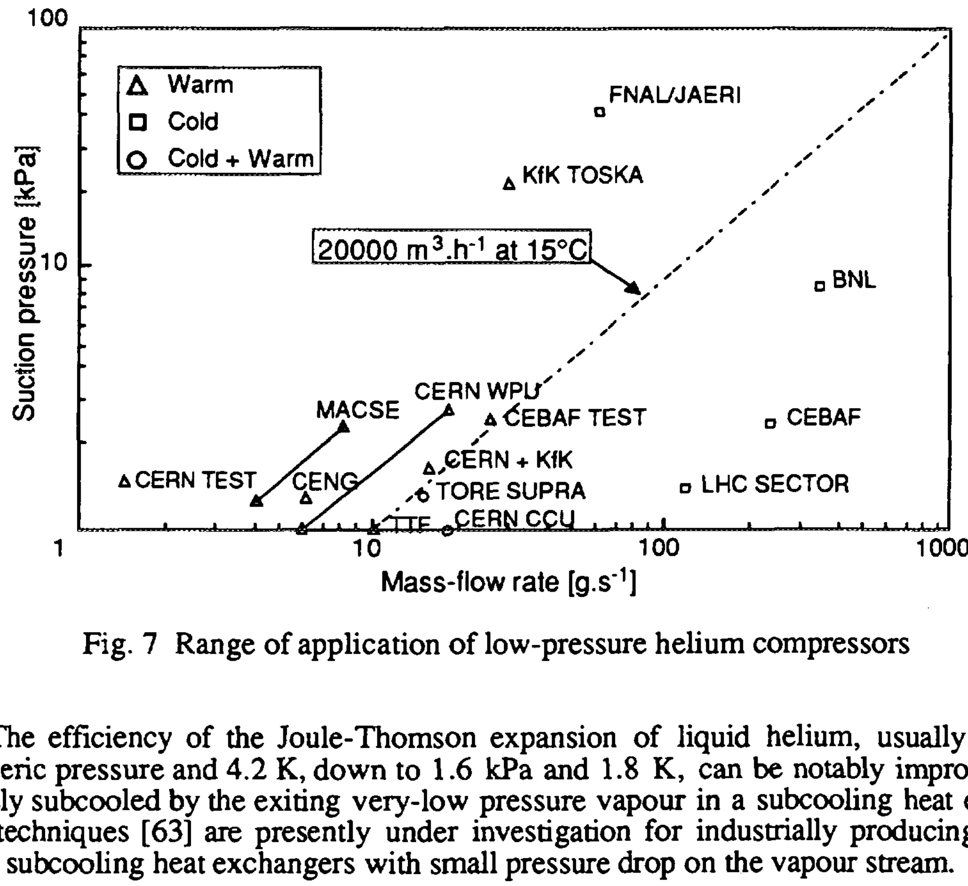


Fig. 7 Range of application of low-pressure helium compressors

The efficiency of the Joule-Thomson expansion of liquid helium, usually tapped at atmospheric pressure and 4.2 K, down to 1.6 kPa and 1.8 K, can be notably improved if it is previously subcooled by the exiting very-low pressure vapour in a subcooling heat exchanger. Several techniques [63] are presently under investigation for industrially producing efficient, compact subcooling heat exchangers with small pressure drop on the vapour stream.

7. SUPERFLUID HELIUM CRYOSTATS

The high thermodynamic cost of low-temperature refrigeration may only be rendered acceptable provided the heat loads at the 1.8 K level are tightly budgeted and contained. Setting aside the dynamic heat loads, produced by powering of the magnets and acceleration cavities, or interactions with the circulating particle beams, the base load in the cryogenic system of a large accelerator is the cryostat heat inleak, falling onto several kilometers (or tens of kilometers) of cold mass. It is therefore very important, in large projects, to design, construct and assemble reproducibly device cryostats with low residual heat inleak.

The structure of accelerator cryostats for devices operating in superfluid helium [64–67] (Fig. 8) reflects these preoccupations. The device cold mass is supported and precisely positioned from the ambient-temperature vacuum vessel by post-type supports, made of non-metallic composites, with several levels of heat interception at intermediate temperatures [68]. In this fashion the residual heat inleak to the 1.8 K level can be kept more than two orders of magnitude smaller than the heat drawn by conduction along the posts from the ambient-temperature environment. The thermal performance achieved in practice, however, critically depends on the quality of the heat intercepts, i.e. on the ability to realize good solid-to-solid thermal contacts under vacuum.

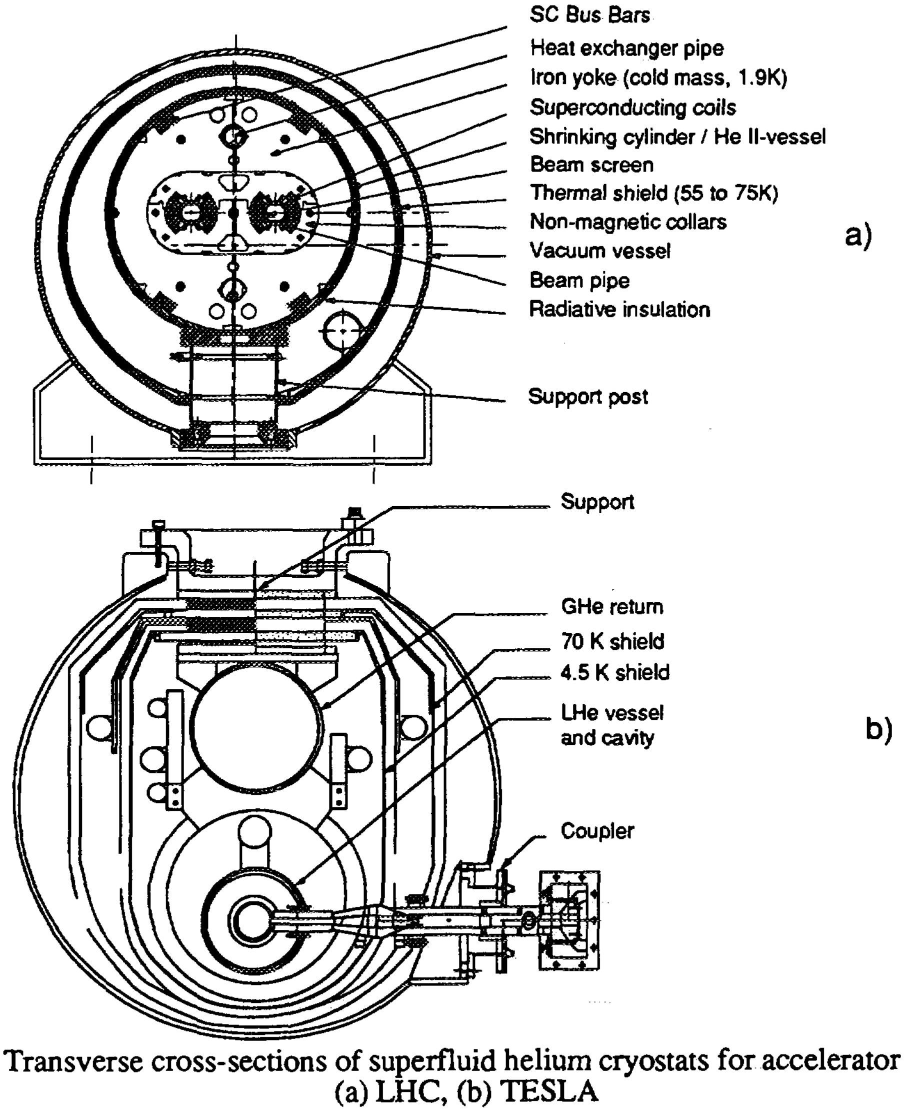


Fig. 8 Transverse cross-sections of superfluid helium cryostats for accelerator devices: (a) LHC, (b) TESLA

Radiation and residual gas conduction in the evacuated insulation space are limited by one or two nested actively cooled shields, wrapped with multilayer reflective insulation. Although single low-emissivity surfaces are in principle sufficient to reduce the radiative heat inleak below 80 K, the use of multilayer systems appears essential to provide robustness of performance against degradation of the insulation vacuum [69], a mishap bound to occur locally around the multi-kilometer circumference of a large accelerator. With proper industrial-scale implementation of such techniques, the distributed heat flux reaching the cold mass may then be kept well below 0.1 $W.m^{-2}$.

The experimental validation of cryostat performance requires precise calorimetric measurements, which may only be performed on well-instrumented, dedicated thermal models [70]. It is however satisfying to observe that industrially-produced cryostats reproducibly perform in accordance with budgeted heat loads [71]. In particular, the leaktightness requirements of superfluid helium enclosures can be met by the standard joining and testing technology employed for helium I cryostats.

8. CONCLUSION

Operating particle accelerators and colliders at 1.8 K, using superfluid helium as a technical coolant, has now become state of the art. The new projects, in construction or in study, nevertheless represent major challenges in view of their large size and quest for reliability and efficiency. This situation has triggered vigorous development programmes in several laboratories, as well as in the cryogenic industry, which have started to produce tangible results. There is no doubt that the coming years will see more progress in the technology of superfluid helium.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the help of R. Decreuse and L. Tavian in the preparation of the manuscript and illustrations.

REFERENCES

[1] G. Claudet & R. Aymar, Tore Supra and helium II cooling of large high-field magnets, Adv. Cryo. Eng. 35A (1990) 55-67

[2] S.W. van Sciver, He II cooling of large superconducting magnet systems, Cryogenics 32 ICEC Supplement (1992) 320-327

[3] H.J. Schneider-Muntau & J.C. Vallier, The Grenoble hybrid magnet, IEEE Trans. Mag. 24 (1988) 1067-1069

[4] J.R. Miller, M.D. Bird, S. Bole, A. Bonito-Oliva, Y. Eyssa, W.J. Kenney, T.A. Painter, H.J. Schneider-Muntau, L.T. Summers, S.W. van Sciver, S. Welton, R.J. Wood, J.E.C. Williams, E. Bobrov, Y. Iwasa, M. Leupold, V. Stejskal & R. Weggel, An overview of the 45-T hybrid magnet system for the new National High Magnetic Field Laboratory, IEEE Trans. Mag. 30 (1994) 1563-1571

[5] S.W. van Sciver, J.R. Miller, S. Welton, H.J. Schneider-Muntau & G.E. McIntosh, Cryogenic system for the 45 Tesla hybrid magnet, Adv. Cryo. Eng. 39A (1994) 375-380

[6] R. Aymar, G. Claudet, C. Deck, R. Duthil, P. Genevey, C. Leloup, J.C. Lottin, J. Parain, P. Seyfert, A. Torossian & B. Turck, Conceptual design of a superconducting tokamak: Tore II Supra, IEEE Trans. Mag. MAG 15-1 (1979) 542-545

[7] G. Claudet, G. Bon Mardion, B. Jager & G. Gistau, Design of the cryogenic system for the Tore Supra tokamak, Cryogenics 26 (1986) 443-449

[8] L.R. Evans, The Large Hadron Collider, paper presented at PAC'95, Dallas (1995)

[9] Ph. Lebrun, Superfluid helium cryogenics for the Large Hadron Collider project at CERN, Cryogenics 34 ICEC Supplement (1994) 1-8

[10] W. Weingarten, Basic properties of superconducting cavities, these proceedings

[11] C.H. Rode, CEBAF cryogenic system, paper presented at PAC'95, Dallas (1995)

[12] Ph. Leconte & M. Promé, ELFE, an electron laboratory for Europe, Proc. EPAC'94 1 (World Scientific, 1994) 181-183

[13] R. Brinkmann, Low-frequency linear colliders, Proc. EPAC'94 1 (World Scientific, 1994) 363-367

[14] G. Horlitz, T. Peterson & D. Trines, A 2 Kelvin helium II distributed cooling system for the 2 x 250 GeV $e^+ e^-$ linear collider Tesla, Cryogenics 34 ICEC Supplement (1994) 131-134

[15] F. Vinen, Physical properties of superfluid helium - a general review, Proc. Workshop on the stability of superconductors in helium I and helium II, IIF/IIR Bulletin, Commission A1/2, Saclay, 1981/6 (1981) 43-51

[16] F. Vinen, Superfluidity, these proceedings

[17] G. Bon Mardion, G. Claudet, P. Seyfert & J. Verdier, Helium II in low-temperature and superconductive magnet engineering, Adv. Cryo. Eng. 23 (1977) 358-362

[18] B. Fallou, J. Galand & B. Bouvier, Dielectric breakdown of gaseous helium at very low temperature, Cryogenics 10 (1970) 142

[19] H. Winkelnkemper, Z. Krasucki, J. Gerhold & T.W. Dakin, Breakdown of gases in uniform electric fields, Paschen curves for hydrogen, carbon dioxide and helium, Electra 52 (1977) 67

[20] B. Rousset & F. Viargues, An alternative cooling scheme for the TeV superconducting linear accelerator project, Cryogenics 34 ICEC Supplement (1994) 91-94

[21] V. Arp, Heat transport through helium II, Cryogenics 10 (1970) 96

[22] G. Bon Mardion, G. Claudet & P. Seyfert, Practical data on steady state heat transport in superfluid helium at atmospheric pressure, Cryogenics 19 (1979) 45-47

[23] S.W. van Sciver, Heat transfer in superfluid helium II, Proc. ICEC8 (IPC Science & Technology Press, 1980) 228-237

[24] P. Seyfert, Practical results on heat transfer to superfluid helium, Proc. Workshop on the stability of superconductors in helium I and helium II, IIF/IIR Bulletin, Commission A1/2, Saclay, 1981/6 (1981) 53-62

[25] S.W. van Sciver, Developments in He II heat transfer and applications to superconducting magnets, Adv. Cryo. Eng. 27 (1982) 375-398

[26] G. Claudet & P. Seyfert, Bath cooling with subcooled superfluid helium, Adv. Cryo. Eng. 27 (1982) 441-449

[27] P. Seyfert, Results on heat transfer to He II for use in superconducting magnet technology, Proc. ICEC9 (Butterworth, 1982) 263-268

[28] G. Claudet, F. Disdier, Ph. Lebrun, M. Morpurgo & P. Weymuth, Preliminary study of a superfluid helium cryogenic system for the Large Hadron Collider, Proc. ICFA Workshop on Superconducting Magnets and Cryogenics, BNL 52006 (1986) 270-275

[29] V. Benda, M. Granier, Ph. Lebrun, G. Novellini, V. Sergo, L. Tavian & B. Vullierme, Cryogenic benches for superfluid helium testing of full-scale prototype superconducting magnets fot the CERN LHC project, Cryogenics 34 ICEC Supplement (1994) 733-736

[30] H. Danielsson, Ph. Lebrun & J.M. Rieubland, Precision heat inleak measurements at 80 K, 4.2 K and 1.8 K, Cryogenics 32 ICEC Supplement (1992) 215-218

[31] S. Caspi, The use of calorimetry in superfluid He II to measure losses in superconducting magnets, Proc. ICEC9 (Butterworth, 1982) 347-350

[32] P.L. Walstrom, Joule Thomson effect and internal convection heat transfer in turbulent helium II flow, Cryogenics 28 (1988) 151-156

[33] G. Claudet, F. Disdier, A. Gauthier, Ph. Lebrun, M. Morpurgo & J. Schmid, Conceptual study of the superfluid helium cryogenic system for the CERN Large Hadron Collider, Proc. ICEC12 (Butterworth, 1988) 497-504

[34] A. Kashani & S.W. van Sciver, Steady-state forced convection heat transfer in He II, Adv. Cryo. Eng. 31 (1986) 489-498

[35] A. Kashani & S.W. van Sciver, Transient forced convection heat transfer in He II, Proc. ICEC11 (Butterworth, 1986) 654-658

[36] P.L. Walstrom & J.R. Maddocks, Pressure drop and temperature rise in He II flow in round tubes, venturi flowmeters and valves, Adv. Cryo. Eng. 33 (1988) 449-456

[37] P.L. Walstrom, J.G. Weisend II, J.R. Maddocks & S.W. van Sciver, Turbulent flow pressure drop in various helium II transfer system components, Cryogenics 28 (1988) 101-109

[38] B. Rousset, G. Claudet, A. Gauthier, P. Seyfert, Ph. Lebrun, M. Marquet, R. van Weelderen & J.L. Duchateau, Operation of a forced-flow superfluid helium test facility and first results, Cryogenics 32 ICEC Supplement (1992) 134-137

[39] B. Rousset, G. Claudet, A. Gauthier, P. Seyfert, A. Martinez, Ph. Lebrun, M. Marquet & R. van Weelderen, Pressure drop and transient heat transport in forced-flow single-phase helium II at high Reynolds numbers, Cryogenics 34 ICEC Supplement (1994) 317-320

[40] H.A. Snyder & A.J. Mord, Flow of superfluid helium in tubes with heated walls, Adv. Cryo. Eng. 37A (1991) 81-88

[41] A.J. Mord, H.A. Snyder & D.A. Newell, End-to-end modelling of helium II flow systems, Cryogenics 32 (1992) 291-299

[42] M. Morpurgo, Design and construction of a pump for liquid helium, Cryogenics 17 (1977) 91-93

[43] A. Hofmann, A. Khalil, H.P. Krämer, J.G. Weisend, R. Srinivasan & B. Vogeley, Investigations on fountain-effect pumps for circulating pressurized helium II, Proc. ICEC11 (Butterworth, 1986) 312-316

[44] S.W.K. Yuan & T.C. Nast, The design of fountain-effect pumps, Adv. Cryo. Eng. 33 (1988) 457-464

[45] P. Kittel, Operating characteristics of isocaloric fountain-effect pumps, Adv. Cryo. Eng. 33 (1988) 465-470

[46] A. Hofmann, A. Khalil & H.P. Krämer, Operational characteristics of loops with helium II flow driven by fountain-effect pumps, Adv. Cryo. Eng. 33 (1988) 471-478

[47] P. Kittel, Losses in fountain-effect pumps, Proc. ICEC11 (Butterworth, 1986) 317-322

[48] G. Klipping & H.D. Denner, The thermomechanical effect of superfluid helium and its technical application, Adv. Cryo. Eng. 35A (1990) 81-93

[49] A.J. Mord & H.A. Snyder, Self-driven He II cooling system for the interaction region focusing magnets at SSC, Adv. Cryo. Eng. 39A (1994) 797-804

[50] A. Hofmann, W. Herz, E. Suesser, B. Vogeley, T. Voekel & G. Zahn, Tests results of a large-size thermomechanical pump, Adv. Cryo. Eng. 39B (1994) 1813-1820

[51] A.J. Mord & H.A. Snyder, Comparison of He II cooling systems for superconducting magnets using self-driven mass flow and zero-flow heat pipes, Cryogenics 32 (1992) 461-465

[52] J. Casas-Cubillos, A. Cyvoct, Ph. Lebrun, M. Marquet, L. Tavian & R. van Weelderen, Design concept and first experimental validation of the superfluid helium system for the Large Hadron Collider (LHC) project at CERN, Cryogenics 32 ICEC Supplement (1992) 118-121

[53] A. Bézaguet, J. Casas-Cubillos, Ph. Lebrun, M. Marquet, L. Tavian & R. van Weelderen, The superfluid helium model cryoloop for the CERN Large Hadron Collider (LHC), Adv. Cryo. Eng. 39A (1994) 649-656

[54] M.M. Kado, Thermal conductance measurements on the LHC helium II heat exchanger pipes, CERN AT/95-34 (CR) (1995)

[55] A. Bézaguet, J. Casas-Cubillos, B. Flemsaeter, B. Gaillard-Grenadier, Th. Goiffon, H. Guinaudeau, Ph. Lebrun, M. Marquet, L. Serio, A. Suraci, L. Tavian & R. van Weelderen, The superfluid helium system for the LHC Test String: design, construction and first operation, paper presented at CEC'95 Columbus (1995)

[56] B. Flemsaeter, Contribution to the dynamic analysis and optimal control of the superfluid helium cooling loop for the LHC magnet string, Diploma thesis, Norwegian Institute of Technology, Trondheim (1995)

[57] G.M. Gistau & G. Claudet, The design of the helium refrigerator for Tore Supra, Proc. ICEC10 (Butterworth, 1984) 288-291

[58] G.M. Gistau, J.C. Villard & F. Turcat, Application range of cryogenic centrifugal compressors, Adv. Cryo. Eng. 35B (1990) 1031-1037

[59] G.M. Gistau, Y. Pecoud & A.E. Ravex, The 300 W - 1.75 K Tore Supra refrigerator cold centrifugal compressor report, Adv. Cryo. Eng. 33 (1988) 675-681

[60] C.H. Rode, D. Arenius, W.C. Chronis, D. Kashy & M. Keesee, 2 K CEBAF cryogenics, Adv. Cryo. Eng. 35A (1990) 275-286

[61] J.Ph. Guignard, Contribution à l'étude de la stabilité de fonctionnement et de l'adaptation de charge des compresseurs centrifuges cryogéniques multiétages, Travail d'option "Machines-Energétique", Ecole des Mines de Paris (1993)

[62] M. Kauschke, C. Haberstroh & H. Quack, Safe and efficient operation of multistage cold compressor systems, paper presented at CEC'95 Columbus (1995)

[63] F. Viargues, G. Claudet & P. Seyfert, Construction and preliminary testing of perforated-plate heat exchangers for use in helium II refrigerators, Cryogenics 34 ICEC Supplement (1994) 325-328

[64] M. Granier, Ph. Lebrun & M. Mischiatti, Design and construction of a superfluid helium cryostat for a ten-meter long high-field superconducting dipole magnet, Cryogenics 30 September Supplement (1990) 98-102

[65] J.C. Brunet, J. Kerby, Ph. Lebrun, P. Rohmig, B. Szeless & L.R. Williams, Design of LHC prototype dipole cryostats, Cryogenics 32 ICEC Supplement (1992) 191-194

[66] F. Alessandria, G. Cavallari, M. Minestrini, T.H. Nicol, C. Pagani, R. Palmieri, S. Tazzari & G. Varisco, Design, manufacture and test of the TESLA-TTF cryostat, paper presented at CEC'95 Columbus (1995)

[67] G. Horlitz, T. Peterson & D. Trines, The TESLA 500 cryogenic system layout, paper presented at CEC'95 Columbus (1995)

[68] M. Blin, H. Danielsson, B. Evans & M. Mathieu, Design, construction and performance of magnet support posts for the Large Hadron Collider, Adv. Cryo. Eng. 39A (1994) 671-677

[69] Ph. Lebrun, L. Mazzone, V. Sergo and B. Vullierme, Investigation and qualification of thermal insulation systems between 80 K and 4.2 K, Cryogenics 32 ICMC Supplement (1992) 44-47

[70] L. Dufay, G. Ferlin, Ph. Lebrun, G. Riddone, J.M. Rieubland, A. Rijllart, B. Szeless & L. Williams, A full-scale thermal model of a prototype dipole cryomagnet for the CERN LHC project, Cryogenics 34 ICEC Supplement (1994) 693-696

[71] V. Benda, L. Dufay, G. Ferlin, Ph. Lebrun, J.M. Rieubland, G. Riddone, B. Szeless, L. Tavian & L.R. Williams, Measurement and analysis of thermal performance of LHC prototype dipole cryostats, paper presented at CEC'95 Columbus (1995)

SELECTED BIBLIOGRAPHY

F. London, Superfluids, Volume II, Macroscopic theory of superfluid helium (John Wiley & Sons, 1954, republished by Dover, New-York, 1964)

J. Wilks, The properties of liquid and solid helium (Clarendon Press, Oxford, 1967)

S.W. van Sciver, Helium cryogenics (Plenum Press, New-York, 1986)

J. Wilks & D.S. Betts, An introduction to liquid helium (Clarendon Press, Oxford, 1987)

Proceedings of the CERN Accelerator School "Superconductivity in particle accelerators" (CERN Report 89-04, 1989)

Proceedings of the symposium "A half century of superfluid helium", Adv. Cryo. Eng. 35A (1990) 1-93

IMPACT OF SUPERCONDUCTORS ON LHC DESIGN

J. Gareyte
CERN, Geneva, Switzerland

Abstract
The LHC is a high field, high luminosity hadron collider. The desire to exploit the existing LEP tunnel and infrastructure in the most cost effective way while reaching the highest possible energy led to the choice of a high field superconducting technology, using NbTi conductors in superfluid helium. This has important consequences for the design and performance of the machine. On the one hand it helps reach the desired large value of the luminosity with the minimum number of circulating particles. On the other hand it makes the machine very sensitive to particle losses and other beam effects. The unavoidable magnetic field errors of superconducting magnets have also a strong influence on the machine performance and must be carefully compensated.

1 INTRODUCTION

The vast majority of large hadron colliders proposed or designed in the last two decades use superconducting magnets. The obvious reason is that of cost. It has been shown that the price per Tm of superconducting magnets is much lower than that of classical ones. In addition, the power bill during operation is reduced, and if one considers the problems linked to the size of the site and of the infrastructure, the reasons to choose a superconducting system are compelling. In the case of the Large Hadron Collider (LHC) the tunnel and the general infrastructure exist already. A strong incentive to reach the highest possible energy on the available CERN site led the designers to use even higher fields than in other projects, thus pushing the technology to its limit. The two-in-one design, in which the superconducting coils around the two beam channels are embedded in the same iron yoke and cryostat, was adopted because it is particularly well-suited to a proton–proton collider in this high-field regime. It reduces significantly the material cost, the tunnel occupancy, and the installation cost.

The only cost-effective way of reaching magnetic fields in the 9 T range is to use the well-proven NbTi conductors but cooled at superfluid helium temperatures. The choice of this high-field superconducting technology for the LHC has consequences of overwhelming importance for many aspects of the design and for the performance of the machine. Some of these consequences follow directly from the properties of the superconductors and from the way the magnets are built. This is the case of the magnetic errors which are much larger in a superconducting machine than in a classical one. Others come indirectly, for example through the very low temperature of the beam enclosure.

On the one hand, the high-field technology helps reach the high luminosity required for physics in the TeV energy range by allowing strong focusing of the beams in the experimental insertions, by reducing the amount of particles necessary, and by making the beams less prone to collective instabilities. On the other hand, it makes the machine extremely sensitive to beam losses and to synchrotron radiation.

After recalling the general parameters of the LHC, we shall consider two very important properties of very high energy hadron colliders, namely luminosity and beam stability, study their scaling with energy and the machine size, and examine the impact of the high-field technology on these properties. Then we shall discuss particular problems directly or indirectly linked to the choice of superconducting magnets, and the solutions proposed in the LHC design to alleviate them. This concerns vacuum, protection against synchrotron radiation, and beam losses. The most important part of the report concerns the effect of magnetic errors which are inherent to superconducting magnets, and the ways to minimize their effects. Finally, we shall briefly discuss the interest of a superconducting RF system in the LHC.

2 LHC PARAMETERS

The LHC is a high-field, high-luminosity proton–proton collider [1], which will also be able to collide heavy ions, for instance lead against lead. At a later stage it will be possible to arrange collisions of protons in LHC against electrons in LEP. In this report we restrict our considerations to the LHC as a high-luminosity proton–proton collider, which is by far the most demanding case.

In our desire to use the highest possible magnetic field, an ambitious goal was set at the beginning of the Research and Development programme, aiming at a maximum field of the order of 10 T. In fact, 10.5 T were reached in 1-m long magnet models, but the programme showed that difficulties and cost increased considerably above say 9.5 T. The LHC design is now based on a maximum 'short-sample limit' field of 9.5 T, with an operating field of 8.4 T. Figure 1 shows the cross-section of the magnet.

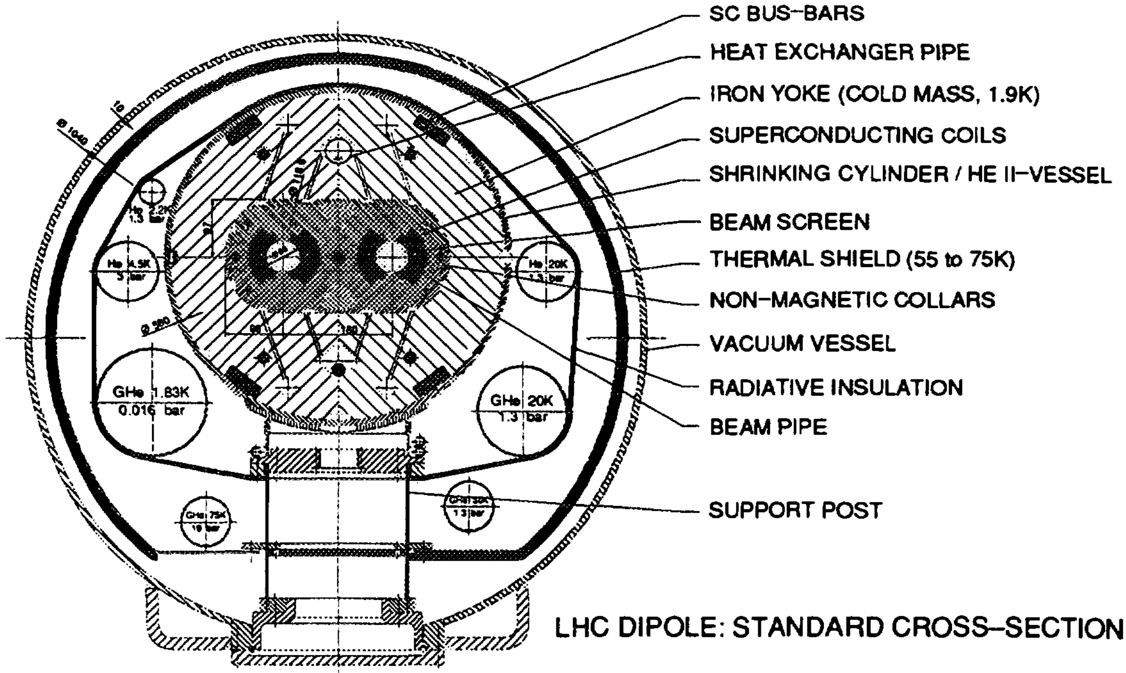


Fig. 1 Dipole magnet cross-section

The luminosity is limited by combined long-range and head-on beam–beam effects to about 10^{34} cm^{-2} s^{-1} for two experiments operating simultaneously. The beam current is about 0.5 A distributed among 2835 bunches of 10^{11} protons each.

Tables 1 and 2 display the main parameters of the machine, while Fig. 2 shows the overall structure of the LHC with the utilization of the eight straight sections.

Table 1
LHC performance parameters

Energy	E	[TeV]	7.0
Dipole field	B	[T]	8.4
Luminosity	L	[$cm^{-2} s^{-1}$]	10^{34}
Beam–beam parameter	ξ		0.0034
Total beam–beam tune spread			0.01
Injection energy	E_i	[GeV]	450
Circulating current/beam	I_{beam}	[A]	0.53
Number of bunches	k_b		2835
Harmonic	h_{rf}		35640
Bunch spacing	τ_b	[ns]	24.95
Particles per bunch	n_b		$1.05\ 10^{11}$
Stored beam energy	E_s	[MJ]	332
Normalized transverse emittance $(\beta\gamma)\sigma^2/\beta$	ε_n	[mrad]	$3.75\ 10^{-6}$
Collisions			
Beta-value at I.P.	β^*	[m]	0.5
r.m.s. beam radius at I.P.	σ^*	[µm]	16
r.m.s. divergence at I.P.	σ'^*	[µrad]	32
Luminosity per bunch collision	L_b	[$cm^{-2} s^{-1}$]	$3.2\ 10^{26}$
Crossing angle	ϕ	[µrad]	200
Number of events per crossing	n_c		19
Beam lifetime	τ_{beam}	[h]	22
Luminosity lifetime	τ_L	[h]	10

Table 2
LHC parameters related to RF

			Injection	Collision
Intrabeam scattering				
Horinzontal growth time	τ_h	[h]	45	100
Longitudinal growth time	τ_p	[h]	33	60
Radiofrequency				
RF voltage	V_{rf}	[MV]	8	16
Synchrotron tune	Q_s		$5.5\ 10^{-3}$	$1.9\ 10^{-3}$
Bunch area (2s)	A_b	[eV.s]	1	2.5
Bucket area	A_{rf}	[eV.s]	1.46	8.7
Bucket half-height	$\Delta p/p$		$1\ 10^{-3}$	$3.6\ 10^{-4}$
r.m.s. bunch length	σ_s	[m]	0.13	0.075
r.m.s. energy spread	σ_e		$4.5\ 10^{-4}$	$1.0\ 10^{-4}$

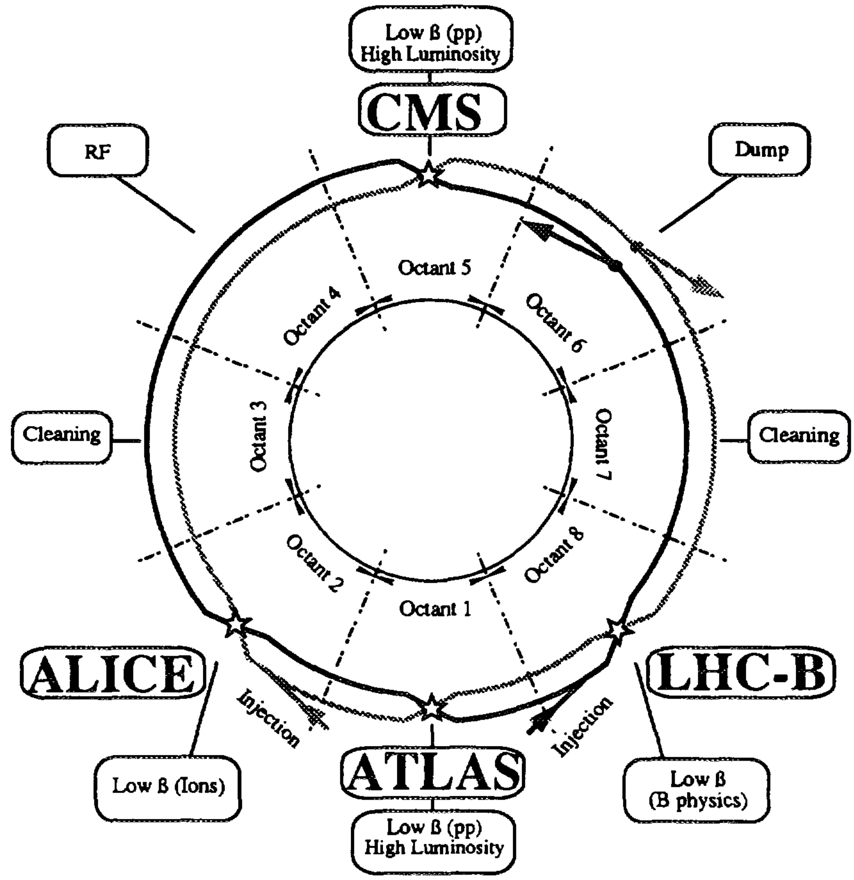


Fig. 2 Schematic layout of LHC

3 SCALING LUMINOSITY WITH ENERGY

Cross-sections for hard processes decrease like γ^{-2} when the energy is increased and therefore, ideally, one would like the collider luminosity to scale like γ^2. The luminosity can be written

$$L = \frac{1}{4\pi}\left(\frac{N}{\varepsilon_n}\right)(Nkf)\frac{\gamma}{\beta^*}, \qquad (1)$$

where N is the number of particles per bunch, k the number of bunches, ε_n the invariant transverse emittance, β^* the value of the betatron function at the collision point, γ the energy of the particles divided by their rest energy and f the revolution frequency. The quantity in the first bracket is proportional to the head-on beam–beam coefficient, and it has a fixed maximum value. The quantity in the second bracket is proportional to the beam current.

Suppose that we want to build a proton–proton collider with an energy about 10 times higher than previously available, and that we want to compare a classical machine to a high-field superconducting one. The ratio of the bending fields in the two cases, taking as examples the LHC and the SPS, is 8.4/1.8 that is about 5. This means that a classical machine of the same energy would be roughly 5 times larger than the LHC, and would require 5 times more protons to provide the same beam current. This is much more demanding for the beam injection and dumping systems.

In addition, with a classical technology the value of β^* cannot be preserved when γ is increased. For this to happen the gradient of the final-focus quadrupoles would have to scale like γ. But since the maximum field at the pole tip is fixed, the gradient can only scale at best like $\gamma^{1/2}$ since the beam dimensions, and hence the required quadrupole aperture, scale like $\gamma^{-1/2}$. On the contrary, the high-field technology allows us to preserve the same value of β^* while increasing the energy. As an illustration, β^* in the LHC at 7 TeV is the same as in the SPS at 315 GeV. Formula (1) shows that it is not easy to increase luminosity like γ^2 as is required by physics, but the above considerations show that it is even more difficult to do so in a classical machine than in a high-field superconducting one.

4 THE RESISTIVE-WALL INSTABILITY

We have seen that in order to reach large luminosities in high-energy colliders one is bound to increase the beam current. This can be limited by many effects like collective beam instabilities, heating of the beam enclosure by induced wall currents or synchrotron radiation, or quenching the magnet system through losses of a tiny fraction of the beam particles. In addition the difficulty of disposing safely of the beam at the end of operation or in case of malfunction increases with the number of particles in the beam. In this section we are going to examine the impact of the choice of superconducting technology on one of the beam collective effects, the resistive-wall instability.

In large machines the transverse resistive-wall instability is dominant and has to be damped by active feedback. This is manageable if the growth rate of the instability is small enough so that the damper can act over several turns. The relevant parameter is thus the growth time τ divided by the revolution period T. This is given by:

$$\frac{\tau}{T} = \frac{2E/e}{\beta Z_T I}, \qquad (2)$$

where E/e is the energy in electronvolts, I the total beam current and β the betatron function. The transverse impedance Z_T is given by:

$$Z_T = \frac{2c}{b^3}\left(\frac{\rho}{\delta}\right)\frac{R}{\omega}, \qquad (3)$$

where c is the speed of light, b the beam chamber half-height, ρ the resistivity of the chamber material and δ the skin depth at the lowest betatron frequency ω, and R the average machine radius. Since ρ/δ scales like $\rho^{1/2} R^{-1/2}$, Z_T scales like $\rho^{1/2} R^{1.5}/b^3$. In addition, comparison of existing machine designs shows that β has a tendency to increase like $R^{1/2}$, so that τ/T scales like $b^3/\rho^{1/2} R^2$.

This reveals the problem of resistive-wall instability in high-energy colliders: since the beam shrinks at high energy, one is tempted to decrease b in order to reduce cost. This can be compensated only by limiting the machine radius R, which is possible with high-field technology, and reducing ρ by coating the beam chamber with a pure metal like copper and cooling it at cryogenic temperatures, which comes naturally in a superconducting machine.

As an illustration the value of τ/T is about 300 in the LHC with a copper-coated beam screen cooled at 20 K, and this gives ample time for the damper to suppress any disturbance before it grows as a result of the instability. With a copper tube of 2 mm thickness at room temperature, τ/T would decrease to about 100 and with a stainless-steel tube it would decrease down to 3. In this condition the LHC would not work at the required high luminosity.

The above considerations show that a classical collider of the same energy as the LHC would have difficulty in delivering the same luminosity: applying the scaling derived above one sees that τ/T is down to about 5 in this case due to the increased radius, even using a thick copper vacuum chamber at room temperature.

5 VACUUM

We have already alluded to the fact that high-energy colliders are bound to have beam pipes of smaller and smaller diameter. Evacuating these pipes down to the very low pressures needed for a proton collider requires a powerful distributed pumping system. In a superconducting machine this is again achieved naturally through cryopumping on the cold pipe.

However, in the LHC, and for the first time, a considerable complication arises due to synchrotron radiation. The total power radiated by each beam is

$$P_s = \frac{Z_0 e^2 c}{3\rho} \gamma^4 kNf , \tag{4}$$

where Z_0 is the impedance of free space and ρ is here the bending radius. It amounts to 3.7 kW in the LHC. We see that a superconducting machine is not an advantage in this respect since the bending radius appears in the denominator.

A flux of about 10^{17} photons per second per metre impinges on the beam pipe, with a characteristic energy of 46 eV. These photons desorb molecules trapped in the bulk of the material. These molecules are subsequently cryosorbed to the wall surface where they are continuously bombarded by photons and ions accelerated by the beam potential. When the surface coverage increases, the equilibrium pressure in the beam pipe runs away, as can be seen on Fig. 3. It is proposed to solve this problem in the LHC by separating the functions between a beam screen surrounding the beam and taking the bulk of the synchrotron radiation and an outside vacuum chamber at 1.9 K playing the role of the cryopump and protected from the beam effects. The beam screen has a large number of small holes to let the molecules released by synchrotron radiation escape to the annular space where they will be definitely cryopumped on the vacuum chamber (Fig. 4).

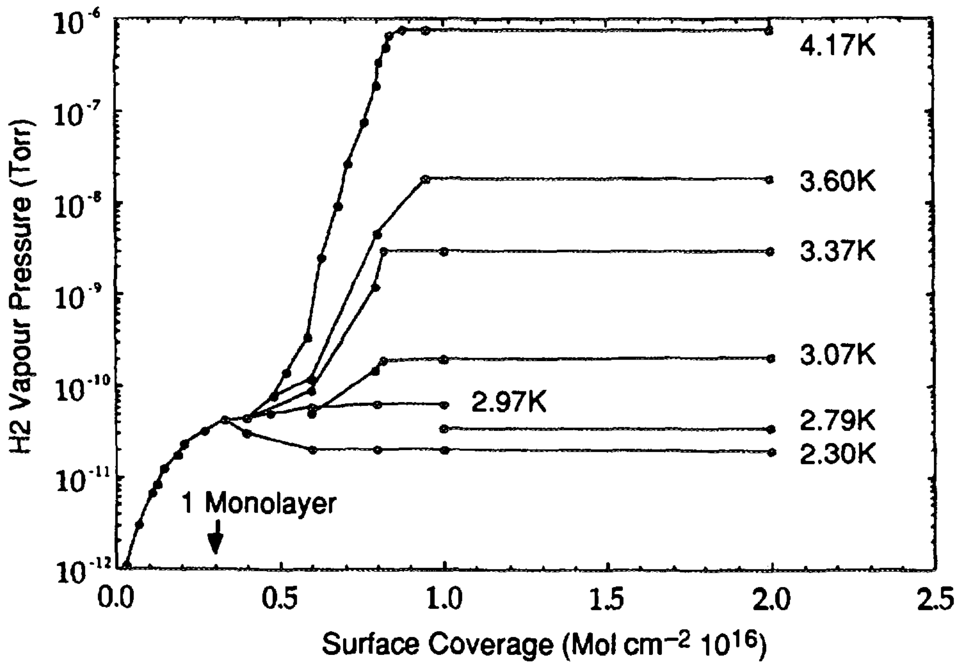


Fig. 3 Adsorption isotherms for H_2 as a function of temperature

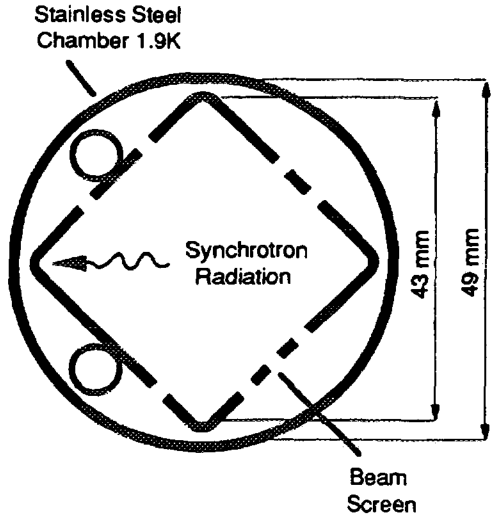


Fig. 4 Beam screen and vacuum chamber

This is a costly and delicate system, but it has the additional advantage of reducing the impact of the cryogenic load due to synchrotron radiation and to the heating produced by beam-induced wall currents. It is sufficient to cool the beam screen at temperatures of the order of 20 K to provide the small coupling impedance which is required for beam stability, and taking the cryogenic load at 20 K is much less costly than at 1.9 K.

6 EFFECT OF BEAM LOSSES

In any collider one tries to minimize the sources of particle loss other than the collisions themselves, in order to maximize beam lifetime and reduce spurious counts in the detectors. In superconducting colliders, one must in addition avoid quenching the magnets. This is not so difficult in the Tevatron or HERA which have large apertures and relatively small beam intensities. It is much more critical in the LHC where the aperture is tight and the energy stored in the beam is enormous (about 500 MJ at collision energy).

The LHC is also special in that it uses superfluid helium. On the one hand this is an advantage because of the very good contact which can in principle be ensured between the conductors and the helium bath, and the large thermal conductivity and good heat capacity of superfluid helium. On the other hand, since one is forced to operate close to the λ point, the temperature margin is reduced compared to machines operating with normal helium. One has to make sure that the cable insulation allows the conductor to be in direct contact with helium, and that this contact persists even when the conductor is heated by particle showers resulting from beam losses. Studies are being pursued to better understand these processes and determine what are the tolerable losses. It seems at present that we should limit particle losses in the superconducting part of the machine to 1.4×10^6 p/s per metre at collision energy. This is an extremely small value compared to the 2.8×10^{14} particles circulating in each LHC beam. In fact, a proper operation of the LHC requires a single beam lifetime of the order of at least 50 hours. In this situation about 1.6×10^9 p are lost per second and could strike the vacuum chamber at any place where the orbit is particularly distorted. Therefore the ratio between the estimated loss rate and the loss rate tolerable in the magnet system is about 1000, which gives an idea of the unprecedented efficiency required from the beam halo cleaning system.

Two out of the eight straight sections of LHC will be devoted to halo cleaning: one for betatron cleaning and one for momentum cleaning. This last one is thought to be necessary in particular to take care of the few per cent of particles which escape RF capture at injection and drift towards the inside at the start of the ramp instead of being accelerated.

Both insertions are made up of warm, classical dipoles and quadrupoles and cover the whole 500 m of the straight section. In this way a sufficient phase advance is available to locate the primary and secondary collimators at strategic places to optimize efficiency. Computer simulations show that this system is capable of the performance required.

7 MAGNETIC ERRORS

The Tevatron and HERA were pioneers in investigating the distortions of the magnetic field inherent to superconducting magnets. In this respect the LHC magnets are not very different from their predecessors. Of course they have a higher field, a smaller bore, and a two-in-one geometry which introduces additional effects. But by and large the very detailed measurements made at HERA can be used with proper scaling to predict the magnetic errors in the LHC. The non-linear errors are measured in terms of the high-order coefficients a_n and b_n of the complex field expansion

$$B_y + iB_x = B_1 \sum_n (b_n + ia_n)\left(\frac{Z}{R_r}\right)^{n-1} , \quad (5)$$

where B_1 is the nominal vertical magnetic field, B_y and B_x are the actual components of the field in the vertical and horizontal planes, R_r = 1 cm is the reference radius, and $Z = x + iy$. The values of the coefficients depend on the design and the size of the magnet coils. They are shown in Table 3.

Table 3

Expected multipole performance, at injection and at 8.40 T (In units of 10^{-4} relative field error at 10 mm)

	At injection, 0.58 T				At nominal operation, 8.40 T			
	Systematic		Random		Systematic		Random	
	norm.	skew	norm.	skew	norm.	skew	norm.	skew
n	b_n	a_n	$\sigma[b_n]$	$\sigma[a_n]$	b_n	a_n	$\sigma[b_n]$	$[a_n]$
2	±0.7	±0.7	0.4	1.0	1.4±0.7	±0.7	0.4	1.0
3	−3.4±0.3*	±0.3	0.5	0.15	1.0±0.3*	±0.3	0.6	0.15
4	±0.2	±0.2	0.1	0.1	±0.2	±0.2	0.1	0.1
5	0.25±0.05*	±0.05	0.08	0.04	0.06±0.05*	±0.05	0.05	0.04
6	−0.004	0.0	0.02	0.01	−0.005	0.0	0.006	0.01
7	−0.026	0.0	0.01	0.01	0.006	0.0	0.009	0.003
8	0.0	0.0	0.005	0.005	0.0	0.0	0.001	0.002
9	0.006	0.0	0.003	0.004	−0.001	0.0	0.001	0.001
10	0.0	0.0	0.002	0.002	0.0	0.0	0.0	0.0
11	0.008	0.0	0.001	0.001	0.008	0.0	0.0	0.0

* Systematic b_3 and b_5 will be compensated by correctors at each dipole end. The b_3 values in this table indicate the magnitude of the persistent current effect at injection and of the yoke saturation at operational conditions, the coil geometry being designed for $b_3 = 0$.

The errors have both systematic and random components. The systematic ones, which are the same in all magnets of a production line, are mainly due to persistent currents. Their magnitudes are calculable but their time dependence is not well understood and must be measured. The random ones are mainly due to manufacturing tolerances and vary from magnet to magnet. These two types of errors produce different kinds of effects, and are treated separately.

The measurements made up to now on the few available LHC magnet prototypes largely confirm the validity of Table 3. The main difference between the LHC and its predecessors lies in the much enhanced sensitivity to errors of the LHC due to its larger size. This point will be illustrated in the section on coupling effects.

8 SYSTEMATIC ERRORS

Persistent currents in the superconducting filaments produce large sextupole and decapole errors in the dipole magnets and dodecapole errors in the quadrupoles, mainly at injection energy. Their most detrimental effect is to render the tune of particles dependent on their betatron amplitude and momentum deviation. This creates a tune spread in the beam with the consequence that relatively low-order betatron resonances (of order 5 to 7) can no longer be avoided and produce beam losses. In the LHC, these effects were minimized by reducing the diameter of the NbTi filaments to 6 μm, but this is not enough to reduce sufficiently the errors in the long dipoles and special correctors have to be used in addition. Studies have shown that a few strong correctors located close to the lattice quadrupoles would not be adequate: the correction must be located closer to the sources of errors to avoid the build-up of high-order interferences. In a previous version of the LHC lattice which had four dipole bending magnets per half-cell, a very efficient scheme featured sextupole and decapole correctors in between pairs of dipoles in the middle of each half-cell. In the final version, which has longer dipoles but only three per half-cell, this solution is no longer practical, and we have instead introduced small 10 cm long correcting coils, one sextupole and one decapole, at each end of the magnets. Figure 5 shows the effectiveness of the correction scheme in reducing the tune dependence on betatron amplitude and momentum.

9 RANDOM ERRORS

The random errors essentially excite betatron resonances. Low-order resonances lead usually to a rapid loss of particles, and have to be avoided. High-order resonances fill the tune plane and cannot be avoided. Their strengths depend on the magnitude of the errors and on their particular distribution around the machine, and increase very quickly with the initial amplitude of oscillation of the particles. They conspire to induce chaotic motion leading eventually to slow beam losses above a certain amplitude, which we call the dynamic aperture. The dynamic aperture, which is of overwhelming importance for the LHC, is not calculable analytically but has to be evaluated through computer simulation. For that, a set of particles with different initial amplitudes is tracked element by element around the machine for a large number of turns. Usually about 10^4 turns are sufficient to detect whether the particle motion is regular or chaotic. Chaotic particles will eventually be lost on the wall, but this may take a long time for weakly-chaotic particles close to the dynamic-aperture limit. For some selected sets of parameters the tracking is pursued up to 10^5 or 10^6 turns, which is the maximum possible in practice with currently available computers. From this one obtains 'survival plots' showing the variation of the survival time as a function of the initial amplitude of the particles in the chaotic region. These plots can be extrapolated to evaluate the maximum amplitude of particles which are likely to survive for 10^7 or 10^8 turns as needed in the real machine. In the majority of cases this amplitude is only slightly above the chaotic boundary, which can therefore be taken as a safe value for the long-term dynamic aperture.

This evaluation has to be made for a large number of realizations of the random errors, which we call 'random seeds'.

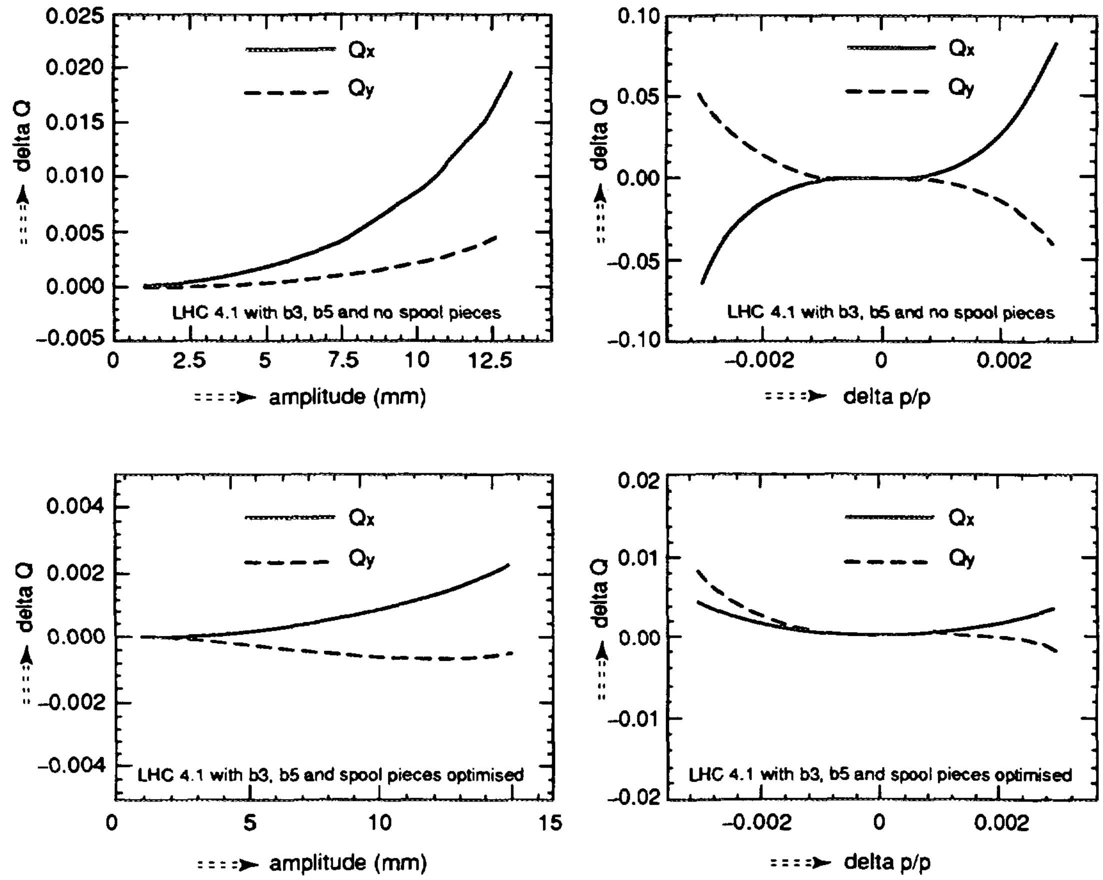


Fig. 5 Tune shift with amplitude (left) and momentum (right) without (up) and with (down) correction of systematic b_3 and b_5 using end magnet coils. Note the change of vertical scales.

The dynamic aperture can be expressed in units of the r.m.s. beam size σ. We believe that a dynamic aperture of at least 6 to 7 σ is necessary for a reliable operation of the machine. The limitation occurs at injection where the non-linearities are the largest due to persistent currents and where the beam has a large emittance. With the same coil design it is possible to reduce the high-order coefficients by increasing the coil diameter. By doing so while keeping the value of the field on the inner coil surface constant, one reduces the coefficients a_n and b_n by the enlargement factor to the power $n - 1$. The initial design of the LHC had a coil diameter of 50 mm. This provided an insufficient dynamic aperture and the coil had to be enlarged to 56 mm, the value of the final design. Also dynamic-aperture studies showed that high-order systematic errors, b_7 and b_9 were too large in the initial magnet design. A detailed study of the conductor's placement allowed a considerable reduction in the value of these coefficients.

10 COUPLING EFFECTS

The a_2 term in the field expansion (5) corresponds to a skew quadrupole (a normal quadrupole rotated by 45°) and provides the lowest order, linear coupling between horizontal and vertical betatron motion. It is absolutely crucial in a hadron collider to correct the effect of a_2 in order to reduce coupling effects, for the following reason. Inspection of a resonance diagram in the tune plane shows that the only areas free of resonances are close to the diagonal, where the vertical and horizontal tunes are almost equal modulo an integer. Coupling terms create a band around the diagonal where it is impossible to distinguish between horizontal and vertical tunes (the eigenmodes are at 45°) and therefore it is impossible to adjust the machine parameters. Since we have to operate very close to the diagonal to avoid non-linear resonances, we must reduce the width of the linear coupling resonance on the diagonal. This is done by placing skew-quadrupole correctors at strategic places in the lattice.

In classical machines coupling is induced by unwanted small rotations of the main quadrupoles, experimental solenoids, and vertical orbit offsets in the chromaticity sextupoles. In the LHC all these effects are present but are dominated by the a_2 term in the field of the main dipoles. This term is generated by an up–down asymmetry of the coils, which it is difficult to reduce below a few hundredths of a millimetre on account of mechanical tolerances.

There is no reason why a_2 should be much different in LHC magnets than in HERA magnets. However, the stop band width, which reflects the sensitivity of the machine to this coupling term, increases linearly with the beta function and also with the machine radius R, because local effects accumulate along the circumference.

Since beta functions tend to increase like the square root of R, the stop-band width increases like $R^{1.5}$, which makes the LHC almost ten times more sensitive than HERA. As a consequence the correction must be extremely precise. It is possible in the LHC to achieve the required precision by installing two correcting skew quadrupoles on each side of the eight straight sections. However, these localized correctors are strong and can themselves produce unwanted side-effects like beta beating which have to be minimized by a suitable choice of the lattice parameters. In addition, the currents in the correctors have to follow the changes in a_2 during the ramp with a precision of the order of 1% which is a real challenge. To alleviate these problems we are looking for ways to reduce a_2 and at the same time we are investigating lattices in which the effect of a_2 would partially compensate itself along the circumference.

11 TIME VARIATION OF MAGNETIC ERRORS

After studies in the Tevatron and then in HERA we know that the persistent currents that flow on either side of superconducting filaments and produce strong sextupole and decapole errors in the dipole magnets decay with characteristic times of about an hour. This has to be taken into account because injecting the beams into the LHC takes seven minutes. In addition we know from experience with previous colliders that the starting time of the injection process with respect to the initialization of the magnets varies from fill to fill owing to various irreproducible events which occur during the procedure.

Since the errors due to persistent currents are supposed to be the same for all magnets of a production line, we shall place measuring coils in reference magnets as was done in HERA and use the information to drive the small correcting lenses in the magnet ends as a first line of attack. Of course there should be one reference magnet and correction system for each ensemble of identical magnets. The last fine-tuning could be made using measurements with the beam. The LHC is particularly sensitive to drifts in the chromaticity $Q' = \Delta Q/(\Delta P/P)$ which is influenced by sextupoles because of its large bunch line density and the large momentum spread $\Delta P/P = 10^{-3}$ at injection. Chromaticity should always remain positive to prevent dense bunches from being lost owing to the transverse head-tail instability. In addition the absolute value of Q′ should also remain small to keep the momentum spread in the beam $\Delta Q = Q' \Delta P/P$ small.

The persistent currents which decay slowly during injection are affected by a rapid change during the early phase of the acceleration when the magnetic field starts increasing.

This is understandable since the *raison d'être* of these currents is to screen the inside of the conductors from the applied magnetic field. This effect should also be corrected by the same means as during injection. But there is another effect which happens during the ramp and which is of great concern in the LHC. It is due to the finite contact resistance between the different strands in the superconducting cable. This allows current loops to develop as soon as the field rises, and these generate magnetic errors which can vary from magnet to magnet, and which are proportional to the ramp rate. To alleviate these effects we carefully optimize the interstrand resistance (it cannot be made too large because it plays a role in balancing the sharing of the current between the strands) and we plan to start the ramp by a gentle progressive transition. The coupling term a_2 is strongly affected by these variations, and we are looking for operational methods to correct the residual coupling online using beam measurements as a last resort.

12 SUPERCONDUCTING RF SYSTEM

The RF system of the LHC is modest in comparison, for instance, with that of LEP. The power lost by the beam through synchrotron radiation is very small and the acceleration from injection to top energy is slow because the superconducting magnets cannot be ramped faster. Therefore the RF voltage is determined primarily by the surface of the longitudinal phase space area required to ensure the long-term stability of the dense LHC bunches. For this a maximum voltage of 8 MV at injection and 16 MV at top energy is sufficient, and it could be provided by classical copper cavities.

However, it is interesting in this domain as well to use superconducting technology. First of all, the necessary cryogenic environment exists already in the LHC. But more importantly, superconducting cavities do not need to be optimized for maximum voltage, contrary to copper cavities, because of their enormous quality factor. A large diameter can be ascribed to the beam pipe, leading to smaller values of R/Q for the fundamental mode as well as for the parasitic high-order modes. As a consequence, a large electromagnetic energy is stored in the cavities to provide the necessary voltage, and this helps in fighting transient beam loading effects. These are important in the LHC because of the large value of the average beam current. They appear during injection each time a new batch is added to the already circulating ones, and also in steady state because the beam is not continuous: there are holes to accommodate the risetimes of the injection and abort kickers.

REFERENCES

[1] The Large Hadron Collider Accelerator Project, CERN/AC/95–05

IMPACT OF SUPERCONDUCTING CAVITIES ON LEP2 DESIGN

D. Boussard
CERN, Geneva, Switzerland

Abstract
The Large Electron Positron (LEP) collider at CERN is being upgraded in energy to about 90 GeV per beam (above the W-pair-production threshold). This requires very large RF voltage (more than 2000 MV) to reach the design energy. In addition, the machine transverse broadband impedance must be kept as small as possible to maximize single-bunch currents and hence luminosity. Superconducting cavities are best suited, if not the only choice, for fulfilling these requirements. After a description of the LEP2 cavities and couplers and their technical difficulties, the problem of stability of the RF system with beam is addressed.

1. THE LEP MACHINE — ENERGY AND LUMINOSITY

LEP, the largest particle accelerator in the world, is an electron-positron collider located close to Geneva (Switzerland). Its circumference, almost 27 km long, straddles the Swiss-French border between Lake Geneva and the Jura mountains. To minimize the risks associated with the difficult geological areas in the Jura limestone, the plane of the machine is not horizontal, but slightly tilted with a maximum slope of 1.42%. Table 1 shows the major LEP parameters, especially those relevant to the acceleration system.

Table 1

A few LEP1 parameters

Circumference		26 658 m
Revolution frequency		11.245 kHz
Injection energy		20 GeV
Operating energy		45 GeV
Number of bunches per beam		4
Intensity per bunch		~ 0.75 mA
RF frequency		352.209 MHz
Harmonic number		31320
Available RF voltage (128 cavities)		340 MV
RF cavity	Length:	2.12 m
	Impedance:	28.5 MΩ (43.8 MΩ with storage cavity)

The present RF system is composed of 128 copper cavities located in the two straight sections at Point 2 and Point 6 [1]. Each cavity is a five-cell structure (total length 2.5λ). The internal diameter of the drift tubes is 100 mm. The cavity design takes advantage of the very long distance between bunches, by storing the RF energy between bunch passages in a low-loss storage cavity coupled to the accelerating structure. In fact the RF energy oscillates between storage and accelerating cavity at $8 f_{rev}$ (f_{rev} being the machine revolution frequency). The storage cavity is a low-loss spherical resonator, which increases the effective shunt impedance of the coupled system by more than 50% as compared to the accelerating cavity alone, and hence reduces the RF power by the same factor. Sixteen cavities are powered via a tree of magic tees, from two high-power klystrons, each driven at one of the resonant frequencies of the coupled system.

LEP was conceived from the very beginning as a machine with an energy capability much higher than the Z_0 energy (45 GeV) which is its presently operating condition. Going higher requires much more RF voltage, because of the very steep increase in synchrotron radiation. The energy lost per turn by an electron or a positron U_0 is given by:

$$U_0 = \frac{4\pi}{3} \frac{r_e}{E_0^3} \frac{E^4}{\rho} \tag{1}$$

where r_e is the classical electron radius, E_0 its rest energy and E its actual energy. ρ is the radius of curvature. At the injection energy of 20 GeV, $U_0 = 4.5$ MeV and at the present energy of 45 GeV, $U_0 = 117$ MeV, whereas $U_0 = 1875$ MeV is needed to reach 90 GeV, above the W pair production threshold. Note that these figures, valid for a bare machine can be somewhat changed if additional wiggler magnets are inserted in the machine. This energy upgrade from 45 to 90 GeV is called the LEP2 programme.

When synchrotron radiation is important, as in the LEP case, the energy and phase oscillations of individual particles, present as in any other machine, are damped. This is because an electron which has for instance a little excess energy compared to equilibrium, will lose more energy [in Eq. (1) the term E^4 largely dominates the $1/\rho$ term] and ultimately reach equilibrium. The damping time τ_e of this exponential decay of energy oscillations is simply given by:

$$\tau_e = \frac{E}{U_0 f_{rev}} \tag{2}$$

This effect is not very important at injection ($\tau_e = 395$ ms) but very strong at 90 GeV ($\tau_e = 4.2$ ms).

If there were no other effects, all electrons and positrons would end up having exactly the same energy (the equilibrium energy). This would assume that the radiated energy was lost in a continuous, smooth way. This is not the case because of the quantum nature of radiation emission; photons are emitted randomly and their energy has a broad spectrum. From the point of view of electron dynamics, this is equivalent to a noisy RF, which is known to produce a gradual increase of the synchrotron oscillation amplitude. The result of the two competing effects, radiation damping and quantum emission, is to produce an equilibrium distribution of the particle's energy which has a gaussian shape with a typical relative width σ_e/E proportional to energy:

$$\frac{\sigma_e}{E} = \frac{E}{E_0}\sqrt{\frac{c_q}{2\rho}} \tag{3}$$

c_q being a constant of the machine.

The equilibrium energy distribution which extends theoretically to infinity is in fact limited by the RF bucket size. That part of the gaussian distribution beyond the RF separatrix on the energy axis is continuously lost. The beam lifetime is therefore limited; the beam intensity decays exponentially with a time constant τ_q (quantum lifetime) given by:

$$\tau_q = \tau_e \left(\frac{\sigma_e}{E}\right)^2 \exp\left(\frac{\Delta E^2}{2\sigma_e^2}\right) \tag{4}$$

where ΔE is the bucket height, depending on the RF voltage V_{RF}, the stable phase angle ϕ_s $(\sin\phi_s = U_0/V_{RF})$ and the energy.

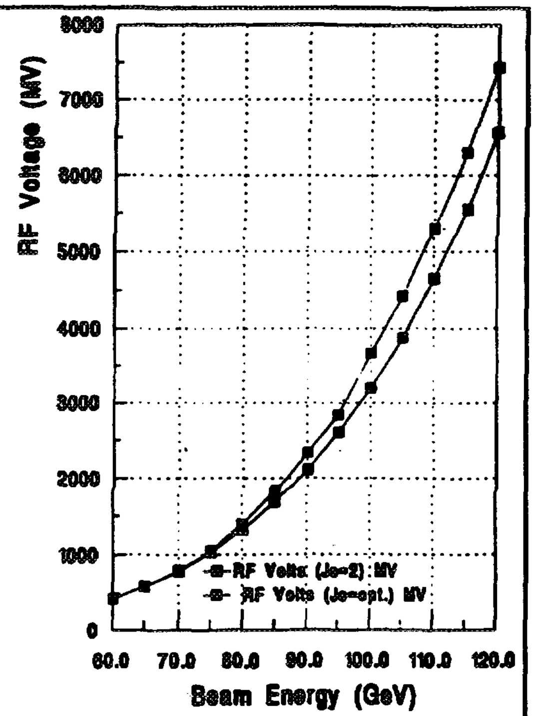


Fig. 1 RF voltage needed for a given beam energy and τ_q = 15h (two different lattices).

All ingredients are now available to evaluate the RF voltage necessary to obtain a given energy at a given quantum lifetime. Figure 1 shows the *result for LEP2 energies and* τ_q = *15 h, with two* different lattices. The necessary RF voltage is close to 2000 MV which, if only copper cavities were used, would require about 600 cavities, that is 1.3 km of accelerating structure and, even more striking, 50 MW of RF power wasted in the cavity walls.

Other constraints are usually put on the RF voltage in LEP:

- At low energy the synchrotron tune $Q_s = f_s/f_{rev}$ (f_s = synchrotron frequency) must be kept constant to avoid synchrobetatron resonances.
- During collisions, a proper balance between the RF stations (at present in points 2 and 6) is necessary to make sure that the average collision energy is the same in all experiments (at points 2, 4, 6 and 8) and to avoid loss of machine dynamic aperture.

Besides the collision energy, the other key parameter of a collider is its luminosity L given by the equation:

$$L = \frac{N_{e^+} N_{e^-} k_b f_{rev}}{4\pi\sigma_x\sigma_y} \tag{5}$$

where N_{e^+}, N_{e^-} are the number of particles per bunch, k_b the number of bunches and σ_x, σ_y their r.m.s. transverse sizes.

In the case of to-day's LEP, the two parameters N and σ are not independent. When two high-intensity bunches collide, (the highly non-linear) space charge transverse forces that one bunch exerts on the other bunch particles, leads to an increase of the transverse beam dimensions (σ_x, σ_y). The effect is characterized by the so-called beam-beam parameter ξ_y (vertical plane, most critical):

$$\xi_y = \frac{N r_e \beta_y^*}{2\pi(E/E_0)\sigma_x\sigma_y} = \frac{N r_e}{2\pi(E/E_0)\varepsilon} \tag{6}$$

Here β_y^* is the vertical beta function of the machine at the collision point and ε the beam transverse emittance. In LEP, it can be shown that ε is proportional to $(E/E_0)^3$, which means that ξ_y decreases strongly with energy.

At 45 GeV, the present LEP operating energy, L is still limited by the beam-beam effect, but this is very unlikely to be the case at 90 GeV, for LEP2. In the situation where beam-beam is not a limitation, Eq. (5) can be written:

$$L = \frac{I_b^2}{4\pi k_b f_{rev} \sigma_x \sigma_y} \tag{7}$$

$I_b = k_b N f_{rev}$: beam current. There the luminosity will be limited by the total beam current and therefore the total power P_{RF} to be delivered to the two beams:

$$P_{RF} = 2\, I_b\, U_0 \tag{8}$$

Equation (7) shows that the beam current should be obtained with as few bunches as possible (k_b small), or in other words the current per bunch should be as high as possible. The limit is now at injection (20 GeV) where the so-called transverse mode coupling instability limits the current ($I_b < I_{th}$) that can be accumulated in a single bunch.

The threshold current:

$$I_{th} = \frac{2\pi f_{rev}}{e} \frac{Q_s E}{\sum \beta_i k_i(\sigma_s)} \tag{9}$$

is inversely proportional to the total transverse machine impedance ($\sum \beta_i k_i$, where k_i is the transverse loss factor of the individual impedance, located at a position where the beta function is β_i). The threshold is proportional also to the synchrotron tune (running at high Q_s imposes a careful control of the synchrobetatron resonances) and to the injection energy. Thanks to new superconducting (sc) cavities installed in the LEP injector (see Section 4), the injection energy will be raised from 20 to 22 GeV, giving a straightforward 10% increase in intensity and hence luminosity.

In LEP, the transverse loss factor is dominated by the copper cavities. This is because k_i is, roughly speaking, inversely proportional to the cube of the vacuum tube diameter (100 mm in the case of the LEP RF cavities). With low-frequency superconducting cavities (352 MHz, iris diameter 250 mm), the transverse loss factor is reduced by a factor 8.5.

In conclusion LEP2 could not work at a useful luminosity without 352 MHz sc cavities having a low transverse impedance. This also justifies the choice of 352 MHz, which was also obviously favoured for reasons of hardware compatibility with the copper RF system.

To reduce the LEP transverse impedance further, it is foreseen that part of the copper RF system will be removed and replaced by sc cavities.

2. THE LEP2 CAVITIES

The LEP2 programme is based on a large number (~ 200) of four-cell, 352 MHz, sc cavities (Fig. 2) [2, 3]. From the beginning of the project the technique of niobium-copper was selected, partly because of the substantial savings made on niobium costs for such a large project, partly because of the high Q_0 values which can be obtained at the nominal operating accelerating field of 6 MV/m at 4.5 K and the absence of hard quenches.

The Nb-Cu technology is however delicate. Sputtering of Nb on the interior of the four-cell copper cavity is made by a magnetron-type discharge between the cylindrical Nb cathode and the cavity walls. The longitudinal position of the discharge along the cavity is controlled by an array of magnetizing coils inside the Nb cathode. The quality of the sputtered Nb layer is checked in a vertical cryostat test on the bare cavity, before any further assembly. The cavities are produced by three different European manufacturers, but all acceptance tests are carried out at CERN.

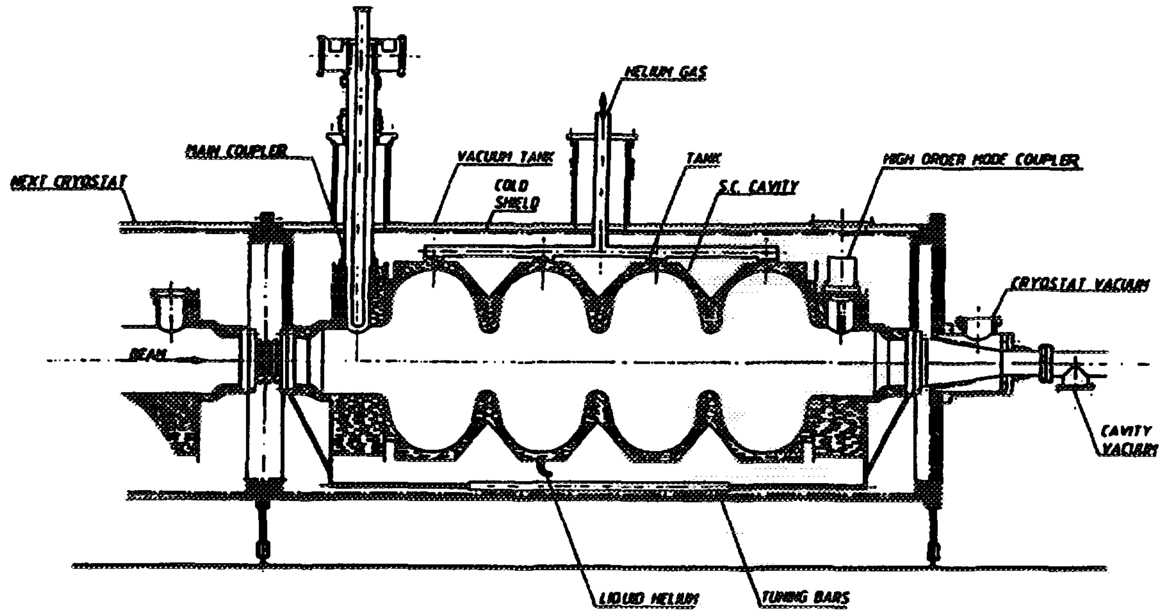


Fig. 2 LEP2 superconducting cavity

Despite the large number of cavities already produced, the success rate of the sputtering process is only about 70% at present. Critical parameters seem to be the chemical preparation of the surface, its exposure to ambient air before sputtering and the temperature of the cavity during coating. Rejected cavities are either rinsed or, in the worst cases, coated again, after stripping off the imperfect Nb layer. It is observed that the second coating is usually better than the first. The typical performance of cavities produced by industry is displayed in Fig. 3, showing a good reproducibility of the *Q(E)* curve among the three independent manufacturers. The slope of the *Q(E)* curve is characteristic of Nb/Cu technology; it is believed to be due either to the small grain size in the Nb layer or to substrate impurities migrating into it.

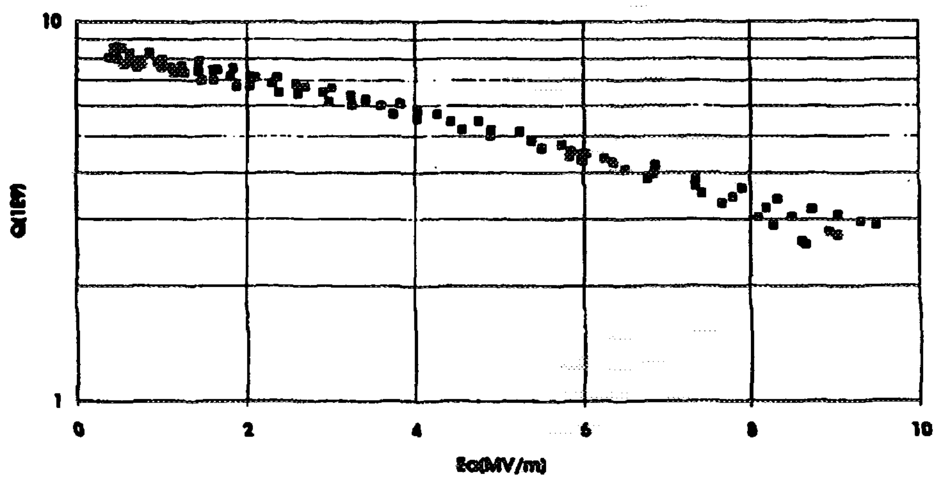


Fig. 3 *Q*(*E*) curves for cavities produced by three manufacturers

Accepted cavities are returned (under vacuum) to the firms for subsequent assembly of the helium tank, tuner bars and cryostat frame. Four such assembled cavities are then connected together in a single module. This is the second delicate operation in the overall fabrication cycle as the cavities have to be opened to air to be connected together and form a single vacuum enclosure. This is done in a clean room (class 100) and requires very experienced personnel to avoid dust contamination of the cavity surfaces. The loss of performance of the cavities after assembly into modules is fairly small ($\leq 10\%$ of Q_0 at 6 MV/m), and most of the industry-produced modules are accepted by CERN. In the case where a degradation cannot be recovered by helium processing or pulse processing [4], the cavity is taken out of the module, water-rinsed and rechecked.

LEP2 cavities are tuned with three longitudinal nickel bars connected to the two cavity end flanges. The temperature of the bars (and their length) is determined by the equilibrium between the cold return helium gas and an electric heater, which provides a slow control of the cavity tune. Fast control is obtained by the magnetostrictive effect of the nickel bars under an applied magnetic field (response time ~ 20 to 50 ms).

Mechanical cavity resonances can be harmful to LEP2 operation. When beam is present, the servo tuner keeps cavity voltage and forward power in phase (see Section 3), thus detuning the cavity. In this situation any tune modulation of the cavity (typically at the frequency of a longitudinal mechanical resonance) leads to an unwanted modulation of the RF voltage, thus limiting the overall performance. Excitation of mechanical cavity resonances (typically around 100 Hz) can be external, usually via the cryogenic system, or intrinsic to the cavity [5]. In the latter case an electroacoustic instability develops due to the dependence of the cavity tune on the RF field (Lorentz force detuning or thermal effects in the helium bath). Large modulations of the RF voltage (> 50%) and cavity phase can be observed. In theory, the fast cavity tuner could suppress this instability, if only one mechanical resonance were present. This is, unfortunately, not the case and the only practical remedy is to run the cavity closer to tune, at the expense of more RF power.

3. LEP2 COUPLERS

The main couplers of LEP2 cavities are of the coaxial line type (Fig. 4). The waveguide-to-coaxial transition is derived directly from the copper cavity coupler design, with a matching element (doorknob) in the waveguide and a warm ceramic vacuum window. The cold-warm transition is via the outer conductor of the coaxial line (thin stainless steel tube, copper coated).

The inner conductor is at room temperature (air-cooled) and forms a 75 Ω line which, compared to the original 50 Ω design, shifts upwards in power the multipacting levels in the coaxial line. The original coupler was variable with a sliding contact on the inner conductor, protected by a folded $\lambda/2$ line (choke) located inside the window area. Despite its interesting ability to adjust the cavity coupling (Q_{ext}) precisely to the waveguide distribution system, the variable coupler has been replaced by a fixed version ($Q_{ext} \cong 2 \cdot 10^6$) to avoid additional multipacting effects in the choke area.

These couplers have been plagued by multipacting effects, especially in the cold outer conductor of the coaxial line. The solution was to bias the inner conductor with a d.c. voltage of 2.5 kV which prevents any electron multiplication effect. The doorknob has been redesigned with a cylindrical kapton insulating foil which exhibits negligible RF losses. During cavity operation in LEP, the bias voltage completely suppresses the vacuum outbursts in the coupler. Another method of suppressing some multipacting levels has been successfully tested on the LEP2 couplers. It consists of injecting a second frequency in the coupler (a few MHz away from the operating RF frequency) with an amplitude of some 10 to 20% of the main RF drive. The resonance conditions necessary for multipacting to develop are strongly perturbed by the presence of this non-synchronous RF field in the coupler. Note that the cavity voltage is not affected, because of its very narrow bandwidth.

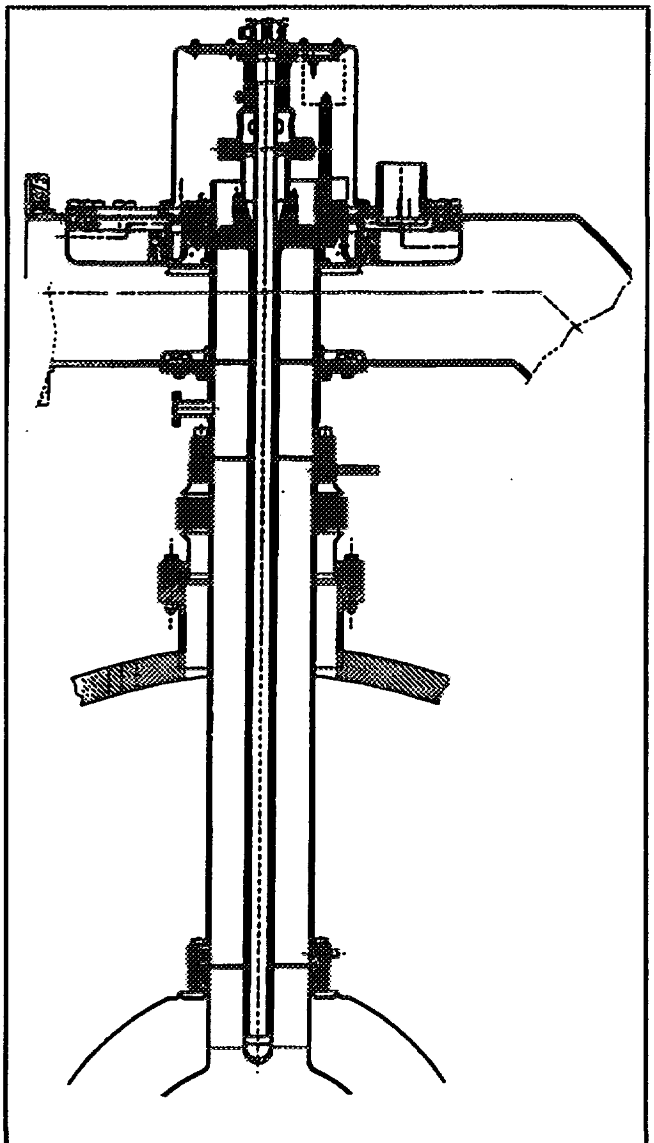


Fig. 4 The LEP2 cavity power coupler

Before installation on the cavities the couplers are processed at room temperature in travelling-wave conditions. Two identical couplers are connected on a strongly overcoupled cavity and RF power up to 200kW cw is transmitted through the couplers. It was found that processed couplers, after being mounted on the cavity (in clean-room conditions) were very difficult to condition again when the cavity was cold. This effect has been confirmed in a dedicated experiment where the outer conductor of a coupler, exposed to air, was cooled at liquid nitrogen temperature: conditioning was much more difficult with a cold than with a warm surface. As a practical conclusion, exposure to air during installation on the cavity should be kept to a strict minimum. It has also been found that baking *in situ* the RF window on the cavity, prior to cooling down, strongly reduces the conditioning time of the coupler. Proper cooling of the vacuum window during operation is important for reducing outgassing in the coupler. The Kovar rings brazed on the ceramic are copper-plated to minimize RF losses; they are brazed under axial constraint to keep a good RF contact during operation. Air cooling of the window is provided through holes in the doorknob.

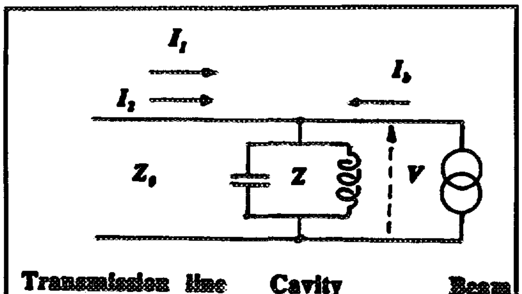


Fig. 5 Equivalent circuit of the sc cavity with beam

The critical parameter of the RF coupler is the peak electric field, which determines the multipacting levels encountered during operation. Figure 5 shows the equivalent circuit of the sc cavity (*L* and *C* elements) together with the coupler line (impedance $Z_0 = Q_{ext} \cdot R/Q$ transformed at the cavity "gap"). The beam is represented by a pure current source $\vec{I}_b$. $\vec{I}_1$ and $\vec{I}_2$ being the forward and reflected current waves on the line (measured at the cavity location or $n \cdot \lambda/2$ away), the following equations describe the circuit:

$$\vec{V} = Z(\vec{I}_1 + \vec{I}_b - \vec{I}_2) \qquad (10)$$

$$\vec{V} = Z_0\left(\vec{I}_1 + \vec{I}_2\right) \tag{11}$$

where $\vec{V}$ is the cavity accelerating voltage and Z the impedance of the parallel LC circuit (purely reactive).

The forward $\left(\vec{I}_1\right)$ and reflected $\left(\vec{I}_2\right)$ waves are given by:

$$2\vec{I}_1 = \vec{V}\left(\frac{1}{Z_0} + \frac{1}{Z}\right) - \vec{I}_b \tag{12}$$

$$2\vec{I}_2 = \vec{V}\left(\frac{1}{Z_0} - \frac{1}{Z}\right) + \vec{I}_b \tag{13}$$

The servo tuner compares the phase of forward wave I_1 and cavity voltage V, and brings the difference to zero by adjusting the value of the imaginary term Z. It follows that:

$$2I_1 = \frac{V}{Z_0} + I_b \sin\phi_s \tag{14}$$

$$2I_2 = \frac{V}{Z_0} - I_b \sin\phi_s \tag{15}$$

where ϕ_s is the stable phase angle ($\phi_s = +\pi/2$ for $\vec{V}$ and $\vec{I}_b$ in opposition).

Figure 6 shows the forward P_f and reflected P_r powers on the coupler line $P_f = \frac{1}{2}Z_0 I_1^2$ and $P_r = \frac{1}{2}Z_0 I_2^2$ respectively. There is matching for $V/Z_0 = I_b \sin\phi_s$, which corresponds to the optimum power transfer to the beam. Where the two waves, forward and reflected, are in phase the peak electric field in the line is proportional to the quantity $|I_1| + |I_2|$. Below the matching point $\left(V/Z_0 > I_b \sin\phi_b\right)$ Eqs. (14) and (15) show that the peak field is proportional to the cavity accelerating voltage V and independent of the beam current. Above, it becomes independent of V, but proportional to $I_b \sin\phi_s$. Going beyond the matching point changes little the power transfer to the beam, but unfortunately results in a rapid increase of the coupler peak field.

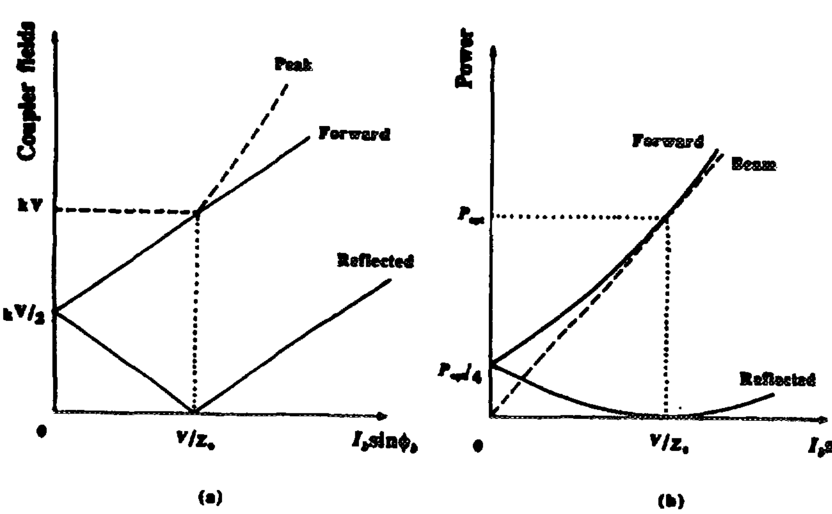


Fig. 6 Forward and reflected fields (a) and powers (b) in the coupler

The variations of Q_{ext} from cavity to cavity (due to mechanical tolerances) can be compensated by an outside $\lambda/4$ fixed transformer located in the upstream waveguide. This does not change the power transfer to the beam but leads to an increase of the coupler peak field.

It can be shown from Eq. (12) that running a cavity on tune ($1/Z = 0$), as was foreseen for suppressing electroacoustic instabilities in the cavity, will require additional power: $\Delta P = \frac{1}{2} Z_0 I_b^2 \cos^2 \phi_b$. In the case of LEP2 this additional power remains acceptable (of the order of 10 kW per coupler). However the peak field in the coupler increases by about 30%. All this shows the importance of conditioning the cavity couplers largely beyond their nominal power.

Each LEP2 cavity is equipped with two higher-order mode (HOM) couplers. They are of the "hook" type [6] where a series notch filter at the RF frequency is established with the inductance of the "hook" and its capacitance to the wall port. This type of HOM coupler is better suited to the Nb/Cu technology of LEP2 cavities. Liquid helium fills the hook tube (niobium material) to keep the notch filter elements superconducting. Adjustment of the notch frequency can be made outside the machine vacuum by elastic deformation of the base of the hook.

The power transmission capability of the HOM assembly is limited, not by the HOM coupler itself but rather by the connecting line, inside the insulation vacuum, between the cold coupler and the warm cryostat enclosure. The solution adopted is to use a rigid 25 Ω coaxial line consisting of two thin-walled stainless-steel tubes, copper-plated. Finger contacts at either end of the tubes allow some mechanical flexibility during the cryostat cooldown. It has been demonstrated experimentally that more than 850 W can be transmitted through the HOM coupler and its line, at 630 MHz (frequency of the dominant longitudinal HOM of the LEP2 cavity). This figure is largely beyond what is expected in LEP2 operation, for the largest beam currents considered; therefore the power capability of the HOM couplers is not a limitation on machine performance.

4. OPERATION WITH BEAM

At present (September 1995) ten niobium-copper modules are installed in LEP and six of them have been operated at nominal field (6 MV/m) and 7.5 mA beam current. Four more modules will go into the machine next month in order to be able to operate LEP at an intermediate energy of 65-70 GeV during a preparatory run in 1995.

Installation of the bulk of LEP2 cavities will take place in 1996, in two steps: 18 modules in spring to reach 81 GeV and another eight later in the year. Operation above the W pair threshold is expected by the end of 1996. A further increase of LEP2 energy up to about 97 GeV is foreseen for 1998 by adding 24 more modules [7].

Concerning operation with beam, the major differences with respect to copper cavities are linked to beam loading and microphonics. Beam-loading effects are much stronger in the case of LEP2 superconducting cavities, because of their much larger impedance at the RF frequency. The total impedance of the copper cavity system, as seen by the beam, is 2460 MΩ at 352 MHz; the same figure for the LEP2 superconducting cavities is more than one order of magnitude higher (~ 10^5 MΩ)

Without any special measures, the beam stability would be marginal. In particular at injection, the LEP2 operating point would be very close to the second Robinson instability limit (Fig. 7). It is recalled that the second Robinson limit corresponds to the case where the beam runs on the crest of the *generator-induced* voltage, i.e. when all RF power is transferred to the beam. In this situation the synchrotron frequency for the coherent dipole mode oscillation vanishes and stability of that mode is lost. To restore stability one could move the normal operating line ($\vec{I}_1$ and $\vec{V}$ in phase, solid line Fig. 7) further away from resonance. Unfortunately, to avoid electroacoustic instabilities, it is better to run closer to the tuned situation. There one would fall completely into the unstable region.

It has already been observed, with only a few modules in the machine that their operation is very dependent on beam parameters. This is not suprising, given the large beam-induced voltage. A typical example was the effect of a trip of a copper cavity unit (16 cavities), which increased the stable phase angle, and as a consequence decreased the sc cavities' voltage. The result was a further increase in ϕ_s, and in some cases a complete beam loss.

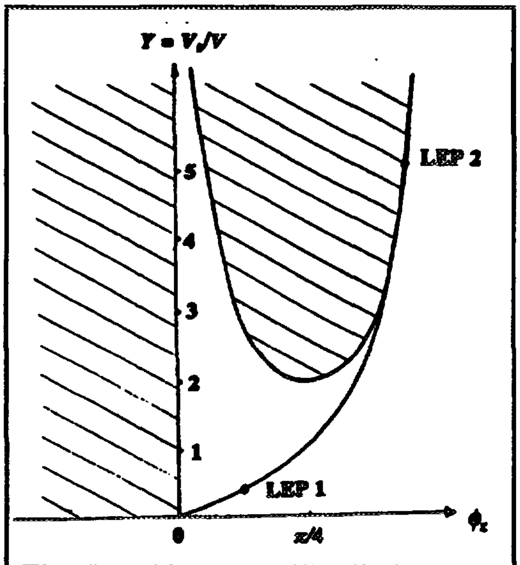


Fig. 7 Robinson stability limits: V_b = beam-induced voltage; ϕ_z = cavity detuning angle.

Microphonic effects are important for LEP2 cavities, because of their narrow bandwidth (±100 Hz), as compared to the LEP copper cavities (±4 kHz) and their mechanical construction (sheet-metal vessel, suspended at either end). Excitation of mechanical vibrations from the cryogenic system or other sources is very difficult to avoid completely. Large oscillations due to electroacoustic instabilities are suppressed by running the cavities closer to the tuned condition. This is achieved by changing the phase set point of the servo tuner on all eight cavities driven by the same klystron. Changing the set point of only the unstable cavity will primarily increase the voltage on that particular cavity (at constant RF drive) and is very likely to enhance the instability. In any case, adjusting the servotuner set points means changing the phase of the cavity with respect to the reference RF, which is undesirable for maximizing the overall available accelerating voltage.

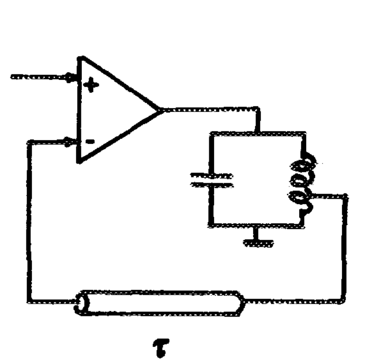


Fig. 8 Principle of RF feedback loop

In order to overcome all these difficulties, a fast RF feedback system is being implemented on all modules now installed in LEP. The total RF voltage seen by the beam when crossing the eight cavities driven by a common klystron, is reconstructed from the field-probe signals of each cavity. Great care must be applied to the calibration of the probes and cable connections (in amplitude and phase) to ensure that the overall vector sum signal is a faithful representation of the RF voltage experienced by the beam. The "vector sum" signal is maintained equal to the demanded RF voltage by the action of an RF feedback loop (Fig. 8) The loop gain, at resonance is about 30 dB, sufficient to reduce the equivalent impedance of sc cavities down to the LEP1 level, and to correct any phase variations on the cavities (either from residual microphonics, or from various phase settings of the servo-tuners).

RF feedback is a very powerful technique for dealing with beam-loading effects. The minimum achievable impedance, at the fundamental cavity frequency, is given by:

$$R_{\min} = 4\frac{R}{Q} f_{RF}\tau \qquad (16)$$

where τ is the overall delay in the feedback path. In the case of LEP2, the overall gain is rather limited by the quality of the "vector sum" reconstruction.

This is not the case for the sc cavities used in the LEP injector (the multipurpose SPS machine, which accelerates protons and leptons on alternate cycles). Here reduction of impedance is critical, especially for the very high-intensity proton beam: it is achieved by proper filtering of all four cavity modes. The loop delay (500 ns) is essentially determined by the distance in frequency between the $3\pi/4$ and π modes (about 1 MHz) which corresponds to a 180° phase rotation. In addition to the short delay RF feedback, a complementary loop with an overall delay of one machine turn (23 μs) is used to further reduce the cavity impedance. Figure 9 shows the pulsed RF waveform obtained on the SPS sc cavities, together with the demanded power from the tetrode power generator.

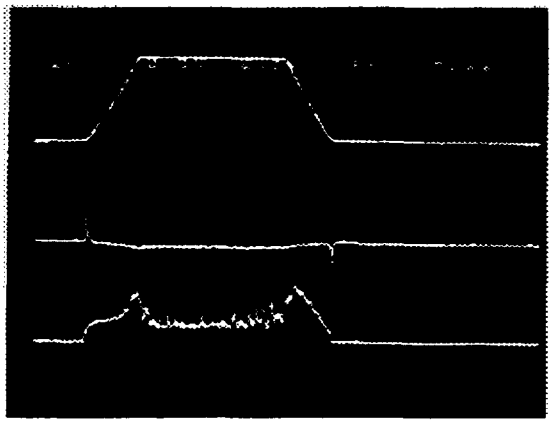


Fig. 9 Pulsed RF waveform on SPS sc cavities (50 ms/div)

Top: Cavity voltage (peak = 7.5 MV)
Middle: Tuner error signal
Bottom: Tetrode drive

REFERENCES

[1] LEP Design Report, Vol. II The LEP Main Ring, CERN-LEP/84-01, June 1984.

[2] P. Bernard, H. Lengeler, E. Picasso, Upgrading of LEP Energies by Superconducting Cavities, CERN/LEP DI/86-29.

[3] P. Bernard et al., Superconducting Cavities for LEP, Status Report, Proc. EPAC, 1988, Rome, Italy.

[4] J.A. Uythoven, First Experience with Pulsed Power Processing of the LEP Superconducting Cavities, CERN LEP2 Note 94-22.

[5] D. Boussard, P. Brown, J. Tückmantel, Electroacoustic Instabilities in the LEP2 Superconducting Cavities, CERN-SL-95-81 (RF) and Proc. 7th Workshop on RF Superconductivity, Gif-sur-Yvette, France, October 1995.

[6] P. Bernard et al., Demountable E/H field HOM for the Nb-sputtered 4-cell LEP Cavity, 5th Workshop on RF Superconductivity, Hamburg, 1991, p. 956.

[7] S. Myers, C. Wyss, LEP2 Energy Upgrade, CERN LEP2 Note 95-34.

NEXT PAGE(S) left BLANK

LIST OF PARTICIPANTS

ADAM, J.	CERN, Geneva, Switzerland
ANTHON, A.	DANFYSIK A/S, Jyllinge, Denmark
BACHER, R.	DESY, Hamburg, Germany
BAUDRENGHIEN, P.	CERN, Geneva, Switzerland
BAUER, P.	TU WIEN, Vienna, Austria
BENDA, V.	CERN, Geneva, Switzerland
BERRIAUD, C.	CEA Saclay, Gif-sur-Yvette, Switzerland
BIALOWONS, W.	DESY, Hamburg, Germany
BIENVENU, G.	LAL, Orsay, France
BLANC, D.	CERN, Geneva, Switzerland
BLOMQVIST, I.	DANFYSIK A/S, Jyllinge, Denmark
BOER ROOKHUIZEN, H.	NIKHEF-K, Amsterdam, Netherlands
BONNETON, M.	L'Air Liquide, Sassenage, France
BOURDON, J.C.	LAL, Orsay, France
BREMER, J.	CERN, Geneva, Switzerland
BRÜCK, H.	DESY, Hamburg, Germany
BRUNET, J-C.	CERN, Geneva, Switzerland
BUGOI, R.	Institute of Physics, Bucharest, Romania
CALATRONI, S.	CERN, Geneva, Switzerland
CASSINESE, A.	Univ.di Napoli "Federico II", Naples, Italy
CHAMPION, M.	DESY, Hamburg, Germany
DE JONGE, L.	CERN, Geneva, Switzerland
DURAND, C.	CERN, Geneva, Switzerland
DWERSTEG, B.	DESY, Hamburg, Germany
ECKOLDT, T.	DESY, Hamburg, Germany
FACCO, A.	INFN/LNL, Legnaro, Italy
FEDURIN, M.	BINP, Novosibirsk, Russia
FRIEDBERGER, G.	TU Wien, Vienna, Austria
FULJAHN, T.	DESY, Hamburg, Germany

GALL, D.	DESY, Hamburg, Germany
GARVEY, T.	LAL, Orsay, France
GURATZSCH, H.	Forschungszentrum Rossendorf, Dresden, Germany
HAGEDORN, D.	CERN, Geneva, Switzerland
HANKE, K.	DESY, Hamburg, Denmark
HAUGE, N.	DANFYSIK A/S, Jyllinge, Germany
HE, P.	CERN, Geneva, Switzerland
JUILLARD, M.	CEA-Saclay, Gif-sur-Yvette, France
KARPPINEN, M.	CERN, Geneva, Switzerland
KIRBY, G.	CERN, Geneva, Switzerland
KNOOPS, S.	CERN, Geneva, Switzerland
KOSKI, A.	DESY, Hamburg, Germany
KOUTCHOUK, J-P.	CERN, Geneva, Switzerland
KREPS, G.	DESY, Hamburg, Germany
KUBO, T.	RIKEN, Saitama, Japan
KUZMINSKI, J.	DESY, Hamburg, Germany
LACARRERE, D.	CERN, Geneva, Switzerland
LECA, G.	Univ. of Bucharest, Romania
LEHMANN, M.	DESY, Hamburg, Germany
MATHEISEN, A.	DESY, Hamburg, Germany
MAYRI, C.	CEA Saclay, Gif-sur-Yvette, France
MODENA, M.	INFN, Frascati, Italy
MÖLLER, W-D.	DESY, Hamburg, Germany
PASSARDI, G.	CERN, Geneva, Switzerland
PERINET-MARQUET, J-L.	CERN, Geneva, Switzerland
PERINI, D.	CERN, Geneva, Switzerland
PETERS, A.	GSI, Darmstadt, Germany
PROUDLOCK, P.	CERN, Geneva, Switzerland
PUTTNER, M.	TU Wien, Vienna, Austria
QUAN-SHENG, S.	DESY, Hamburg, Germany
RENKEN, D.	DESY, Hamburg, Germany

RICHTER, D.	CERN, Geneva, Switzerland
RIDDONE, G.	CERN, Geneva, Switzerland
RUBER, R.	The Svedberg Lab., Uppsala, Sweden
RUSSO, R.	CERN, Geneva, Switzerland
SAFA, H.	CEA-Saclay, Gif-sur-Yvette, France
SAVARY, F.	CERN, Geneva, Switzerland
SCHMICKLER, H.	CERN, Geneva, Switzerland
SCHMIDT, R.	CERN, Geneva, Switzerland
SCHROEDER, C.H.	GSI, Darmstadt, Germany
SEKUTOWICZ, J.	DESY, Hamburg, Germany
SERIO, L.	CERN, Geneva, Switzerland
SIAMBANIS, T.	CERN, Geneva, Switzerland
SIEMKO, A.	CERN, Geneva, Switzerland
SINGER, W.	DESY, Hamburg, Germany
SOLHEIM, N.O.	CERN, Geneva, Switzerland
TAVIAN, L.	CERN, Geneva, Switzerland
TENG, M.	CERN, Geneva, Switzerland
THOMAS, A.	Uni Mainz, Germany
TOMMASINI, D.	CERN, Geneva, Switzerland
TOWNSEND, M.	Oxford Instruments, Eynsham, Oxon, England
TRBOJEVIC, D.	BNL, Upton, NY, USA
TREADGOLD, J.R.	Oxford Instruments, Eynsham, Oxon, England
VEDRINE, P.	CEA Saclay, Gif-sur-Yvette, France
VULLIERME, B.	CERN, Geneva, Switzerland
WEDENIG, R.	TU Wien, Vienna, Austria
WEISZ, S.	CERN, Geneva, Switzerland
WIESER, S.	TU Wien, Vienna, Austria
WILHELMSSON, M.	CERN, Geneva, Switzerland
WITZELING, W.	CERN, Geneva, Switzerland
ZICHY, J.A.	PSI, Villigen, Switzerland