CERN–2004–008
11 September 2004

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

CAS CERN ACCELERATOR SCHOOL

SUPERCONDUCTIVITY AND CRYOGENICS FOR ACCELERATORS AND DETECTORS

Erice, Italy
8–17 May 2002

PROCEEDINGS

Editors: S. Russenschuck, G. Vandoni

GENEVA
2004

ISSN 0007–8328
ISBN 92–9083–230–4

CERN–2004–008
11 September 2004

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

CERN ACCELERATOR SCHOOL

SUPERCONDUCTIVITY AND CRYOGENICS FOR ACCELERATORS AND DETECTORS

Erice, Italy
8–17 May 2002

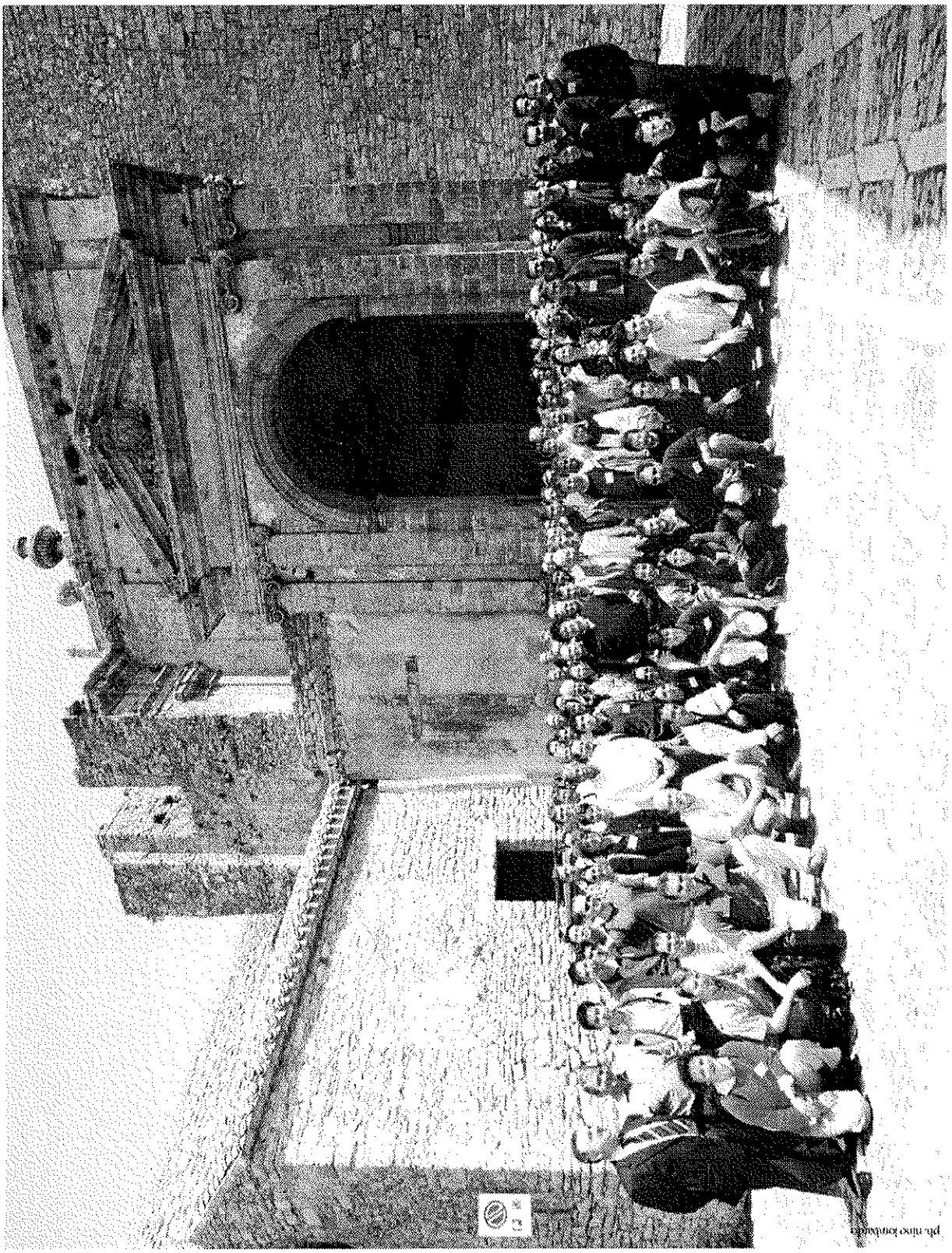
PROCEEDINGS

Editors: S. Russenschuck, G. Vandoni

GENEVA
2004

ABSTRACT

This Report contains the write-ups of the presentations given at the CERN Accelerator School course on Superconductivity and Cryogenics for Accelerators and Detectors held at the Ettore Majorana Foundation and Centre for Scientific Culture in Erice, Italy. The contributions are ordered in six groups, namely theory of superconductivity, magnet design and applications, cavity design, thermodynamics, heat transfer and refrigeration techniques, and vacuum technology.



FOREWORD

This course was conceived in 2000 at Erice, by Philippe Lebrun (head of the LHC division at CERN) and myself, during a workshop dedicated to accelerators at the energy frontier. Working on the challenges that the present accelerators, and hopefully the future ones too, are posing to the applied superconductivity community, we felt the time was more than ripe for a new CAS Course on Superconductivity for Accelerators (the first two were both hosted by DESY in Hamburg, in 1988 and 1995 respectively). This time, however, the scope is essentially wider: cryogenics has been treated as an essential part of the superconducting installation, and the detector magnets community, which shares so many technologies with the accelerator community, has also been invited to participate. The name of the course, **Superconductivity and Cryogenics for Accelerators and Detectors,** stems from these considerations.

I am very grateful to Ted Wilson, Director of CAS, for having accepted this proposal, and to all the 24 lecturers who came to the course and who stayed for many days or the whole duration. I would like especially to thank the speakers that have found the energy and the time to provide the write-up of their lecture!

This volume will certainly constitute a reference over the coming years and has been made possible by the work of the two editors: Stephan Russenschuck (CERN) and Giovanna Vandoni (CERN). In particular I would like to acknowledge the tireless work of Stephan for providing the necessary liaison between authors and publisher in the final phase of editing the proceedings.

I have been delighted that this course, with its 115 attendants from all over the world, has been the most successful of its type and one of the most attended CERN Accelerator School 'specialist courses'. Many scientists, students and professors, found the time to gather in this splendid town and spend 12 days living together, creating a real community dedicated to knowledge and human relations — one of the main purposes of the Erice Centre.

This course has been realized thanks to the dedication and commitment of many people, among others the staff of the CERN Accelerator School and the staff of the Erice Centre: to all of them, and especially to the tireless Signorina Fiorella of the Erice Centre, are due my thanks and the thanks of all participants.

Finally, I would like to express our gratitude to Prof. A. Zichichi for his hospitality at the wonderful Ettore Majorana Foundation and Centre for Scientific Culture (which hosted the first course, organized by the just-starting CAS some 20 years ago!) and for having honoured the course with his presence and a lecture where he shared with us his vision of the challenges that HEP will face in the next decades.

Prof. Lucio Rossi
CERN and Physics Department of the University of Milan
Director of the Course and Local Organizer

CONTENTS

Superconductivity

Peter Schmüser
Institut für Experimentalphysik der Universität Hamburg

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, superconductors in microwave fields 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 an 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.

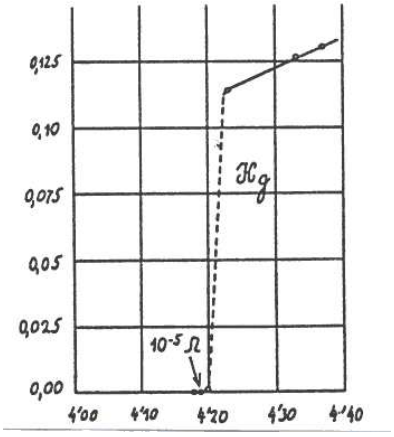


Fig. 1: The discovery of superconductivity by Kamerlingh Onnes.

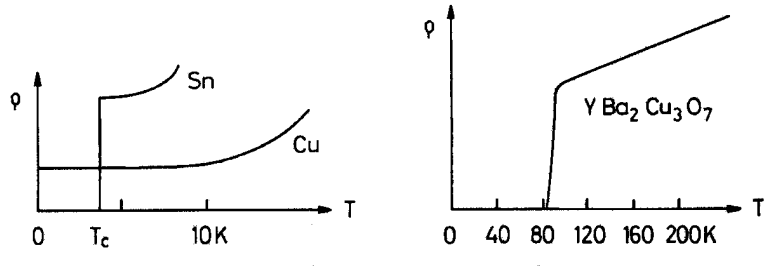


Fig. 2: The low-temperature resistivity of copper, tin and $YBa_2Cu_3O_7$.

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

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 conventional resistance measurement is far too insensitive to establish infinite conductivity, a much better method consists in inducing 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].

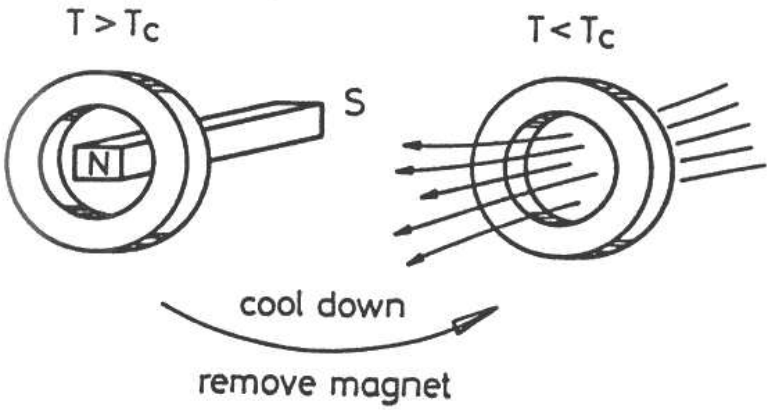


Fig. 3: Induction of a persistent current in a superconducting ring.

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* H_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, H_{c1} and H_{c2}. Below H_{c1} these substances are in the *Meissner phase* with complete field expulsion while in the range $H_{c1} < H < H_{c2}$ 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 EQUATION

We consider a cylinder with perfect conductivity and raise a magnetic field from zero to a finite value H. A surface current is induced whose magnetic field, according to Lenz's rule, is opposed to the applied field and cancels it in the interior. Since the resistance is zero 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 H. 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.

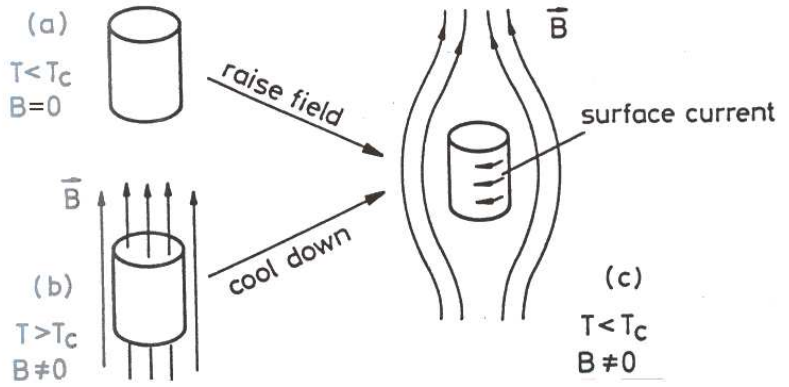


Fig. 4: A lead cylinder in a magnetic fi eld. Two possible ways to reach the superconducting fi nal state with $H > 0$ are sketched. Ideally the length of the cylinder should be much larger than its diameter to get a vanishing demagnetisation factor.

In a (T, H) plane, the superconducting phase is separated from the normal phase by the curve $H_c(T)$ 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.

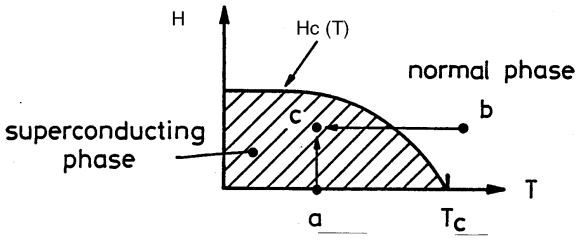


Fig. 5: The phase diagram in a (T, H) plane.

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, H) lies below the curve $H_c(T)$ the material is superconducting and expels the magnetic field, irrespective of by which path the point was reached. If (T, H) 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

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

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 equation

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

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

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

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 is identical to zero. Then one obtains the important London equation

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

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 of type I, 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 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

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

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

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

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*

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

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. Their mass is $m_c = 2m_e$, their charge $-2e$, their density $n_c = n_s/2$. Obviously the penetration depth remains unchanged when going from single electrons to Cooper pairs.

We have now convinced ourselves that the superconductor can tolerate a magnetic field only in a thin surface layer (this is the case for type I superconductors). 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.

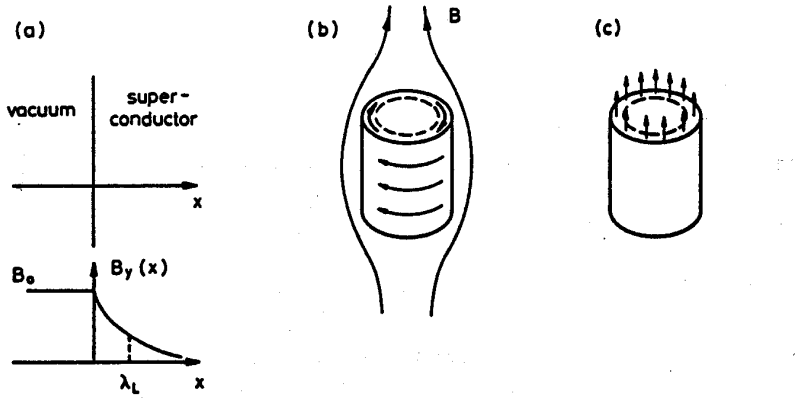


Fig. 6: (a) Exponential attenuation of a magnetic fi eld in a superconducting half plane. (b) Shielding current in a superconducting cylinder induced by a fi eld parallel to the axis. (c) A current-carrying wire made from a type I superconductor.

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.

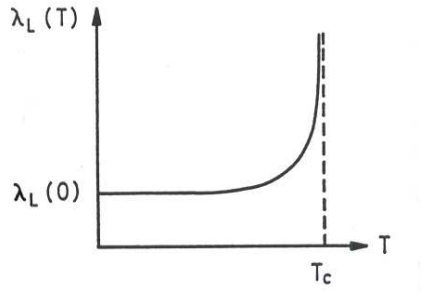


Fig. 7: Temperature dependence of the London penetration depth.

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 $H_c(T)$. 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 Gibbs free energy (see Appendix A)

$$G = U - T\,S - \mu_0 \vec{M} \cdot \vec{H} \tag{5}$$

where U is the internal energy, S the entropy and M the magnetisation of the superconductor (the magnetic moment per unit volume). A measurent of the free energy of aluminium is shown in Fig. 8a. Below T_c the superconducting state has a lower free energy than the normal state and thus the transition normal $\rightarrow$ superconducting is associated with a gain in energy. The entropy of the superconducting state is lower because there is a higher degree of order in this state. From the point of view of the BCS theory this is quite understandable since the conduction electrons are paired and collect themselves in a single quantum state. Numerically the entropy difference is small, though, 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. It should be noted, that also normal conduction is carried by just a small fraction of the valence electrons, see sect. 4.1.

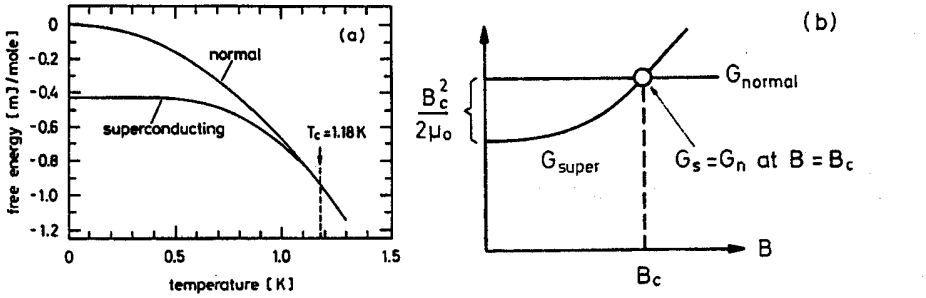


Fig. 8: (a) Free energy of aluminium in the normal and superconducting state as a function of T (after N.E. Phillips). The normal state was achieved by exposing the sample to a field larger than H_c while the superconducting state was measured at $H = 0$. (b) Schematic sketch of the free energies G_{norm} and G_{sup} as a function of the applied field $B = \mu_0 H$.

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:

$$G_{sup} < G_{norm} \text{ for } T < T_c \ .$$

What happens if we apply a magnetic field? A normal-conducting metal cylinder is penetrated by the field so its free energy does not change: $G_{norm}(H) = G_{norm}(0)$. In contrast to this, a superconducting cylinder is strongly affected by the field. It sets up shielding currents which generate a magnetic moment $\vec{m}$ antiparallel to the applied field. The magnetic moment has a positive potential energy in the magnetic field

$$E_{pot} = -\mu_0 \, \vec{m} \cdot \vec{H} = +\mu_0 \, |\vec{m}| \, |\vec{H}| \ . \tag{6}$$

In the following it is useful to introduce the magnetisation M as the magnetic moment per unit volume. The magnetisation of a superconductor inside a current-carrying coil resembles that of an iron core. The 'magnetising' field H is generated by the coil current only and is unaffected by the presence of a magnetic material while the magnetic flux density[1] B is given by the superposition of H and the superconductor magnetisation M:

$$\vec{B} = \mu_0(\vec{H} + \vec{M}) \ . \tag{7}$$

In the following I will call both H and B magnetic fields. For a type I superconductor we have

$$\vec{M}(\vec{H}) = -\vec{H} \quad \text{and } \vec{B} = 0 \tag{8}$$

as long as $H < H_c$. The potential energy per unit volume is obtained by integration

$$E_{pot} = -\mu_0 \int_0^H \vec{M}(\vec{H}') \cdot d\vec{H}' = \mu_0 \int_0^H H'^2 dH' = \frac{\mu_0}{2} H^2 \ . \tag{9}$$

This corresponds to the increase in the Gibbs free energy that is caused by the magnetic field, see Fig. 8b.

$$G_{sup}(H) = G_{sup}(0) + \frac{\mu_0}{2} H^2 \ . \tag{10}$$

Here and in the following G denotes the Gibbs free energy per unit volume. The *critical field* is achieved when the free energy in the superconducting state just equals the free energy in the normal state

$$\frac{\mu_0}{2} H_c^2 = G_{norm} - G_{sup}(0) \ . \tag{11}$$

Since the energy density stored in a magnetic field is $(\mu_0/2)H^2$, an alternative interpretation of Eq. (11) is the following: in order to go 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 the difference between the two free energies at vanishing field. For $H > H_c$ the normal phase has a lower energy, so superconductivity breaks down.

[1] There is often a confusion whether the H or the B field should be used. Unfortunately, much of the superconductivity literature is based on the obsolete CGS system of units 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.

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. 9). Type II conductors are characterized by two critical fields, H_{c1} and H_{c2}, which are both temperature dependent. For fields $0 < H < H_{c1}$ the substance is in the *Meissner phase* with complete exclusion of the field from the interior. In the range $H_{c1} < H < H_{c2}$ the substance enters the *mixed phase*, often also called Shubnikov phase: part of the magnetic flux penetrates the bulk of the sample. Above H_{c2}, finally, the material is normal-conducting. The area under the curve $M = M(H)$ is the same as for a type I conductor as it corresponds to the free-energy difference between the normal and the superconducting state and is given by $(\mu_0/2)\, H_c^2$.

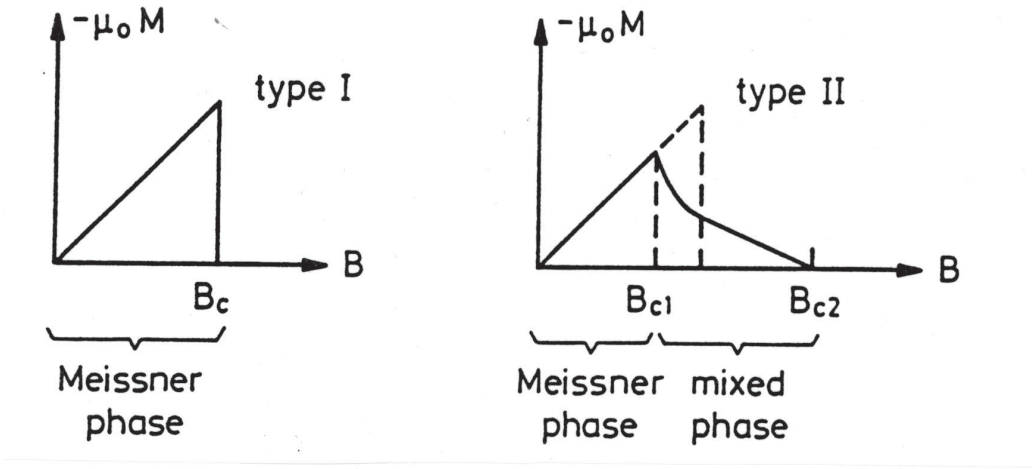


Fig. 9: Magnetisation of type I and type II superconductors as a function of the magnetic field.

It is instructive to compare measured data on pure lead (type I) and lead-indium alloys (type II) of various composition. Figure 10 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. 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. 10 should be identical as the diagram clearly confirms.

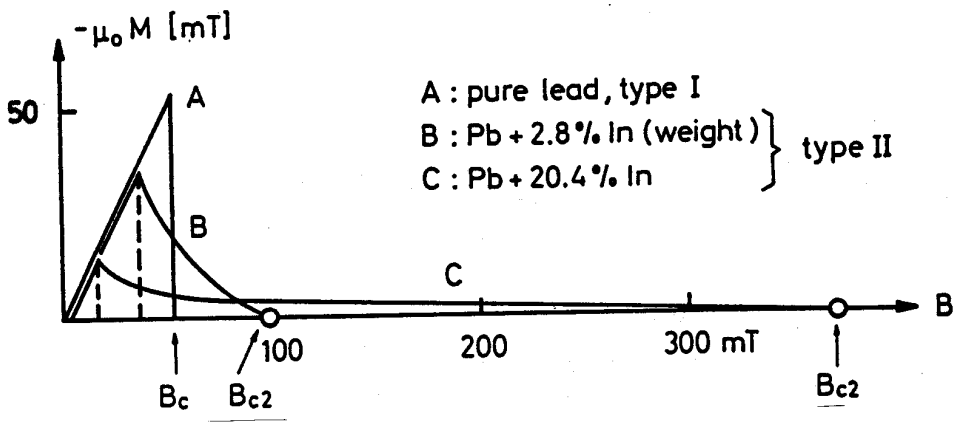


Fig. 10: The measured magnetisation curves [14] of lead-indium alloys of various composition, plotted against $B = \mu_0 H$.

A remarkable feature, which will be addressed in more detail later, 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. 11. Each tube is surrounded with a super-vortex current. The material in between the tubes is field- and current-free.

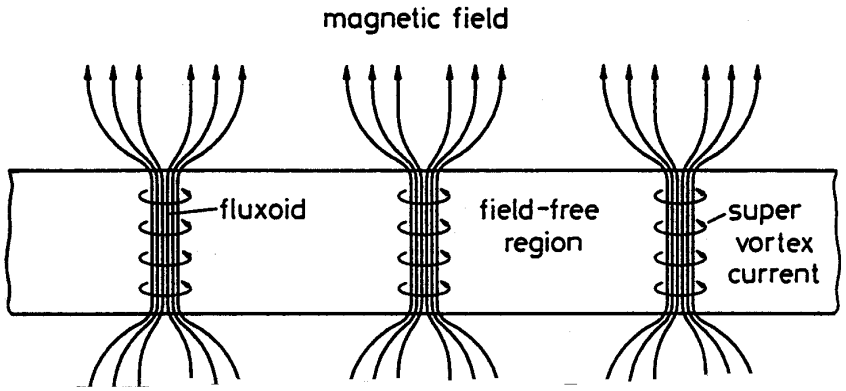


Fig. 11: Flux tubes in a type II superconductor.

The fact that 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. Figure 12 shows that a type II superconducting cylinder in the mixed phase has a smaller magnetic moment than a type I cylinder. This implies that the curve $G_{sup}(H)$ reaches the level $G_{sup}(H) = G_{norm}$ at a field $H_{c2} > H_c$.

In a (T, H) plane the three phases of a type II superconductor are separated by the curves $H_{c1}(T)$ and $H_{c2}(T)$ which meet at $T = T_c$, see Fig. 13a. The upper critical field can assume very large values which make these substances extremely interesting for magnet coils (Fig. 13b).

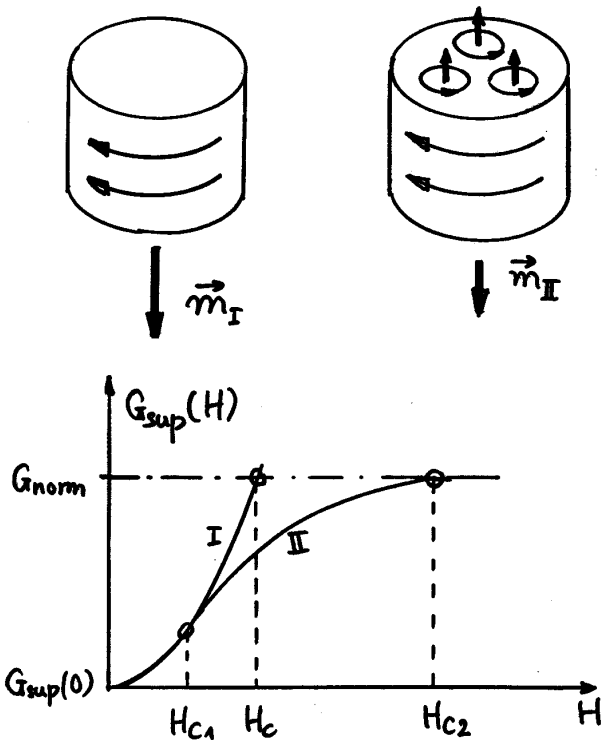


Fig. 12: Top: Magnetic moment of a type I and a type II sc cylinder in a field $H_{c1} < H < H_c$. Bottom: The Gibbs free energies of both cylinders as a function of field.

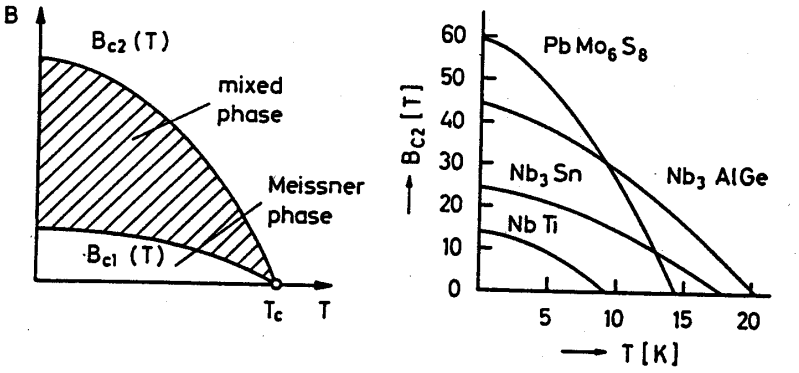


Fig. 13: (a) The phase diagram of a type II superconductor. (b) The upper critical field $B_{c2} = \mu_0 H_{c2}$ of several high-field alloys as a function of temperature.

3.4 When is a superconductor of type I or type II?

3.41 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 B field in the bulk (Fig. 14a) while in the

thin sheet (Fig. 14b) the B field does not drop to zero at the centre. Consequently less energy needs to be expelled which implies that the critical field of a very 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.

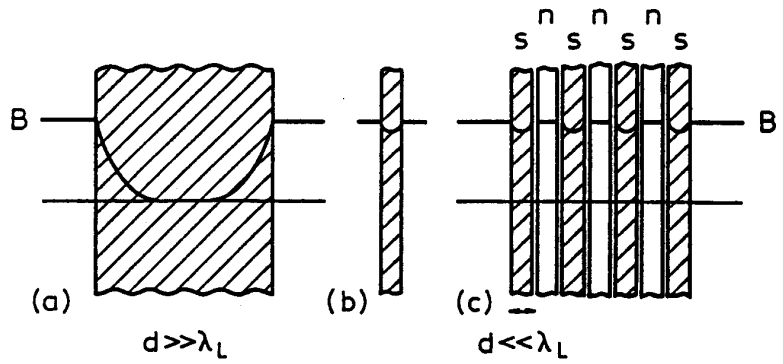


Fig. 14: Attenuation of field (a) in a thick slab and (b) in thin sheet. (c) Subdivision of a thick slab into alternating layers of normal and superconducting slices.

3.42 Coherence length

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

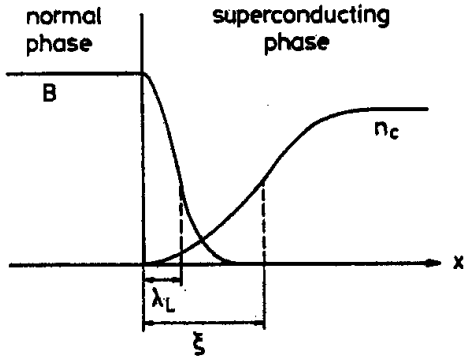


Fig. 15: The exponential drop of the magnetic field and the rise of the Cooper-pair density at a boundary between a normal and a superconductor.

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 *thermodynamic critical field* by the energy relation

$$\frac{\mu_0}{2}H_c^2 = G_{norm} - G_{sup}(0)\ . \tag{12}$$

For type I this coincides with the known H_c, see Eq. (11), while for type II conductors H_c lies between H_{c1} and H_{c2}. The difference between the two free energies, $G_{norm} - G_{sup}(0)$, can be intepreted as the Cooper-pair condensation energy.

For a conductor of unit area, exposed to a field $H = H_c$ 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:

$$\Delta E_{magn} = \frac{\mu_0}{2} H_c^2 \, \lambda_L \ . \tag{13}$$

(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

$$\Delta E_{cond} = -\frac{\mu_0}{2} H_c^2 \, \xi \ . \tag{14}$$

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*

$$\kappa = \lambda_L / \xi \ . \tag{15}$$

The criterion for type I or II superconductivity is found to be

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

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 ℓ of the conduction electrons in the metal. This quantity can be large for a very pure crystal but is strongly reduced by lattice defects and impurity atoms. In alloys the mean free path is generally much shorter than in pure metals so alloys are always type II conductors. In the Ginzburg-Landau theory the upper critical field is given by

$$B_{c2} = \sqrt{2}\, \kappa \, B_c = \frac{\Phi_0}{2\pi \xi^2} \tag{16}$$

where Φ_0 is the flux quantum (see sect. 5.2). For niobium-titanium with an upper critical field $B_{c2} = 10$ T at 4.2 K this formula yields $\xi = 6$ nm. The coherence length is larger than the typical width of a grain boundary in NbTi which means that the supercurrent can move freely from grain to grain. In high-T_c superconductors the coherence length is often shorter than the grain boundary width, and then current flow from one grain to the next is strongly impeded. There exists no simple expression for the lower critical field. In the limit $\kappa \gg 1$ one gets

$$B_{c1} = \frac{1}{2\kappa} (\ln \kappa + 0.08) B_c \ . \tag{17}$$

3.5 Heat capacity and heat conductivity

The specific heat capacity per unit volume at low temperatures is given by the expression

$$C_V(T) = \gamma T + A T^3 \; . \tag{18}$$

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):

$$\gamma = \frac{\pi^2 n k_B^2}{2E_F} \; , \quad A = \frac{12\pi^4 N k_B}{5\Theta_D^3} \; . \tag{19}$$

Here $k_B = 1.38 \cdot 10^{-23}$ J/K is the Boltzmann constant, E_F the Fermi energy, n the density of the free electrons, N the density of the lattice atoms 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 to energy transport. In the BCS theory one expects an exponential rise of the electronic heat capacity with temperature

$$C_{e,s}(T) = 8.5\,\gamma\,T_c\,\exp(-1.44\,T_c/T) \tag{20}$$

The experimental data (Fig. 16a, b) are in good 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.

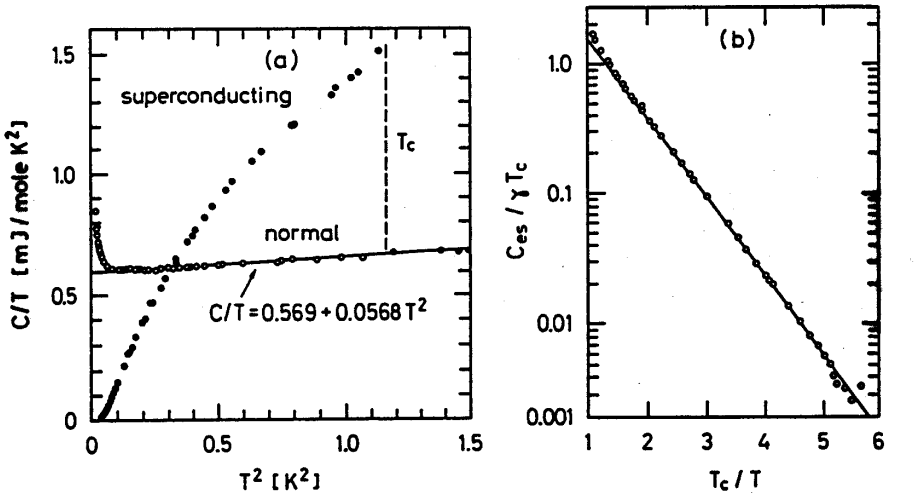


Fig. 16: (a) Specifi c heat $C(T)/T$ of normal and superconducting gallium as a function of T^2 [15]. (b) Experimental verifi cation of Eq. (20).

The heat conductivity of niobium is of particular interest for superconducting radio frequency cavities. Here the theoretical predictions are rather imprecise and measurements are indispensible. The low temperature values depend strongly on the *residual resistivity ratio* $RRR = R(300\,\mathrm{K})/R(10\,\mathrm{K})$ of the normal-conducting niobium. Figure 17 shows experimental data [16].

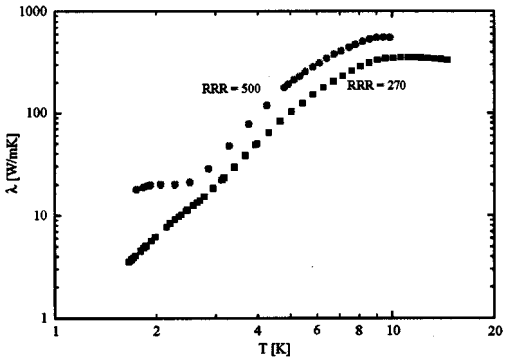


Fig. 17: Measured heat conductivity in niobium samples with $RRR = 270$ and $RRR = 500$ as a function of temperature [16].

4. BASIC CONCEPTS AND RESULTS OF THE BCS THEORY

4.1 The 'free electron gas' in a normal metal

4.11 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 (taking into account the periodic lattice potential leads to the electronic band structure of semiconductors). The energy levels are computed by solving the Schrödinger equation with boundary conditions, 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

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

which amounts to about 0.025 eV at room temperature. Instead, the energy levels are filled with two electrons each up to the Fermi energy E_F. Since the electron density n 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

$$E_F = \frac{\hbar^2}{2m_e}(3\pi^2 n)^{2/3}\ . \tag{21}$$

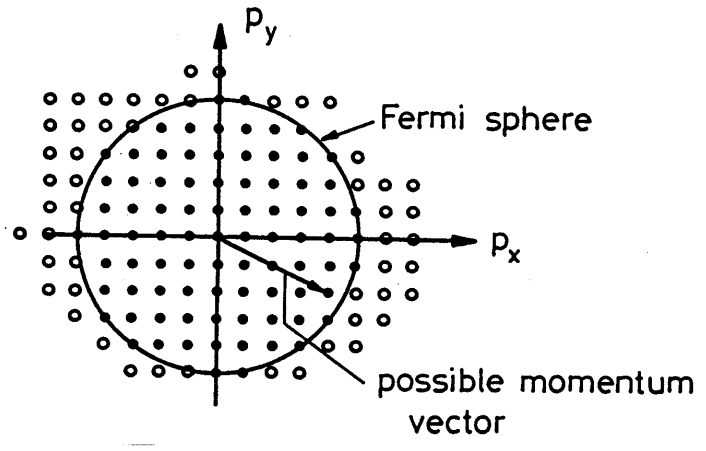


Fig. 18: The allowed states for conduction electrons in the $p_x p_y$ plane and the Fermi sphere. The occupied states are drawn as full circles, the empty states as open circles.

The quantity $\hbar = h/2\pi = 1.05 \cdot 10^{-34}$ Js $= 6.58 \cdot 10^{-15}$ eVs is Planck's constant, the most important constant in quantum theory. In order to remind the reader I will shortly sketch the derivation of these results. Consider a three-dimensional region in the metal of length $L = N\,a$, where a is the distance of neighbouring ions in the lattice and $N \gg 1$ an integer. The Schrödinger equation with potential $V = 0$ and with periodic boundary conditions $\psi(x + L, y, z) = \psi(x, y, z)$ etc. is solved by

$$\psi(x, y, z) = L^{-3/2} \exp(i(k_1 x + k_2 y + k_3 z)) \tag{22}$$

where the components of the wave vector $\vec{k}$ are given by

$$k_j = n_j \frac{2\pi}{L} \quad \text{with} \quad n_j = 0, \pm 1, \pm 2, ... \tag{23}$$

The electron momentum is $\vec{p} = \hbar\vec{k}$, the energy is $E = \hbar^2\vec{k}^2/(2m_e)$. 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.

In the ground state of the metal the energy levels are filled with two electrons each starting from the lowest level. The highest energy level reached is called the Fermi E_F. At temperature $T \to 0$ all states below E_F are occupied, all states above E_F are empty. The highest momentum is called the 'Fermi momentum' $p_F = \sqrt{2m_e E_F}$, the highest velocity is the Fermi velocity $v_F = p_F/m_e$ which is in the order of 10^6 m/s. In the momentum state representation, the occupied states are located inside the 'Fermi sphere' of radius p_F, the empty states are outside.

What are the consequences of the Pauli principle 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

$$\delta p_x = eE_0\,\delta t\ . \tag{24}$$

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 an 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

$$C_e \approx \frac{3}{2} n k_B \cdot \frac{k_B T}{E_F} \, .$$

This explains the linear temperature dependence of the electronic specific heat, see eq. (19).

4.12 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. 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 resistance is nevertheless due to collisions but the collision centres are not the ions in the regular crystal lattice but only the imperfections of this lattice: 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 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.

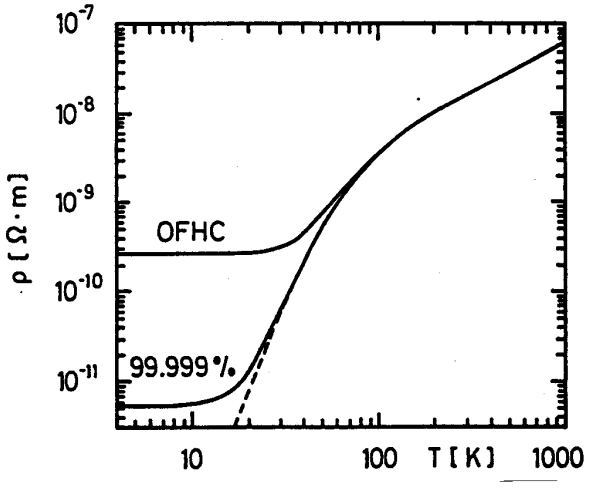


Fig. 19: Temperature dependence of the resistivity of OFHC (oxygen-free high conductivity) copper and of 99.999% pure annealed copper. Plotted as a dashed line is the calculated resistivity of copper without any impurities and lattice defects (after M.N. Wilson [17]).

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 , $E_{pair} < 2E_F$. The mathematics of Cooper pair formation will be outlined in Appendix B.

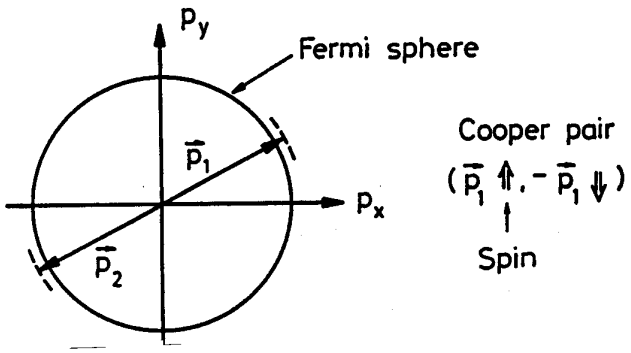


Fig. 20: A pair of electrons of opposite momenta added to the full Fermi sphere.

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: a metal ball is placed on an elastic membrane and deforms the membrane such that a potential well is created. A second ball will feel this potential well and will be attracted by it. So effectively, the deformation of the elastic membrane causes an attractive force between the two balls which would otherwise not notice each other. This visualisation of a Cooper-pair is well known in the superconductivity community (see e.g. [1]) but it has the disadvantage that it is a static picture.

I prefer the following dynamic picture: suppose 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. The Loipe 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

$$d \approx v_F \frac{2\pi}{\omega_D} \approx 100 - 1000\,\mathrm{nm}\ . \tag{25}$$

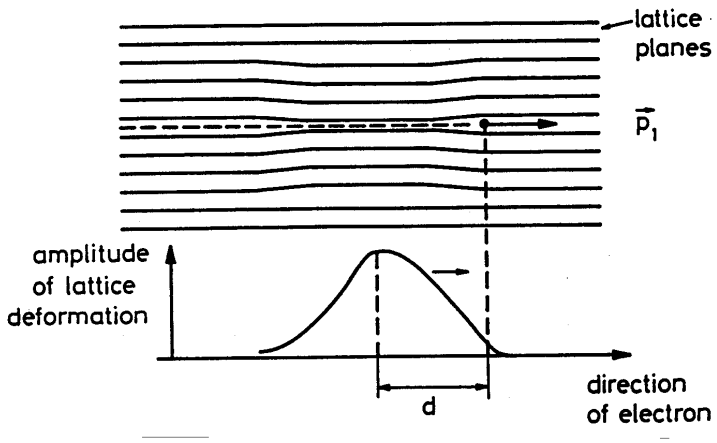


Fig. 21: Dynamical deformation of the crystal lattice caused by the passage of a fast electron. (After Ibach, L¨uth [19]).

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 other composite systems like atomic nuclei or atoms which are tightly bound objects and well-separated from another. 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.

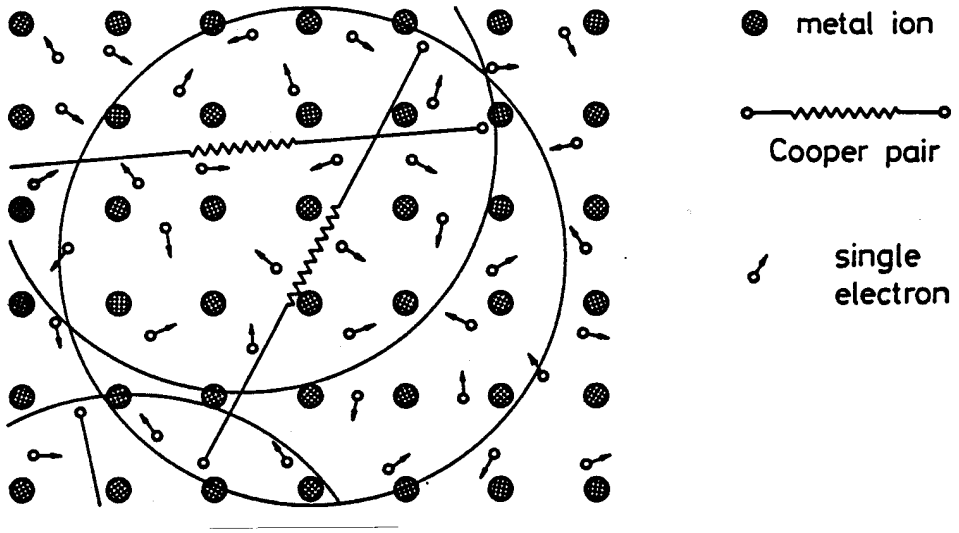


Fig. 22: Visualization of Cooper pairs and single electrons in the crystal lattice of a superconductor. (After Essmann and Tr¨auble [12]).

4.3 Elements of the BCS theory

After Cooper had proved that two electrons added to the filled Fermi sphere are 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

$$\hbar\omega_D = k_B\Theta_D \approx 0.01 - 0.02 \text{ eV} . \tag{26}$$

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 the light wave in a laser in which the photons are all in phase and have the same wavelength, direction and polarisation. The macroscopic photon wave function is the vector potential from which one can derive the electric and magnetic field vectors (see sect. 5.2).

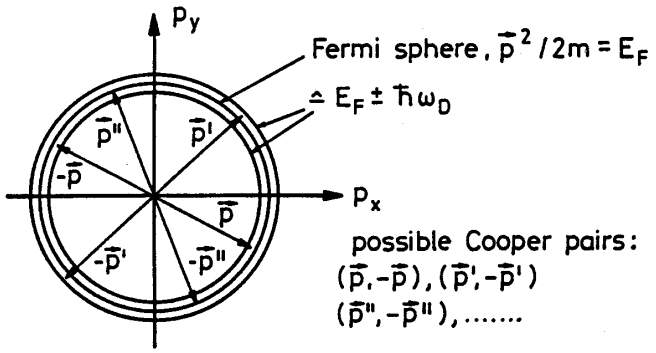


Fig. 23: Various Cooper pairs $(\vec{p}, -\vec{p})$, $(\vec{p}', -\vec{p}')$, $(\vec{p}'', -\vec{p}'')$, ... in momentum space.

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 $\vec{p}$, $\vec{p}'$, $\vec{p}''$, ... 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, see Fig. 24.

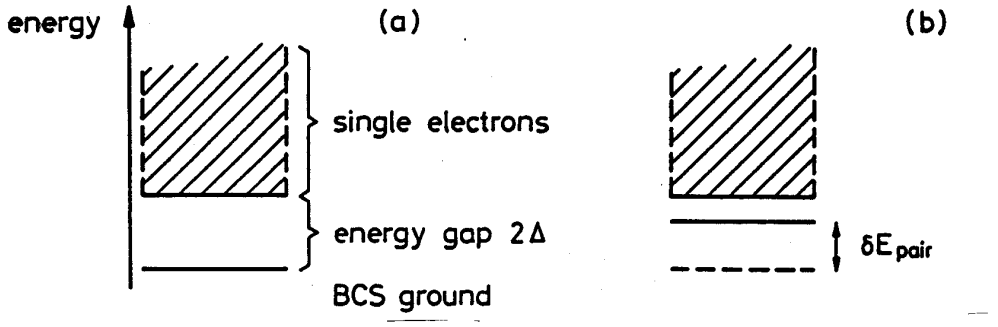


Fig. 24: (a) Energy gap between the BCS ground state and the single-electron states. (b) Reduction of energy gap in case of current fbw.

There is a certain similarity with the energy gap between the valence band 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.

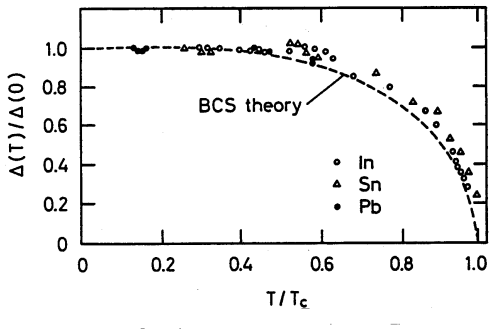


Fig. 25: Temperature dependence of the energy gap according to the BCS theory and comparison with experimental data.

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:

$$\Delta(0) = 2\hbar\omega_D \exp\left(-\frac{1}{V_0 \mathcal{N}(E_F)}\right) . \tag{27}$$

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. (27) 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. The critical temperature is given by a similar expression

$$k_B T_c = 1.14\,\hbar\omega_D \exp\left(-\frac{1}{V_0 \mathcal{N}(E_F)}\right) . \tag{28}$$

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

$$\Delta(0) = 1.76\, k_B T_c . \tag{29}$$

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. (28) 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 from 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.

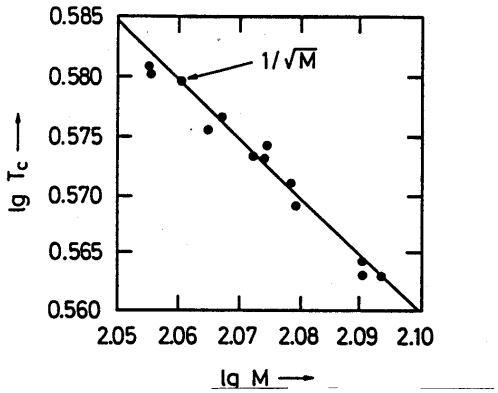


Fig. 26: The critical temperature of various tin isotopes.

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 sect. 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. 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

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

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

$$\vec{J}_s = -n_c\frac{e\hbar}{m_e}\vec{K}\ . \tag{30}$$

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 $\exp(i\vec{K}\cdot\vec{R})$ where $\vec{R} = (\vec{r}_1 + \vec{r}_2)/2$ 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 remain an energy gap provided the kinetic-energy gain $\delta E_{\rm pair}$ 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 and are then broken up into single electrons. 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 is negligible as long as the average thermal energy is smaller than 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 ($2\Delta - \delta E_{pair} > 0$) and the temperature is sufficiently low ($T < T_c(J_s)$).

The supercurrent density is limited by the condition that the energy gain $\delta E_{\rm pair}$ must be less than the energy gap. This leads to the concept of the *critical current density* J_c. The energy of the Cooper pair is, after application of the electric field,

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

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

$$J_s \leq J_c \approx 2e\, n_c\, \Delta/p_F\ . \tag{31}$$

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:

$$H_c(T) = \lambda_L\, J_c(T) \approx \lambda_L\, 2en_c\, \Delta(T)/p_F\ . \tag{32}$$

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 an electrical insulator while a silicon crystal 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 band 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: diamond has no 'resistance' for light. (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).

Finally, I want to give an example for frictionless current flow. The hexagonal benzene molecule C_6H_6 is formed by covalent binding and contains 24 electrons which are localised in σ bonds in the plane of the molecule and 6 electrons in π bonds below and above this plane. The π electrons can move freely around the ring. By a time-varying magnetic field a ring current is induced (benzene is a diamagnetic molecule) which will run forever unless the magnetic field is changed. This resembles closely the operation of a superconducting ring in the persistent mode (see Fig. 27).

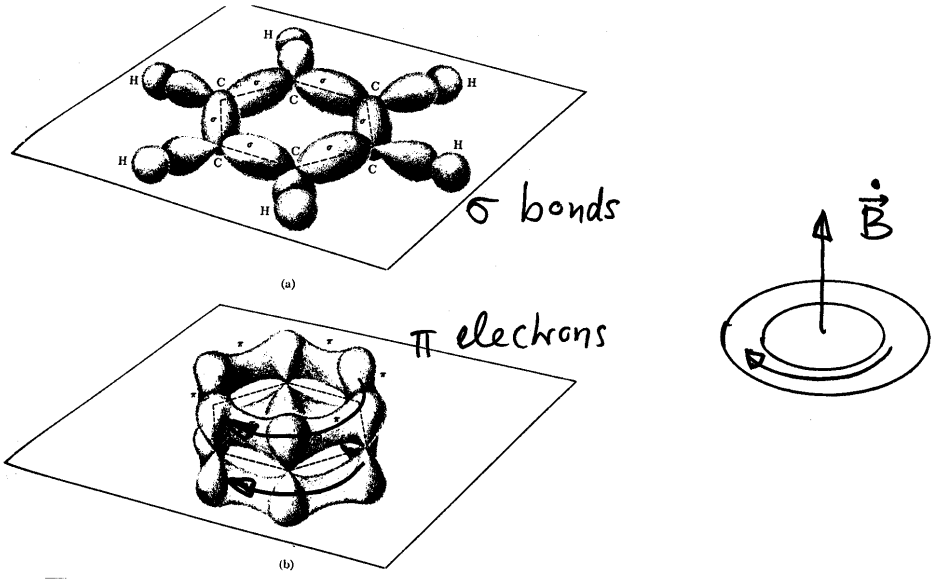


Fig. 27: Persistent ring currents in a benzene molecule and in a superconducting ring which have been induced by a rising fi eld B_z.

5. QUANTISATION OF MAGNETIC FLUX

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:

$$\Phi_{mag} = \int \int \vec{B} \cdot d\vec{F} = \oint \vec{A} \cdot \vec{ds} \, . \tag{33}$$

We apply this to the solenoidal coil sketched in Fig. 28.

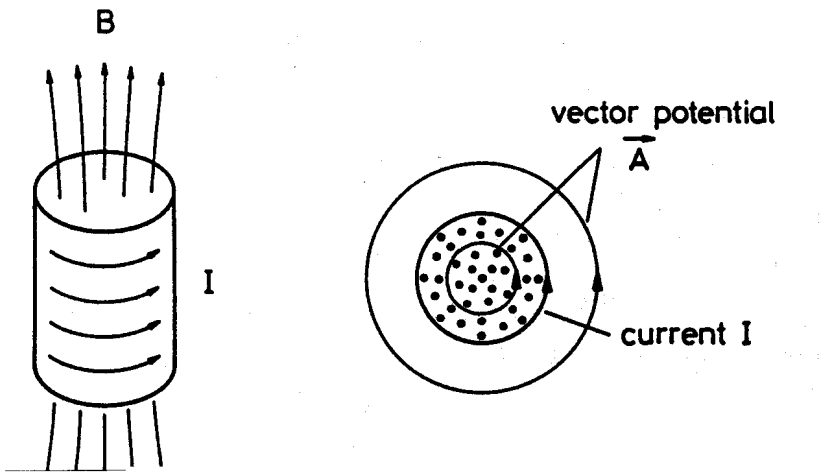


Fig. 28: Magnetic fi eld and vector potential of a solenoid.

The magnetic field has a constant value $B = B_0$ inside the solenoid and vanishes outside if we 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. (33):

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

Evaluating $\vec{B} = \vec{\nabla} \times \vec{A}$ in cylindrical coordinates 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 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 $r > R$ 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 scalar 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

$$\lambda = \frac{2\pi\hbar}{p} \ . \tag{34}$$

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'

$$\vec{p} = m\vec{v} + q\vec{A} \tag{35}$$

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

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

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. 29. 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

$$\delta\varphi = \delta\varphi_0 + \frac{e}{\hbar}\oint \vec{A} \cdot \vec{ds} = \delta\varphi_0 + \frac{e}{\hbar}\Phi_{mag} \ . \tag{36}$$

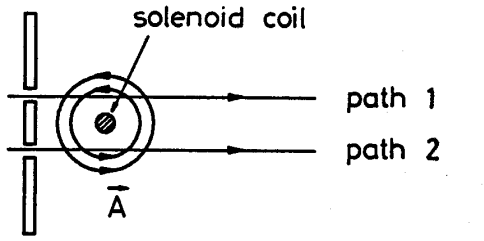


Fig. 29: Schematic arrangement for observing the phase shift due to a vector potential.

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. 30. 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 recording the interference pattern is moved in the vertical direction. Thereby the moving fringes are depicted as inclined lines. The observed shifts are in quantitative agreement with the prediction of Eq. (36).

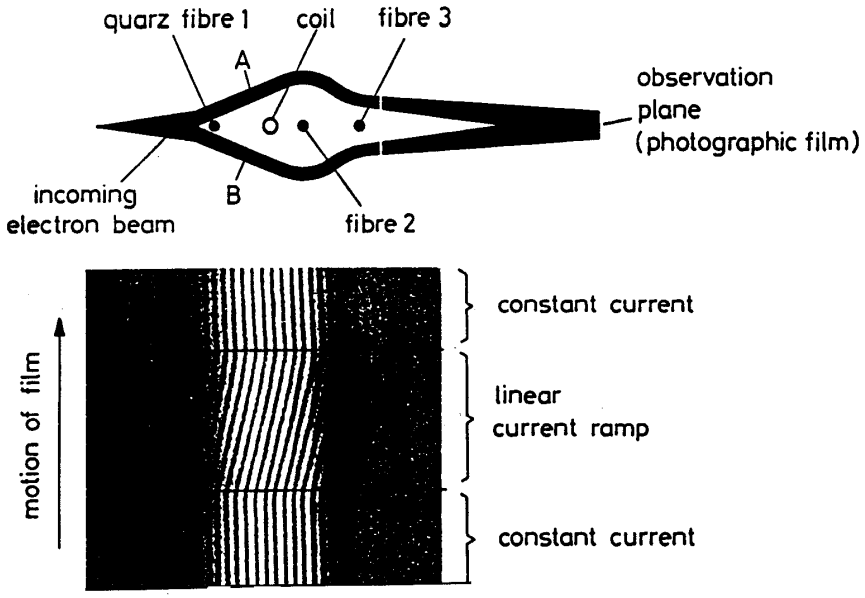


Fig. 30: Sketch of the M¨ollenstedt-Bayh experiment and observed interference pattern.

An interesting special case is the phase shift $\delta\varphi = \pi$ that interchanges bright and dark fringes. According to Eq. (36) this requires a magnetic flux

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

which turns out to be identical to the elementary flux quantum in superconductors, see sect. 5.3. 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. 31). 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.

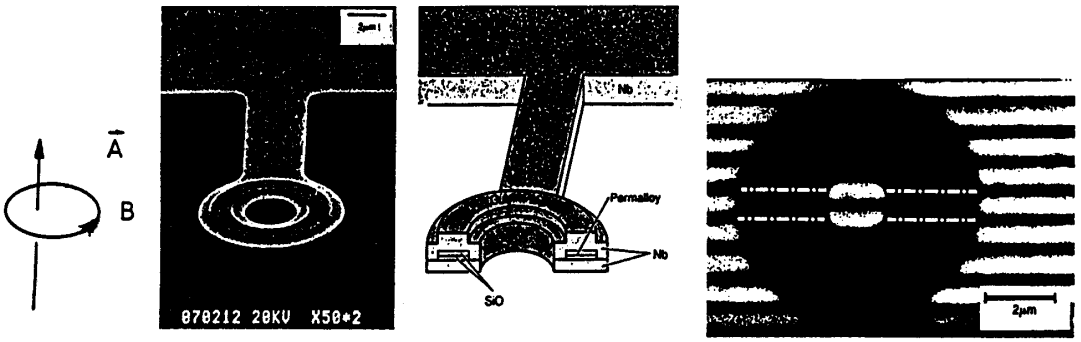


Fig. 31: Observation of Aharonov-Bohm effect using electron holography (after Tonomura [22]). The permanent toroidal magnet, encapsuled in superconducting niobium, and the observed interference fringes are shown.

5.3 Flux quantisation

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. 32. 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

$$\Phi_{\rm mag} = n\,\Phi_0\,, \quad n = 0, 1, 2, \ldots \tag{37}$$

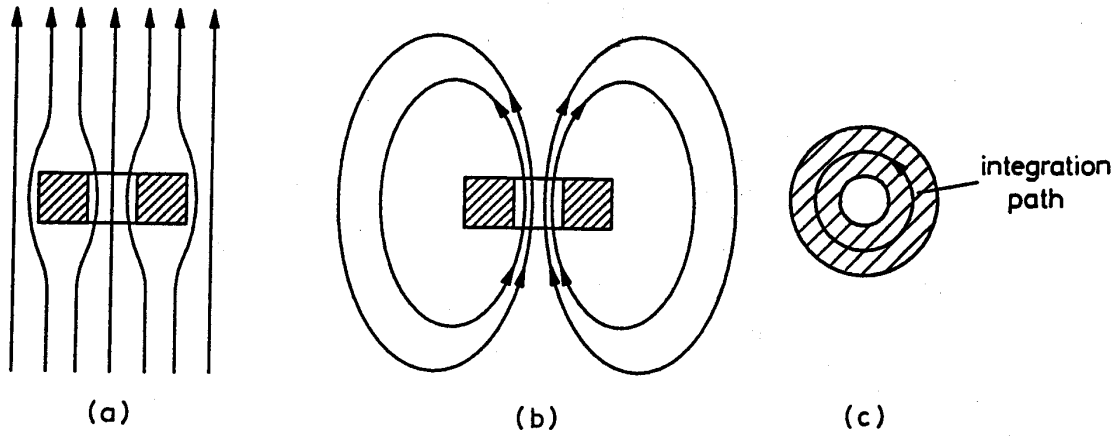


Fig. 32: Trapping of magnetic flux in a ring. First the normal-conducting ring ($T > T_c$) is placed in a magnetic field, then it is cooled down (a) and finally the field is switched off (b). The integration path is shown in part (c).

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

$$\Phi_0 = \frac{h}{2e} \tag{38}$$

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

5.31 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

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

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. (33) we see that the magnetic flux enclosed by the circular path is

$$\Phi_{mag} = \oint \vec{A} \cdot \vec{ds} = \frac{\hbar}{q} \cdot n \cdot 2\pi = n \cdot \Phi_0 \quad \Rightarrow \quad \Phi_0 = \frac{2\pi\hbar}{q} \ . \tag{39}$$

In the BCS theory we have $q = -2e$ and hence $\Phi_0 = h/(2e)$.

5.32 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. 33. 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 and hence 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.

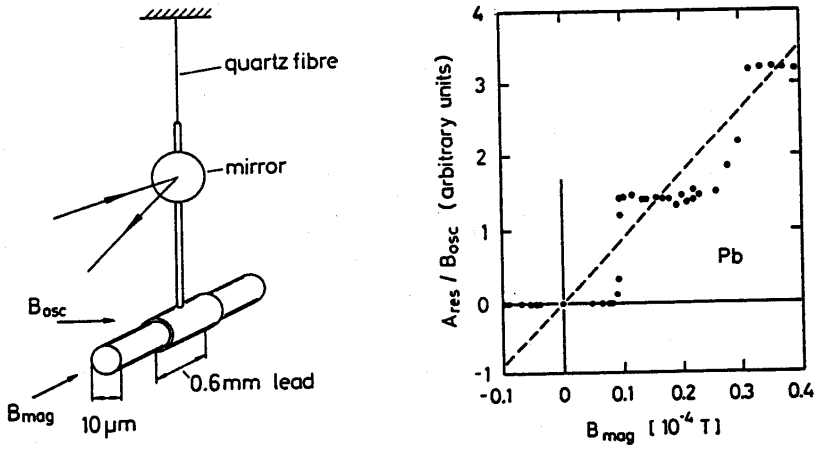


Fig. 33: Observation of flux quantisation [23].

The BCS theory is not directly applicable to high-T_c superconductors[2] which are basically two-dimensional superconductors. In Fig. 34 the flux through a $YBa_2Cu_3O_7$ ring with a weak joint is shown. Flux jumps due to external field variations occur in multiples of $h/2e$ which is an indication that some kind of Cooper pairing is also responsible for the superconductivity in these materials.

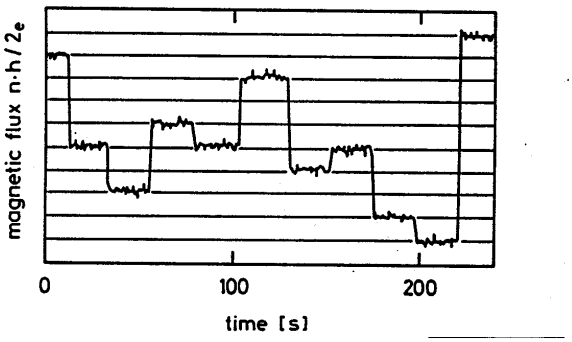


Fig. 34: Flux through an $YBa_2Cu_3O_7$ ring with a weak link [26].

In another interesting experiment a lead strip was bent into ring shape and closed via an intermediate $YBa_2Cu_3O_7$ piece. It was possible to induce a persistent ring current in this combined system of a low-T_c and a high-T_c superconductor.

5.4 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 quadratic or 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

[2]For a review of high-T_c superconductors see ref. [25] and the lectures by R. Fluekiger.

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.

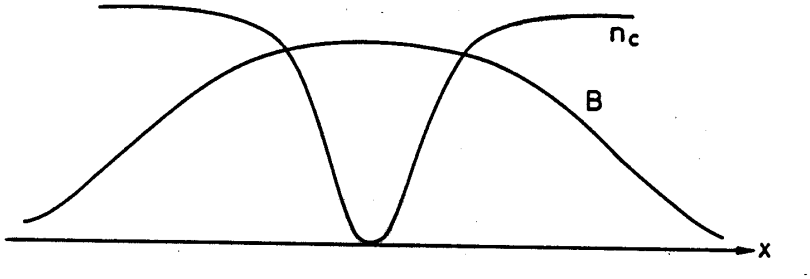


Fig. 35: Schematic cross section of a fluxoid.

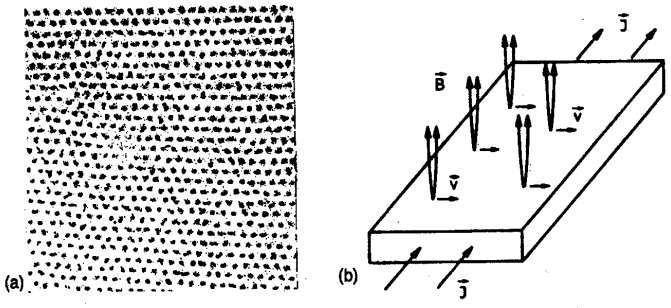


Fig. 36: (a) Fluxoid pattern in niobium (courtesy U. Essmann). The distance between adjacent flux tubes is 0.2 μm. (b) Fluxoid motion in a current-carrying type II superconductor.

The area occupied by a fluxoid is roughly given by $\pi\xi^2$ where ξ is the coherence length. An estimate of the upper critical field is derived from the condition that the fluxoids start touching each other:

$$B_{c2} \approx \frac{\Phi_0}{2\pi\xi^2} \; . \tag{40}$$

An important experimental step was the direct observation of the fluxoid pattern. Essmann and Träuble [12] developed a 'decoration' technique for this purpose. A superconductor sample was cooled by a liquid helium bath with the surface sticking out of the liquid. Iron was evaporated at some distance from the superconductor, and in the helium gas 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. 36a 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.

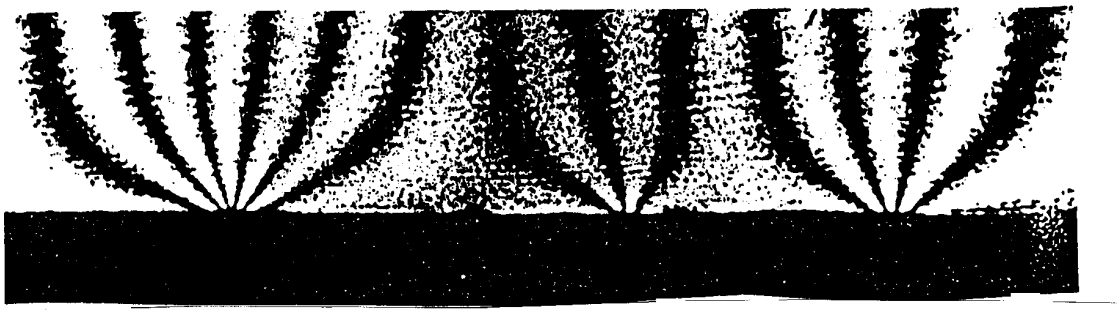


Fig. 37: Holographic image of the magnetic flux lines through a thin lead plate [22].

6. Hard Superconductors

6.1 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 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, 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. The current density $\vec{J}$ exerts a Lorentz force on the flux lines. The 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 (Fig. 36b). 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

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

It is easy to see that $\vec{E}_{equiv}$ 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].

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.

6.2 Magnetisation of a hard superconductor

A type I superconductor shows a reversible response[3] to a varying external magnetic field H. The magnetization is given by the linear relation $\vec{M}(\vec{H}) = -\vec{H}$ for $0 < H < H_c$ and then drops to zero, see Fig. 10. An ideal type II conductor without any flux pinning should also react reversibly. A hard superconductor, on the other hand, is only reversible in the Meissner phase because then no magnetic field enters the bulk, so no flux pinning can happen. If the field is raised beyond H_{c1} magnetic flux enters the sample and is captured at pinning centres. When the field is reduced again these flux lines 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 initial state ($H = 0$ and no captured flux in the bulk material) can only be recovered by warming up the specimen to destroy superconductivity and release all pinned flux quanta, and by cooling down again.

A typical hysteresis curve is shown in Fig. 38a. 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 underlying physical process is diamagnetism. In an accelerator the field is usually not inverted and then the hysteresis has the shape plotted in Fig. 38b.

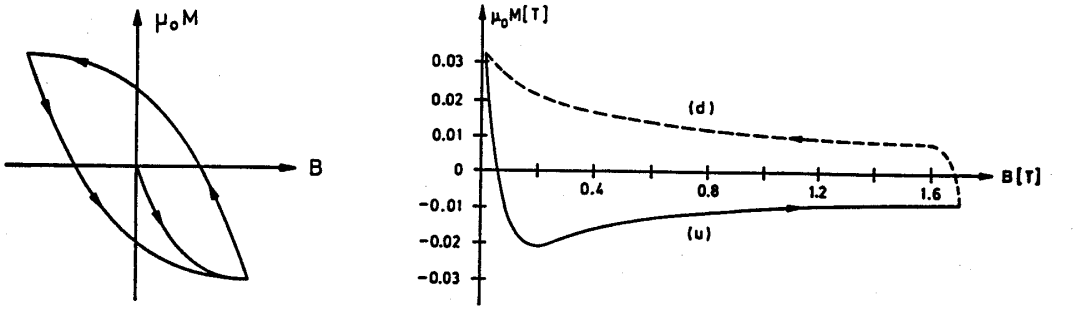


Fig. 38: (a) Magnetic hysteresis of a hard superconductor. (b) Magnetisation hysteresis for the field cycle of accelerator magnets.

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. 39 the sextupole field B_3 at a distance of 25 mm from the dipole axis is plotted as a function of the dipole field $B_1 = \mu_0 H_1$ 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.

[3]This statement applies only for long cylindral or elliptical samples oriented parallel to the field.

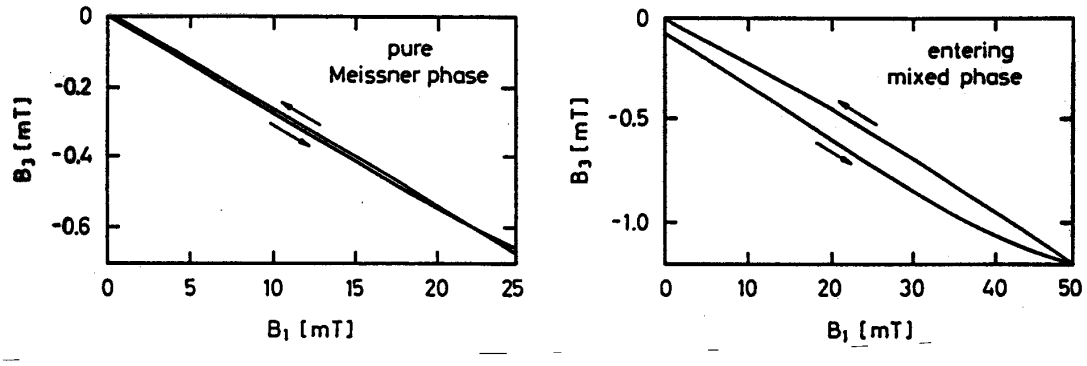


Fig. 39: Sextupole field in a HERA dipole in the Meissner phase and slightly above.

With increasing field the hysteresis widens more and more and is eventually nearly symmetric to the horizontal axis. The sextupole hysteresis observed in the standard field cycle at HERA is plotted in Fig. 40a. A similar curve is obtained for the 12-pole in the quadrupoles.

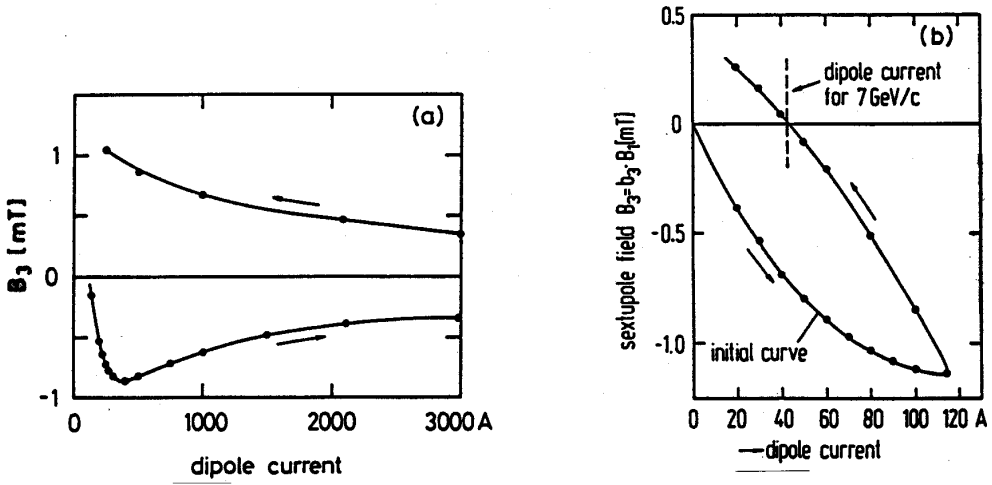


Fig. 40: (a) The sextupole component in the HERA dipoles for the standard field cycle 4.7 T → 0.05 T → 4.7 T. (b) Sextupole field for the first beam test with positrons of 7 GeV.

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 → 112 A → 42.5 A was performed which resulted in an almost vanishing sextupole (see Fig. 40b). A similar procedure was used in the first run with 40 GeV protons, this time with the loop 0 → 314 A → 245 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.

6.3 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 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. 41a. 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. 41b).

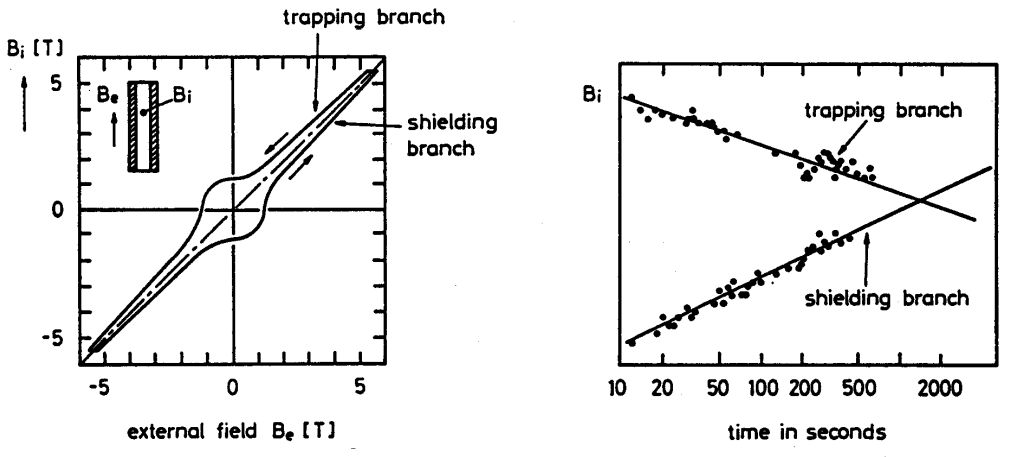


Fig. 41: (a) Hysteresis of the internal field in a tube of hard superconductor. (b) Time dependence of the internal field on the trapping and the shielding branch [29].

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 $\Phi_{av} = n\Phi_0$ 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 force density $F \propto \Phi_{av} J$. This slope reduces the effective potential well depth to $U = U_0 - \Delta U$with $\Delta U \approx \Phi_{av} J a \, l$, see Fig. 42. Here l is the length of the flux bundle. The probability for flux escape increases

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

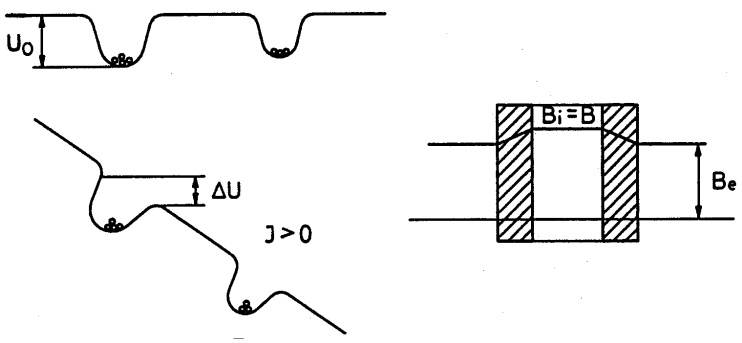


Fig. 42: Sketch of the pinning potential without and with current fbw and fi eld profi le across the NbZr tube.

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 $B_{int} - B_{mext} = \mu_0 J_c\, w$ where J_c is the critical current density at the given temperature and magnetic field and w the wall thickness. Under the assumption $B_{int} - B_{ext} \ll B_{ext}$ 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

$$\frac{dJ_c}{dt} \approx -C \exp\left(\frac{\Delta U}{k_B T}\right) \approx -C \exp\left(\frac{\Phi_{av} a J_c l}{k_B T}\right) \tag{41}$$

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

$$J_c(t) = J_c(0) - \frac{k_B T}{\Phi_{av} a\, l} \ln t\ . \tag{42}$$

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 e.g. [31]. So it seems tempting to 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. 43a.

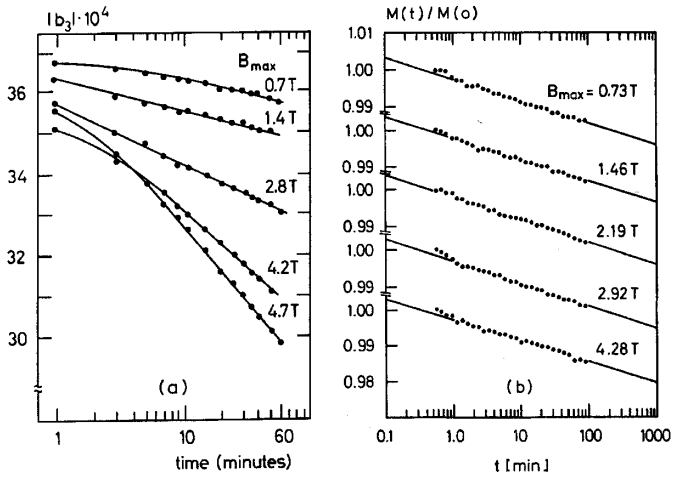


Fig. 43: (a) Decay of the sextupole coefficient in a HERA dipole at a field of 0.23 T for different values of the maximum field in the initialising cycle $0 \rightarrow B_{max} \rightarrow 0.04$T $\rightarrow$ 0.23 T [32]. (b) Magnetisation decay at zero field in a long sample of HERA cable for different values of the maximum field in the initialising cycle $0 \rightarrow B_{max} \rightarrow 0$ [33].

In cable samples this is not the case as is evident from Fig. 43b. 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.

From the data in Fig. 43 it is obvious that thermally activated flux creep can explain only part of the time dependence of multipoles in magnets. The decay rates measured in magnets are usually much larger than those in cable samples, and there is a considerable variation from magnet to magnet. In 1995 experimental results [34] and model calculations [35] were presented showing that the time dependence is due to a complex interplay between magnetisation currents in the NbTi filaments and eddy currents among the strands of the cable. Quantitative predictions are not possible because of too many unknown parameters. For a more detailed discussion see [31].

Flux creep has become an important issue after the discovery of high-temperature superconductors. Figure 44 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 for building magnets cooled by liquid nitrogen.

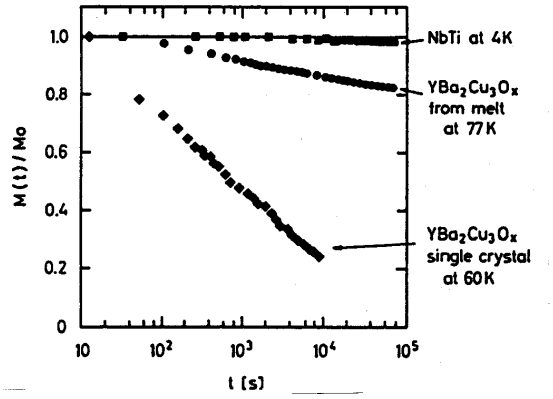


Fig. 44: Comparison of superconductor magnetisation decay due to flux creep in NbTi at a temperature of 4.2 K, in oriented-grained $YBa_2Cu_3O_x$ at 77 K and in a $YBa_2Cu_3O_x$ single crystal at 60 K [36].

7. SUPERCONDUCTORS IN MICROWAVE FIELDS

Superconductivity in microwave fields is not treated adequately in standard text books. For this reason I present in this section a simplified explanation of the important concepts. A similar treatment can be found in [37]. Superconductors are free from energy dissipation in direct-current (dc) applications, but this is no longer true for alternating currents (ac) and particularly not in microwave fields. The reason is that the high-frequency magnetic field penetrates a thin surface layer and induces oscillations of the electrons which are not bound in Cooper pairs. The power dissipation caused by the motion of the unpaired electrons can be characterized by a surface resistance. In copper cavities the surface resistance is given by

$$R_{surf} = \frac{1}{\delta\sigma} \tag{43}$$

where δ is the skin depth and σ the conductivity of the metal.

The response of a superconductor to an ac field can be understood in the framework of the *two-fluid model* [4]. An ac current in a superconductor is carried by Cooper-pairs (the superfluid component) as well as by unpaired electrons (the normal component). Let us study the response to a periodic electric field. The normal current obeys Ohm's law

$$J_n = \sigma_n \, E_0 \, \exp(-i\omega t) \tag{44}$$

while the Cooper pairs receive an acceleration $m_c \, \dot{v}_c = -2e \, E_0 \, \exp(-i\omega t)$, so the supercurrent density becomes

$$J_s = i \, \frac{n_c \, 2 \, e^2}{m_e \omega} E_0 \, \exp(-i\omega t) \; . \tag{45}$$

If we write for the total current density

$$J = J_n + J_s = \sigma E_0 \exp(-i\omega t) \tag{46}$$

[4] A similar model is used to explain the unusual properties of liquid helium below 2.17 K in terms of a normal and a superfluid component.

we get a complex conductivity:

$$\sigma = \sigma_n + i\sigma_s \quad \text{with} \quad \sigma_s = \frac{2\,n_c e^2}{m_e\,\omega} = \frac{1}{\mu_0 \lambda_L^2\,\omega}\ . \tag{47}$$

The surface resistance is the real part of the complex surface impedance

$$R_{surf} = Re\left(\frac{1}{\lambda_L(\sigma_n + i\sigma_s)}\right) = \frac{1}{\lambda_L}\cdot\frac{\sigma_n}{\sigma_n^2 + \sigma_s^2}\ . \tag{48}$$

Since $\sigma_n^2 \ll \sigma_s^2$ at microwave frequencies one can disregard σ_n^2 in the denominator and obtains $R_{surf} \propto \sigma_n/(\lambda_L\sigma_s^2)$. So we arrive at the surprising result that the microwave surface *resistance* is proportional to the normal-state *conductivity*.

The conductivity of a normal metal is given by the classic Drude expression

$$\sigma_n = \frac{n_n e^2 \ell}{m_e v_F} \tag{49}$$

where n_n is the density of the unpaired electrons, ℓ their mean free path and v_F the Fermi velocity. The normal electrons are created by thermal breakup of Cooper pairs. There is an energy gap $E_g = 2\Delta(T)$ between the BCS ground state and the free electron states. By analogy with the conductivity of an intrinsic (undoped) semiconductor we get

$$\sigma_n \ \propto\ \ell\,\exp(-E_g/(2k_BT)) = \ell\,\exp(-\Delta(T)/(k_BT))\ . \tag{50}$$

Using $1/\sigma_s = \mu_0\lambda_L^2\omega$ and $\Delta(T) \approx \Delta(0) = 1.76 k_B T_c$ we finally obtain for the BCS surface resistance

$$R_{BCS} \ \propto\ \lambda_L^3\,\omega^2\,\ell\,\exp(-1.76\,T_c/T)\ . \tag{51}$$

This formula displays two important aspects of microwave superconductivity: the surface resistance depends exponentially on temperature, and it is proportional to the square of the radio frequency.

Eq. (51) applies if the mean free path ℓ of the unpaired electrons is much larger than the coherence length ξ. In niobium this condition is usually not fulfilled and one has to replace λ_L in the above equation by [38]

$$\Lambda = \lambda_L\sqrt{1 + \xi/\ell}\ . \tag{52}$$

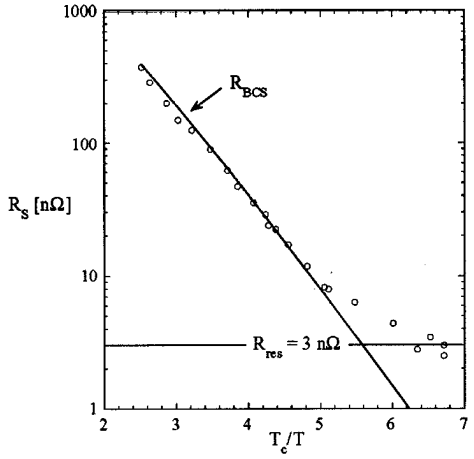


Fig. 45: The surface resistance of a 9-cell TESLA cavity plotted as a function of T_c/T. The residual resistance of 3 nΩ corresponds to a quality factor $Q_0 = 10^{11}$.

Combining equations (51) and (52) we arrive at the surprising statement that the surface resistance does not assume its minimum value when the superconductor is of very high purity ($\ell \gg \xi$) but rather in the range $\ell \approx \xi$. Experimental results [39] and theoretical models [40] confirm this prediction. The effect is also observed in copper cavities with a thin niobium sputter coating in which the electron mean free path is in the order of ξ. At 4.2 K the quality factors in the LEP cavities are indeed a factor of two higher than in pure niobium cavities [41].

In addition to the BCS term there is a residual resistance caused by impurities, frozen-in magnetic flux or lattice distortions.

$$R_{surf} = R_{BCS} + R_{res} \ . \tag{53}$$

R_{res} is temperature independent and amounts to a few nΩ for a clean niobium surface but may readily increase if the surface is contaminated.

For niobium the BCS surface resistance at 1.3 GHz amounts to about 800 nΩ at 4.2 K and drops to 15 nΩ at 2 K, see Fig. 45. The exponential temperature dependence is the reason why operation at 2 K is essential for achieving high accelerating gradients in combination with very high quality factors. Superfluid helium is an excellent coolant owing to its high heat conductivity.

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.

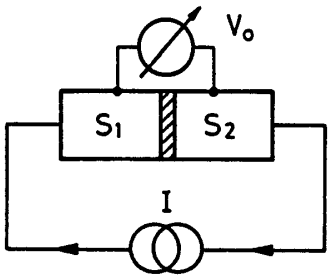


Fig. 46: Schematic arrangement for studying the properties of a Josephson junction.

DC Josephson effect. If the voltage V_0 across the junction is zero there is a dc Cooper-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.

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

$$I(t) = I_0 \sin(2\pi f_J t + \varphi_0) \tag{54}$$

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

$$f_J = \frac{2eV_0}{2\pi\hbar} \ . \tag{55}$$

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 (55) 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 barrier the two Schrödinger equations are coupled

$$i\hbar\frac{\partial\psi_1}{\partial t} = E_1\psi_1 + K\psi_2\,, \qquad i\hbar\frac{\partial\psi_2}{\partial t} = E_2\psi_2 + K\psi_1\,. \tag{56}$$

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

$$\psi_1 = \sqrt{n_1}\exp(i\varphi_1)\,, \quad \psi_2 = \sqrt{n_2}\exp(i\varphi_2)\,. \tag{57}$$

We insert this into (56) 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:

$$\begin{aligned}
\dot{n}_1 &= \frac{2K}{\hbar}\sqrt{n_1 n_2}\,\sin(\varphi_2-\varphi_1)\,,\\
\dot{n}_2 &= \frac{2K}{\hbar}\sqrt{n_1 n_2}\,\sin(\varphi_1-\varphi_2) = -\dot{n}_1\,,\\
\dot{\varphi}_1 &= -\frac{1}{\hbar}\left[E_1 + K\sqrt{\frac{n_2}{n_1}}\cos(\varphi_2-\varphi_1)\right]\,,\\
\dot{\varphi}_2 &= -\frac{1}{\hbar}\left[E_2 + K\sqrt{\frac{n_1}{n_2}}\cos(\varphi_1-\varphi_2)\right]\,.
\end{aligned} \tag{58}$$

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

$$\begin{aligned}
\dot{n}_1 &= \frac{2K}{\hbar}n_1\sin(\varphi_2-\varphi_1) = -\dot{n}_2\,,\\
\frac{d}{dt}(\varphi_2-\varphi_1) &= -\frac{1}{\hbar}(E_2-E_1) = \frac{2eV_0}{\hbar}\,.
\end{aligned} \tag{59}$$

Integrating the second equation (59) yields the Josephson frequency

$$\varphi_2(t) - \varphi_1(t) = \frac{2eV_0}{\hbar}\cdot t + \varphi_0 = 2\pi f_J\cdot t + \varphi_0\,. \tag{60}$$

The Cooper-pair current through the junction is proportional to $\dot{n}_1$. Using (59) and (60) it can be written as

$$I(t) = I_0\sin\left(\frac{2eV_0}{\hbar}\,t + \varphi_0\right)\,. \tag{61}$$

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 $\Phi_{\rm mag}$ threads the area of the loop, the phases differ according to Sec. 5 by

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

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

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

and the current is

$$I = I_0 \sin\phi_0 \ \cos\left(\frac{e}{\hbar}\Phi_{mag}\right) \ . \tag{62}$$

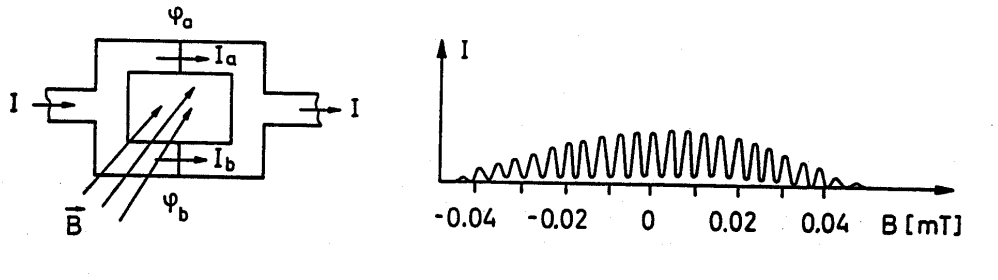


Fig. 47: A loop with two Josephson junctions and the observed interference pattern [42]. The amplitude modulation is caused by the fi nite width of the junctions.

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 $\Delta\Phi_{mag} = \Phi_0$, 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.

A FREE ENERGY IN THERMODYNAMICS

To illustrate the purpose of the free energy I consider first an ideal gas. The internal energy is the sum of the kinetic energies of all atoms

$$U = \sum_{i=1}^{N} \frac{m}{2} v_i^2 = \frac{3}{2}\, N\, k_B T \tag{63}$$

and depends only on temperature but not on volume. The first law of thermodynamics describes energy conservation:

$$dU = \delta Q + \delta W \ . \tag{64}$$

The internal energy increases either by adding heat δQ or mechanical work $\delta W = -p\,dV$ to the gas. For a reversible process one has $\delta Q = T\,dS$ where S is the entropy. Now consider an isothermal expansion of the gas. Thereby the gas transforms heat into mechanical work:

$$dU = 0 \quad \text{for} \quad T = \text{const} \quad \Rightarrow \quad \delta Q = -\delta W = p\,dV\ . \tag{65}$$

The gas produces mechanical work but its internal energy does not change, hence U is not an adequate variable to describe the process. What is the correct energy variable? We will see that this is the Helmholtz free energy, given by

$$F = U - T\,S \quad \Rightarrow \quad dF = dU - SdT - TdS = \delta W - SdT\ . \tag{66}$$

For an isothermal expansion ($dT = 0$) we get $dF = \delta W$, i.e. $-dF = pdV$: the work produced by the gas is identical to the reduction of its free energy.

Now we consider a magnetic material of permeability μ inside a coil which generates a field H. The magnetization is $\vec{M} = (\mu - 1)\vec{H}$. Its potential energy (per unit volume) in the magnetic field is

$$E_{pot} = -\mu_0\,\vec{M}\cdot\vec{H}\ . \tag{67}$$

If the magnetization changes by $d\vec{M}$ the work is $\delta W = \mu_0\,d\vec{M}\cdot\vec{H}$. Defining again the Helmholtz free energy by eq. (66) we get by analogy with the ideal gas $dF = \delta W$, hence F can in fact be used to describe the thermodynamics of magnetic materials in magnetic fields. One drawback is, however, that the magnetization of a substance cannot be directly varied by the experimenter. What can be varied at will is the magnetic field H, namely by choosing the coil current. For this reason another energy function is more appropriate, the Gibbs free energy

$$G = F - \mu_0\,\vec{M}\cdot\vec{H} = U - T\,S - \mu_0\,\vec{M}\cdot\vec{H}\ . \tag{68}$$

For an isothermal process we get $dG = -\mu_0\,\vec{M}\cdot d\vec{H}$. Let us apply this to a superconductor in the Meissner phase. Then $\mu = 0$ and $\vec{M} = -\vec{H}$ from which follows

$$dG_{sup} = \mu_0\,M(H)dH = \frac{\mu_0}{2}\,d(H^2)\ \Rightarrow\ G_{sup}(H) = G_{sup}(0) + \frac{\mu_0}{2}\,H^2\ . \tag{69}$$

This equation is used in Sect. 3.

B 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 $k_F = \sqrt{2m_eE_F}/\hbar$. To the fully occupied Fermi sphere we add two electrons of opposite wave vectors $\vec{k}_1 = -\vec{k}_2$ whose energy $E_1 = E_2 = \hbar^2k_1^2/(2m_e)$ is within the spherical shell (see Figs. 20, 23)

$$E_F < E_1 < E_F + \hbar\omega_D\ . \tag{70}$$

From Sect. 4 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

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

The Schrödinger equation for the two electrons reads

$$-\frac{\hbar^2}{2m_e}(\nabla_1^2 + \nabla_1^2)\,\psi(\vec{r}_1,\vec{r}_2) + V(\vec{r}_1,\vec{r}_2)\psi(\vec{r}_1,\vec{r}_2) = E_{pair}\psi(\vec{r}_1,\vec{r}_2) \tag{71}$$

where V is the interaction potential due to the dynamical lattice polarisation. In the simple case of vanishing interaction, $V = 0$, the solution of (71) 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. (71) with $V = 0$ is a superposition of such functions

$$\psi(\vec{r}) = \frac{1}{L^3}\sum_{\vec{k}} g(\vec{k})\exp(i\vec{k}\cdot\vec{r}) \tag{72}$$

with the restriction that the coefficients $g(\vec{k})$ vanish unless $E_F \le \hbar^2k^2/2m_e \le E_F + \hbar\omega_D$. This function is certainly not an exact solution of the equation (71) 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 (72) into Eq. (71):

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

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

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

Introducing further the transition matrix elements of the potential V

$$V_{\vec{k}\vec{k}'} = \int\exp\left(-i(\vec{k}-\vec{k}')\cdot\vec{r}\right)\,V(\vec{r})d^3r \tag{73}$$

one gets a relation among the coefficients of the expansion (72)

$$g(\vec{k})\left[\frac{\hbar^2k^2}{m_e} - E_{pair}\right] = -\frac{1}{L^3}\sum_{\vec{k}'} g(\vec{k}')V_{\vec{k}\vec{k}'}\ . \tag{74}$$

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.

$$V_{\vec{k}\vec{k}'} = -V_0 \quad \text{for} \quad E_F < \frac{\hbar^2k^2}{2m_e}, \frac{\hbar^2k'^2}{2m_e} < E_F + \hbar\omega_D \tag{75}$$

and $V_{\vec{k}\vec{k}'} = 0$ elsewhere. The negative value ensures attraction. With this extreme simplification the right-hand side of Eq. (74) is no longer $\vec{k}$ dependent but becomes a constant

$$-\frac{1}{L^3}\sum_{\vec{k}'} g(\vec{k}')V_{\vec{k}\vec{k}'} = \frac{V_0}{L^3}\sum_{\vec{k}'} g(\vec{k}') = A\,. \tag{76}$$

Then Eq. (74) yields for the coefficients

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

The constant A is still unknown. We can eliminate it by summing this expression over all $\vec{k}$ and using (76) once more

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

from which follows

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

Dividing by A leads to the important consistency relation

$$1 = \frac{V_0}{L^3}\sum_{\vec{k}} \frac{1}{\hbar^2k^2/m_e - 2E_F + \delta E}\,. \tag{77}$$

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

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

formula (77) becomes

$$1 = V_0\,\mathcal{N}(E_F)\int_0^{\hbar\omega_D} \frac{d\xi}{2\xi + \delta E}\,. \tag{78}$$

The integral yields

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

The energy shift is then

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

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

$$\delta E = 2\hbar\omega_D \exp\left(-\frac{2}{V_0\,\mathcal{N}(E_F)}\right)\,. \tag{79}$$

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

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*, Oxford University Press 1997

[18] L.N. Cooper, Phys. Rev. **104**, 1189 (1956)

[19] H. Ibach and H. Lüth, *Solid State Physics*, 2nd Edition, Springer-Verlag, Berlin 1996

[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] K.-H. Mess, P. Schmüser, S. Wolff, *Superconducting Accelerator Magnets*, World Scientific, Singapore, 1996

[32] H. Brück et al., Contribution to the European Accelerator Conference, Nice 1990

[33] M. Halemeyer et al., IEEE Trans. Appl. Supercond. Vol.3, No.1, 168 (1993)

[34] A.P. Verweij and H.H.J. ten Kate, IEEE Trans. **ASC-5** (1995) 404

[35] L. Krempaski and C. Schmidt, Appl. Phys. Lett. **66** (1995) 1545, and J. Appl. Phys. **78** (1995) 5800

[36] C. Keller et al., Cryogenics **30**, 410 (1990)

[37] H. Padamsee, J. Knobloch and T. Hays, *RF Superconductivity for Accelerators*, John Wiley, New York 1998

[38] B. Bonin, CERN Accelerator School *Superconductivity in Particle Accelerators*, CERN 96-03, ed. S. Turner, Hamburg (1995)

[39] F. Palmer, Proc. 3rd Workshop on RF Superconductivity, K. Shephard (ed.), Argonne Natl. Lab. 1988, p. 309

[40] J. Halbritter, Z. Physik **238** (1970) 466

[41] C. Benvenuti *et al.*, Proc. PAC91, San Francisco (1991), p. 1023.

[42] R.C. Jaklevic et al. Phys. Rev. **140**, A1628 (1965)

PRACTICAL SUPERCONDUCTORS FOR ACCELERATORS - FILAMENTARY CONDUCTORS

Martin N Wilson
Consultant, Brook House, 33 Lower Radley, Abingdon, OX14 3AY, UK.

Abstract
All superconducting accelerators to date have used magnets wound from Rutherford cable. In this first of two lectures, we describe the filamentary superconducting wires used within a Rutherford cable. To discuss the reasons for this fine filamentation of the superconductor, we first describe the critical state model and then use it to show how persistent currents in the superconductor produce magnetization effects, which cause field errors. Coupling between the filaments is then discussed, together with the criterion for twisting to de-couple the filaments. Finally the relationship between magnetization and ac losses in changing fields is described.

1. INTRODUCTION

This is the first of two linked lectures which describe Rutherford cable, the magnet conductor which has been used in all superconducting synchrotron accelerators to date. Fig 1 shows how Rutherford cable consists of a flat twisted cable comprising ~ 30 wires, each of which contains some 5000 – 20,000 fine filaments of superconductor embedded in a matrix of copper. In this first lecture we discuss why such fine subdivision of the superconductor is necessary and how it affects the magnetization and ac loss of the magnet winding. To provide a basis for that discussion, it will first be necessary to get acquainted with the critical state model, which describes how all high field Type II superconductors react to magnetic and electric fields.

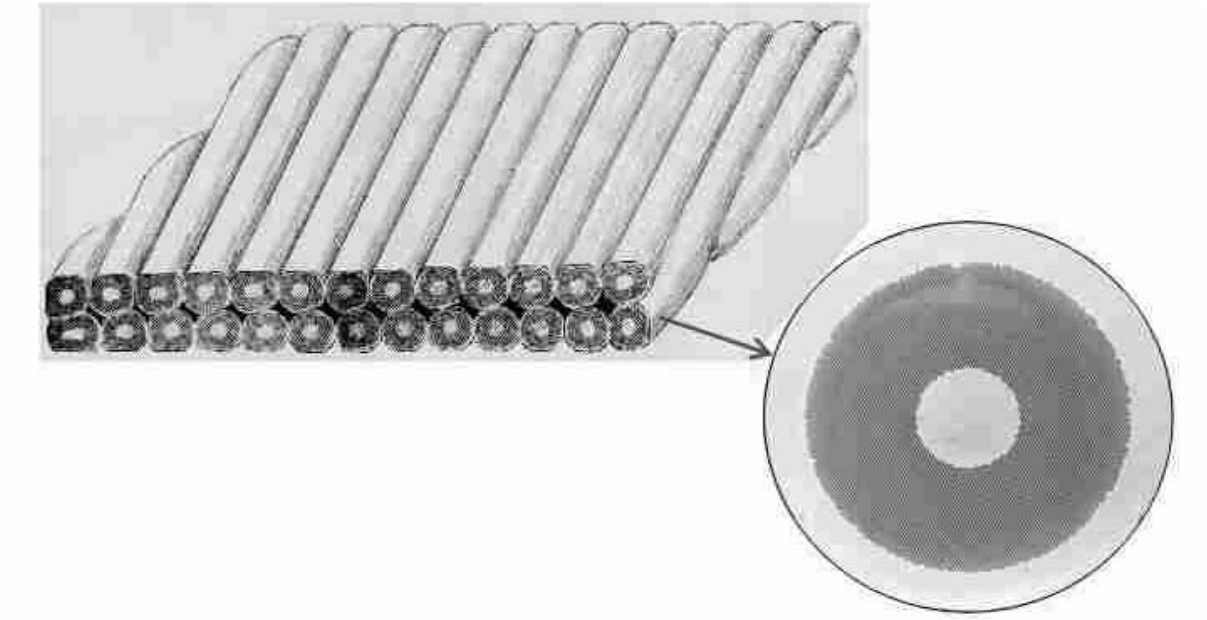


Fig. 1: Rutherford cable, inset shows a single wire comprising ~ 10,000 filaments of NbTi superconductor in a copper matrix.

2. THE CRITICAL STATE MODEL

When a superconductor is exposed to changing magnetic fields, currents are induced to flow in such a way as to shield the interior of the superconductor from the changing field. They are like eddy currents but, because the superconductor has no resistance, they do not decay with time. The path and magnitude of these currents may be predicted using the critical state model, originally proposed by C.P. Bean [1]. The critical state model may be summarized in two statements:

a) all disturbances start at the surface of the superconductor and propagate inwards.

b) all regions within the superconductor have either zero current or current flowing at critical current density J_c.

It is common to discuss problems of this nature in terms of the simplified one dimensional slab model shown in Fig. 2. All the basic physical principles may be discussed in terms of one dimensional arguments; extension to the more realistic situation in two and three dimensions entails much more complication, but provides few further insights.

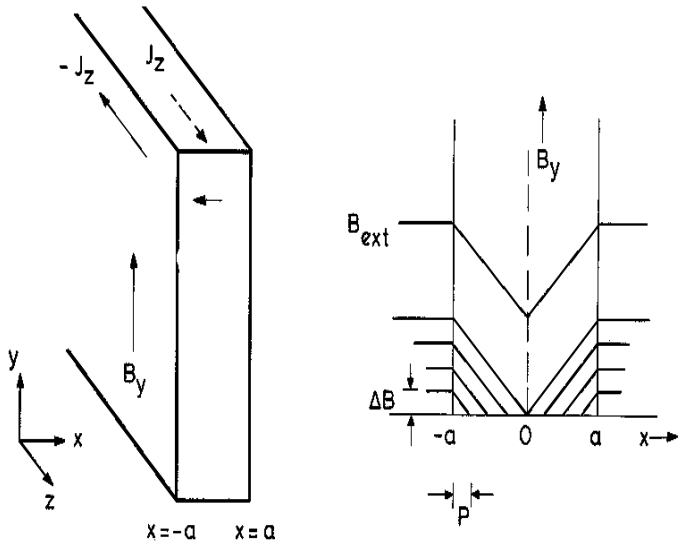


Fig. 2: A field B_y applied to a slab of superconductor induces currents J_z to flow; right hand side shows a plot of B_y through a cross section of the slab, which is infinite in the y and z directions.

As shown in Fig. 2, the increasing field B_y induces screening currents $\pm J_z$ to flow in the slab. According to the critical state model $J_z = J_c$. In one dimension we have:

$$\frac{\partial B_y}{\partial x} = -\mu_o J_z = \mu_o J_c \qquad (1)$$

Thus, when we plot magnetic field within the slab, we see a linear fall-off from the surface until the field reaches zero at the interior, after which the current density becomes zero and the field does not change. As the external field increases, the regions of $\pm J_c$ extend further into the slab until they reach the centre, at which point the slab is said to be *fully penetrated*. Further increases in field produce no change in the current pattern, merely a general increase in field throughout the slab.

Physically we may think of the process as follows. The changing magnetic field produces an electric field at the surface of the slab which drives current in the superconductor slightly beyond critical density until the resistive voltage matches the electric field. When the field change stops, current in the superconductor decays until it falls just below critical density, after which there is no reason for it to decay further, so it stays flowing at J_c.

If we now reduce the external field, patterns like those shown in Fig. 3 are induced. Once again we see the disturbance starting from the surface, where it induces negative currents, which

progressively penetrate into the slab 'wiping out' the original currents. Thus we see that the pattern of currents depends on the history and the direction in which the field is changing; the behaviour is *hysteretic*

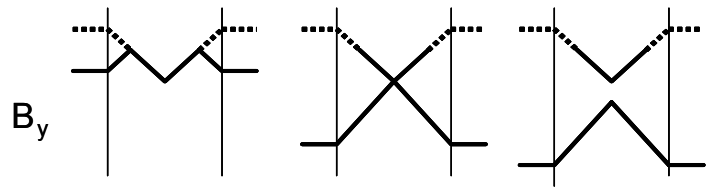


Fig 3: pattern of screening current and internal field when the external field is progressively reduced.

If the slab is thick, so that the screening currents produce a large difference in field between the surface and interior of the slab, the currents can become unstable and collapse. When this happens, magnetic flux rushes into the slab in a process known as a *flux jump.* Details of this process and the criterion of slab width below which the currents are stable, will be presented in the lectures on Stability by H. Ten Kate.

3. MAGNETIZATION

These persistent screening currents produce magnetic effects which can be detected outside the slab. In fact we can define a *magnetization* in the usual way as magnetic moment per unit volume, ie the summation of (current) × (area enclosed) per unit volume. For the infinite slab, when it is fully penetrated, we may integrate the current density across the slab, using the nomenclature sketched in Fig. 4.

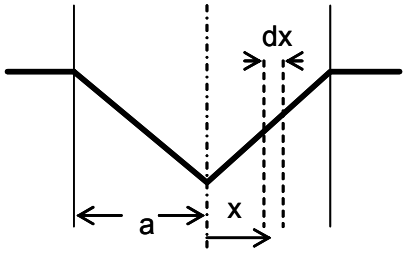


Fig. 4: Nomenclature for the magnetization calculation

Thus

$$M_{slab} = \sum_V \frac{IA}{V} = \frac{1}{a}\int_0^a J_c\, x\, dx = \frac{J_c\, a}{2} \qquad (2)$$

Note that in deriving (2), we are assuming symmetry about the chain dotted line in Fig. 3, often known as the *electric centre line*. For the more practical case of a cylindrical wire aligned perpendicular to the field, the calculation is similar but more complicated and gives the result:

$$M_{cyl} = \frac{4}{3\pi} J_c\, a \qquad (3)$$

Putting the sample into a magnetometer, we can measure its magnetization and plot a loop like that shown in Fig 5, where the field has be repeatedly swept between positive negative values, measuring the magnetization continuously. Of particular interest for synchrotrons is the branch labelled 'inj', for which the field was swept down from high to low positive values and then swept up again from the low value. Note how rapidly the magnetization changes at the beginning of the upward sweep – just at the time when the synchrotron is being injected.

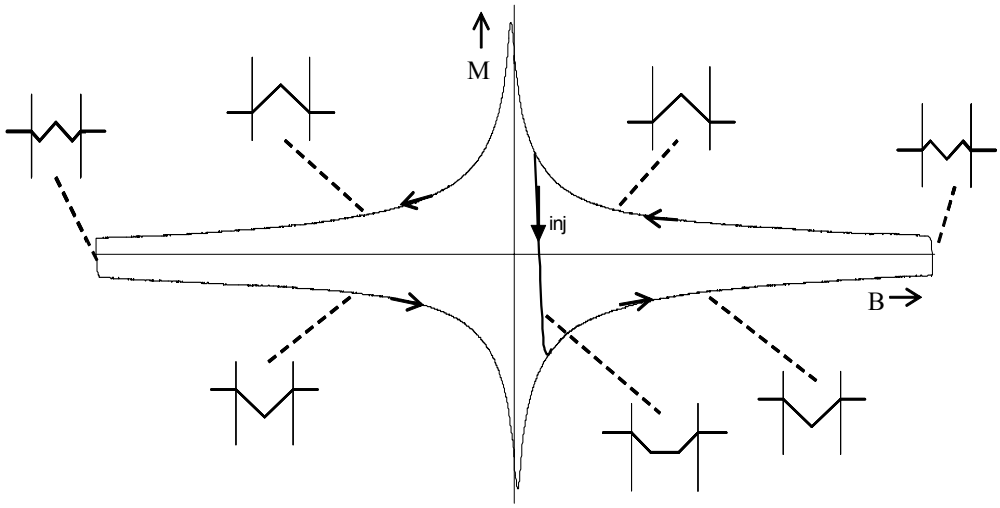


Fig 5: Magnetization curve of a Nb Ti superconducting wire; captions show internal field pattern for each zone of the curve.

In order to reduce magnetization effects, all magnet conductors are made with the superconductor divided into fine filaments. To avoid flux jumping, it is necessary to subdivide below a characteristic diameter, which for NbTi is ~ 50 μm – and this is always done, not just for accelerator magnets. For the more stringent conditions of accelerator magnets however, finer subdivision is required, mainly to reduce the field errors at injection. In addition, for accelerators with a fast ramping cycle, finer subdivision is also needed to reduce the ac losses. There is really no lower limit to the degree of subdivision which would be beneficial, but increasing cost and falling critical current density usually impose a practical lower limit of ~ 6 - 10 μm, and this has been used in all the superconducting accelerators built to date.

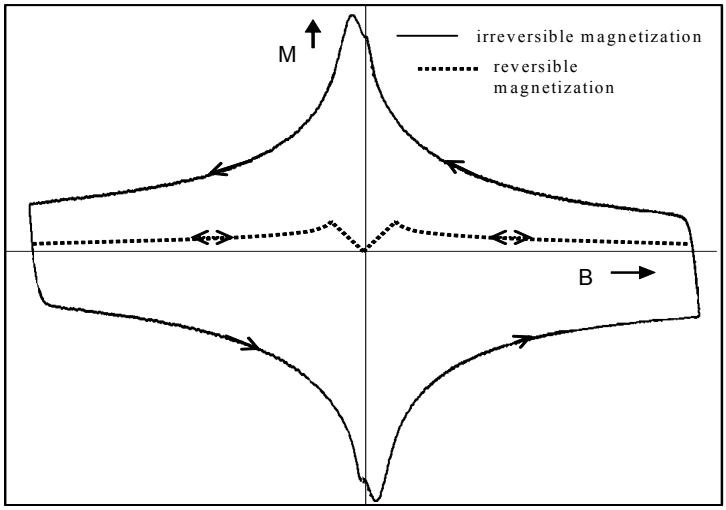


Fig 6: Irreversible and reversible components of magnetization for a Type II superconductor (note: scale of the reversible component increased to make it visible)

Finally, we emphasize that the magnetization discussed here is the *irreversible* magnetization coming from the *bulk* current density. It should not be confused with the *reversible* magnetization, inherent to any type II superconductor and shown in Fig.6. For all practical purposes, the inherent

reversible magnetization may be neglected however because it is very much smaller than the irreversible component (much more so than indicated by Fig. 6, where the scale of the reversible component has been increased for visibility).

4. FILAMENTARY COMPOSITE WIRES

A single 6μm filament of NbTi in 6T field will carry about 75mA – clearly we need a lot of them in parallel to make a synchrotron conductor! For this reason, as well as mechanical robustness, conductors are made in the form of multi-filamentary composite wires containing 5000 – 20000 filaments in a matrix of copper. As will be seen in the lectures on stability and protection, the copper performs several valuable roles in promoting reliable operation of the magnet. Unfortunately it also brings a problem by coupling the filaments together in changing fields, as illustrated in Fig. 7.

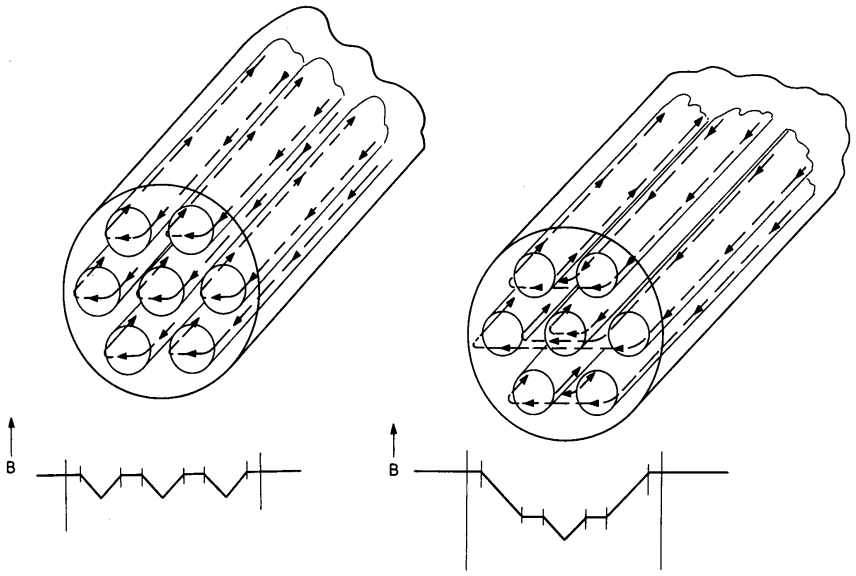


Fig. 7: A composite wire containing 7 filaments of superconductor in a copper matrix; left hand shows screening currents and field profile with the filaments uncoupled, right hand shows the same with filaments coupled.

Ideally we would like the filaments to behave quite independently as shown on the left hand side of Fig. 7, but in changing fields the screening currents flow across the matrix at each end of the wire, thereby screening more field from the interior. When the field change stops, these coupling currents will eventually decay to the pattern shown on the right of Fig. 7, but the time constant for this decay will be days or even years for wires of more than a few metres long. The advantages of fine subdivision are thus completely negated during practical operating timescales.

Fortunately there is a simple solution to this problem – twist the wire so that the coupling currents are forced to reverse every twist pitch. In this way the time constant for decay of the coupling currents may be reduced to a few msec so that, for reasonable rates of change of field, they never have a chance of building up. We now develop the theory of time constants for a twisted multi-filamentary composite wire. Essential to the theory is the assumption that the screening effect of the induced currents is small, so that the rate of change of field within the wire is the same as that outside. Fig. 8 illustrates a twisted multi-filamentary wire exposed to a changing field $B` = dB/dt$. Coupling currents flow along the zig-zag path PQR over many twist pitches before finally crossing over to S at the ends of the wire and returning via a similar zig-zag path down the other side.

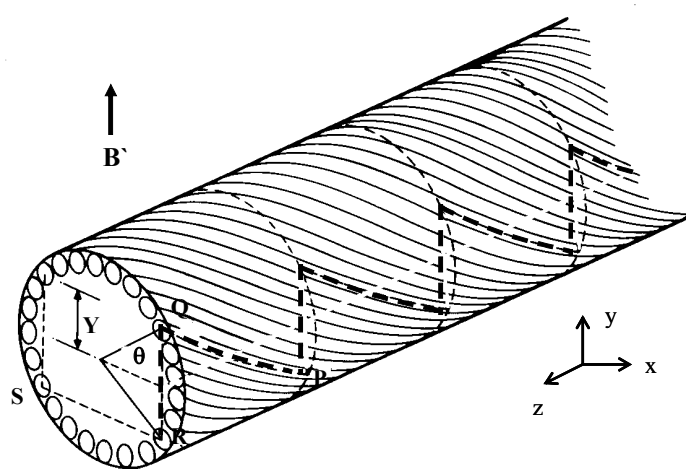


Fig. 8: Coupling current paths in a twisted multi-filamentary composite wire.

Provided the filaments are below their critical current, there can be no electric field along them, so the whole electric field induced by the changing magnetic field must lie along the vertical paths QR.

$$\int_{Q}^{R} E\, dl = \int_{P}^{Q} B_i'\, a\cos\theta\, dz = \frac{2B_i'\, p\, Y}{2\pi} \tag{4}$$

where a is the radius and p is the twist pitch. The expression holds for all y implying a uniform vertical electric field, which will produce a vertical current density:

$$J_y = \frac{B_i'\, p}{2\pi\, \rho_t} \tag{5}$$

where ρ_t is the effective transverse resistivity across the wire. We now consider how this current density through the matrix is collected from and donated to the ring of filaments.

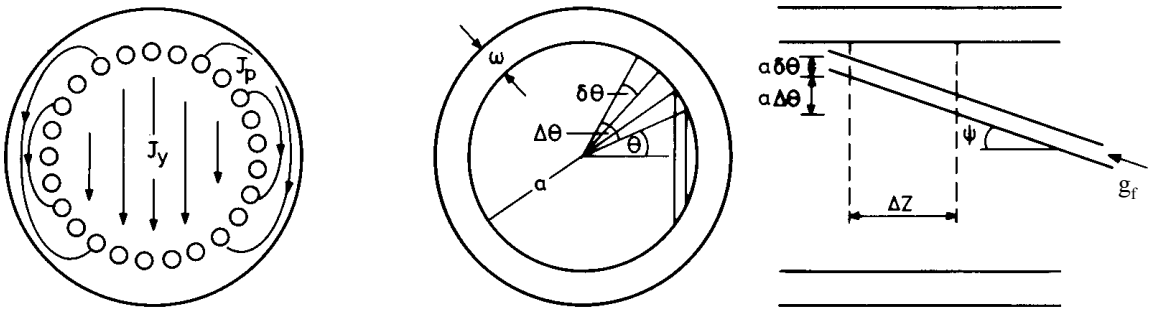


Fig 9: Connection between J_y and linear current density g_f flowing along the filaments.

We approximate the totality of filament currents by a cylindrical current sheet of *linear* (Am^{-1}) current density g_f and consider the current transferred to the matrix from a sector of angular extent $\delta\theta$ over a length in which its azimuthal coordinate changes by $\Delta\theta$.

$$\Delta g_f\, a\, \delta\theta\cos\psi = \frac{B_i'\, p}{2\pi\, \rho_t}\frac{a\,\Delta\theta}{\tan\psi}a\,\delta\theta\sin\theta \tag{6}$$

so that:

$$\frac{\Delta g_f}{\Delta \theta} = \frac{B_i' \; p}{2\pi \; \rho_t} \frac{a \sin\theta}{\cos\psi \tan\psi} \tag{7}$$

by symmetry $g_f = 0$ at $\theta = \pi/2$ so the solution of (7) is:

$$g_f(\theta) = \frac{B_i' \; p}{2\pi \; \rho_t} \frac{a \cos\theta}{\cos\psi \tan\psi} \tag{8}$$

g_f is directed along the filaments; the component along the axis of the wire is:

$$g_z(\theta) = \frac{B_i'}{\rho_t} \left\{ \frac{p}{2\pi} \right\}^2 \cos\theta \tag{9}$$

As discussed in the lectures on Magnetic Field Design, a $\cos\theta$ distribution of current around a cylinder produces a uniform field inside that cylinder. In this case, the induced screening currents produce a uniform field inside our cylinder of filaments, opposing the changing external field $B`$. Thus we may write the internal field:

$$B_i = B_e - \frac{\mu_o}{2} g_z(0) = B_e - \frac{\mu_o}{2} \frac{B_i'}{\rho_t} \left(\frac{p}{2\pi} \right)^2 \tag{10}$$

so that:

$$B_i = B_e - B_i' \; \tau \qquad \text{where} \qquad \tau = \frac{\mu_o}{2\rho_t} \left(\frac{p}{2\pi} \right)^2 \tag{11}$$

Finally we may integrate the (screening current) × (area enclosed) to get the induced magnetization

$$M_e = \frac{4}{\pi \; a^2} \int_0^{\frac{\pi}{2}} g_z(\theta) \; a \cos\theta \; a \, d\theta \qquad \text{and} \qquad M_e = \frac{2}{\mu_o} B_i' \; \tau \tag{12}$$

Note the resemblance between these equations and the standard eddy current formulae. Indeed these currents are really eddy currents in the copper matrix, whose amplitude has been greatly increased by the extra flux linkages through the superconducting filaments.

Thus we see that a filamentary composite wire has two components of magnetization:

a) steady state magnetization coming from the persistent currents flowing *within* filaments.
b) rate dependent magnetization coming from eddy currents flowing *between* filaments.

These two components are illustrated in Fig. 10, which plots the magnetization of a filamentary composite wire at different rates of change of field, the outermost loop being at the fastest rate of change. Note how the steady state component depends on B (via J_c) whereas the rate dependent component is independent of B and depends only on $B`$.

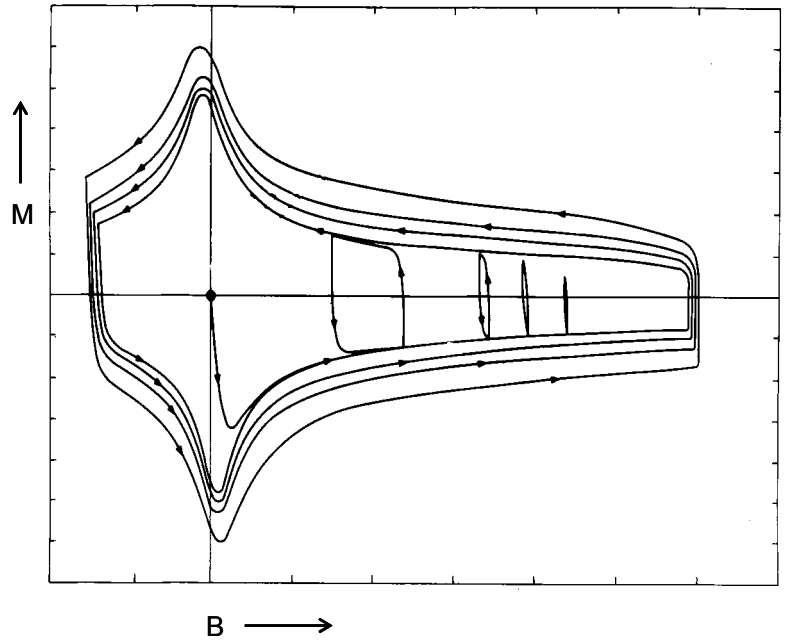


Fig. 10: Magnetization of a filamentary composite wire at different rates of change of field (outermost loop is fastest rate of change).

We can see exactly the same behaviour in the field error of a magnet. Fig. 11 plots the skew quadrupole error of a niobium tin dipole [2] at various rates of change of field.

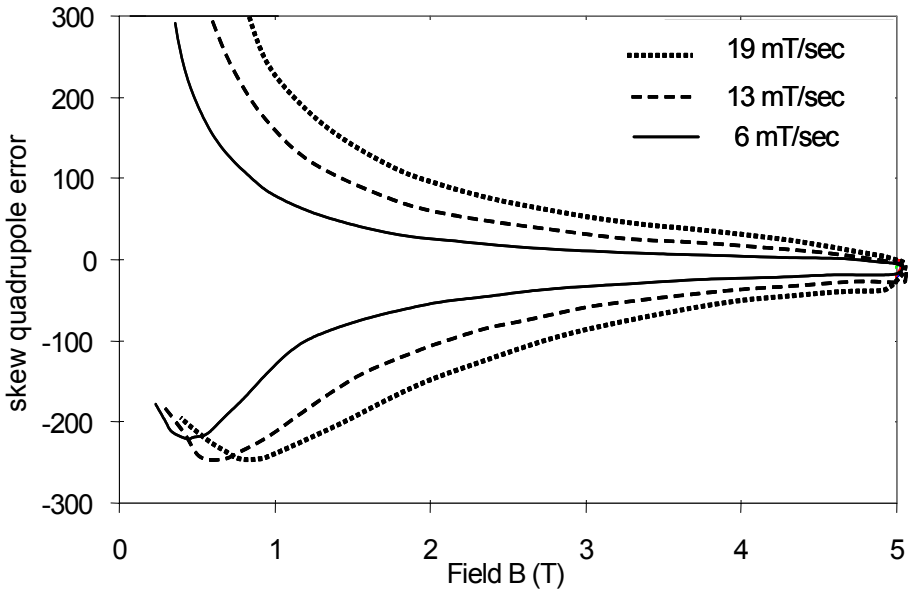


Fig. 11: Skew quadrupole error measured in a niobium tin dipole [2] at different sweep rates.

5. AC LOSSES

We have already seen in the critical state model that a changing magnetic field, by producing an electric field, drives the superconductor into the resistive state. It follows that the changing magnetic field must produce ac loss. In addition, the coupling currents through the copper matrix of a filamentary composite dissipate Ohmic heat. Because the work done to produce these losses is

supplied by the magnetic field, we may calculate them from the external field and magnetization of the sample. The work per unit volume when the sample is taken around a cycle is given by:

$$Q = \oint \mu_o M dH = \oint \mu_o H dM \tag{13}$$

If the magnetization is small compared with the external field (usually the case) we may also write:

$$Q = \oint M dB \tag{14}$$

For a general physics proof of the above, see any text book on Thermodynamics, for an engineering proof see [3]. Thus the ac loss around any ramping cycle is given by the area enclosed within the magnetization loop. Note that the magnetization includes all components, ie persistent currents in the filaments plus coupling between the filaments. For the coupling component of magnetization, which is independent of B, the ac loss when field changes between B_1 and B_2 is simply:

$$Q_e = M_e(B_2 - B_1) \tag{15}$$

For the loss component coming from persistent currents within the filaments, we need to know how the critical current J_c depends on field. This dependency is a function of pinning strength and varies between materials, but for NbTi at fields below ~ 6T, a reasonable fit is given by the Kim Anderson formula.

$$J_c(B) = \frac{J_o B_o}{B + B_o} \tag{16}$$

Thus we may integrate (3) to get the loss between B_1 and B_2:

$$Q_h = \int_{B_1}^{B_2} \frac{4a}{3\pi} \frac{J_o B_o}{B + B_o} dB = \frac{4a}{3\pi} J_o B_o \ln\left\{\frac{B_2 + B_o}{B_1 + B_o}\right\} \tag{17}$$

AC losses are not much of a problem in slow ramping storage rings like LHC, but for a fast cycling fixed target machine like the SIS recently proposed by GSI Darmstadt [4], they are a major design constraint

6. SUMMARY

Persistent currents in superconductors cause problems of flux jumping and magnetization. We can describe them in terms of the critical state model, which also explains the hysteretic behaviour of magnetization. Magnetization in the superconductor is usually the greatest source of field error at injection in superconducting accelerators. Magnetization is proportional to filament size, so conductors for accelerators are made with very fine filaments - typically 5 - 10 μm diameter. Practical conductors are made in the form of filamentary composite wires with the superconductor embedded in a matrix of copper. In changing fields the filaments are coupled together through the copper, thereby losing the benefit of subdivision and introducing a second component of magnetization. Twisting the composite wire serves to reduce the coupling by forcing the coupling currents to reverse every twist pitch. Coupling currents produce a second component of magnetization, which adds to the field error in magnets. They behave like eddy currents with a time constant determined by the twist pitch, so shorter twist pitches mean smaller coupling currents.

In changing fields, eg a synchrotron, the superconductor is driven into a partially esistive state, which means that it produces ac loss and consequent heating. These losses can produce a significant

refrigeration load. Loss is proportional to the area enclosed by the magnetization loop, as plotted in the *B M* plane. So to reduce losses, we must reduce both components of magnetization.

REFERENCES

[1] C. P. Bean, Phys Rev Letters, 8, (1962) 250.
[2] A. den Ouden, IEEE Transactions on Appplied Superconductivity, Vol 7, No 2 (1997) 733.
[3] M.N. Wilson, 'Superconducting Magnets', pub Oxford University Press, 1983.
[4] G.Moritz, C.Muehle (Gesellschaft fur Schwerionenforschung mbH), W.W. Hassenzahl (Advanced Energy Analysis) , M. Anerella, A. Ghosh, W. Sampson, P. Wanderer, E. Willen (BNL), N.N. Agapov, H. Khodzhibagiyan, A.D. Kovalenko (Joint Institute for Nuclear Research) , M.N. Wilson (Oxford Instruments (retired), Proceedings Of The 2001 Particle Accelerator Conference, Chicago, June 2001, p. 211- 214.

PRACTICAL SUPERCONDUCTORS FOR ACCELERATORS - CABLES

Martin N. Wilson
Consultant, Brook House, 33 Lower Radley, Abingdon, OX14 3AY, UK.

Abstract
All superconducting accelerators to date have used magnets wound from Rutherford cable. This second lecture describes why we need cables, the different types of cable and how they are manufactured. In changing fields, the wires in a cable can be coupled via the contact resistances between them. This coupling, which depends strongly on field orientation, adds to the total magnetization and thus to the field error produced in a magnet made from the cable. Non-uniform distribution of current between wires in the cable can cause premature quenching of the magnet and also produce the effect known as 'snap back', which is a sudden change in error field at injection.

1. INTRODUCTION

Why do we need cables? To ensure good tracking of the field with respect to the current, synchrotron magnets are always connected in series. If the stored energy is E, rise time t and operating current I, the charging voltage is

$$V = \frac{2E}{I\,t} \tag{1}$$

For example the dipoles of RHIC at Brookhaven have E = 40kJ/m, t = 75s and operate at I = 5kA, giving ramping voltage per km V_r = 213V. These magnets are wound from a 30 strand cable. If the magnets had been wound from a single strand with I = 167A, the ramping voltage per km would be V_r = 6.4kV, raising some difficult engineering problems. Far more severe problems arise in the rapid cycling synchrotron SIS proposed by GSI Darmstadt, which has E = 40kJ/m, t = 4s and I = 5kA, giving a ramping voltage per km of V_r = 4kV. Here again the magnets are to be wound from a 30 strand cable. The same magnets wound from a single strand would have V_r = 120kV.

Synchrotron magnets clearly need conductors able to carry several kA – and the best way of achieving this is to combine many wires in a cable. Firstly we shall look at some different designs of cable and ways of making them. Then we will discuss the problem of coupling between the strands of a cable, which is in many ways similar to the coupling between filaments in a composite wire, but made more complicated by the anisotropic geometry. Finally we will look at the distribution of current between strands and at some interesting effects coming from non-uniform distributions.

2. TYPES OF CABLE

The main consideration in cable design is the self field which, for a conductor carrying several kA, can be considerable. Fig 1 illustrates the self field lines for a cable, showing that there can be an appreciable flux linkage between those strands inside the cable and those outside. This flux linkage couples the strands together magnetically and has the effect of causing some strands to carry more current than others as the total current is increased. Such non-uniform current distributions are undesirable because these strands will reach their critical current I_c early and quench the whole cable at lower currents than it should be capable of. To promote a uniform current distribution the strands must be fully transposed, ie all strands must change places with each other along their length. As with the filaments in a wire, it is also necessary for the wires to be twisted. Twisting decouples the wires

with respect to an external field and transposition decouples them with respect to their own self field.

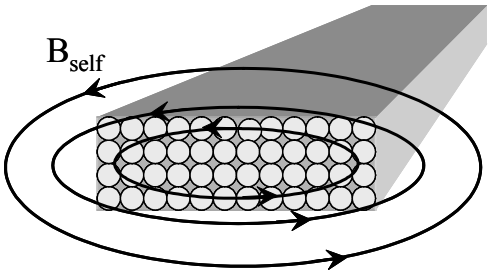


Fig. 1: Lines of self field for a current carrying cable.

Three types of fully transposed cable have been tried in superconducting accelerators and are shown in Fig. 2. Of these three, only the Rutherford cable has succeeded, mainly because it can be compacted to a high density and good dimensional tolerance without breaking the strands.

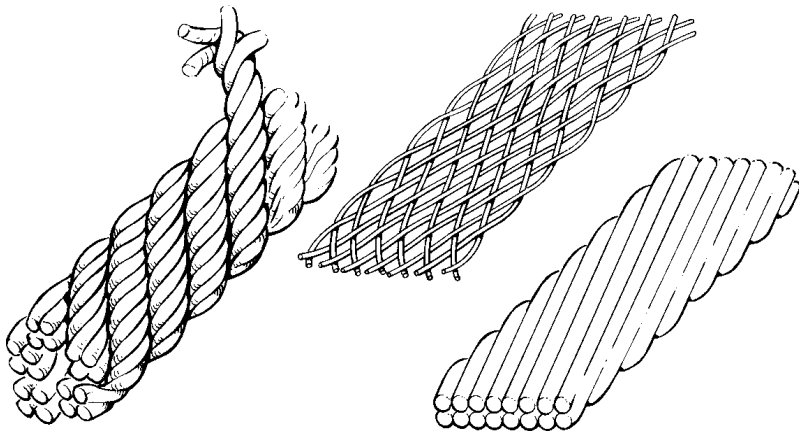


Fig. 2: Three types of accelerator magnet cable: a) rope or Litz cable b) braid c) Rutherford cable.

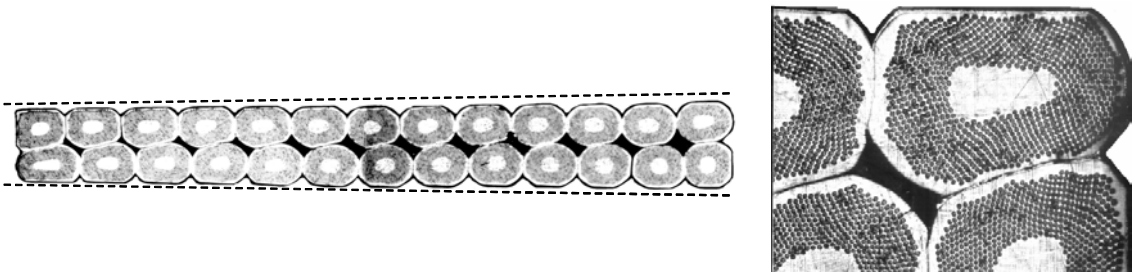


Fig. 3: Cross section and local view of a Rutherford cable for LHC.

Fig. 3 shows a cross section of the cable used in the inner coil of the LHC dipoles; note the small 'keystone angle', which makes the cables fit together better when stacked around a circular aperture.

Fig. 4 shows a typical set-up for making Rutherford cable. The large wheel carries 30-40 bobbins of filamentary composite superconducting wire which is twisted into a hollow tube before passing through the "turk's head" die, consisting of 4 offset rollers arranged around a rectangular aperture which defines the dimension of the cable. With careful setting up, dimensional tolerances of better than 10μm can be achieved. The finished cable may then be annealed before passing to the insulating machine, which typically applies 4-6 wraps of 25μm thick Kapton.

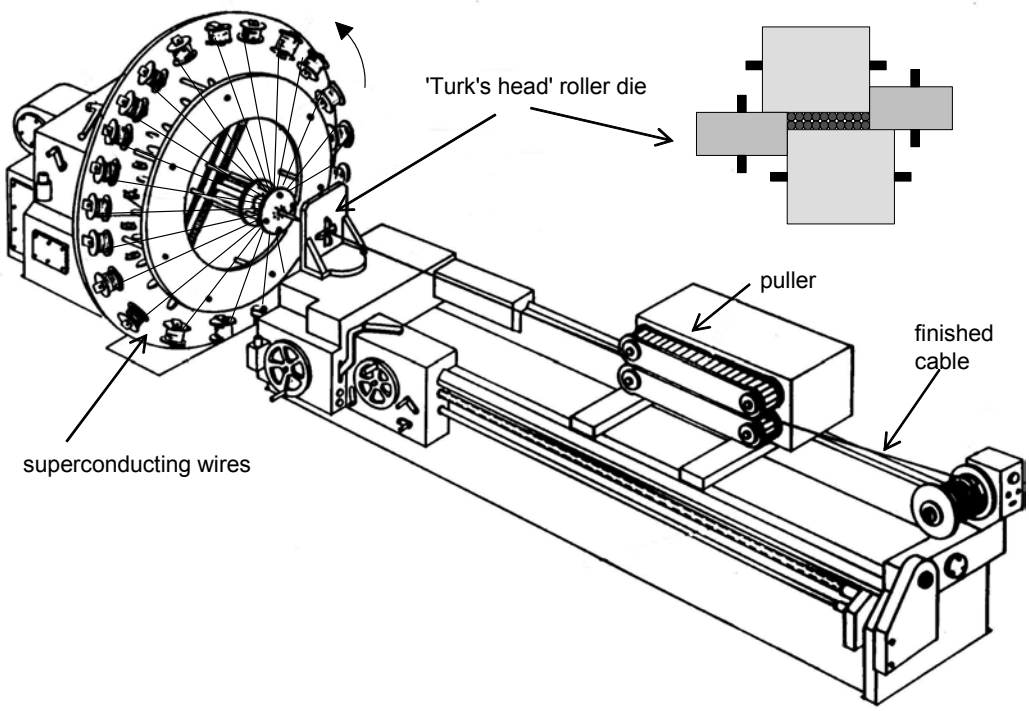


Fig. 4: Manufacture of Rutherford cable.

3. COUPLING IN CABLES

In changing fields, eddy currents are induced to flow via the contact resistances between the strands of a cable. There are two approaches to calculating these currents:

a) network models using finite element software [1]

b) continuum models with analytic solutions

Here we shall concentrate on b), but will first mention the network modelling because it has largely determined the nomenclature in common use. The analytic formulae give an adequate answer for most purposes, but network modelling should be used for precise answers or special cases such as samples which are short enough for end effects to be important. As shown in Fig. 5, the network model assumes a series of nodes which are connected to their neighbours by two types of contact resistance: the crossover resistance R_c and the adjacent resistance R_a. Each resistance is to be associated with that length of cable occupied by a single node.

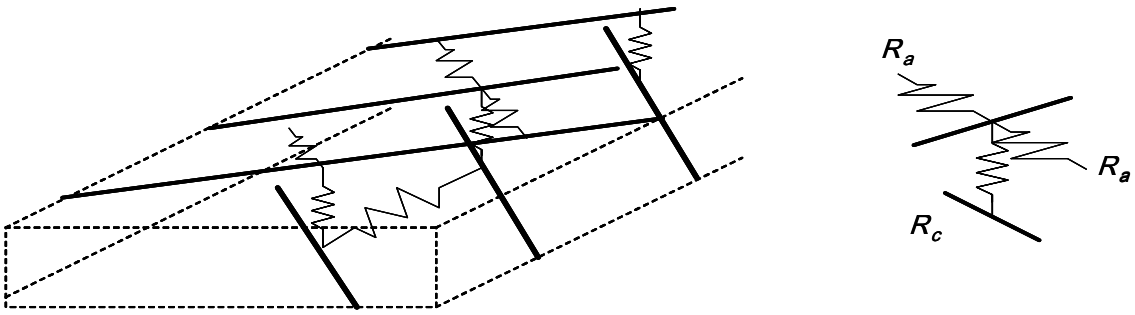


Fig. 5: Network model of a Rutherford cable

The continuum model takes the cable as two layers (top and bottom) of continuous conductor with a uniform resistance per unit area between them. Because of the cable shape and also because R_c and R_a are usually different, the cable behaves very differently depending on the field direction. In the next 3 sections we shall consider three cases:

a) field perpendicular to broad face of cable, coupling via crossover resistance
b) field perpendicular to broad face of cable, coupling via adjacent resistance
c) field parallel to broad face of cable, coupling via adjacent resistance

3.1 Coupling via crossover resistance in perpendicular field

Changing fields perpendicular to the cable face induce diamond shaped loops of current like those shown in Fig.6.

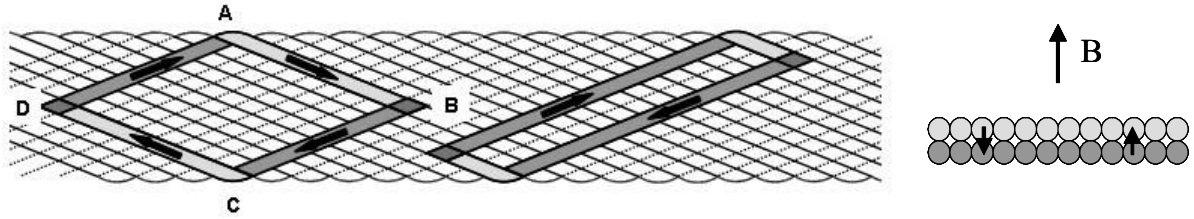


Fig.6: Coupling current loops through crossover resistance induced by field perpendicular to cable face.

Starting at the point A in Fig. 6, current is induced to flow along the top face (shown pale grey) to the point B where it transfers downwards via R_c. It then flows along the bottom face (shown dark grey) to C, where the wire curves around the edge of the cable to the upper face. Then along to D, where it again flows down through the crossover and along the lower face to A, where the wire again curves upwards over the edge of the cable. Thus the loop encloses flux in the area ABCD and has resistance $2R_c$. In addition, there are many other loops like the example shown on the right on Fig.6. To sum the totality of these loops, we switch to a continuum approach.

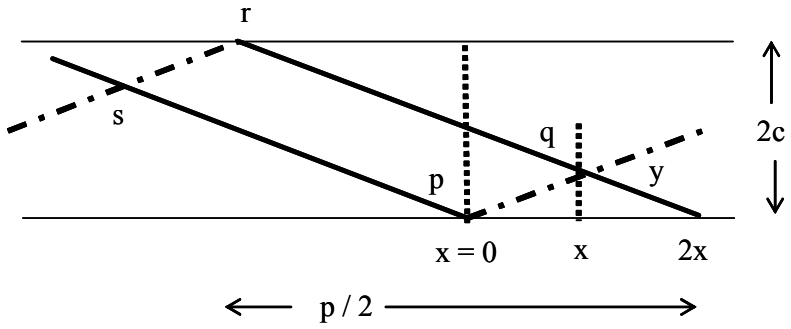


Fig. 7: Geometry of a typical current loop

As shown in Fig. 7 we consider a typical loop *pqrs* which encloses area:

$$A_l = 2\left[xy + 2y\left(\frac{p}{2} - x\right)\right] = p\left[y - \frac{y^2}{2c}\right] \qquad (1)$$

where p is the twist pitch and c is the half width of the cable. We now define a crossover resistance per unit area r_c and calculate the resistance of a crossover contact whose area is defined by a change dy in vertical coordinate for a conductor of width w, as sketched in Fig. 8.

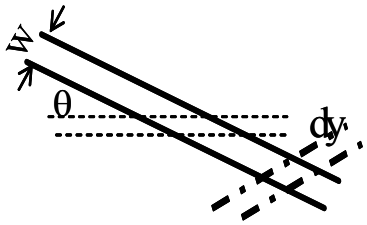


Fig. 8: Area of crossover contact

Thus we find:

$$R_y = r_c \frac{\sin\theta}{w\,dy} \tag{2}$$

$$V = dI_t(y)\,2r_c \frac{\sin\theta}{w\,dy} = B'\,p\left(y - \frac{y^2}{2c}\right) \tag{3}$$

$$\frac{d\,I_t(y)}{dy} = \frac{B'\,w\,p}{2\,r_c \sin\theta}\left\{y - \frac{y^2}{2\,c}\right\} \tag{4}$$

solving for I_t with the symmetry condition that $I_t = 0$ at $y = c$:

$$I_t(y) = \frac{B'\,w\,p}{2\,r_c \sin\theta}\left(\frac{y^2}{2} - \frac{y^3}{6\,c} - \frac{c^2}{3}\right) \tag{5}$$

and the effective *linear* (A.m^{-1}) current density g_x in the x direction is:

$$g_{xt} = \frac{I_t(y)}{w} = \frac{B'\,w\,p}{2\,r_c \sin\theta}\frac{\cos\theta}{w}\left(\frac{y^2}{2} - \frac{y^3}{6\,c} - \frac{c^3}{3}\right) = \frac{B'\,p^2}{8\,r_c\,c}\left(\frac{y^2}{2} - \frac{y^3}{6\,c} - \frac{c^2}{3}\right) \tag{6}$$

which we may integrate across the width of the cable to obtain a magnetic moment per unit volume, ie the magnetization:

$$M = \frac{1}{V}\sum \mu_o\,I\,A = \frac{1}{c\,b}\frac{B'\,p^2}{8\,r_c\,c}\int_0^c \left(\frac{y^2}{2} - \frac{y^3}{6\,c} - \frac{c^2}{3}\right)(c - y)\,dy \tag{7}$$

so $$M_{pc} = \frac{1}{60}\frac{B'}{r_c}p^2\frac{c^2}{b} \tag{8}$$

By analogy with the twisted filamentary wire in Lecture 1, we may also write:

$$M_{pc} = \frac{B'}{\mu_o}\frac{c}{b}\tau_{pc} \qquad \text{where} \qquad \tau_{pc} = \frac{\mu_o}{60}\frac{p^2}{r_c}c \tag{9}$$

In this case however τ_{pc} is not a single time constant, but rather the average of many, because each of the loops we have summated decays with a different time constant [2].

3.2 Coupling via adjacent resistance in perpendicular field

In this case, coupling currents flow in either the top or bottom face, with no current flowing between the two. Current flows along each face symmetrically on either side of the centreline as sketched in Fig. 9.

Fig. 9: Current flow along a face via adjacent resistance in perpendicular field.

First we calculate an effective resistivity along the face with the nomenclature sketched in Fig. 10:

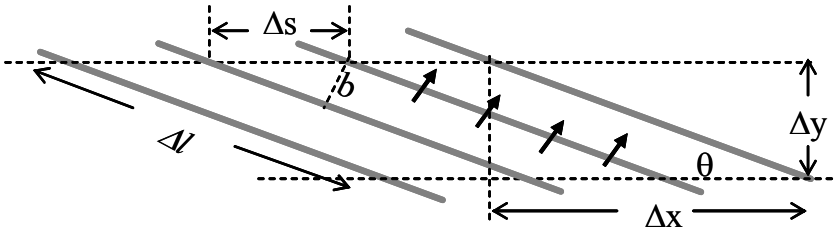


Fig. 10: Calculating the effective resistivity along a face.

Referring to Fig.10, the resistance of a section height Δy over length Δx is:

$$R_x = \frac{\Delta x}{\Delta s}\frac{r_a}{b\,\Delta l} = \frac{\Delta y \sin\theta}{b\cos\theta}\frac{r_a \sin\theta}{b\,\Delta y} = \frac{\sin^2\theta}{\tan\theta}\frac{r_a}{b^2} \tag{10}$$

where r_a is the adjacent resistance per unit area. Let an overall current density J_x be induced by the changing field B`; by symmetry there is no electric field along centre line of cable, so:

$$B'\,y\,\Delta x = J_x(y)\,b\,\Delta y\,R_x \tag{11}$$

Combining (10) and (11) and solving we find:

$$J_x(y) = \frac{B'\,b}{\sin^2\theta\,r_a}\,y \tag{12}$$

Integrating this current over half width of cable and dividing by volume to get magnetization:

$$M = \frac{B'}{c\,b}\frac{b^2}{\sin^2\theta\,r_a}\int_0^c y^2\,dy \tag{13}$$

thus

$$M_{pa} = \frac{B'}{48}\frac{1}{r_a}\frac{p^2 b}{\cos^2\theta} \qquad \text{in practice } cos\theta \text{ so that} \qquad M_{pa} \cong \frac{B'}{48}\frac{p^2 b}{r_a} \tag{14}$$

3.3 Coupling via adjacent resistance in parallel field

Here we have a rather simple situation in which the changing field *B`* induces uniform current density flowing in opposite directions on either side of the cable, just like the slab model in Lecture 1.

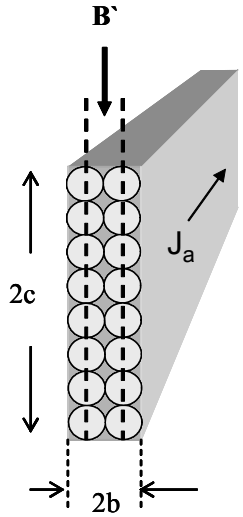


Fig. 11: Coupling in a parallel field.

We assume the current to flow down the centre line of each wire, as shown by the dashed lines in Fig. 11 and consider the flux change enclosed by a length Δx in (the direction of the cable):

$$J_a \, \Delta x \, b \, R_x = V = B' \, \frac{b}{2} \Delta x \tag{15}$$

using (10) again we find:

$$J_a = \frac{B'}{2} \frac{\Delta x}{\Delta y} \frac{1}{R_x} = \frac{B'}{2} \frac{b^2}{\sin^2 \theta \, r_a} \tag{16}$$

the magnetization is thus:

$$M_{la} = \frac{1}{b\,a} J_a b \, a \frac{b}{2} \qquad \text{so with (16)} \qquad M_{la} = \frac{1}{64} \frac{B'}{r_a} \frac{p^2 b^3}{c^2} \frac{1}{\cos^2 \theta} \tag{17}$$

3.4 Comparison of resistances with network model

Finally we calculate the relation between the resistances per unit area of the continuum model and the resistances per node of the network model. To simplify the argument, we assume the wires are ‘square’ in cross section, with side = b in a plane perpendicular to the cable length.

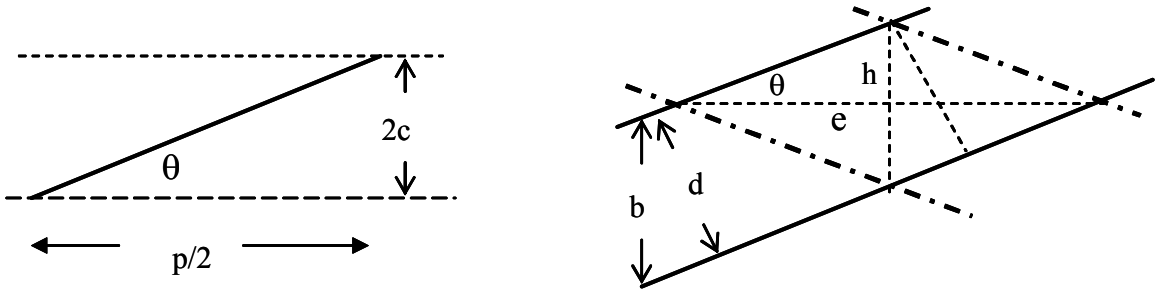


Fig. 12: Geometry of a cross-over.

Referring to Fig. 12, we have:

$$h = \frac{b}{2} \qquad e = \frac{h}{\tan\theta} \qquad \text{and} \qquad \tan\theta = \frac{4c}{p} \tag{18}$$

so that the area of a crossover is:

$$A_c = 4\frac{h}{2}e = 4\frac{b}{4}\frac{b}{2\tan\theta} \qquad \text{and} \qquad A_c = \frac{b^2 p}{8c} \tag{19}$$

Because we have assumed a 'square' wire, the adjacent area of contact is the same as the cross-over area, and so:

$$R_a = \frac{r_a}{A_c} = r_a \frac{8c}{b^2 p} \qquad \text{and} \qquad R_c = \frac{r_c}{A_c} = r_c \frac{8c}{b^2 p} \tag{20}$$

Even when the wires are not square, provided we define *b* in terms of the distance between centres of adjacent wires, the formulae hold good.

4. CONTACT RESISTANCE IN CABLES.

4.1 Measurement of contact resistance.

One obvious way of measuring r_c and r_a would be to measure the dependence of magnetization on rate of change of field B`, but such measurements are expensive and time consuming. Verweij [1] has therefore suggested the much simpler method sketched in Fig. 13.

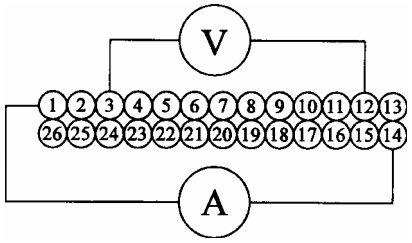


Fig. 13: The VI method of measuring contact resistances in a cable.

Current is fed in at opposite sides of a sample length of cable, usually one twist pitch long, and the voltage on each strand is measured.

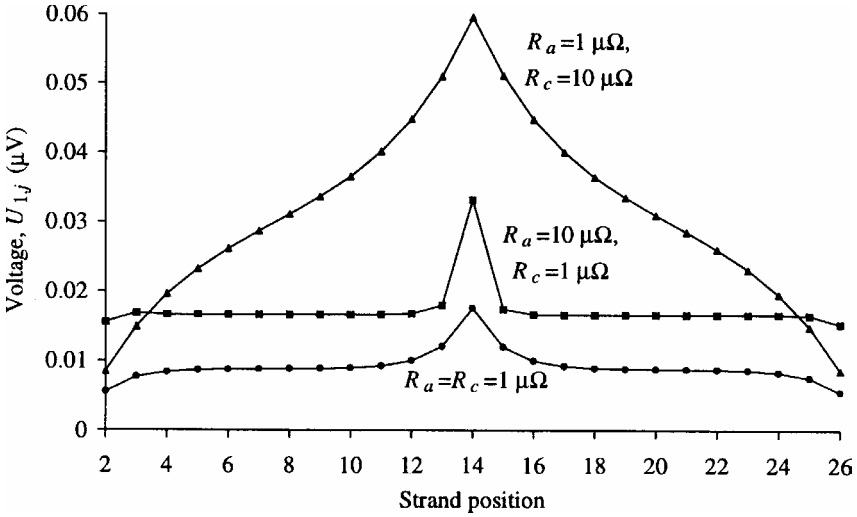


Fig. 14: Computed shapes of voltage versus strand number in the VI cable measurement [1]

To interpret the results however it is necessary to run cases on a network model as described in [1]. Fig 14 shows the results of some of these runs; the shape of the curve depends on the ratio R_a / R_c and the magnitude depends on the size of the resistances.

4.2 Choice of contact resistance

Looking at equations (9), (14) and (17), one might think that the obvious choice would be to make r_c and r_a as large as possible. In this way the coupled magnetization, and its contribution to field error in the magnet, would be made very small. In practice however it has been found that a large resistance between strands in the cable can produce unreliable performance in the magnet. Fig. 15 illustrates this by showing the dependence of quench current on ramp rate for two different magnet types in the SSC booster [3].

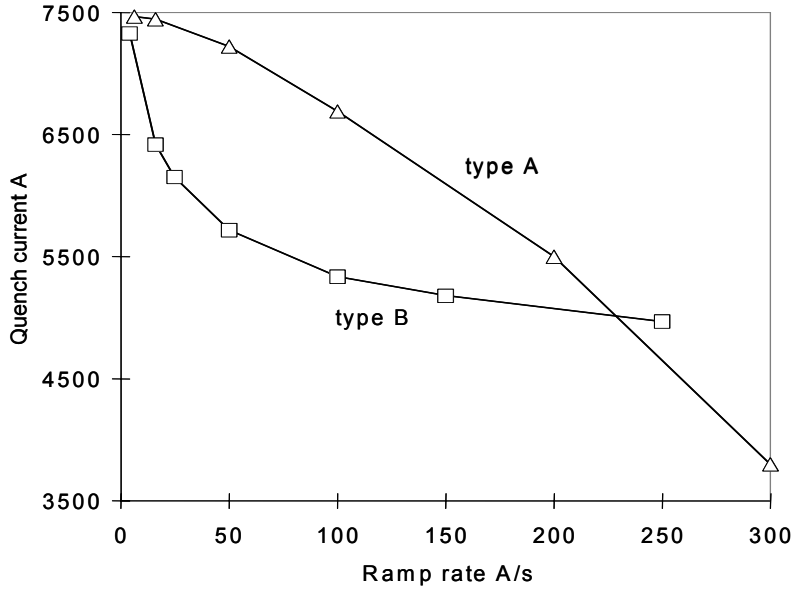


Fig. 15: Dependence of quench current on ramp rate for magnets of the SSC Booster.

Magnets of type A showed the sort of behaviour one might expect. At low ramp rates the quench current does not change much but, at high rates when ac losses start to produce a significant heating and consequent temperature rise, the quench current falls. These magnets have rather low

contact resistances r_c and r_a. Type B magnets, on the other hand, have high values of r_c and r_a. In this case, it is thought that the rapid fall of quench current at low ramp rates is caused by non-uniform distribution of current between wires in the cable, such that one wire hits its critical current before the others and precipitates a quench of the whole cable. Note that at high ramp rates, type B does better than type A, because its ac losses are less. However, temperature rise due to ac loss can easily be reduced by cooling the magnet more efficiently, but we have no way of dealing with non-uniform current distribution.

At the start of section 2, we discussed how fully transposed cables are designed to ensure uniform current distribution – so where does type B behaviour come from? It is thought to come from slight asymmetries between wires in the cable arising at current connections, end turns of the magnet, layer changing 'joggles' etc, but is not fully understood at present. For this reason, we normally choose r_c and r_a to be as small as possible, consistent with achieving acceptably low values of magnetization.

4.3 Control of contact resistance

It is important for the magnet designer to know what contact resistances will be present in the cable when it is actually working in the magnet. Unfortunately they are very variable and strongly affected by factors in coil construction, particularly mechanical stress and the heat treatments which are applied to activate the adhesion between turns. Fig 16 shows some data from CERN on the effect of heat treatment on R_c for wires with different surfaces. It may be seen that bare copper is very sensitive, with heat treatment changing R_c by three orders of magnitude. Plated coatings on the copper are much better, with nickel providing a good high resistance coating and 'Staybrite (AgSn alloy) being good for low resistances.

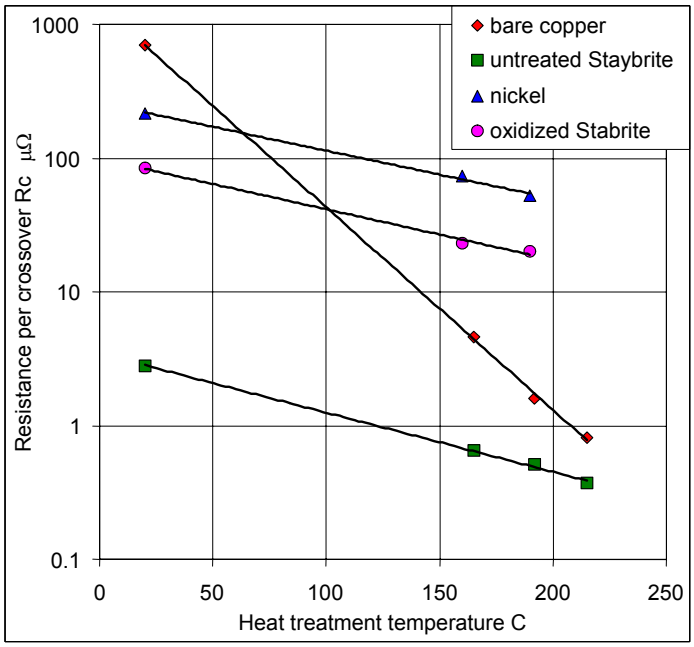


Fig. 16: Contact resistance R_c in LHC dipole cables for different coatings as a function of heat treatment temperature [4].

4.4 Anisotropic resistances

We have seen that, for low magnetization, we want high contact resistance whereas to mitigate the effects of non-uniform current sharing we want low resistance. However the effect of r_c and r_a on magnetization are not the same. Collecting (8) and (14) we find:

$$\frac{M_{pc}}{M_{pa}} = \frac{4}{5}\frac{c^2}{b^2}\frac{r_a}{r_b} \tag{21}$$

Typically the aspect ratio of a cable c / b ~ 8, so magnetization is ~ 60 times more sensitive to rc than ra. As far as we know, current sharing is equally affected by rc and ra. So if we chose rc ~ 50 x ra, we should get the minimum magnetization and best current sharing. We can do this by putting a resistive core inside the cable, for example a stainless steel foil as shown in Fig. 17.

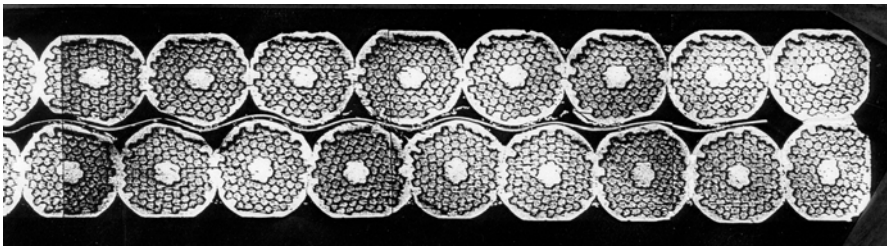


Fig. 17: Rutherford cable with a resistive core foil.

5. LONG RANGE COUPLING CURRENTS

5.1 BICCs

We have already touched on the problem of non-uniform current distribution between the strands of a cable, and this can have some interesting effects on the field profile of a magnet. When the field of an accelerator magnet is measured at low field as a function of distance along the beam direction, it is often found to contain a ripple component. The frequency of this ripple exactly matches the twist pitch of the cable. It is thought that the effect is caused by non uniform current sharing in the cable as the field is increased. When the field is decreased, this non uniform distribution is 'locked in' as a current circulating between the strands of the cable. Verweij [1] has called them 'boundary induced coupling currents' or BICCs. They are thought to be caused by non-uniform flux linkages in the cable, eg at joints, coil ends, manufacturing errors etc. Fig. 18 shows an example of this effect in an SSC dipole at injection and full field. Because they flow over long lengths of cable, BICCs have long time constants ~1000s or longer.

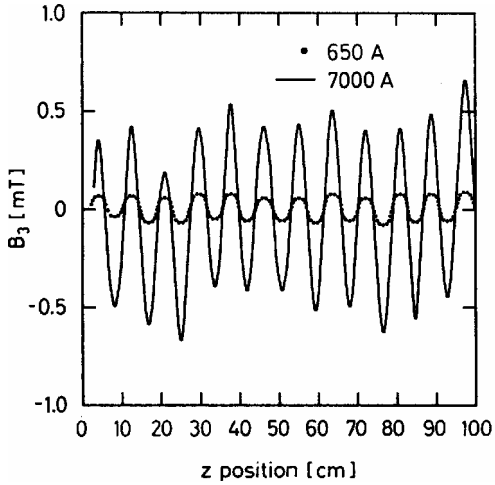


Fig. 18: the sextupole component of field measured in SSC dipole at injection and full field.

5.2 Snap back

The wavelength of the field oscillations is much shorter than that of betatron oscillations, so their effect is averaged out by the beam and will not cause a problem. However there is a rather subtle problem which comes from two factors:

a) the long decay times of BICCs which, because they extend over long lengths of cable, can 1000 seconds or more

b) the interaction between BICCs and other persistent currents

One result of this interaction seems to be a strange effect called 'snap back', which occurs at injection field and is shown in Fig. 19. During the injection platform, when the field is constant, a sextupole field error is seen which decays with a time constant of 1000 seconds or more. When the acceleration ramp is started at the end of injection, this sextupole error snaps back to its original value. Clearly such strange behaviour can play havoc with the beam dynamics at a critical point in the acceleration sequence. If the error is some kind of eddy current effect, it is not difficult to see why it should snap back to its original value as soon as B` is applied. The difficult question is to see where the decay time constant of > 1000s has come from. The only known currents with such a long time constants are BICCs, but the error shown in Fig. 19 is a value which is integrated over very many wavelengths of the ripple shown in Fig 18, so we would expect the positive and negative terms to cancel.

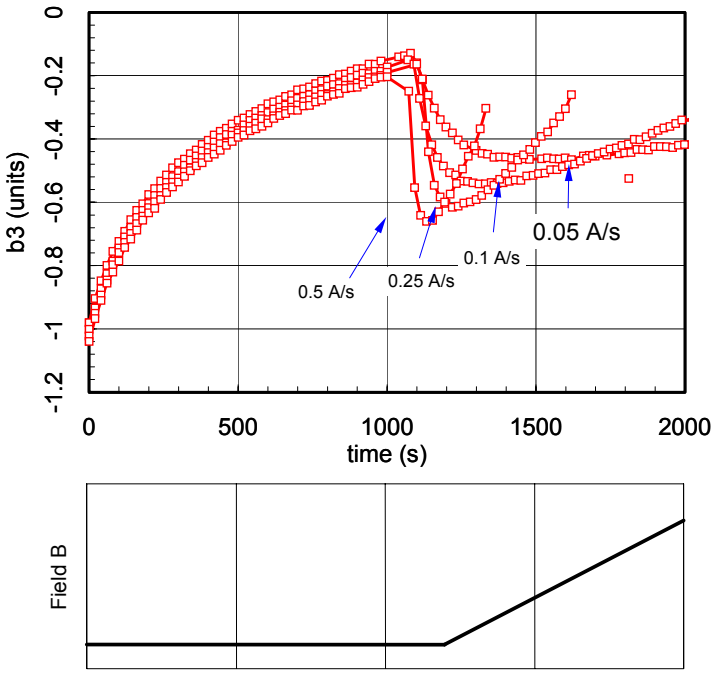


Fig. 19: 'Snap back of the sextupole field component in LHC dipole at the start of ramping; lower curve shows dipole field versus time.

The explanation of how an oscillating field can give a 'dc' effect lies in the nature of the hysteresis loop of an individual filament as described in Lecture 1 and shown for the low field region in Fig 20. In superconducting synchrotrons, it is usual to ramp down from high field A to a field B, which is lower than the chosen injection field, and then ramp up to injection field. In this way, the filaments start out from injection on the lower part of the magnetization loop and avoid the region BC where magnetization changes rapidly. If during the ramp up however, the field were to be lowered again, the magnetization would follow a curve like DE, crossing over from the lower to the upper envelope.

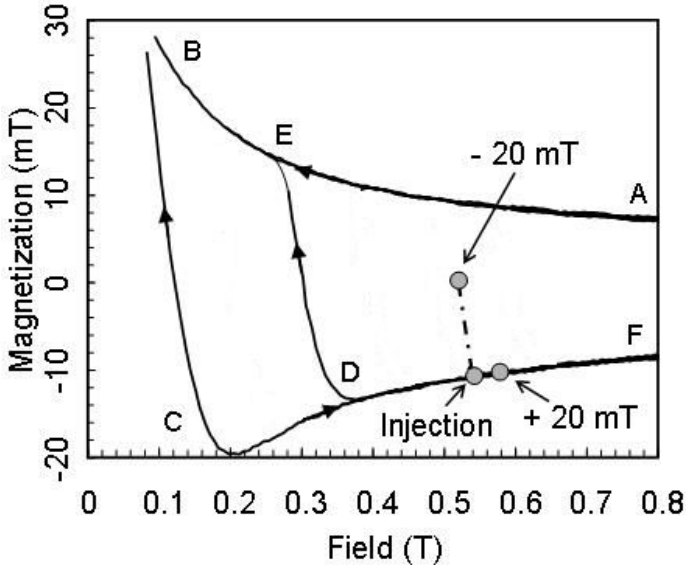


Fig. 20: Local field on a filament subject to a ripple field caused by BICCs.

At the injection field, the cable is carrying BICCs which impose a local ripple field on the filaments and which decay with a long time constant. In the sections where ripple decays in a negative direction, magnetization changes substantially from its injection value to the point '-20mT' but, in sections of positive decay, the magnetization changes very little to the point '+20mT'. Thus we see that the non-linearity of the hysteresis loop has acted as a rectifier to transform the symmetric ripple of the BICC field into a uni-directional field error. All aspects of 'snap back' can be explained by this model, which was first proposed by R. Wolf of CERN [5].

6. CONCLUDING REMARKS

Accelerator magnets need high current because they must be connected in series, so we must use cables of 20 to 40 wires. These cables should be fully transposed to get uniform current sharing between the wires. Wires in the cables become coupled together in changing fields - similar to the filaments in composite wires, but more complicated because of geometry. We distinguish between crossover and adjacent resistances R_c and R_a and fields perpendicular or parallel to the broad face of the cable. R_c coupling is much stronger than R_a. Despite full transposition, the current does not always share uniformly in practical magnets and it get worse for fast ramping. Non-uniform current sharing can give problems of premature quenching. For minimum magnetization and best current sharing, we choose high R_c and low R_a, which may be achieved using a cored cable. Non-uniform current sharing also produces BICCs, which cause a ripple component in the field along the magnet and can also produce the effect of 'snap back' during the injection process.

REFERENCES

[1] A.P. Verweij, 'Electrodynamics of superconducting cables in accelerator magnets', Thesis University of Twente, Netherlands 1995.

[2] A.A. Akhmetov, Physica C, 310 (1998) pp309.

[3] J. Thompkins, C. Haddock and G Snitchler, 'Ramp rate issues in HEB magnets', Chap 31 in SSC Report SSCL-Sr-1235 (1994).

[4] D. Richter, CERN, private communication.

[5] L. Bottura, L. Walckiers and R. Wolf; 'Field Errors Decay and "Snap-Back" in LHC Model Dipoles'. IEEE Trans. Appl. Superconduct., 7 (1997) - pp.602-605

ELECTROMAGNETIC DESIGN OF SUPERCONDUCTING ACCELERATOR MAGNETS

S. Russenschuck
CERN, 1211 Geneva 23, Switzerland

Abstract
The design and optimization of the superconducting magnets for LHC is dominated by the requirement of an extremely uniform field, which is mainly defined by the layout of the coils. Even very small geometrical effects such as the trapezoidal shape of the cable and the grading of the current density in the cable due to varying cable compaction, as well as coil deformations due to collaring and cool down have to be considered in the field calculation. In particular for the 3D case, commercial software has proven inadequate to this task. The CERN field computation program ROXIE was therefore developed for the design optimization of the LHC magnets and includes the method of coupled boundary elements and finite-elements. As with this method the coils do not need to be represented in the finite element mesh, they can be modeled with the required accuracy. This report describes the mathematical foundations of superconducting accelerator magnet design and the fundamentals of numerical field computation methods implemented in the ROXIE code.

1. Introduction

The Large Hadron Collider (LHC) requires high-field superconducting magnets to guide the counter-rotating beams with the desired proton energy of 7 TeV in the existing LEP tunnel with a circumference of 26.6 km. The LHC magnet system consists of 1232 superconducting dipoles and 386 main quadrupoles together with about 20 different types of magnets for insertion and correction. The only way of obtaining a nominal dipole field of 8.3 T with niobium-titanium (NbTi) superconductors is to cool the magnets to 1.9 K where helium takes on the so-called super-fluid state, with zero viscosity and very large heat conductivity. This aids in the cooling of the superconducting wires while reducing the helium flow through the magnets. However, the heat capacity of the superconducting cables is reduced by nearly an order of magnitude (compared to 4.5 K), resulting in a higher temperature rise for a given deposit of energy. Therefore, any movement of the coil must be avoided by the use of an appropriate force-retaining structure, in particular as the forces and the stored energy in the magnets increase with the square of the magnetic field. Fig. 1 shows the critical current surface of the NbTi alloy as a function of current density and field with the load-line of the LHC dipoles. The maximum current is limited by the critical current density in the coils which are exposed to an about 3% higher field than the nominal field in the aperture of the magnet. Important for the magnet performance is the engineering current density J_{E} which also takes into consideration the copper matrix, the filling factor in the cable, and the insulation.

Although a niobium-tin (Nb_3Sn) alloy allows, at 8 T, approximately twice the current density of NbTi, it was not considered for the LHC, as a series production of magnets would have to confront the brittle nature of the material, which requires a *wind-and-react* technique, whereby the conductor containing unreacted niobium and tin is first wound and then heat-treated at a reaction temperature of about 700^{o}C to form the superconducting A15 phase of Nb_3Sn.

The coils of the LHC dipole and quadrupole magnets are wound of Rutherford-type cable of trapezoidal (keystoned) shape. The dipole coils consist of two layers with cables of the same height but of different width. Electrically the two layers are connected in series. The current density in the superconductor of the outer layer, being exposed to a lower magnetic field, is about 40 % higher than in

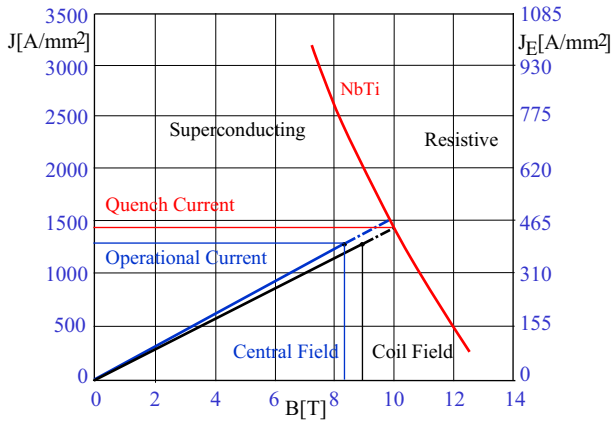


Fig. 1: Critical current density at 1.9 K of the NbTi alloy as a function of the applied field with the load-line of the LHC dipoles. Note that the quench current is limited by the critical surface of the cable being exposed to a higher field than that measured in the aperture of the magnet. Important for the magnet performance is the engineering current density J_{E} (scale on the right hand side) which also takes into consideration the copper matrix, the filling factor in the Rutherford cable, and the insulation.

the inner layer. The conductor for the inner layer consists of 28 strands of 1.065 mm in diameter, the one for the outer layer comprising 36 strands of 0.825 mm in diameter. The outer layer conductor of the dipole is used in both layers of the main quadrupoles.

The strands are made of thousands of filaments of NbTi material (6 and 7 μm in diameter) embedded in a copper matrix, which serves to stabilize the conductor and to take over the current in case of a quench, since superconductors have a high resistivity in the normal state. The filaments are made as small as possible in order to reduce the remanent magnetization effects and increase the stability against flux jumps during excitation, i.e., the release of fluxoids from their pinning centers. The keystoning of the cable is not sufficient to allow the cables to build up arc segments. Copper wedges are therefore inserted between the blocks of conductors. The size and shape of these wedges yield the necessary degree of freedom for optimizing the field quality produced by the coil. Since the field quality is extremely sensitive to coil positioning errors, each layer of the coil is polymerized in a mould at a temperature of 180 ^{o}C for 30 minutes in order to glue the turns firmly together and give the coil its final shape. The size and elastic modulus of each layer is measured to determine pole and coil-head shimming for the collaring. The required shim thickness is calculated such that the compression under the collaring press is about 120 MPa. After the collaring rods are inserted and external pressure is released, the residual coil pre-stress is about 50-60 MPa on both layers. The collars (made of stainless steel) are surrounded by an iron yoke which enhances the magnetic field by about 20%, reduces the stored energy and shields the fringe field. The dipole magnet, its connections, and the bus-bars are enclosed in the stainless steel shrinking cylinder closed at its ends and form the dipole cold-mass, a containment filled with static, pressurized (1 bar) superfluid helium at 1.9 K. The cold-mass, weighing about 24 tons, is assembled inside its cryostat, which comprises a support system, cryogenic piping, radiation insulation, and thermal shield, all contained within a vacuum vessel. The cross-section of the dipole magnet *cold-mass* (version of spring 2000) in its cryostat is shown in fig. 2.

1.1 Field quality in accelerator magnets

The magnetic field errors in the aperture of accelerator magnets can be expressed as the coefficients of the Fourier-series expansion of the radial field component at a given reference radius (in the 2-dimensional case). In the 3-dimensional case, the transverse field components are integrated over the entire length of the magnet. For beam tracking it is sufficient to consider the transverse field components, since the effect of the longitudinal component of the field (present only in the magnet ends) on the particle motion can be

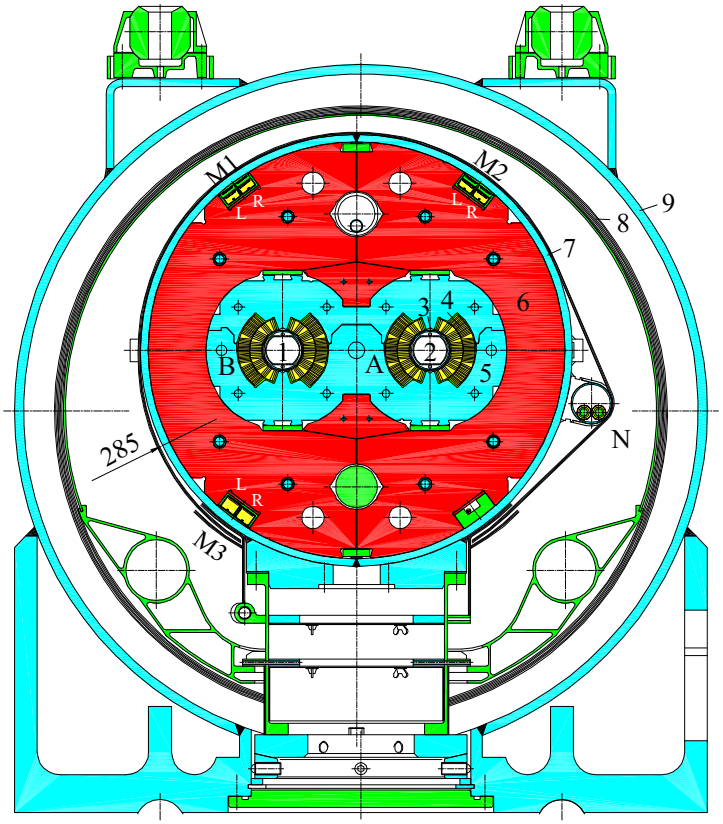


Fig. 2: Cross-section of the dipole magnet and cryostat. 1: Aperture 1 (outer). 2: Aperture 2 (inner). 3: Cold-bore and beam-screen. 4: Superconducting coil. 5: Stainless steel collar (non-magnetic). 6: Iron yoke. 7: Shrinking cylinder. 8: Super-insulation, 9: Vaccum vessel. M3: Busbar for the powering of the MB circuits. M1,M2: Busbar for the powering of the focusing and de-focusing quadrupoles. N: Auxiliary bus-bar for the powering of arc-corrector magnets.

neglected. Assuming that the radial component of the magnetic flux density B_r at a given reference radius $r = r_0$ inside the aperture of a magnet is measured or calculated as a function of the angular position φ, we get for the Fourier-series expansion of the field $B_r(r_0, \varphi) = \sum_{n=1}^{\infty}(B_n(r_0)\sin n\varphi + A_n(r_0)\cos n\varphi)$, with $A_n(r_0) = \frac{1}{\pi}\int_{-\pi}^{\pi} B_r(r_0,\varphi)\cos n\varphi \mathrm{d}\varphi$ and $B_n(r_0) = \frac{1}{\pi}\int_{-\pi}^{\pi} B_r(r_0,\varphi)\sin n\varphi \mathrm{d}\varphi$ for $n = 1, 2, 3, ...$ If the field components are related to the main field component B_N we get for N=1 (dipole field), N=2 (quadrupole), etc.:

$$B_r(r_0,\varphi) = B_N(r_0)\sum_{n=1}^{\infty}(b_n(r_0)\sin n\varphi + a_n(r_0)\cos n\varphi). \tag{1}$$

The B_n are called the *normal* and the A_n the *skew* components of the field given in Tesla, b_n the normal relative, and a_n the skew relative field components. The latter two are dimensionless and are usually given in units of 10^{-4} at a 17 mm reference radius. In practice, the B_r components are calculated at discrete points $\varphi_k = \frac{k\pi}{P} - \pi$, k =0,1,2,..,2P-1 in the interval $[-\pi, \pi[$ and a discrete Fourier transform is carried out:

$$A_n(r_0) \approx \frac{1}{P}\sum_{k=0}^{2P-1} B_r(r_0,\varphi_k)\cos n\varphi_k, \qquad B_n(r_0) \approx \frac{1}{P}\sum_{k=0}^{2P-1} B_r(r_0,\varphi_k)\sin n\varphi_k. \tag{2}$$

The description of the field quality by means of multipoles, Eq. (1), is perfectly in line with magnetic measurements using so-called harmonic coils, where the periodic variation of the flux linkage in radial or tangential rotating coils is analyzed with a Fast-Fourier Transformation (FFT).

The field of a real superconducting magnet deviates from the ideal shape. The three main error sources (geometrical effect, superconductor magnetization and ramp induced effects) can be associated with three types of errors: Systematic errors (average error over the whole LHC ring and in one single aperture) uncertainty (deviations of the systematic error per dipole magnet production line) and random effects (tolerances).

Systematic errors can be classified as follows:

- Errors caused by the shape of the coil winding that can only approximate the ideal $\cos\Theta$ current distribution (section 2.73).
- Remanent fields caused by the so-called persistent currents, induced in the superconducting filaments during the ramp of the magnets to their nominal field value.
- Eddy currents in the multi-strand conductors (interstrand coupling currents).
- Errors from cross-talk in the asymmetric two-in-one magnet design with its common iron yoke and asymmetric iron saturation effects.
- Cool-down of the structure and resulting deformations of the nominal coil geometry.
- Effects from beam-screen, vacuum channel, cryostat, and fringe fields in the coil-end regions, including the effect of bus-bars and interconnections.
- Coil deformations under electromagnetic forces.

Uncertainty errors include:

- Systematic perturbations arising from manufacturing tooling.
- Variations of the properties of the superconducting cable due to different manufacturing procedures.
- Varying properties of steel in yoke and collar laminations depending on the batch.
- Different assembly procedures at the cold-mass manufacturers.
- Torsion and sagitta of the magnet cold-mass.

The random effects mainly arise from:

- Conductor placement errors due to tolerances on coil parts, e.g. insulation thickness, cable keystoning and size of copper wedges.
- Tolerances on yoke parts, e.g. collar outer shape and yoke laminations.
- Manufacturing tolerances and displacements of coil-blocks due to varying elastic modulus of the coil, coil winding procedure, curing, collaring, yoking, etc.
- Alignment tolerances of the magnet system.

1.2 The ROXIE program

The design and optimization of the LHC magnets is dominated by the requirement of an extremely uniform field (no skew field components a_n, no higher-order normal field components b_n, n = 2,3,...), which is mainly defined by the layout of the superconducting coils. For the field calculation it is necessary to consider even very small geometrical effects, such as those produced by insufficient keystoning of the cable, insulation, coil deformations (due to collaring, cool down, and electro-magnetic forces) and variation of the current density in the cable caused by different cable compaction. If the coils had to be modeled in the finite-element mesh, as is the case in most commercial field computation software, it would be difficult to define the current density as this would require a further subdivision of the conductors into a number of radial layers.

For the 3D case in particular, commercial software has proven hardly appropriate for the field optimization of the LHC magnets. The ROXIE (**R**outine for the **O**ptimization of magnet **X**-subsections,

Inverse field calculation and coil **E**nd design) program package was therefore developed at CERN for the design and optimization of the LHC superconducting magnets. The development of the program was driven by the following main objectives:

- To write an easy-to-use program for the design of superconducting coils in two and three dimensions considering field quality, quench margin, and persistent current multipoles.
- To provide for accurate field calculation routines that are specially suited for the investigation of superconducting magnets.
- To integrate the program into a mathematical optimization environment for field optimization and inverse problem solving.
- To integrate the program into the engineering design procedure through interfaces to Virtual Reality, to CAD/CAM systems (for the making of drawings and manufacturing of coil-end spacers), and through interfaces to commercial structural analysis programs.

The modeling capabilities of the ROXIE program, together with its interfaces to CAD/CAM and its mathematical optimization routines, have inverted the classical design process wherein numerical field calculation is performed for only a limited number of numerical models that only approximate the actual engineering design. ROXIE is now used as an approach towards an integrated design of superconducting magnets. The steps of the integrated design process are as follows:

- Feature-based geometry modeling of the coil and yoke, both in two and three dimensions using only a number of meaningful input data to be supplied by the design engineer. This is a prerequisite for addressing these data as design variables of the optimization problem.
- Conceptual coil design using a genetic algorithm, which allows the treatment of combined discrete and continuous problems (e.g. the change of the number of conductors per block). The applied niching method provides the designer with a number of local optima which can then be studied in detail.
- Subject to a varying magnetic field, currents that screen the interior of the superconducting filaments are generated in their outer region. The relative field errors caused by these currents are highest at injection field level and have to be calculated to allow a subsequent part-compensation by geometrical field errors. Deterministic search algorithms are used for the final optimization of the coil cross-section.
- Minimization of iron-induced multipoles using a two-dimensional version of the coupling method between boundary and finite elements (BEM-FEM) developed by ITE Stuttgart and R. Bosch GmbH, Germany.
- Calculation of the peak voltage and peak temperature during a quench.
- Sensitivity analysis of the optimal design through Lagrange-multiplier estimation and the set-up of payoff tables. This provides an evaluation of the hidden resources of the design.
- Tolerance analysis by calculating Jacobian-Matrices and estimation of the standard deviation of the multipole field errors.
- 3D coil-end geometry and field optimization including the modeling and optimization of the asymmetric connection side, ramp and splice region and external connections.
- 3D field calculation of the saturated iron yoke with 3D BEM-FEM computations.
- Production of drawings by means of a DXF interface for both the cross-sections and the 3D coil-end regions.
- End-spacer design and manufacture by means of interfaces to CAD/CAM (DXF, VDA), rapid prototyping methods (laser sinter techniques), and computer controlled 5-axis milling machines.
- Tracing of manufacturing errors from measured field imperfections, i.e., the minimization of a least-squares error function using the Levenberg-Marquard optimization algorithm.

The design process is described in the example of the LHC main quadrupole in [37]. In this report we

will focus on the foundations of the analytical and numerical field computation required for the design process.

2. Analytical Field Computation

The field in superconducting magnets is dominated by the current distribution within the coils. In particular for the higher order field components, the iron yoke plays the role of a mere shielding device. The design process can therefore be split in two parts: The calculation of the coil field using (semi) analytical methods and a subsequent optimization of the iron yoke.

The field in the magnet aperture is governed by the Laplace equation. As a consequence, knowing the normal and skew field components on a given reference radius as in Eq. (1) allows us to re-constitute the entire field in the aperture. The scaling laws so obtained are all we need for the post-processing of data from harmonic coil measurements.

The field harmonics on some reference radius in the aperture of the magnet can be calculated from the current distribution within each strand of the coil by means of Biot-Savart's law and Taylor series expansion. The result combined with the imaging method for line currents in a cylindrical iron yoke reveals the influence of the yoke on the multipoles of different order, the sensitivity to manufacturing errors and symmetry conditions in the magnet. For practical magnet design, however, the analytical results for all strand currents (up to a few thousands) have to be summed-up, hence a semi-analytical method.

2.1 The field equations

For magnetostatic problems ($\frac{\partial}{\partial t} = 0$), Maxwell's equations reduce to

$$\operatorname{curl} \vec{H} = \vec{J}, \qquad \operatorname{div} \vec{B} = 0\,. \tag{3}$$

which can be solved for a given current distribution if we consider the material relations (for non-moving, isotropic materials) which are also called the constitutive equations: $\vec{B} = \mu(H)\vec{H}$ or $\vec{B} = \mu_0(\vec{H}+\vec{M})$ in case there is a remanent magnetization (permanent magnet material) to be considered. The permeability of free space has the value $\mu_0 = 4\pi \cdot 10^{-7}$ H/m.

2.2 The boundary and interface conditions

Subsequently, the closed domain (either 2D or 3D) in which the electro-magnetic field is to be calculated will be denoted as Ω. The field quantities $\vec{B}$ and $\vec{H}$ satisfy boundary conditions on the boundary Γ of the domain Ω. Two types of boundary conditions (prescribed on the two disjunct parts denoted Γ_{H} and Γ_{B} with $\Gamma = \Gamma_{\mathrm{H}} \cup \Gamma_{\mathrm{B}}$) cover all practical cases:

- On the part Γ_{B} of the boundary the **normal** component of the magnetic flux density is given. On symmetry planes parallel to the field, on far boundaries or on outer boundaries of iron yokes surrounded by air (where it can be assumed that no flux leaves the outer boundary) the normal component of the flux density (denoted $\vec{B}_{\mathrm{n}}$) is zero:

$$\vec{B}_{\mathrm{n}} = \vec{B} \cdot \vec{n} = 0 \qquad \text{on}\,\Gamma_{\mathrm{B}}. \tag{4}$$

- On the part Γ_{H} of the boundary the **tangential** component of the magnetic field is given. In many cases (as on symmetry planes perpendicular to the field) and on (infinitely) permeable iron poles (where the field enters at right angle) the tangential component of the field (denoted $\vec{H}_{\mathrm{t}}$) is zero:

$$\vec{H}_{\mathrm{t}} = 0 \qquad \rightarrow \qquad \vec{H} \times \vec{n} = 0 \qquad \text{on}\,\Gamma_{\mathrm{H}}. \tag{5}$$

Consider now the interface between two permeable domains as displayed in fig. 3. If we apply Ampère's law $\oint \vec{H} \cdot \mathrm{d}\vec{s} = \int_A \vec{J} \cdot \mathrm{d}\vec{A}$, to the loop displayed in fig. 3 (left), and let $h \to 0$, then the enclosed current is zero, as in an infinitesimal small rectangle there cannot be a current flow. Therefore the tangential field components are equal, $H_{\mathrm{t}1} = H_{\mathrm{t}2}$ which is equivalent to writing $\vec{n} \times (\vec{H}_1 - \vec{H}_2) = 0$. Because of $\oint \vec{B} \cdot \mathrm{d}\vec{A} = 0$ we get at the interface the continuity of the normal flux density $B_{\mathrm{n}1} = B_{\mathrm{n}2}$ which can also be written as $\vec{n} \cdot (\vec{B}_1 - \vec{B}_2) = 0$. Now from basic geometry (see fig. 3 right):

$$\frac{\tan \alpha_1}{\tan \alpha_2} = \frac{\frac{B_{\mathrm{t}1}}{B_{\mathrm{n}1}}}{\frac{B_{\mathrm{t}2}}{B_{\mathrm{n}2}}} = \frac{\mu_1 H_{\mathrm{t}1}}{\mu_2 H_{\mathrm{t}2}} = \frac{\mu_1}{\mu_2}\,. \tag{6}$$

For $\mu_2 \gg \mu_1$ it follows that $\tan \alpha_2 \gg \tan \alpha_1$. Therefore for all angles $\pi/2 > \alpha_2 > 0$ we get $\tan \alpha_1 \approx 0$, i.e., the magnetic field exits vertically from a highly permeable medium into a medium with low permeability.

2.3 The Lemmata of Poincaré

From vector-analysis it is known that the curl of an arbitrary vector field is source free, i.e., $\mathrm{div}\,\mathrm{curl}\,\vec{g} = 0$ and that an arbitrary gradient field is curl free, i.e., $\mathrm{curl}\,\mathrm{grad}\,\phi = 0$. Reversal of these statements yields the *Lemmata of Poincaré*:

A source free field $\vec{b}$ can be expressed through a vector potential $\vec{a}$.

$$\mathrm{div}\,\vec{b} = 0 \qquad \rightarrow \qquad \vec{b} = \mathrm{curl}\,\vec{a}. \tag{7}$$

A curl free field $\vec{h}$ can be expressed through a scalar potential ϕ.

$$\mathrm{curl}\,\vec{h} = 0 \qquad \rightarrow \qquad \vec{h} = \mathrm{grad}\,\phi. \tag{8}$$

It is necessary that the domain (usually denoted Ω) is topologically not too difficult (precisely, the domain must be star shaped).

2.4 Magnetic potentials

In the aperture of an accelerator magnet (current free region) both the magnetic scalar-potential as well as the vector-potential can be used to solve the field problem as the field is divergence and curl free. For two-dimensional field problems it follows:

$$\vec{H} = -\,\mathrm{grad}\,\Phi = -\frac{\partial \Phi}{\partial x}\vec{e}_x - \frac{\partial \Phi}{\partial y}\vec{e}_y, \tag{9}$$

$$\vec{B} = \mathrm{curl}\,(A_z \vec{e}_z) = \frac{\partial A_z}{\partial y}\vec{e}_x - \frac{\partial A_z}{\partial x}\vec{e}_y, \tag{10}$$

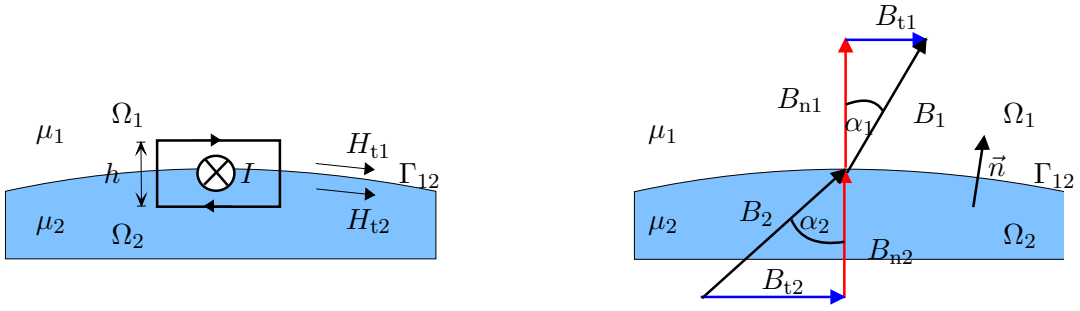


Fig. 3: Interface conditions for permeable media.

As will be shown below, both formulations lead to a scalar Laplace equation. The field governed by the Laplace equation is called *harmonic* and can be expressed by the fundamental solutions of the Laplace equation. Lines of constant vector-potential give the direction of the magnetic field and lines of constant scalar potential define the ideal pole shapes of conventional magnets.

2.41 Reduced magnetic scalar potential

Every vector field can be split into a source free and a curl free part. In case of the magnetic field with $\vec{H} = \vec{H}_\mathrm{s} + \vec{H}_\mathrm{m}$ the curl free part $\vec{H}_\mathrm{m}$ arises from the induced magnetism in ferromagnetic materials and the source free part $\vec{H}_\mathrm{s}$ is the field generated by the prescribed sources and can be calculated directly by means of Biot Savart's law. With $\mathrm{curl}\,\vec{H}_\mathrm{m} = 0$, it follows that $\vec{H} = -\,\mathrm{grad}\,\Phi_\mathrm{m} + \vec{H}_\mathrm{s}$, were Φ_m is called the *reduced* magnetic scalar potential. We get:

$$\begin{aligned} \mathrm{div}\,\vec{B} &= 0\,, && (11)\\ \mathrm{div}\,\mu(-\,\mathrm{grad}\,\Phi_\mathrm{m} + \vec{H}_\mathrm{s}) &= 0\,, && (12)\\ \mathrm{div}\,\mu\,\mathrm{grad}\,\Phi_\mathrm{m} &= \mathrm{div}\,\mu\vec{H}_\mathrm{s}\,. && (13) \end{aligned}$$

The boundary conditions read:

$$\begin{aligned} \mathrm{grad}\,\Phi_\mathrm{m} \times \vec{n} &= \vec{H}_\mathrm{s} \times \vec{n} \quad \text{on } \Gamma_H, && (14)\\ \mathrm{grad}\,\Phi_\mathrm{m} \cdot \vec{n} = \frac{\partial \Phi_\mathrm{m}}{\partial n} &= \vec{H}_\mathrm{s} \cdot \vec{n} \quad \text{on } \Gamma_B. && (15) \end{aligned}$$

While a solution of the obove boundary value problem is possible, the two parts of the magnetic field $\vec{H}_\mathrm{m}$ and $\vec{H}_\mathrm{s}$ tend to be of similar magnitude (but opposite direction) in non-saturated magnetic materials, so that cancellation errors occur in the computation.

2.42 Total magnetic scalar potential

For regions where the current density is zero, $\mathrm{curl}\,\vec{H} = 0$ and the field can be represented by a *total* scalar potential $\vec{H} = -\,\mathrm{grad}\,\Phi_\mathrm{m}$. It follows for regions free of magnetic material and $\mu = \mu_0$:

$$\begin{aligned} \mu_0 \mathrm{div}\ \mathrm{grad}\,\Phi_\mathrm{m} &= 0\,, && (16)\\ \nabla^2 \Phi_\mathrm{m} &= 0\,, && (17) \end{aligned}$$

which is the Laplace equation for the scalar potential. The boundary conditions reduce to:

$$\begin{aligned} \vec{H} \times \vec{n} = \mathrm{grad}\,\Phi_\mathrm{m} \times \vec{n} &= 0 \quad \text{on } \Gamma_H, && (18)\\ \frac{1}{\mu_0}\vec{B}_\mathrm{n} = \mathrm{grad}\,\Phi_\mathrm{m} \cdot \vec{n} = \frac{\partial \Phi_\mathrm{m}}{\partial n} &= 0 \quad \text{on } \Gamma_B. && (19) \end{aligned}$$

Eq. (18), from which follows that Φ_m is constant on the boundary, is called the homogeneous *Dirichlet* boundary condition on Γ_H where $\vec{H}$ is *normal* to the boundary. Eq. (19) that specifies the normal derivative of the system variable is called the homogeneous *Neumann* boundary condition where $\vec{B}$ is *parallel* to the boundary Γ_B.

2.43 Vector-potential formulation with the magnetization seen as an effective current density

With the vector-potential formulation $\vec{B} = \mathrm{curl}\,\vec{A}$ we get

$$\begin{aligned} \mathrm{curl}\,\vec{A} &= \mu_0(\vec{H} + \vec{M}), && (20)\\ \vec{H} &= \frac{1}{\mu_0}\mathrm{curl}\,\vec{A} - \vec{M}, && (21)\\ \frac{1}{\mu_0}\mathrm{curl}\,\mathrm{curl}\,\vec{A} &= \vec{J} + \mathrm{curl}\,\vec{M}\,, && (22)\\ \frac{1}{\mu_0}(-\nabla^2\vec{A} + \mathrm{grad}\,\mathrm{div}\,\vec{A}) &= \vec{J} + \mathrm{curl}\,\vec{M}\,. && (23) \end{aligned}$$

The boundary conditions read:

$$\vec{H} \times \vec{n} = \frac{1}{\mu}(\operatorname{curl} \vec{A}) \times \vec{n} \;\; = \;\; 0 \quad \text{on } \Gamma_H, \tag{24}$$

$$\vec{B}_{\mathrm{n}} = \vec{B} \cdot \vec{n} = \operatorname{curl} \vec{A} \cdot \vec{n} \;\; = \;\; 0 \quad \text{on } \Gamma_B. \tag{25}$$

The condition (25) is equivalent to $\vec{A}_{\mathrm{t}} = 0$, i.e., $\vec{A} \times \vec{n} = 0$ on Γ_B. This is the homogeneous *Neumann* boundary condition on Γ_B. Surface current densities do not appear as long as we have finite conductivity and continuous time dependency. Since the curl (rotation) of a gradient field is zero, the vector-potential is not unique. The gradient of any (differentiable) scalar field ψ can be added without changing the curl of $\vec{A}$:

$$\vec{A}_0 = \vec{A} + \operatorname{grad} \psi. \tag{26}$$

Eq. (26) is called a gauge-transformation between $\vec{A}_0$ and $\vec{A}$. $\vec{B}$ is gauge-invariant as the transformation from $\vec{A}$ to $\vec{A}_0$ does not change $\vec{B}$. The freedom given by the gauge-transformation can be used to set the divergence of $\vec{A}$ to zero $\operatorname{div} \vec{A} = 0$ which (together with additional boundary conditions) makes the vector-potential unique and is is called the Coulomb gauge, as it leads to a Poisson type equation for the magnetic vector-potential. From Eq. (23) we get after incorporating the Coulomb gauge:

$$\nabla^2 \vec{A} = -\mu_0 (\vec{J} + \operatorname{curl} \vec{M}) \; . \tag{27}$$

2.44 Vector-potential formulation for field dependent permeability

An equivalent (but different) formulation for the total vector-potential can be obtained from

$$\operatorname{curl} \frac{1}{\mu} \operatorname{curl} \vec{A} = \vec{J}. \tag{28}$$

where the iron magnetization is taken into account in a multiplicative way through the permeability μ which now depends nonlinearly on the magnetic field and hence $\vec{B} = \mu(\vec{H})\vec{H}$. Introducing a penalty term [34] subtracted from Eq. (28), yields

$$\operatorname{curl} \frac{1}{\mu} \operatorname{curl} \vec{A} - \; \operatorname{grad} \; \frac{1}{\mu} \operatorname{div} \; \vec{A} = \vec{J}. \tag{29}$$

2.5 The Laplace equation

From vector-analysis we know that $\nabla^2 \vec{A}$ reads in Cartesian coordinates:

$$\nabla^2 \vec{A} = (\nabla^2 \vec{A}_x)\vec{e}_x + (\nabla^2 \vec{A}_y)\vec{e}_y + (\nabla^2 \vec{A}_z)\vec{e}_z \; . \tag{30}$$

The Laplace operator acting on a vector in Cartesian coordinates yields a vector that can also be obtained through the application of the operator on the components, i.e., $\nabla^2 A_x = -\mu_0 (J_x + (\operatorname{curl} \vec{M})_x)$ and so on, whereas this is not the case in cylindrical coordinates:

$$\nabla^2 \vec{A} = (\nabla^2 A_r - \frac{1}{r^2} A_r - \frac{2}{r^2} \frac{\partial A_\varphi}{\partial \varphi}) \vec{e}_r + (\nabla^2 A_\varphi - \frac{1}{r^2} A_\varphi + \frac{2}{r^2} \frac{\partial A_r}{\partial \varphi}) \vec{e}_\varphi + \nabla^2 A_z \vec{e}_z. \tag{31}$$

In the two-dimensional case with no dependence on z, ($\frac{\partial}{\partial z} = 0$) and $\vec{J} = J_z \vec{e_z}$, $\vec{A}$ has only a z-component and the Coulomb gauge is automatically fulfilled. Then we get the scalar Poisson differential equation from Eq. (27) or Eq. (29):

$$\nabla^2 A_z = -\mu_0 J_z. \tag{32}$$

For current-free regions Eq. (32) reduces to the Laplace equation which reads in cylindrical coordinates

$$\nabla^2 A_z = r^2 \frac{\partial^2 A_z}{\partial r^2} + r \frac{\partial A_z}{\partial r} + \frac{\partial^2 A_z}{\partial \varphi^2} = 0. \tag{33}$$

2.6 Harmonic fields

A solution of the homogeneous differential equation (33) can be derived with the method of separation and reads: $A_z(r,\varphi) = \sum_{n=1}^{\infty}(E_n r^n + F_n r^{-n})(G_n \sin n\varphi + H_n \cos n\varphi)$. These fields are called *harmonic*. Considering that the field is finite at r = 0, the F_n have to be zero for the vector-potential inside the aperture of the magnet while for the solution in the area outside the coil all E_n vanish. Rearranging Eq. (2.6) yields the vector-potential in the aperture:

$$A_z(r,\varphi) = \sum_{n=1}^{\infty} r^n(\mathcal{C}_n \sin n\varphi - \mathcal{D}_n \cos n\varphi), \tag{34}$$

and the field components can be expressed as

$$B_r(r,\varphi) = \frac{1}{r}\frac{\partial A_z}{\partial \varphi} = \sum_{n=1}^{\infty} n r^{n-1}(\mathcal{C}_n \cos n\varphi + \mathcal{D}_n \sin n\varphi), \tag{35}$$

$$B_\varphi(r,\varphi) = -\frac{\partial A_z}{\partial r} = -\sum_{n=1}^{\infty} n r^{n-1}(\mathcal{C}_n \sin n\varphi - \mathcal{D}_n \cos n\varphi). \tag{36}$$

The solution in Cartesian coordinates can be obtained from the simple transformations $B_x = B_r \cos\varphi - B_\varphi \sin\varphi$, $B_y = B_r \sin\varphi + B_\varphi \cos\varphi$. The coefficients are not known at this stage. They are defined through the (given) boundary conditions on some reference radius or can be calculated from the Fourier series expansion of the numerically calculated (or measured) field in the aperture using the relations

$$A_n = n r_0^{n-1}\mathcal{C}_n \qquad \text{and} \qquad B_n = n r_0^{n-1}\mathcal{D}_n. \tag{37}$$

We finally get for the field components at a given reference radius r_0:

$$B_r(r_0,\varphi) = \sum_{n=1}^{\infty}(B_n \sin n\varphi + A_n \cos n\varphi) = B_N \sum_{n=1}^{\infty}(b_n \sin n\varphi + a_n \cos n\varphi), \tag{38}$$

$$B_\varphi(r_0,\varphi) = \sum_{n=1}^{\infty}(B_n \cos n\varphi - A_n \sin n\varphi) = B_N \sum_{n=1}^{\infty}(b_n \cos n\varphi - a_n \sin n\varphi). \tag{39}$$

The small b_n, a_n are the multipoles related to the main field B_N which is B_1 for the dipole, B_2 for the quadrupole, etc. B_n are given in Tesla and b_n are dimensionless and usually given in units of 10^{-4} at a reference radius of 17 mm. In some documents, e.g., [41], the field strength of the LHC magnets is defined as

$$\mathcal{B}_n = \frac{B_n}{r_0^{n-1}} \tag{40}$$

and $\mathcal{B}_n$ are given in T, T/m, T/m^2 etc. For the scaling of different reference radii we get

$$A_n(r_1) = (\frac{r_1}{r_0})^{n-1} A_n(r_0), \qquad B_n(r_1) = (\frac{r_1}{r_0})^{n-1} B_n(r_0), \tag{41}$$

$$a_n(r_1) = (\frac{r_1}{r_0})^{n-N} a_n(r_0), \qquad b_n(r_1) = (\frac{r_1}{r_0})^{n-N} b_n(r_0). \tag{42}$$

2.61 Dipole, quadrupole and sextupole flux density distributions

Each value of the integer n in the solution, Eq. (38), of the Laplace equation corresponds to a different flux distribution generated by different magnet geometries. The three lowest values, n=1,2, and 3 correspond to a dipole, quadrupole and sextupole flux density distribution. For the dipole field (n=1) we

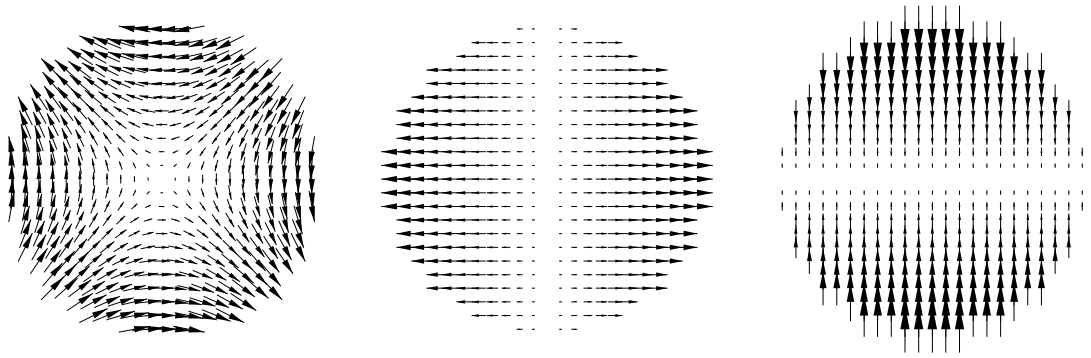


Fig. 4: Field and force distribution inside the aperture of an ideal quadrupole. Left: Magnetic induction of the normal quadrupole ($B_y = gx$, gradient g negative, $B_x = gy$). Middle: x-component of the electromagnetic force field $(\vec{v_z} \times \vec{B})_x$ on a proton beam parallel to the z-axis into positive z-direction. While this quadrupole is defocusing in the y-plane, it is focusing in the x-plane, see figure on the right where $(\vec{v_z} \times \vec{B})_y$ is displayed.

get

$$
\begin{aligned}
B_r &= \mathcal{C}_1 \cos\varphi + \mathcal{D}_1 \sin\varphi, && (43)\\
B_\varphi &= -\mathcal{C}_1 \sin\varphi + \mathcal{D}_1 \cos\varphi, && (44)\\
B_x &= \mathcal{C}_1, && (45)\\
B_y &= \mathcal{D}_1. && (46)
\end{aligned}
$$

This is a simple, constant field distribution according to the values of $\mathcal{C}_1$ and $\mathcal{D}_1$.

Remark: Beware that we have not yet studied the conditions necessary to obtain such a field distribution. We will address this issue in the next chapter.

For the pure quadrupole, n=2 (represented in fig. 4) we get from Eq. (35) and (36):

$$
\begin{aligned}
B_r &= 2\,r\,\mathcal{C}_2 \cos 2\varphi + 2\,r\,\mathcal{D}_2 \sin 2\varphi, && (47)\\
B_\varphi &= -2\,r\,\mathcal{C}_2 \sin 2\varphi + 2\,r\,\mathcal{D}_2 \cos 2\varphi, && (48)\\
B_x &= 2(\mathcal{C}_2 x + \mathcal{D}_2 y), && (49)\\
B_y &= 2(-\mathcal{C}_2 y + \mathcal{D}_2 x). && (50)
\end{aligned}
$$

The amplitudes of the horizontal and vertical components vary linearly with the displacements from the origin, i.e., the gradient is constant. With a zero induction in the origin, the distribution provides linear focusing of the particles. The components of the magnetic fields are coupled, i.e., the distribution in both planes cannot be made independent of each other. Consequently a quadrupole focusing in one plane will defocus in the other. For a normal quadrupole with $\mathcal{C}_2 = 0$ we can calculate the field at some position $x' = x + \Delta x$ which yields

$$B_y = 2\mathcal{D}_2 x + 2\mathcal{D}_2 \Delta x. \quad (51)$$

For a displaced quadrupole, the field contains a constant term, i.e., the second term in Eq. (51). This effect is called feed-down. For the case of the pure sextupole (n=3) we get:

$$
\begin{aligned}
B_r &= 3\,r^2\,\mathcal{C}_3 \cos 3\varphi + 3\,r^2\,\mathcal{D}_3 \sin 3\varphi, && (52)\\
B_\varphi &= -3\,r^2\,\mathcal{C}_3 \sin 3\varphi + 3\,r^2\,\mathcal{D}_3 \cos 3\varphi, && (53)\\
B_x &= 3\mathcal{C}_3(x^2 - y^2) + 6\mathcal{D}_3 xy, && (54)\\
B_y &= -6\mathcal{C}_3 xy + 3\mathcal{D}_3(x^2 - y^2), && (55)
\end{aligned}
$$

which is represented in fig. 5. Displacement of a sextupole magnet creates both a quadrupole and a dipole field component

$$B_y = 3\mathcal{D}_3(x'^2 - y^2) = 3\mathcal{D}_3(x^2 - y^2) + 6\mathcal{D}_3 x\Delta x + \mathcal{D}_3\Delta x^2. \tag{56}$$

Remark: The treatment of each harmonic separately is a mathematical abstraction. In practical situations many harmonics will be present, i.e, many of the coefficients $\mathcal{C}_n$ and $\mathcal{D}_n$ will be non-vanishing. A successful magnet design will, however, minimize the unwanted terms.

2.7 Coil field of superconducting magnets

The problem remains how to calculate the field harmonics from a given current distribution. It is reasonable to focus on the fields generated by line-currents, since the field of any current distribution over an arbitrary cross-section can be approximated by summing the fields of a number of line-currents distributed within the cross-section. The harmonic field, however, cannot account for line currents. We need a special solution of the inhomegenious (Poisson) differential equation.

2.71 The field of a line current

In two dimensions the particular solution of the Poisson equation (32) is

$$A_z = \int_A -\frac{\mu_0 J_z}{2\pi}\ln\left(\frac{R}{a}\right)\mathrm{d}A' = -\frac{\mu_0 I}{2\pi}\ln\left(\frac{R}{a}\right). \tag{57}$$

With the source point $\vec{r}\,' = (r_i, \Theta)$, the field point $\vec{r} = (r_0, \varphi)$, $|\vec{R}| = |\vec{r} - \vec{r}\,'|$, and an arbitrary reference radius a, see fig. 6, the cosine law $R^2 = r_i^2 + r_0^2 - 2r_i r_0\cos(\varphi - \Theta)$ can be rewritten as [28]
$R^2 = r_i^2\,(1 - \frac{r_0}{r_i}e^{i(\varphi-\Theta)})\,(1 - \frac{r_0}{r_i}e^{-i(\varphi-\Theta)})$.
Therefore

$$\ln\left(\frac{R}{a}\right) = \ln\left(\frac{r_i}{a}\right) + \frac{1}{2}\ln\left(1 - \frac{r_0}{r_i}e^{i(\varphi-\Theta)}\right) + \frac{1}{2}\ln\left(1 - \frac{r_0}{r_i}e^{-i(\varphi-\Theta)}\right). \tag{58}$$

With the Taylor series expansion of $\ln(1-x)$ which gives for $|x| < 1$, i.e., for $r_0 < r_i$ inside the aperture of the magnet, $\ln(1-x) = -\sum_{n=1}^{\infty}\frac{1}{n}x^n$, Eq. (57) can be rewritten as

$$A_z = -\frac{\mu_0 I}{2\pi}\ln\left(\frac{r_i}{a}\right) + \frac{\mu_0 I}{2\pi}\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{r_0}{r_i}\right)^n\cos\left(n(\varphi - \Theta)\right). \tag{59}$$

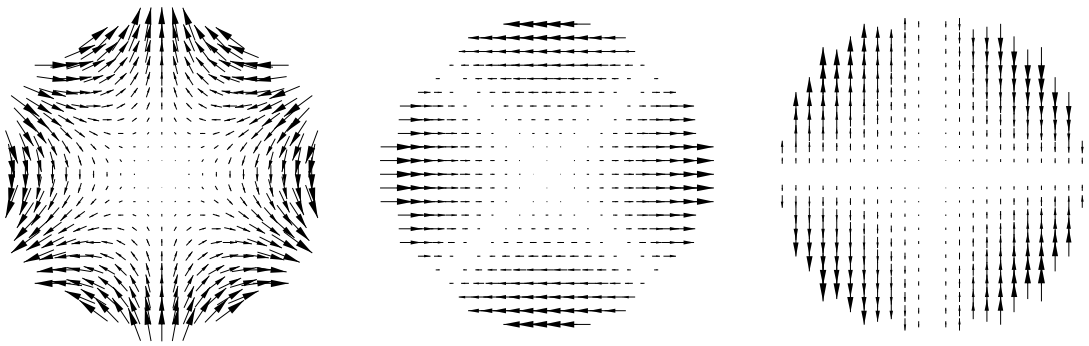


Fig. 5: Field and force distribution inside the aperture of an ideal sextupole. Left: Magnetic induction of the normal sextupole ($B_y = g_s(x^2 - y^2)$, gradient g_s negative, $B_x = 2g_s xy$). Middle: x-component of the electromagnetic force field $(\vec{v}_z \times \vec{B})_x$ acting on a beam parallel to the z-axis into positive z-direction. Right: y-component of the electromagnetic force field $(\vec{v}_z \times \vec{B})_y$.

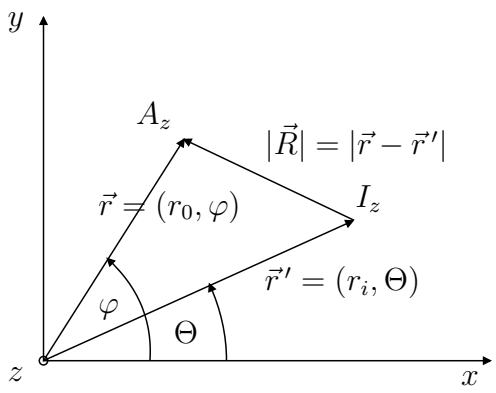


Fig. 6: coordinate system for the calculation of the potential of a line current.

The radial component of the magnetic field in the point (r_0, φ) is then

$$B_r = \frac{1}{r}\frac{\partial A_z}{\partial \varphi} = -\frac{\mu_0 I}{2\pi}\sum_{n=1}^{\infty}(\frac{r_0^{n-1}}{r_i^n})\sin\left(n(\varphi-\Theta)\right) \tag{60}$$

$$= -\frac{\mu_0 I}{2\pi}\sum_{n=1}^{\infty}(\frac{r_0^{n-1}}{r_i^n})(\sin n\varphi\,\cos n\Theta - \cos n\varphi\,\sin n\Theta). \tag{61}$$

Comparison of the coefficients with equation (38) yields (for $r_0 < r_i$):

$$B_n(r_0) = -\frac{\mu_0 I}{2\pi}\frac{r_0^{n-1}}{r_i^n}\cos n\Theta, \qquad A_n(r_0) = \frac{\mu_0 I}{2\pi}\frac{r_0^{n-1}}{r_i^n}\sin n\Theta. \tag{62}$$

2.72 The imaging method for line currents in a cylinder with constant permeability

The effect of an iron yoke with constant permeability and perfect circular inner shape with radius $R_{\rm yoke}$ can be taken into account by means of image currents of the strength and radial position

$$I\,' = \frac{\mu_{\rm r}-1}{\mu_{\rm r}+1}I \qquad r_i\,' = \frac{R_{\rm yoke}^2}{r_i}. \tag{63}$$

which are located at the same angular position. Fig. 7 (left) shows the field distribution for a superconducting coil in a (non-saturated) iron yoke of cylindrical inner shape and the representation of the iron yoke with image current (right).

Remark: Notice that only the field in the aperture of the magnet can be calculated with the imaging method and that the total current inside the yoke has to be zero. Furthermore the yoke has to be of constant permeability and thus local saturation effects cannot be considered.

Including the effect of the imaging currents, the normal multipole coefficients inside the aperture of the magnet reads for **one** line current:

$$\boxed{B_n(r_0) = -\frac{\mu_0 I}{2\pi}\frac{r_0^{n-1}}{r_i^n}\left(1+\frac{\mu_{\rm r}-1}{\mu_{\rm r}+1}(\frac{r_i}{R_{\rm yoke}})^{2n}\right)\cos n\Theta.} \tag{64}$$

This is an important results as it allows to calculate the harmonic content of a field generated by a number of arbitrarily placed current carrying conductors by adding the terms in Eq. (64).

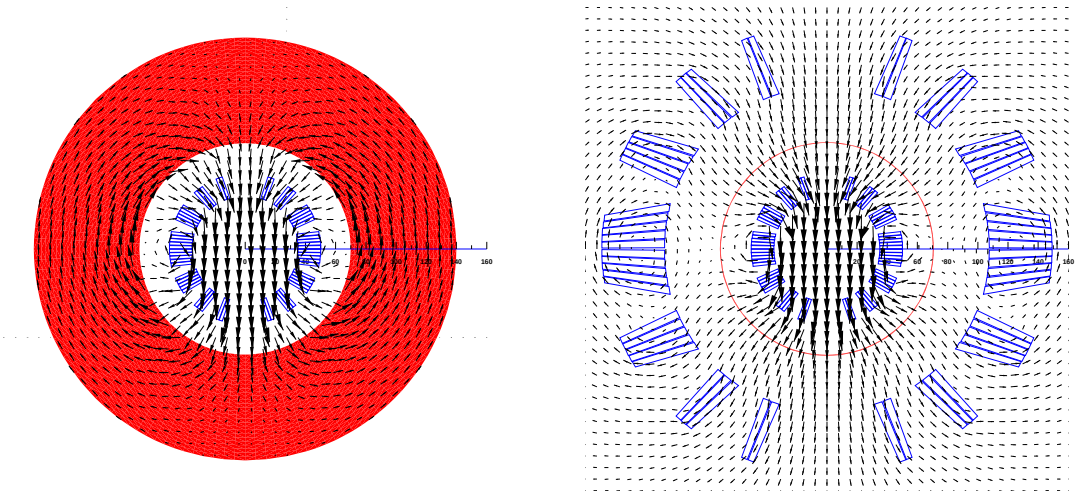


Fig. 7: Left: Field vectors for a superconducting coil in a (non-saturated) iron yoke of cylindrical inner shape (μ_r=2000, constant). Right: The representation of the iron yoke with image currents.

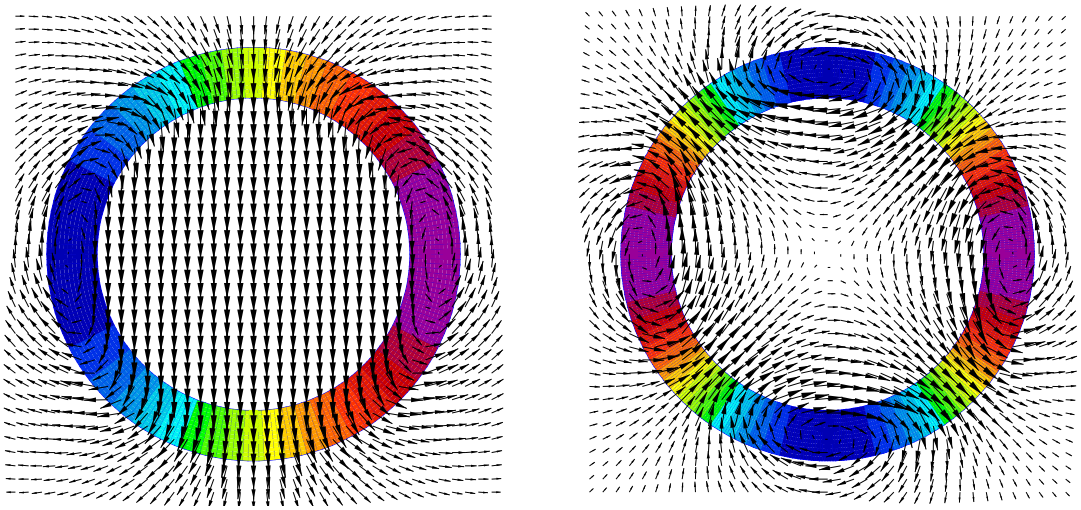


Fig. 8: Shells with $\cos\Theta$ (left) and $\cos 2\Theta$ (right) dependent current density.

2.73 The generation of pure multipole fields

Consider a current shell $r_\mathrm{i} < r < r_\mathrm{e}$ with a current density varying with the azimuthal angle Θ, $J(\Theta) = J_0 \cos m\Theta$, then we get for the B_n components

$$B_n(r_0) = \int_{r_\mathrm{i}}^{r_\mathrm{e}} \int_0^{2\pi} -\frac{\mu_0 J_0 r_0^{n-1}}{2\pi r^n} \left(1 + \frac{\mu_\mathrm{r}-1}{\mu_\mathrm{r}+1}(\frac{r}{R_\mathrm{yoke}})^{2n}\right) \cos m\Theta \cos n\Theta \, r\mathrm{d}\Theta\mathrm{d}r\,. \tag{65}$$

With $\int_0^{2\pi} \cos m\Theta \cos n\Theta \mathrm{d}\Theta = \pi\delta_{m,n}(m,n \neq 0)$ it follows that the current shell produces a pure $2m$-polar field and in the case of the dipole ($m = n = 1$) one gets

$$B_1(r_0) = -\frac{\mu_0 J_0}{2}\left((r_\mathrm{e} - r_\mathrm{i}) + \frac{\mu_\mathrm{r}-1}{\mu_\mathrm{r}+1}\frac{1}{R^2_\mathrm{yoke}}\frac{1}{3}(r_\mathrm{e}^3 - r_\mathrm{i}^3)\right). \tag{66}$$

Obviously, since $\int_0^{2\pi} \cos m\Theta \sin n\Theta \mathrm{d}\Theta = 0$, all A_n components vanish. A shell with $\cos\Theta$ and $\cos 2\Theta$ dependent current density is displayed in figure 8.

2.74 Coil-Block arrangements

As the conductors are keystoned with an insufficient angle to allow for perfect $\cos\Theta$ geometries and the conductors are connected in series, i.e., carry the same current, coil-blocks separated by copper wedges are designed in order to approximate the ideal current distribution. Fig. 9 shows the contribution of the strand currents in a superconducting dipole coil to the B_3 and the B_5 field component as a visualization of Eq. (64). The field errors scale with $1/r^n$ where n is the order of the multipole and r is the mid radius of the coil.

With equation (64), a semi-analytical method for calculating the fields in superconducting magnets is given. The iron yoke is represented by image currents (the second term in the parentheses). At low field level, when the saturation of the iron yoke is low, this is a sufficient method for optimizing the coil cross-section. Under that assumption some important conclusions can be drawn:

- For a coil without iron yoke the field errors scale with $1/r^n$ where n is the order of the multipole and r is the mid radius of the coil. It is clear, however, that an increase of coil aperture causes a linear drop in dipole field. Other limitations of the coil size are the beam distance, the electromagnetic forces, yoke size, and the stored energy which results in an increase of the hot-spot temperature during a quench.
- For certain symmetry conditions in the magnet, some of the multipole components vanish, i.e. for an up-down symmetry in a dipole magnet (positive current I_0 at (r_0, Θ_0) and at $(r_0, -\Theta_0)$) no A_n terms occur. If there is an additional left-right symmetry, only the odd $B_1, B_3, B_5, B_7, ..$ components remain.
- The relative contribution of the iron yoke to the total field (coil field plus iron magnetization) is for a non-saturated yoke ($\mu_\mathrm{r} \gg 1$) approximately $(1+(\frac{R_\mathrm{yoke}}{r})^{2n})^{-1}$. For the main dipoles with a mean coil radius of $r = 43.5$ mm and a yoke radius of $R_\mathrm{yoke} = 89$ mm we get for the B_1 component a 19% contribution from the yoke, whereas for the B_5 component the influence of the yoke is only 0.07%.

It is therefore appropriate to optimize for higher harmonics first using analytical field calculation, and include the effect of iron saturation on the lower-order multipoles only at a later stage.

2.8 Complex Analysis Methods for Magnet Design

With the spread of numerical field computation software now running on desk- and laptop computers, complex analysis methods have become less important for magnet design. However, important applications abound: Complex representation of the field quality (often given as definition of the multipoles), scaling of field measurements and the compensation of the so-called feed down effect due to off-centering

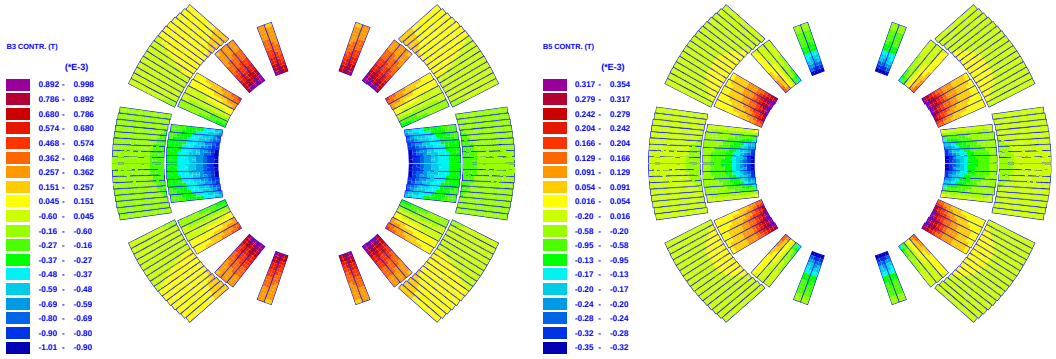


Fig. 9: Contribution of the strand current to the B_3 (left), B_5 (right) field component (visualization of Eq. (64)) at $r_0 = 17$ mm, for a coil-block arrangement which approximates the ideal current distribution.

of the measurement coils (or miss-alignment of magnet chains), and the calculation of fields due to intersecting circles or ellipses as the ideal current distribution for dipole and quadrupole magnets and the basis for a macroscopic model of shielding currents in superconducting filaments.

2.81 Complex potentials

In current and magnetization free, two-dimensional regions both the magnetic scalar-potential as well as the vector-potential can be used to solve Maxwell's equations, equations (9) and (10). Therefore

$$\frac{\partial A_z}{\partial y} = -\mu_0 \frac{\partial \Phi}{\partial x} \qquad \text{and} \qquad \frac{\partial A_z}{\partial x} = \mu_0 \frac{\partial \Phi}{\partial y}. \tag{67}$$

which are the Cauchy-Riemann differential equations of the complex potential

$$W(z) = u(x,y) + iv(x,y) = A_z(x,y) + i\mu_0 \Phi(x,y), \tag{68}$$

which is an analytic function of $z = x + iy$. Derivatives of W are also analytic functions of z.

$$-\frac{\mathrm{d}W}{\mathrm{d}z} = -\frac{\partial A_z}{\partial x} - i\mu_0 \frac{\partial \Phi}{\partial x} = B_y(x,y) + iB_x(x,y) = \widetilde{B} \tag{69}$$

where the field components appear as real functions of x and y and the Cauchy-Riemann equations between them express Maxwell's equations for the two-dimensional case with zero current density. From Eq. (67) follows directly that the Laplace equation holds for both A_z and Φ. The complex potential at a point $z_0 = r_0 e^{i\varphi}$ due to a line current at the position $z = re^{i\Theta}$ is given by

$$W(z_0) = -\frac{\mu_0 I}{2\pi} \ln\left(\frac{z_0 - z}{z_a}\right) \tag{70}$$

where z_a is an arbitrary complex reference point and therefore

$$-\frac{\mathrm{d}W}{\mathrm{d}z} = \widetilde{B} = -\frac{\mu_0 I}{2\pi} \frac{1}{z - z_0} = -\frac{\mu_0 I}{2\pi} \frac{1}{(x - x_0) + i(y - y_0)}, \tag{71}$$

which yields the well known equations for the magnetic field

$$B_y = -\frac{\mu_0 I}{2\pi} \frac{x - x_0}{(x - x_0)^2 + (y - y_0)^2}, \qquad B_x = \frac{\mu_0 I}{2\pi} \frac{y - y_0}{(x - x_0)^2 + (y - y_0)^2}. \tag{72}$$

With the Taylor-series expansion for $|z_0| < |z|$, i.e., inside the circular aperture of the magnet,

$$\frac{1}{z - z_0} = \frac{1}{z(1 - \frac{z_0}{z})} = \frac{1}{z} \sum_{n=1}^{\infty} \left(\frac{z_0}{z}\right)^{n-1} = \sum_{n=1}^{\infty} \frac{z_0^{n-1}}{z^n} \tag{73}$$

we get $B_y + iB_x = -\frac{\mu_0 I}{2\pi} \sum_{n=1}^{\infty} \frac{z_0^{n-1}}{z^n}$. Bringing together the multipole coefficients, eqns. (38) and (39) yield

$$\sum_{n=1}^{\infty} (B_n + iA_n) \frac{z_0^n}{r_0^n} = (B_\varphi + iB_r) = (B_y + iB_x) \frac{z_0}{r_0} = -\frac{\mu_0 I}{2\pi} \sum_{n=1}^{\infty} \frac{z_0^{n-1}}{z^n} \frac{z_0}{r_0}. \tag{74}$$

Therefore $B_n + iA_n = -\frac{\mu_0 I}{2\pi} \frac{r_0^{n-1}}{z^n}$ which is identical to Eq. (64). From Eq. (74) it follows:

$$\boxed{B_y + iB_x = \sum_{n=1}^{\infty} (B_n + iA_n) \left(\frac{z_0}{r_0}\right)^{n-1}} \tag{75}$$

which is sometimes given as a definition of the multipole coefficients and which allows an easy reconstruction of the cartestian components of the magnetic flux density from the calculated or measured field components.

2.82 Feed-down

An interesting example for the use of the complex field representation is the calculation of the ***feed-down*** effect due to an off-centering of the measurement coil with respect to the magnet axis (or a misalignment of the magnet with respect to the beam axis). The transformation law for the field harmonics $c_n \rightarrow c'_n$; $c_n = b_n + i\,a_n$ can be derived for a translation of the reference frame (center of the measurement coil) $z \rightarrow z'$: $z = z' + \Delta z$; $z = x + iy$ as follows: As the field components in both coordinate systems have to be identical we can write

$$\sum_{n=1}^{\infty} c_n \left(\frac{z}{r_0}\right)^{n-1} = \sum_{n=1}^{\infty} c'_n \left(\frac{z'}{r_0}\right)^{n-1} = B_y + i\,B_x \ . \tag{76}$$

The transformation law for the field harmonics then reads (for the proof see Appendix A):

$$\boxed{c'_n = \sum_{k=n}^{\infty} c_k \, \frac{(k-1)!}{(k-n)!\,(n-1)!} \left(\frac{\Delta z}{r_0}\right)^{k-n} .} \tag{77}$$

For example, off-centering of the measurement coil creates a (measured) quadrupole field component which results from the natural sextupole component feed-down. Equation (77) can also be used to determine the center of the measurement coil, e.g., through the elimination of the measured dipole component in quadrupole magnets. In bending magnets the b_{11} component is very insensitive to manufacturing errors while b_{10} is near zero. If c_{11}, c_{12} and c_{13} at a reference radius r_0 are known with sufficient accuracy, the displacement Δz can be calculated from a truncated series of equation (77). Notice, that this holds only for small displacements $\Delta z \ll r_0$.

2.83 The Residue theorem and the field of intersecting ellipses

The *residue theorem* states that if a complex function is meromorph in a domain Ω (i.e., analytic in Ω except at a finite number of isolated singularities) and analytic on the boundary $\partial\Omega$ of Ω then it holds:

$$\oint_{\partial\Omega} f(z)\mathrm{d}z = 2\pi i \sum_n R(z_n) \tag{78}$$

where the $R(z_n)$ are the residuals of the poles of the function at the points z_n, i.e., the constant in the numerator of the single pole, if the complex function is given in the form

$$f(z) = \frac{R(z_n)}{z - z_n} + g(z_n)\,. \tag{79}$$

For the calculation of the two-dimensional magnetic field generated by intersecting ellipses or circles, Beth [7], uses the complex potential

$$F(z) = \widetilde{H} - \frac{1}{2} J(x,y) z^* \tag{80}$$

with $\widetilde{H} = H_y + iH_x$ and the complex conjugate $z^* = x - iy$. Eq. 80 can be rewritten as

$$F(z) = \widetilde{H} - \frac{1}{2} J z^* = H_y + iH_x - \frac{1}{2} J(x - iy) = \underbrace{H_y - \frac{1}{2} Jx}_{u(x,y)} + i\underbrace{(H_x + \frac{1}{2} Jy)}_{v(x,y)}\,. \tag{81}$$

Calculating the partial derivatives of the real valued functions $u(x,v)$ and $v(x,y)$ and inserting into the Cauchy Riemann equations yields the two equations

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = J\,, \qquad \frac{\partial H_y}{\partial y} + \frac{\partial H_x}{\partial x} = 0\,. \tag{82}$$

which are nothing but the Poisson and Laplace equations in two-dimensional Cartesian coordinates. It can thus be shown that the complex potential $F(z) = \widetilde{H} - \frac{1}{2}J(x,y)z^*$ is analytic and that it can be used to solve the field distribution inside current carrying conductors using the residue theorem.

Let a be the minor and b the major half-axis of an infinitely long, elliptic conductor carrying uniform current density J. Let the points on the boundary C' of the ellipse be denoted by the small $z = x + iy$ and the field point by z_0. The domain inside C' is denoted by the capital Z_{in} and the domain outside the ellipse by Z_{out}. This gives the following expression for $\widetilde{H}_{\text{in}}(z_0)$ for $z_0 = x_0 + iy_0 \in Z_{\text{in}}$ [7]:

$$\widetilde{H}_{\text{in}}(x_0, y_0) = \frac{J}{a+b}(bx_0 - iay_0) \tag{83}$$

and the components of the magnetic field within an elliptic conductor with constant current density J are:

$$H_x(x_0, y_0) = -J\frac{a}{a+b}y_0\,, \qquad H_y(x_0, y_0) = J\frac{b}{a+b}x_0\,. \tag{84}$$

For the round conductor $a = b$ we get the well known relations $H_x(x_0, y_0) = -\frac{1}{2}Jy_0$ and $H_y(x_0, y_0) = \frac{1}{2}Jx_0$.

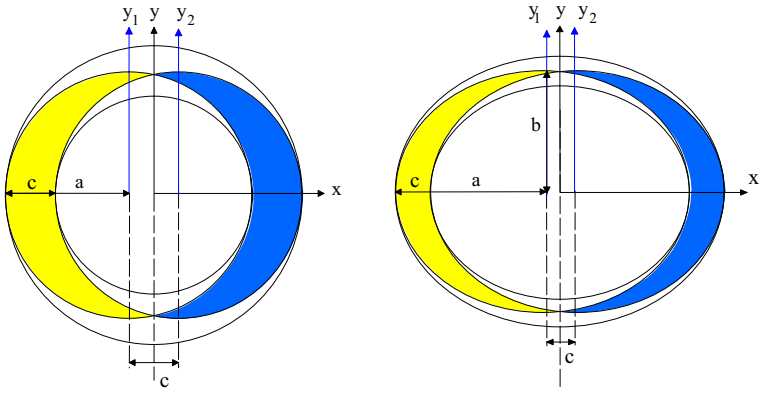


Fig. 10: Intersecting circles and ellipses that create an ideal dipolar fi eld inside the aperture.

As an application of the above result it is now straight forward to show that in the center of two intersecting ellipses or circles with opposite current density a homogeneous field is generated. With the two intersecting circles shifted by c and with the two local coordinate systems $x_1 = x + c/2$, $x_2 = x - c/2$, $y_1 = y$, $y_2 = y$ (positive current in circle 1) it follows:

$$B_x = B_x^{(1)} + B_x^{(2)} = -\mu_0\frac{1}{2}J(y_1 - y_2) = 0, \tag{85}$$

$$B_y = B_y^{(1)} + B_y^{(2)} = \mu_0\frac{1}{2}J\left((x + \frac{c}{2}) - (x - \frac{c}{2})\right) = \mu_0\frac{1}{2}Jc\,. \tag{86}$$

For the intersecting ellipses the B_x component is zero due to the same reasoning, and the y-component takes the value:

$$\boxed{B_y = \mu_0 Jc\frac{b}{a+b}\,.} \tag{87}$$

Remark: This important result has two implications: First it paves the way for designing block coil magnets with constant current density in the conductors and secondly it can reproduce shielding current densities in hard superconductors subjected to changing excitation field, i.e., a macroscopic way to model persistent currents, chapter 5.

3. Numerical Field Computation

Magnets for particle accelerators have always been a key application of numerical methods in electromagnetism. Hornsby [19], in 1963, developed a code based on the finite difference method for the solving of elliptic partial differential equations and applied it to the design of magnets. Winslow [44] created the computer code TRIM (Triangular Mesh) with a discretization scheme based on an irregular grid of plane triangles by using a generalized finite difference scheme. He also introduced a variational principle and showed that the two approaches lead to the same result. In this respect, the work can be viewed as one of the earliest examples of the finite element method applied to the design of magnets. The POISSON code which was developed by Halbach and Holsinger [18] was the successor of this code and was still applied for the optimization of the superconducting magnets for the LHC during the early design stages. In the early 1970's a general purpose program (GFUN) for static fields had been developed by Newman, Turner and Trowbridge that was based on the magnetization integral equation and was applied to magnet design.

When the LHC magnets are ramped to their nominal field of 8.33 T in the aperture, the yoke is highly saturated, and numerical methods have to be used to replace the imaging method. In this case it is advantageous to use numerical methods that allow a distinction between the coil-field and the iron magnetization effects, to confine both modeling problems on the coils and FEM-related numerical errors to the 20% of field contribution from the iron magnetization. The integral equation method of GFUN would be appropriate, however, it leads to a very large, fully populated matrix of the linear equation system.

The method of coupled boundary-elements/finite-elements (BEM-FEM), developed by Fetzer, Haas and Kurz at the University of Stuttgart, Germany, combines a finite-element description using incomplete quadratic (20-node) elements and a gauged total vector-potential formulation for the interior of the magnetic parts, and a boundary element formulation for the coupling of these parts to the exterior, which includes excitational coil fields. This implies that the air regions need not to be meshed at all.

Of course the method is not limited to the accelerator magnet design. In order to explain the principles of numerical field computation the principle steps of the finite element and boundary element techniques are outlined here under:

- Formulation of the physical laws by means of partial differential equations.
- Transformation of these equations into an integral equation with the weighted residual method.
- Integration by parts (using Green's theorems) in order to obtain the so-called weak integral form. Consideration of the *natural* boundary conditions.
- Discretization of the domain into (higher-order) finite elements.
- Approximation of the solution as a linear-combination of so-called shape functions.
- In case of the finite element (Galerkin) method, using the shape functions as the weighting functions of the weak integral form.
- In case of the boundary-element method, another partial integration of the weak integral form results in an integral equation. Using the fundamental solution of the Laplace operator as weighting functions yields an algebraic system of equations for the unknowns on the domain boundary. In case of the coupled boundary-element/finite-element method, the two domains are coupled through the normal derivatives of the vector-potential on their common boundary.
- Consideration of the *essential* boundary conditions in the resulting equation system.
- Numerical solution of the algebraic equations. A direct solver with Newton iteration is used in the 2D case; the domain decomposition method with a $M(B)$ iteration [15] is applied in the 3D case.

In order to understand the special properties of the numerical methods and the reasoning which leads to their application in the design and optimization of superconducting magnets, it is sufficient to concentrate on some aspects of the formulations. We will therefore omit the treatment of the linear equation system.

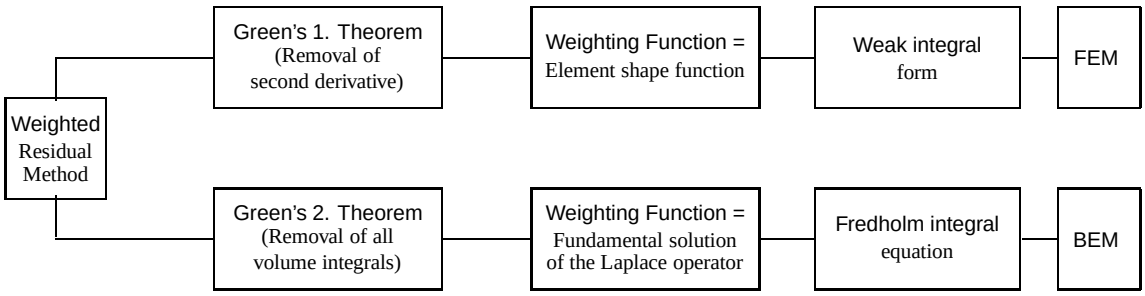


Fig. 11: The role of Green's first and second theorem in the FEM and BEM method.

3.1 Green's identities

The mathematical foundations of numerical field calculation require some knowledge about vector-analysis. We shall recall Green's first and second identity

$$\int_\Omega (\operatorname{grad}\phi \cdot \operatorname{grad}\psi + \phi\nabla^2\psi)\mathrm{d}\Omega = \oint_\Gamma \phi \operatorname{grad}\psi \cdot \vec{n}\mathrm{d}\Gamma \tag{88}$$

$$\int_\Omega (\phi\nabla^2\psi - \psi\nabla^2\phi)\mathrm{d}\Omega = \oint_\Gamma (\phi \operatorname{grad}\psi - \psi \operatorname{grad}\phi) \cdot \vec{n}\mathrm{d}\Gamma \tag{89}$$

with Ω being a 3 dimensional domain with boundary surface Γ. Green's identities are generalizations of the integration by parts rules

$$\int_{x_1}^{x_2} (\phi\,\phi'' + \phi'^2)\mathrm{d}x = [\phi\phi']_{x_1}^{x_2}, \tag{90}$$

$$\int_{x_1}^{x_2} (\phi\,\psi'' - \psi\phi'')\mathrm{d}x = [\phi\psi' - \psi\phi']_{x_1}^{x_2}. \tag{91}$$

to three dimensions. Green's identities play a vital role in numerical field computation as they constitute the junction between the finite element (FEM) and the boundary element (BEM) method as shown in fig. 11.

3.2 Finite element method with total vector-potential formulation

Consider the elementary model problem for a superconducting magnet consisting of two different domains: Ω_i the iron region with permeability μ and Ω_a the air region with the permeability μ_0. The regions are connected to each other at the interface Γ_ai. The non-conductive air region Ω_a may also contain a certain number of conductor sources $\vec{J}$ which do not intersect the iron region Ω_i.

3.21 The weighted residual

Let us start with the total vector potential formulation

$$\operatorname{curl}\frac{1}{\mu}\operatorname{curl}\vec{A} - \operatorname{grad}\frac{1}{\mu}\operatorname{div}\vec{A} = \vec{J} \quad \text{in}\,\Omega. \tag{92}$$

The domain $\Omega = \Omega_\mathrm{a} \cup \Omega_\mathrm{i}$ is discretized into finite-elements in order to solve this problem numerically. For the *approximate* solution of $\vec{A}$ as an interpolation of its values on the nodes of the finite element mesh, the differential equation (92) is only approximately fulfilled:

$$\operatorname{curl}\frac{1}{\mu}\operatorname{curl}\vec{A} - \operatorname{grad}\frac{1}{\mu}\operatorname{div}\vec{A} - \vec{J} = \vec{R} \tag{93}$$

with a residual (error) vector $\vec{R}$. A linear equation system for the unknown nodal values of the vector potential $\vec{A}^{(k)}$ can be obtained by minimizing the weighted residuals $\vec{R}$ in an average sense over the domain Ω, i.e.,

$$\int_\Omega \vec{w}_a \cdot \vec{R}\,\mathrm{d}\Omega = 0, \qquad a = 1,2,3 \tag{94}$$

with the vector weighting functions $\vec{w}_1 = (w_1, 0, 0)^T$, $\vec{w}_2 = (0, w_2, 0)^T$, $\vec{w}_3 = (0, 0, w_3)^T$, where w_1, w_2, w_3 are arbitrary (but known) weighting functions. Forcing the weighted residual to zero yields

$$\int_\Omega \vec{w}_a \cdot \left(\mathrm{curl}\,\frac{1}{\mu}\mathrm{curl}\,\vec{A} - \mathrm{grad}\,\frac{1}{\mu}\mathrm{div}\,\vec{A} \right) \mathrm{d}\Omega = \int_\Omega \vec{w}_a \cdot \vec{J}\,\mathrm{d}\Omega, \qquad a = 1,2,3. \tag{95}$$

3.22 The weak form

With the following generalizations of the integration by parts rule

$$\int_\Omega \left(\mathrm{curl}\,\frac{1}{\mu}\mathrm{curl}\,\vec{A} \right) \cdot \vec{w}_a\,\mathrm{d}\Omega = \int_\Omega \frac{1}{\mu}\mathrm{curl}\,\vec{A} \cdot \mathrm{curl}\,\vec{w}_a \mathrm{d}\Omega - \oint_\Gamma \frac{1}{\mu} \left(\mathrm{curl}\,\vec{A} \times \vec{n} \right) \cdot \vec{w}_a \mathrm{d}\Gamma, \tag{96}$$

$$\int_\Omega \left(-\,\mathrm{grad}\,\frac{1}{\mu}\mathrm{div}\,\vec{A} \right) \cdot \vec{w}_a\,\mathrm{d}\Omega = \int_\Omega \frac{1}{\mu}\mathrm{div}\,\vec{A}\,\mathrm{div}\,\vec{w}_a \mathrm{d}\Omega - \oint_\Gamma \frac{1}{\mu}\mathrm{div}\,\vec{A}(\vec{n} \cdot \vec{w}_a)\mathrm{d}\Gamma, \tag{97}$$

and the consideration of the boundary conditions on Γ, the weighted residual can be transformed to

$$\int_\Omega \frac{1}{\mu}\mathrm{curl}\,\vec{w}_a \cdot \mathrm{curl}\,\vec{A}\,\mathrm{d}\Omega + \int_\Omega \frac{1}{\mu}\mathrm{div}\,\vec{w}_a \mathrm{div}\,\vec{A}\,\mathrm{d}\Omega = \int_\Omega \vec{w}_a \cdot \vec{J}\,\mathrm{d}\Omega \tag{98}$$

with a = 1,2,3. In two dimensions, where $\frac{\partial}{\partial z} = 0$ and the Coulomb gauge is automatically fulfilled, Eq. (98) further reduces to

$$\int_\Omega \frac{1}{\mu}\mathrm{curl}\,\vec{w}_3 \cdot \mathrm{curl}\,\vec{A}_z\,\mathrm{d}\Omega = \int_\Omega \vec{w}_3 \cdot \vec{J}_z\,\mathrm{d}\Omega. \tag{99}$$

Eq. (98) is called the weak form of the vector-potential formulation because the second derivatives have been removed and the continuity requirements on $\vec{A}$ have been relaxed at the expense of an increase in the continuity conditions of the weighting functions. Only this makes possible the use of elements with linear shape functions. Inside these elements the first derivative of the shape functions is a constant and the second derivative vanishes. On the element boundary we find a jump in the first derivative and a Dirac δ-function for the second. Thus there would be a problem in Eq. (98) if the second derivative was present. This level of continuity is termed as C_0 continuous.

Remark: The current density $\vec{J}$ appears on the right hand side of eg. (98). In consequence, when using the FE-method for the solution of this problem the relatively complicated shape of the coils must be modeled in the FE-mesh.

3.23 Finite element shape functions

The domain $\Omega = \Omega_\mathrm{a} \cup \Omega_\mathrm{i}$ is discretized into finite-elements Ω_j, it is "meshed" in common parlance. The method is explained by means of the most simple (2D, linear, triangular) finite elements. Within a particular element j the z component of the vector potential $A_j = A_{z_j}(x, y)$ is approximated by

$$A_j = \alpha_1 + \alpha_2 x + \alpha_3 y. \tag{100}$$

The above functions (with finite support) are called *basis* functions and are defined on the element j only. The approximated potentials $A^{(k)}$ at the three local nodes $k = 1, 2, 3$ of the element j can then

be expressed as $A^{(k)} = \alpha_1 + \alpha_2 x_k + \alpha_3 y_k$ and so forth. This set of 3 equations can be solved for the unknowns $\alpha_1, \alpha_2, \alpha_3$ using Cramer's rule for instance. For a general point $\vec{x}$ within the element j the potential can then be approximated with

$$A_j(\vec{x}) \quad = \quad \alpha_1 + \alpha_2 x + \alpha_3 y = \sum_{k=1}^{3} N_k(\vec{x}) A^{(k)} \tag{101}$$

with the so-called element shape functions (or trial functions) N_k

$$N_k(\vec{x}) = N_k(x, y) == \frac{a_k + b_k x + c_k y}{2S}, \tag{102}$$

where S is the surface of the element and the a_k, b_k, c_k are nothing else than linear combinations of the coordinates of the nodal points. The N_k are the barycentric coordinates of a point $\vec{x}$ with respect to the node k. As we shall see, in the FEM formulation only the element shape functions appear but not the basis functions (100). Obviously, the element shape function has to equal one when the point $\vec{x}$ is joined with the node k; it equals zero when the point is outside the element. Figure 12 shows the element-wise defined (linear) basis functions $A_j = \alpha_1 + \alpha_2 x + \alpha_3 y$ and the shape function N_k in the local node $k = 3$ of element $j = 6$.

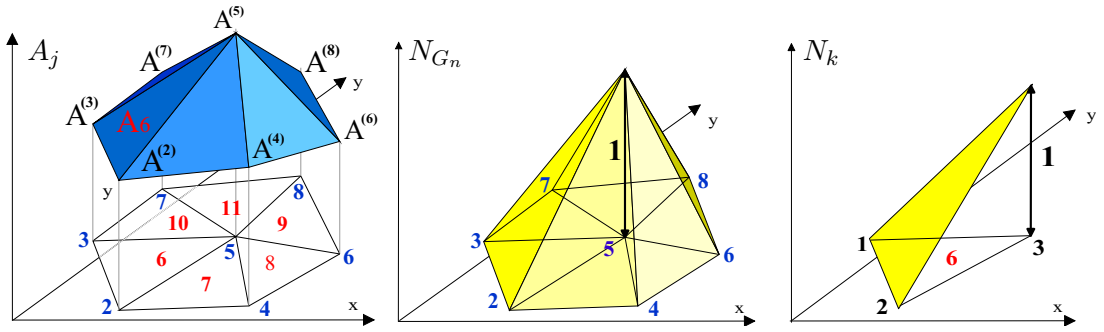


Fig. 12: Left: Element-wise defi ned (linear) basis functions $A_j = \alpha_1 + \alpha_2 x + \alpha_3 y$. Middle: Complete (linear) nodal function N_{G_n} of global node $n = 5$. Right: Shape function (also called trial or interpolation function) N_k in local node $k = 3$ in element $j = 6$.

3.24 Assembling the matrix

So far the basis functions were defined to be scalars. In 3 dimensions the solution is approximated **element-wise** by the potential vectors $\vec{A}_j$ (containing the three scalar components of the vector potential) which are expanded with respect to the element shape functions (or trial functions) $N_k(\vec{x})$ and their nodal values $\vec{A}^{(k)}$ which are to be determined in the solution:

$$\vec{A}_j = \sum_{k=1}^{K} N_k(\vec{x}) \vec{A}^{(k)} \quad \text{in } \Omega_j, \tag{103}$$

where K is the number of nodes of the element Ω_j and $\vec{A} = (A_x, A_y, A_z)^T$. A linear equation system for the unknowns $\vec{A}^{(k)}$ can be obtained by forcing the weighted residual to zero in an average sense over the element Ω_j and integration by parts, which yields the weak integral form for **one** element. The weighting

functions are chosen as the element shape functions $N_l, l = 1, K$ (Galerkin method) and we get:

$$\int_{\Omega_j} \frac{1}{\mu} \mathrm{curl}\, \vec{N}_{la} \cdot \mathrm{curl} \left(\sum_{k=1}^{K} N_k(\vec{x}) \vec{A}^{(k)} \right) \mathrm{d}\Omega + \int_{\Omega_j} \frac{1}{\mu} \mathrm{div}\, \vec{N}_{la} \mathrm{div} \left(\sum_{k=1}^{K} N_k(\vec{x}) \vec{A}^{(k)} \right) \mathrm{d}\Omega = \int_{\Omega_j} \vec{N}_{la} \cdot \vec{J}\, \mathrm{d}\Omega \tag{104}$$

for $l = 1, K$ and $a = 1, 2, 3$, with $\vec{N}_{k1,l1} = (N_{k,l}, 0, 0)^T$, $\vec{N}_{k2,l2} = (0, N_{k,l}, 0)^T$ and $\vec{N}_{k3,l3} = (0, 0, N_{k,l})^T$. The equation system (104) can be written in a matrix representation as

$$[k]\{A\} = \{f\}. \tag{105}$$

The matrix $[k]$ is often called the *stiffness* matrix (by reference to elasticity problems), and $\{f\}$ is the element *force* vector. $\{A\}$ is the vector of the nodal potential function. The stiffness matrix has the form

$$[k] = \begin{bmatrix} k_{11} & \cdots & k_{1K} \\ \vdots & \ddots & \vdots \\ k_{K1} & \cdots & k_{KK} \end{bmatrix} \tag{106}$$

for a element with K nodes (three degrees of freedom per node). The coefficients $k_{lm}(l, m = 1, ..., K)$ in this matrix are 3×3 matrixes and can be calculated by means of Gauss-integration. The Galerkin method has the important property that the stiffness matrix is sparse and symmetric. The equation system for the entire domain Ω can be assembled through the merging of the element (sub)-systems in the resulting equation system

$$[K]\{A\} = \{F\} \tag{107}$$

with the introduction of the so-called nodal functions $N_{G_n}(\vec{x})$ with $N_{G_n}(\vec{x}) = N_k(\vec{x})$ for $\vec{x} \in \Omega_j$, where n is the global node number in the domain Ω.

3.3 Numerical field computation for superconducting magnets

3.31 Quadrilateral higher order elements

The so-called simplex elements (triangular in 2D, tetrahedral in 3D) have the disadvantage that curved shapes can only be modeled by polygonal approximations. The advantage of a higher order approximation of the potentials may be lost due to a rather rough geometric approximation. An alternative are the *isoparametric* elements which have curved sides. They can therefore be found in most commercial software packages. It avoids numerically unfavorable prisms when the geometry is simply extruded into the third dimension.

A mesh generator based on geometrical domain decomposition, which was developed at the University of Stuttgart, Germany [30] has been implemented in the ROXIE program package. The following extensions have been added.

1) Extension of the method to 8 noded (higher order) quadrilateral elements. 2) Parametric input for the definition of design variables for mathematical optimization. 3) Implementation of design features for the definition of material boundaries. 4) Modular magnet geometry input by means of the GNU m4 macro language. 5) A morphing algorithm for optimization and sensitivity studies which avoids remeshing and changing mesh topologies.

The quadrilateral mesh generator relies on the method of geometrical domain decomposition [30]. In a first step, the input geometry is decomposed into areas that are topologically equivalent to disks, i.e.,

holes are eliminated. The shape of the sub-domains is optimized by minimizing a function that depends on the ratios l^2/A, l/U, and the angles of the newly created areas; l is the length of the cut, A is the area, and U is the circumference of the sub-domain. This decomposition is continued until the remaining areas are regarded as simple (when they are similar to triangles or rectangles). These areas are then filled with quadrilateral elements using a modified paving strategy [9]. In this approach an area is filled from the outside to the inside by adding full rows of quadrilateral elements. Finally, a smoothing algorithm is applied. A linear combination of Laplace-smoothing, edge-smoothing and angle-smoothing is used to enlarge small angles, to reduce large angles, and to increase short distances between mesh points while leaving the topology unchanged. As an example, figure 13 shows the meshed iron yoke, insert, and collar geometry of the LHC main dipole.

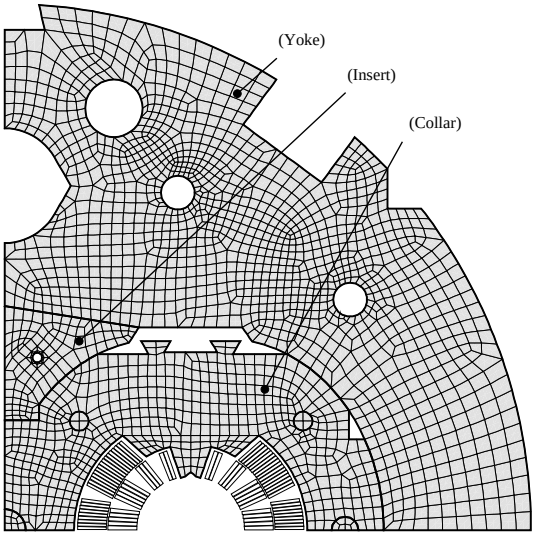


Fig. 13: Quadrilateral higher order fi nite element mesh of the LHC main dipole iron yoke, insert, and stainless steel collar.

3.32 Coupling Method between Boundary Elements (BEM) and Finite Elements (FEM)

The disadvantage of the finite element method is that only a finite domain can be discretized, and therefore the field calculation in the magnet coil-ends with their large fringe-fields requires a large number of elements in the air region. The relatively new boundary-element method is defined on an infinite domain and can therefore solve open boundary problems without approximation with far-field boundaries. The disadvantage is that non-homogeneous materials are difficult to consider. The BEM-FEM method couples the finite element method inside magnetic bodies $\Omega_{\mathrm{i}} = \Omega_{\mathrm{FEM}}$ with the boundary-element method in the domain outside the magnetic material $\Omega_{\mathrm{a}} = \Omega_{\mathrm{BEM}}$, by means of the normal derivative of the vector-potential on the interface Γ_{ai} between iron and air. The application of the BEM-FEM method to magnet design has the following intrinsic advantages:

- The coil field can be taken into account in terms of its source vector potential $\vec{A}_{\mathrm{s}}$, which can be obtained easily from the filamentary currents I_{s} by means of Biot-Savart type integrals without the meshing of the coil.
- The BEM-FEM coupling method allows for the direct computation of the reduced vector potential $\vec{A}_{\mathrm{r}}$ instead of the total vector potential $\vec{A}$. Consequently, errors do not influence the dominating contribution $\vec{A}_{\mathrm{s}}$ due to the superconducting coil.
- Because the field in the aperture is calculated through the integration over all the BEM elements, local field errors in the iron yoke cancel out and the calculated multipole content is sufficiently accurate even for very sparse meshes.

- The surrounding air region need not be meshed at all. This simplifies the preprocessing and avoids artificial boundary conditions at some far-field-boundaries. Moreover, the geometry of the permeable parts can be modified without regard to the mesh in the surrounding air region, which strongly supports the feature based, parametric geometry modeling that is required for mathematical optimization.
- The method can be applied to both 2D and 3D field problems.

3.33 The FEM part

Inside the magnetic domain Ω_i a gauged vector-potential formulation is applied:

$$-\frac{1}{\mu_0}\nabla^2\vec{A} \;=\; \vec{J} + \operatorname{curl}\vec{M} \quad \text{in}\,\Omega_i. \tag{108}$$

Considering that the free current density in the iron yoke is zero, and forcing the weighted residual to zero yields

$$-\int_{\Omega_\mathrm{i}} \frac{1}{\mu_0}\nabla^2\vec{A}\cdot\vec{w}_a \mathrm{d}\Omega_\mathrm{i} = \int_{\Omega_\mathrm{i}} (\operatorname{curl}\vec{M})\cdot\vec{w}_a \mathrm{d}\Omega_\mathrm{i} \quad a = 1,2,3. \tag{109}$$

with $\vec{w}_{1,2,3}$. With Green's first theorem

$$\int_{\Omega_\mathrm{i}} \nabla^2\vec{A}\cdot\vec{w}_a \mathrm{d}\Omega_\mathrm{i} = -\int_{\Omega_\mathrm{i}} \operatorname{grad}(\vec{A}\cdot\vec{e}_a)\cdot\operatorname{grad} w_a \,\mathrm{d}\Omega_\mathrm{i} + \oint_\Gamma \frac{\partial\vec{A}}{\partial n_\mathrm{i}}\cdot\vec{w}_a \,\mathrm{d}\Gamma \tag{110}$$

and the identity

$$\int_{\Omega_\mathrm{i}} \operatorname{curl}\vec{M}\cdot\vec{w}_a \mathrm{d}\Omega_\mathrm{i} = \int_{\Omega_\mathrm{i}} \vec{M}\cdot\operatorname{curl}\vec{w}_a \mathrm{d}\Omega_\mathrm{i} - \oint_\Gamma (\vec{M}\times\vec{n}_i)\cdot\vec{w}_a \,\mathrm{d}\Gamma \tag{111}$$

we get for the weak form

$$\frac{1}{\mu_0}\int_{\Omega_\mathrm{i}} \operatorname{grad}(\vec{A}\cdot\vec{e}_a)\cdot\operatorname{grad} w_a \,\mathrm{d}\Omega_\mathrm{i} - \frac{1}{\mu_0}\oint_{\Gamma_\mathrm{ai}} \left(\frac{\partial\vec{A}}{\partial n_\mathrm{i}} - (\mu_0\vec{M}\times\vec{n}_\mathrm{i})\right)\cdot\vec{w}_a \,\mathrm{d}\Gamma_\mathrm{ai} = \int_{\Omega_\mathrm{i}} \vec{M}\cdot\operatorname{curl}\vec{w}_a \,\mathrm{d}\Omega_\mathrm{i} \tag{112}$$

with a = 1,2,3. The continuity condition of the tangential component of the magnetic field at the interface between iron (FEM domain) and air is equivalent to

$$\frac{\partial\vec{A}_\mathrm{i}}{\partial n_\mathrm{i}} - (\mu_0\vec{M}\times\vec{n}_\mathrm{i}) + \frac{\partial\vec{A}_\mathrm{a}}{\partial n_\mathrm{a}} = 0, \tag{113}$$

where $\vec{n}_\mathrm{i}$ is the normal vector on Γ_ai pointing out of the FEM domain Ω_i, and $\vec{n}_\mathrm{a}$ is the normal vector on Γ_ai pointing out of the BEM domain Ω_a. The boundary integral term on the boundary between iron and air Γ_ai in (112) serves as the coupling term between the BEM and the FEM domain. Let us for the moment *assume* that the normal derivative on Γ_ai

$$\vec{Q}_{\Gamma_\mathrm{ai}} = -\frac{\partial\vec{A}^\mathrm{BEM}_{\Gamma_\mathrm{ai}}}{\partial n_\mathrm{a}} \tag{114}$$

is given. If the domain Ω_i is discretized into finite elements Ω_j (C^0-continuous, isoparametric 20-noded hexahedron elements are used) and the Galerkin method is applied to the weak formulation, then a nonlinear system of equations is obtained

$$\begin{pmatrix} [K_{\Omega_\mathrm{i}\Omega_\mathrm{i}}] & [K_{\Omega_\mathrm{i}\Gamma_\mathrm{ai}}] & 0 \\ [K_{\Gamma_\mathrm{ai}\Omega_\mathrm{i}}] & [K_{\Gamma_\mathrm{ai}\Gamma_\mathrm{ai}}] & [T] \end{pmatrix} \begin{pmatrix} \{\vec{A}_{\Omega_\mathrm{i}}\} \\ \{\vec{A}_{\Gamma_\mathrm{ai}}\} \\ \{\vec{Q}_{\Gamma_\mathrm{ai}}\} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{115}$$

with all nodal values of $\vec{A}_{\Omega_\mathrm{i}}$, $\vec{A}_{\Gamma_\mathrm{ai}}$ and $\vec{Q}_{\Gamma_\mathrm{ai}}$ grouped in arrays

$$\{\vec{A}_{\Omega_\mathrm{i}}\} = (\vec{A}^{(1)}_{\Omega_\mathrm{i}}, \vec{A}^{(2)}_{\Omega_\mathrm{i}}, \dots), \quad \{\vec{A}_{\Gamma_\mathrm{ai}}\} = (\vec{A}^{(1)}_{\Gamma_\mathrm{ai}}, \vec{A}^{(2)}_{\Gamma_\mathrm{ai}}, \dots), \quad \{\vec{Q}_{\Gamma_\mathrm{ai}}\} = (\vec{Q}^{(1)}_{\Gamma_\mathrm{ai}}, \vec{Q}^{(2)}_{\Gamma_\mathrm{ai}} \dots). \quad (116)$$

The subscripts Γ_ai and Ω_i refer to nodes on the boundary and in the interior of the domain, respectively. The domain and boundary integrals in the weak formulation yield the stiffness matrices $\big[K\big]$ and the boundary matrix $\big[T\big]$. The stiffness matrices depend on the local permeability distribution in the nonlinear material. All the matrices in (115) are sparse.

3.34 The BEM part

By definition, the BEM domain Ω_a contains no iron, and therefore $\vec{M}$ = 0 and $\mu = \mu_0$. The governing equation

$$\nabla^2 \vec{A} = -\mu_0 \vec{J} \quad (117)$$

decomposes in Cartesian coordinates into three scalar Poisson equations to be solved. For an approximate solution of these equations and the weighted residual forced to zero yields:

$$\int_{\Omega_\mathrm{a}} \nabla^2 A w \mathrm{d}\Omega_\mathrm{a} = -\int_{\Omega_\mathrm{a}} \mu_0 J w \mathrm{d}\Omega_\mathrm{a}. \quad (118)$$

Employing Green's second theorem yields

$$\int_{\Omega_\mathrm{a}} A \nabla^2 w \mathrm{d}\Omega_\mathrm{a} = -\int_{\Omega_\mathrm{a}} \mu_0 J w \, \mathrm{d}\Omega_\mathrm{a} + \int_{\Gamma_\mathrm{ai}} A \frac{\partial w}{\partial n_a} \mathrm{d}\Gamma_\mathrm{ai} - \int_{\Gamma_\mathrm{ai}} \frac{\partial A}{\partial n_a} w \mathrm{d}\Gamma_\mathrm{ai} \, . \quad (119)$$

In Eq. (119) it is already considered that all the boundary integrals on the far field boundary $\Gamma_{\mathrm{BEM}\infty}$ vanish. Now the weighting function is chosen as the fundamental solution of the Laplace equation, which is in 3D

$$w = u^* = \frac{1}{4\pi R}, \qquad \frac{\partial w}{\partial n_a} = q^* = -\frac{1}{4\pi R^2}, \qquad \nabla^2 w = -\delta(R) \quad (120)$$

we get the Fredholm integral equation of the second kind:

$$\frac{\Theta}{4\pi} A + \int_{\Gamma_\mathrm{ai}} Q_{\Gamma_\mathrm{ai}} \, u^* \, \mathrm{d}\Gamma_\mathrm{ai} + \int_{\Gamma_\mathrm{ai}} A_{\Gamma_\mathrm{ai}} \, q^* \, \mathrm{d}\Gamma_\mathrm{ai} = \int_{\Omega_\mathrm{a}} \mu_0 J u^* \, \mathrm{d}\Omega_\mathrm{a}. \quad (121)$$

The right hand side of Eq. (121) (the last remaining domain integral) is a Biot-Savart-type integral for the source vector potential A_s.

The components of the vector potential $\vec{A}$ at arbitrary points $\vec{r}_0 \in \Omega_\mathrm{a}$ (e.g. on the reference radius for the field harmonics) have to be computed from (121) as soon as the components of the vector potential $\vec{A}_{\Gamma_\mathrm{ai}}$ and their normal derivatives $\vec{Q}_{\Gamma_\mathrm{ai}}$ on the boundary Γ_ai are known. Θ is the solid angle enclosed by the domain Ω_a in the vicinity of $\vec{r}_0$.

For the discretization of the boundary Γ_ai into individual boundary elements $\Gamma_{\mathrm{ai},j}$, again C^0-continuous, isoparametric 8-noded quadrilateral boundary elements (in 3D) are used. In 2D 3-noded line elements are used. They have to be consistent with the elements from the FEM domain touching this boundary. The components of $\vec{A}_{\Gamma_\mathrm{ai}}$ and $\vec{Q}_{\Gamma_\mathrm{ai}}$ are expanded with respect to the element shape functions, and (121) can be rewritten in terms of the nodal data of the discrete model,

$$\frac{\Theta}{4\pi} \vec{A} = \vec{A}_\mathrm{s} - \{\vec{Q}_{\Gamma_\mathrm{ai}}\} \cdot \{g\} - \{\vec{A}_{\Gamma_\mathrm{ai}}\} \cdot \{h\}. \quad (122)$$

In (122), g results from the boundary integral with the kernel u^*, and h results from the boundary integral with the kernel q^*. The discrete analogue of the Fredholm integral equation can be obtained from (122) by successively putting the evaluation point $\vec{r}_0$ at the location of each nodal point $\vec{r}_j$. This procedure is called point-wise collocation and yields a linear system of equations,

$$\left[G\right]\left\{\vec{Q}_{\Gamma_{\mathrm{ai}}}\right\} + \left[H\right]\left\{\vec{A}_{\Gamma_{\mathrm{ai}}}\right\} = \left\{\vec{A}_{\mathrm{s}}\right\}. \tag{123}$$

In (123), $\left\{\vec{A}_{\mathrm{s}}\right\}$ contains the values of the source vector potential at the nodal points $\vec{r}_j, j = 1, 2, \ldots$. The matrices $\left[G\right]$ and $\left[H\right]$ are asymmetric and fully populated. Equation (123) gives exactly the missing relationship between the Dirichlet data $\left\{\vec{A}_{\Gamma_{\mathrm{ai}}}\right\}$ and the Neumann data $\left\{\vec{Q}_{\Gamma_{\mathrm{ai}}}\right\}$ on the boundary Γ_{ai}.

An overall numerical description of the field problem can be obtained by complementing the FEM description (115) with the BEM description (123) which results in

$$\begin{pmatrix} \left[K_{\Omega_{\mathrm{i}}\Omega_{\mathrm{i}}}\right] & \left[K_{\Omega_{\mathrm{i}}\Gamma_{\mathrm{ai}}}\right] & 0 \\ \left[K_{\Gamma_{\mathrm{ai}}\Omega_{\mathrm{i}}}\right] & \left[K_{\Gamma_{\mathrm{ai}}\Gamma_{\mathrm{ai}}}\right] & \left[T\right] \\ 0 & \left[H\right] & \left[G\right] \end{pmatrix} \begin{pmatrix} \left\{\vec{A}_{\Omega_{\mathrm{i}}}\right\} \\ \left\{\vec{A}_{\Gamma_{\mathrm{ai}}}\right\} \\ \left\{\vec{Q}_{\Gamma_{\mathrm{ai}}}\right\} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \left\{\vec{A}_{\mathrm{s}}\right\} \end{pmatrix}. \tag{124}$$

It can be shown [21] that this procedure yields the correct physical interface conditions, the continuity of $\vec{n} \cdot \vec{B}$ and $\vec{n} \times \vec{H}$ across Γ_{ai}.

3.4 Numerical examples

3.41 Field quality in collared coils

Using double aperture stainless-steel collars with a relative permeability of $\mu_{\mathrm{r}} = 1.0025$ creates asymmetries in the magnetic field in the case of warm measurements of the collared coil assembly where only one aperture is powered. As the BEM-FEM method does not require the meshing of the coil (which can therefore be modeled with the required accuracy) and does not require a "far field" boundary condition which would influence considerably the results of this unbounded field problem the BEM-FEM method is specially qualified for the calculations.

The additional field errors generated by the asymmetric stainless steel collar are a Δb_2 of -0.239 units in 10^-4 (at 17 mm reference radius), Δb_3=-1.173, Δb_4=-0.012 and Δb_5=0.305.

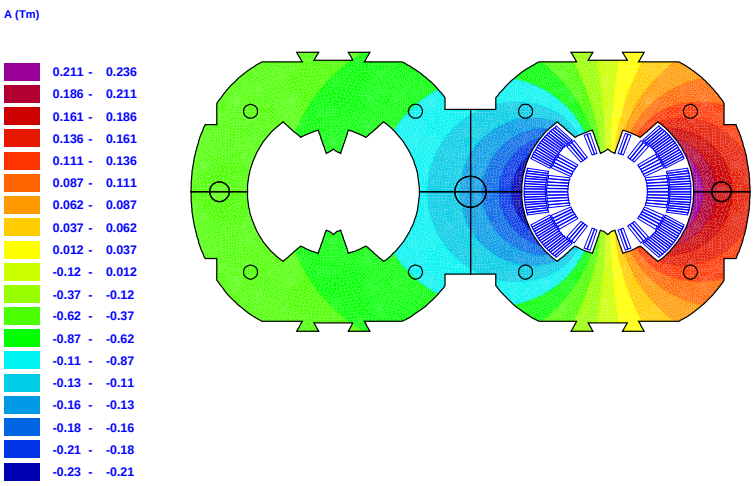


Fig. 14: Geometric model of one dipole coil powered for warm measurement in the combined collar structure assuming a constant relative permeability of $\mu_{\mathrm{r}} = 1.0025$. The fi gure displays the magnetic flux density in the collars.

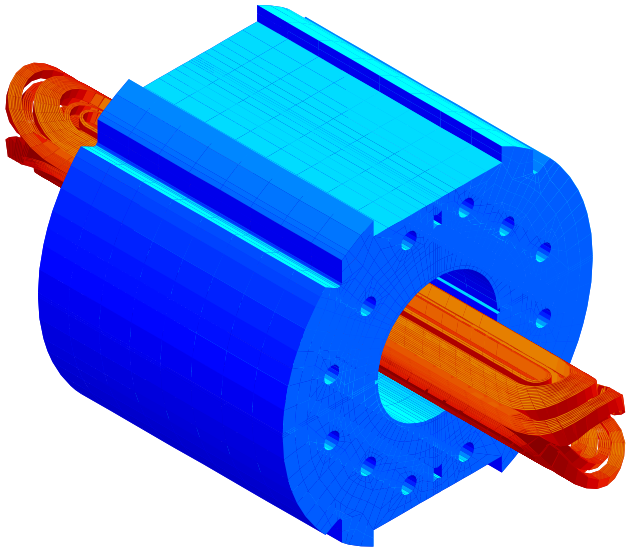


Fig. 15: ROXIE model of the 1m long coil-test facility (CTF).

3.42 End fields in the dipole short models

As was stated in the introduction, the LHC requires a major effort to guarantee that the superconducting dipoles will perform to specifications. For this reason an R&D program for the development of the superconducting dipoles was started in 1995. The aim of the program is to study the influence of individual coil parameters, pre-stress in the coil, collar material and yoke structure on a series of otherwise identical model dipoles. A maximum of turn-around and testing efficiency is achieved by reducing the length of the models from 14 m to approximately 1 m and by manufacturing single aperture models with only one coil in a re-usable iron-yoke. The short models have a coil length of 1.05 m and a magnetic yoke of only 402 mm approximately centered in the magnet in order to leave the configuration in the end and connection region unchanged with respect to the long prototypes.

The magnetic field homogeneity of the models is systematically measured in a vertical test set-up where the magnet is suspended inside a cryostat. A drawback of the short length of the dipole models is that end effects influence the magnetic field quality in the center of the magnet. As the rotating pick-up coils deliver multipoles averaged over their length, and the pick-up in the center of the magnet has a length of 200 mm (half the length of the magnetic yoke) the interpretation of measurements becomes difficult. In order to study systematic effects in the field quality, it is necessary to calculate, with a high precision, the 3D multipole field errors in these magnets as a function of the z-position. Fig. 15 shows the geometric model of the coil-test facility (CTF). To prove the reliability of the method, the results of the 3D field calculation are compared to the measurements on the single aperture model MBSMS21. The multipoles are computed as a function of the excitation current over a length of 550 mm along the magnet bore. The computed dipole field $B_1(z)$ and the multipoles $b_n(z)$ are used to compute the average multipoles over the length spanned by the measurement pick-up coils. This averaging step is necessary to obtain quantities comparable to the measured values. The results of the measured and simulated normal dipole and normal sextupole component are shown in fig. 16 together with the 2D calculations for the long magnet. It can be seen that the short models show a globally different saturation behavior compared to the long dipole prototypes.

3.43 End field in long dipole prototype magnets

In order to reduce the peak field in the coil-end and thus increase the quench margin in the region with a weaker mechanical structure, the magnetic iron yoke in LHC prototype magnets was replaced by non-

magnetic stainless steel laminations approximately 100 mm from the onset of the ends. The BEM-FEM coupling method was therefore used for the calculation of the end-fields. The computing time for the 3D calculation is in the order of 5 hours on a DEC Alpha 5/333 workstation. The iterative solution of the linear equation system converges better in the case of a high excitational field than in the case of the injection field with its non-saturated iron yoke. It is therefore still impossible to apply mathematical optimization techniques to the 3D field calculation with iron yoke. However, as the additional effect from the fringe field on the field quality is low, it is sufficient to calculate the additional effect and then partially compensate with the coil design, if necessary. It has already been explained that the BEM-FEM coupling method allows the distinction between the coil field and the reduced field from the iron magnetization. Fig. 17 shows the field components along a line in the end-region of the twin-aperture dipole prototype magnet (MBP2), 43.6 mm above the beam-axis in aperture 2 (on a radius between the inner and outer layer coil) from z = -200 mm inside the magnet yoke to z = 200 mm outside the yoke. The iron yoke ends at z = -80 mm, the onset of the coil-end is at z = 0.

4. Integral Quantities of the Field Solutions

The calculation of stored energy and inductance is vital for the calculation of the quench behavior of superconducting magnets. The integration of the magnetic energy density in the entire domain is not practical as with the BEM-FEM coupling method the air region is not meshed at all. We will therefore derive more practical equations both for the linear and the nonlinear case. We will show that for the calculation of the voltage drop across the terminal during a quench the differential inductance has to be considered if the iron yoke is saturated at high field. This is not important for the LHC main dipoles but can become dominant for insertion quadrupoles.

4.1 Linear circuits

4.11 Stored energy

In the linear case, the stored magnetic energy in a volume V is given by the integral

$$W = \frac{1}{2}\int_V \vec{H}\cdot\vec{B}\,\mathrm{d}V\,. \tag{125}$$

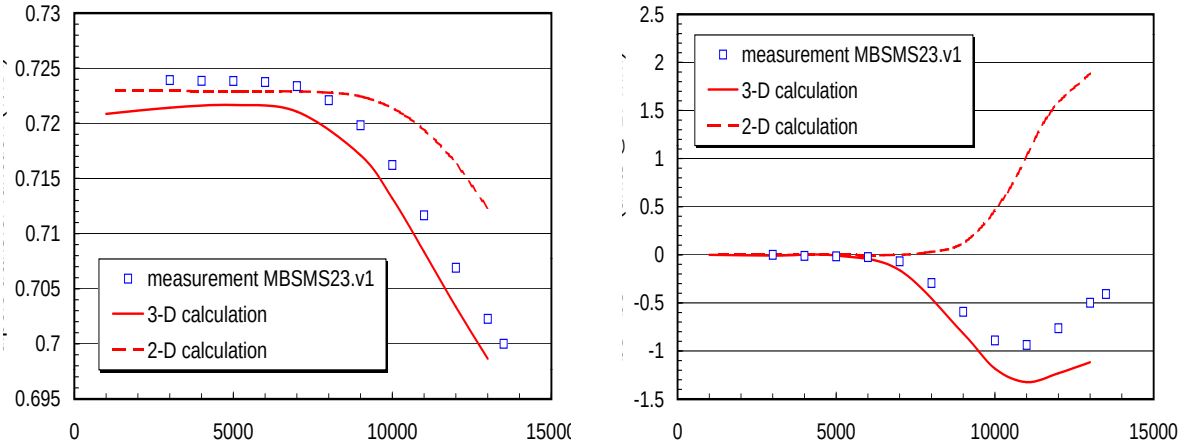


Fig. 16: Measured and computed field components B_1 (left) and b_3 (right) as a function of the excitation (between injection and nominal field) averaged over the length of the measurement pick-up coil (200 mm).

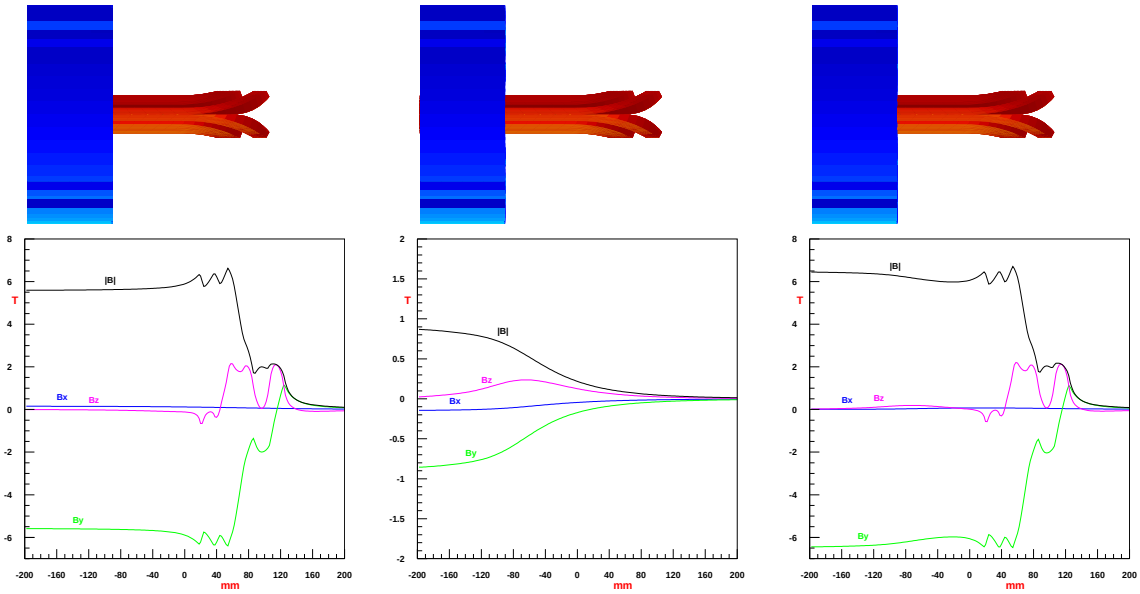


Fig. 17: Magnetic flux density at nominal current along a line at x = 97mm, y = 43.6 mm (above the beam-axis of aperture 2 on a radius between the inner and the outer layer coil) from z = -200 mm inside the magnet yoke to z = 200 mm outside the yoke. The iron yoke ends at z = -80 mm, the onset of the coil-end is at z = 0. Left: coil field, Middle: reduced field from iron magnetization, Right: total field. Note the different scales and the relatively small contribution from the yoke.

Because of $\mathrm{div}\,(\vec{A}\times\vec{H}) = \vec{H}\cdot\mathrm{curl}\,\vec{A} - \vec{A}\cdot(\mathrm{curl}\,\vec{H})$, Eq. (125) can be rewritten as:

$$\begin{aligned} W &= \frac{1}{2}\int_V \vec{H}\cdot\mathrm{curl}\,\vec{A}\,\mathrm{d}V = \frac{1}{2}\int_V \mathrm{div}\,(\vec{A}\times\vec{H})\,\mathrm{d}V + \frac{1}{2}\int_V \vec{A}\cdot\mathrm{curl}\,\vec{H}\mathrm{d}V \\ &= \frac{1}{2}\oint(\vec{A}\times\vec{H})\cdot\mathrm{d}\vec{a} + \frac{1}{2}\int_V \vec{A}\cdot\mathrm{curl}\,\vec{H}\mathrm{d}V\,. \end{aligned} \tag{126}$$

The term $\frac{1}{2}\oint(\vec{A}\times\vec{H})\cdot\mathrm{d}\vec{a}$ vanishes on the far-field boundary as $A\propto 1/r, H\propto 1/r^2, \mathrm{d}a\propto r^2$.

Inner surfaces (which have to be considered from both sides) are not contributing to the total energy as long as they carry no surface currents ($\vec{J}_a = 0$, which is true in case of finite conductivity and excitation without jump discontinuity). Proof:

$$\begin{aligned} W &= \frac{1}{2}\oint(\vec{A}\times\vec{H})\cdot\mathrm{d}\vec{a} = \frac{1}{2}\sum_i\int_{a_i}[\vec{A}\times(\vec{H}_2-\vec{H}_1)]\cdot\vec{n}_2\,\mathrm{d}a_i \\ &= \frac{1}{2}\sum_i\int_{a_i}\vec{A}\cdot[(\vec{H}_2-\vec{H}_1)\times\vec{n}_2]\,\mathrm{d}a_i = \frac{1}{2}\sum_i\int_{a_i}\vec{A}\cdot\vec{J}_a\,\mathrm{d}a_i \end{aligned} \tag{127}$$

which is zero for $\vec{J}_a = 0$. The magnetic energy can then be calculated with

$$W = \frac{1}{2}\int_V \vec{A}\cdot\mathrm{curl}\,\vec{H}\,\mathrm{d}V = \frac{1}{2}\int_V \vec{A}\cdot\vec{J}\,\mathrm{d}V. \tag{128}$$

For the 2-dimensional calculation of coil cross-sections made of k multi-filamentary strands, the stored energy per unit length can be calculated as

$$\boxed{W/l = \frac{1}{2}\sum_k A_{z_k} I_k.} \tag{129}$$

In this expression A_{z_k} refers to the vector potential due to the currents (other than the I_k in the strand) which produce the field $\vec{B}_k$ in that strand and therefore neglects the magnetic energy in the strand.

4.12 The energy inside a strand

From Ampère's law $\oint \vec{H}\cdot \mathrm{d}\vec{s} = \int \vec{J}\cdot \mathrm{d}\vec{a}$ we get for a round wire with radius r_0 and current I the equation $H = \frac{Ir}{2\pi r_0^2}$. For magnetically neutral materials of constant permeability, the energy in a cylinder of radius r, thickness $\mathrm{d}r$ and length l is $\mathrm{d}W = \frac{1}{2}BH\,2\pi rl\mathrm{d}r = \frac{1}{2}\mu_0 H^2\,2\pi rl\mathrm{d}r = \frac{\mu_0 lI^2}{4\pi r_0^4}r^3\mathrm{d}r$. Therefore the total energy in the strand is

$$\frac{W}{l} = \frac{\mu_0 I^2}{4\pi r_0^4}\int_0^{r_0} r^3\mathrm{d}r = \frac{\mu_0 I^2}{16\pi} \tag{130}$$

which is independent of the strand radius. For one aperture of the LHC main dipole the stored energy at 8.33 T (linear calculation with equivalent inner radius of the iron yoke of 98 mm) is 237 KJ/m. The inner layer consists of 15 turns (have to be considered $\times$ 4 for one aperture), with each turn containing 2$\times$18 strands with a current of 320 A each. The outer layer consists of 25 turns, with each turn containing 2$\times$14 strands with a current of 411.5 A each. The energy stored in the strands can then be calculated to 4.3 J/m and can, indeed, be neglected.

4.13 Self and mutual inductance

With Eq. (128) it yields

$$W = \frac{\mu_0}{8\pi}\int_V\int_{V'}\frac{\vec{J}(\vec{r})\cdot\vec{J}(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\mathrm{d}V'\,\mathrm{d}V. \tag{131}$$

For a set of n closed current loops with current densities $\vec{J}_i(\vec{r}), (i = 1, 2, ..., n)$, $\vec{J}(\vec{r}) = \sum_{i=1}^{n}\vec{J}_i(\vec{r})$ we get

$$\begin{aligned} W = \sum_{i=1}^{n}\sum_{j=1}^{n}W_{ij} &= \frac{\mu_0}{8\pi}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_V\int_{V'}\frac{\vec{J}_i(\vec{r})\cdot\vec{J}_j(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\mathrm{d}V'\,\mathrm{d}V \\ &= \frac{\mu_0}{8\pi}\sum_{i=1}^{n}\sum_{j=1}^{n}I_iI_j\int_V\int_{V'}\frac{\vec{J}_i(\vec{r})\cdot\vec{J}_j(\vec{r}\,')}{I_iI_j|\vec{r}-\vec{r}\,'|}\mathrm{d}V'\,\mathrm{d}V. \end{aligned} \tag{132}$$

With the mutual inductances defined as

$$L_{ij} = \frac{\mu_0}{4\pi I_iI_j}\int_V\int_{V'}\frac{\vec{J}_i(\vec{r})\cdot\vec{J}_j(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|}\mathrm{d}V'\,\mathrm{d}V\,, \tag{133}$$

Eq. (132) can be rewritten as

$$\boxed{W = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}L_{ij}I_iI_j.} \tag{134}$$

We see from Eq. (133) that the inductance only depends on the coil geometry. It also shows the symmetry $L_{ij} = L_{ji}$. For $i = j$ the coefficients (133) are called self-inductance. The inductances can be calculated directly using Eq. (133) or by calculating the stored energy and comparing with (134).

The self and mutual inductances can be derived by powering one single coil at a time (e.g., coil i, containing k individual wires and total current I) and calculating the stored magnetic energy according to Eq. (129). This gives $L_{ii} = \frac{2W_{\mathrm{single}}}{I^2}$, ($W_{\mathrm{single}} = W_{ii}$). Subsequently, powering any two coils i and j with the same current I yields (because of the symmetry of the mutual inductances) $L_{ij} = \frac{1}{2}\left(\frac{2W_{\mathrm{double}}}{I^2} - L_{ii} - L_{jj}\right)$ with ($W_{\mathrm{double}} = W_{ij} + W_{ji}$).

4.14 The magnetic flux

In section 2.81 we showed that outside any current-carrying conductor the field can be represented through a magnetic scalar potential $\vec{H} = -\operatorname{grad}\Phi$. If we now assume one single closed loop of a wire with negligible cross-section (i.e., $\mathrm{d}V$ does not contain the conductor and hence the inner energy of it), the stored magnetic energy can be calculated by integrating

$$W = \frac{\mu_0}{2}\int_V \vec{H}^2 \mathrm{d}V = \frac{\mu_0}{2}\int_V (\operatorname{grad}\Phi)^2 \mathrm{d}V, \tag{135}$$

where V is the (current free) external volume. Using the identity $\operatorname{div}(\Phi \operatorname{grad}\Phi) = \Phi\nabla^2\Phi + (\operatorname{grad}\Phi)^2$, considering that $\nabla^2\Phi = 0$, and applying Gauss' theorem $\int_V \operatorname{div}\vec{g}\mathrm{d}V = \int_A \vec{g}\cdot\mathrm{d}\vec{a}$ yields:

$$W = \frac{\mu_0}{2}\int_V (\operatorname{grad}\Phi)^2 \mathrm{d}V = \frac{\mu_0}{2}\int_V \operatorname{div}(\Phi\operatorname{grad}\Phi)\mathrm{d}V = \frac{\mu_0}{2}\int_a \Phi\operatorname{grad}\Phi\cdot\mathrm{d}\vec{a}. \tag{136}$$

Now the surface is split up into a surface at infinite distance and the surface of the current loop a_s which, itself, consists of an upper (direction of $\vec{n}_1$) and a lower surface (direction of $\vec{n}_2$):

$$\begin{aligned} W &= \frac{\mu_0}{2}\int_{a_\infty} \Phi\operatorname{grad}\Phi\cdot\mathrm{d}\vec{a} + \frac{\mu_0}{2}\int_{a_\mathrm{s}} \Phi\operatorname{grad}\Phi\cdot\mathrm{d}\vec{a} \\ &= 0 + \frac{\mu_0}{2}\int_{a_\mathrm{s}} [\Phi_1\operatorname{grad}\Phi\cdot\vec{n}_1 - \Phi_2\operatorname{grad}\Phi\cdot\vec{n}_2]\,\mathrm{d}a. \end{aligned} \tag{137}$$

Because of the continuity of B_n we get $\mu_0\operatorname{grad}\Phi\cdot\vec{n}_1 = \mu_0\operatorname{grad}\Phi\cdot\vec{n}_2 = \mu_0\operatorname{grad}\Phi\cdot\vec{n} = -B_n$ and therefore

$$W = \frac{\mu_0}{2}\int_{a_\mathrm{s}} \operatorname{grad}\Phi\cdot\vec{n}\,\mathrm{d}a = \frac{1}{2}\int_{a_\mathrm{s}} (\Phi_2 - \Phi_1)B_n\mathrm{d}a\,. \tag{138}$$

The surface of the current loop a_s can be regarded as a double layer of fictitious magnetic charges, on which $\Phi_2 - \Phi_1$ is constant, i.e., $\Phi_2 - \Phi_1 = I$ and therefore

$$W = \frac{1}{2}\int_{a_\mathrm{s}} (\Phi_2 - \Phi_1)B_n\mathrm{d}a = \frac{1}{2}\int_{a_\mathrm{s}} IB_n\mathrm{d}a = \frac{1}{2}I\psi\,. \tag{139}$$

Comparing the result (139) with $W = \frac{1}{2}L_{11}I^2 = \frac{1}{2}(L_{11}I)I$ yields $\psi = L_{11}I$. For multiple conductors we get from (139):

$$W = \frac{1}{2}\sum_{i=1}^{n} I_i\psi_i \tag{140}$$

Because of Eq. (134) it follows that $\psi_i = \sum_{j=1}^{n} L_{ij}I_j$. With Eq. (128) and Eq. (139) we also find for a single coil:

$$\psi = \frac{\int_V \vec{A}\cdot\vec{J}\mathrm{d}V}{I}. \tag{141}$$

If the currents are time dependent, the induced voltage will be

$$U_i = -\frac{d\psi_i}{dt} = -\sum_{j=1}^{n} L_{ij}\frac{dI_j}{dt} \tag{142}$$

and for a single coil $U = -L\frac{dI}{dt}$ which is also called a "back emf".

4.2 Non-linear circuits

4.21 Magnetic energy

In the non-linear case, the increment of the magnetic energy in a volume V is given by $\delta W = \int_V \vec{H} \cdot \delta\vec{B}\,\mathrm{d}V$. Because of $\mathrm{div}\,(\delta\vec{A} \times \vec{H}) = \vec{H} \cdot \mathrm{curl}\,\delta\vec{A} - \delta\vec{A} \cdot \mathrm{curl}\,\vec{H}$ Eq. (4.21) can be rewritten as

$$\begin{aligned} \delta W &= \int_V \vec{H} \cdot \mathrm{curl}\,\delta\vec{A}\,\mathrm{d}V = \int_V \mathrm{div}\,(\delta\vec{A} \times \vec{H})\,\mathrm{d}V + \int_V \delta\vec{A} \cdot \mathrm{curl}\,\vec{H}\mathrm{d}V \\ &= \oint (\delta\vec{A} \times \vec{H}) \cdot \mathrm{d}\vec{a} + \int_V \delta\vec{A} \cdot \mathrm{curl}\,\vec{H}\mathrm{d}V\,. \end{aligned} \tag{143}$$

The term $\oint(\delta\vec{A} \times \vec{H}) \cdot \mathrm{d}\vec{a}$ vanishes on the far-field boundary as $A \propto 1/r, H \propto 1/r^2, \mathrm{d}a \propto r^2$. Therefore $\delta W = \int_V \vec{J} \cdot \delta\vec{A}\mathrm{d}V$ and

$$\boxed{W = \int_V \int_{A_1}^{A_2} \vec{J} \cdot \mathrm{d}\vec{A}\mathrm{d}V.} \tag{144}$$

4.22 Self inductance in non-linear circuits

In the presence of iron parts with saturation dependent magnetization, we can still define the self inductance as

$$L(I) = \frac{\psi(I,t)}{I(t)}, \tag{145}$$

with the flux linkage depending on I, t and the current I being a function of t. Therefore the back emf. is

$$U(t) = -\frac{\mathrm{d}\psi(I,t)}{\mathrm{d}t} \tag{146}$$

and the stored energy is

$$\frac{\mathrm{d}W}{\mathrm{d}t} = I\frac{\mathrm{d}\psi}{\mathrm{d}t} \quad \rightarrow \quad \mathrm{d}W = I\frac{\mathrm{d}\psi}{\mathrm{d}t}\mathrm{d}t = I\mathrm{d}\psi \quad \rightarrow \quad W = \int_0^{\psi} I\mathrm{d}\psi\,. \tag{147}$$

With the total differential of ψ: $\mathrm{d}\psi = \frac{\partial\psi}{\partial t}\mathrm{d}t + \frac{\partial\psi}{\partial I}\mathrm{d}I$ it follows:

$$\begin{aligned} |U(t)| &= \frac{\mathrm{d}\psi(I,t)}{\mathrm{d}t} = \frac{\partial\psi(I,t)}{\partial t} + \frac{\partial\psi(I,t)}{\partial I}\frac{\mathrm{d}I(t)}{\mathrm{d}t} \\ &= L(I)\frac{\mathrm{d}I(t)}{\mathrm{d}t} + I(t)\frac{\mathrm{d}L(I)}{\mathrm{d}I}\frac{\mathrm{d}I(t)}{\mathrm{d}t} = \left(L(I) + I(t)\frac{\mathrm{d}L(I)}{\mathrm{d}I}\right)\frac{\mathrm{d}I(t)}{\mathrm{d}t}\,. \end{aligned} \tag{148}$$

If the induced voltage is measured during the ramping of the magnet, then the so-called differential inductance

$$\boxed{L_d(I) = L(I) + I\frac{\mathrm{d}L(I)}{\mathrm{d}I}} \tag{149}$$

is obtained and can be listed for the range of current values. The stored energy can then be calculated from the differential inductance with

$$W = \int_0^{I_0} IL_d(I)\mathrm{d}I\,. \tag{150}$$

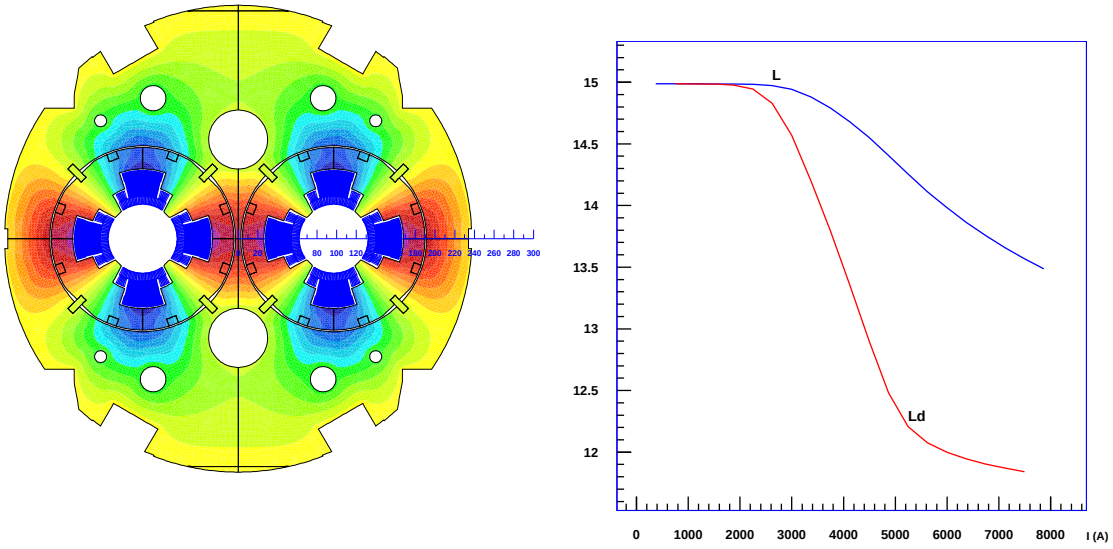


Fig. 18: Left: Magnetic vector potential in the iron yoke of the LHC insertion quadrupole (MQXA). Right: Self inductance in (mH/m) and differential inductance for the insertion quadrupole as a function of the excitational current.

Fig. 18 shows the vector potential in the iron yoke of the LHC insertion quadrupole MQXA and the self and differential inductance as a function of the excitational current. From Eq. (145), using the chain rule, one can easily verify that

$$\boxed{L_d(I) = \frac{\mathrm{d}\psi(I)}{\mathrm{d}I} \ .} \tag{151}$$

5. Superconductor Magnetization

In NbTi conductors, normal conducting deposits of a titanium-rich phase, dislocations and grain boundaries serve as *pinning centers* for the flux tubes. The pinning of flux tubes is instrumental in achieving high critical current densities. However, the very phenomenon that makes type II (hard) superconductors useful in high field magnets is responsible for magnetic hysteresis and thus for field dependent multipole errors. For the calculation of field errors in superconducting magnets due to filament magnetization a couple of "ingredients" are necessary which will be described in this section.

- A phenomenological model for filament magnetization including hysteresis modeling and taking into consideration the field dependence of the critical current density.
- The combination of these models (on the strand level) with methods of numerical field computation, for the consideration of iron saturation effects, and
- an iteration scheme to calculate the feed-back of the magnetization on the field distribution within the coil.

5.1 The critical state model

According to the *critical state model*, Bean [5], a hard superconductor tries to expel a varying external field by generating a bipolar current distribution of the critical density J_c, which depends on the local field level and the temperature. This phenomenological model takes into account that the maximum current density in the conductor is directly related to the maximum pinning force. Although the critical current

density decreases with field in all real superconductors, the Bean model assumes a field-independent critical current density to simplify the mathematical treatment of the magnetization problem. The limitations of the Bean model stem from the idealization of the electrical field versus current $\vec{E}(\vec{J})$ relation and from the fact that the explicit solution of the Maxwell equations are only possible for simple shapes of the superconductor.

Consider a superconducting slab (infinitely long in y and z-direction) which is supposed to be previously unexposed to a magnetic field and which is therefore said to be in the *virgin state*. An external field applied parallel to the slab creates a field inside the slab according to Ampère's law (in one dimension)

$$\operatorname{curl} \vec{H}_{\mathrm{s}} = \frac{\partial H_y}{\partial x}\vec{e}_z = J_z\vec{e}_z = \vec{J}_{\mathrm{c}}. \tag{152}$$

The slope of the field inside the slab is therefore equal to J_{c}, positive where J_{c} is positive (directed out of the paper plane) and negative where J_{c} is negative. At $H_{\mathrm{p}} = J_{\mathrm{c}}d/2$, where d is the thickness of the slab, the entire slab is in the critical state and H_{p} is called the penetration field.

The critical state model was originally derived for a slab of superconducting material and experimentally confirmed by Coffey [11] who mapped the field distribution in a test sample (with a little gap) of hard superconductors using a Hall probe. The slab model has been modified for cylindrical filaments by M. Wilson [42], c.f. fig. 19. Recall that a current distribution of intersecting ellipses creates an ideal dipole (shielding) field inside. Following Wilson [42] and Mess [28], and assuming a constant (shielding) current density, this current distribution is approximated by a shell with an elliptic inner boundary, see fig. 19. This boundary has the half axis a equivalent to the filament radius and the small half axis

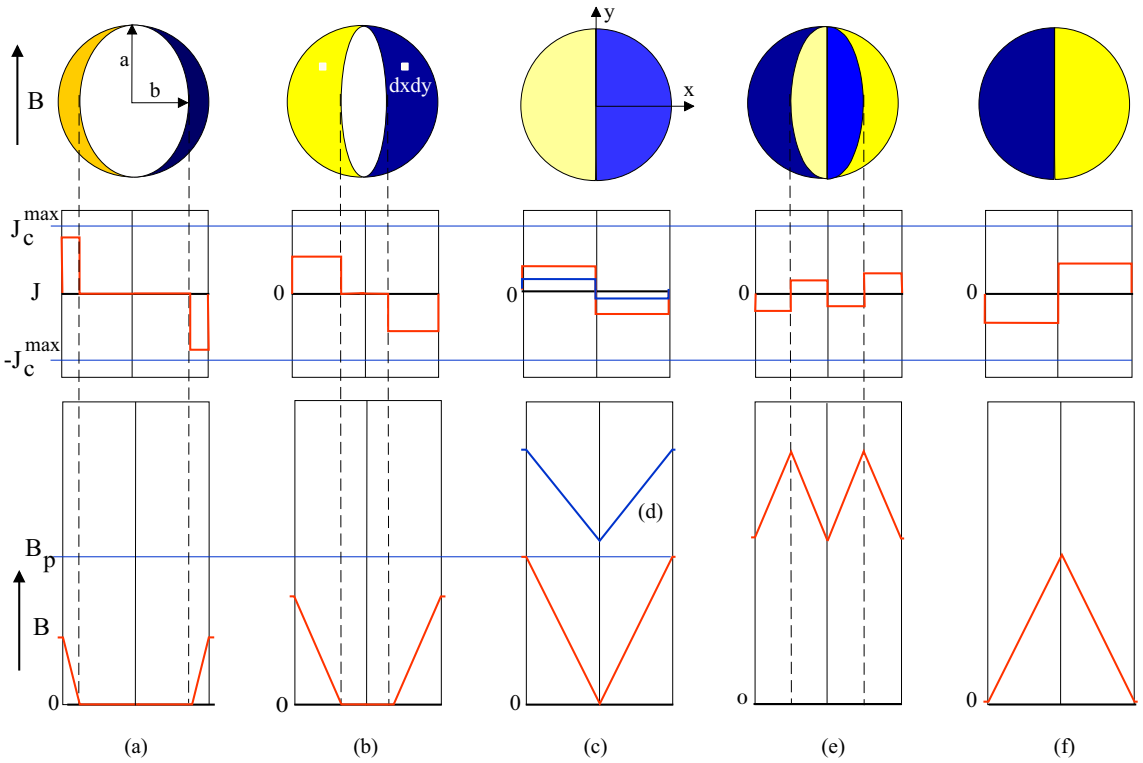


Fig. 19: The Bean model as modified by Wilson for cylindrical conductors. Notice the scaling of the critical current density with the strength of the external field B_{e}. To simplify the mathematical treatment, the current density is assumed to be constant within the cylinder but dependent on the external field, resulting in different slopes for the penetrating and trapped field inside the filament.

b, see fig. 19,a. The contribution of two area elements $\mathrm{d}x\mathrm{d}y$ at the locations (x, y) and $(-x, y)$ to the shielding field is

$$\mathrm{d}B_\mathrm{s} = -2\frac{\mu_0 J_\mathrm{c}}{2\pi\sqrt{x^2+y^2}}\cos\Theta\mathrm{d}x\mathrm{d}y \tag{153}$$

with $\cos\Theta = x/\sqrt{x^2+y^2}$ we get

$$B_\mathrm{s} = -\frac{\mu_0 J_\mathrm{c}}{\pi}\int_{-a}^{a}\left[\int_{u(y)}^{v(y)}\frac{x}{x^2+y^2}\mathrm{d}x\right]\mathrm{d}y \tag{154}$$

where $u(y) = b\sqrt{1-\frac{y^2}{a^2}} = \frac{b}{a}\sqrt{a^2-y^2}$ resulting from the equation of the ellipse and $v(y) = \sqrt{a^2-y^2}$ resulting from the equation of the circle. Integration yields:

$$B_\mathrm{s} = -\frac{2\mu_0 J_\mathrm{c} a}{\pi}\left(1 - \frac{b}{a}\frac{\arcsin\sqrt{1-\frac{b^2}{a^2}}}{\sqrt{1-\frac{b^2}{a^2}}}\right). \tag{155}$$

The maximum field that can be shielded from the center of the filament is called the *penetration field* where the current distribution resembles two half cylinders (small half axis of the ellipse b=0), see fig. 19 c.

$$B_\mathrm{p}(B,T) = \frac{\mu_0 J_\mathrm{c}(B,T)d}{\pi}, \tag{156}$$

with the filament diameter d. Wilson proposes to use the following dependency of J_c on the (external) field at constant temperature which can be expressed by the relation [20]

$$J_\mathrm{c}(B) = \frac{J_0 B_0}{B_0 + |B|} \tag{157}$$

where J_0 and B_0 are fit constants determined by magnetization measurements ($J_0 \approx 50\mathrm{kA/mm}^2$, $B_0 \approx 0.29\mathrm{T}$). For the calculation of the resulting magnetic moment $m\,[\mathrm{A}\cdot\mathrm{m}^2]$ the above exercise can be repeated considering again the area elements and the magnetic moment generated by the resulting current loop: $\mathrm{d}m = -J_\mathrm{c}\,2x\,l\,\mathrm{d}x\mathrm{d}y$ and therefore

$$m = -2J_\mathrm{c} l\int_{-a}^{a}\left[\int_{u(y)}^{v(y)} x\mathrm{d}x\right]\mathrm{d}y = -\frac{4}{3}J_\mathrm{c}(1-\frac{b^2}{a^2})a^3 l \tag{158}$$

The magnetization (magnetic moment per unit volume) $[\mathrm{A/m}]$ is $M = \frac{m}{\pi a^2 l} = -\frac{4}{3\pi}J_\mathrm{c}(1-\frac{b^2}{a^2})a$. The magnetization of a strand containing only fully penetrated filaments can then be calculated to

$$\boxed{M_\mathrm{p}(B,T) = -\lambda\frac{2}{3\pi}J_\mathrm{c}(B,T)d} \tag{159}$$

and has the same direction as the applied field ($\vec{M} = M\vec{B}/|B|$), which varies over the cross-section of the magnet winding. With $\lambda = 0.29$ (the filling factor for the filamentary superconductor) and a filament diameter $d = 6\mu\mathrm{m}$, the magnetization $\mu_0 M$ is about 9 mT at 0.53 T injection field level, where the critical current density of NbTi is 19000 A/mm^2. If the field is raised above the penetration field level, the shielding current distribution is maintained but with field penetrating into the filament and therefore M_p decreases proportionally to the critical current density.

5.2 Magnetization model with varying current density

The Wilson model has been modified, to allow for varying current densities within the filament. The filament is described as a set of nested intersecting ellipses with varying current density. Each layer produces a dipole screening field which screens the field inside the superconductor and therefore increases the critical current density in the inner layers, c.f. fig. 21.

5.21 The $J_c(B,T)$ dependence

The dependence of the critical current density on the modulus of the magnetic induction $B = |\vec{B}|$ is given by the following fit [4], where $B_c = B_{c0}(1-(T/T_{c0})^{1.7})$:

$$J_c(B,T) = \frac{J_c^{\mathrm{ref}} C_0 B^{\alpha-1}}{(B_c)^{\alpha}} \left(1-\frac{B}{B_c}\right)^{\beta} \left(1-\left(\frac{T}{T_{c0}}\right)^{1.7}\right)^{\gamma} . \tag{160}$$

The fit parameters for the LHC main-magnet cables are a critical current density at $4.2\,\mathrm{K}$ and $5\,\mathrm{T}$ of $J_c^{\mathrm{ref}} = 3\cdot 10^9\,\mathrm{A/m^2}$, an upper critical field $B_{c0} = 14.5\,\mathrm{T}$, a critical temperature of $T_{c0} = 9.2\,\mathrm{K}$, a normalization constant $C_0 = 27.04\,\mathrm{T}$ and the fit parameters $\alpha = 0.57$, $\beta = 0.9$ and $\gamma = 2.32$.

For small magnetic inductions B, where the persistent currents influence the field quality most, Eq. (160) strives for infinity with $B^{\alpha-1} = B^{-0.43} \sim 1/\sqrt{B}$ for $B \to 0$. For the computation of the induction inside the filament, Eq. (160) is approximated around the actual value of the applied field $B_{\mathrm{out}} = |\vec{B}_{\mathrm{out}}|$ with the following function (T is constant):

$$J_c(B) \sim J_c(B_{\mathrm{out}}) \frac{\sqrt{B_{\mathrm{out}}}}{\sqrt{B}} \equiv \frac{\mathcal{F}(B_{\mathrm{out}})}{\sqrt{B}} . \tag{161}$$

5.22 The geometry and screening field of nested intersecting circles

Fig. 22 shows on the left the geometry of nested intersecting circles. For illustration the number of intersecting circles is $n = 3$. On the right the geometry of nested intersecting ellipses with $n = 5$ is displayed.

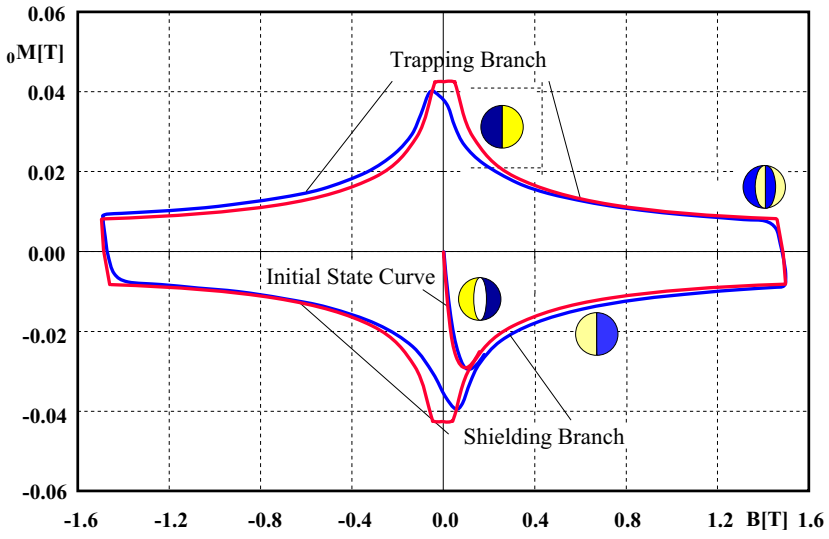


Fig. 20: Measured (blue) and calculated (red) hysteresis of a LHC strand subjected to a varying field with patterns of the persistent currents according to the critical state model as applied by Wilson to cylindrical filaments. It can be seen that the Wilson model fails to explain the "peak shifting" of the maximal magnetization at low field values.

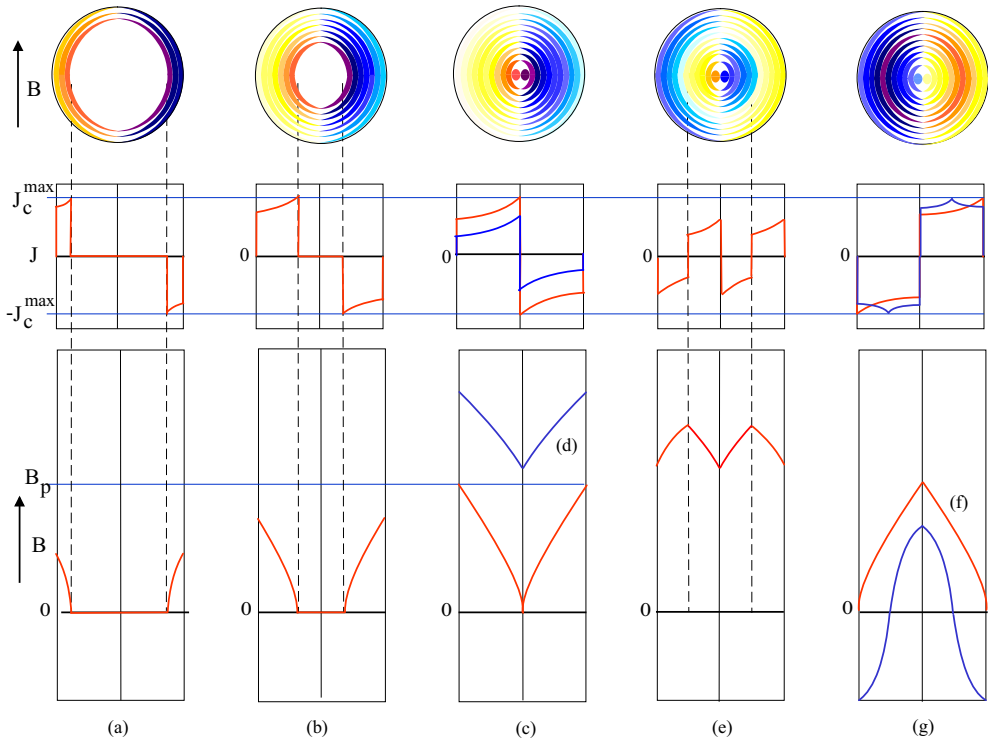


Fig. 21: Magnetization model with nested intersecting ellipses with current density depending on the fi eld distribution within the fi lament.

The thickness c of the current carrying shell is then constant for all circles $c = \frac{U}{n}$ and the radius of the i-th circle is $a_i = U - (2i-1)\frac{c}{2}$. In case of intersecting ellipses the other semi-axis can be calculated by $b_i = a_i \frac{V}{U}$. The shielding field (denoted t) of two intersecting ellipses is constant in the aperture

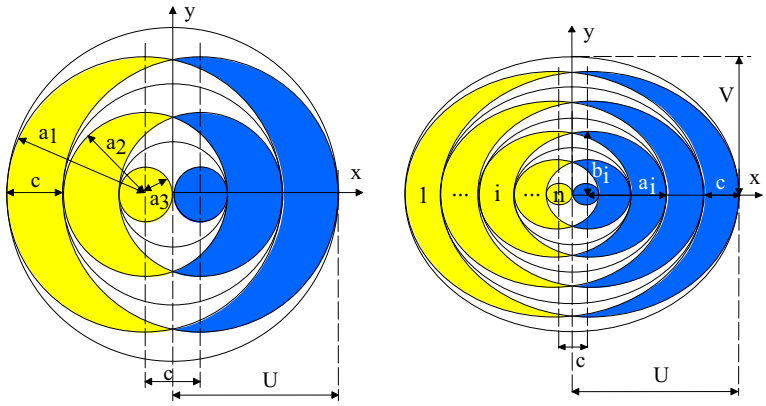


Fig. 22: Left: Geometry of nested intersecting circles. For illustration $n = 3$. Right: Geometry of nested intersecting ellipses ($n = 5$). In the semi-analytical model $n \to \infty$.

and can be calculated to $t = \mu_0 J_c c \frac{V}{U+V}$, where U and V are the half-axes of the ellipses and c is the displacement between them (which is identical with the thickness of the current layer on the x-axis). Easier relations result for the intersecting circles of radius r where the shielding field is given by

$$t = 0.5 \mu_0 J_c c r \, . \tag{162}$$

Let now q be the relative penetration parameter which is zero at the surface of the filament and equals one in the center. The uniform dipole field produced by two intersecting circles with opposite current densities shifted by the relative distance $\Delta q = q_2 - q_1$, can then be expressed as

$$|\Delta\vec{t}| = \frac{\mu_0 r}{2} \int_{q_1}^{q_2} J_\mathrm{c}(B(q))\,\mathrm{d}q, \tag{163}$$

where $\Delta\vec{t}$ is the shielding field, r is the filament radius and q_1, q_2 are the relative penetration parameters that limit the shielding current layer. Such pairs of circles are nested inside concentric circles. This equation will now be used to find a differential equation for the differential shielding $\mathrm{d}\vec{t}(\mathrm{d}q)$. We get from (163) and (161):

$$\mathrm{d}B(q) = \xi\mu_0\,\mathcal{H}\,J_\mathrm{c}(B(q))\,\mathrm{d}q = \frac{\xi\mu_0\,\mathcal{H}\,\mathcal{F}(B_\mathrm{out})\,\mathrm{d}q}{\sqrt{B(q)}}\,, \tag{164}$$

where r is the filament radius. The geometry factor $\mathcal{H} = r/2$ corresponds to the ideal screening field of two intersecting circles. As a refinement $\mathcal{H} = r(2-2\ln 2) = 0.614\,r$, correcting for the little spaces that are left when a round filament is filled with a series of intersecting circles (see fig. 22). The parameter ξ equals -1 in case of ramping up and $\xi = 1$ for ramping down. In the first case, the orientation of the magnetic moment of the screening current is opposite to the orientation of the outside field B_out and B decreases inside the filament.

Equation (164) is a differential equation for $B(q)$, considering the dependence of J_c on $B(q)$, that can be solved with the known boundary condition $B(0) = B_\mathrm{out}$ in a closed analytical form, yielding:

$$B(q) = \left(B_\mathrm{out}^{3/2} + \frac{3}{2}\,\xi\,\mathcal{H}\,\mathcal{F}(B_\mathrm{out})\mu_0\,q\right)^{2/3} \tag{165}$$

Figure 23 shows $B(q)$ according to Eq. (165) and the dependence of the critical current density $J_\mathrm{c}(B(q))$ on the penetration depth q. The dotted line shows the magnetic induction for a constant current density $J_\mathrm{c}(B_\mathrm{out})$ and demonstrates the importance of using Eq. (165) instead. The magnetic induction at $q = 0$ equals the external field B_out. The shown field distribution is a fully penetrated state, on the shielding branch reached after increasing the external field from negative field values to $B_\mathrm{out} = 0.08\,\mathrm{T}$ ($\xi = -1$). As is shown in Figure 23, this results in a decreasing field $B(q)$ along the penetration depth which produces an increase of $J_\mathrm{c}(B(q))$ along q. At $B(q) = 0$, the critical current density reaches its maximum value. There the strong increase of J_c produces a sharp decline of $B(q)$. The course of $J_\mathrm{c}(q)$ shows the importance of expressing J_c as a function of q rather than assuming a constant value.

5.23 The peak-shifting in the hysteresis curve

From the expression for the magnetic induction $B(q)$ inside the filament, the magnetization due to the radial slice of current $J_\mathrm{c}(q)$ between the penetrations q_i and q_{i+1} is derived. Individual slice magnetizations are needed to describe the hysteresis after changes of the ramp direction. Such a change ($\mathrm{d}B_\mathrm{out}/\mathrm{d}t$ changes sign) will produce a new layer of screening currents with opposite polarity (ξ switches sign). For small changes the *new* current layer will penetrate the filament only from $q_1 = 0$ to $q_2 \leq 1$ while the currents inside persist. The values for q_i are calculated using Eq. (165). For minor excitation loops, the magnetization is obtained as the superposition of n different layers,

$$M = \sum_{i=1}^{n} M_i = \sum_{i=1}^{n} \int_{q_i}^{q_{i+1}} \mathcal{M}_i(q)\,\mathrm{d}q\,. \tag{166}$$

In Eq. (166) M_i denotes the modulus of the magnetization $\vec{M}_i$; it has negative values if the orientation is opposite to the external field. $\mathcal{M}$ is the magnetization density upon q.

$$M_i = \frac{4r\xi}{\mu_0\pi} \int_{q_i}^{q_{i+1}} J_c(B(q))(1-q)^2 \, \mathrm{d}q = \frac{4r\xi\,\mathcal{F}}{\mu_0\pi} \int_{q_i}^{q_{i+1}} \frac{(1-q)^2}{\sqrt{B(q)}} \, \mathrm{d}q. \tag{167}$$

Equation (167) can be solved analytically and (together with Eq. (165)) yields a closed expression for the filament magnetization:

$$M_i = \frac{4rB(q)}{5\pi\mathcal{F}^2\mu_0^3\mathcal{H}^3} \left[B_{\mathrm{out}}^3 + \xi\mathcal{H}\mathcal{F}\mu_0 \left(\left(5 - 4q + \frac{5}{4}q^2 \right) \xi\mathcal{H}\mathcal{F}\mu_0 - (q-4)\, B_{\mathrm{out}}^{3/2} \right) \right] \Bigg|_{q=q_i}^{q=q_{i+1}} \tag{168}$$

From the expressions for the magnetic induction $B(q)$, Eq. (165), and the magnetization M, eqs. (167) and (166), both parameters can be computed as a function of the penetration depth q. Figure 24 (upper plot left), shows the values for increasing external fields B_{out} ($\xi = -1$) for the virgin curve creating one layer extending from $q_1 = 0$ to $q_2(B_{\mathrm{out}})$. Depending on B_{out}, the field decreases until a certain penetration depth, where complete screening of the external field is obtained. The remaining part of the filament stays field free. In the lower plot the contribution $m(q)$ of a slice $\mathrm{d}q$ to the total magnetization M is presented. The value of M can be obtained by integrating the presented curves (see indication on the plot).

Figure 24 upper right, illustrates the same quantities as on the left hand side, but for a filament that has already experienced a negative outside field before (different history) and hence is fully penetrated. Since the currents inside the superconductor persist, there is a remaining negative field $B(q)$ inside, whereas in the case of the virgin curve the field remains zero for $q > q_2(B_{\mathrm{out}})$. The lower plot in fig. 24 right, also explains why the maximum magnetization does not occur at $B_{\mathrm{out}} = 0$: The magnetization is given by the integrated area M under the $m(q)$ curve, which is biggest for small values of $B_{\mathrm{out}} \neq 0$. This characteristic behavior has also been observed in measurements (see fig. 25), and is in good agreement with the calculations.

5.24 Calculation of B_{p1}

The magnetic induction B_{p1} denotes the value of the outside magnetic induction where the modulus of the filament magnetization passes through its first maximum when ramping up on the virgin curve (see

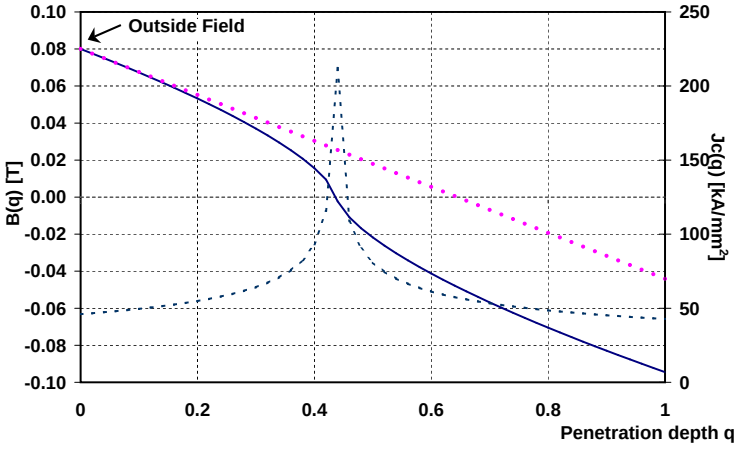


Fig. 23: Magnetic induction $B(q)$ as a function of the penetration depth q (continuous line). The dashed line denotes the current density $J_c(q)$. The dotted line shows the magnetic induction for a calculation assuming a constant current density.

fig. 25). Since the magnetization has been calculated in a closed analytical form (Eq. (168)), we can now derive the maximum of the virgin curve by solving $\mathrm{d}M(B_{\mathrm{out}})/\mathrm{d}B_{\mathrm{out}} = 0$. For the virgin curve the magnetization consists of only one layer, $n = 1$, that penetrates from $q_1 = 0$ to

$$q_2 = 2B_{\mathrm{out}}^{3/2}/(\mathcal{H}\,\mathcal{F}(B_{\mathrm{out}})\mu_0)\,, \tag{169}$$

obtained by solving $B(q_2) = 0$. In this region of $B_{\mathrm{out}} = B_{p1}$ we find that $\mathcal{F}'(B_{\mathrm{out}}) \cong 0$ and hence we obtain

$$B_{p1} \;\cong\; (\mathcal{H}\,\mathcal{F}(B_{p1})\,\mu_0)^{2/3}\,\frac{(15-5\sqrt{5})^{1/3}}{2}\,, \tag{170}$$

$$q_2 \;\cong\; \sqrt{\frac{5}{6}-\frac{5\sqrt{5}}{18}} \cong 0.46\,. \tag{171}$$

The recursive Eq. (170) yields a good estimate for B_{p1} after few iterations. From Eq. (171) can be seen that the maximum modulus of the magnetization does occur at a penetration depth of $q_2 \cong 0.46$ rather than in the fully penetrated state. This fact is illustrated by the lower plot of fig. 24 right, where the area will be maximal for $q \to 0.46$ (solid line). Note, that the value of q_2 is independent of the critical current fit, provided $\mathrm{d}\mathcal{F}(B_{\mathrm{out}})/\mathrm{d}B_{\mathrm{out}} \cong 0$ (i.e. the critical current diverges with $1/\sqrt{B_{\mathrm{out}}}$, for $B_{\mathrm{out}} \to 0$, see Eq. (161)).

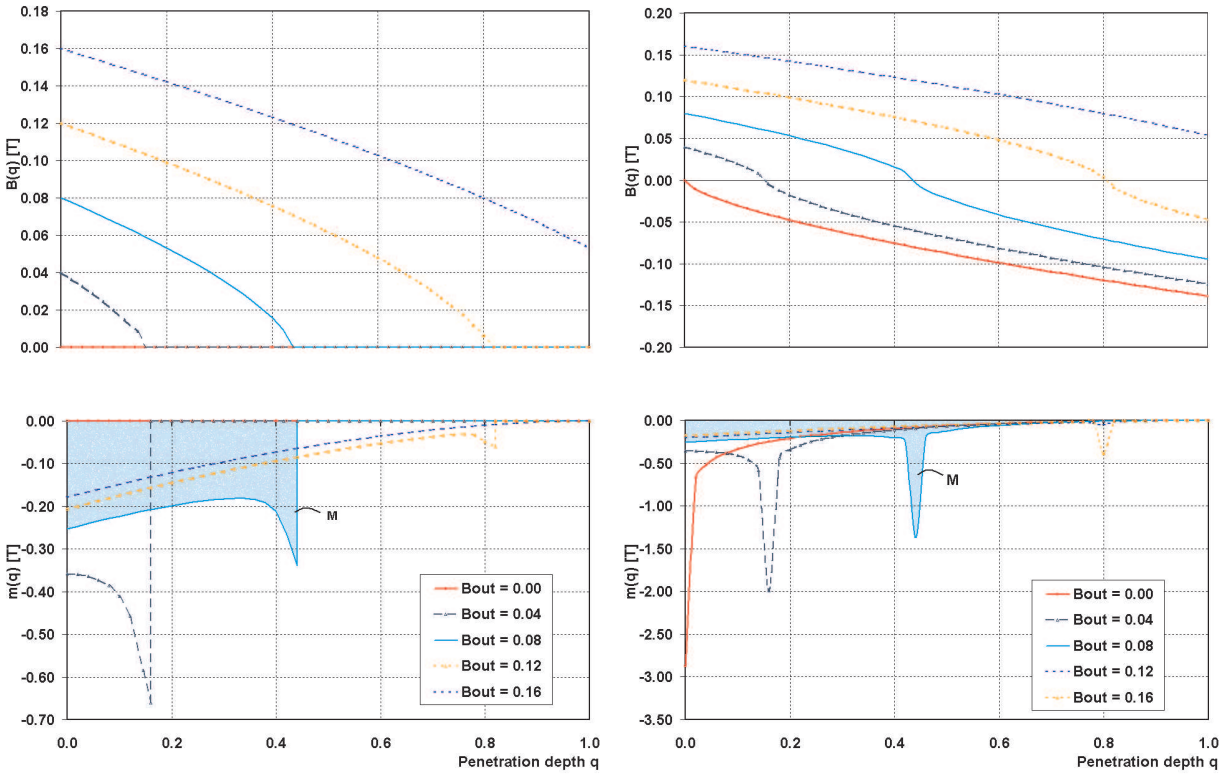


Fig. 24: Left (top and bottom): Magnetic induction $B(q)$ and the magnetization contribution $m(q)$ as a function of the penetration depth q for the virgin curve. Right (top and bottom): Course of the magnetic induction $B(q)$ and the magnetization contribution $m(q)$ as a function of the penetration depth q for a filament already been exposed to a magnetic field before (non-virgin curve).

5.25 The hysteresis loop for a multifilamentary strand

Figure 25 presents computations of the filament magnetization according to Eq. (168) multiplied with the filling factor λ. The virgin curve and several hysteresis loops are displayed. The comparison of the calculation and the measured magnetization of a superconducting strand (dashed line) shows good agreement apart from the region of B close to zero. There the difference between the magnetizations of one filament and a whole strand becomes significant. Since the outside field for each filament varies slightly due to the position in the strand cross-section and is additionally influenced by the field arising from the screening currents in the neighboring filaments, B_{out} will be different for each filament according to its exact position. This results in a spread of filament magnetizations and hence in a smoothening of the region of B close to zero. Since the Rutherford cables used in LHC magnets consist of many individual strands, this region will be smoothened out automatically due to the differences of B_{out} at the individual strand positions within the coil.

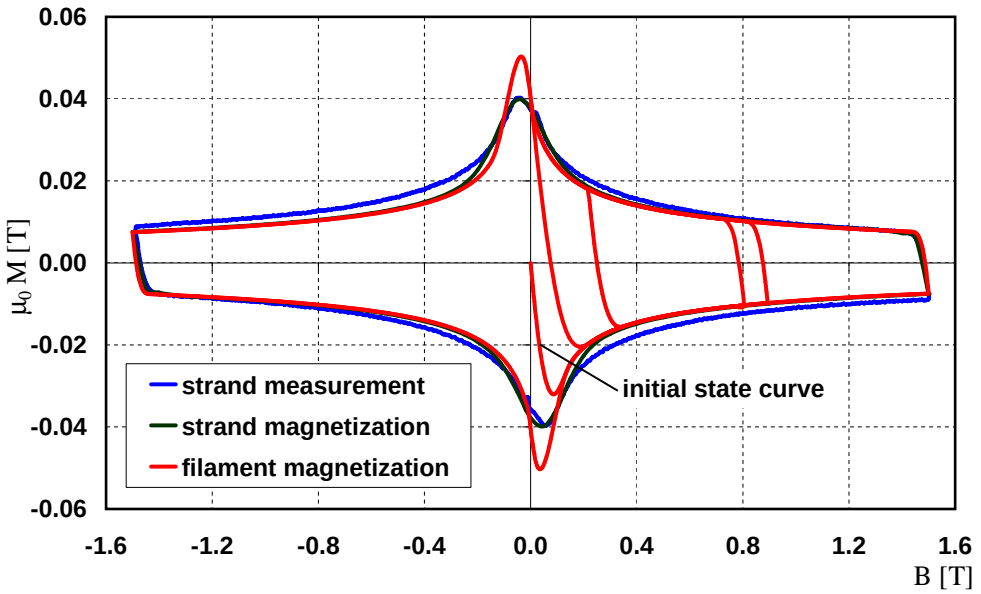


Fig. 25: Computed magnetization curve for filament a filament diameter of $r = 3.5\,\mu\mathrm{m}$, filling factor $\lambda = 1/2.95$, and temperature $T = 1.9\,\mathrm{K}$) compared with measurements from a superconducting strand [45].

5.26 Strand magnetization in the LHC main dipole coil

The external field, seen by individual filaments, depends on their position in the coil geometry. Filaments in the outer layer of the coil (close to abscissa) experience low fields (dark blue regions in fig. 26, left), but high variations in the field if saturation effects cause a movement of this low field region within the coil. Filaments in the inner coil layer experience higher field (red and purple regions in fig. 26, left, which is increasing with inverse dependency to the positioning angle. The modulus of the superconducting filament magnetization in the coil cross-section is shown in fig. 26, right. Even at nominal field there are filaments in the coil cross-section remaining still non-fully penetrated.

5.3 Field errors due to the filament magnetization

The magnetic moment per unit length of a strand with cross-section A, $\vec{m}/l = A\vec{M}$ can be represented by a small dipole of line-currents with the intensity $-I_s$ and I_s spaced by a distance S apart and located perpendicular to the field direction. The magnetic moment of such a single dipole $m/l = I_s \cdot S$ must equal the magnetic moment arising from persistent currents, i.e., $I_s = \frac{MA}{S}$ where S can be chosen as

the strand diameter. For a strand surface of approx. A = 0.5 mm^2 we get about 5 A, which compares to about 20 A of transport current in each strand at injection field level.

A more elegant method is to calculate the vector potential at a point $\vec{r}$ from a magnetic moment at point $\vec{r}\,'$ (for the coordinate system see fig. 6) using the identity

$$A_z(\vec{r}) = \frac{\mu_0 \vec{m}}{2\pi} \times \operatorname{grad}_{\vec{r}\,'} \ln \frac{|\vec{r} - \vec{r}\,'|}{a} \tag{172}$$

which holds for two-dimensional problems. With Eq. (57) and Eq. (59) we get

$$\ln \frac{|\vec{r} - \vec{r}\,'|}{a} = \ln(\frac{r_i}{a}) - \sum_{n=1}^{\infty} \frac{1}{n} (\frac{r_0}{r_i})^n \cos(n(\varphi - \theta)) \,. \tag{173}$$

and with the gradient in two dimensional cylindrical coordinates $\operatorname{grad}_{\vec{r}\,'} = \frac{\partial}{\partial r_i}\vec{e}_{\vec{r}\,'} + \frac{1}{r_i}\frac{\partial}{\partial\theta}\vec{e}_\theta$ it follows that

$$\begin{aligned} &\operatorname{grad}_{\vec{r}\,'} \ln \frac{|\vec{r} - \vec{r}\,'|}{a} = \\ &\qquad = \frac{1}{r_i} \left[(1 + \sum_{n=1}^{\infty} (\frac{r_0}{r_i})^n \cos(n(\varphi - \theta))) \, \vec{e}_{\vec{r}\,'} + \sum_{n=1}^{\infty} (\frac{r_0}{r_i})^n \sin(n(\varphi - \theta)) \vec{e}_\theta \right] . \end{aligned} \tag{174}$$

Introducing this result in Eq. (172) and calculating the cross-product yields

$$A_z = \frac{\mu_0}{2\pi r_i} \vec{e}_z \left[m_{\vec{r}\,'} \sum_{n=1}^{\infty} (\frac{r_0}{r_i})^n \sin(n(\varphi - \theta)) - m_\theta (1 + \sum_{n=1}^{\infty} (\frac{r_0}{r_i})^n \cos(n(\varphi - \theta))) \right] . \tag{175}$$

With $B_r(r_0, \varphi) = \frac{1}{r_0}\frac{\partial A_z}{\partial \varphi}$ and $\sin(n\varphi - n\theta) = \sin n\varphi \, \cos n\theta - \cos n\varphi \, \sin n\theta$, $\cos(n\varphi - n\theta) =$

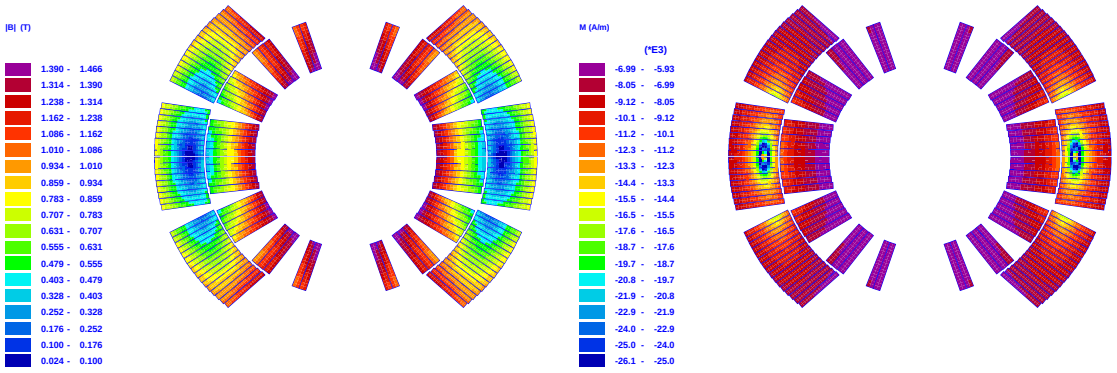


Fig. 26: Left: Modulus of magnetic fi eld in the coil. Right: Modulus of superconducting fi lament magnetization, both at about 1.3 T fi eld level ramped from zero on the virgin curve.

$\cos n\varphi\ \cos n\theta + \sin n\varphi\ \sin n\theta$, it follows that

$$\begin{aligned} B_r(r_0,\varphi) &= \frac{\mu_0}{2\pi r_0 r_i}\left[m_{\vec{r}'} \sum_{n=1}^{\infty} (\frac{r_0}{r_i})^n\, n(\cos n\varphi\ \cos n\theta + \sin n\varphi\ \sin n\theta) \right. \\ &\quad \left. - m_\theta \sum_{n=1}^{\infty} (\frac{r_0}{r_i})^n\, n(-\sin n\varphi\ \cos n\theta + \cos n\varphi\ \sin n\theta) \right] \\ &= \frac{\mu_0}{2\pi r_0 r_i}\left[\sum_{n=1}^{\infty} n(\frac{r_0}{r_i})^n (m_{\vec{r}'}\ \cos n\theta - m_\theta\ \sin n\theta) \cos n\varphi \right. \\ &\quad \left. + \sum_{n=1}^{\infty} n(\frac{r_0}{r_i})^n (m_{\vec{r}'}\ \sin n\theta + m_\theta\ \cos n\theta) \sin n\varphi \right] . \end{aligned} \tag{176}$$

For the multipole coefficients we finally obtain

$$A_n = \frac{\mu_0}{2\pi} \frac{r_0^{n-1}}{r_i^{n+1}} n(m_{\vec{r}'}\ \cos n\theta - m_\theta\ \sin n\theta)\,, \tag{177}$$

$$B_n = \frac{\mu_0}{2\pi} \frac{r_0^{n-1}}{r_i^{n+1}} n(m_{\vec{r}'}\ \sin n\theta + m_\theta\ \cos n\theta)\,. \tag{178}$$

The contribution of the strand magnetization to the B_3 field component is displayed in fig. 27.

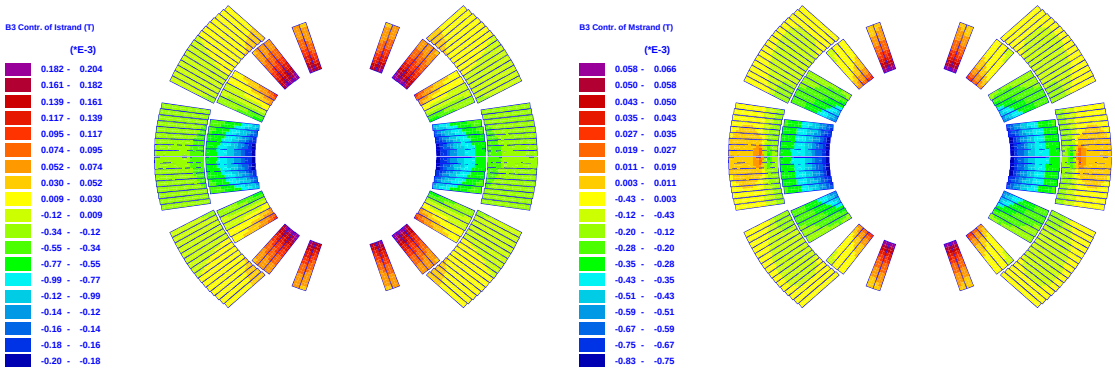


Fig. 27: Left: Contribution of the strand current to the B_3 fi eld component. Right: Contribution of the strand magnetization to the B_3 fi eld component (both at about 1.3 T fi eld level ramped from zero on the virgin curve and at 17 mm reference radius).

6. Summary

The design and optimization of superconducting accelerator magnets requires numerical methods which allow the accurate modeling (both in two and three dimensions) of the coil with its keystoned cables, wedges and insulation, and which limit the numerical errors to the effect from the iron yoke. Semi-analytical methods for the calculation of the excitational field from the coil combined with the BEM-FEM coupling method and higher order quadrilateral finite elements provide the required accuracy of the field solution. Although this report cannot provide in depth treatment of the fundamentals in field theory and numerical field computation, we have motivated the development of the CERN field-computation program ROXIE which is used as an approach towards an integrated design of superconducting magnets.

7. Appendix A: Feed-down of multipole components

In order to derive the transformation law for the field harmonics $c_n \rightarrow c'_n$ ($c_n = b_n + i\,a_n$) for a translation of the reference frame $z \rightarrow z'$: $z = z' + \Delta z$ ($z = x + iy$), we consider the field components to be identical in both coordinate systems.

$$\sum_{n=1}^{\infty} c_n \left(\frac{z}{r_0}\right)^{n-1} = \sum_{n=1}^{\infty} c'_n \left(\frac{z'}{r_0}\right)^{n-1} = B_y + i\,B_x \ . \tag{179}$$

Using the binomial series

$$\left(\Delta z + z'\right)^{n-1} = \sum_{k=1}^{n} \frac{(n-1)!}{(n-k)!\ (k-1)!} \left(z'\right)^{k-1} \Delta z^{n-k} \ , \tag{180}$$

the left hand side of equation (179) becomes

$$\begin{aligned} \sum_{n=1}^{\infty} c_n \left(\frac{z}{r_0}\right)^{n-1} &= \sum_{n=1}^{\infty} \frac{c_n}{{r_0}^{n-1}} \left(z' + \Delta z\right)^{n-1} \\ &= \sum_{n=1}^{\infty} \frac{c_n}{{r_0}^{n-1}} \sum_{k=1}^{n} \frac{(n-1)!}{(n-k)!\ (k-1)!} \left(z'\right)^{k-1} \Delta z^{n-k} \ . \end{aligned} \tag{181}$$

For the coefficients $\mathcal{C}_{k_0}$ of one particular power of z', e.g., $(z')^{k_0-1}$ we get

$$\mathcal{C}_{k_0} = \sum_{n=k_0}^{\infty} \frac{c_n}{{r_0}^{n-1}} \frac{(n-1)!}{(n-k_0)!\ (k_0-1)!} \Delta z^{n-k_0} \tag{182}$$

$$= \frac{1}{{r_0}^{k_0-1}} \sum_{n=k_0}^{\infty} c_n \frac{(n-1)!}{(n-k_0)!\ (k_0-1)!} \left(\frac{\Delta z}{r_0}\right)^{n-k_0} \ . \tag{183}$$

Merging equations (182) and (181) yields (indices changed: $n \rightarrow k$ and $k_0 \rightarrow n$)

$$\begin{aligned} \sum_{n=1}^{\infty} c_n \left(\frac{z}{r_0}\right)^{n-1} &= \sum_{n=1}^{\infty} \left(z'\right)^{n-1} \mathcal{C}_n \\ &= \sum_{n=1}^{\infty} \left(z'\right)^{n-1} \frac{1}{{r_0}^{n-1}} \sum_{k=n}^{\infty} c_k \frac{(k-1)!}{(k-n)!\ (n-1)!} \left(\frac{\Delta z}{r_0}\right)^{k-n} \\ &= \sum_{n=1}^{\infty} c'_n \left(\frac{z'}{r_0}\right)^{n-1} \ . \end{aligned} \tag{184}$$

Comparing the coefficients of $\left(\frac{z'}{r_0}\right)^{n-1}$ leads to the transformation law for the complex field harmonics

$$c'_n = \sum_{k=n}^{\infty} c_k \frac{(k-1)!}{(k-n)!\ (n-1)!} \left(\frac{\Delta z}{r_0}\right)^{k-n} \ . \tag{185}$$

References

[1] M. Aleksa, S. Russenschuck and C. Völlinger Magnetic Field Calculations Including the Impact of Persistent Currents in Superconducting Filaments IEEE Trans. on Magn., vol. 38, no. 2, 2002.

[2] Armstrong, A.G.A.M. , Fan, M.W., Simkin, J., Trowbridge, C.W.: Automated optimization of magnet design using the boundary integral method, IEEE Transactions on Magnetics, 1982

[3] Bossavit, A.: Computational Electromagnetism, Academic Press, 1998

[4] L. Bottura, A Practical Fit for the Critical Surface of NbTi, 16th International Conference on Magnet Technology, Florida, 1999

[5] Bean, C.P.: Physics review letters, 1962

[6] T. Beier et al., Feature-based image metamorphosis, Computer Graphics 26(2), 1992.

[7] Beth, R. A.: An Integral Formula for Two-dimensional Fields, Journal of Applied Physics, Vol. 38, Nr. 12, 1967

[8] Binns, K.J., Lawrenson, P.J., Trowbridge, C.W.: The analytical and numerical solution of electric and magnetic fields, John Wiley & Sons, 1992

[9] T.D. Blacker et al., Int. J. Numer. Method. Eng., 32, 811-847, 1991.

[10] Buckel, W.: Supraleitung, VCH, Weinheim, 1990

[11] Coffey, H.T.: Distribution of magnetic fields and current in Type II superconductors, Cryogenics 7, 1967.

[12] Brebbia, C.A.: The boundary element method for engineers, Pentech Press, 1978.

[13] Braess, D.: Finite Elemente, Springer, 1991

[14] Fetzer, J., Abele, S., Lehner, G.: Die Kopplung der Randelementmethode und der Methode der finiten Elemente zur Lösung dreidimensionaler elektromagnetischer Feldprobleme auf unendlichem Grundgebiet, Archiv für Elektrotechnik, 76(5), 1993.

[15] Fetzer, J., Kurz, S., Lehner, G.: Comparison between different formulations for the solution of 3D nonlinear magnetostatic problems using BEM-FEM coupling, IEEE Transactions on Magnetics, Vol. 32, 1996

[16] E. Gröller, Interactive transformation of 2d vector data, Technical University Vienna, Institute for Computer Graphics.

[17] Halbach, K.: A Program for Inversion of System Analysis and its Application to the Design of Magnets, Proceedings of the International Conference on Magnet Technology (MT2), The Rutherford Laboratory, 1967

[18] Halbach, K., Holsinger R.: Poisson user manual, Techn. Report, Lawrence Berkeley Laboratory, Berkeley, 1972

[19] Hornsby, J.S.: A computer program for the solution of elliptic partial differential equations, Technical Report 63-7, CERN, 1967

[20] Kim, Y.B., Hempstead, C.F., Strnad, A.R.: Critical currents in hard superconductors, Physics review letters, 1962

[21] Kurz, S., Fetzer, J., Rucker, W.M.: Coupled BEM-FEM methods for 3D field calculations with iron saturation, Proceedings of the First International ROXIE users meeting and workshop, CERN, March 16-18, 1998

[22] Kurz, S., Russenschuck, S.: The application of the BEM-FEM coupling method for the accurate calculation of fields in superconducting magnets, Electrical Engineering, 1999

[23] Kurz, S., Russenschuck, S.: Different Approaches for Accurate Magnetic Field Computation in Superconducting Magnets Using BEM-FEM Coupling, Conference on the Computation of Electromagnetic Fields, COMPUMAG, Sapporo, 1999.

[24] Kurz, S., Russenschuck, S.: Accurate Calculation of Magnetic Fields in the End Region of Superconducting Accelerator Magnets using the BEM-FEM Coupling Method, PAC99, New-York

[25] S. Kurz and S. Russenschuck, The application of the BEM-FEM coupling method for the accurate calculation of fields in superconducting magnets, Electrical Engineering, 1999.

[26] M.S. Lubell: Empirical Scaling Formulas for Critical Current and Critical Field for Commerical NbTi, MT16, 1999.

[27] Mathematica 4.1 is a trademark by Wolfram Research Inc.

[28] Mess, K.H., Schmüser, P., Wolff S.: Superconducting Accelerator Magnets, World Scientific, 1996

[29] Le Naour S., et. al.: Magnetization measurement on LHC superconducting strands, Applied superconductivity Conference, Palm Springs, USA, 1998.

[30] G.D. Nowottny, Netzerzeugung durch Gebietszerlegung und duale Graphenmethode, Shaker Verlag, Aachen, 1999.

[31] Orlando, T.P., Delin, K.A.: Foundations of applied superconductivity, Addison-Wesley, 1991

[32] M. Pekeler et al., Coupled Persistent-Current Effects in the Hera Dipoles and Beam Pipe Correction Coils, Desy Report no. 92-06, Hamburg, 1992

[33] Perin, R.: Field, forces and mechanics of superconducting magnets, CERN Accelerator School, CERN Yellow Report, 96-03, ISBN 92-9083-084-0

[34] Preis, K., Biro, O., Magele, C.A., Renhart, W., Richer, K.R., Vrisk, G.: Numerical analysis of 3D magnetostatic fields, IEEE Transactions on Magnetics, 1991

[35] W.H. Press et al., Numerical Recipes in C: The Art of Scientific Computing, 2nd edition, Cambridge University Press, 1992

[36] Rischmüller, V., Fetzer, J., Haas, M., Kurz, S., Rucker, W.M.: Computational efficient BEM-FEM coupled analysis of 3D nonlinear eddy current problems using domain decomposition, In Proc. of the 8th International IGTE Symposium, Graz, Austria, September 1998.

[37] Russenschuck, S., et. al.: Integrated Design of Superconducting Accelerator Magnets - A case study of the main quadrupole, The European Physical Journal, Applied Physics, 1, 1998

[38] Y. Saad, M.H. Schultz: GMRES: A Generalized Minimal Residual Algorithm For Solving Nonsymmetric Linear Systems, SIAM Journal of Scientific Statistical Computing, 1986

[39] H.A. Van der Vorst: Bi-CGSTAB: a fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems, SIAM Journal of Scientific Statistical Computing, 1992

[40] The LHC study group, Large Hadron Collider, The accelerator project, CERN/AC/93-03

[41] The LHC study group, The Large Hadron Collider, Conceptual Design, CERN/AC/95-05

[42] Wilson, M.N.: Superconducting Magnets, Oxford Science Publications, 1983

[43] Wille, K.: Physik der Teilchenbeschleuniger und Synchrotronstahlungsquellen, Teubner, 1992

[44] Winslow, A.A.: Numerical solution of the quasi-linear Poisson equation in a non-uniform triangular mesh, Journal of computational physics, 1, 1971

[45] Wolf, R.: Persistent currents in LHC magnets, IEEE Transactions on Magnetics, 1992

FIELD MEASUREMENTS

L. Bottura, K.N. Henrichsen
CERN, Geneva, Switzerland

Abstract
The measurement of the magnetic field is often the final verification of the complex design and fabrication process of a magnetic system. In several cases, when seeking high accuracy, the measurement technique and its realization can result in a considerable effort. This note describes most used measurement techniques, such as Nuclear Magnetic Resonance, fluxmeters and Hall generators, and their typical range of application. In addition some of less commonly used techniques, such as magneto-optical, SQUIDs, or particle beams methods, are listed.

1. INTRODUCTION

Magnetic systems are designed today thanks to powerful computational tools that achieve remarkable accuracy and are indeed vital for proper optimization. However, the most straightforward way to verify that the required field strength, field shape and desired homogeneity have been achieved is still a direct magnetic measurement. In fact, when striving to reach accuracy in the magnetic field of the order of 100 ppm or better, as needed for example in accelerator and especially NMR magnets, the measurement of the magnetic field is the only way to guarantee that the target is reached. This is especially true when dealing with superconducting magnets, where material properties as well as the magnet geometry are not exactly known at the operation temperature. In addition the magnetic field can be strongly affected by the non-linear, diamagnetic behaviour of the superconducting material.

In addition to magnet technology, the measurement of magnetic fields is a key science in various technical areas. High sensitivity field measurements are used to detect ores underground, mines, submarines in the ocean, heart and brain activity, flux distribution inside superconductors, data stored on magnetic support. Directional field measurements are used to direct vehicles on magnetic tracks, help in the navigation by detecting minute changes in the earth magnetic field, monitor interplanetary activity. Wireless detection of a magnetic field is used to encode and control movements, identify proximity conditions and count items on production lines. Albeit impressive, this is only a much reduced sample of the technical applications requiring sensing and measuring the magnetic field.

This note describes principle and typical applications of most common magnetic measurement techniques and devices. Obviously, given the vastness of the domain, it is not thought as an exhaustive review on magnetic measurement methods. These methods are a typical topic of discussion in dedicated international journals [1], conferences [2,3,4] and workshops [5]. Additional information for researchers as well as new users entering this field can be found in [6] and [7]. In this note we focus mainly on the field measurement methods and devices that are best suited for superconducting magnets and materials, with special attention to operation at cryogenic temperatures, merely listing other interesting or promising methods for the sake of completeness.

2. MEASUREMENT TECHNIQUES

Various different methods are available for the measurement of the magnetic field. The best method for an application can be selected balancing the requirements of the field to be mapped to the typical:

- field measurement range,
- reproducibility and accuracy,

- mapped volume and field geometry,
- time bandwidth

of the method. In addition the measurement environment can impose stringent conditions such as temperature or geometric constraints that can largely drive the selection of the most adapted method. In the sections below we will review the most known field measurement techniques. We will briefly outline the basic principles of the measurement method, discuss the typical measurement techniques and classify the method in terms of range, accuracy, mapped volume and bandwidth.

2.1 Magnetic Resonance

2.1.1 Measurement principle

Magnetic resonance phenomena (Nuclear Magnetic Resonance, NMR, and Electron Paramagnetic Resonance, EPR[1]) provide today the most reliable standards for the measurement of a homogeneous magnetic field commonly achieving accuracy of 0.1 ppm and better in controlled conditions. For this reason NMR is considered today the primary standard for calibration. Nuclear magnetic resonance was first observed in molecular beams in 1938 by Rabi and co-workers [8]. A few years later, in 1946, the phenomenon was observed in liquids and solids by two independent teams [9,10,11].

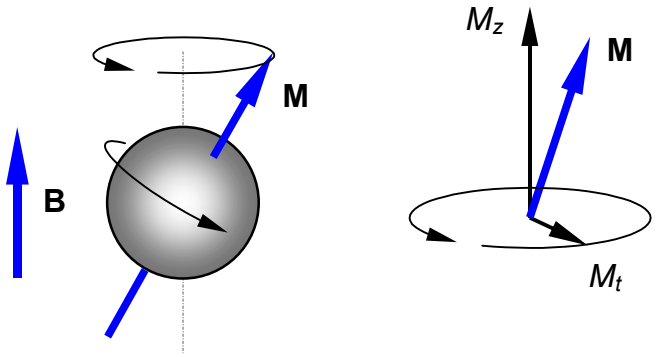


Figure 1. Schematic representation of a particle with magnetic moment in a background magnetic field. The particle precesses with a frequency f. The magnetic moment has a component M_z directed parallel to the external field B and a rotating component M_t in the plane normal to the field.

The principle of the measurement of a magnetic field using magnetic resonance can be intuitively explained starting with the observation that when a particle with a magnetic moment and angular momentum is placed in a background magnetic field with strength B it precesses around the direction of the field, see Fig. 1. The frequency f of the precession, also called Larmor frequency, is proportional to the magnetic field:

$$f = \gamma B \tag{1}$$

where the proportionality constant γ is called the gyromagnetic ratio. This last is a property of the particle that can be traced to standard values such as those reported in Table 1. As an example, for the proton the value of γ is 42.576396(3) MHz/T [12], implying that a proton in a 1 T field precesses at approximately 42.6 MHz.

In general the magnetic moment precession in a sample of many particles is not coherent, and the total magnetization of the sample is parallel to the applied field. It is however possible to use an RF pulse to

[1] Electron Paramagnetic Resonance (EPR) is also referred to in the literature as Electron Spin Resonance (ESR).

rotate the magnetization vectors and achieve coherence. This is the initial step of a typical spectroscopy experiment, when an RF pulse of proper frequency, intensity and direction flips the magnetic moments of the particles. After the end of the pulse the precessing magnetic moments loose coherence in a process called Free Induction Decay (FID). The typical time scale of the FID is determined by spin-lattice and spin-spin relaxation times, and is usually indicated as T_2*. The time constant T_2* can range from a few ms to several seconds, depending on the nature of the sample and specific factors such as the presence of field gradients.

Table 1
Values of the gyromagnetic ratio γ for most common particles and nuclei in magnetic resonance.

particle.	γ (MHz/T)
e^-	28.025 x 10^3
^1H	42.576396(3)
^2H	6.535
^{13}C	10.71
^{14}N	3.08
^{19}F	40.08
^{23}Na	11.27
^{27}Al	11.093
^{31}P	17.25

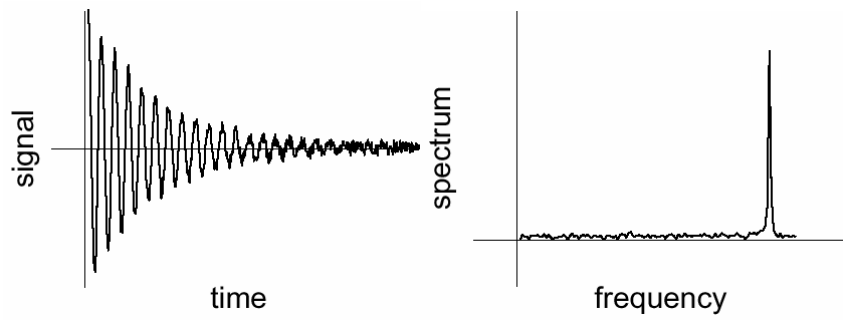


Figure 2. Schematic representation of the signal generated by the decay of the transverse magnetization M_t following an RF pulse (left). The periodic oscillation happening at the Larmor frequency is damped with a time constant T_2* that depends on the nature and state of the sample. A Fourier analysis, schematically shown on the right, gives the oscillation frequency *f* of the nuclear resonance.

Magnetic resonance is very sensitive to the local value of the magnetic field and therefore to field gradients in the sample. A change of the local field in the sample causes a change of the resonant frequency, following Eq. (1). This property is used to achieve encoding of the spatial position through known field gradients in magnetic resonance imaging. However for precise field measurement care must be taken that the field gradient within the resonating sample is kept below tight bounds.

The small field gradients naturally present in a sample made of nuclei bound in molecules are responsible for two other effects that result in a shift of the resonant frequency. Once the nuclei are bound in molecules the field at the nucleus can be affected by shielding (or deshielding) originating from the electron orbitals. This small change in the background field is visible in a shift of the resonating frequency typical of the molecule, called chemical shift. An additional split of the resonant frequency lines is visible when the nuclei are bound closely enough in the molecule. In this case the local field at a nucleus is affected by the alignment of the spin of the neighbouring nuclei. The resulting shift in the resonant frequency is then said to originate from spin-spin coupling. The chemical shift and spin-spin pattern of several molecules is known and can be used as a marker for

spectroscopy. The above effects are small, but become significant and must be taken into account when trying to calibrate the gyromagnetic ratio for a given particle or nucleus to ppb levels.

2.1.2 Field measurement with NMR and EPR

In NMR and EPR devices the measurement of the field is based on the detection and measurement of the precession frequency f of particles contained in a sample placed in the field to be measured. The corresponding field value is then deduced using the proportionality Eq. (1). In general the measurement of a frequency can be performed with very high accuracy, typically fractions of ppm, thus leading to the known accuracy of the method. Different techniques can be used for the measurement of the precession frequency [13]:

Continuous Wave (Q-meter). The sample is inside a coil that is a part of an L-C system tuned on a frequency range f_{min} to f_{max} that determines the range of field B_{min} to B_{max} that can be measured with the probe. The circuit, shown schematically in Fig. 3, is excited by an RF signal, oriented normal to the background field B_0 to be measured. The RF signal initially scans the range of frequency of the probe. When the RF excitation frequency f_{RF} is equal to the precession frequency f_0 for the particle in the external field the L-C circuit absorbs part of the RF signal. This *resonant* condition can be detected by measuring the decrease of the transmission of the RF signal. The corresponding excitation frequency gives then a measurement of the external field through the knowledge of the gyromagnetic ratio. To improve the measurement of the resonant frequency a small, symmetric modulation field ΔB can be added to the background field. The modulation field causes the precession frequency to shift continuously in the range f_0-$\gamma\Delta B$ to f_0+$\gamma\Delta B$. The resonance condition is used in this case to generate a pulse freezing the voltage applied to the field modulation coil. The average voltage during one modulation period is then used as feed-back signal to change the RF frequency. With this expedient the RF excitation f_{RF} is *latched* onto the precession frequency f_0 and thus tracks the external field at all times.

Impedance (Z-meter). Instead of tracking the precession frequency, it is possible to measure a change in the impedance of the sample at the frequency f_0. The sample is in this case a part of a balanced bridge that must include the line between the instrument and the sample itself. At the precession frequency the impedance changes due to the resonance. The change in impedance generates an error signal proportional to the change of phase. Similarly to the continuous wave method, the VCO modifies the oscillation frequency and can thus lock onto the field. As this detection can be faster, this allows faster tracking of a changing field. In addition it can work at higher frequency, in the GHz range, allowing the measurement of higher field strength.

Pulsed Magnetic Resonance. In this case a single RF pulse induces coherence in the magnetic moments. The FID of the sample is measured following the RF pulse and is Fourier analysed to determine the resonance frequency, see Fig. 2. Given the long times available for detection (T_2* can be several seconds) the measurement of the frequency can be very precise in the case of a sample with a single resonant frequency. The interest of this method is that it in addition it can detect resonance at different frequencies as it is the case for particles bound in molecules. In this case the precession frequency is modified by the presence of other particles. For this reason pulsed NMR techniques are commonly used in spectroscopy and imaging techniques.

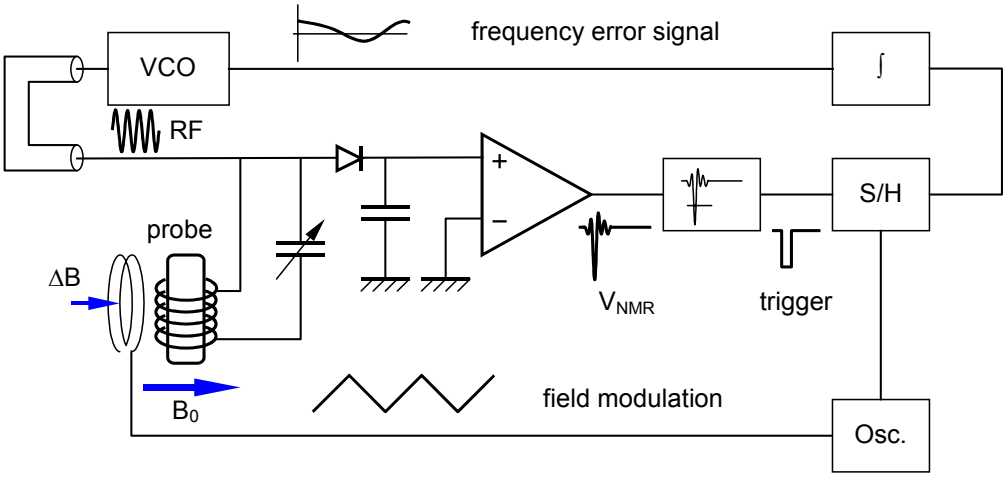


Figure 3. Principle of resonant frequency measurement using the continuous wave (Q-meter). The probe is placed in the background field B_0 and is equipped with an RF excitation coil and a modulation coil that provides the necessary field modulation ΔB for latching the resonance frequency [13].

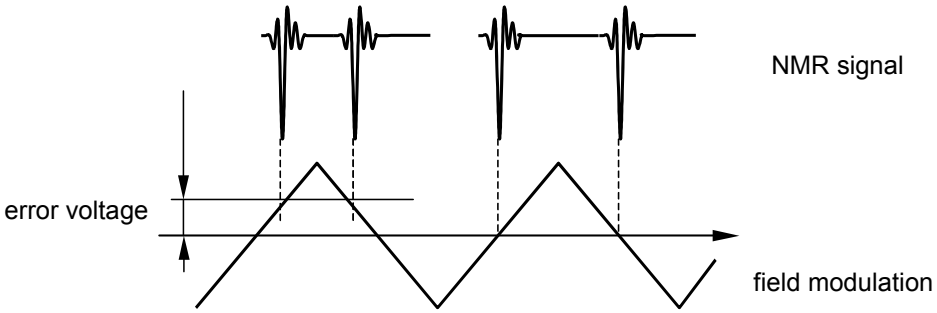


Figure 4. Schematic representation of the latching principle. The resonance peaks detected during field modulation generate an error voltage proportional to the distance of the peaks from the zero-crossing of the field modulation. This error voltage is used to generate a negative feed-back on the RF excitation frequency [13].

Other techniques for special applications. The continuous wave and impedance methods discussed above are the most common techniques for the measurement of uniform magnetic field. In addition, pulsed NMR measurements have been practiced for field mapping in accelerator magnets using a multiple set of samples placed on a circle with the diameter of the bore to be mapped [14]. Probes made from ^{3}He and ^{27}Al have been used even at cryogenic temperatures [15]. The equipment for this type of measurements is at the prototype level and not yet commercially available. Finally, it should be mentioned that a rather exotic method of NMR measurements is to detect the spin resonance in water flowing in laminar conditions in a small tube [16,17,18]. This method has given remarkably good results at low fields for which NMR equipment is not yet commercially available. In addition, it is possible through this technique to perform measurements in environments of strong ionizing radiation such as in particle accelerators. The technique was tested for measurements in the bending magnets installed in LEP, achieving a resolution of 0.1 μT in the range from 0.5 mT up to 112.5 mT [19]. The operation of this type of equipment is rather complicated due to the relatively long time delays in the measurement process.

2.1.3 Range and accuracy

The range of measurement by magnetic resonance is determined by the range of practical frequencies for the electronics of the instrument. Using common RF technology it is possible to detect efficiently a resonance in the 1 to 100 MHz range. Choosing the appropriate material for the probe, the minimum

field that is measured can be as low as fractions of mT (e.g. 0.5 mT using an EPR probe) up to more than 10 T (e.g. 14 T using a deuterium NMR probe). The measurement accuracy depends mostly on the method used to detect the precession frequency and the time taken for the measurement. As a general rule, fast measurements are affected by larger uncertainty on the measured frequency. Continuous wave measurements can reach 0.5 ppm accuracy for 1 s measurement time, and 0.1 ppm for 10 s measurement time. In pulsed NMR the uncertainty on the frequency measured can reach a few ppb, given the long relaxation time available for the data acquisition. A possible source of error is the standard for the gyromagnetic ratio. Although γ is constant in a broad range of temperature, care should be taken to adopt the value that corresponds to the actual state of the particle in the sample. As an example the official value of γ for protons, given by NIST with an uncertainty of 70 ppb, is affected by a 3-ppm shift once the proton is measured in rubber. Once calibration is achieved, however, the knowledge of the value of the gyromagnetic ratio is affected by an error that is much smaller than the resolution of the frequency measurement.

The typical volume of a NMR probe is of the order of 10 to 100 mm^3. The lower limit is set by signal intensity and signal-to-noise requirements and practical manufacturing considerations, while to upper limit comes from the necessity to have a homogeneous field within the probe. One of the most known properties of field measurement through magnetic resonance is that the field needs to be constant in time and uniform through the sample. The effect of a field gradient through the sample is to broaden and lower the resonance peak in the frequency response. This makes on one side latching to the resonant frequency more difficult, while on the other side precision in the determination of the frequency is inevitably lost. The maximum field gradient that can be dealt with depends on the size and material used in the probe and on its range. The limit on the field gradient for a typical probe size with linear dimensions of a few mm is between 10 and 100 ppm/mm. Magnetic fields with higher gradients can still be measured adding small compensation coils. The compensation coils are designed to generate a gradient opposite to the one present in the field to be measured. They are placed around the probe, with the probe centered in the point of zero field. Adjusting the current flowing through the gradient coil it is possible to *shim* the field to be measured thus achieving the desired improvement of the resonance signal. The frequency scan and latching, such as used in the Q-meter technique described above, takes place slowly, typically a few Hz. The corresponding tracking speed is limited in the range of 1 % relative variation rate of field per second. Open loop techniques can be used for faster measurements [13].

2.2 Fluxmeter

2.2.1 Principle

The fluxmeter principle is based on the magnetic induction law in its integral form. If we take the ideal coil of Fig. 5 consisting of a filamentary winding with contour Γ the magnetic flux φ linked with the coil is given by:

$$\varphi = \int_S \mathbf{B} d\mathbf{S} \tag{2}$$

where **B** is the magnetic field, S is an arbitrary surface bound by the contour Γ and d**S** is the normal to the surface. A variation in time of the flux linked by the coil induces a loop voltage proportional to the rate of change, or:

$$V = -\frac{d\varphi}{dt} \tag{3}.$$

A fluxmeter detects this voltage and uses it to measures the flux variation. The magnetic field in the area enclosed by the coil can then be deduced from the flux measurement.

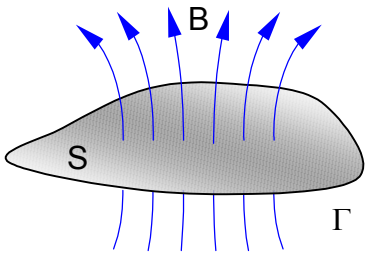


Figure5. Flux linkage with an ideal, filamentary induction coil. The coil contour Γ defines a surface **S**. The flux is defined as the integral of the component of the magnetic field **B** normal to the surface **S**.

The fluxmeter is probably the oldest of the field measurement techniques, and was employed by Weber in the middle of the 19th century to study the direction of the Earth magnetic field [20]. It is a relatively simple method, highly linear, and only requires an induction coil and a voltage measurement. In addition the fluxmeter method provides a very accurate measurement of the field direction, which in some cases (e.g. accelerator magnets) is of greatest importance. As shown in Eqs. (2) and (3) a voltage can be induced on the coil either by a variation of the magnetic field d**B**/dt, or by a change in the orientation of the coil d**S**/dt. In the case of a magnetic field variation in time the coil is usually kept static, and the fluxmeter provides a measurement of the flux change between two given instants. The second technique, i.e. a movement of the coil, is usually employed in static magnetic fields. In this case the field measurement depends, among others, on the precision of the movement. For this reason the most commonly used movements used are simple translations, 180 degrees flip or continuous rotation around a single axis. Fixed and moving coils have been routinely used also for cryogenic measurements.

2.2.2 Measurement of the Magnetic Flux

The measurement of the change of the magnetic flux in a fluxmeter requires forcibly an integration of Eq. (3):

$$\varphi_{end} - \varphi_{start} = -\int_{t_{start}}^{t_{end}} V\,dt \qquad (4).$$

where t_{start} and t_{end} are the time at the start and at the end of the measurement, φ_{start} and φ_{end} are the corresponding values of the magnetic flux. In practice the integration of the coil voltage can be performed using several different techniques:

Analog Integrator. The electronic, analog integrator makes use of the integrating property of an RC circuit. It was the first integrator achieving high speed and good accuracy, and remains today one inexpensive but effective way to measure flux changes. The principle of the electronic integrator, also called Miller integrator, is shown in Fig. 6. For a circuit with ideal characteristics, the signal V_{out} at the output of the DC amplifier is proportional to the integral of the input signal V_{in}. The proportionality constant is $1/RC$:

$$V_{out} = -\frac{1}{RC}\int_{-\infty}^{t} V_{in}\,dt \qquad (5).$$

Connecting an induction coil to the input, and a voltmeter at the output of the circuit it is thus possible to obtain the flux change.

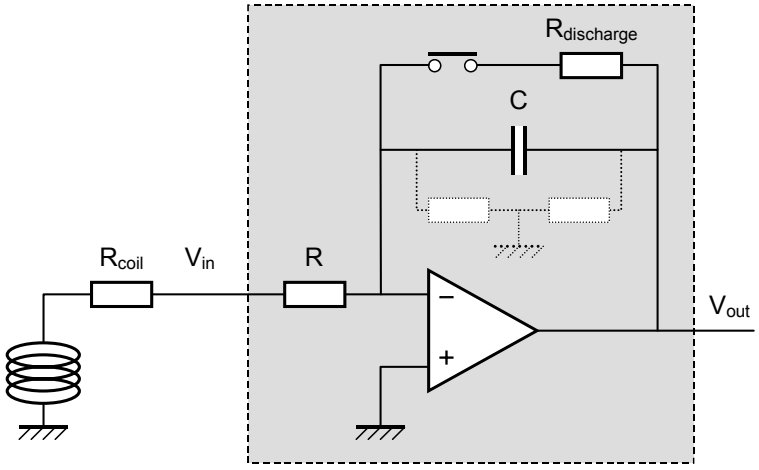


Figure 6. Principle of an electronic analog integrator for the flux measurement. The output signal V_{out} is proportional to the integral of the input signal V_{in}.. Schematically shown the switch and discharge resistor needed to reset the integrator and the parasite resistance responsible for the current leakage between the capacitor plates and to ground.

The most critical element in the analog integrator is the capacitor and in particular the change of its properties with temperature as well as the leakage current. Thermal effects can be controlled through temperature stabilization of the electronics, where the typical level required is better than 0.1 °C. Careful protection and shielding of the capacitor is essential to reduce the voltages across the surface resistances. Another parasitic but important effect is the finite resistance of the coil R_{coil} that acts as an additional load for the integrator, thus changing the proportionality constant between input voltage and output integral. Provided that the effects above are minimised, the limit to the precision achievable with the integrator is then set by the dielectric absorption of the capacitor.

Digital Integrator. A very precise measurement of the flux change can be obtained using hybrid, digital integrators based on the Voltage-to-Frequency Conversion (VFC). The principle of this method, originally developed at CERN, is shown in Fig. 7.

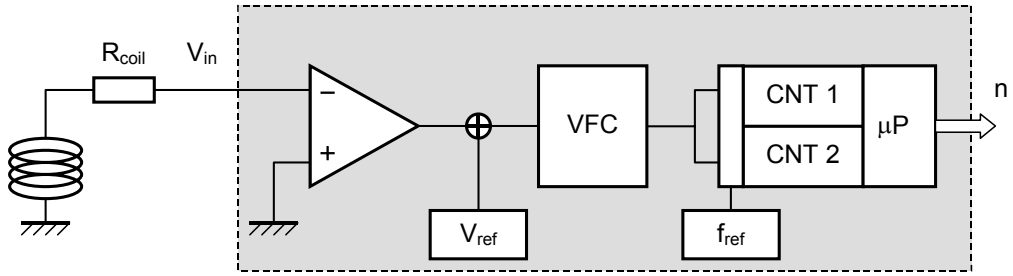


Figure 7. Principle of a digital integrator based on voltage-to-frequency conversion. The digital output n generally available on a computer bus, is proportional to the integral of the input signal V_{in}.

The voltage from the induction coil, after proper conditioning, is sent to a VFC whose output is an AC signal with frequency f proportional to the input voltage V_{in}. The AC signal is then entered in a counter that accumulates the number n of AC pulses during a measurement period starting at t_{start} and ending at t_{end}. The frequency f of the AC signal is equal by definition to the time derivative of the number of

pulses dn/dt, and the output of the counter is, apart for a proportionality constant K_{VFC}, a digital measurement of the integral of the input voltage:

$$n = \int_{t_{start}}^{t_{end}} f dt = K_{VFC} \int_{t_{start}}^{t_{end}} V_{in} dt \qquad (6).$$

The digital integrator achieves high precision thanks to the conversion to frequency domain. The limiting elements in this concept are the stability and linearity of the VFC, and the resolution of counting operation that depends on the maximum operation frequency of the VFC. Hybrid technology VFC's have linearity and stability of better than a few ppm over the whole range of input voltage. The typical maximum frequency of operation is 1 MHz. In order to make the circuit practical some additional features are added to the basic scheme described above. Commercial VFC circuits work only with single polarity voltage, e.g. 0 to 10 V, while the signal from an induction coil can have both polarities. The best way to restore the dual polarity capability is by shifting the input voltage by a precise and stable reference whose effect is to place the input zero exactly in the middle of the VFC range. This offset is then eliminated after counting, subtracting the counts from a reference AC source oscillating at exactly half of the maximum frequency of the VFC. Another technical detail that allows to avoid dead times during the transfer of the result from the buffer of the counter to the downstream circuitry is to use two parallel counters working in alternance. This technique is very effective and results in the cancellation of cumulative errors.

Numerical Integration. Another possibility to integrate the voltage signal from an induction coil is to measure at sufficiently high rate the coil voltage and resort afterwards to numerical techniques. This is possible nowadays thanks to precision Digital Volt-Meters (DVM) or Analog-to-Digital Converters (ADC) commercially available. The principle of a DVM or ADC is to sample the input voltage at a given rate, defined by an electronic trigger. The voltage sample V_i, together with the information on the sampling time t_i and interval Δt_i, can be used to perform numerical integration and thus obtain the flux change at any time during sampling.

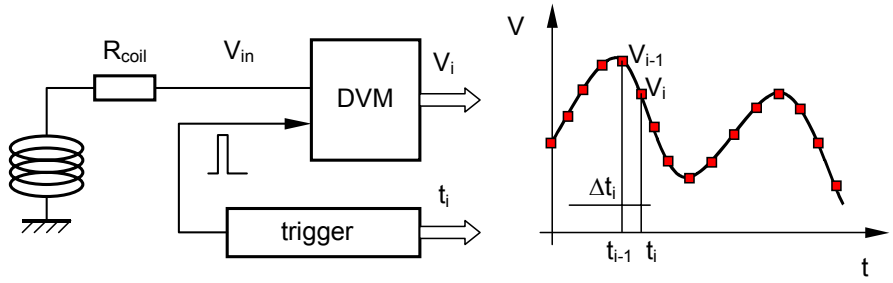


Figure 8. Principle of integration based on digital voltage measurement, showing schematically an approximation for the voltage readings. The DVM delivers instantaneous voltage measurements V_i at precisely defined times t_i. This information is collected by a computer that performs the numerical integration.

The technique is extremely powerful as it can make use of advanced numerical techniques for integration and interpolation. It should be remarked, however, that it generally requires higher sampling rate than analog or digital integration to resolve the fine structure of the voltage waveform in presence of parasitic perturbations or any significant noise. In addition it requires the precise knowledge of the time interval between samples. This can be affected by unknown dead times during data transfer from the DVM or ADC to the following electronics. As for digital integrators it is possible to obviate this last problem by using two voltage acquisitions in parallel and working in alternance.

2.2.3 Induction Coils

The sensor used in a fluxmeter is a stationary or moving pick-up coil. Coils are wound using an insulated conductor of small diameter on non-magnetic and non-conducting support. Suitable supports are composite materials such as fiberglass reinforced epoxy, or amorphous materials such as glasses and ceramics. Mechanical stiffness and low thermal expansion of the support are important requisites to achieve good and stable coil calibration. Similarly the coil winding technique, including the application of proper tension to the wire, is a key point for stable and reproducible results. Finally, good control of the winding geometry is important to calibrate the coil sensitivity to higher order harmonics of the field. High winding uniformity can be achieved using multi-filamentary wire or photo-lithographic techniques [21].

We consider the ideal case of an induction coil of cross section S obtained from N turns tightly wound. The winding can be treated as filamentary if its dimensions are negligible compared to the coil size. If this ideal coil is placed in a uniform magnetic field B oriented normal to the cross section the flux linked with the coil is:

$$\varphi = NSB = \kappa B \tag{7}$$

where we indicate with κ the coil sensitivity that depends on the coil geometry and the number of turns. The number of turns is selected to achieve a sufficient flux linkage to bring the acquisition electronics (integrators or voltmeters) in the useful working range. The flux reading obtained with one of the methods outlined above can be readily converted into field once the coil sensitivity κ is known. In practice the flux linkage depends not only on the field strength and the coil cross-section and number of turns, but also on the shape of the field lines as well as on the coil winding geometry. Different coil geometries can be used depending on the shape of the field to be measured. Often the winding is designed to compensate for field gradients and thus obtain a good measurement of the average field strength in a point or in a cross section. In several cases, however, the coil is specifically built for the measurement of field gradients or higher order harmonics. In all cases the coil sensitivity κ must be determined through an independent calibration procedure.

Coils can be broadly classified based on their form and extension in space as follows [21]:

- point coils, dedicated to the measurement of the magnetic field at a small point in space, ideally infinitesimal;
- line and area coils, used for measurement of integrated field along a line or on a flat region of space;
- harmonic coils, obtained in general as assembly of line or area coils, and designed to be sensitive only to particular spatial harmonics of the field.

Point coil. The simplest point coil is a cylindrical coil wound around a small core. Because of symmetry a cylindrical coil is insensitive to field gradients, and in general to all even field harmonics. Choosing appropriately the ratio of coil thickness to height, as shown in Fig. 9, it is also possible to wind cylindrical coils insensitive to the third spatial harmonic [22]. If the field to be sampled is nearly homogeneous the point coil measures the average magnetic field in the coil volume and can be considered as a close approximation to the ideal case of Eq. (8). A special type of point coil is the fluxball [23]. The fluxball is a coil with spherical shape, wound with a uniform density of turns per unit length along its axis and an increasing density of turns in radial direction, see Fig. 10. This configuration is sensitive to the average field in the sphere and insensitive to all higher order spatial harmonics and can thus be considered as the closest approximation to an ideal point coil. The winding configuration of the ideal fluxball can be approximated for practical purposes using sets of concentric solenoids with appropriate dimensions and turns.

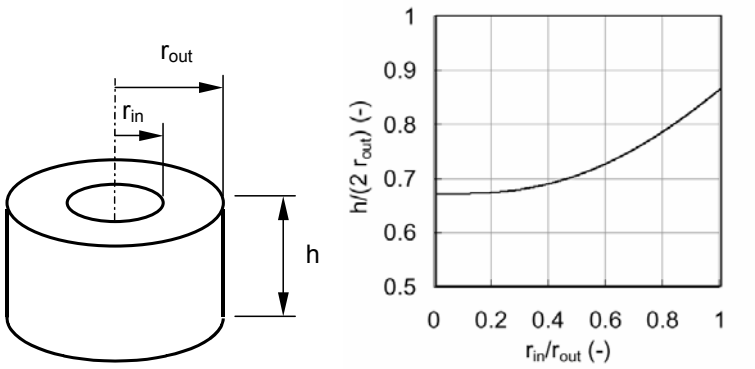


Figure 9. Schematic representation of a cylindrical point coil. The plot shows the ratio of inner to outer radius that cancels the third spatial harmonic of the field, plotted as a function of the ratio of coil height to coil external diameter (from [22]).

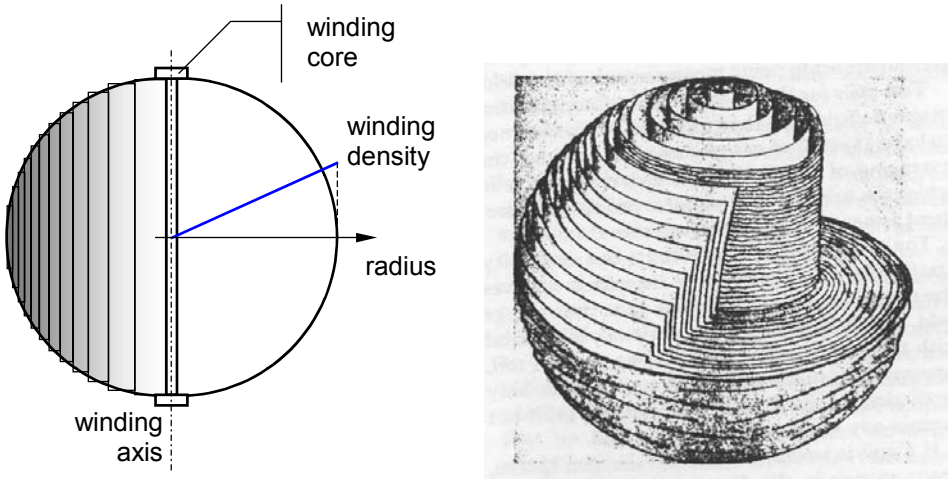


Figure 10. Cross section of an ideal fluxball. The winding density is uniform along the coil axis (vertical) and increases proportionally to the radius across the winding (horizontal). This ideal configuration can be reproduced by proper arrangement of concentric solenoids, as suggested in the picture. Also shown a sketch of its original construction, reproduced from [23].

Line and area coils. Line coils are designed to provide the integrated value of the field along a straight or curved line. Their width is much smaller than the coil length. As for point coils if the winding has a symmetric, rectangular cross section then the flux linked with the coil does not depend on the even harmonics of the field. In addition, by properly selecting the ratio of winding thickness and height the coil can be made insensitive to the third spatial harmonic [24,21]. Any coil that has comparable winding width and length or height can be classified as area coil. Area coils have the same properties of line coils, but they sample a large area of space. Line and area coils of long length and slender construction are routinely used for the measurement of the field in accelerator magnets. The reason is that in spite of their considerable length (e.g. 15 m for the LHC dipoles) the effect of an accelerator magnet on the particle beam, whose betatron oscillation happens on a much longer scale, can be assimilated to a localised optical element of equivalent strength and negligible dimension. Under this condition it is enough to know the integrated field strength and homogeneity of the magnet under examination. Long rectangular coils, translated horizontally during the measurement, were usually employed and are still used in magnets with a wide horizontal aperture and limited gap height. Similar coil geometries are used for rotating coils measurements in the narrow, circular bore of superconducting accelerator magnets. A few typical examples are shown in Fig. 11.

Figure 11. Examples of typical induction coils used for accelerator measurements, by courtesy of J. Billan (CERN). The coils have length ranging from 25 cm to more than 2 m much larger than its extension in the cross section (10 mm to 50 mm), and windings of negligible dimensions with respect to the overall size of the coil (typically 0.5 x 0.5 mm).

Harmonic coil. Shaping the coil winding, or interconnecting coils suitably placed in space, it is possible to achieve sensitivity only to a selected number of spatial harmonics of the field. We call such a winding or set of windings a harmonic coil. Harmonic coils are a particularly useful expedient when investigating the field homogeneity. Indeed the sensitivity to field errors can be largely increased if the main field component is cancelled using a suitable array of coils. A simple example of a coil only sensitive to even harmonics is a *gradient* coil, consisting of a set of two identical windings mechanically mounted either side by side or on top of each other and wound in opposite direction. The total signal from the gradient coil is insensitive to the average field, as the signal produced by the two windings is identical in amplitude and opposite in sign. A field gradient, instead, produces signals of identical amplitude and sign on both windings thus resulting in a net output. Another example of harmonic coil that has been used in the measurement of the two-dimensional field of accelerator magnets is the Morgan coil [25]. The coil winding in a Morgan coil is subdivided in $2M$ blocks that are placed with uniform angular spacing on a cylinder. The contiguous blocks have opposite winding direction and form M ideal loops connected in series. This configuration is only sensitive to harmonics of order $(2k+1)M$, with k integer. An example of the cross section of a Morgan coil sensitive to the octupole harmonic B_4 (fourth harmonic $M=4$) is shown schematically in Fig. 13.

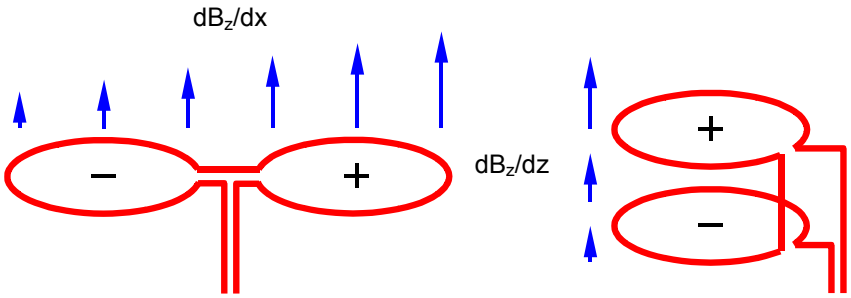


Figure 12. Schematic representation of different coils sensitive to a field gradient in the vertical component of the magnetic field B_z. Gradient coils are used to measure the second spatial harmonic of the field.

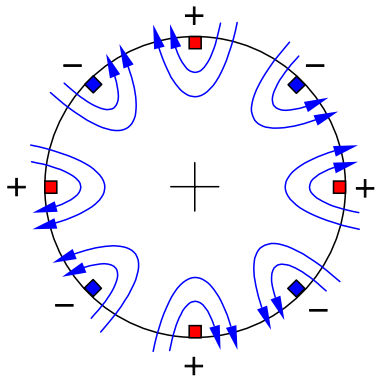


Figure 13. Cross-section of a Morgan coil sensitive to the fourth harmonic of the field (M=4, octupole term B_4). The winding direction enters or exits the plane of the page alternatively, according to the signs indicated. Also shown schematically the field lines for an octupolar field.

2.2.4 Coil calibration

One of the main sources of error in the fluxmeter is the uncertainty on the sensitivity of the coil to the field measured. Because of tolerances and imperfections during coil manufacturing the real winding shape has inevitable deviations from the ideal one. The solution is to calibrate the coil in reference magnets using a field reference such as an NMR probe. In addition to the equivalent surface of the coil it is often necessary to measure the orientation of its median plane with respect to mechanical references on the coil support. This is a mandatory calibration step if the coil has to be used for the measurement of field direction. Gradient coils are calibrated in magnets with known gradients or alternatively using controlled displacements in an unknown gradient magnet. Erratic signals from wire loops must be avoided during the calibration process as well as during measurements. In case that the field to be mapped has strong variations along the length of the coil, it may be necessary in addition to examine the change of the effective width. Temperature variations and long term geometric stability can affect the quality of the calibration that may thus require periodic checking. For cryogenic applications, and especially at liquid helium temperature (4.2 K), it is necessary to take into account the large thermal contraction from room temperature to the measurement conditions (several 1000 ppm in terms of coil surface). This is usually done either using empirical contraction coefficients or calibrating at liquid nitrogen (77 K) and neglecting the residual thermal contraction below 77 K. Precise calibration at cryogenic temperature is for obvious reasons more difficult to achieve and leads to an overall decrease in the expected accuracy.

2.2.5 Field measurement with fluxmeters

As we anticipated, the field measurement using a fluxmeter is generally performed either keeping the induction coil fixed, in case of changing field, or moving the coil in a static field. In the case of fixed coil measurements, the integrator is reset at the start of the measurement (e.g. discharging the capacitor in the analog integrator). In the simplest case the integration starts and continues till the end of the measurement, when the flux change is finally readout. The flux change is then converted to the desired value of field change using the known coil calibration, e.g. the surface for a uniform field as in Eq. (8). Note that according to Eq. (4) a fixed coil measurement can only give the flux *change* between the start and the end of the measurement and is therefore an intrinsically relative measurement. Indeed, an absolute measurement is possible only in the case that the magnetic field at the start is known to be zero. In practice the integrator can be readout or electronically triggered to deliver the flux change at any desired time during the measurement period, thus providing the waveform of the field change. For fixed coil measurements it is of primary importance to reduce the voltage offset at the input of the integrator generated by parasitic currents in the electronics or thermal voltages. An input voltage offset translates in an apparent field drift that often cannot be distinguished from the physical field change. This effect becomes important for long integration times, normally a few seconds to minutes for state-of-the art technology. Offset trimming must be performed electronically in well-controlled steady-state conditions (i.e. when the induced voltage on the coil is zero) and is often supplemented by digital reduction of residual drift. The latter is possible if the measurement spans phases when the field is constant, and the input voltage of the integrator is thus the residual offset, in addition to the period of changing field, when the input voltage contains both contributions from induced voltage and offset.

When the field is constant the necessary induction signal is generated moving the induction coils. The integrator is reset before the movement and is readout at the end of the movement. Because the magnetic field is solenoidal, the final reading depends only on the initial and final positions of the coil and does not depend on the path taken for the movement. Starting with the coil outside the magnet (e.g. in a zero-Gauss chamber) and finishing with the coil in the magnet bore, the initial field is known to be zero so that the total flux change can be used to deduce the absolute value of the field. This *moving coil* technique has been successfully used to measure field strength in accelerator dipole magnets taking care that the final position of the coil is normal to the field. An alternative technique consists in *flipping* the coil, rotating it by 180 degrees during the measurement. In this case the flux change during the measurement is equal to twice the flux linked with the coil in the initial position. Taking again the case of a homogeneous, dipole field as an example, and having placed the coil initially normal to the field, it is possible to use the flux change to deduce the absolute field strength. As in the case of fixed coil measurements, both moving and flip coil techniques are sensitive to the adjustment of the voltage offset. In these cases however the initial and final states are by nature stationary and it is easier to identify residual drifts. In addition first-order drift correction can be obtained by proper combination of movements including direction reversals, thus exploiting physical symmetries. Finally in both cases the resolution can be greatly increased by using differential measurements, where a pair of coil is connected in opposition, with one coil moving and the other stationary. This expedient can be used to compensate fluctuations in the magnet excitation current and to provide higher sensitivity when examining field quality.

An alternative method based on induction measurement is the *stretched wire* technique [26,27] that can be regarded as an extension of the moving coil. In its simplest variant a thin wire of high strength, non-magnetic material (e.g. W or CuBe) is stretched inside the bore of the magnet. A second wire, laid fixed either inside the bore or outside the magnet, provides the return path for this single-turn coil. The stretched wire is moved in the magnetic field using precision stages and the signal generated by the flux cut during the movement is integrated as in the case of a moving coil. The accuracy in this case is mainly determined by the precision of positioning. Sensitivity and resolution can be increased using

multi-wire arrays. This method is well suited to geometry measurements, to the absolute calibration of quadrupole fields and in particular to measurements in strong magnets with very small aperture.

Rotating coils is finally one of the most successful methods in the measurement of field and field quality for accelerator magnets. Devised since 1954 [28,29] the method has become now widely used for magnets with cylindrical bore. The induction coil is placed on a circular support and is rotated in the field to be mapped. The coil angular position is measured by an angular encoder, rigidly connected to the rotating support. As the coil rotates in the field it cuts the flux lines and a voltage is induced at the terminals. Integrating the induced voltage between predefined angles using one of the techniques described above it is hence possible to obtain the flux change as a function of angular position, as shown schematically in Fig. 14. If the field measured is 2-D in the cross section of the magnet measured, with negligible variation along the magnet length, it can be shown [30] that a Fourier analysis of the angular dependence of the measured flux leads naturally to coefficients directly proportional to the so-called multipole coefficients of the field [31]. In turn the multipole coefficients of the field can be related directly to linear and non-linear accelerator beam properties, thus explaining the wide acceptance of the rotating coil method for mapping accelerator magnets.

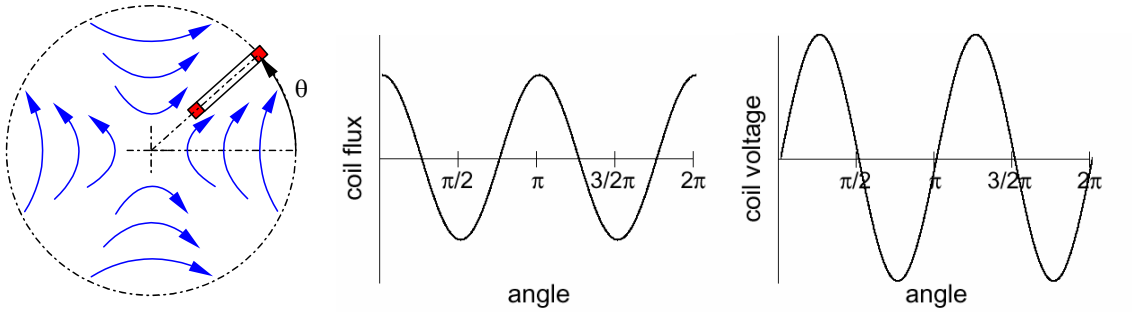


Figure 14. Principle of the rotating coil method showing an induction coil rotating in a quadrupole field and the corresponding flux and induced voltage plotted as a function of angular position θ.

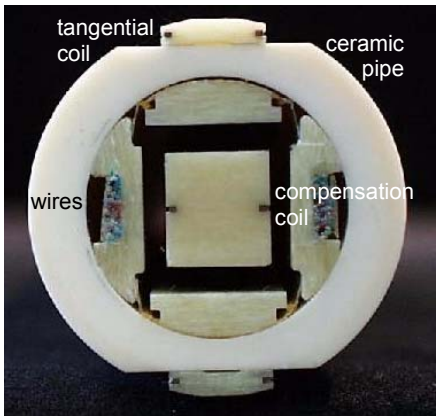


Figure 15. Cross section of the rotating coil used at CERN for the measurement of the 15-m long superconducting LHC dipoles. The upper and lower coils, mounted on a hollow ceramic haft, are used for the measurement of the main field strength and direction. The signal from the central coil is used to compensate analogically the contribution of the dipole component without affecting higher order harmonics.

In order to facilitate the realisation of a practical rotating coil system the angular encoder can be used to trigger the integrator so that integrator delivers the flux change between adjacent angular positions. This method eliminates the time dependence and in particular the influence of variations of the rotation speed, greatly relaxing requirements for uniform rotation. Differential measurements are also greatly beneficial to increase the resolution of high-order multipoles, generally several orders of magnitude smaller than the main field in an accelerator magnet. This is realised using a set of

compensation coils mounted on the rotation support. The signal from the compensation coils is used to suppress the strong contribution from the main field, either analogically or digitally. The residual signal, usually called *compensated* or *bucked* signal, can be analysed in Fourier series to obtain the higher order multipoles. A cross section of a typical rotating coil is reported in Fig. 15, showing the arrangement of main and compensation coils on the stiff ceramic support. The signal from the rotating coils is transmitted either through slip-rings or through flexible cables to the integrator. In the first case the coil can rotate continuously in one direction, while in the second case it must be returned to the start position. Using this return rotation for a *backward* measurement, and averaging its result to the one obtained during the *forward* rotation it is possible to draw advantage from this "*washing-machine"* operation to reduce systematic errors such as those due to mechanical torsion.

2.2.6 Range and accuracy

An induction coil is a linear device, whose sensitivity can be designed to fit the field to be measured. Hence the fluxmeter is a method potentially covering a very broad range of field measurement. Its main limitation originates from thermal and environmental noise, electronic drifts and the quality of the mechanics in the case of moving or rotating coils. In practice the application of the induction method is limited to minimum field levels above 0.1 μT, with no fundamental limitation at the upper end.

The first error source in a fluxmeter is the coil sensitivity. Coil sensitivity to a uniform field, i.e. the effective surface, can be calibrated to typical accuracy ranging from 10 to 100 ppm, depending on the size and shape of the coil. Gradient coils can be calibrated to better than 1000 ppm. The accuracy of integration depends strongly on the integration technique used. Analog integrators reach accuracy of 100 ppm and resolution of a few ppm. The resolution is generally limited by the noise and offset of the input amplifier. Digital integrators can reach linearity of better than 10 ppm and overall accuracy of 100 ppm. The resolution of digital integrators is affected by noise and offset of the input amplifier and in addition depends on the discretization noise of the counters. This last usually imposes the most severe limit, resulting in a resolution proportional to the clock frequency and to the inverse of the integration time. As an example, a digital integrator with a 1 MHz clock and integrating for 1 s has a resolution of 0.25 ppm. If the integration time is reduced to 1 ms the resolution decreases to 250 ppm. For DVM's and ADC's the accuracy of the reading can be very high, of the order of 1 ppm and better, and the digital stage is usually sized not to limit the resolution of the reading. In this case the largest error source comes from the jitter in the integration time intervals, from uncorrected dead time during reading and from the numerical approximation of the voltage integral. Overall accuracy of numerical integration methods is then comparable to the performance of digital integrators, in the 100 ppm range.

The volume sampled by induction coils can also be vastly different. Coils can be practically wound around a 1 mm^3 core to sample point regions. At the other extreme rotating coils with diameter in excess of 1 m have been used to sample a few m^3 of space. The time response of the fluxmeter depends on the bandwidth of the electronics reading and integrating the coil signal. In the case of analog integrators and numerical integration the voltage readout can be fast, up to 50 kHz For digital integrators the balance among integration time and resolution usually limits the acquisition frequency to the range of 100 Hz.

2.3 Hall generator

2.3.1 Principle

Hall generators measure the magnetic field through its effect on the free path of charge carriers in a conducting or semi-conducting material. In a Hall generator an electric current flows in a small section of a suitable material placed in the region where the magnetic field **B** must be mapped. For simplicity

we assume here that the material is a thin slab and we consider only the magnetic field component $B\cos(\theta)$ normal to the wide face of the slab, as shown in Fig. 16. The charge carriers moving in the material are subjected to a force **F** transverse to the direction of the electric current, with a magnitude:

$$F = qvB\cos(\theta) \tag{8}$$

where q is the charge of each carrier and v the average drift velocity of the carrier. This force tends to curve the direction of flow of the charge carrier and causes the polarization of the material in the direction of **F**. The resulting electric field, whose effect at equilibrium balances the transverse force from the applied magnetic field, is the Hall voltage. The magnitude of the Hall voltage V_H is proportional to the magnetic field component normal to the surface of the generator and to the electric current I exciting the Hall generator:

$$V_H = GR_H IB\cos(\theta) \tag{9}$$

where θ is the angle between the direction of the field and the normal to the generator, the factor R_H is the Hall coefficient characterizing the material used in the generator, and G is a geometric factor depending on the size and shape of the generator itself. It is interesting to remark that, in accordance with Eq. (9), the Hall generator provides a steady-state signal proportional to the magnetic field.

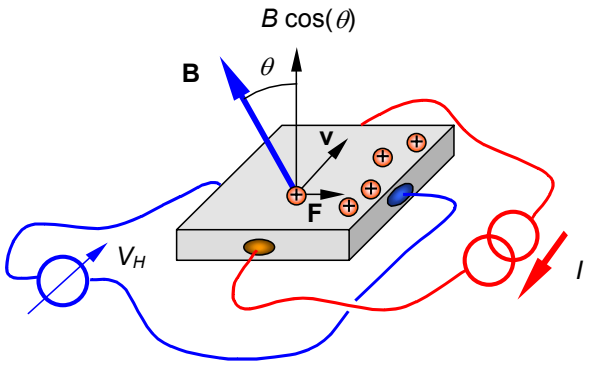


Figure 16. Schematic representation of the force acting on majority charge carriers (assumed to be holes in the example considered) in an ideal Hall generator and appearance of the Hall voltage.

The Hall effect [32] was firstly observed in metals, such as iron, gold and silver. The Hall coefficient of these materials, the ratio between transverse voltage and applied magnetic field, is relatively small. In an ideal Hall generator the Hall coefficient is proportional to the mobility of the charge carriers in the material, and inversely proportional to its conductivity. Because metals exhibit low mobility and high conductivity they are not suitable for efficient Hall generators, and in principle low conductivity materials should be preferred. On the other hand, and especially for temperature stability reasons, it would be preferable to use materials with a small resistance to the flow of the control current. It was only with the use of semiconductors [33] that the right balance could be achieved among mobility and conductivity and Hall generators became practical for field measurement. Nowadays Hall generators are built using either un-doped semiconductors such as Si or compounds of the type $A^{III}B^{V}$, i.e. formed by elements of groups III and V of the periodic table, such as InSb, InAs and GaAs.

2.3.2 Non-linearity and parasitic effects

Practical Hall generators are fabricated using the same technology applied for the production of thin semiconductors, often directly within integrated circuits that provide sturdy packaging and easy connection through soldering pads. Because of their finite, small dimension the flow of current within the material is much more complex than the simple picture discussed above. The terminals connection

act as discontinuities perturbing the current flow (current terminals) or affecting the voltage pick-up. When the size of the connections is large with respect to the dimension of the Hall generator itself, the connections can partially shunt the current and thus deform the ideal electric field pattern.

The finite geometry of the contacts and the shunting of the Hall voltage influence the electric field pattern in different ways a different applied magnetic field. As a result a Hall generator has a non-linear response to the field, usually identified as a field dependence of the geometric factor G. In addition to this geometric non-linearity, a Hall generator has also a material non-linearity that is commonly attributed to the Hall coefficient R_H. By proper choice of geometry and material it is possible to partially compensate the two effects achieving linear response to better than 1 % over a wide range of field. In this respect the Hall generator of the cruciform type [34] shows a better linearity and has a smaller active surface than the classical rectangular generator. Its magnetic center is better defined, so that it is particularly well suited for measurements in strongly inhomogeneous fields.

Another source of non-linearity in the characteristics of Hall generators is the variation of temperature. Both the charge mobility and material conductivity are temperature dependent, so that the Hall coefficient R_H has a temperature dependency that can be of the order of 100 to 1000 ppm/°C. The control current in the Hall generator is associated with a small heat source that must be controlled to achieve high precision.

As discussed above, the voltage reading in a Hall generator depends on the direction of the magnetic field. The maximum Hall voltage is measured when the magnetic field is in a direction normal to the generator, and ideally the Hall voltage is zero when the magnetic field is in the plane of the device. In reality any anisotropy in the generator leads to a non-zero voltage also when the field is in the plane of the sensor. This effect is often referred to as *planar* Hall effect, and depends on the field strength and on the angle φ between the field and the direction of the electric current in the generator [35,36,37] giving rise to an additional voltage at the terminals:

$$V_{planar} = V_{HP} B^2 \cos(2\varphi) \tag{10}$$

where V_{HP} is a coefficient typical of the material used and of the geometry of the generator. The typical planar Hall voltage in a generator built with InSb semiconductor is shown in Fig. 17.

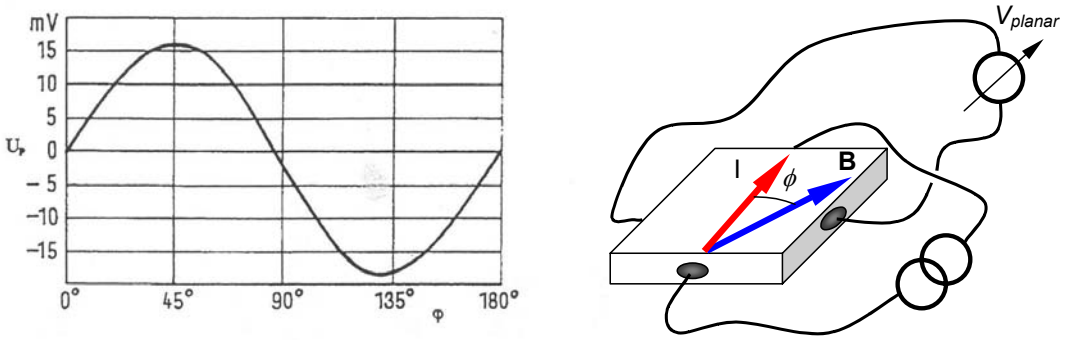


Figure 17. Measurement of planar Hall voltage in an InSb Hall generator in a background field of 1 T, plotted as a function of the angle between the field and the direction of the current in the generator (reproduced from [37]).

Magnetic field gradients can also have an effect on the Hall voltage, mainly because the generator has a finite size. The effect of field variation in the sensitive region of a Hall generator is however small and is generally neglected [36].

Special Hall generators for use at cryogenic temperature are also commercially available. Although they show a very low temperature coefficient, they unfortunately reveal an additional problem at low temperatures. In these conditions the so-called Shubnikov-de Haas effect [38,39] is responsible for a field dependent oscillation of the Hall coefficient causing as a deviation from the linear sensitivity. The deviation may amount to about one per cent at high fields, depending on the type of semiconductor used for the Hall generator. An example of this effect is shown in Fig. 18. The Shubnikov-de Haas effect adds a serious complication to the calibration.

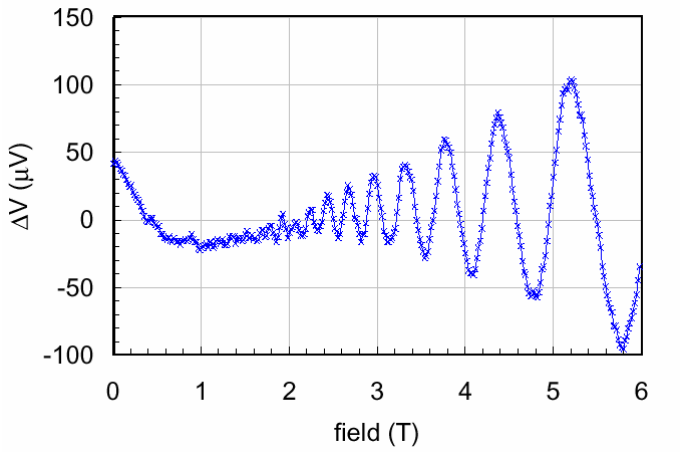


Figure 18. Shubnikov-de Haas effect in a Hall generator at 4.2 K. The plot shows the deviation ΔV of the Hall voltage from a straight line as a function of the applied field. For comparison, the Hall generator produces on average 10 mV/T.

A further parasitic effect is associated with the presence of a voltage offset even at zero applied magnetic field. The offset can be of geometric origin (misalignment of the voltage terminals with respect to the electric equipotential lines), associated with rectifying effects at the metal-semiconductor connection, stress induced or originating from gradients in material properties (e.g. doping density). The offset voltage is temperature dependent, and in addition temperature gradients at the connections can generate thermo-electric voltages. Voltage offsets can be significant, 0.1 to 1 mV are typical values, and hence need to be controlled or compensated to achieve good precision.

2.3.3 Field measurement with Hall generators

A Hall generator is a four-terminal device. The control current is generally supplied by a highly stable AC or DC source, while the Hall voltage is read by a high impedance voltmeter or is fed to the acquisition electronics (amplification or conditioning). It is useful to remember that the Hall generator input is not isolated from the output. In fact the impedance among input and output is generally of the order of the input resistance. In practice the current source must be isolated from the electronics that reads the Hall voltage to prevent stray current flowing through the Hall generator in the electronics. This is often achieved using a differential-input amplifier for signal conditioning.

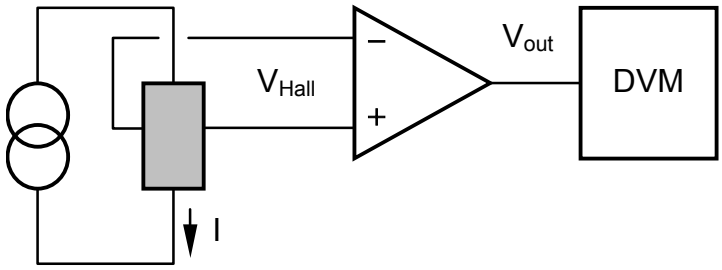


Figure 19. Principle of use of a Hall generator showing the current source and readout electronics.

AC excitation and lock-in techniques are used to increase the noise rejection of the reading of the Hall voltage. In this case the source current is modulated while the voltage is readout synchronously with the excitation. The frequency of the excitation is low, a few Hz to a few hundred Hz, and is chosen to reject typical sources of systematic noise such as 50 Hz and the related harmonics.

The measured Hall voltage must be converted to the value of the field using a calibration function. To achieve precision of the order of 1 % and better this function must be established through an independent measurement in a known field. The non-linearity in the relation between field and Hall voltage is in fact one of the major drawbacks of Hall generators. Precision measurements require in addition the knowledge of the temperature and the relative correction of the calibration. Alternatively, temperature compensation in a thermostat surrounding the Hall generator can be used. In this case the calibration measurement is performed at the same working temperature.

Another major issue to be considered for the calibration and the measurement is the alignment of the Hall generator. The Hall generator is a directional device, sensitive to the field component normal to the generator. An angle between the field and the normal to the surface of the generator results in an apparent reduction of the field strength. This can be empirically corrected by tilting the Hall generator until the maximum reading is obtained. Alternatively, one can establish a mechanical reference that can be reproduced between calibration and measurement. Three-dimensional arrays of Hall generators, also available as single chips sensitive to the three components of the magnetic field, once complemented with a suitable calibration, are the best adapted devices for field strength measurement.

Finally, it is important for precision measurements that parasitic effects, and especially voltage offset, are minimised or compensated. The offset is best removed with the Hall generator placed in a zero-field chamber, providing an ideal volume with no magnetic field. The offset compensation can be electronic or digital. In any case temperature stabilization is necessary to prevent the drift in the offset caused either by changes in temperature or temperature gradients.

2.3.4 Range and accuracy

Thanks to the latest technological development in integrated circuit design and manufacturing, Hall generators have become inexpensive devices that are widely used for large-scale applications and field mapping. The directionality and small active area of the Hall generator make it to an ideal instrument for mapping magnetic fields in a very wide range of strength and shape. Other attractive features are the wide dynamic range and the possibility of static operation under constant excitation current. The range of field that can be practically measured is related to the voltage measurement capability and the sensitivity of the Hall generator. The sensitivity in turn varies depending on the material, the thickness of the generator and the size of the active area. Commercially available generators have typical sensitivities in the range of 10 mV/T to 1 V/T. The typical field measurement range for these generators is above 1 mT. Accurate voltage measurement is possible below the μV level using precision voltmeters, thus giving a field resolution better than 0.1 μT.

The absolute accuracy of the measurement is mostly governed by the non-linearity and its temperature dependence, alignment errors, and the parasitic effects discussed earlier. Commercial equipment has typical accuracy 1000 ppm of reading. Custom-made calibration, temperature control and various compensation techniques can be used to achieve an improvement of a factor 10, giving an absolute accuracy of 100 ppm of reading. An additional issue in Hall generators is long-term stability. A well-designed Hall-probe assembly can be calibrated to a long-term accuracy of 100 ppm. The stability may be considerably improved by powering the Hall generator permanently and by keeping its temperature constant.

The active area of the Hall generator determines the size of the volume mapped. Typical volumes are in the range of 0.01 mm^3 to 0.1 mm^3. Quoted time resolutions for Hall generators are below 1 ns, as typical of semiconductors electronics. In general however the main measurement bandwidth limitation is determined by the precision voltmeter. Commercial devices have maximum bandwidth extending up to 500 Hz.

2.4 Magneto-resistors

Conducting materials exhibit a change in resistivity once exposed to a magnetic field. This effect, discovered by W. Thomson in mid 19th century [40], was exploited to measure the strength of the field already at the beginning of the 20th century. The change in resistance in conductors is due to two contributions, a *physical* contribution related to the intrinsic conductance properties of the material, and a *geometric* contribution due to the bending of the carrier path under the magnetic force and thus of the same origin as the Hall effect. The magneto-resistance associated with intrinsic changes in conductance property is usually small at room temperature, of the order of 0.5 to 2.5 % increase at 1 T in semiconductors such as InSb and InAs that are commonly used to build magneto-resistors. The origin of the geometric contribution is of dual nature with respect to the Hall effect and can be successfully exploited by properly shaping the current path in the material.

To explain how this is done we need to recall that in a Hall generator the bending of the current flow gives rise to a polarization in the direction transverse to the current. At a sufficient distance from the electrodes, and once the transverse electric field has built up, its effect is to balance the force due to the magnetic field. Therefore in a Hall generator in steady-state operation the current path is nearly straight between the two terminals, and the resistance has only a modest variation with field.

If, on the other hand, the Hall generator is built with a very short length then the charge cannot build up at the sides, and the transverse field is negligible while the current path curves increasingly as a function of the applied field. In this situation the Hall voltage is small while the longitudinal resistance increases significantly with field due to the increase of the length traversed by the charge carriers. To reach an appreciable total resistance, and thus improve the sensitivity to the magnetic field, several of these short Hall generators must be piled in series. One of the possible ways to achieve this is to distribute microscopic needles of a high conductivity material such as NiSb in a base material with large Hall effect such as in InSb. The needles are placed with their long axis normal to the current flow and thus short the Hall voltage transverse to the current. In practice the needles delimit ideal Hall plates with a small ratio between width and length, as shown schematically in Fig. 20. The change in the length of the current path under the action of the magnetic field between two adjacent needles is responsible for the increase in resistance.

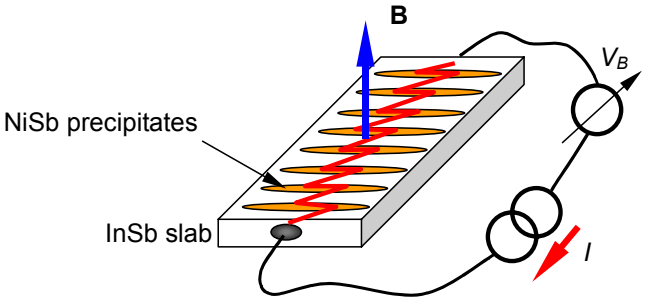


Figure 20. Schematic representation of a magneto-resistor containing needle-shaped low resistance precipitates (e.g. NiSb) in a semiconductor matrix (e.g. InSb). The needles act as shorting bars for the Hall voltage, and redirect the current flow, shown as the *zig-zag* path inside the slab. The total resistance growth with field is proportional to the change in the length of the current path in the semiconductor.

Magneto-resistors of this type have a highly non-linear resistance increase, in first approximation proportional to the square of the field strength, or:

$$\frac{R_B}{R_0} \propto B^2 \tag{11}$$

where R_0 and R_B are the resistances in zero field and in the applied field respectively. For this reason magneto-resistors are relatively insensitive to low fields. In addition they are sensitive only to the field component in the direction normal to the slab, but cannot distinguish field polarity. Finally they have large temperature coefficients, 2500 ppm/C typical, and can be affected by large property variations from device to device. For the above reasons up to recent years magneto-resistors did not find widespread application, in spite of the advantage of being a two-terminal device and the easy manufacturing with integrated circuit technology. The recent discovery of the *giant magneto-resistance* effect [41] as well as technological developments have improved the situation. Highly symmetric magneto-resistors are built today using thin layers of ferromagnetic and non-magnetic materials in Wheatstone bridges for high sensitivity detection of magnetic field gradients. The symmetry is achieved by manufacturing the magneto-resistors using photolithography on the same substrate. The resistive bridge configuration largely removes temperature dependence of the balanced voltage, as well as the effect of uniform magnetic fields. Often a magnetic element is added in the circuit to produce a bias field that displaces the working point of the magneto-resistors to a region of acceptable linearity. In this configuration magneto-resistors can be used as position and proximity sensing devices, as well as devices to measure the electrical current in nearby conductors. In this range of applications magneto-resistors are suitable for the detection of fields in the range of 1 μT to a few mT. In resistive bridge configuration, using suitable feed-back circuitry for compensation, the sensitivity can be as low as 100 pT. Calibration properties and long term stability are similar to Hall generators, although the intrinsic non-linearity of the sensors lowers the absolute precision that can be achieved to the level of fraction of % (1000 ppm). The typical bandwidth of magneto-resistive sensor is extremely large, from DC to several GHz, and the volume mapped can be extremely small, down to 100 μm^2, e.g. for photo-lithographically patterned giant magneto-resistive sensors.

2.5 Fluxgate magnetometer

A fluxgate magnetometer measures the magnetic field through its effect on the magnetization of a thin ferromagnetic core with a very high magnetic permeability and nearly rectangular hysteresis. The measurement principle is illustrated in Fig. 21.

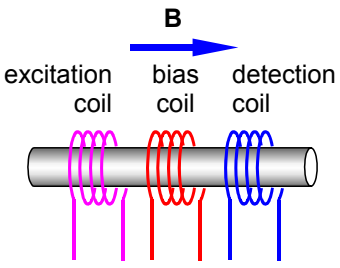


Figure 21. Principle configuration of a fluxgate magnetometer.

In its basic version, also called “peaking strip” [42], it consists of three coils wound around a ferromagnetic core: an excitation winding that produces an AC magnetic field well above the saturation field of the core, a detection winding that picks-up the excitation voltage and is used to detect the zero field condition and to generate an error signal proportional to the field, and a DC bias coil that improves the linearity of the device by creating and maintaining zero field in the core. The current in the bias winding is adjusted based on the error signal from the detection winding. At

equilibrium, once the error signal is zero, the field to be measured is equal (and opposite) to the field generated by the bias coil. The magnetometer reading provides hence a signal proportional to the field.

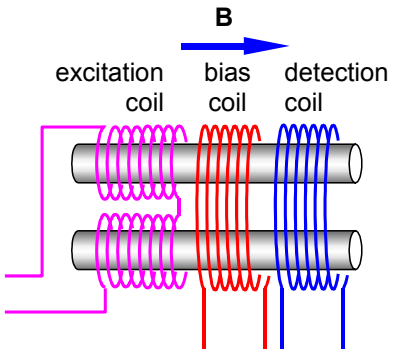


Figure 22. Fluxgate magnetometer built from two ferromagnetic needles.

In practice the coils are wound coaxially in subsequent layers to minimise stray field and maximise mutual inductances. Two ferromagnetic needles are used in order to avoid the distortion of the background field due to the presence of the probe, see Fig. 22. The excitation coil is wound around each needle, but the direction of the winding of the coils is reversed. The AC current in the excitation coil magnetizes the two needles with the same strength, but opposite direction. The detection and bias coils are wound surrounding the two ferromagnetic cores and the excitation coil. The detection coil thus senses the voltage induced by the change of magnetization in both needles. In absence of an external field the total voltage seen by the detection coil is zero because the magnetic field in the two needles have the same strength but opposite direction. If on the other hand the cores are aligned with an external magnetic field, one core will produce a magnetic field in the same direction and will reinforce it, while the other will subtract to it. This difference generates a measurable signal in the detection coil that appears as a second harmonic of the excitation current and whose strength is proportional to the external magnetic field (see Fig. 23). As discussed above, the bias coil is powered with a DC current proportional to the error signal from the detection coil.

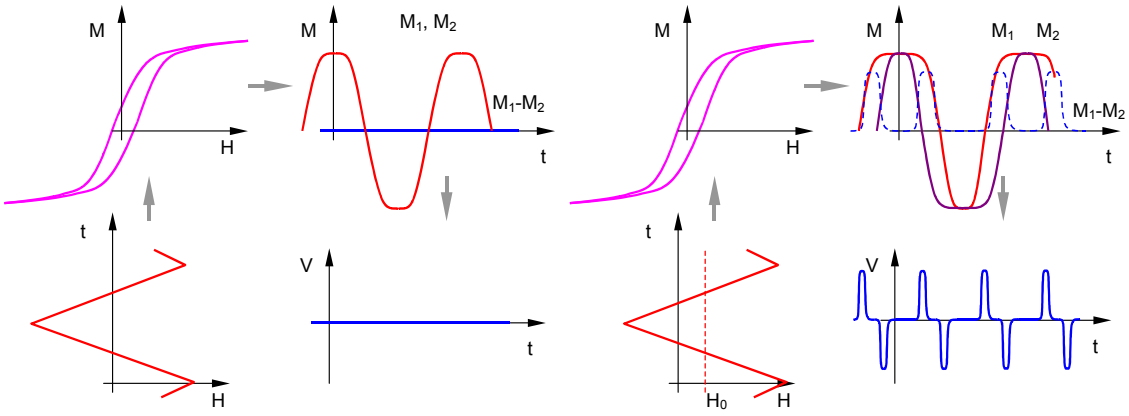


Figure 23. Working principle of the two-needles fluxgate magnetometer of Fig. 21 explained using a graphical construction. The magnetization of the two needles M_1 and M_2 is plotted in absence (left) and presence (right) of an external field $B_0 = \mu_0 H_0$. The voltage on the detection coil V is proportional to the time derivative of the difference of the two magnetizations M_1-M_2.

A fluxgate magnetometer of this type provides a measurement of the field component in the direction of the two needles and is a highly sensitive directional device. The bias coil is often combined with the

detection coil by powering it directly with a DC component proportional to its AC reading. Combining several magnetometers it is possible to sample the field in different directions, e.g. in three dimensions. The ferromagnetic needles can be substituted by a toroidal core partially covered by the detection coil. In this case the field component detected is the one normal to the direction of the winding of the detection coil. Using two detection coils wound at 90 degrees on the toroidal core it is possible to sense simultaneously two components of the field.

The first fluxgate magnetometers were built in the 1930's and largely developed during World War II. One of the applications was submarine detection from low-flying aircraft. The improvement of magnetic materials and detection electronics has lead to an improvement of sensitivity up to 0.1 nT. Because of the excellent sensitivity, small volume and weight and low cost, the fluxgate magnetometer is commonly used in the measurement of weak magnetic fields, such as earth field or rock layers. It can be used to detect small anomalies in magnetization such as generated by leakage earth current in power lines or the magnetic signature of vessels and weapons. Due to its directionality it is also used as a navigation instrument, to measure the geomagnetic field as well as for space applications, e.g. in the mapping of the interplanetary and planetary magnetic field.

Finally, it is worth to remember that the Direct Current Current Transformer (DCCT), commonly used for high-precision measurement of current, is based on the same principle of the fluxgate magnetometer. In the case of a DCCT the magnetic core has a toroidal shape, with the excitation and detection/bias coils wound around the toroid, as shown schematically in Fig. 24. The DCCT senses the weak magnetic field produced by a current conductor passing in the toroidal core. By careful arrangement of the conductor the magnetic field generated is perfectly linked with the magnetic core, while external fields have no influence. The signal on the bias coil is thus proportional to the current flowing in the conductor.

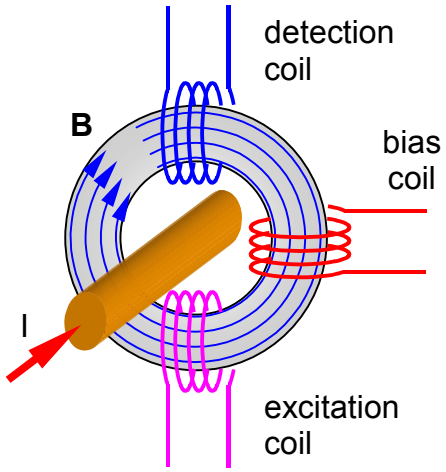


Figure 24. Principle of a DCCT, using the fluxgate magnetometer concept to measure the current passing in a conductor.

The range of fluxgate magnetometer is limited by the capability of the bias coil. The maximum field that can be measured is of the order of a few mT. The field range can be extended significantly by actively cooling the bias coil. The accuracy of the reading depends mainly on the calibration and on the stability of the device. Typical accuracy of commercial devices is in the 0.1 % to 1 % range (a few 1000's ppm). The volume sampled by the commercial probes ranges from a few to a few tens of mm^3. Finally, the bandwidth is typically from DC conditions to 1 kHz.

2.6 Magneto-inductance

Magneto-inductive magnetometers are relatively new devices [43,44], with the first patent issued in 1989. The magneto-inductive sensor detects a change in the inductance of a coil wound around a magnetic core whose permeability changes linearly over the range of measurement. This coil is an element of an L-R relaxation oscillator, as shown schematically in Fig. 25. The oscillation frequency is proportional to the external field to be measured. A DC bias can be applied to the coil to shift the working point of the magnetic core in a linear range of operation. The frequency can be monitored and measured to derive the corresponding field value by calibration. As for the fluxgate magnetometer, the sensor can be arranged to detect not only the magnitude but also the direction of the field. These magnetometers are gaining in popularity as they are easy in construction and relatively inexpensive. Magneto-inductive magnetometers have a limited operating range of typically 1 mT with a resolution in the μT range. Absolute accuracy is limited by calibration and stability issues.

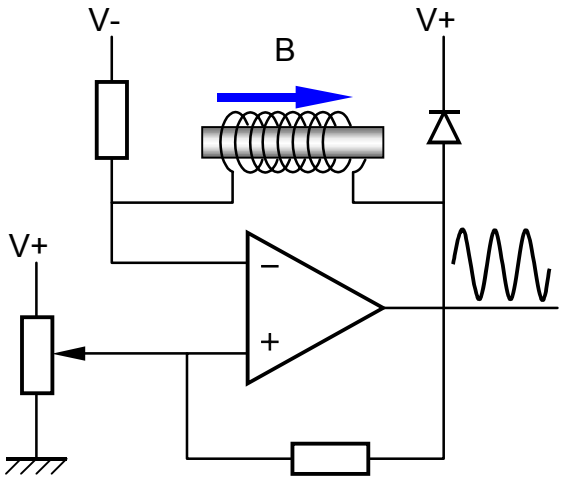


Figure 25. Principle of a magneto-inductive sensor. The sensor is the coil wound around a magnetic core. The oscillation frequency is proportional to the inductance of the sensor which in turn is affected by the applied field.

2.7 Superconducting Quantum Interference Device (SQUID)

2.7.1 Principle and applications

The Superconducting Quantum Interference Device (SQUID) is at present the device with the highest sensitivity to the magnetic field. The working principle of a SQUID rests on two physical phenomena typical of superconductors and accounted for only by quantum mechanics: flux quantization and Josephson tunneling. Flux quantization originates from the relation between magnetic flux and the phase of the wave function describing paired electrons in a superconductor. In the case of a superconducting ring encircling a normal-conducting region the change of phase δ along a closed path Γ that includes the hole can be expressed as a function of the magnetic flux φ linked with the same path as follows:

$$\delta = 2\pi \frac{\varphi}{\varphi_0} \tag{12}$$

where the quantity φ_0 is the so called flux quantum:

$$\varphi_0 = \frac{h}{2e} \tag{13}$$

whose value in SI units is 2.07×10^{-15} Wb. In order to avoid interference in the wave function, and thus to have a finite probability for the paired state, the phase difference must be an integer multiple of 2π. The consequence is that the magnetic flux inside the superconducting ring can only take integer multiples of φ_0. It is for this reason that magnetic field penetration in a bulk superconductor can only take place discretely, with the field entering in frozen flux lumps, the fluxoids, enclosing a magnetic flux equal to φ_0.

The second phenomenon, Josephson tunneling, is named after Brian Josephson who predicted in 1962 that pairs of superconducting electrons in two bulk superconducting material would tunnel through a thin, normal conducting barrier separating them [45]. The assembly of the superconducting materials and the normal conducting barrier is called a *Josephson junction*. Because of the pairing mechanism, a DC superconductive current can flow across the junction. The maximum current I that can flow without voltage development depends on the phase difference δ between the wave functions describing the paired electrons on the two sides of the resistive barrier. The dependence is periodic in δ:

$$I = I_c \sin(\delta) \tag{14}$$

where I_c is the maximum critical current of the junction. Above the critical current a resistive voltage is developed and the junction behaves as an ohmic conductor.

A SQUID consists schematically of a superconducting ring separated by one (RF SQUID) or two (DC SQUID) Josephson junctions. Although RF SQUIDs were the first to be used for magnetometry, present systems are largely based on DC SQUIDs. For this reason we will limit our discussion to DC SQUIDs only. Details and a more precise treatment on the principles of operation of SQUIDs can be found in [46] and [47].

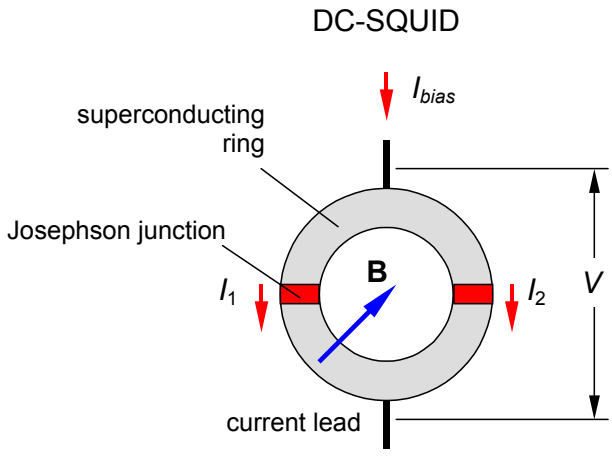


Figure 26. Schematic geometry of a DC-SQUID consisting in a superconducting ring with two Josephson junctions.

In the case of the DC SQUID shown in Fig. 26 the critical current of the ring is given by the parallel of the critical currents of the two Josephson junctions I_1 and I_2. These in turn depend through the relation expressed in Eq. (13) on the phase differences δ_1 and δ_2 of the wave functions of paired electrons across the junctions. In the presence of an external magnetic field the integral δ of the phase of the wave function of paired electrons around the ring depends on the flux linked by the ring in accordance with Eq. (12), where now however we need to consider the phase jumps across the two junctions:

$$\delta + \delta_1 - \delta_2 = 2\pi \frac{\varphi}{\varphi_0} \tag{15}$$

The flux quantization condition, requesting that the total phase change is an integer multiple of 2π, applies to the above equation and imposes a constraint between the single-phase jumps across the junctions. This constraint, combined with the periodic dependence of the maximum currents of each junction expressed by Eq. (14), results in a flux-dependent modulation of the critical currents of the two junctions. It can be shown in particular that the total critical current of the ring, under simple conditions of identical junctions and negligible loop inductance, is given by:

$$I = 2I_c \left| \cos\left(\pi \frac{\varphi}{\varphi_0} \right) \right| \tag{16}$$

which is a periodic function with period φ_0 and maximum value $2I_c$ shown in Fig. 27. To measure flux, the SQUID is powered as shown in Fig. 26 with a bias current I_{bias} close to the maximum critical current. The resistive voltage V appearing at the terminals of the SQUID is then also a periodic function of the magnetic flux, as shown schematically in Fig. 27.

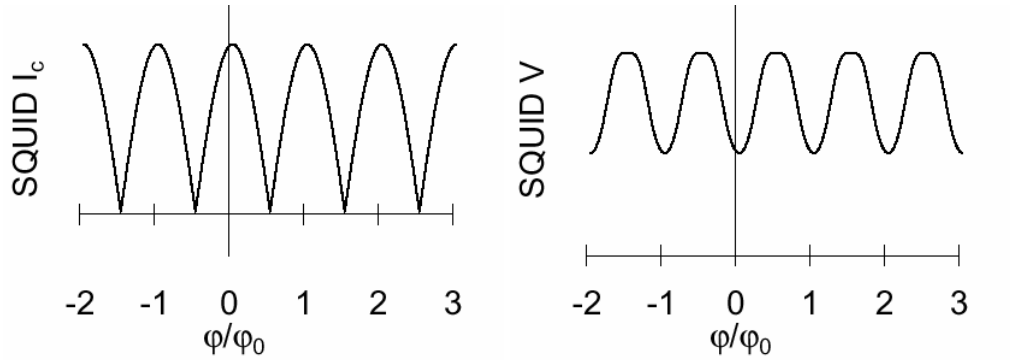


Figure 27. Critical current (left) and voltage (right) of a DC-SQUID as a function of the magnetic flux threading the ring.

Detecting the resistive voltage change it is hence possible to measure the flux linked by the SQUID. In practice the flux variation is measured using the voltage from the SQUID as an error signal for a feed-back coil magnetically coupled with the SQUID that adjusts until the feed-back flux cancels the external flux. The flux generated by the feed-back coil is then used as a measurement of the external flux. This zeroing technique makes possible to measure flux variations that are small fractions of the flux quantum.

Based on the measured flux change it is possible to derive the magnetic field as in the case of an induction coil. The main factors that govern the sensitivity of a SQUID are therefore the flux resolution and the SQUID area. The practical size of a SQUID is about 0.01 mm^2, and the field variation that can be detected by a bare device is in the pT range. The absolute accuracy of the device depends on the knowledge of the SQUID surface, but is usually not the main concern, as SQUIDS are mainly used for field measurements because of their extraordinary sensitivity. In order to further increase the sensitivity of the simple SQUID magnetometer described above, the flux seen by the superconducting ring can be amplified using a superconducting transformer. This consists of a detection coil wound with a superconducting wire connected to an input coil mutually coupled with the SQUID as shown in Fig. 28. Choosing properly the ratio of area and turns of the detection and input coils it is possible to amplify any field variation at the detection coil by a factor up to one

thousand. This gives a detection capability at the level of 10 fT for coil sizes of a few cm^2. Due to their exceptional sensitivity, SQUIDs are well suited for the measurement of the tiny magnetic fields associated with heart (100 pT) and brain (100 fT) activity. In addition they are used for fine magnetometry, non-invasive measurements and flaw detection. The bandwidth of a SQUID magnetometer is typically from DC to a few tens of kHz. Given the present rate of development in high-temperature superconductors and superconducting electronics, SQUIDs are expected to become more widespread in the future.

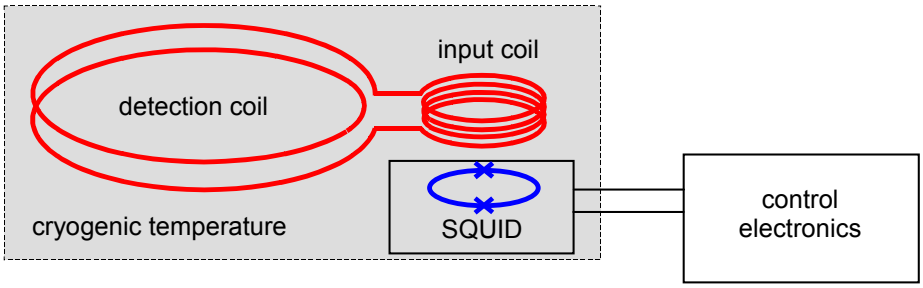


Figure 28. Coupling of a SQUID with a superconducting transformer consisting of a detection coil and an input coil that concentrates the magnetic flux at the SQUID.

2.8 Particle beams

The motion of elementary charged particles is another sensitive method to measure and map magnetic field over large volumes. As an example, low-energy electron beams are systematically used to map the field surfaces of thermonuclear fusion experimental devices with magnetic confinement. A diffused electron beam is injected in the plasma volume. The motion of the electrons takes place along the field lines, on the nested magnetic surfaces. Using transparent fluorescent meshes to detect the electrons it is possible to obtain a cross section of the field surfaces. The resulting maps are generally used to verify the regularity of the magnetic surfaces and the absence of magnetic island that could generate an instability in the thermonuclear plasma.

A second example is a method for the precise measurement of the beam position with respect to the magnetic center of quadrupole magnets installed in particle accelerators that has been developed over the last decade [48,49]. The procedure consists of modulating the field strength in individual lattice quadrupoles while observing the resulting beam orbit oscillations. Local DC orbit distortions are applied in the search for the magnetic center. This so-called K-modulation provides a perfect knowledge of the location of the particle beam with respect to the center of a quadrupole. In addition, it may provide other very useful observations for operation and adjustment of the accelerator [50]. This is obviously of particular importance for superconducting accelerators [51]. It is very difficult to provide a superconducting quadrupole magnet with a direct optical reference to its magnetic center, so errors caused by changes of temperature profiles and other phenomena may build up as time passes. The method may be further improved by synchronous detection of the oscillation, so that its phase can be identified. The sensitivity of the detection is impressive. Experience from LEP [52] showed that an absolute accuracy of 0.05 mm in both the vertical and horizontal plane could be obtained. Furthermore it was observed that a modulation of the quadrupole field of about 300 ppm could be clearly detected, which means that measurements may be carried out on colliding beams while particle physics experiments are taking place.

2.9 Magneto-optical measurement methods

Light propagation in matter is affected by the presence of a magnetic field, and this can be exploited to detect and measure the magnetic field itself. As light-based system have generally a high degree of electromagnetic compatibility and are fast, magneto-optical measurement find their application in

special and extreme operating conditions where measurement speed is important. The magneto-optical effect that is most widespread for field measurement is the Faraday rotation of the polarisation plane of light. The principle of an optical magnetometer based on Faraday rotation is shown in Fig. 29. A light beam is polarised on a plane before entering the detector proper. The detector itself consists of a transparent material (e.g. an optical fiber) placed in the magnetic field to be detected. The polarisation plane is rotated by the magnetic field component parallel to the direction of light propagation. The rotation θ of the polarisation plane is linearly proportional to the integrated field over the detector material, or:

$$\theta = v\int Bdl \tag{17}$$

vhere v is the material dependent Verdet constant. The light emerging from the detector is then analysed with an analyser and measured on a light intensity detector. As standard practice, the analyser is set at right angle with respect to the plane of polarisation, and the detector is used to find the minimum of the transmitted light intensity, thus increasing the resolution. This measurement principle has shown to be useful for measurements of transient magnetic fields [53,54], especially when high speed (sub-μs time scale) and ruggedness (e.g. explosive magnetic field generators) are important. It is however less convincing with respect to accurate measurement of field strength, as in general the rotation of the polarisation plane through practical thicknesses of optical materials is small, as shown by the values of the Verdet constant reported in Table 2. The uncertainties in the measurement of the angle of rotation of the polarisation plane, combined with the wavelength dependence of the Verdet constant, limit the typical accuracy of magneto-optical field measurements to fractions of percent at best (a few thousands of ppm).

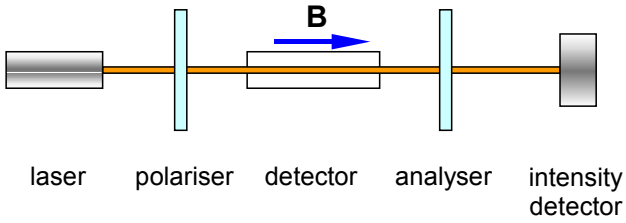


Figure 29. Working principle of magneto-optical field measurement based on the Faraday effect.

Table 2
Typical values of Verdet constant for optical materials, measured in visible light.

material	v (rad/T m)
water	3.8
glass	4…6
dense flint	25
quartz	4.8
NaCl	12
CaF_2	2.5
benzene	8.7

2.10 Field Imaging

The best known method for field imaging is made by spreading iron powder on a horizontal surface placed near a magnetic source, thus providing a simple picture of the distribution of flux lines. Another very classical way of observing flux-line patterns is to place a free-moving compass needle at

different points in the volume to be examined and note the direction of the needle. This compass method was applied, long before the discovery of electromagnetism, for studies of the variations in the direction of the earth's magnetic field. The magneto-optical rotation of the plane of polarization of polarized light (the Faraday effect described previously) is a classical method for the visualization of magnetic fields. A transparent container filled with a polarizing liquid and placed inside the magnet gap may visualize for example the field pattern in a quadrupole by observation through polarization filters placed at each end of the magnet. The rotation of the plane is proportional to the field strength and the length of the polarizing medium and may give a certain indication of the field geometry. A similar effect may be obtained by observing the light transmission through a colloidal suspension of magnetic particles, subject to the field [55,56]. An example of field imaging obtained using this method is shown in Fig. 30. The field pattern reported there has been obtained at RHIC using a ferro-fluid colloidal suspension contained in a cell placed in the warm bore of a quadrupole operated at a gradient of 75 T/m during cold test [57].

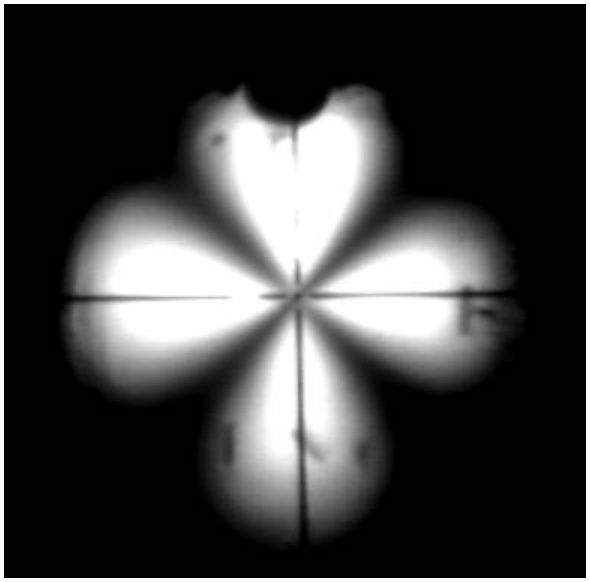


Figure 30. Polarised light pattern obtained using a ferro-fluid colloidal cell placed in the warm bore of a RHIC quadrupole powered at 75 T/m during tests at cryogenic temperature [57].

2.11 Other techniques

A plethora of different, alternative techniques can be used to measure magnetic field. Many of these techniques were either popular in the past, and have passed from practice, or are becoming increasingly attractive thanks to technological development, or again they are specially suited for particular applications. A few examples are listed below, once again not aiming at completeness:

Floating wire measurements were quite popular in the past [58]. If a current-carrying conductor is stretched in a magnetic field, it will curve subject to the electromagnetic force and describe the path of a charged particle with a momentum corresponding to the current and the mechanical tension in the wire. A flexible annealed aluminum wire was used in order to reduce effects of stiffness and gravity. This method has now been entirely replaced by precise field mapping and simulation of particle trajectories by computer programs.

Cantilever magnetometers sense the strength of a magnetic moment in a field gradient through the measurement of the magnetic force or the magnetic torque.

Magnetostriction has been used [59,60] to measure the magnetic field either detecting the magnetostrictive strain using a piezo-electric sensor or measuring the motion of a magnetostrictive ribbon using a tunneling tip.

Magneto-transistors and magneto-diodes are being developed. Semiconductors are sensitive to the magnetic field either through the Hall effect, or through the Suhl effect, a process by which the holes injected into an N-type semiconducting filament are deflected to the surface when a strong transverse magnetic field is applied to the semiconductor.

3. SUMMARY

Figure 31 reports an order-of-magnitude summary of the capability of the methods discussed in this note, in terms of measurement accuracy and measurement range.

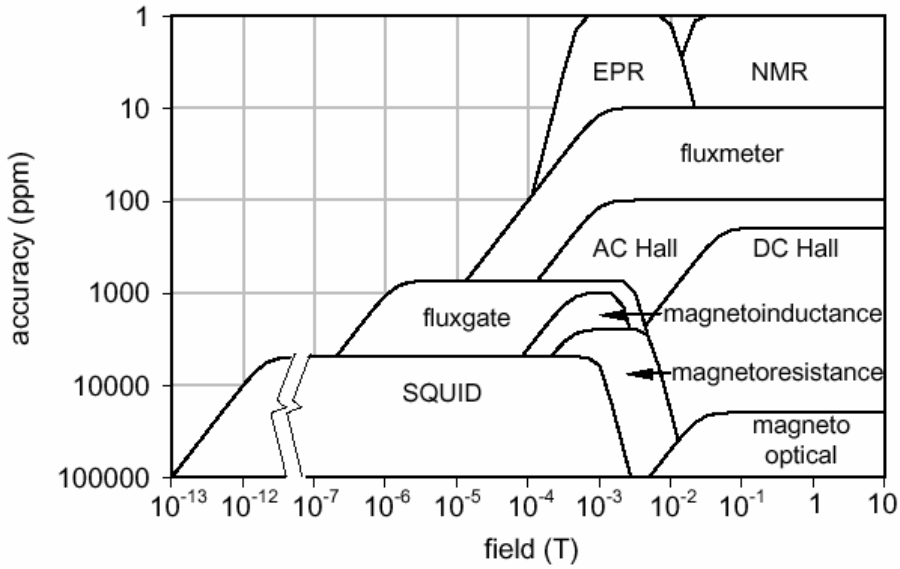


Figure 31. Overview of typical accuracy attainable with different magnetic measurement methods depending on the measured field. Note the jump by 5 orders of magnitude at the lower end of the magnetic field scale.

Proven measurement methods and powerful equipment is readily available for most of the measurement tasks related both to normal and superconducting magnets. It is therefore prudent to examine existing possibilities carefully before launching the development of a more exotic measurement method. Many unnecessary costs and unpleasant surprises can be avoided by choosing instrument, which are commercially available. The measurement methods described above are complementary and a well-chosen combination of two or more of these will certainly meet most requirements.

REFERENCES

[1] Nuclear Instruments and Methods in Physics Research – Section A, Elsevier Science.
[2] Proc. of the International Magnet Technology Conferences, MT-1 (1965) to MT-17 (2001).
[3] Proceedings of Particle Accelerator Conferences, PAC-1 (1965) to PAC-19 (2001).
[4] Proceedings of European Particle Accelerator Conferences, EPAC-1 (1988) to EPAC-18 (2002).
[5] Proc. of the International Magnetic Measurement Workshops, IMMW-1 (1977) to IMMW-12 (2001).

[6] Proceedings of CERN Accelerator School on Magnetic Measurement and Alignment, S. Turner ed., CERN-92-05, 1992.
[7] Proceedings of CERN Accelerator School on Measurement and Alignment of Accelerator and Detector Magnets, S. Turner ed., CERN-98-05, 1998.
[8] J.J. Rabi, J.R. Zacharias, S. Millman, P. Kusch, "A new method of measuring nuclear magnetic moment", Phys. Rev., **53**, 318, 1938.
[9] F.F. Bloch, W.W. Hansen, M. Packard, "Nuclear Induction", Phys. Rev., **69**, 127, 1946.
[10] F.F. Bloch, W.W. Hansen, M. Packard, "The Nuclear Induction Experiment", Phys. Rev., **70**, 474, 1946.
[11] E.M. Purcell, H.C Torrey, R.V. Pound, "Resonance Absorption by Nuclear Magnetic Moments in a Solid", Phys. Rev., **69**, 37-38, 1946.
[12] NIST Journal, **95** (5), 521, 1990.
[13] C. Reymond, "Magnetic Resonance Techniques", CERN Accelerator School on Measurement and Alignment of Accelerator and Detector Magnets, CERN-98-05, 219-231, 1998.
[14] W.G. Clark, T. Hijmans, W.H. Wong, "Multiple Coil Pulsed NMR Method for Measuring the Multipole Moments of Particle Accelerator Bending Magnets", J. Appl. Phys. **63**, 4185-4186, 1988.
[15] W.G. Clark, J.M. Moore, W.H. Wong, "Multiple Coil Pulsed Magnetic Resonance Method for Measuring Cold SSC Dipole Magnet Field Quality", Supercollider 2: Proc. of the 2nd International Industrial Symposium on the Supercollider, Miami Beach, USA 405-414, 1990.
[16] G. Suryan, "Nuclear Resonance in Flowing Liquids", Proc. Indian Acad. Sci., A33, 107-111, 1951.
[17] C. Sherman, "High-Precision Measurement of the Average Value of a Magnetic Field over an Extended Path in Space", Rev. Sci. Instr., **30**, 568-575, 1959.
[18] J.M. Pendlebury, K. Smith, P. Unsworth, G.L. Greene, W. Mampe, "Precision Field Averaging NMR Magnetometer for Low and High Fields, using Flowing Water", Rev. Sci. Instr., **50**, 535-540, 1979.
[19] L. Jansak, J. Kokavec, J. Kvitkovic, P. Galbraith, K.N. Henrichsen, "Low Field NMR Magnetometry using Flowing Water", Proceedings of the International Conference on Measurement, Measurement 97, Smolenice, Slovakia, 328-331, 1997.
[20] W. Weber, "Ueber die Anwendung der Magnetischen Induktion auf Messung der Inklination mit dem Magnetometer", Annalen der Physik, **2**, 209-247, 1853.
[21] M.I. Green, "Search Coils", CERN Accelerator School on Measurement and Alignment of Accelerator and Detector Magnets, CERN-98-05, 143-173, 1998.
[22] M.W. Garrett, "Axially Symmetric Systems for Generating and Measuring Magnetic Fields", Part I, Appl. Physics, **22** (9), 1091, 1951.
[23] W.F. Brown, J.H. Sweer, "The Fluxball", Rev. Sci. Instr., **16**, 276, 1945.
[24] B. de Raad, "Dynamic and Static Measurements of Strongly Inhomogeneous Magnetic Fields", Ph.D. Thesis, Delft University, 1958.
[25] G.H. Morgan, Proc. of 4th Int. Mag. Techn. Conf., USAEC CONF-720908, 787, 1972.
[26] Z.R. Wolf, et al., "Magnetic Measurement Instrumentation Development at the SSC", Supercollider **4**, 467-474, 1992.
[27] D. Zangrando, R.P. Walker, "A Stretched Wire System for Accurate Integrated Magnetic Field Measurements in Insertion Devices", Nucl. Instr. and Meth., A **376**, 275-282, 1996.
[28] W.C. Elmore, M.W. Garrett, "Measurement of Two-Dimensional Fields", Part I: Theory, Rev. Sci. Instr., **25**, 480-485, 1954.
[29] I.E. Dayton, F.C. Shoemaker, R.F. Mozley, "Measurement of Two-Dimensional Fields, Part II: Study of a Quadrupole Magnet", Rev. Sci. Instr., **25**, 485-489, 1954.
[30] A.K. Jain, "Harmonic Coils", CERN Accelerator School on Measurement and Alignment of Accelerator and Detector Magnets, CERN-98-05, 175-217, 1998.
[31] R.A. Beth, "Complex Representation and Computation of Two-Dimensional Magnetic Fields", J. Appl. Phys, **37**, 2568-2571, 1966.

[32] E.H. Hall, “On a New Action of the Magnet on Electric Currents”, Am. J. Math., **2**, 287, 1879.
[33] G.L. Pearson, “A Magnetic Field Strength Meter Employing the Hall Effect in Germanium”, Rev. Sci. Instr., **19**, 263, 1948.
[34] J. Hauesler, H.J. Lippmann, ”Hallgeneratoren mit Kleinem Lineariserungsfehler, Solid State Electron”, **11**, 173-182, 1968.
[35] C. Goldberg, R.E. Davis, “New Galvanometric Effect”, Phys. Rev., **94**, 1121-1125, 1954.
[36] B. Berkes, “Hall generators”, CERN Accelerator School on Magnetic Measurement and Alignment, CERN-92-05, 167, 1992.
[37] J. Kvitkovic, “Hall generators”, CERN Accelerator School on Measurement and Alignment of Accelerator and Detector Magnets, CERN-98-05, 233-249, 1998.
[38] J. Babiskin, “Oscillatory Galvanomagnetic Properties of Bismuth Single Crystals in Longitudinal Magnetic Fields”, Phys. Rev., **107**, 981-992, 1957.
[39] H.P.R. Frederikse, W.R. Hosler, “Oscillatory Galvanomagnetic Effects in n-Type Indium Arsenide,” Phys. Rev., **110**, 880-883, 1958.
[40] W. Thomson, “On the Magnetization of Electric Conductivity of Metals”, Philosoph. Trans., **146**, 736-751, 1856.
[41] M.N. Baibich, et al., “Gian Magnetoresistance of (001)Fe/(001)Cr Magnetic Superlattices”, Phys. Rev. Lett., **61**, 2472-2475, 1988.
[42] J.M. Kelly, “Magnetic Field Measurements with Peaking Strips”, Rev. Sci. Instr., **22**, 256-258, 1958.
[43] M.J. Caruso, T. Bratland, C.H. Smith, and R. Schneider, “A New Perspective in Magnetic Field Sensing, Sensors Online”, December 1998, http://www.sensorsmag.com/articles/1298/mag1298/index.htm.
[44] N.H. Kim, T. Hawks; “Digital compass and magnetometer having a sensor coil wound on a high permeability isotropic core”, US Patent No. 4,851,775, 1998.
[45] B.D. Josephson, Phys. Lett., 1, 251-253, 1962.
[46] J. Clarke, “SQUID Sensors: Fundamentals, Fabrication and Applications”, H. Weinstock ed., 1-62, Kluwer Dordrecht, 1996.
[47] P. Carelli, G. Paterno, “Macroscopic Quantum Interference: DC-SQUID, in Principles and Applications of Superconducting Quantum Interference Devices”, A. Barone ed., World Scientific, 1992.
[48] D. Rice, G. Aharonian, K. Adams, M. Billing, G. Decker, C. Dunnam, M. Giannella, G. Jackson, R. Littauer, B. McDaniel, D. Morse, S. Peck, L. Sakazaki, J. Seeman, R. Siemann, R. Talman, “Beam Diagnostic Instrumentation at CESR”, IEEE Trans. on Nucl. Sci., **30**, 2190-2192, 1983.
[49] P. Rojsel, “A Beam Position Measurement System Using Quadrupole Magnets Magnetic Centra as the Position Reference”, Nucl. Instr. and Meth., A **343**, 371-382, 1994.
[50] R. Brinkmann, M. Boge, “Beam-Based Alignment and Polarization Optimization in the HERA Electron Ring”, 4th European Particle Accelerator Conference, London, 938-940, 1994.
[51] J. Deregel, P. Genevey, J.-M. Rifflet, P. Galbraith, K.N. Henrichsen, “Proposal of a K-Modulation System for the LHC Quadrupoles”, CERN LHC Project Report 4, 1996.
[52] I. Barnett, A. Beuret, B. Dehning, P. Galbraith, K. Henrichsen, M. Jonker, M. Placidi, R. Schmidt, L. Vos, J. Wenninger, I. Reichel, F. Tecker, “Dynamic Beam Based Alignment”, presented at the 6th Beam Instrumentation Workshop, Vancouver, Canada, 1994.
[53] J. Malecki, M. Surma, J. Gibalewicz, “Measurements of the intensity of transient magnetic fields by the Faraday effect”, Acta Phys. Polon., 16 (1057) 151-156.
[54] J.L. Robertson, D.T. Burns, D. Hardie, “Measurements of a magnetic field integral using the Faraday effect”, Nucl. Instr. And Meth., 203 (1982) 87-92.
[55] J.K. Cobb, J.J. Muray, “Magnetic Center Location in Multipole Fields”, Nucl. Instr. and Meth., 46, 99-105, 1967.

[56] D. Trbojevic, P. Cameron, G.L. Ganetis, M.A. Goldman, R. Gupta, M. Harrison, M.F. Hemmer, F.X. Karl, A. Jain, W. Louie, S. Mulhall, S. Peggs, S. Tepikian, R. Thomas, P. Wanderer, "Alignment and Survey of the Elements in RHIC", 1995 Particle Accelerator Conference, Dallas, Texas, USA, 2099-2021, 1995.
[57] A. Jain, Private Communication, BNL, 2002.
[58] L.G. Ratner, R.J. Lari, "A precision system for measuring wire trajectories in magnetic fields", Proc. International Symposium on Magnet Technology, Stanford, USA (1965) 497-504.
[59] M.D. Mermelstein, A. Dandridge, "Low-Frequency Magnetic Field Detection with a Magnetostrictive Amorphous Metal Ribbon", Appl. Phys. Lett., 51 (7), 545-547, 1987.
[60] J.H. Wandass, J.S. Murday, R.J. Colton, "Magnetic Field Sensing with Magnetostrictive Materials Using a Tunneling Tip Detector, Sensors and Actuators", 19, 211-225, 1989.

DETECTOR MAGNET DESIGN

T. M. Taylor
CERN, Geneva, Switzerland

Abstract
The place of magnets in detectors for particle physics experiments is explained. The magnet is often the most visible and expensive component of he experiment, and it is important to ensure that it is appropriately optimized – both as a major piece of engineering and as a complement to the other experimental equipment. The approach to this process is described, with indications of how to call on the significant experience which been accumulated on the subject. The salient features of magnets past and present are given as examples, as is the direction of current development in the field.

1. INTRODUCTION

In most particle physics experiments a magnet is used to identify the particles emanating from the interaction. While uncharged particles traverse the field without being affected, particles having a positive charge will be deflected to one side and ones having a negative charge to the other. This provides the first information as to the identity of the particle. The deflection of the charged particle depends on the strength of the field, the length of the trajectory in the field and the momentum of the particle. The size and strength of the field can be measured, so by measuring the deflection one can find the momentum. This is the aim of the exercise. Ideally for the experiment the magnetic field should be present, preferably in a direction perpendicular to the trajectories of the particles, without there being any hardware to produce it. And it should not be expensive. On the other hand, with the increasing energy of particle accelerators and the need for greater and greater precision in the measurements, the experiments require stronger and stronger fields in larger and larger volumes. Many of these so-called detector magnets therefore rely on superconducting windings. The aim of the lecture is to describe the features of this class of magnet so that should you ever be confronted with having to design one you will be able to point yourself in the right direction.

2. UNDERSTANDING THE REQUIREMENT

This is the first and probably the most important step in the exercise. You have to really know what the potential user thinks he needs – or rather what he ought to think he needs! In fact, in order to choose the most appropriate design for the magnet, you are strongly advised to understand how the rest of the detector system works, and the experiment would be wise to incorporate you into the decision making structure which optimises the experiment as a whole.

2.1 How the particle is deflected

The force F exerted on a particle of mass m carrying charge q and traveling at velocity v in a magnetic field B is given by Equation (1)

$$\underline{F} = q\underline{v}\times\underline{B} \tag{1}$$

Assuming the field to be uniform and perpendicular to the motion, the particle will follow a trajectory which is the arc of a circle of radius *r* where

$$r = mv^2/|F| \tag{2}$$

The momentum of the particle, p, is equal to mv, so after some simple algebra we can write Equation (3), representing the angle φ through which the trajectory is deflected

$$\varphi \approx L/r = qB_{\perp}L/p \tag{3}$$

where L is the path length in the magnetic field having component $B_{\perp}$ perpendicular to the path, and Equation (4) for s, representing the sagitta of the trajectory

$$s \approx qB_{\perp}L^2/8p \tag{4}$$

In practice the track detectors record the trajectories in the field, and sometimes before and/or after the particle exits the field, so measurements are made of s and φ. The precision with which the momentum can be determined will depend on the precision of these measurements and the magnitude of L and B. We note for future reference that L is more effective than B. It is also clear that to be efficient the direction of the field in the detector magnet should subtend as large an angle as possible to the direction of the trajectories. In fact the ideal field shape for the detector would be a thick spherical shell of field enclosing the point of interaction – which is physically impossible! In the next section we look at some configurations that are realistic and enumerate their strong and weak points.

There are two more definitions that are useful to know in the context of matching the magnet to the experiment. The first of these concerns the "transparency" of material to the passage of elementary particles. This is measured in radiation lengths, where one radiation length is the thickness of material that is required to impede the passage of 1/e of the particles. It is denoted by X_0 and is proportional to $A/\rho Z^2$, where A = atomic weight, ρ = density and Z = atomic number. For plastic one radiation length is about 300 mm, for aluminium it is 90 mm and for copper 14.5 mm. The second concerns the "reach" of detectors in terms of relativistic kinematics; thus pseudo-rapidity η is dimensionless and is defined as

$$\eta = 0.5\ ln\{(E + p_{//})/(E - p_{//})\} \tag{5}$$

where E is the total energy of the particle and $p_{//}$ is its longitudinal momentum. For $\theta = 40^0$, $\eta = 1$, and for $\theta = 15^0$ and 5^0, $\eta = 2$ and 3 respectively.

2.2 Attributes of practical magnet geometries

Basically three configurations can be considered for detector magnets, namely solenoids, toroids and dipoles.

2.2.1 Solenoids

Solenoids are conceptually simple, elegant and very effective. The vast majority of recent 4π detectors at colliders have relied on solenoidal type magnets, producing a cylindrically symmetric field having the same axis as the colliding beams. They are also used for fixed target and cosmic ray experiments. The reason for this is easy to understand: the symmetric 2-D field is fairly uniform and this facilitates reconstruction of the events; there is no material within the field volume to give rise to spurious secondary interactions, and the magnetic forces are relatively easy to contain. In addition, thanks to the large number of magnets that have been made and the experience with their operation, the associated technology is mature and steadily progressing.

The length of the trajectory of a charged particle emanating at zenithal angle (i.e. the angle between the trajectory and the axis) θ from an interaction on the axis of a long solenoid of radius R producing a magnetic field B, is $R/\sin\theta$. The component of field perpendicular to the trajectory is

$Bsin\theta$. Using Equations (3) and (4) the change of angle, $\delta\theta_{solenoid}$, and the sagitta, $s_{solenoid}$ of the trajectory are therefore

$$\delta\theta_{solenoid} \approx qBR/p \tag{6}$$

and

$$s_{solenoid} \approx qBR^2/8psin\theta \tag{7}$$

The "analyzing power" depends on the layout of the detector and is obtained by some combination of measurements of sagitta and changes in angle; this is reduced at small θ because of the finite length L of the solenoid. In fact in a solenoid the momentum is usually analyzed by measuring tracks inside the magnet, and the momentum resolution (δp/p)solenoid scales as follows:

$$(\delta p/p)_{solenoid} \sim psin\theta/BR^2 \tag{8}$$

Within the limits of known technology and transportable size, the cost of the solenoid of length Λ is roughly proportional to ΛR^2B^2. It is therefore clear that, as concerns resolution, it can be preferable to invest in size than in central field.

The basic technology for building superconducting solenoids is established, so it is relatively straightforward to design one for incorporation into a given detector. This will be addressed in another section. This geometry is also adapted to making "thin" magnets through which particles pass with little likelihood of interacting, making for clean reconstruction of the events.

When used on a collider the effect of the solenoid on the circulating beams is small but must not be neglected. The accelerator has to be provided with skew quadrupoles to correct for the coupling of vertical and horizontal betatron oscillations, and the orbit correction scheme must include capacity for compensating the effect of a finite crossing angle of the beams.

2.2.2 Toroids

In theory a toroidal magnetic field is ideal, both for a 4π detector at a colliding beam facility and for forward and fixed target detectors. The field is symmetric and perpendicular to the particle motion. The integrated field along the trajectory of a particle increases as θ decreases, which is favorable because the likelihood of having high energy particles to analyze also increases as θ decreases. Charged particles produced on the axis are deflected in a plane. There is no iron yoke. And there is no field along the axis of the beams.

For an ideal toroid contained between current sheets at radii R_i *(inner)* and R_o (outer) and with field B_i at the inside radius, the deflection, $\delta\theta_{toroid}$ and the sagitta, s_{toroid} of the trajectory within the field, are given by

$$\delta\theta_{toroid} \approx (qB_iR_i/psin\theta)ln(R_o/R_i) \tag{9}$$

and

$$s_{solenoid} \approx qB_iR_i^2/10psin\theta \tag{10}$$

the latter assuming a typical ratio of 3 for R_o/R_i. For toroids the change in angle is usually weighted in favour for the momentum measurement, and the resolution can be expressed as

$$(\delta p/p)_{toroid} \sim p/BL \approx p\ sin\theta/B_iR_iln(R_o/R_i) \tag{11}$$

In this idealized case the optimum resolution is obtained with $R_o/R_i \approx 3$, but this optimum is quite flat and is only reduced by 5% for a ratio of 2; a ratio of 1.5 leads to a reduction of 25%. The optimum is with a ratio of about 4 and more strongly peaked if the sagitta measurement is used [1].

So why have we seen so few toroids? It is not (only) because the toroid is a harder magnet to build (which it is, as we shall see). It is mainly because of the difficulty of making in practice anything resembling an ideal toroid, i.e. to make the inner conductor sheet and its supporting structure sufficiently transparent, or to divide it up so as to cover a sufficiently small proportion of the azimuth, and to make it sufficiently long. This material creates confusion in that some particles are absorbed or interact with the structure, an effect that can easily outweigh the benefits of conceptual elegance. Various studies made over the last 25 years have concluded that the best practical approximation for a high field (and therefore superconducting) toroid, covering the central region of the detector of a collider experiment, consists of a small number of lumped, flat pancake coils (typically eight). With care these can be made to cast shadows over as little as 30% of the azimuth, which can be considered acceptable. It implies, however that a significant fraction of the central detector, in particular the electromagnetic calorimeter, should lie within the inner radius of the toroid.

In point of fact the catalogue of advantages of the toroid compared with the solenoid must be read with circumspection. Iron yokes double as calorimeters; the compensation of the effects of solenoids on the circulating beams is fully understood and is not onerous; and for a lumped racetrack toroidal coil geometry the field is far from uniform. Moreover, the field on the inside conductor of a typical lumped toroid is about four times the maximum field in the useful part of the magnet (compared with about 1.3 times for a typical solenoid), which is obviously highly unfavorable in the case of a superconducting coil.

2.2.3 Dipoles

The momentum resolution of a dipole with field B and length L is of course simply

$$(\delta p/p)_{dipole} \sim p/BL \tag{12}$$

While dipoles are not appropriate for the central part of 4π detectors, they are the magnets of choice for dedicated "forward" detectors that concentrate on the cone of particles emanating in the forward direction, up to $\theta \approx 300$ mrad, as well as for fixed target detectors. These magnets are quite challenging. Because of the large aperture and short length the field is highly 3-dimensional, and in order to control the stray field the geometry of the yoke and coils needs to be carefully studied. In addition, a dedicated compensation scheme must be provided for the circulating beams. The preferred saddle shape for the coils is much easier to achieve with a resistive than with a superconducting coil. Unless a suitable geometry of the field can be achieved using either circular or eventually racetrack coils, it is unlikely that any experiment in today's era could afford a superconducting version. In order to maximize the integrated bending field and minimize the stray field, if the magnet has iron poles these are often made sloping, to match closely the required acceptance and maximize BL.

2.2.4 Other geometries

Alternatives to the dipole for the forward region are the quadrupole and the toroid. The toroid provides better acceptance than the dipole, but suffers from its contribution to background through multiple scattering, and is obviously no use for very small angles. An advantage is that it does not require compensation. The quadrupole is open and is easier to compensate than the dipole, but the field reduces with reducing θ, and it needs to be complemented farther downstream with a dipole.

In particular cases other geometries may be considered. These are almost always based on astute combinations of flat circular or racetrack coils, usually without an iron yoke. The field is strongly inhomogeneous, but can be accurately calculated analytically, and while onerous, tracking through it is feasible using modern computers. The design of the supporting structure for these coils can be quite challenging, and the space required must be taken into account in the design of the experiment.

3. TECHNOLOGICAL ASPECTS OF DETECTOR MAGNETS

Ever since practical superconductors appeared they have been used for detector magnets; they were being used for this purpose long before they were first used in accelerators. There is therefore a wealth of experience which has been built up over the years and which should be tapped if you want to make the most efficient design. Most of this experience has been gained through the building of solenoids, and we shall give most attention to this geometry.

3.1 Boundary conditions

3.1.1 The requirement

It is again stressed just how important it is to clearly establish what is required by the experiment. In order to iterate to a reliable magnet design that is matched to the rest of the experiment this can involve a lot of discussion. Cost and timescale must also be included right from the start.

3.1.2 Operating conditions

It is important to be fully aware of the conditions under which the magnet is expected to operate, as this can have a profound effect on how to go about its design. The vast majority of detector magnets operate in a quasi-DC fashion, for example, and their design can be simplified accordingly. Other important parameters are whether or not the field should be reversible, the expected length of runs, the expected number of hours of operation per year, and the expected lifetime of the magnet.

3.1.3 Cooling supply

What means are there available for cooling the magnet? Is there a refrigerator, and if so what capacity will be available for servicing the new magnet? Would it be expected to cool from dewars? Is a cryo-cooler an option to consider?

3.1.4 Conductor

Detector magnets rely on multifilamentary superconducting material that is in good electrical and thermal contact with a normal conductor of sufficient cross section to bypass the current due to a microscopic perturbation, to allow the affected superconductor to regain its superconducting state or to buy time for the quench protection circuitry to do its job. This stabilizing material is a pure metal with low electrical resistivity and high residual resistivity ratio (RRR). The magneto-resistance of the material must also be taken into account.

3.1.5 Protection

The cross-section of high purity aluminium required for quench protection is given by the adiabatic criterion

$$J^2 = G(\Theta)VI/E \tag{13}$$

where

$$G(\Theta) = {}_0\int^{\Theta max}\{C_p(\Theta)/\rho(\Theta)\}d\Theta \tag{14}$$

and J is the current density in the stabilizer, I is the operating current, V is the protection voltage, E is the stored energy, C_p is the specific heat of the conductor, $\rho(\Theta)$ is its resistivity and Θ_{max} is the peak temperature during a quench. The adiabatic criterion assumes no heat conduction from the conductor hot spot and is therefore very conservative.

The conductor typically works with a temperature margin of 1 - 3 degrees. The coils are secured against movement by impregnation. Quenches should not happen. If they do occur, due for example to energy release in cracking epoxy or a deficiency in the insulation creating a warm spot, then the design must be such as to quickly spread the heating by inducing a general quench and/or dump the stored energy in an external resistor. In this way the temperature rises uniformly throughout the coil,

and the maximum temperature in the coil can be limited to less than 80 K, say, by a judicious combination of

i) Extracting the stored energy.

ii) Firing quench heaters.

iii) Quench back.

iv) Incorporating pure aluminium heat shunting sheets.

The final choice of the conductor is made following detailed calculation of all possible scenarios in the environment of the magnet, including effects such as friction and work of fracture, which can lead to energy releases in the coil. To give an order of magnitude, in recent magnets the stability margin for a sudden release of energy is a few joules, down from about 10 J in the case of the LEP solenoids.

The ratio of the stored energy, E, to the mass, M, of the winding, called the E/M ratio, is a convenient parameter for indicating the "safety" of the design and the attention which has to be paid to quench protection. Based on the results of refined calculation and validating testing, this ratio is being gradually pushed from the range 3 – 5 in past designs to around 10 and more in the most audacious designs today.

Perhaps the most likely place for a quench to occur is in the bus work, due to conductor movement in the magnetic field. Such a quench is hard to detect with voltage taps, and recent practice is to install a thin, insulated, cupro-nickel stabilized superconductor in good thermal contact along the bus, to provide the signal; it works in the same way as a helium level gauge.

3.1.6 Model and prototype work

There is usually just one opportunity to build the complete magnet, and model and prototype work is limited to validation of choices of design concepts and components. It is vitally important to include this essential activity into the general plan. The well being of the whole experiment depends on the successful and timely completion of the magnet, so it is customary to request that the results of the model work be exposed to critical review.

3.1.7 Reliability

Because of the one-off nature of these magnets, large safety factors are usually applied and the construction is followed up with great attention. Any failure risks to be a catastrophic failure, and it will put in jeopardy the whole experiment. These superconducting magnets do however have a history of very good reliability.

3.1.8 Size

Bigger may be better for momentum resolution, but watch out for threshold sizes that lead to a jump in cost. An example of this is the supplementary civil engineering work that may be required to house the magnet. Another limitation is that imposed by transport. Beyond 6 - 7 m in diameter implies in-situ winding.

3.2 Magnet design and manufacture

3.2.1 Magnetic design

The initial design is best explored using hand calculations. The next step may be a partial optimisation using a 2-D finite element program, which can easily be tweaked to give 10% accuracy (and a lot better for solenoidal geometries, which are essentially two-dimensional). Only when you have a good idea of what is required should you plunge into the 3-D modelling, which can involve a heavy input overhead. It is important to couple this magnetic design to mechanical calculations from the start,

because the structure required to support these coils has also to be carefully optimised with regard to mass and cost. It is most important to take possible buckling into account.

3.2.2 Cooling scheme

The first magnets for detectors were bath-cooled or cooled with helium circulating in a cooling channel, analogous to that of a water-cooled resistive magnet. The problems associated with the electrical insulation and the manifolds were overcome, with some difficulty, and the magnets gave good service. However, for a magnet that has to work at a constant field there are no AC losses and it is not necessary to put the liquid helium in contact with the conductor in order to ensure that its temperature remains low enough to work reliably as a superconductor. Detector magnets are virtually DC magnets, and taking advantage of this fact leads to a major simplification in the design, namely the separation of the helium from the electrical circuit, with the cooling ensured by conduction. The advantages of indirect cooling can be listed as follows:

i) All electrical circuits and connections can be in the insulation vacuum, with the consequent reduction in the risk of arcing (beware the Paschen effect).

ii) There are no (or few) electrically insulating sections in the helium cooling circuit, so the complete circuit can be of welded construction with increased reliability.

iii) The volume of helium required to fill the circuit is greatly reduced.

There are of course also some disadvantages:

i) There must be a good and uniform thermal contact between the cooling pipes and the coil, across a layer of good electrical insulation. This calls for thoroughly reliable gluing techniques.

ii) The cooling circuit must be carefully designed to ensure both that the coil can be maintained at close to liquid helium temperature, and that it can also be cooled down in a reasonable time.

iii) Inevitably resistive joints in the winding must either be so good as to dissipate negligible power or be provided with local cooling, which complicates the cooling circuit.

The advantages however outweigh the disadvantages and practically all detector magnets are now indirectly cooled.

3.3 Choice of materials

3.3.1 Conductor

The superconductor is standard NbTi multifilamentary conductor in a copper matrix. It is used either drawn as such or cabled into a flat Rutherford cable. To render the conductor sufficiently cryo-stable to allow practical protection, it is either soldered into a channel in an extruded copper or aluminium bar, or co-extruded with pure aluminium to form the stabilized conductor for use in these magnets. While cryo-stabilized NbTi is and has been the universal choice until now for the superconducting material, one does start to think of HTS (BSSCO 2223 or 2212) for very special applications, especially if a cryo-cooler is favoured. High purity aluminium is the preferred medium for stabilization, because of its low electrical resistivity and high RRR that is not greatly affected by magnetic field, and because it is light in weight. It is however very soft, and in certain applications it needs strengthening: this will be discussed in the section on solenoids. Examples of conductors are shown in Fig. 1, in which we note the introduction of reinforcement in recent years.

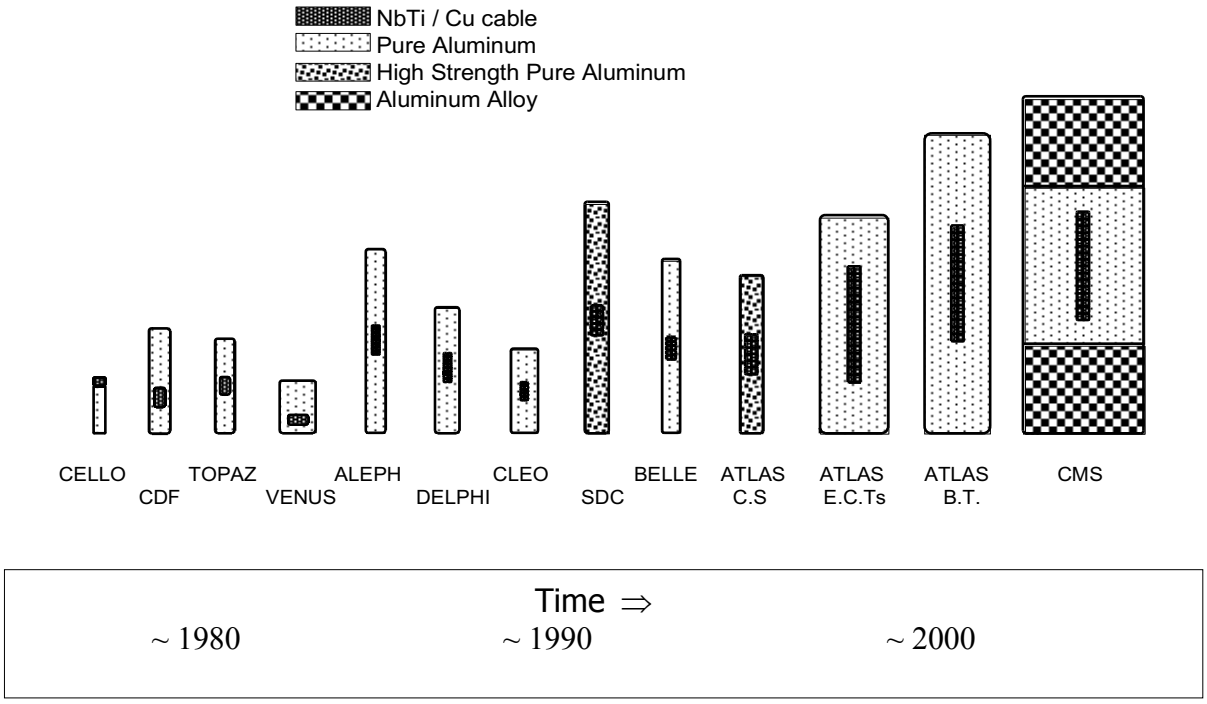


Fig.1 Progress in the development of aluminium stabilized superconductor (courtesy A. Yamamoto).

3.3.2 Supporting structure

As the coil is mainly aluminium, in order to minimize forces due to differential thermal contraction the coil supporting structure is also made of aluminium, alloyed to give structural strength. The preferred alloy is typically 5083. The support of the cold mass to the room temperature vacuum insulation vessel requires a material having a high ratio of strength to thermal conductivity, with titanium or stainless steel being frequent choices.

3.3.3 The cooling circuit

This consists of an aluminium tube welded either to an aluminium sheet in good thermal contact with the coil, or to channels machines into the aluminium coil support structure. The good thermal contact requirement implies vacuum impregnation with epoxy resin, and all of the associated precautions (e.g. voids to be filled with glass fibre) to preclude a build up of resin thickness that may lead to formation of cracks during the cooldown.

3.3.4 Electrical insulation

Glass fibre tape, sometimes in combination with polyimide film, is wrapped around the conductor to provide turn-to-turn insulation and placed between coil and support structure/cooling plate to provide ground insulation. Depending on the geometry of the coil and its size, the resin for impregnating the coil may be pre-impregnated into the glass tape.

3.3.5 Thermal insulation

Having indirect cooling for the coil simplifies the cooling circuit, but it means that particular attention has to be paid to the thermal insulation in order to avoid hot spots. It usually consists of a helium gas-cooled screen, also made of aluminium alloy, wrapped with a blanket of multi-layer super-insulation.

3.3.6 Electrical joints and connections

The conductor should be supplied in long piece lengths to avoid having to make splices in the coil. For large coils this may not be practical and a few splices may have to be accepted: these are made before insulation and winding the conductor, by removing half of the stabilizer from the ends to be joined, soldering the strands, welding or soldering the aluminium, and machining. This must be done over a length of about 20 times the equivalent diameter of the superconductor in order to ensure low resistance. In a coil that is wound in sections these can be connected simply by welding the aluminium stabilizer of side-by-side tails over a suitable length (at least 20 times the equivalent diameter of the whole conductor). The connections to the current leads are made by removing the aluminium and soldering the bared copper of the strands to the terminals.

3.4 Geometry specific features

We now pass to features of the design that depend specifically on the magnet type. The parameter lists of the LHC detector magnets, as well as other details including pictures, and the progress in their manufacture and assembly can be viewed on the Web [2].

3.4.1 Smooth solenoid

Smooth is used here to describe a continuous solenoid in contrast to one made up of a succession of thick, lumped coils, which will be treated in the category of "Other geometries". The coils are wound on the inside of a cylindrical aluminum alloy former and the radial magnetic force is taken by the winding itself, in combination with the cylinder. The coil is indirectly cooled by supercritical helium circulating in tubes welded to the cylinder. The solenoids for the LHC differ from previous solenoids for colliders, however, in that, due to higher fields, there is higher stress in the winding and it has been necessary to strengthen the conductor.

CMS is a general-purpose detector optimized to run at the highest luminosity at LHC. It is designed with a single solenoid magnet surrounded by a return yoke acting as the external muon spectrometer [3]. The solenoid coil is designed to provide a field of 4T in a clear bore of 6 m and a magnetic length of 12.5 m, and the operating field is to be 3.5 ~ 4 T. The maximum stored energy is 2.6 GJ; the coil cold mass is 225 tons. This is three times the stored energy of the BEBC magnet, which is the largest superconducuting magnet ever built, and more than an order of magnitude greater than previous solenoids for collider experiments. The coil design has been made drawing on previous experience, but has required the following new features:

i) It is self-supporting, obtained by use of reinforced conductor allowing a strain of 0.15 %.

ii) It has four layers, and will be assembled on-site from five transportable modules.

The aluminium-stabilized superconductor is reinforced with high strength aluminium alloy, attached by electron-beam welding. This innovative technique, which has required significant R&D to ensure that the superconductor is not damaged by overheating during any phase of the process, allows an overall hoop stress of 105 MPa at a strain of 0.15 %. Coil winding is done using the "inner winding" technique. The coil will be cooled indirectly using two-phase helium flow in cooling tubes on the support cylinder, with circulation ensured by the thermo-siphon technique, following the experience with ALEPH. This is one of the magnets for which E/M is expected to go beyond 10, as can be seen in comparison with other solenoid magnets in Fig. 2.

The iron yoke and end caps of the CMS magnet are assemblies of iron plates weighing a total of 10'000 tonnes, and have also been an engineering challenge for the magnet designers. These components are instrumented as hadron calorimeters, but are considered to be part of the magnet. The work required for the design of the flux return circuit should not be underestimated.

The ATLAS central solenoid magnet is to provide an axial magnetic field of 2 T in a warm bore of 2.3 m in diameter in the central tracker region [4,5]. Since the coil is placed in front of the liquid-

argon electro-magnetic calorimeter, the solenoid coil must be as thin and transparent as possible for the best calorimeter performance, with minimum interaction of particles in the coil. The solenoid coil design features:

i) High-strength aluminum stabilized superconductor with uniform reinforcement.

ii) Pure-aluminum strip technique for uniform energy absorption in the thin coil.

iii) A cryostat shared with the liquid argon (LAr) calorimeter, to save on wall material.

Since the hadron calorimeter acting as magnetic return yoke is located far enough away, the coil is designed to be an iron-free, air-cored solenoid. Extensive efforts have been made to reinforce the aluminum stabilizer while keeping sufficiently low electrical resistivity of the aluminium which is characterized by a residual resistivity ratio (RRR). An optimum solution has been found by using a combined process of "micro-alloying" and "cold-work hardening"[6]. Together with the use of a common cryostat with the LAr calorimeter, this has made it possible to achieve a total radiation thickness of only 0.66 X_0. The coil is to be cooled by using two-phase forced flow helium with a possible back-up option of thermo-siphon cooling.

3.4.2 Toroids

While many studies of superconducting toroids have been made, until recently none ever reached the final approval stage, so there has not been the same progression in experience specific to the design of magnets with toroidal geometry as there has been for the solenoids. In addition to the solenoid described above, the ATLAS experiment for the LHC will employ three toroids: a Barrel Toroid and two End Cap Toroids as shown schematically in Fig. 2.

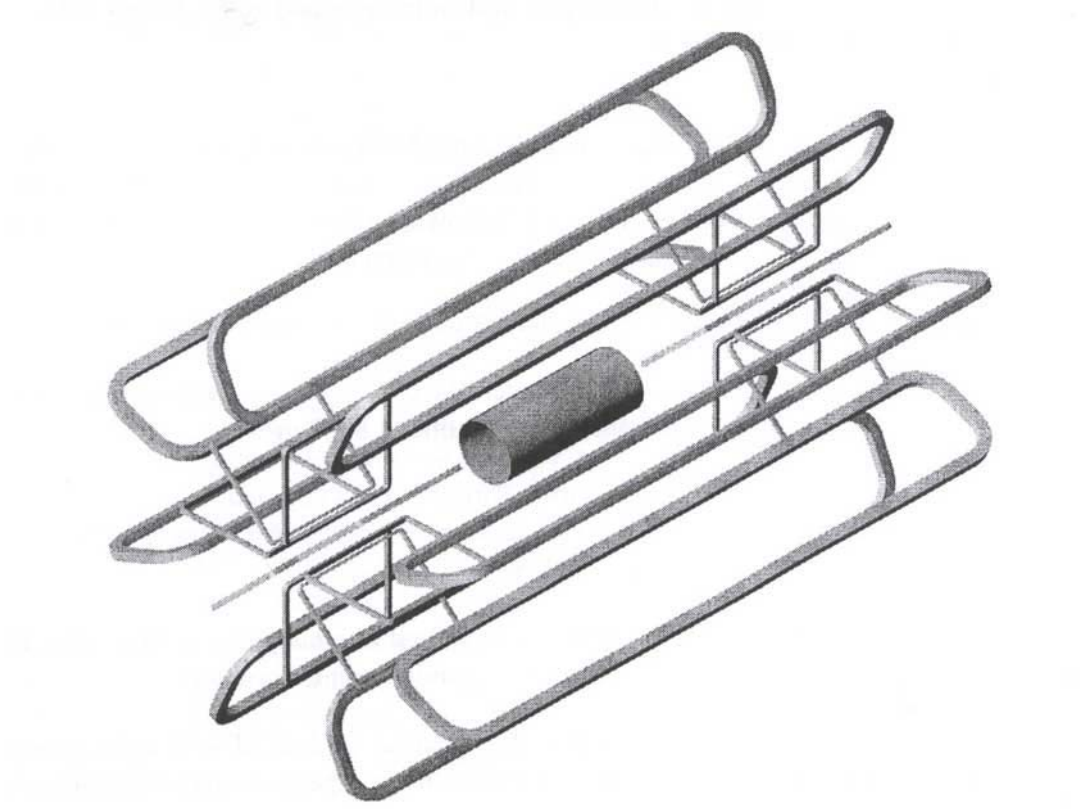


Fig. 2 Schematic view of the arrangement of coils forming the ATLAS magnet system.

The Barrel Toroid (BT) consists of eight coils assembled symmetrically around the beam axis [7]. A double pancake coil is wound in racetrack shape and placed in an individual coil case to support the electromagnetic force. Every coil is installed in its own cryostat and mechanically linked together with 8 support rings from the warm structure to maintain mechanical stability. The ATLAS BT has an inner bore of 9.5 m, outer diameter of 20 m and length of 25 m. It provides a deflection power (BL) of 2 - 6 Tm depending on the angle. The air-cored toroid with individual cryostats allows the installation of muon chambers in the magnetic field and provides measurements of both the sagitta and the deflection. The stored energy of the BT is about 1300 MJ.

The End Cap Toroid (ECT) is also optimized with eight coils, with its azimuthal position rotated by 22.5° with respect to the BT to allow for radial overlap and to optimize the bending power in the interface region of both coil systems [8]. In contrast to the BT, the eight coils are assembled to be a single cold mass. The double pancake coils wound in the coil cases are linked by using a cold web structure in a single aluminum cryostat with the inner radius of the coil placed as near as possible to the beam axis to provide the best coverage of the forward angles. This creates greater simplicity in mechanical support of the magnetic force. On the other hand, it does not allow installation of the chambers in the toroidal field area, so the momentum measurement relies entirely on a measurement of deflection produced by the field integral of 4 - 7 Tm over the pseudo-rapidity range 1.5 to 3.0. The stored energy of each ECT is about 240 MJ.

Besides differing from the ideal smooth toroid by virtue of their lumped coils, both BT and ECT feature pancakes that are very thick in the radial direction, which reduces the peak field on the inner leg. Whereas the coils of the BT feature the same cross section for the inner and outer legs, the coils of the ECT feature a grouped outer leg and extensive spreading of the conductors forming the inner leg, thereby improving the utilization of the superconductor. These magnets have required extensive mechanical engineering studies concerning both the securing of the coils in their casings and the external structures which must safely support the magnets against a variety of forces: gravity, and three categories of Lorentz force – within the coil, between the coils and between the toroids. The asymmetric forces due to offsets and to quenching or one or more pancakes have also had to be thoroughly understood in the framework of the integrated design of these magnets.

Both the BT and the ECT are to be cooled by means of a forced flow of two-phase helium, provided by cold pumps.

3.4.3 Other geometries (dipoles etc.)

The ideal geometry for a dipole detector magnet is the window frame magnet. The problems with this are the support of forces on the flat current sheets and the flat-sided vacuum vessel, and the complicated winding and cryostat shapes due to the upturned coil ends. A more practical dipole would feature a pair of racetrack coils in individual cryostats surrounding the iron poles; another possibility is a large bore version of an accelerator-type cosθ magnet. Both of these latter types have been made, but the complication is such that, except in very specific cases, it is better to consider a normal resistive magnet. Both the Alice and LHCb experiments at the LHC started by considering superconducting dipoles, and gravitated to resistive versions after seeing the likely costs – and this despite the fact that they consume about 4MW apiece.

Other arrangements that can be considered are groups of thick lumped coils. One proposal for a collider featured a single circular coil coaxial with the beams and centered on the collision point. The experiment took advantage of the open nature of this geometry and although non-uniform the analytic 2-D nature of the field renders the on-line analysis of tracks relatively easy to cope with. The next possibility is a Helmholtz pair of coils, with or without an iron yoke: this will provide a more uniform field, but with poorer access due to the structure required to keep the coils apart. Several such coils can also be assembled to form a solenoid: in certain experiments the possibility of access between the coils may outweigh the disadvantages as compared with a smooth solenoid. The unfavourable field

direction of a solenoid in the far forward region can be offset by the incorporation of a long, tapered iron "nose" with a central hole for the beam pipe. The iron will saturate but will nevertheless suck in sufficient flux to improve significantly the momentum resolution at small forward angles. For muon spectroscopy the nose could also be magnetized azimuthally and instrumented. As concerns magnets for use in space, ensembles of two (or more) thin solenoids in conjunction with racetrack coils to cancel the magnetic moment would appear to have promising features [9]. There will usually be a definite cost advantage in designing with circular coils, in simple symmetrical arrangements.

4. CASE STUDY – THE ECT

If this section we go through a typical technical proposal for a detector magnet to give an idea of the work involved in its preparation. For this illustration the case of the ATLAS end cap toroid has been chosen, but it could be any other such magnet.

4.1 Requirement, design constraints, design concept

What the experiment needs in terms of momentum resolution and coverage; space restrictions; and the design concept based on experience and affordability, together with indications of where R&D is required. The design is based on a system of 8 SC coils each manufactured as a pair of single layer pancake windings

4.2 Magnetic design

Using OPERA 2- and 3-D, calculate peak fields, stored energies, inductances, magnetic forces and field integral performance. Iterate.

4.3 Conductor

Choose a conductor operating at 60% of critical current along the peak field load line. The peak field is 4.4 T, and the chosen operating current is 20 kA. Choose a reasonable temperature margin, ~2 K. Choose a convenient cable geometry to meet this requirement, which is Icrit(5 T, 4.2 K) = 50 kA, and gravitate to 36 strands of 1.3 mm diameter with Cu:SC =1.35. Verify that piece length needed (1 km) is reasonable. Calculate required cross-section of aluminium – 7 mm x 70 mm would allow extrusion according to experience and be adapted to the coil winding. Choose RRR on cost (but > 500). Do quench analysis.

4.4 Cold mass

The cold mass consists of the 8 coils together with the supporting structure. Between the coils use webs to take the inward forces of up to 600 t per coil, and cross braces to stabilize the structure against magnetic and gravitational loads. The cold mass weighs about 130 tonnes. Elaborate a coil design concept. A coil assembly consists of a centre plate with a coil winding on either side, completed with two outer plates in a sandwich-like structure. The winding formers are fixed to the centre plate. The whole boxed coil including ground insulation in impregnated with epoxy. The coils have clamp bars fitted around the outer circumference. Cooling circuits are attached to the outer plates. Perform complete F.E. analysis and iterate. Do bonding tests to choose epoxy. Make calculations of the thermal system, with regard to cool-down, temperature gradients, steady state etc. Iterate. Study the manufacturing sequence and tooling requirements.

4.5 Cryostat

The design is based on a single large cryostat, comprising vacuum vessel, radiation shields, super-insulation, service turret, coldmass supports and cryostat mounting rail. The ECT is to be inserted into the end of the BT, so a castellated shape has been adopted for the vacuum vessel. The outer vessel consists of 16 prefabricated shell units the central bore tube is a smooth cylinder which also supports shielding. Make a detailed structural analysis with comfortable constraints on stresses (< 50 MPa) and deflections (< 5 mm), and taking into account all loads, including the axial load due to interaction with the BT, and offsets. Iterate until reaching a satisfactory proposal.

4.6 Assembly and installation

The ECT is too large to be transported to CERN as a single unit, so the design has to be modular. Choose components to minimise manipulation and manufacturing steps at CERN. Establish space requirements for assembly and testing. Study how to move the complete tested units from the test area to the pit. Iterate.

4.7 Cryogenic system

It is proposed to allow 4 weeks to cool an ECT to 4.5 K. To achieve will require that the coils and radiation shield be pre-cooled to 100 K with high pressure helium gas (cooled by liquid nitrogen) at 35 g/s and 12 g/s respectively, maintaining a constant temperature differential of 40 K. Cooling from 100 K to 4.5 K will take less than 3 days. Continuous cooling will require circulating 2-phase helium by means of cold pumps.

4.8 Vacuum systems

The volume to be evacuated is very large (384 cubic metres), and the area of super-insulation is such that outgassing will give a high load during the initial pump down. The stray magnetic fields must also be taken into account when choosing the type of pump an its location.

4.9 Instrumentation

Transducers are needed for control and monitoring. System parameters to be monitored are Temperature, forces, voltages/currents, cryogen flows and pressures, magnetic fields, and coldmass/toroid positions.

4.10 Electrical systems

The current is the same in the 3 toroids, so it has been decided to power them in series. Besides the power converter, the circuits include a slow dump circuit for running down the current under normal or certain fault conditions, and a fast dump circuit for fast energy extraction. The latter would be activated if there were to be a quench or other serious system fault such as failure of the pumped 2-phase helium supply or a massive vacuum failure. The dump system comprises circuit breakers and a large external resistor.

4.11 Fault analysis

Fault analysis is an important part of the design of these magnets, and particularly in the case of toroids in which the quenching of one or particular combinations of the lumped coils can give rise to unpleasant asymmetric forces. This has had to be studied in detail, and structures dimensioned to take the buckling forces in the worst scenarios. The analysis also draws attention to the importance of stringent quality assurance throughout manufacture.

REFERENCES

[1] T. M. Taylor, “The choice of spectrometer magnets for LEP”, Physica Scripta, Vol. 23, 459-464 (1980)

[2] http://cern.web.cern.ch/CERN/

[3] A. Herve et al., “Status of the CMS magnet (MT-17)”, IEEE Transactions on Applied Superconductivity, Vol. 12, No. 1, 385-390 (2002)

[4] H. ten Kate, “The superconducting magnet system for the ATLAS detector at CERN”, IEEE Transactions on Applied Superconductivity, Vol. 10, No. 1, 347-352 (2000)

[5] A. Yamamoto et al., “Progress in ATLAS central solenoid magnet”, IEEE Transactions on Applied Superconductivity, Vol. 10, No. 1, 353-356 (2000)

[6] K. Wada et al, “High-strength and high-RRR aluminium-stabilized superconductor”, IEEE Transactions on Applied Superconductivity, Vol. 10, No. 1, 1012-1015 (2000)

[7] A. Dael et al., “Synthesis of technological developments for the B0 model coil and the ATLAS barrel toroid coils”, IEEE Transactions on Applied Superconductivity, Vol. 10, No. 1, 361-364 (2000)

[8] E. Baynham et al., “Engineering status of the superconducting end cap toroid magnets for the ATLAS experiment at LHC”, IEEE Transactions on Applied Superconductivity, Vol. 10, No. 1, 357-360 (2000)

[9] A. Yamamoto, “Superconducting magnets advanced in particle physics”, Nuclear Instruments and Methods in Physics Research, Section A, Vol. 453, 445-454 (2000)

A THIN SUPERCONDUCTING SOLENOID MAGNET FOR BALLON EXPERIMENTS IN SPACE SCIENCE

Akira Yamamoto
KEK, High Energy Accelerator Research Organization, Tsukuba, 305-0801, Japan

Abstract
A very thin and transparent superconducting solenoid magnet is being developed to investigate cosmic-ray antiparticles in the Universe. The uniform magnetic field is provided in a cosmic-ray particle detector system, BESS-Polar, for high-altitude long duration ballooning at Antarctica. The coil is wound with high-strength aluminum stabilized superconductor which has been recently much advanced in mechanical strength while keeping electrical resistivity low enough for the coil stability. The progress of the high-strength aluminum stabilized superconductor and the development of the thin solenoid magnet are described.

1. INTRODUCTION

The Balloon-borne Experiment with a Superconducting Solenoid Magnet **S**pectrometer, BESS, has been carried out, as a US-Japan space science cooperation program, since 1993 [1-2]. Figure 1 shows a picture just before launching of the BESS spectrometer, in northern Canada, 2002. It aims at highly sensitive search for cosmic-ray antiparticle and for the novel primary origins such as evaporation of Primordial Black Holes (PBH) initiated in the early history of the Universe [3.4]. The search for antihelium to investigate the asymmetry of particle/antiparticle composition in the Universe is also a major fundamental question in the history of the Universe [5].

Fig. 1 Balloon launching of the BESS superconducting magnet spectrometer. It provides a central magnetic field of 1T maintained constant in persistent mode.

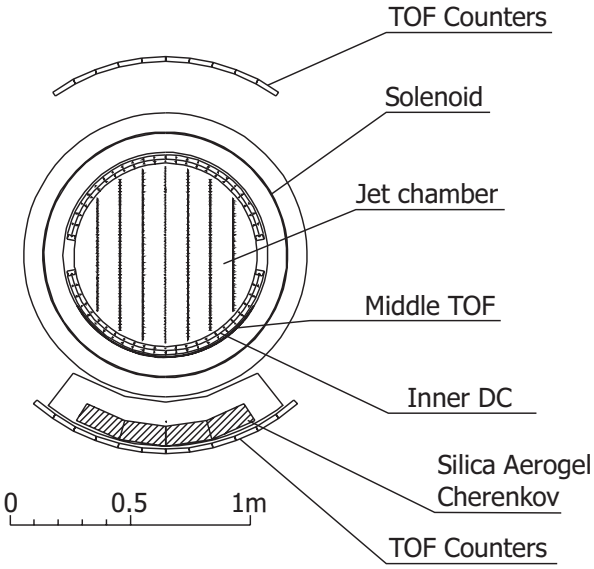


Fig. 2 Cross section of the BESS-Polar spectrometer with a thin superconducting solenoid magnet.

The BESS spectrometer features a cylindrical detector configuration with a thin and transparent superconducting solenoid magnet. Most cosmic ray particles can freely traverse the spectrometer [6,7]. The unique cylindrical configuration enables the spectrometer to have a large geometrical acceptance of ~ 0.3 m^2.sr efficiently in a compact spectrometer volume. The BESS spectrometer with a total weight of 2 ton has been launched by using a large balloon, and flown at high altitudes above 36 km, in northern Canada, as an ideal location to measure low energy cosmic rays with the minimum effect from the earth's magnetic field which comes in/out in the vertical direction. Seven flights have been successfully carried out so far, measuring, for the first time, the low energy spectrum of antiprotons, and searching for antihelium at the highest sensitivity ever achieved [8.9,10].

To extend the scientific objectives in this ideal experimental environment, the "BESS-Polar program" is being prepared to realize long duration balloon flights around the south pole in Antarctica, where one can expect the maximum flux of low-energy cosmic-ray charged particles coming down to the earth, and continuous day-light helps guarantee very stable floating of the balloon [11]. The BESS-Polar spectrometer is being developed to achieve an ultimately transparent spectrometer to maximize the sensitivity and the resolution of the low energy particles and a very compact spectrometer to be launched in Antarctica.

The BESS-Polar thin superconducting solenoid is designed to provide a central magnetic field of 1 T in a cylindrical volume of 0.9 m in diameter and 1 m in length with a half wall material density of < 1 g/cm^2 for most of cosmic-ray particle to pass through without interactions in the coil wall, and to allow a continuous operational time period of ≥ 2 weeks in floating. An acceleration load (shock) of 10 G must be taken into account in the design against dynamic ballooning conditions. Figure 2 shows a cross section of the general BESS-Polar spectrometer. The thin solenoid coil is being developed as a key component in the spectrometer. This seminar gives a brief introduction for the thin solenoid and fundamental technology such as high strength aluminum stabilizer, as well as progress in development of the extremely thin superconducting solenoid to be launched for long duration flights at Antarctica.

2. THIN SOLENOID MAGNETS

2.1 Fundamentals

The concept of the thin superconducting solenoid magnet is to provide maximize magnetic field with minimum amount of material in the coil and cryostat. The radiation thickness, X, is a useful scale the

thinness and transparency of the solenoid coil and it is determined with an exponential decay (1/e) of electron energy loss in its passing through the material [12]. In large-scale solenoid magnets, it is primarily determined by the thickness of material for the conductor stabilizer and the coil support cylinder. As a simple example, the mechanical hoop stress and necessary thickness for an ideal solenoid coil (with no radial field component) are given by

$$\sigma_h = t^{-1} \bullet r \bullet p = t^{-1} \bullet r \bullet B^2/2\mu_0$$

$$X = t / X_0 = \{r \bullet B^2/2\mu_0\}\sigma_h^{-1} \bullet X_0^{-1}$$

where p is magnetic pressure, t is coil thickness, r is coil radius, B is magnetic field, σ_h, is hoop stress, and X_0 is one radiation length of the material, respectively. In case of the air-core solenoid having the radial field component, additional axial stress, σ_z (~ $1/2\sigma_h$), needs to be taken into account. In a simply conservative approach, a stress intensity, $\sigma_{SI} = \sigma_h - \sigma_z \sim 1.5\sigma_h$, should be taken into account in the necessary coil thickness. The strength of the aluminum stabilizer and support cylinder is required to be sufficiently stronger than σ_h or σ_{SI}. In an ultimately thin solenoid design as described below, an approach can be made by using the so-called "equivalent stress", given by $\sigma_{eq} = [\{(\sigma_h - \sigma_z)^2 + \sigma_h^2 + \sigma_z^2\}/2]^{1/2}$, which should be typically 10~20 % lower than σ_{SI}.

The thickness and the stress of the coil are related with a ratio of the electromagnetic stored energy, E, to the effective coil mass, M, defined by

$$E/M = \{\int B/2\mu_0 \, dv\} / \{d \bullet V_{coil}\} \sim \sigma_h / 2d$$

where d is effective density of the coil to be able to absorb the stored energy. The E/M ratio in the coil approximately equivalent to "enthalpy" of the coil, H, and it determines the average coil temperature after the stored energy absorption during a quench as follows

$$E/M = H(T_2) - H(T_1) \sim H(T_2).$$

where T_2 is average coil temperature after the full energy absorption in the quench, and T_1 is the initial temperature. The parameter of E/M ratios of 5, 10, and 20 kJ/kg approximately correspond to ~ 65, 80 K, and 100 K, respectively, as shown in Fig. 3.

From a viewpoint of superconducting stability, the small resistivity, ρ, of pure-aluminum at 4.2 K is a very important advantage. A useful parameter is the so-called "minimum quench energy" (MQE) of the superconducting coil. In a simplified one-D model this may be characterized by

$$MQE = \gamma C \bullet \Delta T \bullet A (J_n \rho)^{-1} \bullet \{2 L_0 T \bullet \Delta T\}^{1/2}$$

where γC is the specific heat of the stabilizer, ΔT is temperature difference between normal state and superconducting state, A is cross sectional area of stabilizer, J_n is the current density in the stabilizer after normal transition, L_O is the Lorentz number ($=2.4 \times 10^{-8}$ $\Omega W K^{-2}$), and T is the average temperature during the transition. The lower resistivity contributes to increase the stability and reduces the amount of stabilizer needed in the coil.

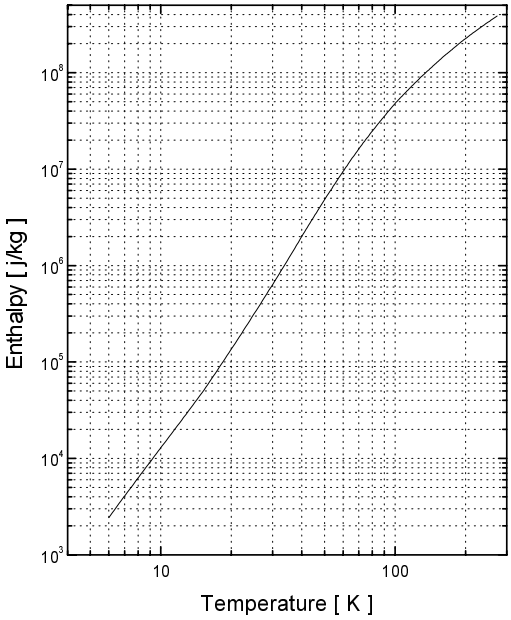


Fig. 3 Enthalpy of aluminum as a function of temperature.

In summary, the coil thickness scales with $r \cdot B^2/\sigma_{SI}$ or $r \cdot B^2/(E/M)$, and the aluminum stabilizer is required to be mechanically strong but with low electrical resistivity at 4.2 K (i.e. having large residual resistivity ratio, or RRR).

2.2 Aluminum stabilized superconductor

Aluminum stabilized superconductor has been developed since the1970s, and widely used in large-scale detector magnets for high energy physics [12, 13]. Since1980s, a special effort has been devoted to the development of high strength aluminum stabilized superconductor having low resistivity at 4.2 K, in a series of R&D programs for ASTROMAG [14], SDC [15], ATLAS Central Solenoid [16, 17], and BESS-Polar [18,19]. The development was focused on achieving uniform reinforcement of the conductor/stabilizer in order to minimize the strain in the coil and maintain mechanically fully elastic conditions. Ideally it should be as strong as the support cylinder.

The high strength aluminum stabilizer has been developed using a combined technology of "micro-alloying" with "cold-work hardening" [12, 18]. The micro-alloying contributes to mechanical reinforcement and its long-term stability against annealing effects in subsequent processes such as coil curing. A list of possible additives for this micro-alloying is summarized in Table 1, in comparison with impurities to be eliminated. Some additive metals such as Mg, Si, Cu, and Zn have a common feature of contributing little to electrical resistivity under the so-called "solid-solution" condition. The additive provides internal stress, and weakly traps dislocations in the base metal. It contributes to the mechanical reinforcement with smaller contribution to the resistivity. On the other hand, a metal such as Ni has a different feature of much lower solubility. Beyond this threshold, the additive is crystallized and/or precipitated according to the process in the base metal as shown in Fig. 4 [18].

Table 1. Possible additives into pure Al in micro-alloying process

Additive	A	d [g/cm^3]	Solubility [% wt]	Resistivity contribution (in solution / crystallization) [10^{-12} Ω.m/wppm]	
• *solid solution:*					
Mg	24	2.4	17.4	0.54	0.22
Si	28	2.6	1.65	0.7	0.088
Cu	64	8.9	6	0.35	0.03
Zn	65	7.1	83@ 400C	0.10	0.023
• *crystallization/precipitation*					
Ni	59	8.8	0.05 @640C <0.006 @<500C	0.81	0.061
•*major impurities to be eliminated*					
Cr	52	6.9	0.72	4.4	0.18
Mn	55	7.2	1.5	3.5	0.34
Fe	56	7.9	0.05	2.6	0.058

The crystallized/precipitated additive creates barriers against dislocation with little increase of resistivity. Therefore, the crystallized/precipitated condition allows a much large amount of additive for mechanical reinforcement while keeping the low electrical resistance of the alloyed metal. Cold work hardening with permanent deformation plays a major role in the mechanical reinforcement. The strain hardening results from the generation of lattice defects and dislocations during plastic deformation. The cold work predominantly reinforces the aluminum.

The development progressed first with the solid-solution approach by using Si (in ASTROMAG) and Zn (in SDC). The further advances have been successful with the crystallization/precipitation approach by using Ni (in ATLAS). Figure 5 shows how the reinforcement has progressed in these various approaches. The aluminum stabilizer with 0.1 % Ni gives a yield strength of 110 MPa at 4.2 K and RRR = 590 ($\rho = 4.7 \times 10^{-11} \Omega m$ @ 4.2K, 0 T), after 21 % cold reduction and curing at 130 °C for 150 hrs. Aluminum stabilizer with Ni 0.1 % becomes as strong as typical copper stabilizer, having YS = 100 MPa, and RRR = 100, and provides much better stability and is lighter. A record yield strength of 167 MPa was achieved with Ni of 2 % for RRR = 190. An overall yield-strength of 240 MPa was achieved with NbTi/Cu contribution for the BESS-Polar Solenoid as described below.

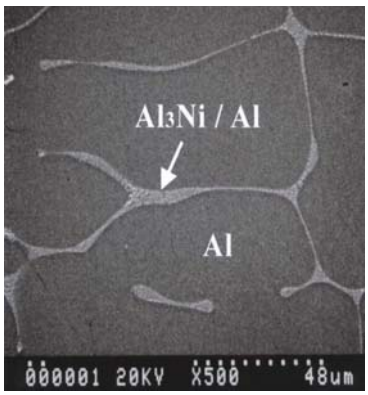


Fig. 4 Enlarged view of aluminum stabilizer alloyed with Ni of 0.5 %.

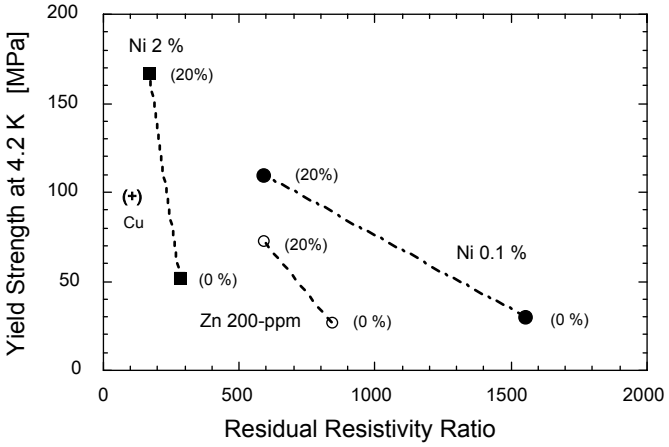


Fig. 5 Yield strength of Al-stabilizer compared with ordinal Cu –stabilizer.

2.3 Fast quench propagation by using pure aluminum strips

When a quench occurs, full energy absorption into the coil needs to be taken into account for safety in the worst case design. In order to eliminate excessive thermal stress in the coil, the stored energy needs to be absorbed as uniformly as possible into the coil with fast quench propagation. A technique to increase the axial quench propagation velocity has been developed by using pure aluminum strips in the BESS balloon magnet project [26]. If adiabatic conditions are assumed, the longitudinal (along the conductor) quench propagation velocity V_ϕ is given by

$$V_\phi = (J/\gamma C) \bullet \{L_0 T_s / (T_s - T_0)\}^{1/2}$$

where J is the current density, γC is the volumetric specific heat, L_0 is the Lorentz number, T_s is the wave-front temperature and T_0 is the initial operational temperature [20]. The relative axial (transverse to the conductor) velocity is given by

$$V_z/V_\phi = \sim \{k_Z / k_\phi\}^{1/2}$$

at low temperature, as shown in Fig. 7, and this helps to enhance the axial quench velocity and thus reduce the peak temperature in the coil

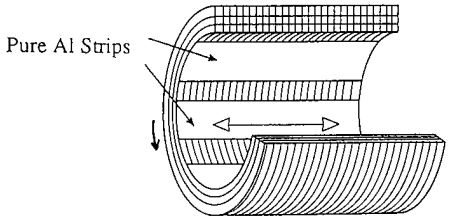


Fig. 6 Concept of Al-strip quench propagator.

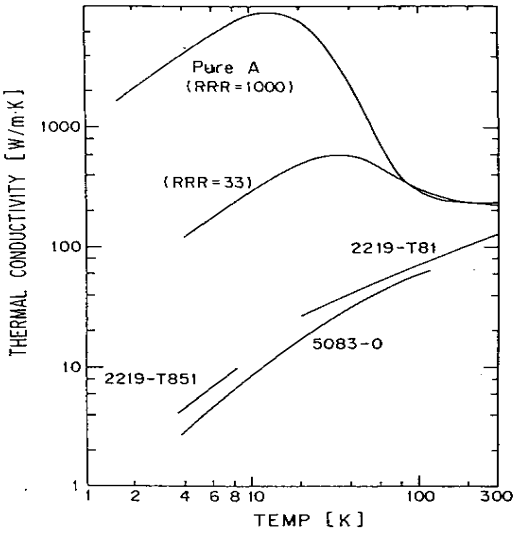


Fig. 7 Thermal conductivity of pure-aluminum and aluminum alloys.

2.4 E/M Ratio

The E/M ratio is a useful parameter to scale the relative transparency of the coil as described before [12, 21]. Figure 8 shows E/M ratios in various detector solenoid magnets. The E/M ratio of magnets using pure aluminum stabilizer is usually around 5 kJ/kg. The SDC prototype using high-strength aluminum stabilizer achieved approximately 10 kJ/kg, with an experimental result of a peak coil temperature of 95 K in the full energy dump test, and the BESS-Polar prototype model has reached ~13 kJ/kg with a peak coil temperature of 150 K, as described later.

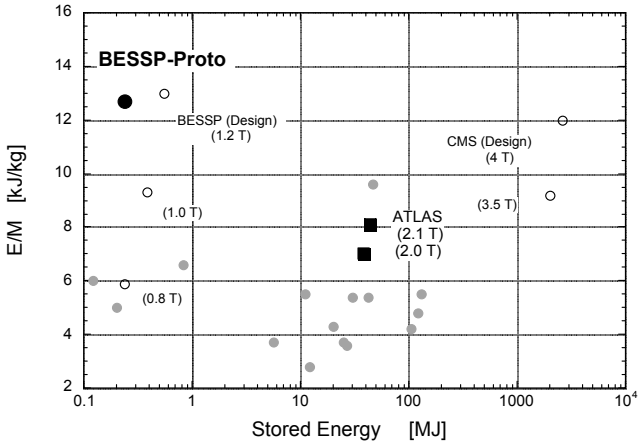


Fig. 8 E/M Ratio as a function of stored energy in various solenoid magnets in particle detectors.

3. BESS-POLAR SOLENOID DESIGN

3.1 General Design

The BESS-Polar thin superconducting solenoid coil has been designed by using the high-strength aluminum stabilized superconductor described above. The coil and conductor design are described in [xx]. A two-layer solenoid coil is designed to provide a central field of 1.2 T with a field uniformity of $dB/B \leq \pm 9$ % in a particle tracking volume of 0.75 m in diameter and 1 m in length. The coil has a radial thickness of 3.4 mm and a mass density of 1 g/cm^2 corresponding to a radiation thickness of 0.056 X_0. Including the cryostat, the total material becomes 2.3 g/cm^2, i.e. a radiation thickness of 0.1 X_0.

Figure 9 shows the cross section of the BESS-Polar superconductor with a cross sectional area of 1.1 mm x 0.8 mm before insulation. The superconductor has been optimized with a Ni content of 5,000 ppm with an area reduction of 15 %. A yield strength of 103 MPa at 4.2 K with RRR = 125 was obtained after a heat cycle up to 130 °C. When the contribution from NbTi/Cu conductor is included, the overall yield strength reaches 240 MPa, which allows the coil to be fully self-supporting. The high-strength aluminum stabilized superconductor enables a simple and very thin coil design. The superconductor is strong enough for the solenoid coil with two layers to be fully self-supporting.

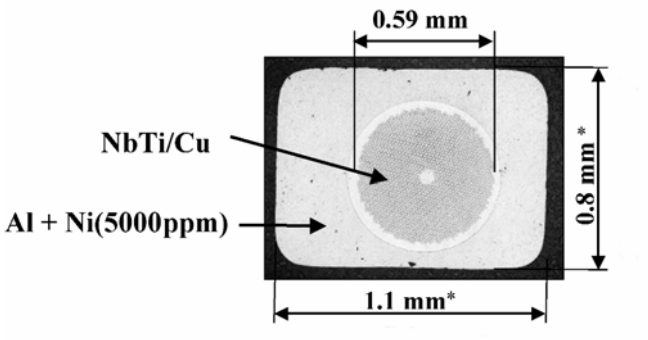


Fig. 9 Cross section of the aluminum stabilized superconductor for BESS-Polar thin Solenoid.

The inner layer solenoid coil is wound with the conductor in the vertical position (edge-wise) for the whole length. The outer layer is wound with the conductor in the horizontal position (flat-wise) in the central region, while the axial ends are wound in the vertical position (edge-wise) for an optimization that minimizes the material in the central region and shapes the field. The maximum field in the coil is calculated to be 1.72 T for a central field of 1.2 T, at 571 A. It corresponds to a load-line ratio of 72 % at 4.2 K. The equivalent stress of ~180 MPa (estimated by the simplified equation above) is supported by the conductor itself. A finite element analysis shows a consistent result of 174 Mpa. This is well below the measured yield strength of 240 MPa at 4.2 K. Assuming an effective coil density of 3,500 kg/m^3, an E/M of 13 kJ/kg is to be achieved in the BESS-Polar solenoid design. During the ballooning period, the coil will be operated at a maximum of 1.0 T to ensure stability under various dynamic conditions, with acceleration loads of up to 10 G.

Both inner and outer surfaces of the coil were covered with thin axial pure-aluminum strips with a thickness of 0.1 mm in order to provide conductive cooling for the coil and to enhance axial quench propagation. It gives a maximum coil temperature of ~ 180 K after the absorption of full energy into the coil in case of a quench. High-strength aluminum shells are placed between two layers to align the coil layers and to reinforce the axial rigidity and stability against the axial magnetic force of the coil. This component also fulfils the function of connecting the coil both mechanically and thermally to the LHe reservoir.

The coil is placed in a cryostat and cooled with thermal conduction through aluminum shells/strips thermally linked to a LHe reservoir at one end. The cryostat consists of this LHe reservoir, radiation shield, vacuum vessel, and mechanical support of the coil and shield. The outer vacuum vessel design is based on the use of a corrugated cylinder to minimize the wall thickness while maintaining an appropriate buckling pressure of 3.5 bar. Based on the design described above, a total radiation thickness of 0.11 Xo including the cryostat has been achieved.

4. DEVELOPMENT

The development has been carried out in three stages, basic R&D, a prototype, and the main solenoid. The prototype coil was wound with the same cross section as the main coil, but having half in the length. The excitation test of the prototype coil was carried out in the vertical position in a test cryostat. The coil reached 696 A corresponding to 120 % of the full current, with the first spontaneous quench. It resulted in the peak coil temperature ~ 165 K with the full safe energy absorption in the coil. The mechanical stress in the coil reached the level corresponding to a central field of 1.2 Tesla in the main coil. This ensures technical feasibility of the main solenoid to run at a maximum of 1.0 T during the long duration balloon flights. The main coil winding has been completed as shown in Fig. 11.

The BESS-Polar spectrometer is to be launched in 2004 in Antarctica or Arctic polar circle, as shown in Fig. 12.

Fig. 11 The BESS-Polar thin solenoid coil winding completed.

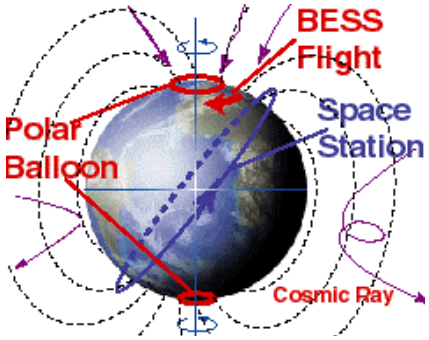


Fig. 12 Polar ballooning compared with the current BESS flight in northern Canada. The earth magnetic flux lines show how cosmic rays come into the earth's environment.

5. SUMMARY

A very thin superconducting solenoid magnet is being developed for long duration, high altitude ballooning at Antarctica. The technology of high-strength aluminum stabilized superconductor has enabled the achievement of a solenoidal field of 1 T in a magnetic volume of 0.9 m in diameter with a wall material density of 1 g/cm^2 (or a radiation thickness of 0.06 Xo), and with an E/M ratio of 14 kJ/kg. The total material including the cyostat is achieved to be 1.1 Xo. The performance has been successfully demonstrated in a prototype model. The technology of high-strength aluminum stabilized superconductor would be a promising approach for superconducting magnets to be applied in future astro-particle physics experiment in space.

REFERENCES

[1] S. Orito KEK Report 87-19, Proc. of ASTROMAG workshop, pp. 119 – 125, (1987).

[2] A. Yamamoto et al., "Conceptual design of a thin superconducting solenoid for particle astrophysics", IEEE Trans. Mag., 24, pp. 1421-1424, (1988).

[3] A. Yamamoto, et al., "Balloon – borne experiment with a superconducting solenoidal magnet spectrometer", Adv. Space Res., 14 29, pp. 75-82, (1994).

[4] Y. Ajima, et al., "A superconducting solenoidal spectrometer for a Balloon-Borne Experiment", Nucl. Instr. Meth., A 443 71, (2000).

[5] K. Yoshimura, et al., "Observation of cosmic-ray antiprotons at energies below 500 MeV", Phys. Rev. Lett., 75, pp. 3792-3795, (1995).

[6] S. Orito et al., "Precision measurement of cosmic-ray antiproton spectrum", Phys. Rev. Lett., 84, pp. 1078-1082, (2000).

[7] T. Maeno et al., "Successive measurements of cosmic-ray antiproton spectrum in a positive phase of the solar cycle", Astroparticle Phys., 16, 2, pp. 121 – 128, (2001).

[8] J.F. Ormes, et al., "Antihelium in cosmic rays: A new upper limit and its significance", Astrophysics J, 482, L187-L190, (1997).

[9] T. Saeki et al., "A new limit on the flux of cosmic antihelium", Phys. Lett., B442, pp. 319-324, (1998).

[10] A. Yamamoto et al., "BESS progress and future prospect towards long duration flights", Adv. Space. Res., 30 , (5) p1253.(2002).

[11] A. Yamamoto et al., "A thin superconducting solenoid for particle astrophysics", IEEE Trans. Applied Super. Vol. 12, p438.(2002)

[12] A. Yamamoto et al., "Development towards ultra-thin superconducting solenoid magnets for high energy particle detectors", Nucl. Phys., B (Proc. Suppl.), 78, pp. 565-570, (1999).

[13] K. Wada et al., "Development of high-strength and high-RRR aluminum-stabilized superconductor for ATLAS thin superconducting solenoid", IEEE Trans. Appl., super., 10, No. 1., pp. 1012-1015, (2000).

[14] A. Yamamoto, "Superconducting magnets advanced in particle physcis", Nucl. Instr. Meth., A 453, pp. 445 – 454, (2002).

[15] I. L. Horvath et al., "The CMS conductor", IEEE Trans. Applied. Super., Vol. 10, No. 1, pp. 395-398, (2000).

[16] Y. Makida et al., "Development of an ASTROMAG test coil with aluminum stabilized superconductor", IEEE Trans. Mag., Vol. 27, No. 2, 1944 - 1947, (1991).

[17] M. Wilson, “Superconducting Magnets”, Oxford University Press.

[18] Y. Makida et al., “A thin superconducting magnet for balloon borne experiment”, to be published in Int. Conf. Space Technology and Science. Shimane, Japan, (2002).

VERY HIGH FIELD MAGNETS

Lucio Rossi
University of Milan – Physics Department, Italy
CERN, LHC Division, Geneva, Switzerland

Abstract
In the 1970s superconductivity dramatically changed the prospects of circular accelerators, allowing the construction of accelerators with a magnetic field greater than 2 tesla. This culminated with four large particle accelerators built or under construction in the last thirty years. By using NbTi, magnets with an operative field spanning from 3 to 9 T have already been successfully built. The LHC has plans to use this technology to accelerate protons to 7 TeV/beam. To attain higher energy in hadron colliders a change in magnet technology is envisioned that will require an incremental increase in the dipole field to arrange dipole fields between 11 and 15 tesla. This paper will review the status of conductor development, magnet design and the issues associated with needs for high energy machines.

1. INTRODUCTION

Since the beginning when superconductivity [1] was utilized to enhance magnet technology the push has been on to achieve higher fields. There is a general consensus in the community that a next step in energy, at least for heavy particles, needs to be accomplished by an increase in both the dipole magnetic field, see Fig.1, and the accelerator size. In the case of the SSC, an optimization study based on the freedom of choice for both a new tunnel and a new site showed that a field considerably above the previous projects (6.5 T against the 4 T Tevatron field) was needed. At the time of the LHC study, in the late 1980's, the possible use of higher field magnets was also discarded as too expensive [2]. However, the recent VLHC proposal, with the freedom of a new tunnel, envisages an increase in tunnel length and an increase of magnetic field, even at the cost of utilizing new Nb_3Sn technology

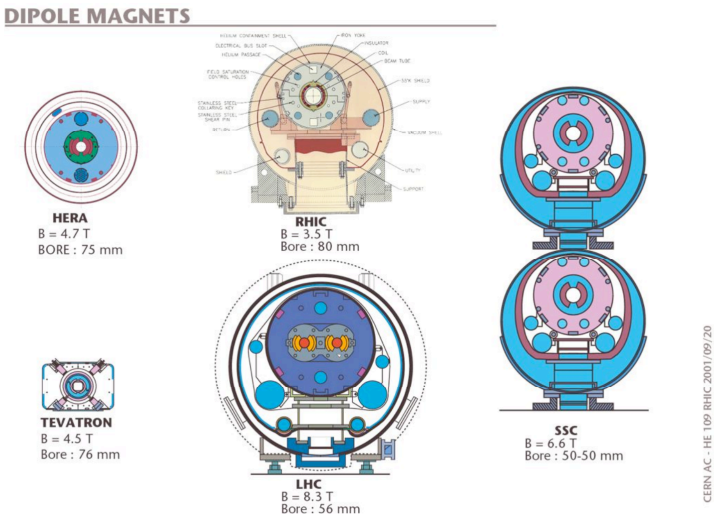


Fig.1 Cross section of the dipoles of the main projects based on superconducting magnets

As we shall see there are reasons other than the final beam energy that push the community to develop high field magnets: beam focusing, beam damping, etc. The HEP community of course has to take maximum profit from the technology developed for other applications of superconductivity (high field laboratory magnets, NMR high field magnets, thermonuclear fusion reactor, etc.). In this paper we shall concentrate on the issues that are particular to HEP accelerators, and consider applications requiring high field magnets (i.e. more than 10 T at 1.9 K or more than 8 T at 4.2 K or higher temperatures).

2. THE DEMAND FOR HIGH FIELD

1.1 High energy hadron colliders

The push toward high magnetic fields is the consequence of the following relationship, valid for relativistic particles:

$$\mathbf{E \approx 0.3\ B\ R},$$

where E is the particle energy in TeV, B is the bending (dipole) magnetic field in tesla and R is the bending radius in km. We have to consider that a tunnel, see Fig. 2, cannot be completely filled with dipole magnets and that a typical filling factor of bending field length to total tunnel length is 2/3 or less. This provides room for acceleration sections, beam focusing sections, beam preparation prior to interaction regions, beam dump etc.. Therefore if we take as an example the LHC configuration, the total length of 27 km reduces actually to an effective magnetic length of 17.5 km, with a bending radius of 2.8 km in the 1232 dipoles. This implies that going beyond 10 TeV will require an operative field of about 12 T.

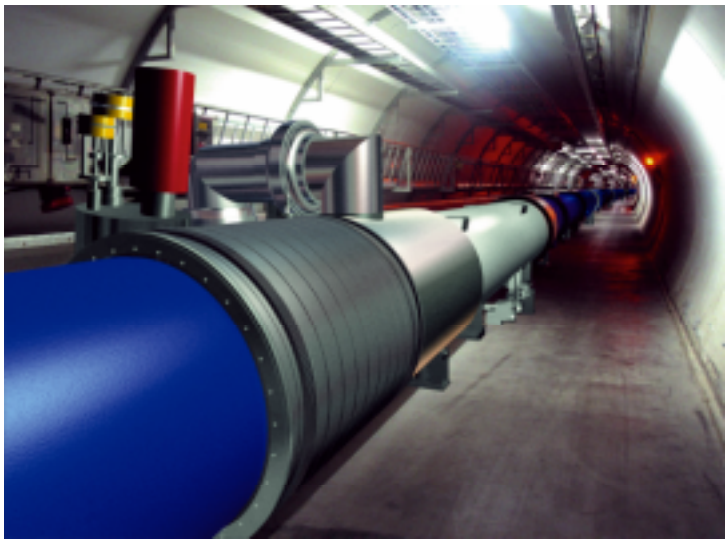


Fig.2 LHC tunnel showing the basic repetition string: one 6 m-long straight section, main quadrupoles with other magnets (sextupoles, octupoles and correctors) and cryogenics services, alternated with three 15 m-long main dipoles. (darker vessels).

It is important to note that hadrons radiate very little compared to electrons, so the gain in energy has to be accomplished with circular accelerators. However the radiation level will substantially increase above the level of the LHC, which is the first hadron collider with non negligible synchrotron power emission. The power emitted poses the following challenges:

The power need to be removed, via a suitable thermal screen, see Fig.3. The presence of a beam screen precludes the design of magnets with very small bores. (which would be very desirable for cost reasons).

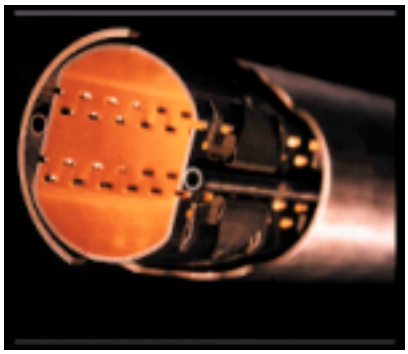


Fig.3 Beam screen in the present LHC cold bore tube.

The electron beam cloud phenomenon [3], see Fig. 4, can pose serious problems which may require different technology than that presently used in the present LHC.

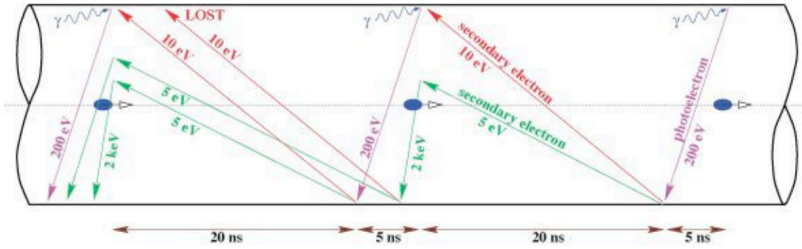


Fig.4 Schematics of the generation of the electron cloud phenomenon: synchrotron radiation photon emitted by the beam extract photoelectron from the wall, with possible subsequent avalanche effect (in some conditions of beam and synchronization).

When comparing the potentiality of such machines with others we have to take into consideration that hadrons are not point-like and a 10 TeV/beam collider is approximately equivalent to a 1□TeV/beam electron-positron collider. The collisions are much dirtier for hadron colliders than for electron colliders, but the discovery potential is superior.

1.2 High luminosity colliders

Luminosity is the other parameter, in addition to energy, that determines the performance of a collider. The collision optics can be improved if powerful quadrupoles are employed. For example, in the present LHC, the so-called low-beta triplet, i.e. the four quadrupoles nearest the interaction region, have peak fields of about 8 tesla, quite comparable to the main dipoles. It is foreseen that in order to improve the luminosity performance of the LHC, a second phase, using higher gradient (G > 250 T/m) or larger aperture quadrupoles (Coil bore > 90 mm) will have to be considered. This implies that the peak field in the coils will be in excess of 10 T, a value that certainly requires the use of very high field conductors and techniques, similar to the requirements for high field dipoles for high energy colliders.

According to the optic design, powerful dipoles, with fields ranging from 10 to 15 tesla and large apertures (coil bore > 90 mm) may be needed to separate the two beams after crossing. In this case certain schemes, very advantageous from the beam dynamics point of view, require that these separation dipoles should be able to withstand a massive heat deposition due to the debris of the collision itself: the power to be removed from the magnet can be of the order of 5-50 W/m, which is a major challenge in itself, even more difficult than achieving very high fields. In Fig. 5

one possible solution for the LHC luminosity upgrade requiring high field magnet technology is shown.

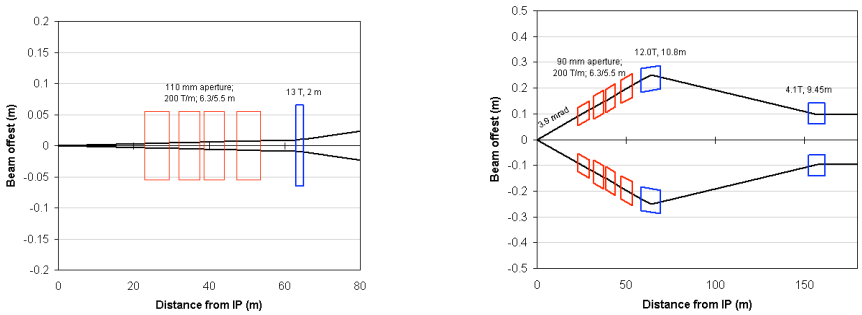


Fig.5 A possible new schematics of the LHC upgraded luminosity insertion regions. To the left the scheme is similar to the present LHC but with stronger magnets, to the right there is a completely new scheme with strong quadrupoles and very strong separator dipoles (from the American proposal LARP: courtesy of J. Strait, FNAL).

Actually, the demand for powerful quadrupoles is not limited to circular accelerators: very high energy e^+e^- colliders in the TeV range may need powerful large aperture quadrupoles with high peak fields in the coils. Indeed the Tesla Project foresees the use of quadrupoles with apertures and gradients similar to the main quadrupole of the LHC. Gradients must be obtained without iron to enable the insertion of the quadrupole within the detector itself, to be as close as possible to the interaction point. Since the detector will be equipped with a large solenoid, the solenoidal field would saturate the iron and generate forces which would be difficult to handle. Moreover, the ironless quadrupole must work at 4.2 K, so the gradient cannot be obtained with present LHC technology; such fields under such condition require the use of high field technology. The sketch of the detector is shown in Fig. 6.

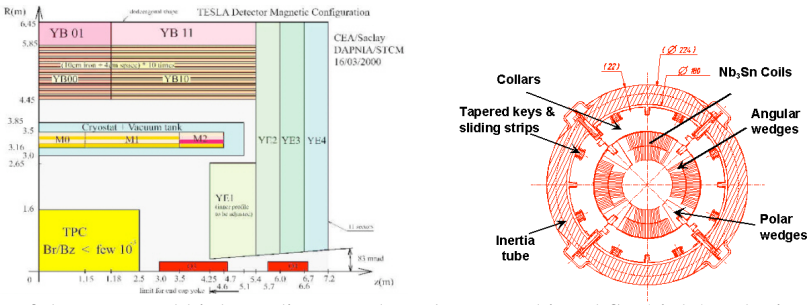


Fig.6 Sketch of the proposed high gradient quadrupoles wound in NbSn (right) to be inserted in the detector for TESLA (left). The two boxes on the axis below the iron cap are the two ironless final focusing quadrupoles: note that they start 3 m far, only, from the interaction point! (courtesy of F. Kircher, CEA, Saclay).

1.3 Muon colliders

For the past few years a fraction of the HEP community has been studying the possibility of reaching high energy by acceleration of muons [4]. Like electrons they are point-like, with all the advantages for detector technology of a clean interaction, but being 200 times more massive they radiate much less. In principle it is possible to reach the multiTeV by means of a ring of 0.5-2 km in diameter, using all the technology of large hadron colliders. A Muon collider is a machine that poses immense challenges, especially in the generation and handling of the muon and its decay products. Even a neutrino factory [5] see Fig. 7 which can be seen as a first step towards a real muon collider is a very difficult machine, but appear to be a good candidate as a benchmark for such technology.

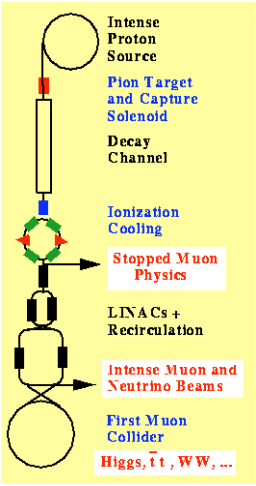


Fig. 7 A possible scheme for a neutrino factory.

In such accelerators there are two domains where high field magnet technology is required [6]: the source of muons is a target bombarded by an intense proton beam where muons are generated and immediately trapped and focused. Here solenoidal fields of the order of 20 T, with bore of 0.5-1 m are necessary! To make things even more complicated, the solenoid must be heavily shielded against radiation, and must be designed to be capable of absorbing (and/or removing) kW of radiation without endangering its cryogenic operation.

The main ring of the accelerator needs dipoles and quadrupoles with fields that vary according to the particular design between 5 to 12 T. What makes high field technology appealing, even in intermediate fields of 7-8 T, is the fact that such dipoles require very large apertures, more than 100 mm, to provide room for proper shielding against radiation and for other reasons such as beam dynamics. Superior field properties and higher critical temperatures that provide a lot of stability against heat deposition are typical in superconductors such as Nb_3Sn which justify their higher cost above the classic NbTi.

3. ADVANTAGE FOR THE SYSTEMS OF USING HIGH FIELD

Going toward high fields has at least one advantage from the beam dynamics point of view, since we enter a synchroton radiation dominated region, when B > 10-11 T. Indeed, the beam damping time scales as $1/(B^2 \rho E)$, where B is the dipole field and E is the beam energy. In the case of a fixed radius (as a possible upgrade of the LHC), one can read the relation as $\tau \mu 1/B^3$.

This advantage is not without its costs: indeed the synchrotron radiation must be removed before it impinges on the magnet performance, operating at cryogenic temperatures, where removing 1 W corresponds to about 500-1000 W at the plug. Handling of this may be the most challenging issue of accelerators dominated by synchrotron radiation.

The American project VLHC [7], where it has been proposed that a 200 km long tunnel would be filled with magnets operating at 10-11 T has taken this into account in a very brilliant way: by deciding to have moderate high fields with an aperture of approximately 45 mm, and a large curvature radius of more than 20 km, they let the radiation exit ahead almost collinear with the beam, without touching any shield. Then, as the radiation escapes the magnets and the distance between beam and its synchrotron radiation becomes measurable, the radiation is intercepted by an absorber which remains in the vacuum at room temperature, making the heat removal much easier, see Fig. 8. This mechanism also has the advantage of drastically reducing the problem of the electron cloud effect, triggered mainly by electrons emitted by the beam tube when it is hit by a photon.

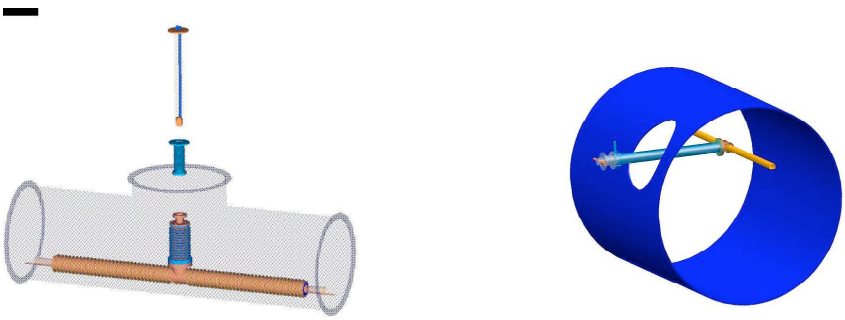


Fig. 8 VLHC radiation absorber put, at the edge of the magnet.

4. DETECTOR MAGNETS

Historically, detector magnets have been the first place where superconductivity has been applied in real devices, for HEP. However, the push for high field is limited and no one has ever proposed fields in excess of 8-10 T for such a device. The reasons are two:

A detector magnet is usually a single magnet system of large if not huge, dimensions. Atlas with its 25 m long barrel toroid coils (see Fig. 9) is a major example of this [8]. CMS, a more classical solenoid surrounded by iron yoke, also features considerable dimensions, with an inner diameter of 6 m for a length of 12 m [9]. This size, and volume make it difficult and costly to construct using high field superconductors. Indeed a magnet of similar (even bigger) size of 12 T is imagined for ITER, whose cost exceeds that of the entire LHC project and whose complexity and size would make it impossible to run physics experiments (where the magnet, important as it can be, is only a service to the various detector layers and their electronics that are installed in the global detector).

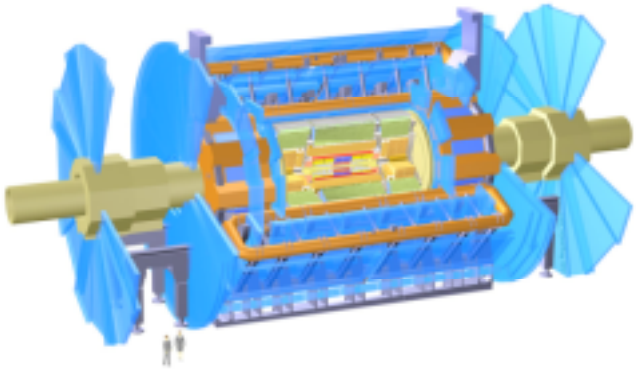


Fig. 9 The Atlas detector with its huge barrel toroid.

Magnetic fields provide electric charge identification, and momentum spectrometry. If the spectrometry is carried out by tracking the particles in the field, one can show that the momentum resolution scales as $\Delta p/p \propto 1/BL^2$, with L being the length of the trajectory measured in the magnetic field. This shows clearly that a larger volume is more effective than the use of higher magnetic fields, and explains why detectors are usually so huge. The field is however weaker than in accelerator magnets, ranging from 1-2 T of toroids like Atlas (where however the coil peak field is about 4 T) to the 4 T central field of CMS (where the peak field is almost 6 T). In the detector of the TESLA project the solenoid that is envisioned is very similar to the CMS one and can be considered as a size limit for detector magnets.

5. ADEQUATE SUPERCONDUCTOR

5.1 The first request: current density

Well before modern synchrotron-collider magnets were proposed in the seventies, the basic idea of modern conductor dominated magnet was proposed [10]. Looking at Fig10 one notes that in

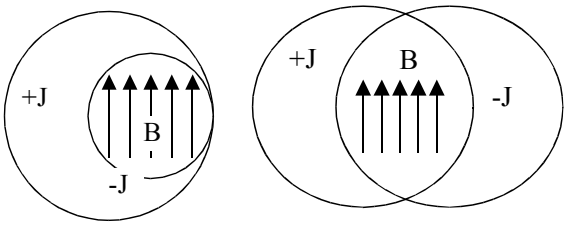


Fig. 10 Cylinder with opposite current, equivalent to cosθ net current distribution (right), generating perfect dipole field in the inner current free region.

the current free region, i.e. where the two opposite currents flow, the perfect dipole field is created, whose magnitude depends on the current density J and on the distance between the two centers, *d*: $B = \mu_0 Jd/2$. It is worth noticing that such a field is exactly half of that generated by a solenoid with the same thickness *d* and the same current density: $B_{sol} = \mu_0 Jd$. In other words, a dipole field of 13-15 T is certainly not less difficult than an NMR magnet of 23 T (the GHz dream!).

Magnets with a current cross section like the one in Fig11, are called cosθ magnets, and if we want B = 10 T in a 50 mm bore this requires J = 400 A/mm2. But J cannot be the superconductor J, since the coil volume is composed of different materials, including voids to be filled with resin or cryogen. Without going into details, a good breakdown of J for High Field Magnets is as follows:

35% superconductor cross sections

35% stabilizer (protection as well)

15% copper wedge

15% insulation and voids (to be filled with resin, if impregnated)

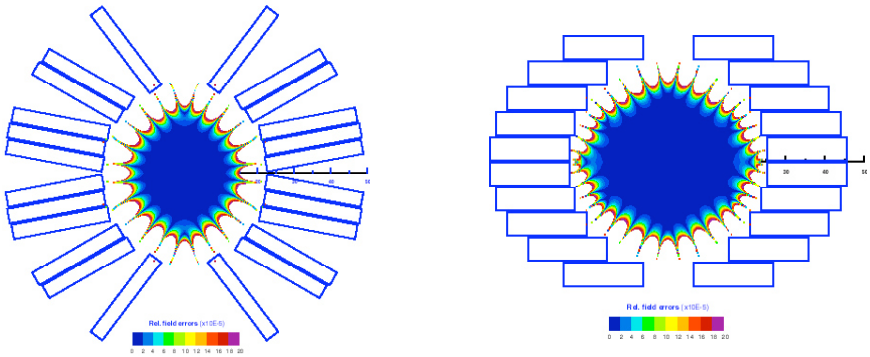


Fig. 11 Practical approximation with cables of a cosθ current distribution. (Courtesy of A. Zlobin, FNAL)

Therefore, for an ideal magnet we need at least 3 times the engineering (or overall) current density of 400 A/mm^2. Typically 20-30% "inefficiency" of the design must be taken into account, since cables can never perfectly fill the cosθ profile. Therefore the superconductor element must operate at 4 times higher J_{sc} than the overall J.

Without going into details, we can introduce the concept that the critical current density of the superconductor, J_c, at short sample field and temperature must be at least ~ 40% higher at the

operating point: that means we need about a 20% margin on the load line, i.e. the short sample of the 10 T magnet must be around 12 T.

Putting all these considerations together we may conclude that for a 10 T operating dipole magnet requires J_c ≈ 2200-2400 A/mm^2.

The Fig 12 shows the J_c vs. field for the best industrial NbTi and Nb_3Sn and HTS (this last however still in prototypal phase), at 4.2 K. Clearly, such a goal is intrinsically beyond the possibility of NbTi. Of course in principle one can compensate low J with thicker windings (for dipoles, whereas for higher order like quadrupole this is not true even by principle). But this is impossible in practice for two reasons:

- it is not economical
- the coil becomes too thick and the mechanical integrity degrades as the e.m. forces are restrained much too far from where they are generated.

In another plot we compare the performance for all current practical conductors, with an emphasis on NbTi as it is widely used and reliable, Nb_3Sn has shown these characteristics in an

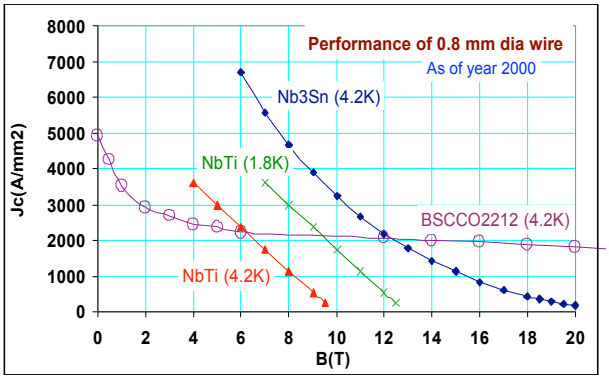


Fig. 12 Best critical current density for Industrial NbTi and Nb_3Sn and for prototypes of Bi-2212.

industrial environment but not in large quantities, while BSCCO2212 is still in experimental phase and only one company is really putting an effort into it.

There is no doubt that today and probably for the remainder of this decade, the only practical way to build a real magnet beyond 10 T is by using Nb_3Sn.

5.2 Progress in Nb_3Sn conductors

Around 1995 the available J_c for Nb_3Sn was around 1300 A/mm^2 available in ITD conductor (Internal Tin Diffusion) or MJR (Modified Jelly Roll). That value was less than the record J_c = 1600 A/mm^2 obtained by the Dutch ECN in 1993 with PIT (Powder In Tube) conductor. That current density was not widely publicized because of the line of fabrication was discontinued. Nb_3Sn current densities are usually quoted at 12 T and 4.2 K, a reference used by the HEP magnet community.

After several years of R&D conductors with current densities around 1800-1900 A/mm2 became available. In 1998 the US started a program sponsored by the Department of Energy, aimed at developing a conductor with a current density of 3000 A/mm^2. In Fig 13 the 2001 results are reported for various materials vs. field (EM reached those values in 1998 and then discontinued the activity on Nb_3Sn).

Today, conductors at J = 2500 A/mm^2 are available, an increase by a factor two in less than one decade.

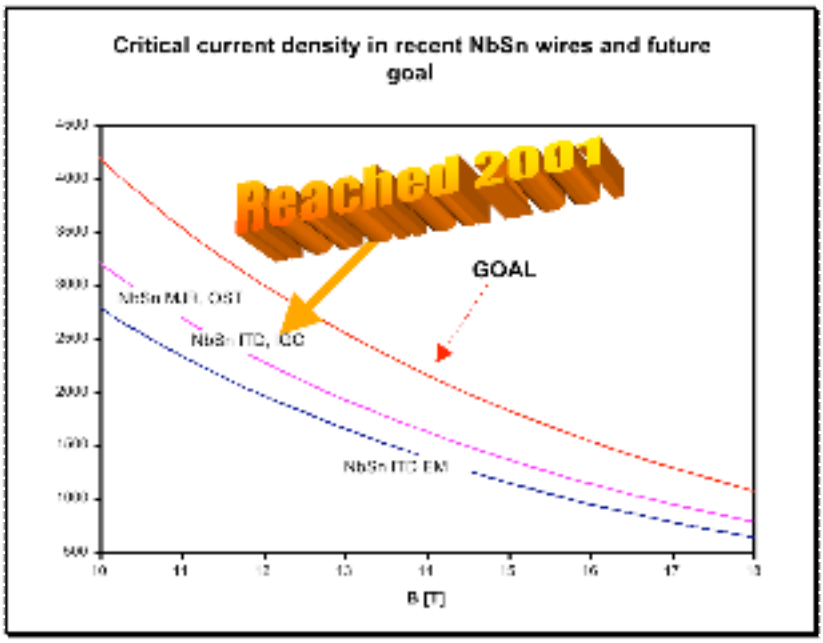


Fig. 13 Best J_c for different companies and goal set in 1999 by the DOE R&D program.

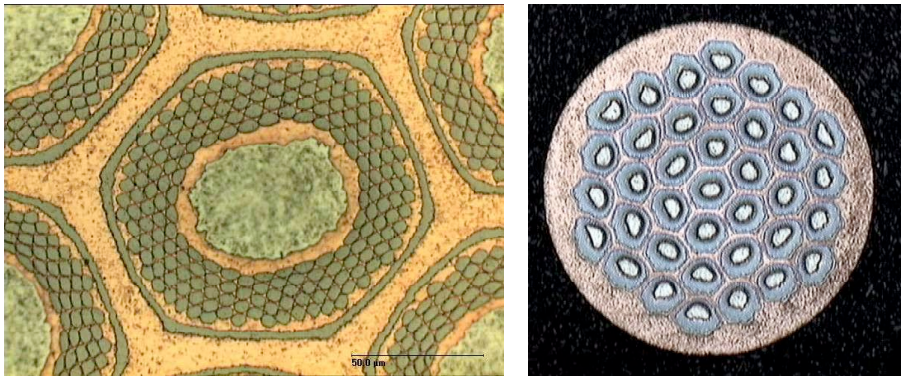


Fig. 14 Cross section of a wire and a zoom over a sub-element of the Nb_3Sn developed in 1996-98 by EM for an INFN-LASA (Milan) program.

Although the rush for high J_c, was fully justified, the conductor development is far from being finished. Other parameters such as the effective filament size are very important in accelerator magnets since they control magnetization effects and therefore the remnant field. By comparison the LHC NbTi conductor has a filament diameter of 6-7 micron. It is commonly understood that a future high field hadron collider will utilize filament size of the order of 20 micron, maybe 40 micron, but not more. All the results quoted above are with materials built via techniques that lead to filament size of 100 microns or more. A noticeable exception is the PIT, a conductor program that has re-started at the University of Twente,(NL) and could potentially reach 2500 A/mm^2 and 20 micron filament size.

In Fig. 14, 15 and 16, pictures of the various wires developed for High Field in recent years are reported.

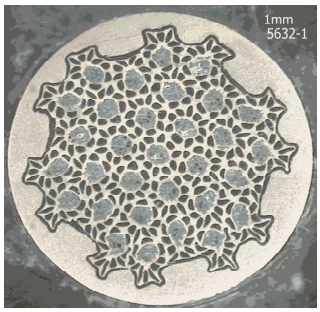


Fig. 15 Cross section of a wire developed by OST for the DOE program (Courtesy of R. Scanlan, LBNL).

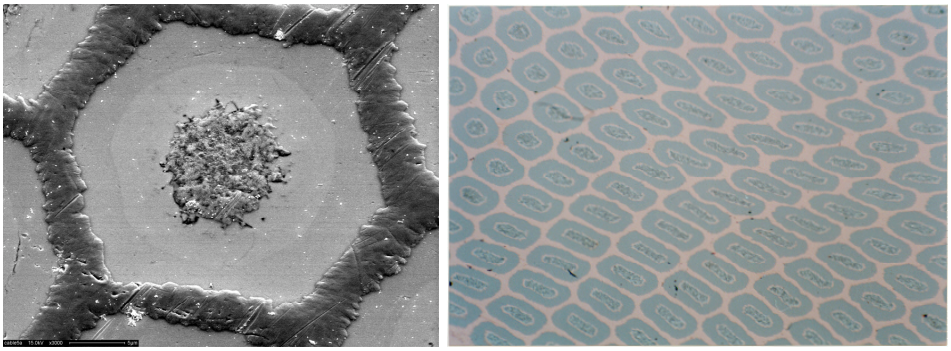


Fig. 16 Cross section of the wire developed by SMI (NL) with the PIT technique (courtesy of A. den Ouden).

Finally, it is worth mentioning the encouraging very preliminary results on making Rutherford cables with HTS materials. Rutherford cables have a geometry that requires a high compaction factor, leading to heavy deformation of the strands. However, as shown in Fig. 17 a good cable has been made at LBNL by cabling Bi-2212 strands. This resulted in a final current density of 2000 A/mm^2 after careful heat treatment.

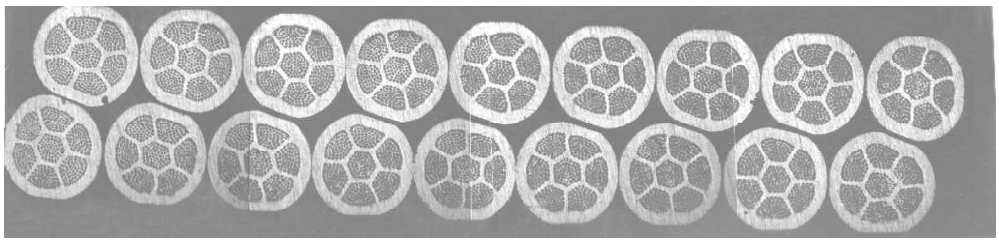


Fig. 17 Cross section of one of the first successful cable done by R. Scanlan (LBNL) with Bi-2212 wire from Showa (J).

An excellent and updated review of progress in the high field conductor for HEP can be found in [11].

5.3 Cost Issue for Conductor

One concern for high field conductor is its cost. The following data are applicable today (always referred at 12 T, 4.2 K) [12]:

Nb_3Sn (Internal Sn)--\$5.48/kA-m (almost OK!)

Nb_3Sn (MJR)--\$7.74/kA-m

Nb_3Sn (PIT)--\$28.94/kA-m

Bi-2212 (PIT)--$57.00/kA-m

Goal : < 2 k$/kA-m (like LHC at 10 T, 1.9 K).

Clearly, there is still a long way to go with HTS, and in part also for Nb_3Sn via PIT. For ITD the road is not long and the remaining reduction factor of 2-3 can be achieved just by the increase volume production. Indeed, today the quantity of Nb_3Sn is negligible in comparison with the NbTi production for the LHC and assuming a factor of two reduction in cost is therefore more than justified.

6. CLASSICAL AND NEW MAGNET DESIGN

Fig.1 is an overview of typical superconducting magnet cross sections used in HEP accelerators. In table 1 we can find some characteristics of accelerator projects. One can see that in little less than 30 years the maximum field almost doubled and the size became more and more important (although the use of the Twin aperture concept in LHC permitted a much more compact design with respect to the ill-fated SSC). Size means cost: to give an idea of the costs that are involved in future colliders, one can use table 2 where the cost per tesla-meter (the bending strength) for two beam magnets has been evaluated by an international panel [13].

From there one can infer that a saving on magnetic systems is mandatory in order to reduce the cost of such accelerators!

For this reason examining different coil configurations and conductors are key points in the R&D toward high field.

Table 1. Main HEP accelerator projects employing superconducting magnets.

Project (Lab)	Energy (TeV/beam)	Operating dipole field (T)	Tunnel length (km)	Status
CBA[□] (BNL)	0.4	5	3.8	Cancelled in 1983
Tevatron[†] (FNAL)	0.98	4	6.3	Operated in 1987
HERA (DESY)[*]	0.92	5.3	6.3	Operated in 1989
SSC (Dallas)	20	6.8	87	Cancelled in 1993
RHIC (BNL)	0.1/nucleon	3.4	3.8	Operated in 1999
LHC (CERN)	7	8.3	27	Operation in 2007

□ last name of the early "Isabelle".
† operation of superconducting magnets began in 1983 at lower field. 0.98T were reached in 2001.
* after upgrading in 1998 from 0.82 TeV, 4.7 T
heavy ion accelerator

Table 2 – Estimation of cost of the cryo-magnetic system for a future very high energy hadron collider into 2200 km tunnel length (from [13]). ELN is the Eloisatron project..

	B_o	Energy/beam	Bending cost	Dipole cost	Cryog. Cost	Total cost
Name	Tesla	TeV	$/d.Tm	G$	G$	G$
Pipetron	1.8	12.9	1500	0.41	0.50	0.90
RHIC type	5.4	38.6	2441	1.98	1.36	3.34
RHIC enh.	6.8	48.6	2720	2.77	1.45	4.22
LHC enh.	9.5	67.9	4400	6.27	3.17	9.44
ELN	12.5	89.3	6000	11.25	4,20	15.45

6.1 The classical cos design

Coil design of magnets that are utilized or proposed for real projects are very similar, as shown in Fig. 1, sharing the basic layout previously described. In this configuration the forces are like the one shown in Fig. 18.

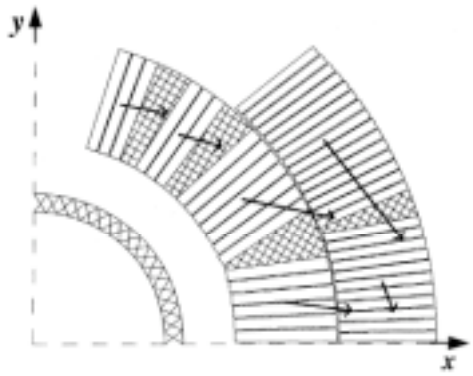


Fig. 18 Forces in a cos dipole (only one quadrant).

Restraining e.m. forces is basically achieved by placing the coils within a cavity. That cavity is made of stainless steel or aluminum collars surrounded by an iron yoke. The magnet assembly is enclosed within a cylindrical shell that may or may not provide additional structural confinement, for example by welding under stress. This is well illustrated in the cross section of the LHC main dipole magnet, Fig. 19.

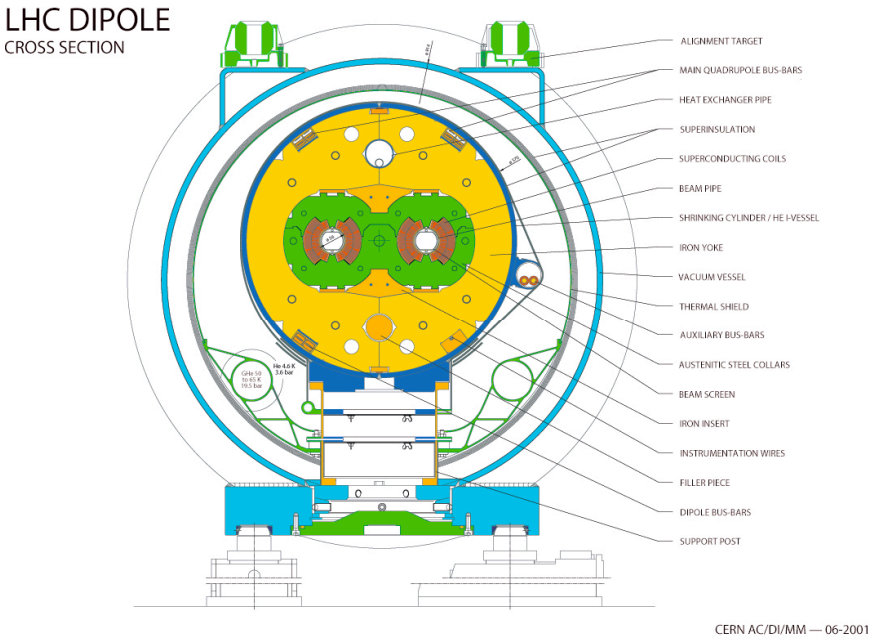


Fig. 19 Cross section of the LHC Main Dipoles

This cos layout has many advantages, among other the best magnetic efficiency but it has two main disadvantages:

- the force in the horizontal and vertical plane are coupled, and so is the way to control them;

- with higher fields the forces tend to growth to almost unmanageable limits. One can see from Fig. 20 that, for the force level obtained so far, a simple mechanism to cope with higher forces is not easy. If the coil gets thicker the effectiveness of the supporting collars decreases.

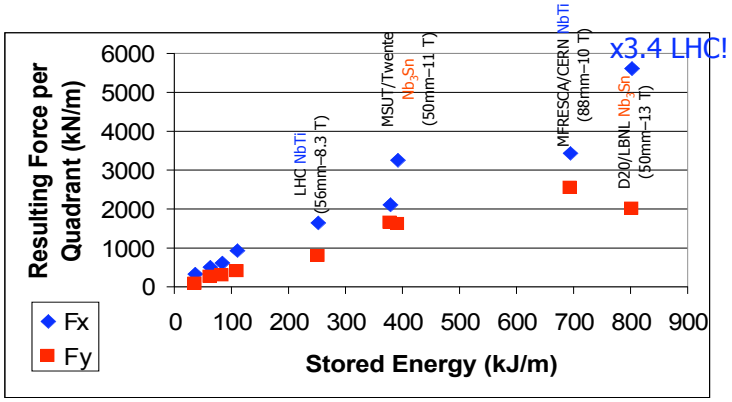


Fig. 20 Resultant of the horizontal and vertical force per quadrant for different projects. Most are NbTi wound dipoles except MSU and D20 that go beyond the 10 T threshold and are wound in Nb_3Sn (Courtesy of Arnaud Devred, CEA-Saclay).

However there are many magnet designs that keep maintaining this approach even beyond 10 T. In this regards we have to stress that there are two class of magnets for high fields:

- the main magnets (arc magnets), which determine the energy and cost consideration for the entire machine.
- the magnets in the insertion regions (near collisions) that mainly control the intensity. These last type of magnets are in small number: In the LHC there are about 20 km of main magnets and few hundreds meters of special magnets.

Of course similar design principles may not apply to both class of magnets and for the second class, performance is not necessary coupled to low cost. For example candidates for insertion magnets for the LHC upgrade are shown in Fig. 21 for dipoles and Fig. 22 for quadrupoles [14].

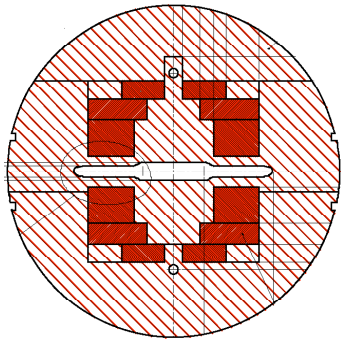


Fig. 21 Separation dipole proposed by Fermilab for the high luminosity upgrade of LHC. Note the opening in the mid plane, designed to cope with the high radiation deposition. (Courtesy of J. Strait, FNAL)

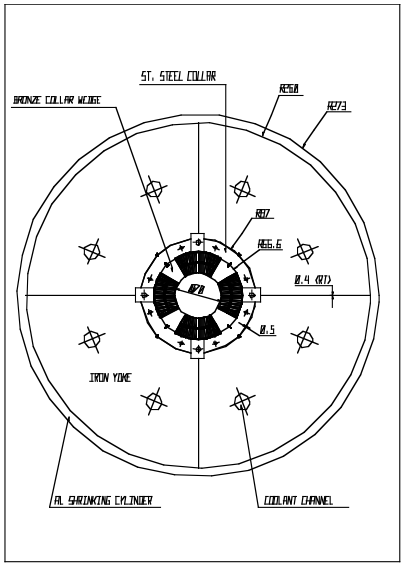


Fig. 22 Quadrupole proposed in 1997 by INFN-LASA Milan, later an INFN-LASA and LBNL joint proposal, for the LHC luminosity upgrade. [14].

For the first class, main magnets, there is a certain consensus (but not unanimity) that 12-14 T may be the field practical limit in Nb_3Sn magnets based on cos. Indeed, a main dipole magnet with a 11 T field has been engineered by Fermilab for the VLHC project as shown in Fig. 23.

It is worth to emphasize that the record field for main dipoles is 13.5 T, reached by LBNL in 1997 with the D20 magnet, reported in Fig.20 as the highest force cos ever built: however this excellent result, based on a conductor with half the current density presently available, is shadowed by the long training required to reach that field and by the use of construction technique not really scalable to long magnets and large production.

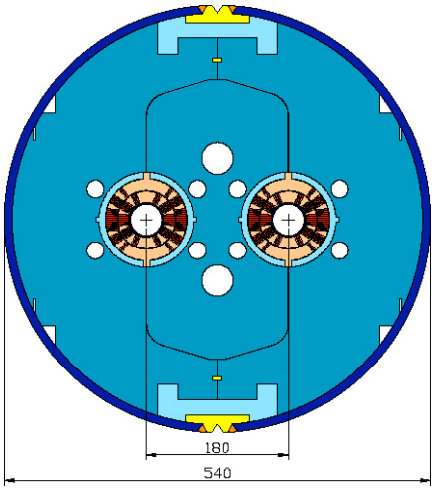


Fig. 23 Cross section of the Nb_3Sn, 11 T (operative) of a type of Main Dipole proposed for the VLHC (courtesy of A. Zlobin, Fermilab)

6.2 Novel designs

The energy and free thinking (not the money!) that followed the collapse of the SSC, has triggered and stimulated in the USA a fair number of novel ideas, or better a renewal of old ideas that have generated novel designs.

6.2.1 Block dual dipole

This is one of the first novel designs, proposed in recent years by P. McIntyre of the Texas A&M University [15]. In Fig. 24 a cross section of the dipole is reported (a dual bore version is available, too). The conductor, in form of tape or rectangular cable, is bent in the easy way and the coils are subdivided into many small blocks by means of strips and ribs made of Inconel to intercept horizontal forces. Vertical stresses are very moderate and by means of a strip/rib structure the peak horizontal stress is quite moderate, 50-70 MPa, a feature that looks very attractive for Nb_3Sn and for HTS. Coils are small and manageable and may possibly be manufactured from a single tape with sufficient size and current carrying capacity.

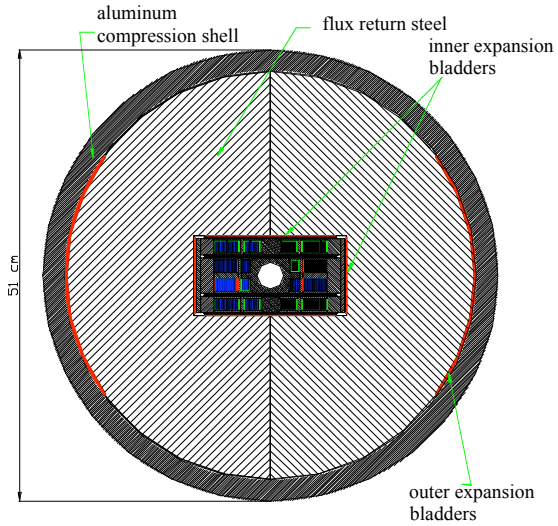


Fig. 24 Cross section of the single version of the dual block coil (courtesy of P. McIntyre, Texas A&M University).

However coil ends must be bent by a small amount in the non-easy direction to allow for the beam passage through the midplane. This bend is easier if the coil aperture can be kept small, some 25 mm or so. With such a gap this magnet can yield more than 16 T.

6.2.2 Common Coil Dipole Magnet

This very special design, first proposed by BNL and now the main project also at LBNL, relies on simple race-track coils constituting the leading current of one coil aperture and the return current on the same side of the other aperture, rather than the return on the other side of the same aperture. In this way the two apertures are strongly coupled, the field at the ends have a very bad shape, the two apertures must be aligned vertically rather than horizontally but the coils are really simple and fully planar. It's the ideal design to use HTS or LTS tapes. Moreover, as typical for window frame coils (rectangular block) the forces are easy to contain. The forces are mainly horizontally and require very moderate vertical prestress. Winding cost is low and assembly of coils becomes easy and flexibile, with the possibility of easily changing the coil configuration. Therefore, this design seems excellent in carrying out an R&D program where many coils have to be built and tested.

With this dipole configuration, it looks like fields in the 15-20 T range are possible [16], within an attractive 40 mm aperture. Provided that the conductor has enough current density pre-stress becomes manageable as well, In Fig. 25 a sketch of the magnet is shown.

This design still needs to solve many problems, the main ones being related to field quality. Slanted coils may be required within the main coil block, just below and above the beam tube. Other schemes are also possible in order to correct field errors at the straight section and the coil ends. All these trimmings add up complexity to a very simple and elegant design.

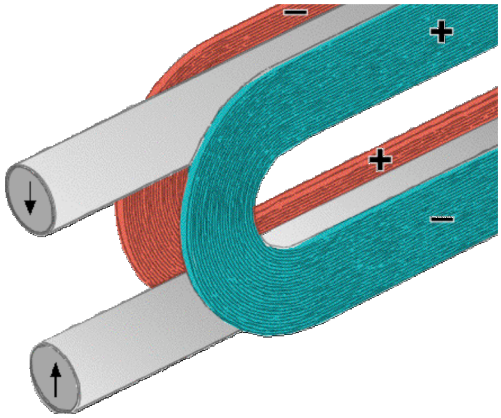


Fig.25 Basic idea of the common coil as proposed first by R. Gupta (BNL). The main coils are arranged such as not to cross the beam tubes.

6.2.3 R&W Common Coil Dipole

Fermilab is also investigating the possibility of building a Common Coil Dipole. The magnet is very similar to the previous one, except that it is based on React and Wind coils, trying to take full advantage of the easy bending. The distance between apertures is such that the curvature radius of the racetracks is about 90 mm. This choice of R&W looks quite attractive for HTS tapes, especially Bi-2223 and future Y-123, which unfortunately do not yet have sufficient current carrying capability. An extensive R&D program is going on to test current degradation vs. bending radius of Nb_3Sn single wire and flat cables. Conductor are reacted at intermediate radii's between the final winding radius and an infinite radius, i.e. straight cable, in order to determine the effective maximum strain experienced by the material after its last thermal treatment [17]. An artistic view of the ends of this dipole is shown in Fig. 26. Recent simplified models (so called race-tracks) went beyond 75% of the maximum load line current, a result that give room for some optimism in this route.

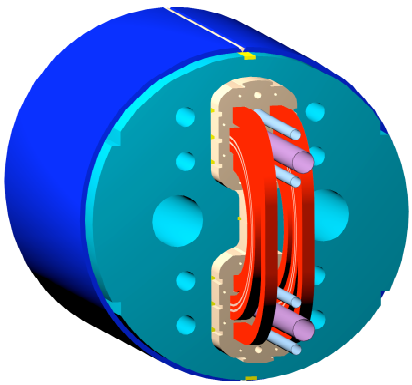


Fig. 26 Artistic view of the ends of the common coil under development at Fermilab with R&R technique (courtesy of G. Ambrosio, FNAL).

The R&D for Common coil has today its center at LBNL where recently a record field of 14.5 T has been reached with a particular version of common coil, although we have to stress that the bore is very small: 25 mm. This magnet relay on a structure that is a variant proposed by S. Caspi, based on an idea stemming from an INFN-LBNL collaboration on low beta Nb3Sn quadrupoles for the LHC [14]: to provide pre-stress through bladders, as illustrated in the Fig. 27, [18].

This route has been apparently very successful, and the LBNL design apparently can contribute to economical and affordable magnets in the range of 15 T.

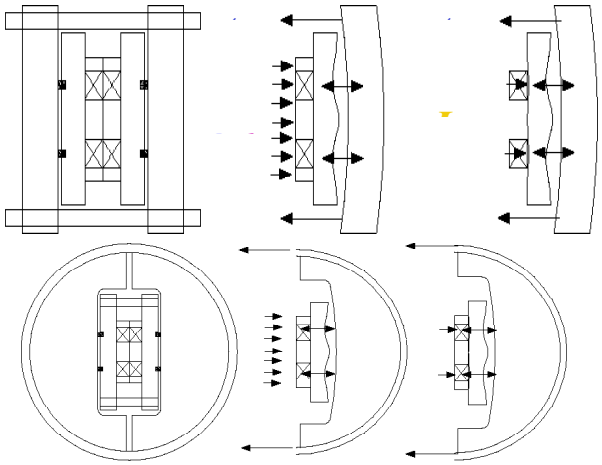
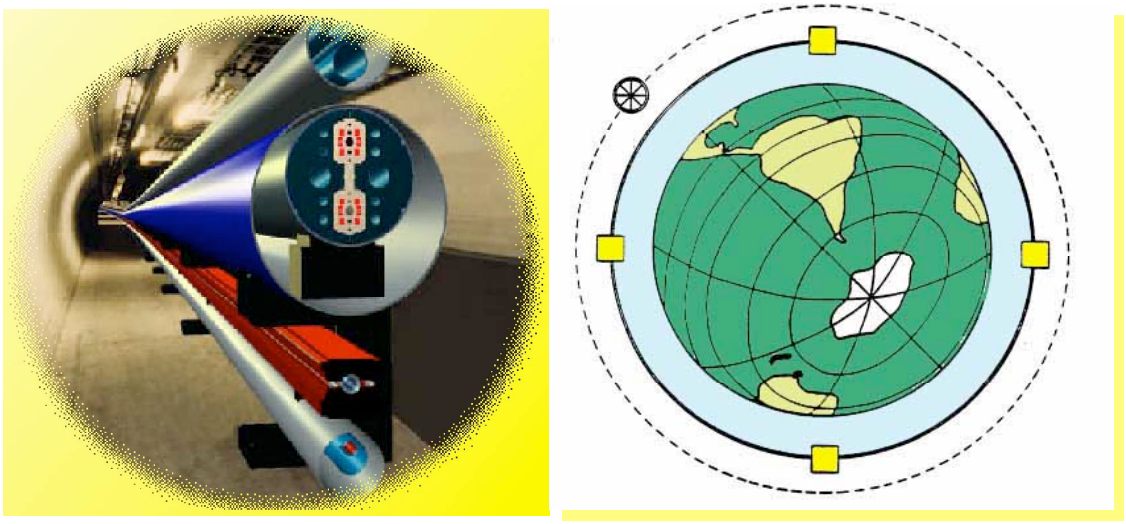


Fig. 27 Principle of functioning of bladders (the square elements among coils and mechanical supports) in the case of lateral plates or of containment by shells. When the magnetic forces go up the internal reactive stress are unloaded and the coils do not move. Courtesy of S. Caspi (LBNL).

7. CONCLUSIONS

This overview shows that, despite all the energy and resources that today are concentrated towards the LHC, the HEP accelerators community has not stop working on the next step, in particular in the US where there are more people and resources available.

Many people ask if there will be a next step: much will depend on the success, both technically and economically of the LHC. Certainly there will be room for high field magnets in the insertion regions for a luminosity upgrade of the LHC. Nobody knows if there will be a next superconducting magnet super-collider. The author can only refer to Fig. 28 where the 200 km VLHC tunnel (filled with common coil magnets) is depicted near the largest possible dream, and one may selected its own farthest energy frontier, having in mind that the one who does not dream never moves. Whatever the decision, the community should provide the best plan and best design, or few of them, to be ready to seize the opportunity when it will come!



Fig, 28 VLHC tunnel (200 km, filled with common coil magnets) and the largest possible ring for accelerators!

8. ACKNOWLEGEMENTS

The author wishes to thank G. Ambrosio (Fermilab), S. Caspi (LBNL), A. Devred (CEA), R. Gupta (BNL), P. McIntyre (Texas University), A. den Ouden (Twente University), R. Scanlan (LBNL), J. Strait (Fermilab), A. Zlobin (Fermilab) and many other colleagues of CERN and other labs for the helpful discussions and for generous and prompt supply of materials for this paper. Further particular thanks to S. Caspi (LBNL) for reading the manuscript and to C. Bosteels (CERN) for helpful assistance in editing it.

REFERENCES

[1] M.N. Wilson., Superconductivity and Accelerators: the Good Companions, *IEEE Trans. on Applid Superconductivity*, **9** (1999), pp.111-121.

[2] R. Perin and D. Leroy, Superconducting magnets for particle accelerators (dipoles, multipoles), in Handbook of Applied Superconductivity, Bernd Seeber editor, Institute of Physics Publishing, 1998, p. 1302.

[3] G. Rumolo, F. Ruggiero, F. Zimmermann,"Simulation of the electron-cloud build up and its consequences on heat load, beam stability and diagnostics" Phys. Rev. ST Accel. Beams 4 (2001) 012801

[4] K.T. McDonald, "Muon colliders: Status pf R&D and future plans" Proceedings of the 1999 Particle Accelerator Conference, NY 1999, p. 310.

[5] M.S. Zisman, "Status of Neutrino factory and Muon collider R&D" Proceediings of the 2001 Particle Acceleretor Conference, Chicago 2001, p. 359.

[6] M. A. Green et al., The Use of Superconducting Solenoids in a Muon Collider, IEEE Trans. on Superconductivity, **9** (June 1999), No.2, p. 1049-1052

[7] Design Study for a Staged Very Large Hadron Collider, Fermilab-TM-2149, June 4, 2001

[8] H.H.J. ten Kate, on behalf of the Atlas Collaboration, The Superconducting Magnet System for the ATLAS Detector at CERN, *IEEE Trans. on Applied Superconductivity,* **11**, (2001) pag. 347-352.

[9] F. Kircher et al., CMS Coil Design and Assembly, IEEE Trans. on Superconductivity, **12** (March 2002) No.1 p. 395-398

[10] I.I. Rabi, A Method of Producing Uniform Magnetic Fields, Rev.Sci.Instr.,Vol5, p78-79, Feb.'34

[11] R.M.Scanlan and D.R.Dietderich, "Progress and Plans for the U.S. HEP Conductor Development Program", presented at the 2002 Applied Superconductivity Conf., August 2002, Houston, TX, USA, and accepted for publication on IEEE on Superconductivity

[12] R. M. Scanlan, LBNL, private communication

[13] L. Rossi, Hadron Colliders at the highest energy and luminosity, World Scientific Publ., 1998, pp-133-140.

[14] S. Caspi, et al., Design of a Nb_3Sn High Gradient Low-Beta Quadrupole Magnet, Proc. of MT-15 Conference, Science Press Beijing, 1998, pp. 175-178.

[15] A. Abreau, et al., Block-coil Dipole for Future Hadron Collider, *IEEE Trans. on applied Superconductivity*, **9** (1999), pp.705-708.

[16] R. Gupta, et al., A High Field Magnet Design for a Future Hadron Collider, *IEEE Trans. on applied Superconductivity*, **9** (1999), pp.701-704

[17] G. Ambrosio et al., R&D for a Single –Layer Nb3Sn common Coil Dipole Using the React-and-wind Fabrication Techniques, *IEEE Trans. on Applied Superconductivity*, 12, (2002) pag. 39-42.

[18] S. Caspi, S. Gourlay, R. Hafalia, A. Lietzke, J. ONeill, C. Taylor and Al. Jackson, "The use of Pressurized Bladders for Stress Control of Superconducting Magnets" IEEE Trans. on Applied Superconductivity Vol. 11 No. 1, March 200, pages 2272-2275.

SURFACE EFFECTS IN SCRF CAVITY

H. Safa
CEA Saclay, DSM/DAPNIA91191 Gif-sur-Yvette, France

Abstract
The surface properties of a superconducting material submitted to a high radiofrequency field in the GHz range are studied. First, it is shown how the theoretical BCS surface resistance of an ideal superconductor is increased taking into account the 100 K effect, the trapped magnetic flux, the embedded impurities or the grain boundaries. Consequently, the experimental non-zero residual resistance is fully understood. Then the high field behavior is described. It is shown that thermal instabilities caused by surface defects may severely degrade the quench field value. Experimental evidence and its corresponding explanation are given for the quench field modification due to surface treatments (chemistry, electro polishing, heat treatments) or surface aspects (oxidation, roughness). Finally, the study and understanding of field emission enables today to closely approach the theoretical critical magnetic field in a bulk niobium-superconducting cavity.

1. INTRODUCTION

Superconductivity is basically a surface effect. Although the electron pairing occurs throughout all the superconductor bulk, current can only circulate in a very thin surface layer of tens of nanometers, resulting in extremely high current densities (J can exceed 10^{12} A/m^2). Therefore, any surface imperfection will disturb the optimum current flow. And if one looks closely at a clean metallic surface, he could easily notice it is quite far from a smooth pure ideal surface. There are grains having different crystalline orientation with a lot of impurities and defects inside the crystal structure. They are separated by grain boundaries. The surface is rough and on top of the metal, an oxide layer of a few nm is grown with different oxygen contents which structure contains even more imperfections than the metal itself. The oxide/vacuum interface itself is not perfect. One finds apart from adsorbed gases, a great deal of dust particles lying on the surface. As a matter of fact, it is quite astonishing that real metallic surfaces full of imperfections can give such a high performance as is now currently obtained in SCRF cavities over square meters of superconducting surface.

2. THE SURFACE RESISTANCE

In a radiofrequency field having a pulsation ω, the electromagnetic surface fields are exponentially decaying in the superconducting material over a depth of the order of the London penetration depth λ [**1**]. Because at finite temperature, the pairing is not complete, the order parameter Ω, square of the electronic wave function ψ, is not equal to one. Therefore, losses are generated by the motion of non condensed electrons through collisions (with a collision time τ), whereas the Cooper pairs will flow out of phase of the electric field without inducing any dissipation.

To calculate the RF losses, one may conveniently use the two fluid model in which the total current is the sum of the normal conducting electrons and the superconducting current

$$\mathbf{J} = -\frac{i}{\omega\mu\lambda^2}\,\chi\,\mathbf{E}, \text{ With } \chi = \Omega + i\,\omega\tau(1-\Omega) \text{ being the complex order parameter.}$$

Solving Maxwell equations in the superconductor, namely $\Delta\mathbf{E} = \frac{\chi}{\lambda^2}\mathbf{E}$, leads to the following equation obeyed by the normalized admittance $Y = (i\,\omega\mu\lambda)\frac{H}{E}$:

$$\lambda \frac{\partial Y}{\partial z} + Y^2 = \chi \qquad (1)$$

Integrating the losses inside the material gives a total dissipation proportional to the square of the magnetic surface field H_0 (quadratic behavior). The coefficient is called the surface resistance R_s.

$$P = \frac{1}{2} \int_0^{\infty} Re\left(\mathbf{E}.\mathbf{J}^*\right) dz = \frac{1}{2} R_S H_0^2$$

The solution of equation (1) will give the surface resistance of the superconductor. In the case (most frequently assumed) where the order parameter Ω and the collision time τ are uniform in the London penetration depth, the solution is straightforward

$$R_S = Re\left(\frac{i\,\omega\,\mu\,\lambda}{\sqrt{\chi}}\right) \qquad (2)$$

In principle, this should never be the case, as the very top surface layer is far from being perfect and uniform. But that approximation is often used because it simplifies the calculation. Now the order parameter Ω varies exponentially with temperature starting theoretically from unity at zero temperature and vanishing at the critical temperature T_c. There, different theories predict different temperature variation, depending on the superconductor parameters, defects and impurities.

2.1 BCS Resistance

The BCS theory [**2**] predicts the occurrence of an energy gap opening Δ in the vicinity of the Fermi energy below a critical temperature T_c and calculates its variation with temperature, following a fixed electron-phonon attractive potential (-V). For each temperature T, the gap Δ is solution of the equation

$$N\,V \int_0^{\hbar\omega_D} \frac{th\left(\frac{\sqrt{\xi^2+\Delta^2}}{2\,k\,T}\right)}{\sqrt{\xi^2+\Delta^2}}\, d\xi = 1 \qquad (3)$$

where N is the electron density of states at the Fermi level and ω_D the maximum (or Debye) phonon pulsation of the crystal. The famous relation $\Delta_0 = 1.764\ kT_c$ between the energy gap at zero temperature and the critical temperature can be deduced from equation (3).

Giving a decrease of the normal conducting (or unpaired) electrons in the form of $e^{(-\Delta/kT)}$ at low temperatures, a good approximation of equation (2) writes

$$R_S = \frac{\omega^2\,\mu\,\lambda\,\tau}{2} \frac{(1-\Omega)}{\Omega^{3/2}} \qquad (4)$$

Equation (4) is what generally people refer to when they quote "BCS" surface resistance value in RF field. The temperature variation is hidden in the order parameter variation with Δ accordingly to equation (3). Apart from very specific cases, the BCS variation with frequency (increase as the frequency square) is nicely confirmed experimentally. This frequency variation is the reason why too high frequencies are not favored for superconductivity application. As a matter of fact, there is also a critical frequency ($\hbar\omega_c = 2\Delta$) above which the superconducting state is destroyed. Therefore, for low critical temperature material, typical RF applications lye in the GHz domain.

Theoretically, R_s should steadily approach zero at zero temperature, the order parameter Ω being equal to unity. However, most experimental measurements show that a non-zero surface resistance, called the residual resistance, is still obtained at very low temperatures. It will be shown hereafter that most (if not all) of the many different contributions to this residual resistance has been now understood. As an example, the temperature variation for a typical niobium surface resistance is plotted in the next figure.

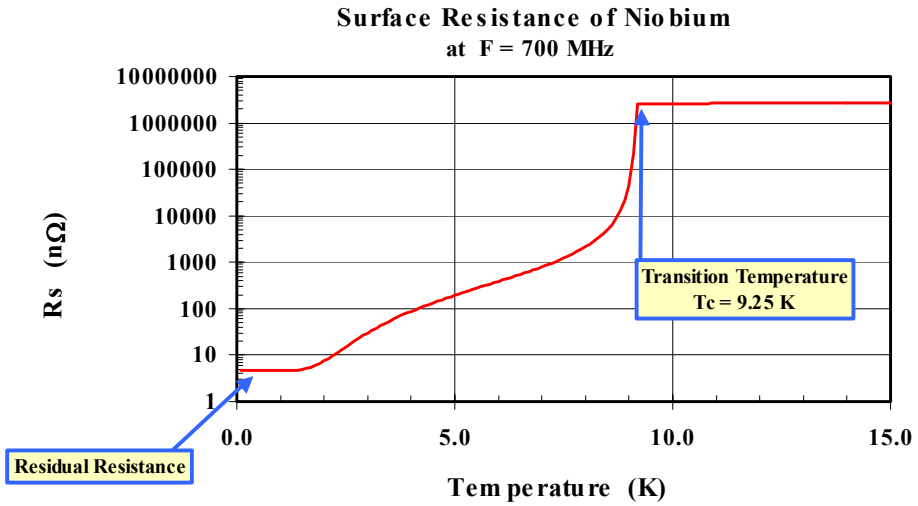


Fig. 1 Temperature variation of the surface resistance of niobium at a frequency of 700 MHz. Note the six orders of magnitude drop between the critical and zero temperature.

Impurities in the superconductor play an important role in the surface resistance. At these cold temperatures, they are the main limitation to the increase of the mean free path of electrons $l = v_F \tau$, thus limiting the collision time τ. The RRR ratio of the electrical resistance of the material in the normal state between room temperature and cold temperature is a figure of merit related to the amount of impurities (each impurity bringing though a different contribution). Equation (4) indicates that the surface resistance is proportional to τ, which in turn is proportional to the RRR value. Therefore, surprisingly enough, at a given temperature, the surface resistance of a superconductor *increases* with the purity of the material. This is called the "clean" limit. But if one decreases the RRR too much (for example by addition of a high impurity content), then the collision time will be effectively diminished but at the expense of degrading the superconducting properties (Ω will then be equally decreased). The critical temperature and the gap value will both decrease leading to a "dirty" superconductor and a higher surface resistance. Therefore, for each working temperature, there is an optimum RRR value reducing the surface resistance as illustrated in Fig.

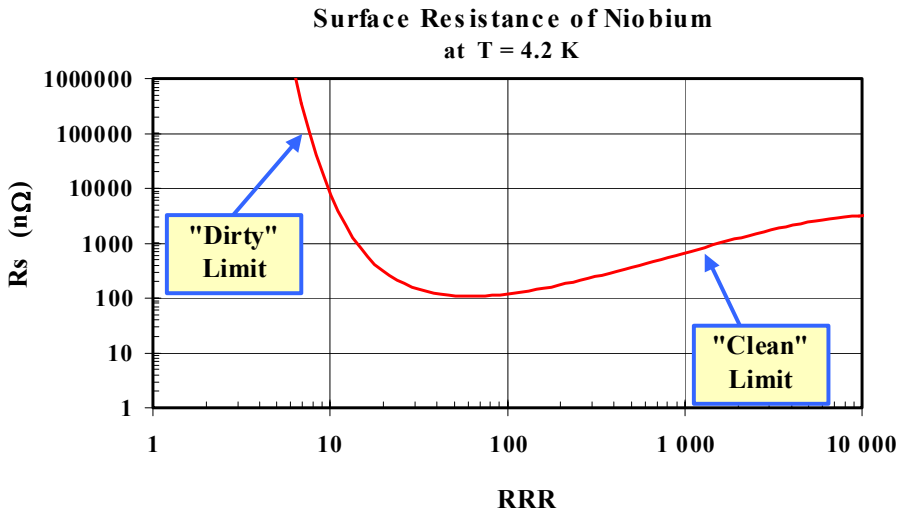


Fig. 2 Surface resistance as a function of RRR. In the clean limit, high RRR leads to high collision time. In the dirty limit, the superconducting state is gradually destroyed.

2.2 The 100 K effect

During early days of SCRF work, measured surface resistance of a cavity could vary by more than one order of magnitude from test to test [**3**]. This puzzling non-reproducibility has been shown to be due to the formation of a hydride at the cavity surface [**4**]. It turns out that the niobium-hydrogen phase diagram shows a thermodynamic phase forming at low temperature (between 100 K and 160 K) which is a rather bad superconductor with a critical temperature of 2.8 K. Depending on the cooling time and the duration in the vicinity of 130 K, a cavity with a high hydrogen content would exhibit either a slight or an important Q-degradation (the quality factor Q of a SCRF cavity is inversely proportional to the surface resistance) as shown in **Error! Reference source not found.**. The cure is to fire the cavity in a vacuum furnace at a temperature close to 800°C, degassing all the hydrogen out of the material. A heat treated cavity will not show any Q degradation even after staying a long time at 100 K.

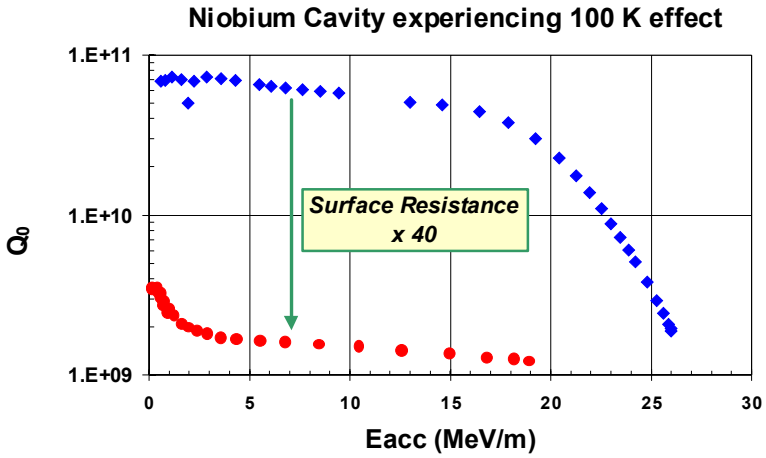


Fig. 3 A niobium cavity with a high hydrogen content experiencing the 100 K effect. The surface resistance may be 40 times higher than a fast cool-down.

2.3 Flux Trapping

In a type II superconductor, a static magnetic field Bext (like the earth one) may squeeze down in vortex upon crossing the superconducting transition, each vortex enclosing exactly the quantum flux $\Phi_0 = \frac{\pi \hbar}{e} = 2.068\ 10^{-15}$ Wb. According to Ginzburg-Landau theory [**5**] by minimizing the energy, vortex are normal flux tubes containing a magnetic critical field H_{c2} and having a size equal to the coherence length ξ of the superconducting Cooper pairs ($\xi^2 = \frac{\hbar^2}{8\,m}\frac{\Omega_\infty}{\mu H_c^2}$). It has been shown [**6**] that all the magnetic flux is trapped in the superconductor in form of vortex. The number of vortex per unit area is then $n = \frac{B_{ext}\,.\,S}{B_{c2}.\left(\pi\,\xi^2\right)}$. In an alternating field, vortex motion is lossy [**7**] and leads to a dissipation equivalent to that of a normal conducting zone. The corresponding additional surface resistance is

$$R_s = R_n \frac{B_{ext}}{B_{c2}}$$

In the earth magnetic field of 40 μT, the corresponding surface resistance is 280 nΩ for a RRR 200 niobium at a frequency of 1 GHz. This is why a highly effective static magnetic shielding is required for superconducting cavities.

2.4 Granular Superconductivity

A polycrystalline material is formed of grains separated by grain boundaries. It is well known that these boundaries are non-perfect and may exhibit a weakened superconductivity as compared with the bulk crystal structure. In fact, one may take advantage of this weakening by forming

Josephson Junctions [8] with quite amazing electronic properties and applications. The grain boundary maybe characterized by its specific resistance value G given by the Ambegaokar-Baratoff [9] relation $G \cdot J_c = \frac{\Delta}{e}$, J_c being the critical current density.

In a RF field, the surface current is forced to pass through many grain boundaries and the material can be viewed as an array of weak links having different critical current density. Various model have been established [10,11] all showing an increase of the surface resistance with the applied electromagnetic field. It is important to point out that this contribution to the surface resistance is highly non-quadratic, the "residual" resistance increasing with the field.

2.5 Very high Q cavity

After nailing down each term contributing to the residual resistance, namely the 100 K effect [**12**], the influence of the static magnetic field, the granular superconductivity, and the impurities (or RRR) effect [**13**], very small residual resistance can be experimentally achieved. Fig shows an example of a niobium cavity where Q values higher than 2.10^{11} are effectively measured. The remaining residual resistance is smaller than 0.5 n□. Even though, this residual tiny part can be explained if one takes into account physical effects that are generally neglected, such as the RF losses in the end-flanges (in stainless steel), the surface roughness (very important at high frequencies), the dielectric losses of the oxide layers or the coupling losses. Moreover, some residual static magnetic field still remains, as the magnetic shielding of the cryostat is not perfect.

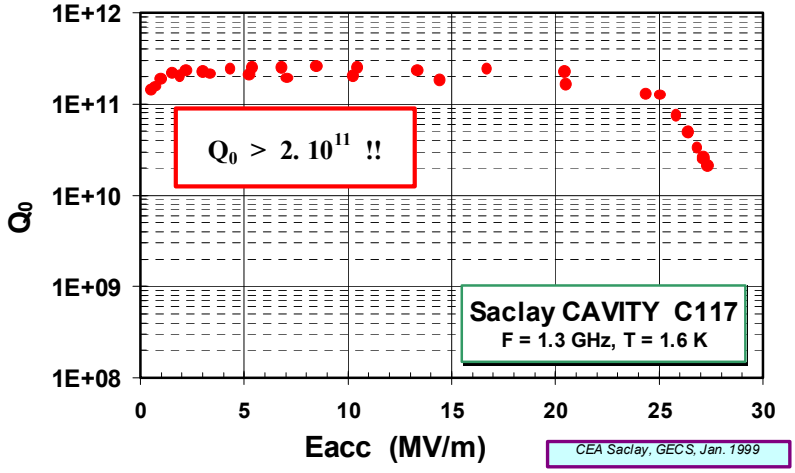


Fig. 4 Residual resistance as low as 0.5 nΩ is actually measured on large area cavities, giving an intrinsic quality factor Q_0 exceeding 2.10^{11}.

3. CRITICAL FIELD

3.1 Phase Diagram

The usual phase diagram of a type II superconductor shows vortex penetration above the first critical field B_{c1} and complete loss of superconductivity above B_{c2}. In a radiofrequency field, it is however stated that the critical magnetic field is the so-called superheating field B_{sh} [**14**], generally exceeding the critical thermodynamic field Bc (199 mT for niobium [**15**,**16**]). It is believed that the nucleation time for creating and moving a vortex in the material is higher than the RF period [**17**,**18**]. But until now, the experience has failed to confirm this hypothesis [**19**], especially at low temperatures (T = 0K). This superheating field should be the ultimate magnetic field a perfect superconductor at (T = 0 K) would sustain before switching to the normal state.

3.2 The Quench : A Thermal Instability

But the actual surface temperature T is finite which will somewhat lower the critical field. Slowly increasing the energy in a SCRF cavity, a sudden “breakdown” is observed at a given field, damping almost immediately all the stored energy. This is defined as the quench field and is mainly a

thermo-magnetic physical process. First, the magnetic critical field decreases with temperature as $B_c(T) = B_{c0}\left[1-\left(T/T_c\right)^2\right]$. On top of that, the quench is not a purely magnetic phenomenon. It is a thermal driven process involving thermal instability due to power dissipation in RF field. Heat flux generated at the inner surface ($Q = \frac{1}{2} R_s H^2$) has to be conducted through the niobium wall to be removed by the helium bath (Fig). Given the wall thickness t, the thermal conductivity Λ and the Kapitza conductance hk, the inner surface temperature T will be solution of the equation:

$$(T - T_{bath}) = Q\left[\frac{t}{\Lambda} + \frac{1}{h_k}\right]$$

Both the surface resistance Rs (equation 4) and the thermal conductivity Λ (only unpaired electrons may conduct heat) are rapidly varying with temperature and the above equation might give no solution in the superconducting state above a given field. This defines the "uniform" quench field level, which may vary according to the heat properties of the material. For example, the quench field is much higher in the superfluid regime ($T_{bath} < T_\lambda = 2.17$ K) than in the boiling helium regime due to the excellent heat removal of a superfluid bath.

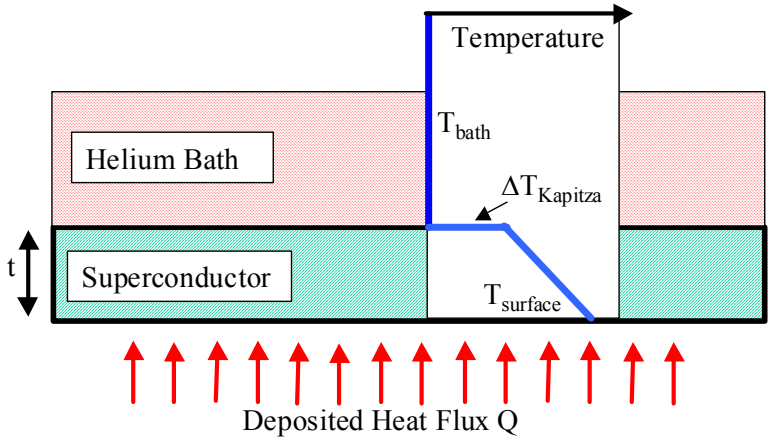


Fig. 5 A SCRF cavity is cooled by the helium bath through the bulk material.

Accordingly, the quench may also vary with the niobium purity (as determined by the RRR), the heat treatment (under vacuum at 800°C for hydrogen removal or from 1000°C to 1400°C for post purification) or the chemistry and preparation. However, the observed quench field of cavities is usually lower than the expected calculated one assuming the uniform heating case. At the same time, some localized heating spots may be detected by temperature mapping indicating that a thermal instability is occurring driven by a micron-size defect. More complete thermal analysis confirm that this is indeed the case and can explain why the quench changes with the heat properties of the helium bath [**20**,**21**,**22**,**23**].

Purifying the niobium from its impurities and reducing the defects during the sheet production has led to tremendous improvement in the average maximum field limit. The quench field value vary with the niobium purity (RRR), the heat treatment and the cavity preparation (chemistry and temperature). Quench fields exceeding 170 mT have been measured on several electropolished single-cell cavities, closely approaching the thermodynamic critical field of niobium Bc = 199 mT.

3.2.1 RRR

It has been experimentally observed quite early in SRF community [**24**,**25**] that the quench field improves when one uses higher RRR material sheet for cavity fabrication. Fig shows Q vs. Eacc curves obtained for different RRR cavities having the same cavity shape at the same frequency (1.3 GHz). The quench field definitely increases with the RRR. A low RRR cavity (reactor grade RRR30) quenches at 9 MV/m. Whereas for the highest RRR (> 500), the quench value could not be reached even at the maximum field exceeding 30 MV/m. Note also that the Q value at low field

steadily increases with the RRR as pointed out in previous observations [**26**]. For low RRR values, the Q curve starts showing a rather strong slope at low field. This slope indicates a non-quadratic loss increase with field and can be observed whenever the superconductor surface is degraded. For example, a cavity as received after fabrication will exhibit that behavior due to the damage layer following the mechanical cold work during forming. Only a heavy chemistry removal (above 100 µm) will get rid of that slope. In the same manner, thin film deposited superconductors [**27**] (like niobium on copper) or seamless cavities [**28**,**29**] (hydroformed or spun) will almost always exhibit that kind of slope starting at low fields.

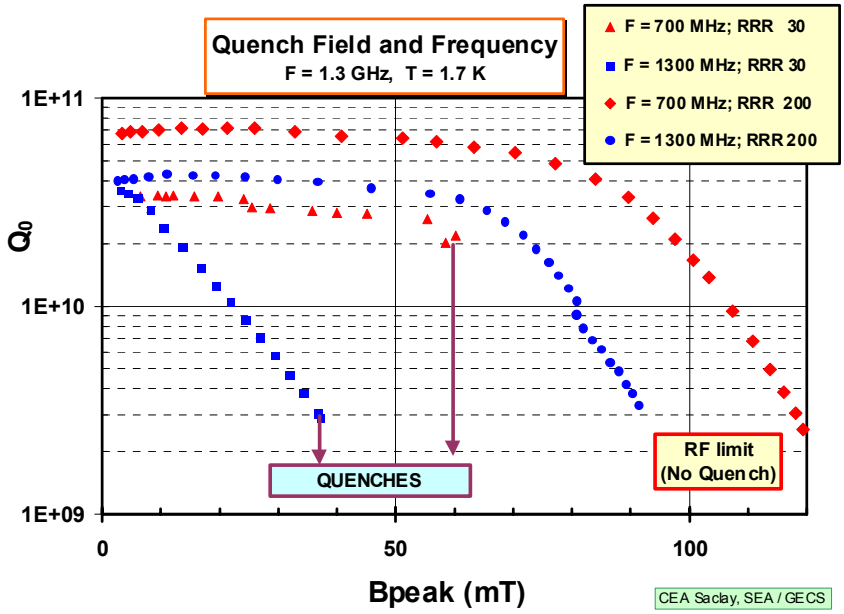


Fig. 6 The quench limit strongly depend on the niobium RRR value.

The RRR ratio related to impurities contents, especially light ones (O, C, N), is obviously an important parameter. Generally speaking, a poor RRR material exhibits poor superconducting properties. Major superconducting parameters can be affected like the critical temperature Tc (loosing for example -1 K/at. % of oxygen [**30**,**31**]), the critical field Hc or even the residual resistance [**26**]. Hence, one needs to get very high RRR material to approach the pure niobium theoretical limits. And the purer is the better. Also, the high RRR material favors the heat conduction through the cavity wall as the thermal conductivity steadily increases with the purity. So the temperature drop between the helium bath and the inner surface is lower for high RRR metal, helping to thermally stabilize local defects at high fields.

However, these general statements, while always true, have to be evaluated for each specific case. For example, the critical temperature reduction is very small for all RRR higher than 200 ($\delta T_c <$ 0.01 K) and can only be measured by extremely precise experiments. In the same manner, the defect stabilization is for example frequency dependent as stated and calculated previously [**32**]. This is also supported by some experimental results. Q-curves and quench field values are different at different frequencies even for cavities made from the very same RRR sheets and prepared in the same manner (same chemistry). Although a low frequency cavity has a much bigger area, and consequently, has a much higher probability to include larger defects, the defect stabilization is more effective at lower frequency, and hence the quench field is higher, as expected from simulations. In a similar way, baking will diffuse some light impurities from the very outer niobium surface inside the material, as shall be seen in the following discussion. The layer where the diffusion take place will exhibit a degraded RRR and consequently its superconducting properties will be degraded. That is why the quench value, the BCS resistance at 4.2 K and the Q_0 value at low field are all three lowered after baking. Baking at higher temperature (or for longer time duration) will also result in additional degradation of the RRR in the diffused layer and is completely consistent with the observation.

3.2.2 Heat Treatment

Two different types of heat treatment of the niobium cavity in a vacuum furnace have been extensively used. The "medium temperature" heat treatment (typically 800°C, 2 hours) is generally used to avoid the Q-disease [4]. Over 600°C, hydrogen starts getting out fast from the metal. Lowering the hydrogen concentration avoids the formation of the hydride thermodynamic phase during cool down at temperatures between 160 K and 100 K. The result of this heat treatment on the quench value is not established. Surprisingly, some cavities do show an increase in performance while most of the time, the quench does not change. The second "high temperature" heat treatment is used to purify the niobium from its light impurities (O, C and N) using an evaporated getter (for example Ti). Heating temperature range from 1000°C to 1400°C depending on the available time. After that treatment, the RRR is higher (actually very high RRR can be achieved [33]) and consequently the quench field increases. Significant improvement in field value can be obtained following that heat treatment (Fig).

Unfortunately, that high temperature treatment has a major drawback. The grain growth is so important at those temperatures that the mechanical properties of the niobium are seriously degraded. The yield strength decreases, the cavity soften and that induces additional problems mainly in pulsed operation due to the frequency shift with field (so-called Lorentz force detuning). The large grain size (a few mm) will also induce uneven etching upon using the standard buffered chemical bath (BCP). Because different grains have different crystallographic orientations, the grain structure is highly revealed, their boundaries are quite well marked showing sharp edges and steps from one grain to another. That is the reason why one would prefer the lower temperature (1000°C) treatment to minimize this drawback. It is also the reason why a study of a new type of chemical etching bath have been started to minimize the uneven etch among grains on heat treated cavities. Although the field was improved, a Q-slope appears at high field without any sign of field emission or X-ray signal [**34**]. These anomalous losses at high fields have been shown to be more or less uniformly distributed on the cavity wall.

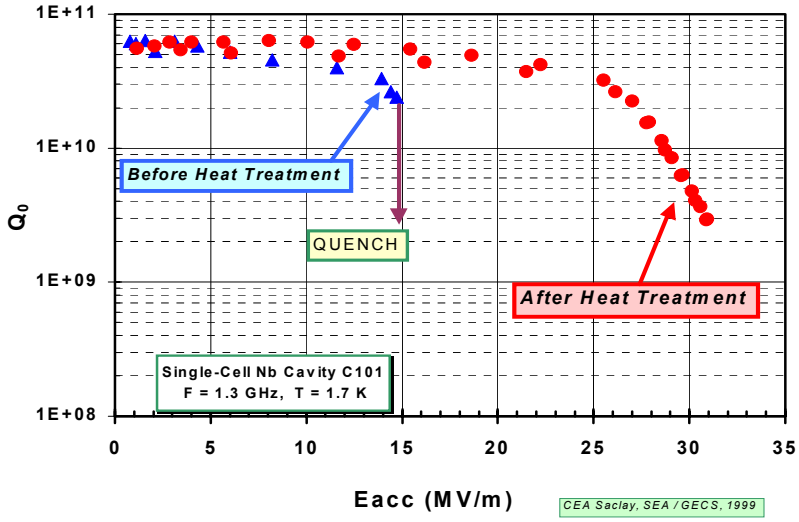


Fig. 7 A high temperature heat treatment improves the quench field value.

3.3 Surface Roughness

It has been shown that tiny local defects may induce a premature quench of the whole cavity. The weakest point on the surface will therefore determine the critical surface field. A surface obtained after a BCP is far from regular and the roughness is usually quite bad (about 3 µm rms). But the main point is not actually the roughness itself but the possible ***local*** enhancement of the magnetic field due to sharp edges, like those of grain boundaries revealed by etching (Fig). RF calculation [**35**] do show that a magnetic field enhancement is expected right on the edges, with about the right order of magnitude. Due to the difference in the enhancement factor, the ratio between the macroscopic field and the local field is lower for the BCP cavities, leading to a lower apparent accelerating quench field.

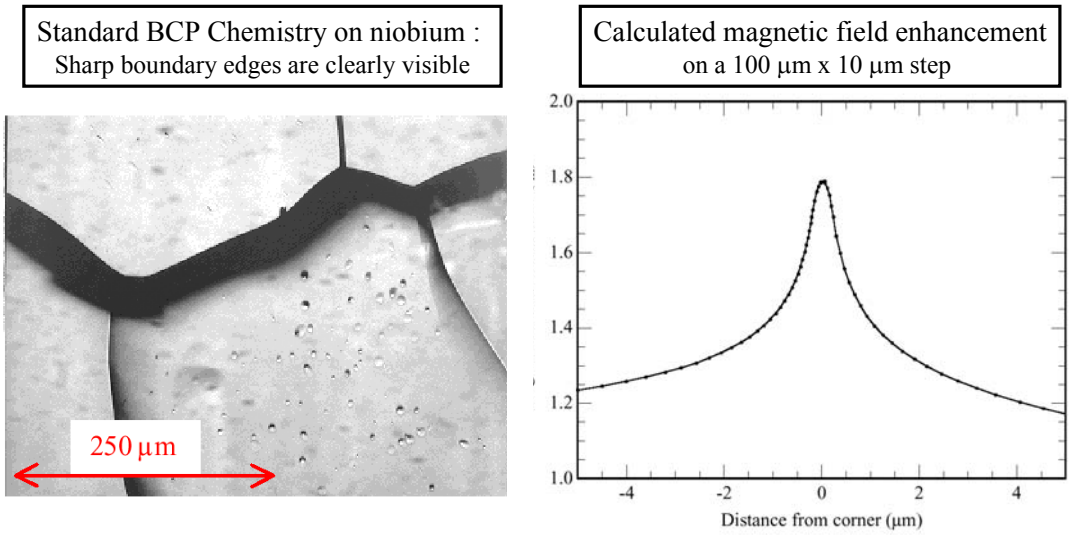


Fig. 8 Micrographic picture of a BCP etched niobium sample [from 41] and magnetic field enhancement calculation at the edge of a step in a RF cavity [from 35].

3.3.1 Electropolishing

Electropolishing (EP) is a chemical treatment that smoothens the sharp edges of the material grain. Consequently, a higher quench field would be expected with this preparation as compared to the standard BCP chemistry. And this is actually what is experimentally observed. This technique was first applied by KEK in the mid 90's: Accelerating gradients as high as 40 MV/m have been obtained on EP cavities even with "moderate" RRR values of around 200 [**36**,**37**]. Electropolishing have been used since long time in Japan and have also been investigated elsewhere [**38**], but other limiting factors at that time were inhibiting the high field capability of this chemical preparation. The Q-slope at high fields on EP cavities happened to be much less of a problem than on the standard BCP chemistry used in most other labs. In fact, it turned out that the cavity preparation procedure used at KEK usually includes a final baking at a temperature around 85°C.

The KEK results have been since widely confirmed, particularly through crosscheck laboratory measurements at KEK, Saclay, DESY or CERN [39]. A specific DESY/CERN/Saclay collaboration has been setup to evaluate the benefit of electropolishing with a large number of single-cell cavity tests. Detailed description of the procedure used and the different results obtained can be found elsewhere [40]. Some typical results are plotted in **Error! Reference source not found.** and can be summarized as follows: The average quench field exceeds 150 mT, significantly higher than the fields obtained with any usual BCP chemistry. Also the reduction in Q-slope at high fields after a 120°C baking under vacuum is much more effective.

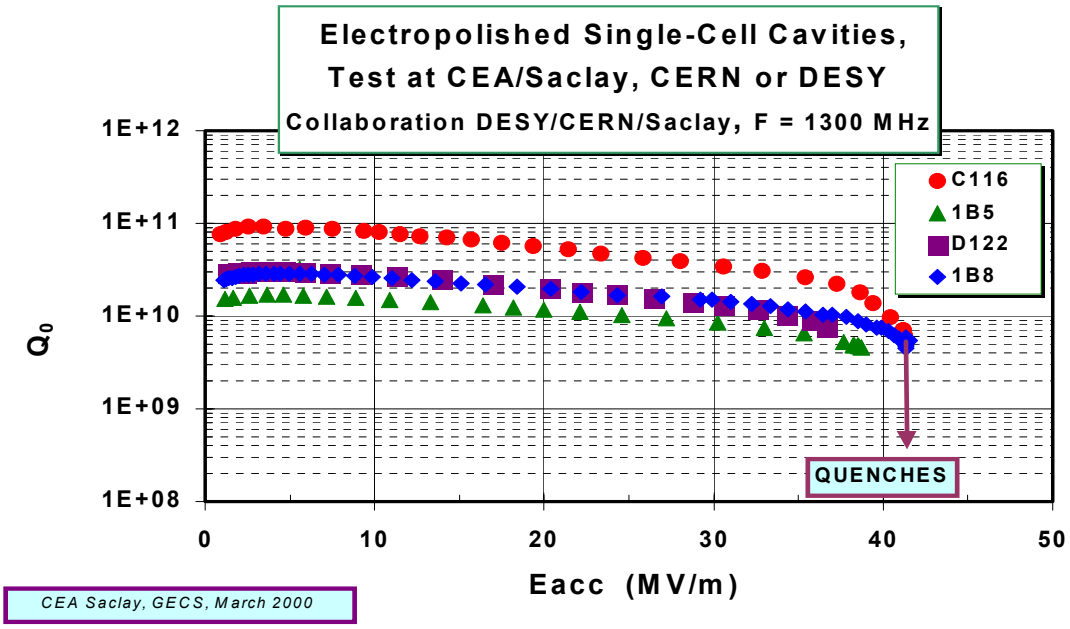


Fig. 9 Results of four different electropolished single-cell cavities. The average quench field value is around 40 MV/m.

3.3.2 Other Chemistry

Since a few years, other type of chemical baths, different from the standard one, have been investigated, mainly at CEA Saclay [41]. Different acid mixtures or preparation have been tried first on samples and then on single-cell cavities. Very interesting results have been obtained, and the main feature is the fact that the quench field value of a SCRF cavity can be moved if the chemistry is changed. As a matter of fact, the microscopic observations made on samples etched with these different baths are also quite different (grain boundaries, step heights, shining and roughness are quite different). For illustration, a cavity has been successively prepared using three different chemical baths. And the corresponding quench field have been measured each time to be different for each different chemical preparation.

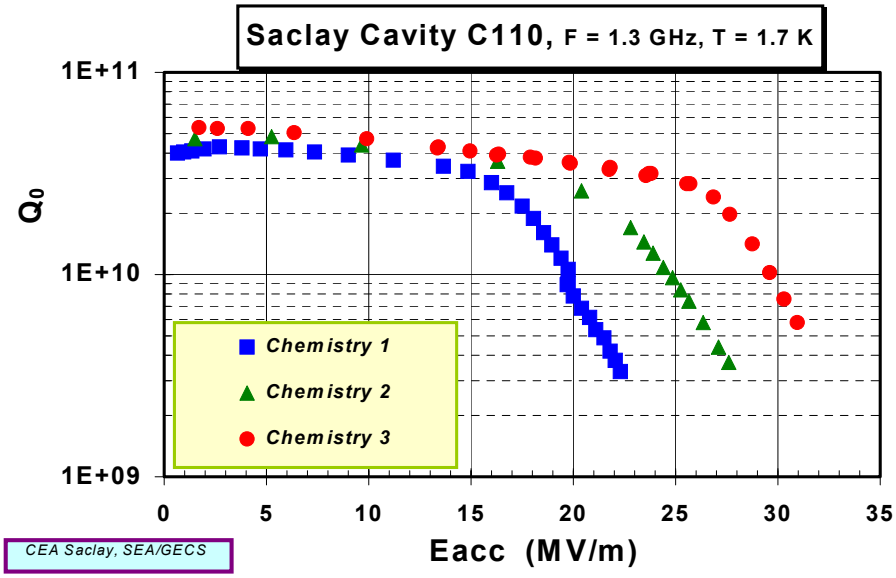


Fig. 10 The quench field of a cavity may vary with the chemical bath preparation. That particular cavity was not heat treated at high temperature.

4. HIGH FIELD BEHAVIOR

4.1 Field Emission

4.1.1 Phenomenology

Field Emission (FE) is a quantum mechanics process allowing a conduction electron to tunnel through the work function barrier Φ (generally a few eV) into vacuum if a high enough electric field is applied to the surface. The probability of crossing through the barrier has been first calculated by Fowler and Nordheim (FN) [**42**] giving the expected electron current density J as a function of the applied electric field E.

$$J = \frac{1.54\,10^{-6}\,E^2}{\Phi}\; e^{-\left(\frac{6.83\,10^9\,\Phi^{3/2}}{E}\right)} \qquad (5)$$

Note the exponential increase of current with field. The electric field at which the current starts to be measurable using some electronic device (usually 1 pA) is the field emission threshold. Each surface will have its FE threshold. According to equation (5), the expected field threshold for a typical niobium surface should be 1 GeV (Φ = 4 eV, J = 10^{10} A/m^2). But the experimental one deduced from a FN plot (current versus field) is about 100 times lower! Moreover, it has been identified [**43**] that the current is only emitted from localized spots on the surface called "emitter sites". These emitters have equivalent surface size of a few nm^2. A lot of work have been initiated in many SCRF labs to try to identify the nature of the emitter sites [**44**,**45**,**46**]. It was found, using specific apparatus (like modified SEM in DC field), that most of the emitters were due to surface defects. They can be separated in two categories: scratches and particle contamination (Fig.). A scratch can be caused by touching the surface with a hard tip (a tool or even a nail hit may induce a scratch). Some surface material is then pushed aside forming a micron-size protrusion. Particle contamination comes from the ambient environment and all kind of elements is to be found there (from C to In).

Fig. 11 Example of a scratch and a particle on a niobium surface.

Although all scratch and particle show a reduced FE threshold, the geometrical enhancement due to their geometry is not sufficient to explain the 100 ratio observed. These micron-size defects would explain at the most a factor of 10 in aspect ratio (height to radius) field enhancement. But an improvement in the analysis can account for the additional order of magnitude using the "tip-on-the-tip" model [**47**,**48**]. This model (Fig.) is based on the fact that a micron-size particle (or scratch) has an irregular surface and assumes that there is at least one sharp nanometer protuberance on it. A very rough estimate would lead to explain the overall enhancement field factor $\beta = \beta_1 . \beta_2$ of the order of 100. This is supported by experimental evidence using a high resolution SEM (Fig.) and by the nanometer surface area deduced from FN plots. Specific experiments using spherical shaped particles do show that their FE threshold is about 10 times larger than irregular shaped particles of the same kind.

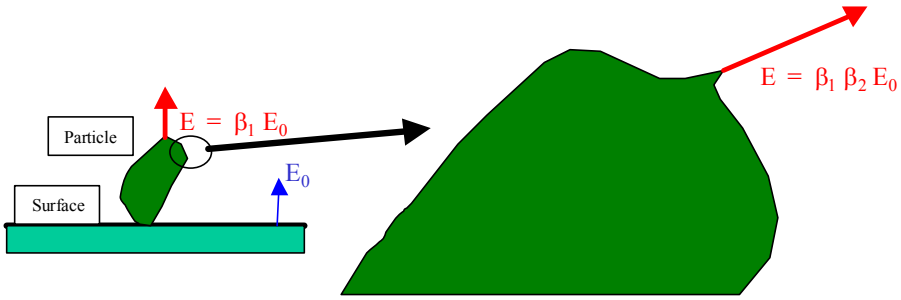


Fig. 12 "Tip-on-Tip" model explaining the geometrical local enhancement factor.

Fig. 13 SEM picture showing a local enlargement (right picture) of the tip on a scratch. That experimentally supports the "tip-on-tip" existence.

The behavior in an alternating RF field has been specifically studied at Saclay using a special cavity [49,50]. It has been demonstrated that while the basic FE mechanism is still the same as in DC (the RF period being much larger than the electron tunneling time), specific features like the RF heating of the particle make the overall FE behavior quite different in RF. In particular, it has been explained how and why High Peak Power Processing (HPPP) [51], applying very short RF pulses, is quite effective.

4.1.2 How to avoid field emission?

The only way to avoid field emission is to be as clean as possible in the surface preparation. After the chemistry, the cavity is rinsed in ultrapure filtered water and then brought inside a clean room. There, a quite effective experimental procedure for removing particle contamination has been found: the use of a high pressure rinsing with ultrapure water [**52**]. The effectiveness of the high pressure rinsing has been demonstrated [**53**] and explained [**54**,**55**]. After that, the mounting has to be extremely careful and clean and the cavity sealed inside the clean room. In order to avoid re-contamination of the surface, the cavity must not be opened to air after mounting. It should be very slowly evacuated (particle can be swept by a too high differential pressure) and then installed for test or operation, remaining all the time under vacuum. By doing so, very high electric surface fields are achievable (over 70 MV/m), even on large surface area, as shown in the following figure.

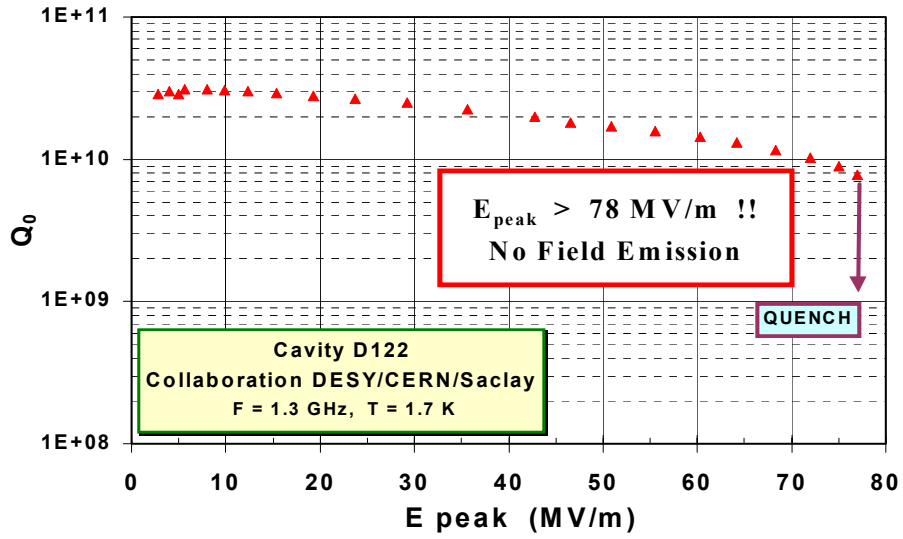


Fig. 14 High electric field emission free cavity (no X-ray, no electron signal).

4.1.3 How to suppress field emission?

Generally, there is no way to completely suppress field emission, especially upon contamination of the surface. However, there are some techniques to try to increase the FE threshold field that might be valuable in real accelerator environment. One of the oldest technique is called helium processing [**56**]. By introducing a small amount of helium gas inside the cavity (around 10^{-5} mbar) and maintaining the highest possible field could sometimes result in a slight and progressive increase in the achievable gradient. This has been extensively used on the CEBAF accelerator where the FE was limiting the accelerating field at around 5 to 7 MV/m [**57**]. The physical explanation is the ionization of the helium gas by the electron emission and the subsequent bombardment of the emitter site by these ions gaining energy in the RF field. But this might eventually lead to destruction of the site or even production of other artificial emitters. This has also been observed and sometimes, especially when starting from a rather good surface, helium processing may lead to the decrease of the threshold, degrading the performance. RF and HPPP processing are generally more efficient but could only be applied provided one has a very high power RF source available. Other different techniques have been also tried but they require a cavity opening and are not suitable for application in the machine.

4.1.4 Multipacting

Multipacting is a resonance phenomenon appearing in RF at precise energies depending on the geometry. Electrons impacting the wall may induce secondary electrons that in turn, if the electric RF field is at the right phase, are accelerated in the cavity. The secondary yield of the surface is therefore quite important to sustain the resonance. The oxide layer in particular is known to enhance this secondary coefficient. If the cavity geometry is properly designed, the multipacting barriers are usually easily crossed and will not hinder the cavity performance.

4.2 Non-Quadratic Losses

4.2.1 Anomalous Losses

Anomalous (non-quadratic) losses appear when a cavity is submitted to surface fields exceeding 100 mT, thus rapidly limiting the performance [58]. Recent experiments are now bringing some light on the mechanism of this Q degradation at high fields. Baking the cavity at moderate temperatures (80 to 150°C) reduces the effect [59]. This clearly suggests a diffusion process. Analysis of oxygen diffusion in the bulk, reducing the oxygen concentration on the surface, at the expense of a slight degradation in the first 10 to 50 nm is consistent with the experimental observations. The measured surface resistance at T = 4.2 K is also reduced suggesting a lower RRR in the London penetration depth.

Q-slope in many cavities is observed even at low field whenever superconductivity is degraded. This is the case for low RRR material but also for thin films [**60**] or seamless cavities [**61**]. This slope

is generally attributed to defects (impurities, dislocations from cold work, grain boundaries [**11**]) in the very first 50 nm from the inner surface. Incidentally, this is the reason why, up to now, any trial to replace niobium by another higher critical temperature material has failed. Even thin film technology has ended up using pure niobium films [**62**], with results however significantly lower than for thick metal. The fact that only a mild baking (at 100°C) may affect the high field Q-slope is also pointing towards a surface effect. In addition, while after baking a high pressure rinsing have no effect, a pure HF bath only a small one, a 1 μm chemistry is enough to remove all the baking effect and brings back the cavity to a similar state where it was prior to baking.

4.2.2 Oxide Layer

On the other hand, while studying the very first oxide layer growth on a single crystal niobium, I. Arfaoui [**63**] have shown using UV-ray photon spectroscopy that the two first layers of the niobium metal on the surface right underneath the oxide, are heavily charged in oxygen dissolved atoms. He calculated that the upper two first layers of the niobium metal contain over 10% of oxygen (Fig.). If now one considers a very thin layer (2 nm) of a heavily degraded superconductor on top of an extremely pure niobium, one can analyze the overall behavior as a two-layer superconductor model. Because of the non-uniformity in the material depth, the order parameter Ω has to be calculated [**26**] following equation (1).

Due to the fact that the first heavily degraded layer thickness is much smaller than the coherence length, the proximity effect makes Cooper pairs survive even in the bad layer. That is true up to a given magnetic field, B_{c1}, much lower than the superheating field in the pure niobium region and this is most probably the value at which the Q-slope starts showing up.

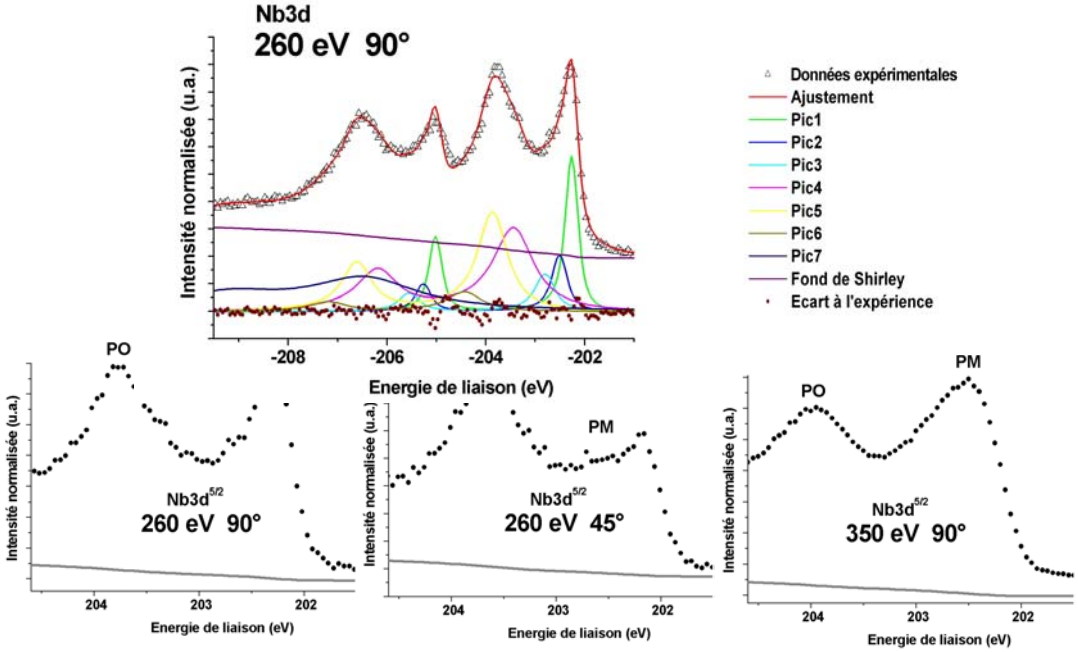


Fig. 15 The metal and oxide lines are clearly separated when using UV photon spectroscopy. By analyzing different angle and energy spectra, the contribution of the sub-oxide layer can be deduced.

Now the baking effect. When heating at relatively low temperatures (100°C), and apart from hydrogen (which is mobile even at room temperature), only the light impurities embedded in the niobium metal may very slightly diffuse. For example, oxygen can travel up to 50 nm in distance after heating at 120°C for 48 hours (Fig.). But that is enough to affect the superconducting properties of the cavity.

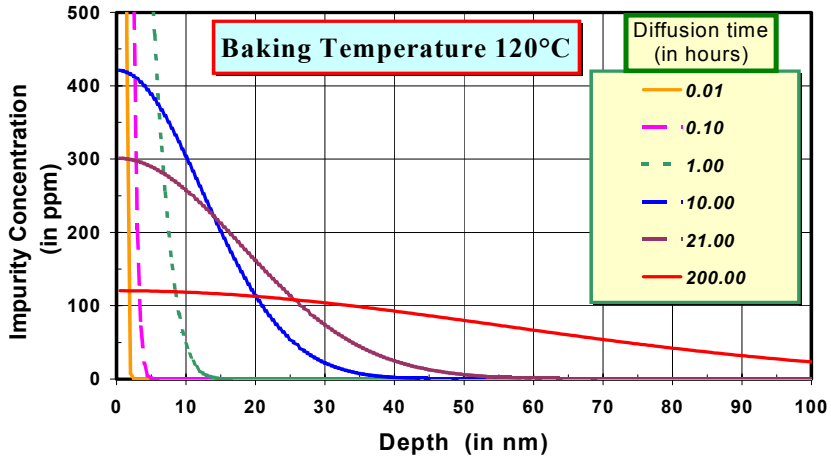


Fig. 16 Diffusion of oxygen in the niobium metal. After 200 hours at 120°C, the heavily contaminated surface will be almost uniformly diffused inside the material up to a depth exceeding 50 nm.

As a consequence of baking, the "bad" superconducting thin layer is spread over the whole bulk niobium at distances of the order of the London penetration depth (50 nm). Instead of having a very good and pure superconductor covered by a "bad" layer, one ends up with a more or less homogeneous superconductor with an "averaged" impurity content in the electromagnetic field penetration region of interest. The Q-slope is therefore decreased, at the expense of a reduced quench field and a higher residual resistance. An optimum in the "pollution" can be found in order to reach the maximum achievable quench field.

This picture is summarized in the next figure (Fig.) and is consistent with all the experimental observations described above.

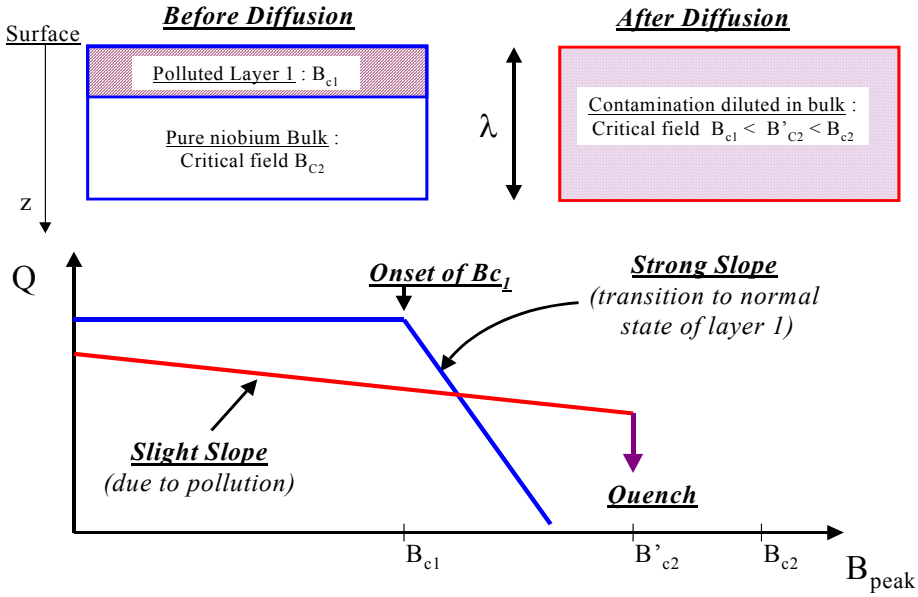


Fig. 17 Schematic of the thin "polluted" layer explaining the Q-slope behavior. After baking, the pollution is diluted up to a depth of the order of the London penetration.

If this scheme is confirmed, the oxide growth on the niobium metal has to be of great importance. The polluted layer underneath the oxide is connected to the sub-oxides appearing at the interface metal-oxide (Nb_4O and Nb_6O). The chemistry used should also affect the amount of dissolved oxygen, together with the preparation right after removing the cavity from the chemical bath. One must remember that after removing the cavity from the bath, an acid film is still lying on its

surface. After rinsing in ultrapure water, the very first oxide layer is growing quite rapidly in just a few minutes. These steps may be crucial for the formation of the polluted layer.

5. CONCLUSION

Despite the many surface issues encountered in SCRF cavity, the theoretical limits of niobium are now at reach. Using defect free bulk niobium sheets with low impurities contents (high RRR), etching with a good chemistry (smooth surface), using high pressure rinsing and mounting in a clean room, bring the performance of quench field over 170 mT with a very low residual resistance (less than 0.5 nΩ).

REFERENCES

[**1**] P. Schmuser, lecture on Superconductivity, these proceedings

[**2**] Bardeen, Cooper & Schrieffer, Phys. Rev. , 108, p. 1175 [1957]

[**3**] H. Padamsee & al., 4th Workshop on RF Superconductivity, Tsukuba, Japan [1989]

[**4**] B. Bonin, R. Roth, Particle Accelerators, 40, p. 59 [1992]

[**5**] V. L. Ginzburg & L.D. Laudau, Zh. Eksp. & Teor. Fiz., 20, p.1064 [1950]

[**6**] "Flux trapping in Superconducting cavities", C. Vallet, M. Boloré, B. Bonin, H. Safa, EPAC, Berlin [1992]

[**7**] J. Bardeen & J. Stephen, Phys. Rev., Vol. 139, 4A [1965]

[**8**] B.D. Josephson, Phys. Rev. Lett., 1, p. 251 [1962] and Advances in Physics, 14, p.419 [1965]

[**9**] V. Ambegaokar & A. Baratoff, Phys. Rev. Letters, 10, p.479 [1963]

[**10**] T.L. Hylton & al, Appl. Phys. Letters, 53, p. 1343 [1988]

[**11**] B. Bonin and H. Safa, "Power Dissipation at high fields in granular RF superconductivity", Superconductor Science & Technology, 4 , p. 257 [1991]

[**12**] B. Aune & al, "Degradation of niobium superconducting RF cavities during cooling time", LINAC Conference, Albuquerque, USA [1990]

[**13**] H. Safa, "Surface Resistance of a superconductor", 5th Workshop on RF Superconductivity, Hambourg, Germany, p.711 [1991]

[**14**] G. Müller, "Superconducting niobium in high RF magnetic fields", 3rd Workshop on RF Superconductivity, Argonne, Ill., USA [Sept. 1987], p.331

[**15**] H. A. Leupold & H. A. Boorse, Physical Review, 134, 5A, p. A1322, [1964]

[**16**] J. Ferreira Da Silva & al., Physica 41, p. 409 [1969]

[**17**] R.B. Flippen, Phys. Rev. Letters, 17, p.193 [1965]

[**18**] J.I. Gittleman & B. Rosenblum, J. of Applied Physics, Vol. 39,6 [1968]

[**19**] T. Hays, 8th Workshop on RF Superconductivity, Padova, Italy, (1997) p.789

[**20**] H. Safa, “An analytical approach for calculating the quench field in superconducting cavities”, Proc. of the 7th Workshop on RF Superconductivity, Gif-sur-Yvette, France (Oct. 1995), p.413

[**21**] G. Muller, “Superconducting niobium in high RF magnetic fields”, Proc. 3rd Workshop on RF Superconductivity, Argonne, Ill., USA [Sept. 1987], p. 331

[**22**] H. Padamsee, “Calculations for breakdown induced by large defects in superconducting niobium cavities”, CERN/EF/RF, 82-5 [1982]

[**23**] R. Roth, Proc. of EPAC, Berlin (1992)

[**24**] H. Padamsee, "The technology of Nb production and purification", 2nd Workshop on RF Superconductivity, CERN, Geneva, Switzerland (1984) p. 339

[**25**] H. Safa, “Influence of the RRR of niobium on the RF properties of superconducting cavities”, Advances in Cryogenic Engineering, Vol. 43, (1998)

[**26**] H. Safa, “Surface Resistance of a superconductor ”, Proc. of the 5th Workshop on RF Superconductivity, Hambourg (August 1991), p.711

[**27**] C. Benvenuti & al, 3rd Workshop on RF Superconductivity, Argonne, p. 445 [1987]

[**28**] J.L. Kirchgessner, 3rd Workshop on RF Superconductivity, Argonne, p. 553 [1987]

[**29**] C. Hauviller, IEEE PAC Conference, Chicago [1989]

[**30**] W. DeSorbo, Phys. Rev. 132, p.107, [1963]

[**31**] C.C. Koch & al., Phys. Rev. B, Vol. 9,3, p.888 [1974]

[**32**] H. Safa, “Influence of the RRR of niobium on the RF properties of superconducting cavities”, Advances in Cryogenic Engineering, Vol. 43, [1998]

[**33**] H. Safa & al., "Advances in the purification of niobium by solid-state gettering with titanium", Journal of Alloys and Compounds, 232 (1996) p. 281-288

[**34**] H. Safa, "High gradients in SCRF cavities", Particle Accelerators, Vol. 60, p. 219 [1998]

[**35**] J. Knobloch & al., 9th Workshop on RF Superconductivity, p. 77, Santa Fe, NM, USA [1999]

[**36**] E. Kako & al., 7th Workshop on RF Superconductivity, Gif-sur-Yvette, France, [1995] p. 425

[**37**] K. Saito & al., Proc. of the 8th Workshop on RF Superconductivity, Abano Terme, Padova [1997], p.472

[**38**] K. Saito & al., 4th Workshop on RF Superconductivity, KEK, Japan, [1989] p. 635

[**39**] E. Kako & al., 9th Workshop on RF Superconductivity, p. 179, Santa Fe, NM, USA [1999]

[**40**] L. Lilje & al., 9th Workshop on RF Superconductivity, p. 74, Santa Fe, NM, USA [1999]

[**41**] C. Antoine & al., 9th Workshop on RF Superconductivity, p. 109, Santa Fe, USA [1999]

[**42**] R.H. Fowler & L. Nordheim, Proc. Roy. Soc. London A, Math. Phys. Sci., 119, p. 173 [1928]

[**43**] R.J. Noer, Applied Phys. A, 28, 1, [1982] and all references therein.

[**44**] See for instance CERN, Cornell, Wuppertal and Saclay work on FE.

[**45**] R. Noer, 6th Workshop on RF Superconductivity, CEBAF, Newport News, USA, [1993] 236

[**46**] J. Tan, "Field Emission studies at Saclay and Orsay", 7th Workshop on RF Superconductivity, Gif-sur-Yvette, France, [1995] 105

[**47**] M. Jimenez & al., J. Phys. D: Appl. Phys., 27, p. 1038 [1994]

[**48**] B. Bonin, "Field Emission studies at Saclay",6th Workshop on RF Superconductivity, CEBAF, Newport News, USA, [1993] 1033

[**49**] J. Tan, PhD Thesis, University Paris VI, [1995]

[**50**] M. Luong, "Study on conducting protrusions enhanced field emission", 8th Workshop on RF Superconductivity, Padova, Italy [1997]

[**51**] J.H. Graber, PhD Thesis, Cornell University [1993]

[**52**] P. Bernard & al., 3rd EPAC, Berlin, Germany, (1992) 1269

[**53**] P. Kneisel & al., 6th Workshop on RF Superconductivity, CEBAF, Newport News, VA, USA (Oct. 1993), p. 628

[**54**] J. Tan & al., "Field Emission and High Voltage Cleaning …", Proc. of the 7th Workshop on RF Superconductivity, Gif-sur-Yvette, France (Oct. 1995), p. 365

[**55**] M. Luong & al., "Understanding and Processing of the Field Emission Enhanced by Conducting Protrusions", 8th Workshop on RF Superconductivity, Padova, Italy, [1997]

[**56**] H.A. Schwettman & al., J. Applied Physics, 45, p.914 [1974]

[**57**] C. Reece & al., "Improvement of the Operational Performance of SRF cavities via in-situ Helium Processing…", Proc. of Particle Accelerator Conf. PAC97, Vancouver, Canada [1997]

[**58**] H. Safa, "High gradients in SCRF cavities", Particle Accelerators, Vol. 60, p. 219 [1998]

[**59**] B. Visentin & al., 9th Workshop on RF Superconductivity, Santa Fe, USA, (1999)

[**60**] W. Weingarten, "Progress in thin films techniques", 7th Workshop on RF Superconductivity, Gif-sur-Yvette, France, [1995] p. 129

[**61**] V. Palmieri, "Seamless Cavities", Proc. of the 8th Workshop on RF Superconductivity, Abano Terme, Padova [1997], p.553

[**62**] C. Wyss, Proc. of Particle Accelerator Conf. PAC91, San Francisco, USA [1991]

[**63**] I. Arfaoui, Ph. D. Thesis, "Stochiometry, structure and high electric field strength of a ultra thin film of $NbO_{x\approx1}$ on Nb(110)", Ecole Centrale paris, March [2001]

RF CAVITIES FABRICATION

D. Proch
Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany

Abstract
In this paper the production of Niobium cavities is described. We will cover the extraction of Niobium from ore, the purification by electron beam melting and the fabrication of Nb sheets. The individual steps of Niobium sheet fabrication are described in detail: deep-drawing of cups, EB welding cells and dumb-bells, RF tuning of these units, preparation of the end groups, final welding, firing at 1400°C with Ti, frequency tuning, chemical etching, high pressure water cleaning and dust free assembly. Advanced technologies like spinning, hydroforming, back extrusion and electro-polishing will be commented. Finally the Niobium sheet fabrication will be compared with Cu-Nb coating technology.

1. INTRODUCTION

Superconducting cavities are (or were) in operation in many storage rings (HERA [1], LEP [2], Tristan [3], KEKB [4], CESR [5]) or linacs (Jefferson Lab [6], TFF-FEL [7]). At present the superconducting proton linac for the SNS spallation source is under construction [8]. In total more than 1000 meter of superconducting cavities have been operated and delivered about 5 GeV of accelerating voltage [9]. TESLA [10] is a proposal for a superconducting linear collider using more than 20000 Niobium cavities. Many other projects using superconducting cavities are under discussion (light sources, muon colliders,…). Most existing cavities are made from Niobium sheet metal, coating Cu resonators with a thin film of sputtered Niobium is an alternative if highest accelerating gradients are not required.

The fabrication of Niobium cavities has become a mature technology. Several companies are qualified as competent producers. The fabrication process of resonators for electron acceleration (relative velocity beta = v/c =1) is described with special reference to the TESLA design, especially considering mass production aspects. The design for medium beta application as for protons looks very similar. The shape of resonators for low beta application (like heavy ions) is different but the fabrication principles are the same.

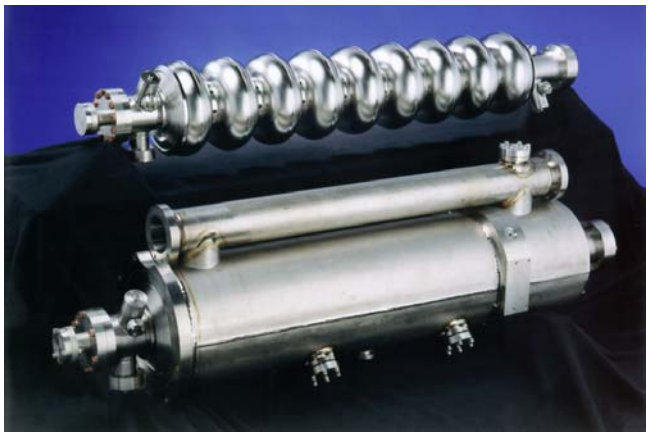


Fig. 1 Modified TESLA 9-cell resonator (with LHe tank at lower picture) as example for a β = 1 structure for acceleration of electrons (courtesy of Accel Instruments GmbH).

Fig. 2 "Split" ring resonator of Argonne laboratory, an example for a $\beta \leq 1$ structure to accelerate heavy ions .

2. PRODUCTION OF NIOBIUM

2.1 From ore to Niobium metal

Niobium and tantalum always occur in association with one another in nature, and this both as isomorphous niobate and tantalite of manganese and iron and in the form of higher grade niobite and tantalite. The latter contain 53 to 84% Ta_2O_5 and 47 to 78% Nb_2O_5. The most important reserves are the pyrochlorines (calcium niobate) out of Brasilia (about 80%) and Canada (about 10%) and zinc- or tin-slags containing niobium and tantalum out of Zaire, Nigeria and GUS as well as stibiotantalite from West Australia.

The most important commercial product for steel-production is the alloy Ferroniobium. Another important production part are optical glasses with are endowed with up to 20% Nb_2O_3, because Nb_2O_3 gives glass a very high refractive index.

The ores are processed to concentrates by opening up with melted alkalis or hydrofluoric acid. Then the niobium and tantalum are separated from one another using an older process of fractionated crystallisation. Modern separating methods are based on solvent extraction with MIBK (methyl isobutyl ketone, 4-methyl pentan-2-one). The next step is the chlorination. Then either a distillation and reduction with hydrogen takes place or an extraction by complete fusion electrolysis is performed to produce pure Nb. In industrial processes the niobium halides are converted to niobium oxide and this is reduced by carbon or aluminum to niobium powder or pellets. As a result of the high melting point of niobium, 2468 ^{o}C, much higher than that of most other materials, evaporation of impurities by electron beam melting in a vacuum typically better than $3 \cdot 10^{-4}$ mbar for the first melt and $2 \cdot 10^{-6}$ mbar in the last melt is very effective. Thus commercial grade niobium can be achieved by a minimum of four EB melting steps. The complete world raise amounts to 26000 t per year Nb metal (status 2001).

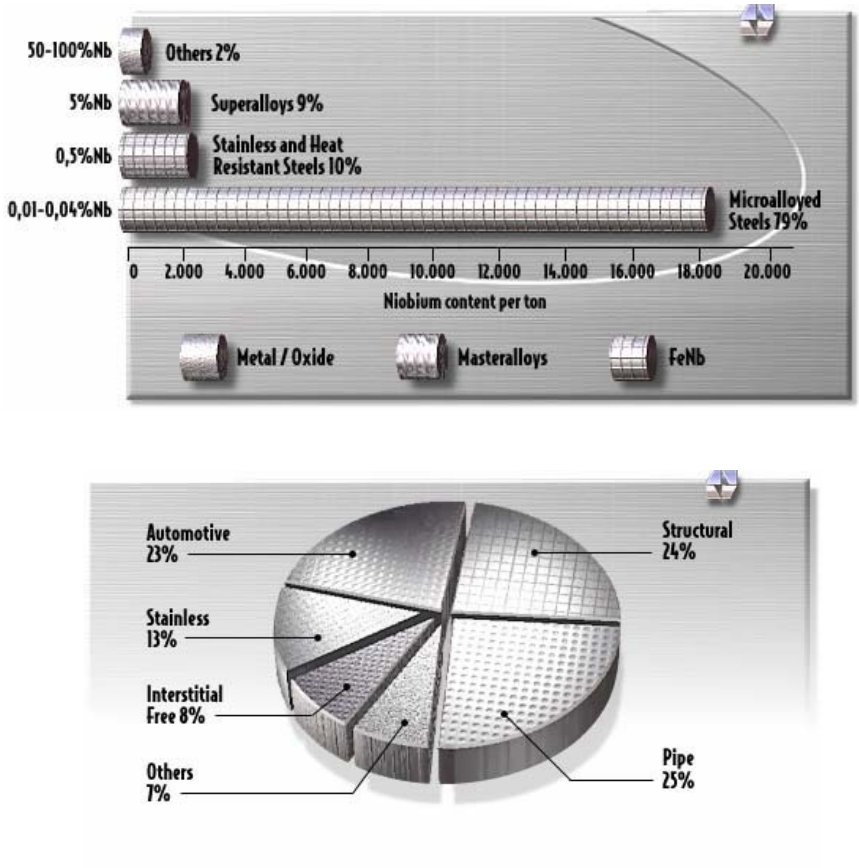


Fig. 3 Worldwide use of niobium (upper), use of niobium in steel industry (lower graph) (Courtesy of CBMM, Brasil).

2.2 Niobium purification and sheet production

The production of high grade niobium (RRR=300) for superconducting high frequency cavities is very sensitive to processing parameters. Therefore, it is absolutely necessary to have a very good analytical equipment for detecting the trace elements and a lot of exact physical test methods are required to find out the correlations between them. It was a challenging task for industry to produce an extremely pure niobium in a tonnage scale based on economic process steps. The niobium ingot is outgassed by several melting cycles in a high vacuum electron beam furnace (see Fig. 4, 5). The interstitial oxygen, nitrogen and carbon contamination is reduced to a few wppm (weight parts per million). The Nb ingots are forged and rolled into sheets (of 2.8 mm thickness in the case of TESLA cavities). After rolling, the Nb sheets are first degreased and cleaned by chemical etching. The sheets are then annealed for 1-2 hours at 700-800 °C in a vacuum oven at 10^{-5}-10^{-6} mbar to achieve full recrystallisation and a uniform grain size of about 50 µm. Fig.6 describes the main sequence of the production of high quality Niobium.

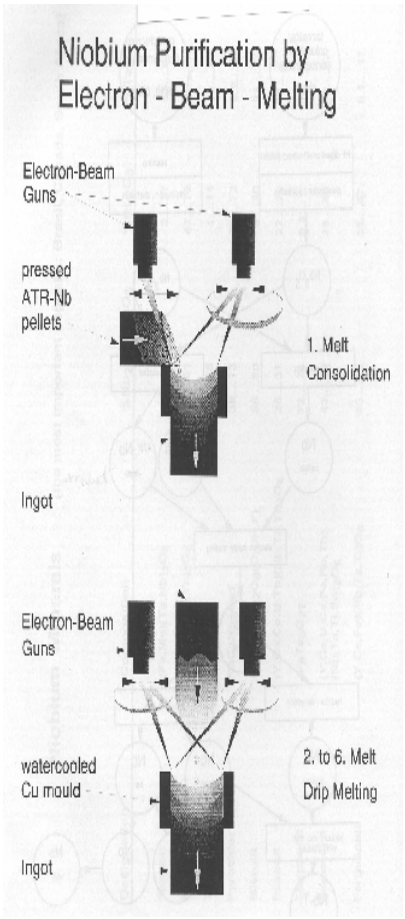


Fig. 4 Principle arrangement of the niobium purification by electron beam (EB) melting (courtesy of W.C. Heraeus, Germany).

Fig. 5 Picture of an EB melting furnace (courtesy of NIM, China).

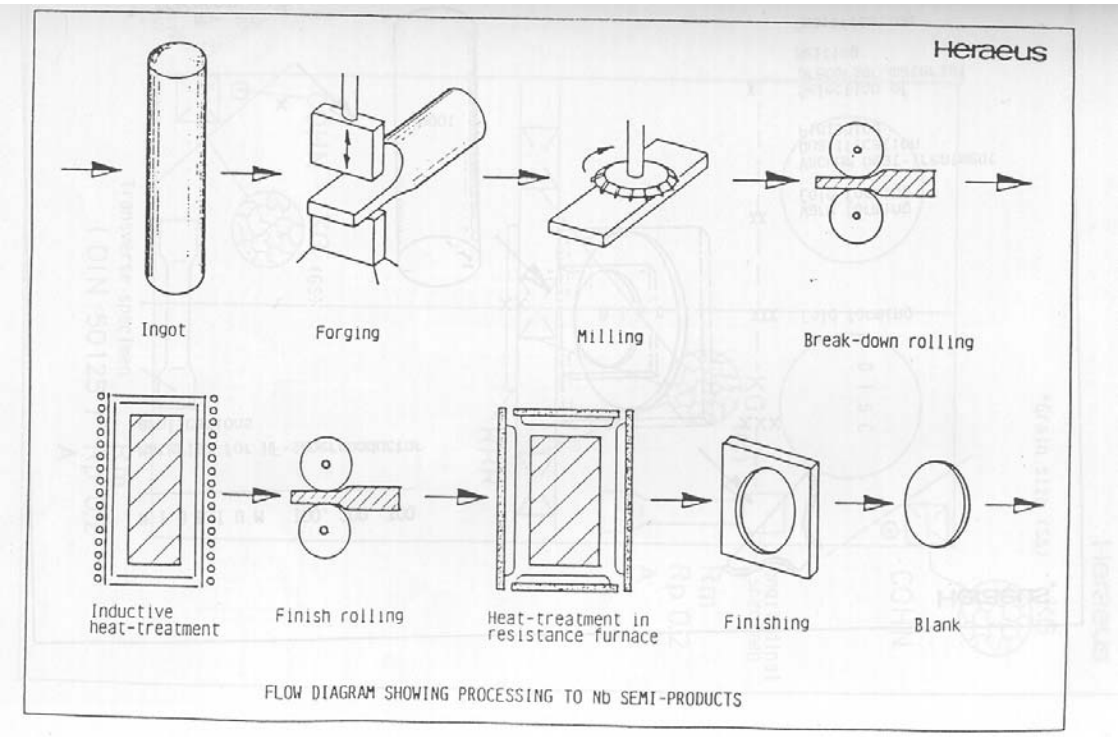


Fig. 6 Flow diagram of Nb sheet production (courtesy of W.C. Heraeus, Germany).

2.3 Niobium specification

The niobium specification for the TESLA cavities is given as typical example for the required quality (see table 1). The most important metallic impurity in niobium is tantalum, with a typical concentration of 500 ppm. The interstitially dissolved gases (mainly oxygen) act as scattering centers for the unpaired electrons and reduce the thermal conductivity.

The finished sheet must have "deep draw" quality to assure a smooth forming process. This is accomplished by the final recrystallisation heating. A resulting grain size of about 50 µm (corresponding to a ASTM grain size number of 4) is required. The sheet must be fully recrystallised to avoid inclusions of "hard" material, which will result in rupture during the forming process.

The mechanical properties (yield strength, tensile strength, hardness) should be isotropic in the plane of the sheet. Otherwise the cups will have an oval "spring back" after forming (by deep drawing or spinning, see chapter 4), i.e. they will not be of rotational symmetry. Therefore cross rolling is specified for the production.

Table 1
Technical Specification for Niobium used in TESLA Cavities.

Impurity content in ppm (wt)				Mechanical Properties	
Ta	= 500	H	= 2	Residual resistivity ratio RRR	= 300
W	= 70	N	= 10	grain size	~ 50 µm
Ti	= 50	O	= 10	yield strength	> 50 MPa
Fe	= 30	C	= 10	tensile strength	> 100 MPa
Mo	= 50			elongation at break	30 %
Ni	= 30			Vickers hardness HV 10	= 50

3. NIOBIUM MATERIAL INVESTIGATIONS

3.1 Control of Niobium quality

Several sample tests have to be carried out at the company in order to document the specified Niobium quality: chemical analysis of residual impurities during specified steps of production, metallurgical cuts to verify the required grain size and degree of recrystallisation, yield strength measurements on some sample sheets. Correct recrystallisation and resultant uniform yield strength data are crucial for a good forming process of the Niobium sheets. These data might vary between different production charges, therefore hardness measurements are done on all sheets as a cheap and fast global check. In case of too much variation of hardness further investigations are required for detailed analysis and/or additional heating cycles must be done. Yield strength measurements are done in orthogonal directions on the sheet to verify isotropic mechanical behaviour.

3.2 Visual inspection of Nb sheets

Niobium sheets should be visually inspected for obvious surface defects like scratches, voids, cracks,… In case of any findings the area should be ground, slightly etched and inspected again. Chemical etching is done by a volume mixture of 1 part hydrofluoric acid (50 % concentration), 1 part of nitric acid (70% concentration) and 2 parts of phosphoric acid (85% concentration). The removal rate of this "buffered chemical polish (BCP) at 15 °C is about 2 microns on the surface per minute. Immersing the Nb sheet in water over night will detect iron particles by the occurrence of a rust spot. In this case a sulfuric acid treatment will dissolve the iron and the inspection should be repeated.

3.3 Eddy current scanning

For a large scale cavity production the visual inspection should be replaced by an appropriate scanning technology. For this purpose an eddy current scanning apparatus was developed by the Bundesanstalt fuer Materialforschung (BAM) in Berlin and DESY [11]. In the xy scanning system the frequency used is 100 kHz corresponding to a penetration depth of 0.5 mm in niobium at room temperature. The maximum scanning speed is 1 m/s. The scanning probe containing the inducing and receiving coils floats on an air pillow to avoid friction. The machined base plate contains holes to evacuate the space between this plate and the Nb sheet. The atmospheric pressure is sufficient to flatten the 265 x 265 mm^2 niobium sheets to within 0.1 mm, which is important for a high sensitivity scan.

Fig. 7 Test set-up for eddy current check of Nb sheets (apparatus with a rotating plate holder).

In the meantime, an improved eddy-current scanning device has been designed and built at BAM which operates similar to a turn table and allows for much higher scanning speeds and better sensitivity since the accelerations of the probe head occurring in xy scans are avoided. A two frequency principle is applied in the new system. Scanning with high frequency (about 1 MHz) allows detection of surface irregularities while the low frequency test (about 150 kHz) is sensitive to bulk inclusions. The high and low frequency signals are picked up simultaneously. Very high sensitivity is achieved by signal subtraction. Further increase of sensitivity is under development by the use of a SQUID element as eddy current detector.

More than 1000 Nb sheets for the TTF cavities have been inspected. Typically 5% of the sheets have been rejected; about half of it presented severe inclusions whereas the other half showed suspicious scanning results. These sheets were used for trimming the deep drawing tools or for some other material investigations like parameter optimization of electropolishing or EB welding.

3.4 SQUID Scanning

Some Nb sheets of TESLA dimensions with material inclusions, found by eddy current scanning, were tested with a SQUID gradiometer system. The SQUID system features higher sensitivity. An iron particle with a dimension less then 50µm, not detectable with the eddy current method, was easily found inside the Nb sheet. Detection of rather deep inclusions in the material or investigation of the inside surface of closed parts (cavities, tubes) is possible because of the high penetration depth of the SQUID method. These investigations were conducted in collaboration between the FIT Messtechnik GmbH and DESY. Recently a new research program was started in collaboration between WKS Messtechnik GmbH, W.C. Heraeus GmbH, IAP University of Giessen and DESY to develop a SQUID scanning apparatus for industrial use.

3.5 Other material investigation

Other methods of material controls have been applied to check the quality of Niobium sheets:

- Neutron activation to search for Ta inclusions
- Scanning electron microscopy with energy dispersive X-ray analysis (EDX) to identify foreign elements
- X-ray radiography to find foreign inclusions
- X-ray fluorescence to identify foreign elements.

These techniques are not suitable for a fast 100% scanning quality control but are very helpful to investigate identified performance limitations in superconducting cavities and guide R&D to upgrade the Niobium material performance.

4. FABRICATION OF CUPS

4.1 Forming of cups by deep drawing

Cavities are constructed from niobium sheets by deep drawing or spinning of half cells followed by trim machining and electron beam welding.

Deep drawing is a process whereby a blank is pressed into shape using a set of dies as shown in Fig.8. These dies are machined out of aluminum alloy because other materials like steel tend to friction welding. The starting blank for deep drawing is a Nb disk cut from a sheet. It is bolted across the female die with a hold-down plate. With an appropriate torque on the bolts, the outer edge of the Nb is constrained without tearing at the clamped edges. Clean oil is painted onto the Nb for lubrication. Alternatively an elastomer sheet replaces the oil film. The male die is placed in position and the assembly is squeezed in a hydraulic press. For a 1.3 GHz cavity about 100 t are applied. To

get the curvature required at the iris, the nose of the cup can be more precisely shaped by a further step of coining.

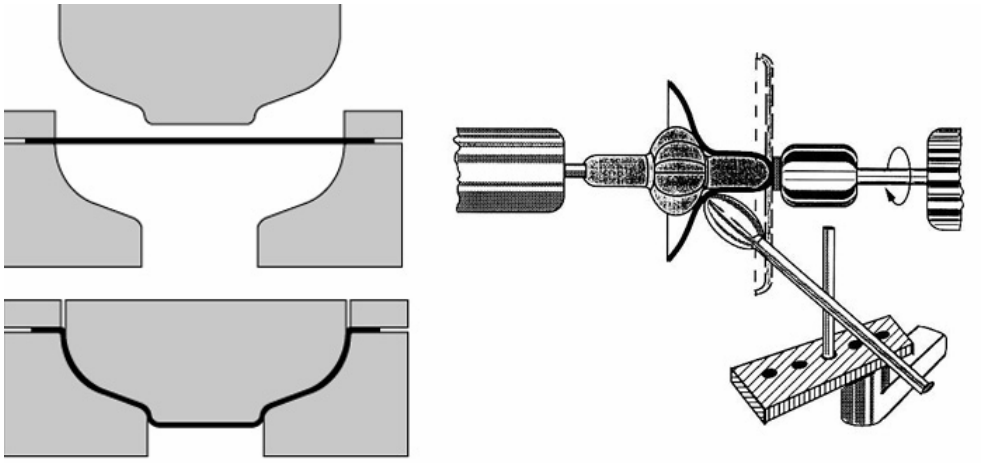


Fig. 8 Principle arrangement of deep drawing (left) and spinning (right) technology (courtesy of H. Padamsee and E. Palmieri).

The described deep drawing process requires a tight thickness tolerance of the sheet material. If the sheet is too thin, the contour of the cup will vary because the gap between male and female die is too large. For a too thick material the sheet might rupture. An alternative forming technique uses an elastomer cushion as part of the female tool. Hereby the inner surface (RF side) of the cup will closely match the male die independent of the sheet thickness.

4.2 Forming of cups by spinning

Spinning is a well known technique in forming metal bodies of rotational symmetry. The metal disc is pushed over the rotating mandrel as depicted in Fig. 8. The effort for tooling is small and no high pressure hydraulic press is needed. However, the contour and the wall thickness of the cup will not be as precise as by deep drawing. Therefore this technology is preferred to produce in a fast and cheap way prototype cavities. Also, at low frequencies (500 MHz and lower) with large size cups, spinning is preferable to avoid the effort of large hydraulic equipment.

4.3 Frequency measurement of cups

The contour, or better, the frequency of the cups must be measured to iterate the shape of the deep drawing tools and to control the precision of the forming process during series production. For this purpose the cups are cut at the equator and iris plane with some mm excessive length for welding shrinkage and possible length tuning. The resonance frequency of a cup is measured with shorting plates (made from Niobium to avoid contamination of the cup end faces by foreign material) the iris and equator plane. A good RF contact is crucial for precise and reproducible measurements.

The measured spread of cup frequencies (9-cell cavities for TESLA-TTF) was smaller by a factor of two for the drawing method with an elastomer cushion as compared to using the standard deep drawing tool (as shown in Fig. 8).

5. ELECTRON BEAM WELDING

5.1 Welding parameters

All of the parts are electron-beam welded together in a vacuum better than $5 \cdot 10^{-5}$ Torr. The weld parameters are chosen to achieve full penetration welds using a defocused electron beam. For

example, to achieve full penetration in 1.6 mm thick Nb, a typical set of electron-beam welding parameters are V = 50 kV, I = 37 mA at a 46 cm/min weld speed**.** One technique to achieve a well defined and reproducible defocused beam is to deflect the beam periodically so that a shape of a rhombus or ellipse is "printed" on the weld area. Welding from the inside (RF side of the cavity) is preferred to get a high quality surface. For geometric reasons welding from the outside might be needed. In this case the welding parameters must be optimized to produce a smooth underbead at the RF surface. It is important not to use a focused beam (as is usually done with electron beam welding) to avoid weld spatter. Also, the vapor column from a focused beam may leave voids in the molten Nb welding seam.

The welding parameters must be selected not to "freeze in" impurities, which are usually found in the front end of the molten zone. For example, a slow reduction of the welding power is required (the Niobium part still under rotation) to get a smooth ending of the welding seam.

The surface geometry of the Niobium parts to be welded together might be flat or contain a small step. In the latter case alignment is easier but chemical cleaning of the Niobium parts before welding might erode the step. The chemical cleaning of the welding area (several μm) should be done immediately before the welding process in order to avoid contamination by storage or additional handling.

5.2 Welding cups and dumb bells

After frequency measurement of the cups (see Fig. 9) the iris and equator plane is cut to the right length (see section 7, Frequency tuning). Weld shrinkage must be included. Furthermore it is advisable to include some excessive length at the equator for compensation shape deformation of the cup due to the later welding of the stiffening ring. After cleaning, two cups are welded at the iris. In a second step the stiffening rings are welded from the outside. Welding parameters are critical not to introduce too much heat with resultant cup deformation but to produce a mechanically rigid connection.

At this stage the frequency of the dumb bells should be measured (with an arrangement similar to the cup frequency measurement shown in fig. 9) and the final length trimming at the equator can be done. The next finished dumb bell will be welded at the equator forming a double dumb bell. In this way the inner cells of the cavity will be produced.

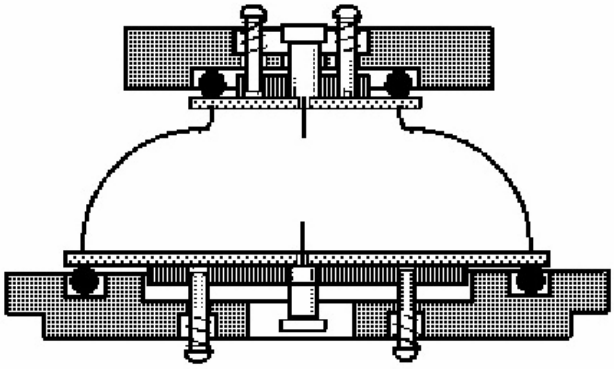


Fig. 9 Principle arrangement of frequency measured of a Nb cup.

5.3 End group welding

The end group consists of the outer cavity cup, the beam pipe, input coupler extension, higher order mode couplers, flanges and transition part to the Helium container ("conical head disk"). Niobium pipes are fabricated by spinning or back extrusion. The flanges are made from NbTi in order to keep

mechanical hardness after the high temperature firing of the cavity at 1400°C with Ti. If this firing is not required a stainless steel flange (conflat sealing design) can be brazed to the Niobium.

All parts of the end group are cleaned and electron beam welded according to the cup welding. The cavity will be completed by welding the equator of the end cells (see Fig. 11, 12).

5.4 Alternative welding sequence

Sequential welding of cups, dumb bells, multi dumb bells and end groups is an established fabrication procedure. If handling and welding parameters are explored in detail, all dumb bells and the end groups can be welded in one welding cycle. This will reduce welding costs and is inevitable for large scale mass production.

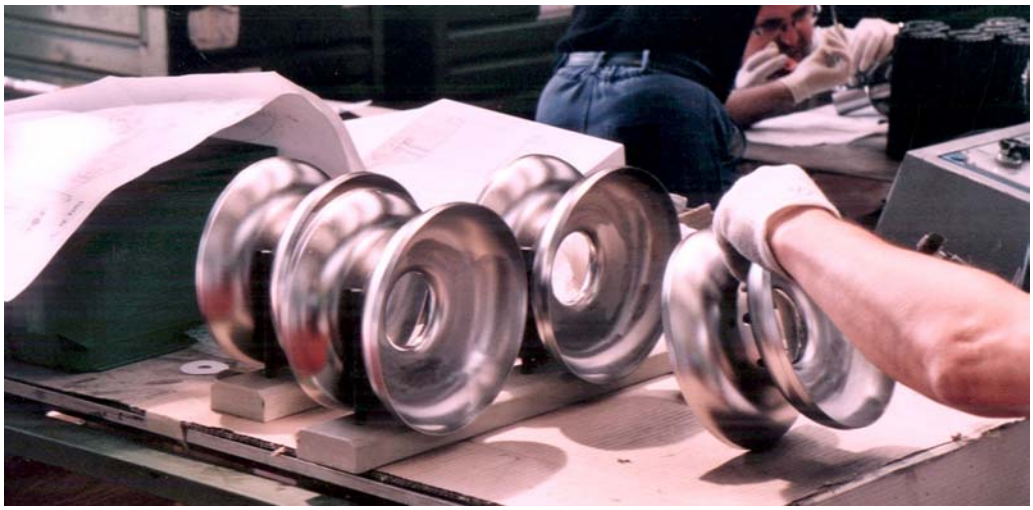


Fig. 10 Several dumb bells during optical inspection (courtesy of ZANON).

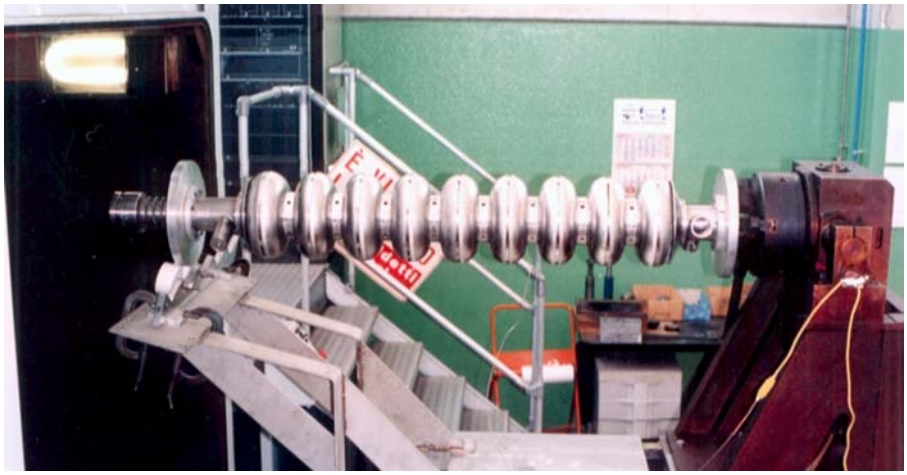


Fig. 11 Nb cavity fabrication (courtesy of CERCA): several dumb bells and two end groups in front, two completed 9-cell TTF cavities after electron beam (EB) welding and EB welding machine in the background.

Fig. 12 Completed TESLA 9-cell cavity leaving the EB welding machine (courtesy of ZANON).

6. ALTERNATIVE PRODUCTION METHODS

6.1 Niobium sputtering on Copper

The sputter technology has been developed at CERN [12]. More than 300 cavities for LEP have been produced by industry. Copper cavities are fabricated similar to the Niobium sheet technology. The inner surface is coated with a several μm thick Niobium film by magnetron sputtering. These cavities do not show magnetic loss by the earth magnetic field, operate at higher quality factor than Niobium cavities but start to degrade above 10 MV/m accelerating gradient. The sputter technology is an attractive alternative fabrication method for moderate gradient application. This is especially true at lower frequencies (below around 500 MHz) where the price of Niobium for the large cavities dominates the fabrication costs. The principle sputter arrangement is shown in Fig. 13.

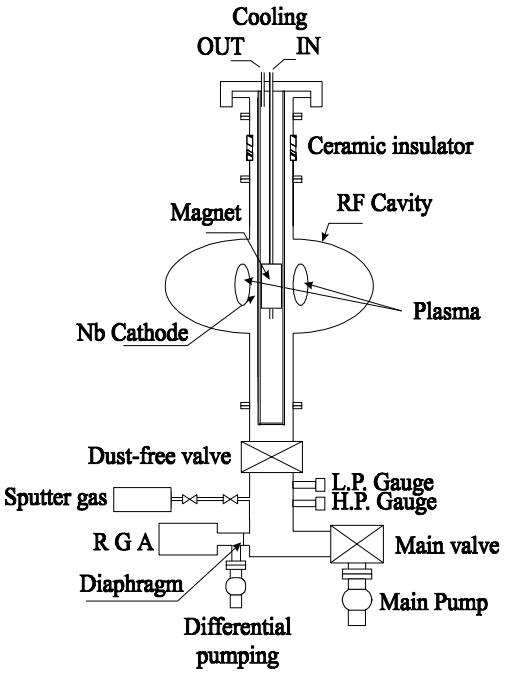


Fig. 13 Principle arrangement of the Nb-Cu sputter coating process (courtesy of C. Benvenuti, CERN).

6.2 Hydroforming

There is hope to form the body of a cavity (excluding the end groups) by hydraulic expansion of a Niobium tube (see Fig. 14). This method would avoid welds in the region of high magnetic and electric fields and promises to save production costs. In the past, all experiments failed because of early rupture of the tube. The maximum elongation of the tube material was too small. Furthermore, the material properties of the tube differ in the longitudinal and circumferential direction. Intermediate annealing could solve this problem but is not attractive because of effort and cost. Recent developments in tube production resulted in a considerably improved quality to allow hydraulic expansion of a cell without annealing. Several single cell cavities have been fabricated and were successfully tested [13]. In Fig.15 the first 2 cell hydroformed cavity is shown. The fabrication of a 9-cell TESLA cavity is expected soon.

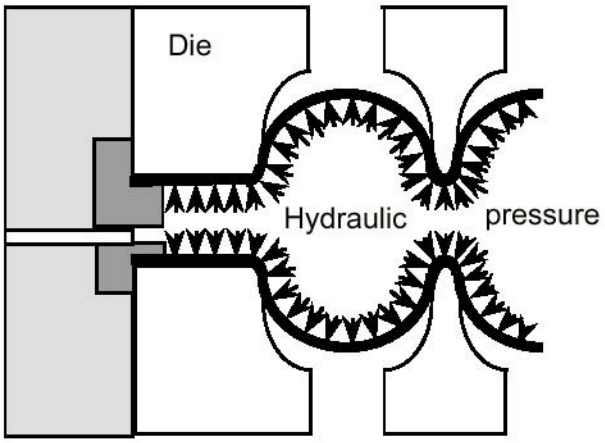


Fig.14 Principle arrangement of hydro-forming a cavity from a tube. It is important to coordinate the hydraulic pressure and the movement of the die for the expansion of the tube without rupture.

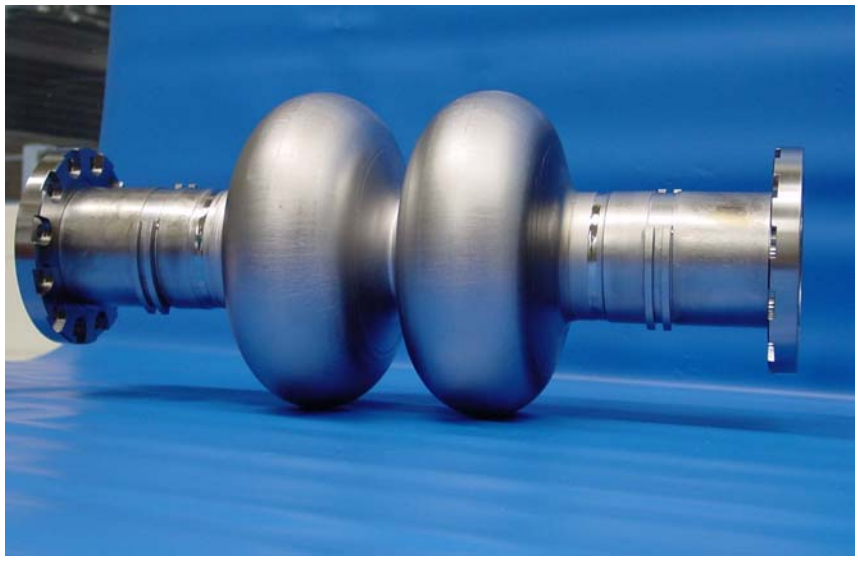


Fig. 15 Two cell Nb cavity fabricated by hydro-forming without intermediate annealing (courtesy W.Singer, DESY).

With the same technology a tube with a sandwich material of solid Copper and Niobium was formed to a single cell. Hereby the outer Copper will stiffen the inner Niobium so that the stiffening ring might be avoided. The inner Niobium sheet could be made thinner than the present 2.8 mm thickness. This would reduce the costs for the expensive Niobium. Furthermore an improved cooling of the superconducting RF surface by the copper can be expected if the interface between Niobium and Copper is thermally matched. A possible drawback is the observation that a quench in such a sandwich material tends to reduce the quality factor of a superconducting cavity. A possible explanation of this effect is that the magnetic flux is frozen when the cavity is thermally cycled above Tc during the quench.

6.3 Spinning

Spinning is successfully applied to fabricate Niobium cups (see Fig. 8). Several single and multicell cell cavities were spun from Niobium sheets. Measurements concluded that good RF performance can be expected but there is need for considerable material removal of the inner surface. Also keeping the wall thickness constant seems to be more difficult than in the case of hydroforming.

7. FREQUENCY TUNING

After fabrication the cavity resonance frequency as well as the electrical field flatness is measured. The cavities are tuned by slightly changing the length of the individual cell. This is accomplished by a mechanical or hydraulic operated ring which is mounted at the outside of the iris gap. Lengthening will increase the resonance frequency because the cavity contour is modified at the electric field region.

Correcting the frequency will result in a new cavity length, which cannot be changed at this stage. A possible length correction must be done at the earlier stage when trimming the cup or dumb bell length. Here a longer cell or dumb bell will decrease the resonance frequency because the magnetic field region is altered. Therefore, the cavity resonance frequency after welding must be extrapolated from the measured cup, or better from the dumb bell frequency, and trimming must be carried out accordingly. By this method the TESLA 9-cell resonators could be adjusted within ±1 mm length.

In the accelerator, fine tuning (of about ±200 kHz) is accomplished by a tuning rig: the total cavity length is adjusted to establish the exact accelerator frequency.

8. POST PURIFICATION

If accelerating fields greater than 20 MV/m are desired, the thermal conductivity of the Nb in the completed cavity may be improved by a "post-purification" step. The purity of the Nb is increased by solid state gettering of the interstitial oxygen using titanium at a temperature of 1400°C. The foreign metal is vapor deposited on the Nb surface and, in the same step, the high temperature decreases the diffusion time of the oxygen in Nb. Over a few hours, oxygen collects in the deposited layer. Ti diffuses 50-100 μm along the grain boundaries, requiring a subsequent heavy etch. Typically, the RRR improves by a factor of 2-3 in a few hours primarily due to the removal of interstitial oxygen.

The cavity must be cleaned before heating to avoid diffusion of impurities into the bulk. Chemically etching away ~ 30μm is generally a sufficient cleaning.

The heat treatment also homogenizes the Nb as indicated by the reduction of magnetic flux pinning centers found by magnetization measurements.

A severe drawback of post-purification is the considerable grain growth of the niobium: post-purified cavities are vulnerable to plastic deformation and have to be handled with great care.

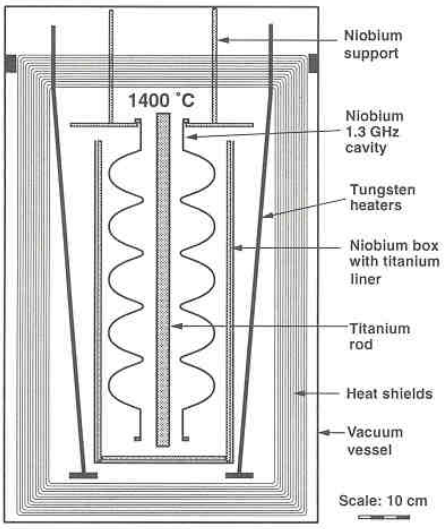


Fig. 16 1400 ^{0}C furnace to post-purify the Nb cavity by Ti solid state diffusion.

9. CHEMICAL CLEANING

9.1 Buffered chemical polish

A layer of 100-200 μm is typically removed in several steps from the inner cavity surface to obtain good RF performance in the superconducting state. If solid state gettering at 1400°C is applied, the main cleaning is done after the heating process because Ti must be removed from the surface, too.

The standard method applied for niobium cavities is called Buffered Chemical Polishing (BCP), and uses an acid mixture of HF (48 %), HNO_3 (65 %) and H_3PO_4 (85 %) in the volume ratio of 1:1:2. The acid is cooled (typically to 15 ºC) and etches at a rate of about 1 μm per minute. A potential danger during chemical treatment of Nb is the "Q-disease" which is known to be caused by hydrogen contamination. Experience has shown that hydrogen contamination can be avoided by keeping the acid temperature ≤18 ^{0}C.

Chemical etching of Nb parts is done by immersing them completely in the acid bath for the required time. Complete cavities are filled with acid using a "closed" system. Hereby the amount of acid is kept small and cooling can be accomplished by a heat exchanger in the circulating acid. Valves can be opened and closed remotely to fill the cavity with cooled acid, drain the acid, and rinse the cavity with ultrapure water. The rinsed water is sent to a neutralization tank.

9.2 Electropolishing

The advantage of electro-polishing as compared to buffered chemical polishing is a smooth, mirror-like surface finish with no sharp steps at the grain boundaries.

The Niobium cavity is placed horizontally and rotates about its axis with 0.5 to 1 rpm. It is half filled with electrolyte consisting of H_2SO_4 and 40% HF in a ratio of 85:10 by volume. Nb is the positive electrode (anode) and the cathode is made from Al. The cathode must be enveloped by a porous Teflon membrane to keep Hydrogen bubbles away from the Nb. Depending on details of the electropolishing process it might be necessary to remove the hydrogen by heating in a vacuum furnace at about 800^{o}C after electro polishing. The thickness of the removed Nb is estimated from the total charge and cross-checked with an ultrasonic thickness gauge.

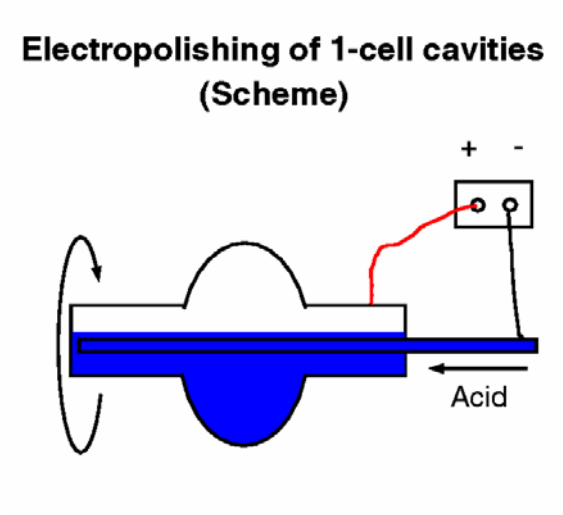


Fig. 17 Principle arrangement of electropolishing.

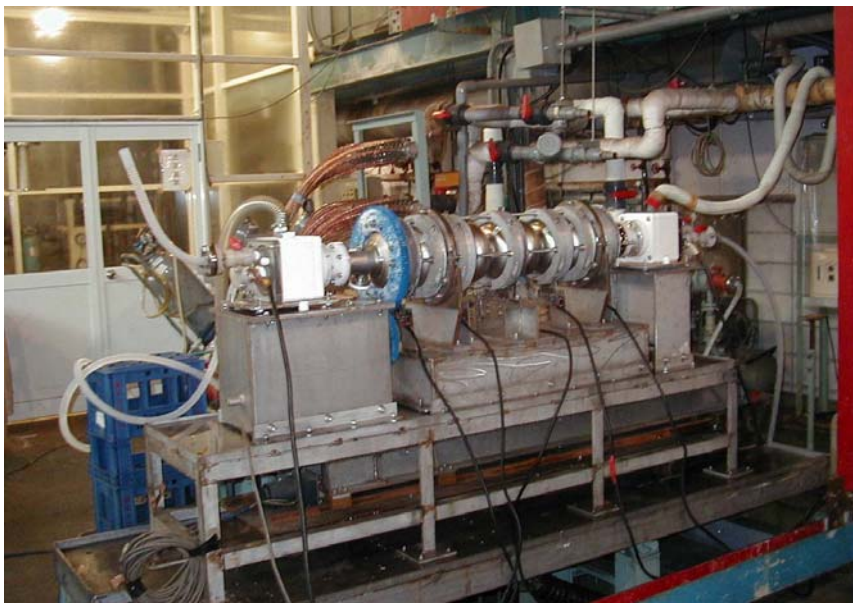


Fig. 18 Electropolishing of a 9-cell cavity (courtesy of Nomura Plating, Japan).

10. FINAL CLEANING AND ASSEMBLY

10.1 High pressure water cleaning

After acid etching or electro-polishing, the cavity is placed in a closed loop with the ultrapure water system for several hours. The resistivity of the water should be close to theoretically pure (> 18 M-cm) and the water inlet should be filtered to eliminate particles >0.3 μm. Water is recirculated for several hours through the cavity in series with the water purification system to continuously and thoroughly remove any chemical and particulate residue from the Nb surface. As last step the cavity is cleaned by high pressure rinsing (100 bar) with ultrapure water through multiple steel nozzles each having a 0.3 mm diameter orifice. With ultraclean water inside the cavity, the plumbing attached to

the cavity is pinched off and the cavity is carried into the dust-free clean room where the water is drained. Thus, the cavity surface only comes in contact with filtered air after the final ultrapure water rinse.

10.2 Clean assembly

The cleaned surface is exposed only to filtered air in a Class 10-100 clean room. The laboratory workers in the vicinity need to wear special particulate-free clothing and follow strict protocols to reduce particulate generation.

After rinsing, the surface is dried by filtered nitrogen gas through the cavity. Once the surface is dry, all the cavity's ports are sealed, and the cavity is attached cleanly to its test stand and evacuated. One can also assemble the wet cavity and dry it by applying heat while pumping.

Care must be taken to ensure that the vacuum system is thoroughly clean and dust free. The cavity itself is evacuated slowly to avoid turbulent flow and reduce the risk of contaminants from the vacuum system reaching the cavity. It is essential to maintain dust free conditions, while attaching the input and output couplers to the cavity.

Fig. 19 Cavity preparation; string and module assembly.

11. COST CONSIDERATION, MASS PRODUCTION FOR TESLA

Worldwide there are several qualified companies for Niobium delivery and cavity production. The production cost will depend on the size (frequency) and number of cells. As an example, the prize for a TESLA 9-cell cavity (1.3 GHz) is approximately 50 k euro at the year 2000 index (quantity of 20, no tooling). The fabrication price is dominated by the welding costs (50%). An industrial study for the fabrication of 20000 cavities for TESLA predicts considerable savings: e.g. the use of a three chamber welding equipment and the application of tools for multiple parts welding will reduce the average welding time of a cavity by a factor 5.

There is no shortage of raw Niobium material even for the required 500 cavities for TESLA. Refining the Niobium to the RRR 300 quality might need some investment in melting facilities to supply the required amount in the requested time frame of 3 years.

ACKNOWLEDGEMENTS

I wish to thank all my colleagues for contributions and discussions. Support of pictures and data from the following companies is greatfully acknowledged: Accel (Germany), CBMM (Brasil), CERCA (France), Heraeus (Germany), NIM (China), Nomura Plating (Japan), Zanon (Italy).

REFERENCES

[1] B. Dwersteg et al, operating Experience with Superconducting Cavities in HERA, Proc. 4th EPAC, London (1994) 2039

[2] G. Cavallari et al, Acceptance Tests of Superconducting Cavities and Modules for LEP from Industry, Proc. 4th EPAC, London (1994) 2042 - 2044

[3] S. Noguchi et al, Recent Status of the TRISTAN Superconducting RF System (Proc. 4th EPAC, London (1994) 1891 - 1893

[4] Shin-Ichi Kurokawa, B-Factory Commissioning and First Results, Proc. Pac 2001, Chicago (2001)

[5] S. Belomestnykh et al, Running CESR at High Luminosity and Beam Current with Superconducting RF System, Proc. 7th EPAC, Vienna (2000) 2025 – 2027

[6] C. Reece et al, Performance Experience with the CEBAF SRF Cavities, Proc. IEEE PAC, Dallas (1995) 1512

[7] J. Andruszkow et al, first Observation of Self-Amplified Spontaneous Emission in a Free-Electron Laser at 109 nm Wavelength, DESY 00-066, ISSN 0418-9833 (2000)

[8] T.E. Mason, The Spallation Neutron Source: A Powerful Tool for Materials Research, Proc. Pac 2001, Chicago (2001)

[9] D. Proch, Superconducting cavities for accelerators, Rep. Prog. Phys. 61 (1998) 431-482

[10] TESLA – Technical Design Report, DESY 2001-011/ECFA 2001-209/TESLA Report 2001-23/TESLA FEL 2001-05 (2001)

[11] W. Singer et al, Diagnostic of Defects in high Purity Niobium, Proc. 8th SRF Workshop, Abano Terme, LNL-INFN (rep) 133/98 (1998) 890

[12] C. Benvenuti et al, Superconducting Niobium Sputter-coated Copper Cavity Modules for the LEP Energy Upgrade, Proc. IEEE PAC, San Francisco (1991) 1023

[13] W. Singer et al, Hydroforming of TESLA Cavities, Proc 7th EPAC, Vienna

BIBLIOGRAPHY

A.W. Chao, M. Tigner (ed.), Handbook of Accelerator Physics and Engineering (World Scientific, 1999) ISBN 9810235003

H. Padamsee, J. Knobloch, T. Hays, RF Superconductivity for Accelerators (John Wiley & sons, Inc., 1998) ISBN 0-471-15432-6

DESIGNING SUPERCONDUCTING CAVITIES FOR ACCELERATORS

Hasan Padamsee
Cornell University, Ithaca, NY 14853

Abstract
Rapid advances in the performance of superconducting cavities have made RF superconductivity a key technology for accelerators that fulfil a variety of physics needs: high energy particle physics, nuclear physics, neutron spallation sources, and free electron lasers. New applications are forthcoming for frontier high energy physics accelerators, radioactive beams for nuclear astrophysics, next generation light sources, intense proton accelerators for neutron, neutrino and muon sources.

1. INTRODUCTION

The goal of this paper is to discuss design choices for superconducting cavities for various accelerator applications. Two classes of considerations govern structure design. The particular accelerator application forms one class, and superconducting surface properties the other. Designing a superconducting cavity is a strong interplay between these two classes. Typical accelerator driving aspects are the desired voltage, the duty factor of accelerator operation, beam current or beam power. Other properties of the beam, such as the bunch length, also play a role in cavity design. Typical superconducting properties are the microwave surface resistance and the tolerable surface electric and magnetic fields. These properties, which are also discussed in different lectures at the Erice School [1,2,3], set the operating field levels and the power requirements, both RF power as well as AC operating power, together with the operating temperature.

Accordingly, the plan of this article is to briefly describe a number of distinct accelerator applications and the structures which emerged. To understand the evolution we first discuss the key electromagnetic properties of accelerating structures leading to an analysis of the power requirements of superconducting accelerators. The behavior of superconducting surfaces exposed to high surface electric and magnetic fields provide a guide to tailoring the cavity shape to achieve desirable values of key properties for accelerator performance. Both accelerator and surface issues govern the choice of the cavity shape, beam aperture, number of cell per structure as well as the choice of the RF frequency. Mechanical properties also play a role in the design aspects. Finally, input and output power coupling issues interact with cavity design, but these are covered in a different lecture[4].

The discussion here is an overall summary and review of design aspects. We refer the reader to the reference text [5] and review article [6] for a more thorough discussion of many of the design topics and their intimate relationships to the physics of RF superconductivity.

Figure 1 shows a variety of superconducting accelerating cavities, ranging in frequency from 200 MHz to 3000 MHz and ranging in number of cells from one to nine. Most are cavities fabricated from pure sheet niobium and some, especially at frequencies below 500 MHz, are made of copper sputtered with a micron thin layer of niobium. Cavity fabrication issues are discussed in other lectures at the Erice School [3]. All the cavities of Figure 1 are intended for accelerating particles moving at nearly the velocity of light, i.e. $v/c = \beta \approx 1$. Accordingly, the period of a long structure (or the accelerating gap) is $\lambda/2$, where λ is the RF wavelength. Particles moving at $v \approx c$ will cross the gap in exactly a half RF period to receive maximum acceleration.

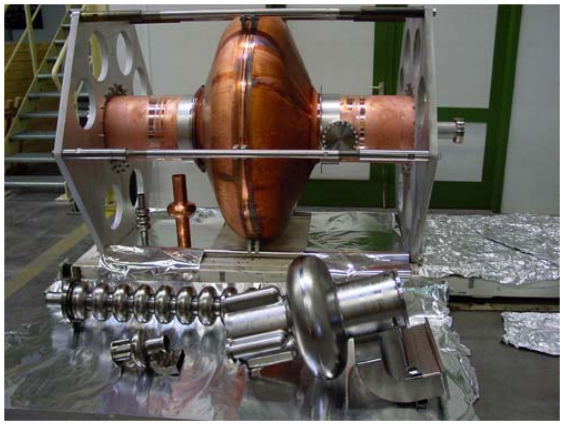


Fig. 1 A spectrum of superconducting cavities.

2. ACCELERATOR REQUIREMENTS AND EXAMPLE SYSTEMS

Superconducting cavities have found successful application in a variety of accelerators spanning a wide range of accelerator requirements. High current storage rings for synchrotron light sources or for high luminosity, high energy physics with energies of a few GeV call for acceleration voltages of less than 10 MV, and carry high CW beam currents up to one amp. Figure 2 [7] shows the accelerating structure based on a 500 MHz, single cell cavity that evolved for the Cornell storage ring CESR/CHESS. The cavity was fabricated from pure sheet niobium. Four such systems provide the needed voltage of 7 MV and beam power of more than one MW. Similar systems are under construction to upgrade the beam current of the existing Taiwan Light Source (SRRC), and for the new Canadian Light Source (CLS). The accelerating gradient choice for all these cases is 7 MV/m or less.

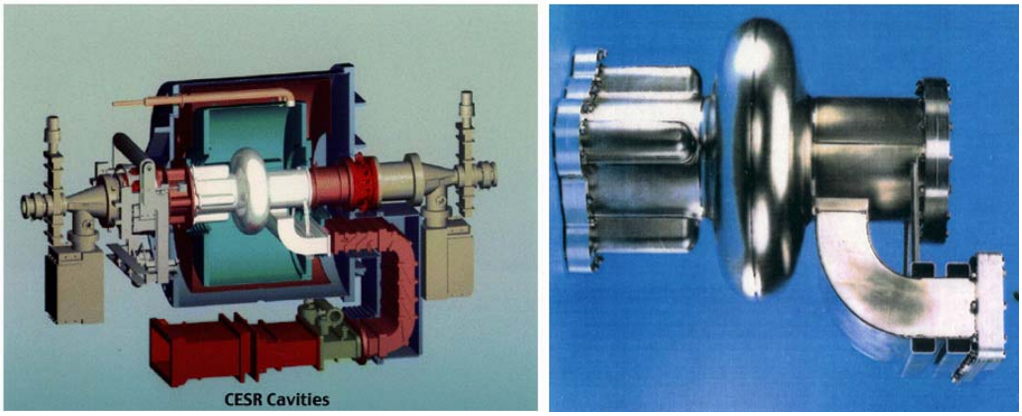


Fig. 2 (Left) 3D-CAD drawing of the CESR superconducting cavity cryomodule . (Right) 500 MHz Nb cavity.

Near the energy frontier, LEP-II at CERN called for an accelerating voltage for nearly 3 GV to upgrade the beam energy from 50 to 100 GeV per beam, with a beam current of a few mA. With a frequency choice of 350 MHz, dominated by higher order mode (HOM) power loss and beam stability considerations, a 4-cell structure emerged[8]. To build 300 such units there was considerable savings in material cost by fabricating the cavity out of copper and coating it with niobium by sputtering. The LEP-II cavities (Figure 3) operated successfully at an average gradient of 6 MV/m.

A one GeV CW linac forms the basis for CEBAF, a 5-pass recirculating accelerator providing 5-6 GeV CW beam for nuclear physics[9]. The total circulating beam current is a few mA. Developed at Cornell, the 5-cell, 1500 MHz cavities (Figure 4) are also fabricated from solid sheet niobium. CEBAF cavities operate at an average accelerating field of 6 MV/m.

Fig. 3 A 4-cell, 350 MHz Nb-Cu cavity for LEP-II

Fig. 4 A pair of 5-cell Nb cavities developed at Cornell for CEBAF.

All the above accelerators run CW at 100% duty factor. The first pulsed superconducting linac will be for the Spallation Neutron Source (SNS) at Oak Ridge. 6-cell niobium cavities at 804 MHz will accelerate a high intensity (≈10 mA) proton beam from 200 MeV to 1000 MeV. Figure 5 shows the medium β =0.64 cavity that resembles a β = 1 cavity that is squashed [10]. The duty factor for SNS is 6% and the RF pulse length is one ms. With recent improvement in cavity gradients the anticipated gradient is near 15 MV/m. Besides spallation neutron sources SNS technology could become suitable for high intensity proton linacs for various applications, such as transmutation of nuclear waste or generation of intense muon beams.

The dream machine for the future will be a 500 GeV energy frontier linac colliding electrons and positrons, upgradable to one TeV. As we will see, refrigerator power considerations drive the duty factor of operation to one percent. The average beam current is about 10 μA. A 9-cell niobium cavity design (Figure 6) has emerged from the TESLA collaboration[11]. With gradients improving steadily over the last decade, the choice of 25 MV/m will lead to 20 km of cavities for the 500 GeV machine. TESLA technology is likely to become the basis for the free electron lasers providing high brightness beams with wavelengths from the infra-red to ultraviolet and ultimately x-rays.

For the far future, acceleration of muons will also benefit from superconducting cavities[12]. A neutrino factory providing an intense neutrino beam from decaying muons may be the first step towards a muon collider that will penetrate the multi-TeV energy scale. At low energies (< a few GeV), where the muons have a large energy spread, the RF frequency has to be very low, e.g 200 MHz, leading to gigantic structures. Once again economics will favor thin film Nb-Cu cavities over

sheet Nb cavities. For comparison, a single cell Nb-Cu cavity at 200 MHz (Figure 1) dominates the size of superconducting cavities for the variety of accelerator applications discussed.

Fig. 5 b = 0.6, 6-cell cavity for SNS, frequency 804 MHz

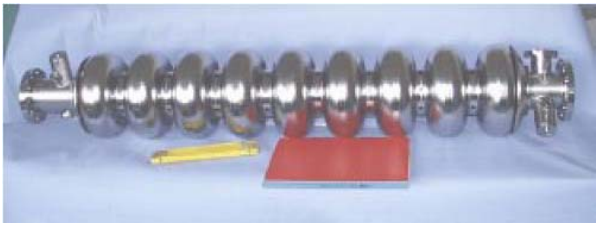


Fig. 6 1300 MHz 9-cell cavity for TESLA

3. BASICS OF ACCELERATING STRUCTURES

3.1 Accelerating field

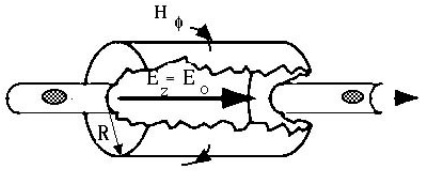


Fig. 7 Pill-box resonator

Only simple structures can be calculated analytically, such as a cylinder with no beam holes (Figure 7), referred to as the "pill-box cavity." For our purposes, the analytic calculations of a simple cylindrical cavity are convenient to define the important performance parameters of superconducting cavities. For a cylinder of length d and radius R using cylindrical co-ordinates (ρ, ϕ, z), the electric (E_z) and magnetic (H_ϕ) fields for the TM_{010} mode are given by:

$$E_z = E_0\, J_0\left(\frac{2.405\rho}{R}\right) e^{-i\omega t}, \quad H_\phi = -i\sqrt{\frac{\varepsilon_0}{\mu_0}}\, E_0\, J_1\left(\frac{2.405\rho}{R}\right) e^{-i\omega t} \qquad (1)$$

where all other field components are zero. J_0 and J_1 are Bessel functions. The angular resonant frequency is given by:

$$\omega_{010} = \frac{2.405\, c}{R}$$

which is independent of the cavity length.

First we determine the accelerating field, E_{acc}. Assume an electron travelling nearly at the speed of light (c). It enters the cavity at time t = 0 and leaves at a time t = d/c. To receive the maximum kick from the cavity, the time it takes the particle to traverse the cavity is to equal 1/2 an RF period (T_{RF}), i.e.

$$t = \frac{d}{c} = \frac{1}{2}\, T_{RF} = \frac{\pi}{\omega}$$

In this case, the electron always sees a field pointing in the same direction. The accelerating voltage (V_{acc}) for a cavity is

$$V_{acc} = \left| \int_{z=0}^{z=d} E_{el}\, dz \right|$$

For an electron accelerator with energy >10 MeV, it is sufficiently accurate to use v = c, so that t(z) = z/c. Thus

$$V_{acc} = \left| \int_{z=0}^{z=d} E_z\,(\rho = 0, z)\, e^{i\omega\, z/c} dz \right|$$

$$V_{acc} = E_0 \left| \int_{z=0}^{z=d} e^{i\omega z/c} dz \right| = d\; E_0\, \frac{\sin\left(\frac{\omega\, d}{2c}\right)}{\frac{\omega\, d}{2c}} = d\, E_0\, T$$

At 1.5 GHz RF frequency, d = c/λ = 10 cm, V_{acc} simplifies to

$$V_{acc} = (2/\pi)\; d\;\; E_0$$

The average accelerating electric field (E_{acc}) that the electron sees during transit is given by

$$E_{acc} = \frac{V_{acc}}{d} = \frac{2E_0}{\pi}$$

3.2 Peak Fields

To maximize the accelerating field, it is important to minimize the ratios of the peak fields to the accelerating field by selecting a suitable cavity geometry. For the TM_{010} accelerating mode in a pill-box cavity

$$E_{pk} = E_0, \quad H_{pk} = \sqrt{\frac{\varepsilon_0}{\mu_0}}\, J_1(1.841) E_0 = \frac{E_0}{647\,\Omega}$$

Thus we obtain the following ratios:

$$\frac{E_{pk}}{E_{acc}} = \frac{\pi}{2} = 1.6, \quad \frac{H_{pk}}{E_{acc}} = 2430\, \frac{\mathrm{A/m}}{\mathrm{MV/m}} = 30.5\, \frac{\mathrm{oersted}}{\mathrm{MV/m}}$$

Figure 8 shows the electric and magnetic field profiles for a real cavity shape. The peak field ratios for a realistic structure are much larger than for the pill-box. For example, for the TESLA cavity, E_{pk}/E_{acc} = 2.0 and H_{pk}/E_{acc} = 42 Oe per MV/m.

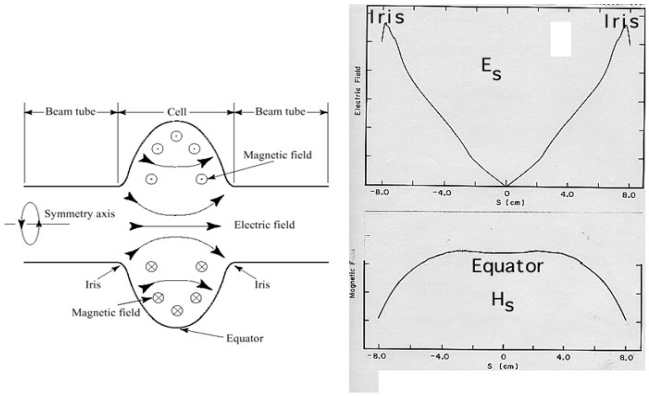


Fig. 8 Electric and magnetic field profiles for single cell cavity.

3.3 Power Losses and Q_0

In order to support the fields in the cavity, currents flow within a thin surface layer of the cavity walls. If the surface resistance is R_s, the power dissipated per unit area (P_a) due to Joule heating is

$$P_a = \tfrac{1}{2} R_s H^2$$

The two most salient characteristics of an accelerating cavity are its average accelerating field, Eacc, and the intrinsic quality factor Q0. We just discussed Eacc. The quality (Q0), is related to the power dissipation by the definition of Q0

$$Q = \omega \frac{\text{Energy Stored}}{\text{Power dissipated}} = \frac{\omega U}{P_c}$$

where U is the stored energy and P_c is the dissipated power. Since the energy stored in the electric field is equal to that stored in the magnetic field, the total energy in the cavity is given by the power dissipated as,

$$U = \tfrac{1}{2}\mu_0 \int_v |H|^2 dv \qquad P_c = \tfrac{1}{2} R_s \oint_s |H|^2 d\mathbf{s}$$

where the first integral is taken over the volume of the cavity, and the second over the surface. Thus

$$Q = \frac{\omega \mu_0 \int_v |H|^2 dv}{R_s \oint_s |H|^2 d\mathbf{s}} \qquad Q = \frac{G}{R_s} \qquad G = \frac{\omega \mu_0 \int_v |H|^2 dV}{\oint_s |H|^2 d\mathbf{s}}$$

Here G is called the "geometry factor" for the cavity shape. For a pill box in the TM010 mode, G = 257 □. Scaling arguments show that the ratio of the integrals □V |H|2 dV / □S |H|2 ds must scale linearly with a, or alternatively, inversely with the mode frequency. Therefore the geometry constant, G only depends on the cavity shape and not its size. A typical observed surface resistance for a well prepared superconducting Nb cavity is Rs = 20 n□. Thus we have a Q_0 value of

$$Q = \frac{G}{R_s} = 1.3 \times 10^{10}$$

For a typical cavity length of d = 10 cm and RF frequency of 1.5 GHz, the cavity radius is R = 7.65 cm. For an accelerating voltage of 1 MV, the following values result for the important features of a superconducting cavity:

$$E_{acc} = \frac{V_{acc}}{d} = 10 \text{ MV/m}$$

$$E_{pk} = E_0 = \frac{\pi}{2} E_{acc} = 15.7 \text{ MV/m}$$

$$H_{pk} = 2430 \frac{\text{A/m}}{\text{MV/m}} E_{acc} = 24.3 \text{ kA/ m} = 305 \text{ Oersted}$$

$$U = \frac{\pi \varepsilon_0 E_0^2}{2} J_1^2 (2.405)\, d\, R^2 = 0.54 \text{ J}$$

$$P_c = \frac{\omega U}{Q} = 0.4 \text{ W}$$

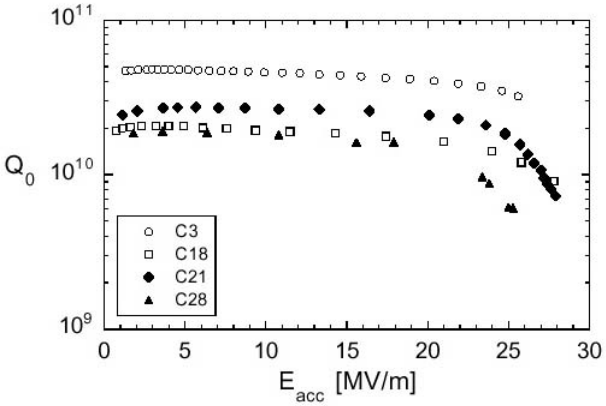


Figure. 9 Q_0 vs E curves for 9-cell, 1.3 GHz cavities. When x-rays are also present, the Q drop at high fields is attributed to field emission [11].

The performance of a superconducting cavity is evaluated by measuring the Q_0 as a function of the cavity field level. These curves bear tell-tale signs of the activities inside the cavity. Figure 9 shows Q_0 vs E curves for some high performance TESLA multi-cell structures.

3.4 Shunt Impedance

Two important quantities for cavity design are the Shunt Impedance R_a and the geometric shunt impedance R_a/Q. R_a is defined in analogy to Ohm's law from the losses in a cavity at a given accelerating voltage:

$$R_a = \frac{V_{acc}^2}{P_c}$$

Ideally we want the shunt impedance to be large for the accelerating mode so that the dissipated power is small. This is particularly important for copper cavities, where the wall power dissipation is a major issue and we wish to have as large an accelerating field as possible. For the TM_{010} mode we have:

$$R_a = \frac{4 \mu_0 d^2}{\pi^3 R_s \varepsilon_0 J_1^2 (2.405) R[R+d]} = 2.5 \times 10^{12} \Omega$$

From the definition of Q the ratio R_a/Q turns out to be:

$$\frac{R_a}{Q} = \frac{V_{acc}^2}{\omega U}$$

The geometric shunt impedance is independent of the surface resistance. The geometric shunt impedance is crucial for determining the beam-cavity interaction in the fundamental and HOMs. Since the ratio V_{acc}^2/U scales inversely with the cavity's linear dimensions, R_a/Q is independent of cavity

$$\frac{R_a}{Q} = 150\,\frac{d}{R} = 196\Omega$$

frequency, and only depends on the cavity geometry. For the pill-box TM_{010} mode we have:

Beam holes reduce the shunt impedance and enhance the peak surface fields relative to the pill-box case so that for a realistic cavity shape, R_a/Q drops by a factor of 2. A typical number for the R/Q of sc cavity cell is 100 Ω.

Finally we define the shunt impedance per unit length, r/Q, which gives the power dissipated per unit length at a given accelerating field Eacc. r/Q increases linearly with frequency.

$$r/Q = \frac{R/Q}{L} \qquad\qquad \frac{P}{L} = \frac{E_{acc}{}^2}{(r/Q)^* Q_0}$$

For real structures with contoured shapes, beam apertures and beam pipes, it is necessary to use field computation codes, such MAFIA and Microwave Studio. Figure 10 shows the electric and magnetic fields computed by Microwave Studio for the accelerating mode of a pillbox cavity with a beam hole, and for a round wall cavity. Such codes are also necessary for computing the fields in the higher order modes of a cavity that can have an adverse effect on beam quality or cause instabilities. Figure 11 shows the electric and magnetic fields of the first monopole HOM. Beam induced voltages are also proportional to the R/Q of HOMs.

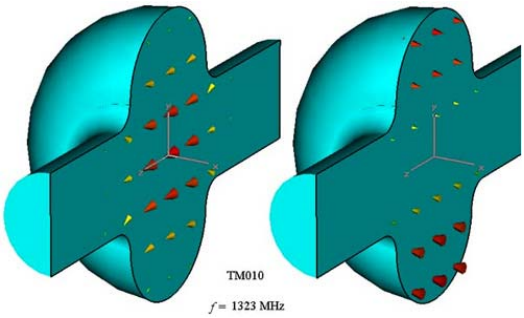


Fig.10 (Left) Electric and (Right) Magnetic fields for a round cavity with beam holes.

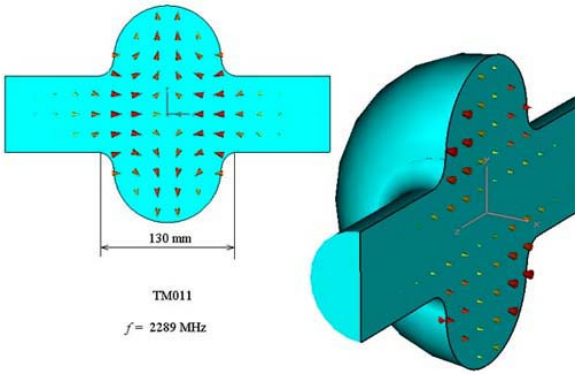


Fig.11 (Left) Electric and (Right) Magnetic fields for a the first monopole HOM.

3.5 Multicell Cavities

A multicell cavity is a structure with several cells coupled together. As with any set of coupled oscillators there are multiple modes of excitation for the full structure for every given mode of a single cell. The frequencies (f) of these modes can be cast via Equation 1 in terms of the single cell resonant freqency (f_0) and a cell-to cell coupling strength (k) via a coupled LC oscillator model (Figure 12). A measurement of the highest and lowest frequencies directly yields the value of the cell-to-cell coupling (Equation 2). Out of the N modes in the TM010 pass band of a N-cell structure, the accelerating mode is the one where the fields in the neighboring cells are π radians out of phase with each other so that each cell provides the same acceleration kick to a velocity-of-light particle that crosses each gap in one-half RF period (Figure 12). Having equal fields in each cell maximizes the overall accelerating voltage and minimizes the peak fields in each cell. A flat field profile is only achieved when the cells are properly tuned relative to each other. As Equations 1 & 2 show, a large number (N) of cells or a small the cell-to-cell coupling (k) reduces the spacing between the accelerating mode and its nearest neighbor, making tuning more difficult, and making the field profile more sensitive to any cell-to-cell frequency differences that arise from manufacturing tolerances. Therefore, in an application that demands high total voltage (e.g TESLA, ≈ TeV), a high number of cells is desirable, and it is important to increase the aperture and decrease the cell-to-cell distance to increase the cell-to-cell coupling coefficient.

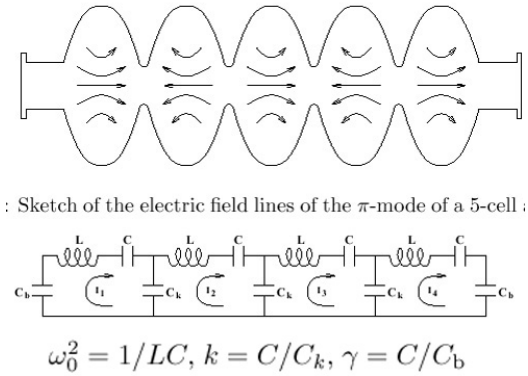


Fig. 12 Circuit model for a 5-cell cavity.

$$\left(\frac{f_m}{f_0}\right)^2 = 1 + 2k\left[1 - \cos\left(\frac{m\pi}{N}\right)\right] \tag{1}$$

If we measure $f^{(N)}$ and $f^{(1)}$, this becomes

$$k = \frac{\frac{1}{2}\left[\left(f^{(N)}\right)^2 - \left(f^{(1)}\right)^2\right]}{2\left(f^{(1)}\right)^2 - \left(f^{(N)}\right)^2\left[1 - \cos\left(\pi/N\right)\right]}. \tag{2}$$

3.6 Traveling Wave vs Standing Wave

Superconducting structures operate in the standing wave mode for which the power required to establish the fields is commensurate with wall losses. Hence the RF power that must be supplied is usually comparable to the beam power. Since the peak fields to accelerating field ratios are high in the standing wave, it is worth examining how much gain in peak fields can be expected by operating in the traveling wave mode. Figure 13 shows the reduction in peak surface magnetic field for traveling wave mode operation the 2π/3, π/4 and π/3 modes. Although the 25% magnetic fild reduction seems attractive at first glance, it comes with the price of smaller gap cells and thus more cells per meter. Consequently the cost of such structures would increase substantially over the standing wave structure, resulting in little overall gain in the total capital cost of a traveling wave superconducting accelerator. Moreover the traveling wave power required to establish the fields would need to be dumped at the end of the structure or recirculated. The first option is wasteful, defeating the main advantage of superconducting cavities, while the second option would increase structure complexity and cost.

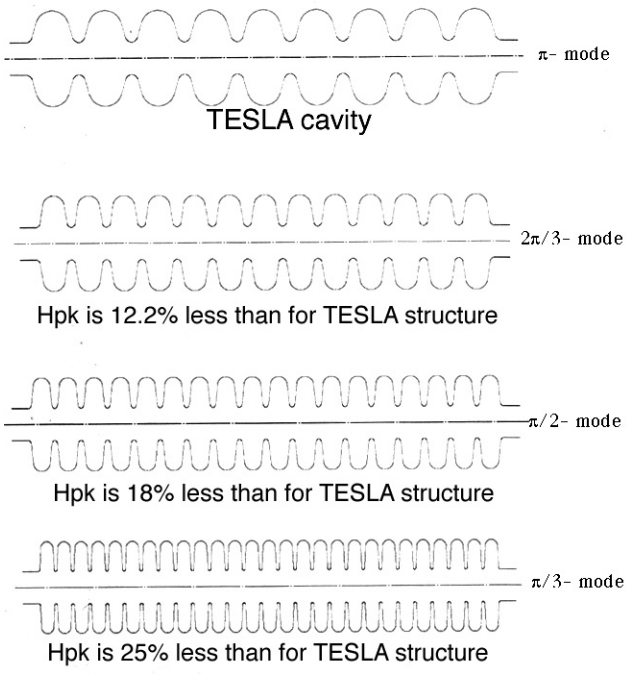


Fig. 13 Structure geometry for travelling wave and accompanying decrease of Hpk.

4. RF SUPERCONDUCTIVITY BASICS: SURFACE RESISTANCE

As explained in the reference text [5], the remarkable properties of superconductivity are attributed to the condensation of charge carriers into Cooper pairs, which then move without friction; hence the zero resistance. At T = 0 K, all charge carriers condense. At higher temperatures, pairs break up. The fraction of unpaired carriers increases exponentially with temperature, as $e^{-\Delta/kT}$, until none of the carriers are paired above T_c – the normal conducting state. Here 2Δ is the energy gap of the superconductor, i.e., the energy needed to break up the pairs. In this simplified picture, known as the London two-fluid model, when a DC field is turned on, the pairs carry all the current, shielding the applied field from the normal electrons. Electrical resistance vanishes since Cooper pairs move without friction.

In the case of RF currents, however, dissipation does occur for all T > 0 K, albeit very small compared to the normal conducting state. While the Cooper pairs move without friction, they do have inertial mass. For high frequency currents to flow, forces must be applied to bring about alternating directions of flow. Hence an AC electric field will be present in the skin layer, and it will continually accelerate and decelerate the normal carriers, leading to dissipation, proportional to the square of the RF frequency. A simplified form of the temperature dependence of Nb for $T_c/T > 2$, and for frequencies much smaller than $2\Delta/h \approx 10^{12}$ Hz is:

$$R_s = A\,(1/T)\,f^2 exp\,(-\Delta(T)/kT) + R_0$$

Here A is a constant that depends on material parameters, discussed in another lecture at the Erice School[1]. Based on the very successful BCS theory, expressions for the superconducting surface impedance have been worked out in terms of material parameters. The operating temperature of a superconducting cavity is usually chosen so that the temperature dependent part of the surface resistance is reduced to an economically tolerable value. R_0, referred to as the residual residual resistance, is influenced by several factors such as the ambient DC magnetic field environment of the cavity or the overall hydrogen gas content of the Nb material.

5. EXAMPLE 1, POWER CONSIDERATIONS FOR STORAGE RINGS

For a few GeV electron storage ring as in a light source or a B-factory, consider a 500 MHz single cell cavity of the geometry in Figure 14 (left), with half wavelength gap about 0.3m, R/Q = 89 Ohm, and G = 270 Ohm. Start with a modest CW voltage of one MV from the cavity, at a gradient about 3 MV/m. For a copper cavity, the Q_0 is 45,000 from the surface resistance of copper (Equation 3), the shunt impedance Ra = 4 MΩ, giving a dissipated power 250 kW. This would result in overheating of the copper cell. Water-cooled copper cavities at this frequency can safely dissipate about 40 kW. To bring dissipation to tolerable level it is essential to either lower the gradient or to raise the Ra/Q. Figure 14 (right) shows a normal conducting cell shape with Ra/Q = 265 Ω, for which the dissipation drops to 80 kW/cell. For a typical klystron efficiency of 0.5, the AC wall power is 160 kW per cell. But with the small beam hole, HOMs cannot propagate down the beam pipe and many protrusions have to be added to the cell to remove HOMs which make harmful interactions with the beam (Figure 15). While the small beam hole helps to raise the R/Q of the fundamental mode as desired, it also raises the R/Q of the HOMs, increasing the danger of beam instabilities. Figure 16 compares the R/Q of HOMs for the large and small beam hole cases.

$$R_n = \sqrt{\pi f \mu_0 \rho} = 6m\Omega \qquad (3)$$

For a superconducting Nb cavity of the shape of Figure 14, the BCS surface resistance at 4.2 K leads to a Q of $2x10^9$, R = 0.5 teraΩ, and dissipated power in the wall of about 2 watt. Taking into account refrigerator efficiency (at 4.2 K) of 1/350, the AC wall plug power is 0.7 kW. In this case the

static heat leak and cryogen transfer lines will dominate the dynamic heat load. Even if these contributions total 45 W, the AC wall power due to the refrigerator will be remain a factor of 10 smaller than for the copper cavity. The power economy of superconductivity opens the possibility of raising the gradient to say 10 MV/m, and reducing the number of cells by a factor of 3. Together with the small HOM impedance of the large beam hole the reduced number of cells improves beam quality, avoids beam instabilities and allows higher currents.

To complete the picture, electrons in a 5 GeV storage ring have a synchrotron radiation energy loss of about one MV per turn. For one amp beam current, the beam power that must be supplied through the cavities is one MW. The corresponding AC wall-plug power is 2 MW due to klystron efficiency. In CESR, the over-voltage factor to overcome quantum fluctuations from synchrotron radiation is about 7, so that the total voltage requirement is not 1 MV, but 7 MV. This can be met by four single cell units (Figure 1), each providing about 2 MV at a gradient of 7 MV/m. Each unit has a dynamic heat load of 50 watt and a static heat plus transfer line heat load of 50 W. The total cryogenic heat load is 400 watt, corresponding to an AC power demand of 140 kW. The copper cavity solution (as for PEP-II at SLAC) would call for 9 units (as in Figure 15), each dissipating 80 kW. Taking into account the klystron efficiency (0.5), the structure associated AC power is 1.44 MW, the same factor of 10 higher than the superconducting case, as discussed. But on adding the AC power associated with the beam power (1 MW) to both solutions, one obtains AC power equal to 3.44 MW for the normal-conducting case versus 2.14 MW for the superconducting case. The large beam power reduces the superconducting advantage from a factor of 10 to just 60%. The impedance reduction remains the dominant advantage.

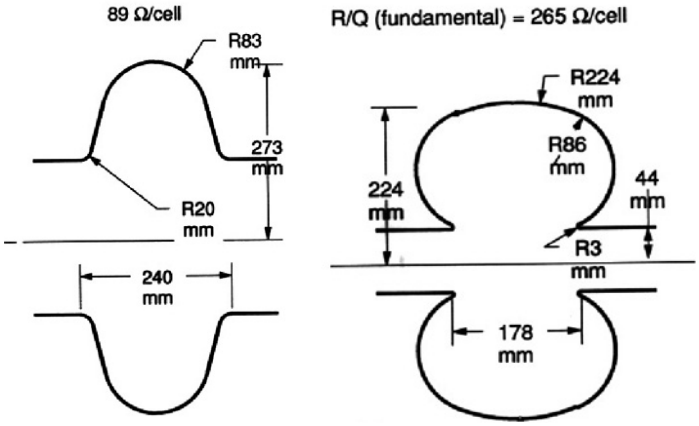


Fig. 14 A comparison of typical shapes used for (Left) superconducting and (Right) normal conducting cavities.

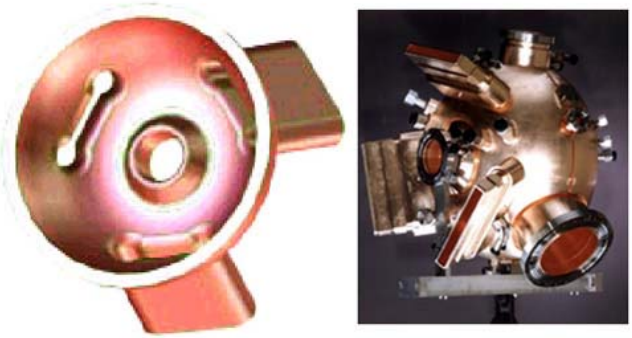


Fig. 15 (Left) Inside view of a copper cell showing reentrant iris and apertures for HOM couplers. (Right) Full view of a copper cell as for PEP-II (SLAC) with HOM couplers[13].

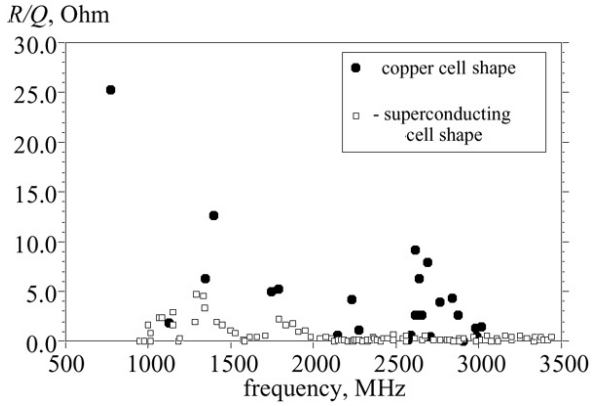


Fig. 16 A comparison of the shunt impedances of a copper cell shape with a superconducting cavity cell shape[7].

6. RF SUPERCONDUCTIVITY BASICS: HIGH FIELD BEHAVIOR

As mentioned, the accelerating field, E_{acc}, is proportional to the peak electric (E_{pk}) as well as the magnetic field (H_{pk}) on the surface of the cavity. Therefore important fundamental aspects of superconducting cavities are the maximum surface fields that can be tolerated without increasing the microwave surface resistance substantially, or without causing a catastrophic breakdown of superconductivity. The ultimate limit to the accelerating field is the RF critical magnetic field, above which the superconducting phase can no longer exist. The RF critical field is related to the thermodynamic critical field. In the process of a phase transition to the normal conducting state, a phase boundary must be nucleated. Because of the rapidly changing RF fields (ns time scale), it is possible for the Meissner state to persist above the thermodynamic critical field (H_c) for Type I superconductors, and above the lower critical field (H_{c1}) for Type II superconductors. Such a metastable situation can be expected up to a superheating critical field, $H_{sh.} > H_c$ (Type I) > H_{c1} (for type II). It is important to note that the RF critical field does *not* depend on H_{c2}. Therefore high field magnet materials, such as Nb-Ti, do not offer correspondingly higher operating fields for superconducting cavities. Indeed for RF superconductivity, it is essential to always operate in the Meissner state. For the most popular superconductor, niobium, H_{sh} is about 0.23 T. These surface fields translate to a maximum accelerating field of 55 MV/m for a typical niobium. The exact values depend on the detailed structure geometry.

Typical cavity performance is significantly below the theoretically expected surface field limit. One important phenomenon that limits the achievable RF magnetic field is “thermal breakdown” of superconductivity, originating at sub millimeter-size regions of high RF loss, called defects. When the temperature outside the defect exceeds the superconducting transition temperature, T_c, the losses increase substantially, as large regions become normal conducting (see Figure 16). Measures to overcome thermal breakdown are to improve the thermal conductivity of niobium by purification or to use thin films of niobium on a copper substrate cavity.

The Q_0 vs E_{acc} curve (Figure 9) only gives information on the average behavior of the RF surface. To resolve the local distribution of RF losses and identify various mechanisms, temperature mapping proves to be a powerful diagnostic technique. A chain of rotating carbon thermometers, or an array of fixed thermometers, samples the temperature of the outer wall of the cavity. The temperature map of Figure 16 shows a hot spot that leads to thermal breakdown, and the SEM micrograph reveals a 50 μm culprit copper particle.

In contrast to the magnetic field limit, we know of no theoretical limit to the tolerable surface electric field. Fields up to 220MV/m have been imposed on a superconducting niobium cavity without any catastrophic effects [14]. However, at high electric fields, an important limitation to the

performance of superconducting cavities arises from the emission of electrons from local spots in the high electric field regions of the cavity. This is a problem endemic to all high voltage devices. Poweris absorbed by the electrons and deposited as heat upon impact with the cavity walls. Copius x-rays are emitted due to bremstrahlung. At high fields the exponential drop in Q_0 with field suggests that field emission is the dominant limiting mechanism, provided x-rays are also observed (Figure 9).

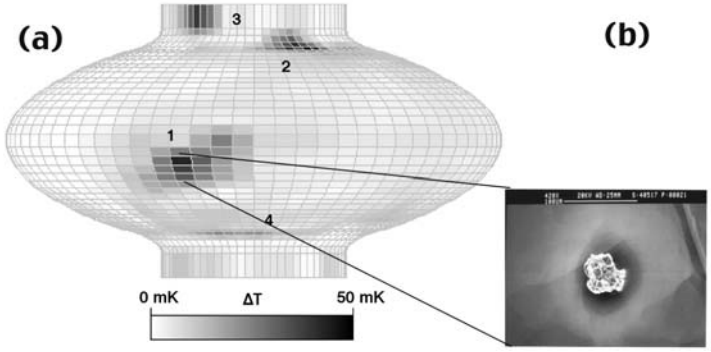


Fig. 16 (Left) Temperature map at 400 Oe of a 1.5 GHz, single cell cavity showing heating at a defect site, labelled #1 and field emission sites labelled #2, 3, and 4. (b) SEM micrograph of the RF surface taken at site #1 showing a copper particle [5].

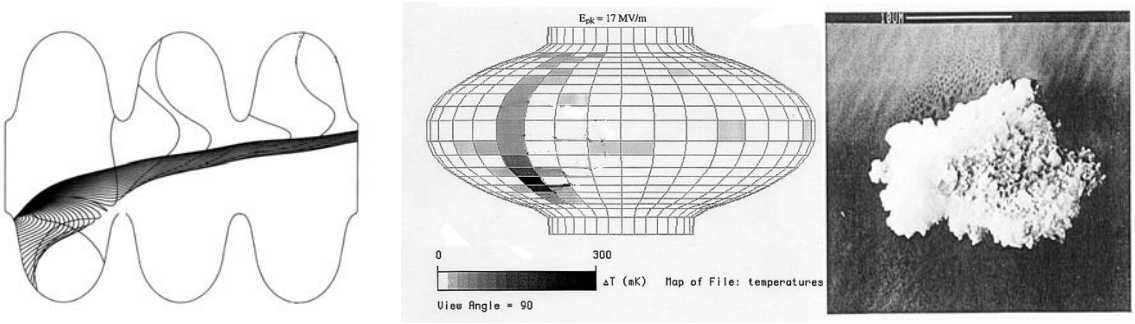


Fig. 17 Calculated electron trajectories in a 3-cell 1.5 GHz cavity operating at Epk = 50 MV/m. Here the emitter is located in an end cell, just below the iris, where the surface electric field is 44 MV/m. Note that a significant number of electrons emitted during the early part of the RF cycle bend back and strike the wall near the emitter. (Middle) Temperature map from the heating of impacting electrons. (Right) A micron size foreign particle found at an emission site.

Figure 17 shows electron trajectories in one RF period and the heating due to their impact with the cavity wall. A typical emiiter is a microparticle contaminant. When emission grows intense at high electric fields it can even initiate thermal breakdown. In many cases, intense field emission eventually leads to momentary voltage breakdown of the vacuum in the cavity. This has mostly a beneficial effect for superconducting cavities, known as conditioning. After a voltage breakdown event, it is usually possible to raise the electric field until field emission grows intense once again at another spot on the cavity surface. We have learned much about the nature of field emission sites and made progress in techniques to avoid them as well as to destroy them by conditioning with high voltage breakdown[5].

In the early stages of the development of superconducting cavities, a major performance limitation was “multipacting”. This is a resonant process in which an electron avalanche builds up within a small region of the cavity surface due to a confluence of several circumstances. With the invention of the round wall cavity shape, multipacting is no longer a significant problem for velocity-

of-light structures. The essential idea to avoid multipacting is to gradually curve the outer wall of the cavity – hence the round wall profile.

If the shape is not rounded, one-surface multipacting will severely limit the cavity performance (Fig. 18). An electron emitted from one region of the surface (usually the outer cylindrical wall) travels in a cyclotron orbit in the RF magnetic field, and returns to near its point of origin. Upon impact it generates a secondary electron which mimics the trajectory of the primary. An exponential build up occurs if the round-trip travel time of each electron is an integer multiple of an RF period, i.e., the electron returns in the same phase of the RF period when it is generated. For the build up to persist, the secondary emission coefficient must be greater than one. This is true for a niobium surface when the electron energy is between 50 eV and 1000 eV [5]. During their excursion into the RF fields, the electrons must gain enough energy from the electric field to generate secondaries on impact. When these conditions are met, an electron avalanche occurs to absorb the RF power, making it impossible to raise the fields by increasing the incident RF power. The electrons impact the cavity walls, which leads to a large temperature rise, thermal breakdown, and in some cases a momentary gas discharge. When the cell shape is rounded (Fig. 18) the electrons are forced to the equator region where the electric field is too low for the electrons to gain sufficient energy to regenerate. The avalanche is arrested.

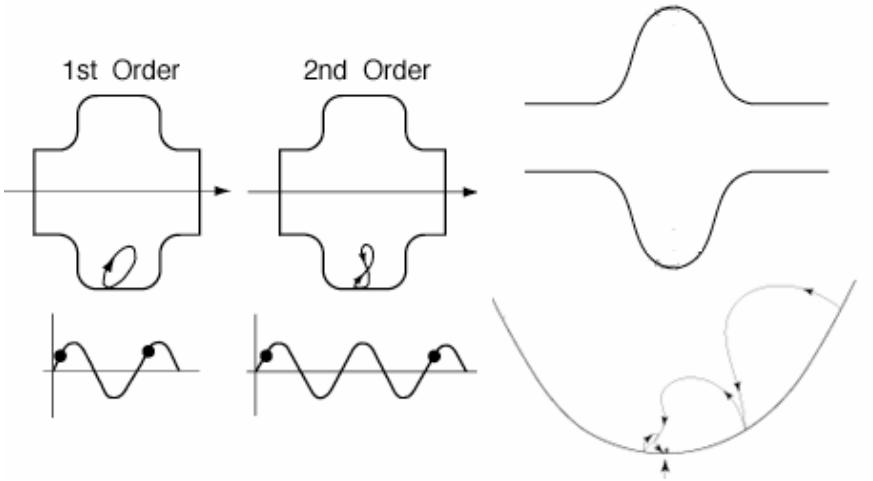


Fig. 18 (Left, upper) In a cavity with a nearly pill-box like shape, electrons can multiply in the region shown. The electron returns in an integer multiple of RF periods. The number of RF periods is called the order of multipacting. (Left, lower) The start and end phases of the electron that returns in one and two RF periods. A travel time of one RF period results in first order multipacting. (Right) With a rounded cavity shape, the electrons drift to the zero field region at the equator. Here the electric field is so low that the secondary cannot gain enough energy to regenerate. Multipacting stops.

7. CHOICE OF CAVITY SHAPE

There are many factors which influence cavity shape. Multipacting is a key factor that governs the overall rounded contour of the cavity profile. Beam dynamics considerations control the size of the aperture. To lower the peak electric field it is necessary to round the iris region with circular or elliptical arcs. Peak magnetic field considerations influence the shape of the cavity in the large diamter (equator) region where the magnetic field is strongest. Elliptical arc segments increase the strength of the cavity against atmospheric load (see mechanical considerations below) and also provide a slope for efficient rinsing of liquids during surface etching and cleaning.

Mechanical considerations also influence the cavity shape. To prevent the cavity from deforming excessively or even collapsing under atmospheric load it is important to avoid flat regions in the cavity profile (Figure 19). Elliptical segments tend to be strongest. For pulsed operation at high gradients the Lorentz force may be strong enough to cause cell deformation demanding larger thickness walls or stiffening rings near the iris (see Figure 19). Typical detuning coefficients are in the range of a few $Hz/(MV/m)^2$. Additional stiffening of multicell structures may be necessary to raise the frequency of mechanical vibration modes. As an extreme example figure shows exaggerated structure deformations for the lowest frequency vibration modes for a 200 MHz 4-cell cavity.

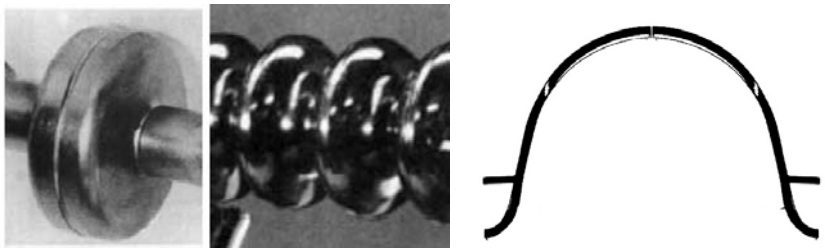


Fig. 19 (Left) Flat versus (Middle) elliptical cavity profile. (Righ) Stiffening rings for the TESLA cavity.

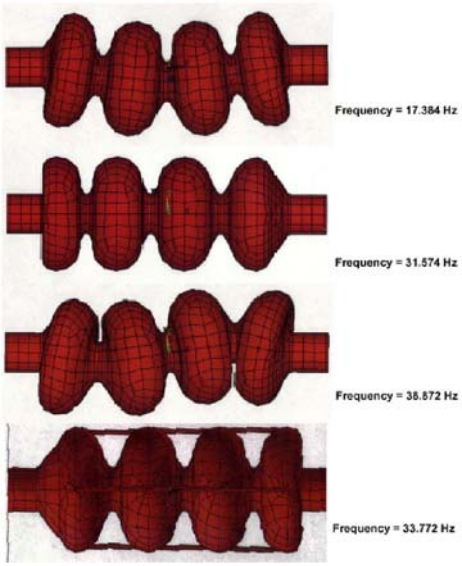


Fig. 20 Mechanical resonant modes of a 4-cell, 200 MHz cavity with 8 mm wall thickness. The low resonant frequencies spell trouble in the form of microphonics. Reducing the number of cells or stiffening is essential.

8. SUPERCONDUCTOR CHOICE

For a material to be useful for accelerators, the primary requirements are a high transition temperature, T_c, and a high RF critical magnetic field, H_{sh}. Among the elemental superconductors, niobium has the highest T_c and the highest RF critical field. Accordingly, it is a most attractive choice for accelerator cavities. Successful cavities have been made from sheet Nb, or by sputtering Nb onto a copper cavity. The realm of superconducting compounds has been much less explored because of technical complexities that govern compound formation. In looking at compound candidates, it is important to select a material for which the desired compound phase is stable over a broad composition range so that formation of the compound is more tolerant to variations in experimental conditions, making it possible to achieve the desired single phase over a large surface area. Nb_3Sn is a promising material. The T_c is 18 K and the RF critical field is 0.4T, twice as high as for Nb. On fundamental grounds the higher field opens up the possibility of accelerating gradients higher than allowed for niobium cavities.

However, the performance for Nb_3Sn cavities to date is far lower than for niobium cavities. The new HTS are even further from the microwave performance level desired for application to accelerators.

A strong motivation for using thin films of Nb on copper cavities is to provide increased stability against thermal breakdown of superconductivity. The thermal conductivity of copper at 4.2 K is between 300 to 2000 W/m-K, depending on the purity and annealing conditions, as compared to the thermal conductivity of 300 RRR niobium, which is 75 W/m-K at 4.2 K. The cost saving of niobium material is another potential advantage, significant for large-size, e.g. 350 MHz cavities, as for LEP-II, or for future projects, which aspire to make 200 MHz cavities, where each cell is more than one meter in diameter.

9. EXAMPLE LINEAR COLLIDERS, GRADIENT AND POWER ISSUES

Consider a one TeV cm energy linear collider, the ambition of many international accelerator collaborations. Based on the latest progress in TESLA gradients (Figure 21) we can confidently select Eacc = 30 MV/m[15], resulting in a 33 km active length linac. TESLA cryomodules have achieved a filling factor of 0.75, so that the real estate length of superconducting linac would be 45 km. Because of a higher gradient choice (50 MV/m, loaded gradient), a normal conducting linear collider would be shorter, about 25 km.) At a Q_0 of 10^{10} we determine a dynamic heat load of 90 watt/m to yield a preposterous, total dynamic heat load of 3 MW at 2 K. The capital cost for such a titanic refrigerator would exceed 10 billion dollars, and the AC power to run it would exceed 3 GW, comparable to a nuclear power plant. Hence a superconducting TeV energy linac must run in the pulsed mode with a duty factor of about one per cent, cutting the dynamic load to just 1 W/m, and the total dynamic cryogenic load to 33 kW.

Or we must look toward improvements that could lead to Q values of 10^{11}. Such Q values have in fact been reached [16] in single cells at 1.6 K (see Figure 22), but remain to be demonstrated in full scale structures inside accelerator cryomodules.

Continuing with the one per cent duty factor scenario, other important heat loads are the static heat and a fraction of the HOM power deposited at low temperature. Assuming another one watt/m for these contributions, and taking into account the entire accelerator length of 44 km, the grand total cryogenic heat load is 33kW + 44 kW = 77 kW. At 2 K the refrigerator efficiency is 1/750, leading to a AC wall plug power load of 58 MW.

As in the case of the storage ring example above, we now include the beam power. For a superonducting linear collider the typical value for the total beam power is 30 MW[11]. At a klystron efficiecncy of 70% the RF installation then calls for an AC power of 43 MW. The grand total AC power becomes 100 MW. An important figure of merit, the efficiency of AC power to beam power conversion, is an attractive 30%. Typical efficiencies of normal-conducting linac options range around 10% due to the large RF power needed to fill copper structures to high gradients [17]. Beam powers are also kept low to manage HOMs. Hence a superconducting linac opens the route to high luminosity via high beam powers, rather than squeezing the spot size at the collision point to nano-meter dimensions as for normal conducting linacs.

The peak beam power for the superconducting linac is 30 MW divided by the duty factor, which comes out to 3 GW, or roughly 100 kW/m. Here is another advantage of the superconducting option. Because of the long filling time allowed by the low cavity wall losses, cavities can be filled slowly (ms) reducing the peak RF power requirement compared to a normal conducting linear collider. Here filling times must be of the order of ns, and peak power in the order of hundred MW per meter to achieved high gradients (70 MV/m). The peak total RF power falls in the multi-Terawatt regime.

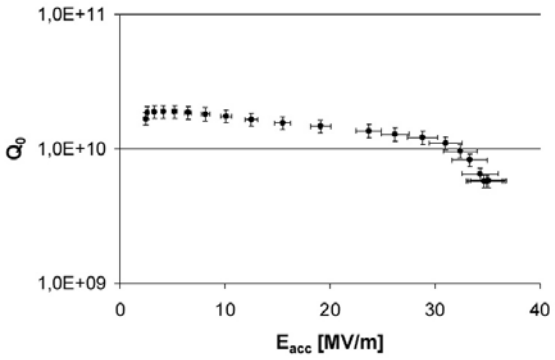


Fig. 21 A TESLA cavity which reached Eacc = 35 MV/m after electropolishing[15].

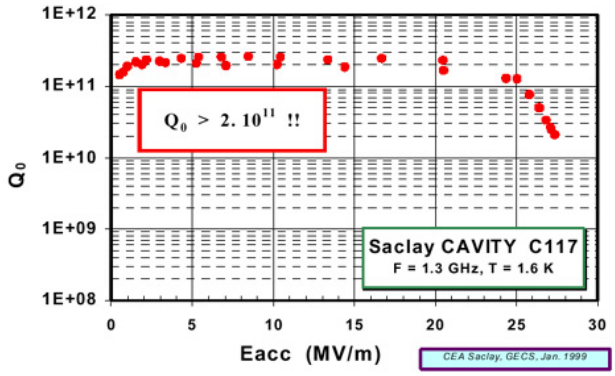


Fig. 22 [16] A single cell cavity which reached Eacc = 25 MV/m at a Q_0 greater than 10^{11}.

10. CONCLUSIONS

When designing a superconducting cavity, many trade-offs must be made between accelerator requirements and cavity performance issues. Reviewing the choice of RF frequency, a high frequency is better because structures are smaller, easier to handle and the per meter structure cost is lower. Small surface area means fewer defects that can cause thermal breakdown and fewer emitters that can cause field emission. Higher r/Q (ohm per meter) means reduced dynamic heat load to liquid helium for the same operating gradient and the same length of accelerator. Smaller volume, high frequency structures contains less stored energy at a given field and lead to reduced AC wall plug power. Finally, a topic we have not discussed here, higher frequency structures have less capture of field emitted electrons in beam pipe, reducing the possibility of polluting the beam. Power economy considerations make it important to keep superconducting cavities clean enough that dark currents are very low.

A low frequency choice is better because the superconducting state (BCS) surface resistance (α f^2) is lower resulting in higher Q at fixed operating temperature. The r/Q (per meter) of HOMs are lower, which is better for beam stability and HOM power loss. Since each cell is longer there are fewer cells per meter. For the same number of cells per structure the end assoicated costs are less.

Reviewing considerations that govern aperture choice, a large aperture is better because of larger beam clearance and greater cell-to-cell energy coupling which reduces field non-uniformity caused by errors in cell shape, allowing a larger number of cells per cavity unit. A large aperture also improves coupling of power from input coupler to cells in the presence of beam loading. For higher mode effects, a large aperture reduces wakefields both long- and short-range resulting in better beam quality and allowing higher beam current. On the other hand, a small aperture is better because it reduces

Epk/Eacc and Hpk/Eac. The resulting higher R/Q in accelerating mode also means lower RF heat load into liquid helium.

Turning to number of cells, a small number of cells per structure unit makes it easier to obtain a flat field profile and leads to less power for each window/coupler to deliver. Again, a smaller cavity area means fewer defects that can lead to thermal breakdown and fewer emitters that cause field emission. There is less chance of trapped HOMs and it is easier to remove HOM from couplers on the beam pipe which intercept fields in the end cells only. On the other hand, a large number of cells is better because it minimizes system costs by reducing wasted space between cells and minimizes the effects of fringing field (which lower R/Q).

To recap the salient results of the storage ring and linear collider examples, superconducting cavities excel in applications requiring continuous wave (CW) or long-pulse accelerating fields above a few million volts per meter (MV/m. Since the ohmic power loss in the walls of a cavity increases as the square of the accelerating voltage, copper cavities become uneconomical when the demand for high CW voltage grows with particle energy. A similar situation prevails in applications that demand long RF pulse length, or high RF duty factor. Here superconductivity brings immense benefits. The surface resistance of a superconducting cavity is many orders of magnitude less than that of copper. After accounting for the refrigerator power needed to provide the liquid helium operating temperature, a large net gain factor remains to provide many advantages.

Copper cavities are limited to gradients well below a few MV/m in CW and long-pulse operation because the capital cost of the RF power and the AC-power operating cost become prohibitive. For example, several MW/m of RF power are required to operate a copper cavity at 5 MV/m. There are also practical limits to dissipating high power in the walls of a copper cavity. The surface temperature becomes excessive causing vacuum degradation, stresses and metal fatigue due to thermal expansion. On the other hand, copper cavities offer higher accelerating fields ($\approx$ 50 MV/m) for short pulse (μs) and low duty factor ($<$ 0.1%) applications. For such applications it is important to provide abundant peak RF power (e.g. 100 MW/m in order to reach the high fields.

There is another important advantage that SRF cavities bring to accelerators. The presence of accelerating structures has a disruptive effect on the beam, limiting the quality of the beam in aspects such as energy spread, beam halo, or even the maximum current. Because of their capability to provide higher voltage, SRF systems can be shorter, and thereby impose less disruption. Due to their high ohmic losses, the geometry of copper cavities must be optimized to provide a high electric field on axis for a given wall dissipation. This requirement tends to push the beam aperture to small values, which disrupts beam quality. By virtue of low wall losses, it is affordable to design an SRF cavity to have a large beam hole, reduce beam disruption and provide high quality beams for physics research.

There has been much progress in understanding the gradient and Q_0 limitations in superconducting cavities. Through better understanding, new techniques have been developed to overcome the limitations. Producing high gradients and high Q_0 with Nb cavities demands excellent control of material properties and surface cleanliness. As a result of the improved understanding and the invention of new treatments, there has been much progress in reducing the spread in gradients that arises from the random occurrence of defects and emitters. Prescreening the starting material by eddy current scanning reduces the number of defects that can cause thermal breakdown. High RRR, high thermal conductivity Nb reduces the impact of any remaining defects. It will be important to aim for higher RRR in large area cavities, where there is a high chance of defects and contamination. High pressure rinsing greatly reduces the number of field emitters. High pulsed power processing destroys accidental field emitter contaminants. This technique will continue to be necessary in order to realize - in accelerators - the high intrinsic gradient potential of SRF cavities. There is now excellent prognosis for reaching 25 – 35 MV/m for future colliders. The road to 40 MV/m is opening up. Although most successful cavities are based on Nb, some exploratory work has been carried out on other materials.

ACKNOWLEDGEMENTS

The author is deeply indebted to Valery Shemelin for providing results presented here from Microwave Studio calculations and for peak fields in traveling wave structures.

REFERENCES

[1] P. Schmueser, these proceedings

[2] H. Safa, these proceedings.

[3] D. Proch, these proceedings.

[4] R. Parodi, these proceedings.

[5] Padamsee H., Knobloch J. and Hays T., *RF Superconductivity for Accelerators* (John Wiley and Sons, New York, 1998).

[6] Proch D., Rep. Prog. Phys. **61**, IOP Pub. (1998) p. 1.

[7] Belomestnykh S., *Proceeding of the 1999 Particle Accelerator Conference* ed. by A. Luccio et al (1999) p. 272.

[8] Brown P. et al., *Proceedings of the 9th Workshop on RF Superconductivity* (1999) ed. By B. Rusnak, paper MOA001.

[9] Reece C. et al., *Proceedings of the 9th Workshop on RF Superconductivity* (1999) ed. By B. Rusnak, paper MOA004.

[10] Ciovati et al, *Proceedings of the 10th Workshop on RF Superconductivity* (2001) ed. By S. Noguchi, paper PT016.

[11] Trines D., *Proceedings of the 9th Workshop on RF Superconductivity* (1999) ed. By B. Rusnak, paper FRA 003.

[12] Padamsee H, *Proceedings of the 10th Workshop on RF Superconductivity* (2001) ed. By S. Noguchi, paper FRA 003.

[13] Rimmer R et al, *Proceeding of the 1997 Particle Accelerator Conference* ed. by M. Comyn et al (1997) p. 3004.

[14] Delayen J. Shepard K. W.. Appl. Phys. Lett., **57**, p. 514 (1990).Tajima T., *Proceeding of the 1999 Particle Accelerator Conference* ed. A. Luccio et al (1999) p. 440.

[15] Lijle L, DESY, private communication.

[16] Safa, H, *Proceedings of the 10th Workshop on RF Superconductivity* (2001) ed. S. Noguchi, paper TL003.

[17] Loew, International Linear Collider Technical Report (1995).

COUPLERS AND HOM DAMPERS

R. F. Parodi
INFN-Genoa

Abstract
A superconducting cavity ready for the installation in a particle accelerator need to be equipped with a quite complex suite of ancillary devices often as complex and challenging as the cavity itself. Fundamental power coupler(s) (used to feed power to the accelerating beam) and HOM couplers (used to extract the idle RF power left from the circulating beam on the cavity modes different from the accelerating one) are examples of that kind of devices. Tuners (used to keep the frequency and phase of the cavity locked to the master oscillator used as a pace maker for the beam bunches) are a further category of ancillaries. All types, if not properly designed and built, will probably degrade the cavity performance beyond any possibility of easy recover. The lecture will introduce the different implementation of couplers and tuners and the related design constraints.

1. INTRODUCTION

A successful S/C cavity meeting the accelerator's specifications is only the glamorous part of the painful task of building a S/C RF accelerating structure.Once defined the cavity electromagnetic design, checked for the best compromise among accelerating field, peak surface fields (electric and magnetic) versus accelerating field ratios, power dissipation at low temperature, mechanical stiffness against atmospheric pressure and radiation pressure (Lorentz force) and so on, you have accomplished about 30% of the task.

Before building an accelerator ready S/C accelerating structure you have to design the cryogenics of your module, deciding, depending on the frequency choice, to use normal or sub-atmospheric normal liquid helium, or to go below the helium lambda point (2.19 K) and go superfluid., either sub-atmospheric @ 50 mbar or pressurized @ 1000 mbar. This part of the design is quite cumbersome, painful and account for roughly another 30% of the design effort. No magic rule exists; You have to start with a design and go dawn drawing all the lines of the Blue Print till You find a really serious road block. At this point, you have to change the design to find a new solution overcoming the roadblock; at some point, a new bock will appear; and the design of the cryogenic system will proceed in an iterative way until to the end, hopefully.

In general the cryogenic design account for roughly another 30% of the total design.

The third broad category of items in the successful design of an accelerating module account for all the ancillaries needed to feed energy to the accelerated beam via RF fields in the cavity. The system must also keep the synchronism between the RF waves and the accelerated bunches. Last we need to avoid any dangerous interaction between the different bunches (through the cavity RF fields) in the accelerator leading to a dramatic loss of beam intensity.

This last part of the design accounts for about the remaining 40% of the design task and accounts for the Main Couplers, the Cavity Tuners and the High Order Mode (HOM) dampers.

Since any flaws of any components connected to the superconducting cavities can, and most likely will, degrade the cavity performance beyond any possibility of easy recover, the attention to the details of the design, fabrication and assembly of the couplers, dampers and tuners is at least as

important as that for the cavities themselves. In the past, cavity performance was often degraded between tests in a vertical dewar and the running operation in cryomodules. Components that can contribute to this degradation were improperly designed fundamental power coupler and HOM dampers. Failure of the tuners is one of the most likely failures putting a superconducting cavity out of operation.The design a “good coupler” or a “good HOM damper” (a working one) need a lot of skill in different fields as Cryogenics, Mechanical engineering, material science, vacuum technology and RF electromagnetic field modeling.

Beginning in the early ’80, work on fundamental power couplers and HOM dampers has been at the forefront of technology. Several papers have addressed the issues related to these components before [0],[1], [2], [3], [4], [5], [6], [7],[8].

Since the technical solution for the design and the construction of the couplers have strongly progressed in the past years; to avoid compete with better teachers (my teachers) and compelled by the page limit, I prefer to address in my lecture the technological issues of the coupler design and construction. I suggest the quoted references, and mainly the Lengeler’s lecture [0], as the main reference for the people interested in a deeper insight of the theoretical background lying below the coupler’s subject.

1.1 Coupling power to the beam

The designs of the main coupler are largely imposed by specific requirements associated with the characteristics of the accelerated beam. This discussion will mostly address accelerator applications and the issues related to high-power operation.

Traditionally, one of the attractive features of superconducting RF cavities has been the low losses that enable operation in continuous wave (CW) with relatively high gradients and considerably lower dissipation than for normal-conducting cavities. So much so, in some areas of accelerator physics, CW operation and pulsed operation have often in the past automatically identified one type of technology or the other. Accordingly, the parameters and the design requirements imposed on couplers have also been dictated by the average power levels and the peak power levels that the couplers would have to handle.The RF power dissipation of a superconducting accelerating cavity ranges in the 10 to 200 Watt interval depending upon the operating frequency and temperature.

The Quality Factor Q_o of a cavity is proportional to the ratio between the stored energy U and the power dissipation P:

$$Q = \frac{\omega U}{P} \qquad (1)$$

From equation (1), a rough estimation of the dissipation can be obtained remembering

$$Q = \omega \frac{\int_v \frac{B^2}{\mu}\, dV}{R_S \int_S H_{//}^2\, dS} \qquad (2)$$

$$P = \omega \frac{\int_V H^2 dV}{Q} \le \frac{\omega H^2 V}{Q} \qquad (3)$$

For a LepII cavity, operating at 350Mhz at an accelerating field of 7 MV/m the Q_o value is $3x10^{9,}$ the a volume is 0.8 m^3, the peak magnetic field B 28mT.Assuming the magnetic field constant on all the volume V(excess approximation) we get from (3) a dissipated power of ~30 Watt.

The primary task of a coupler is an efficient transfer of power from a generator to a “load” (cavity and beam). From this perspective, a coupler can be considered as a properly designed transition in an otherwise perfectly matched transmission line, by which a properly determined energy admission rate can be delivered to the beam.

The power transferred to the beam can span a very broad range depending on the accelerator’s application. The beam power can be as low as few watts (heavy ion linacs as Atlas or RIA in USA, ALPI in Italy). Or reach the hundred of Kilowatt range in High Luminosity storage rings as B-factories (~100 KW for a LEPII cavity up to 300KW for KEK-B): even higher beam power is foreseen in the next generation of Synchrotron Radiation sources and High Intensity linacs for Neutron Spallation, Neutrino Factories and Muon Colliders.

For the above reasons the couplers will be designed for a broad range of coupling coefficients, usually with a strong over coupling factor (ratio between the power delivered to the beam and the power dissipated in the cavity walls) ranging from 100 to 10000. Superconducting cavity couplers must work reliably on a wide range of coupling coefficients depending on the different operating conditions of the accelerator. During the operation the coupler have to sustain strong variations of standing wave ratios (VSWR), high peak voltages and high peak currents

1.2 The coupler as a Vacuum barrier and a low losses Cryogenic transition.

Because of their proximity to superconducting cavities, most couplers must play also other roles, which substantially complicate their design and possibly limit their performance. Two of these “derived” or “secondary“ functions are

Vacuum barriers; between atmospheric pressure at room temperature, and low-temperature vacuum at extremely low pressures.

Thermal transitions between room temperature RF transmission systems and the low-temperature superconducting cryogenic environment, with or without dynamic heat loading generated by the RF.

The first function is played by radiofrequency windows (usually low loss tangent ceramics but also in some case [9] thin foils of high mechanical strength organic polymers as Polymide) Great deal of care needs to be used in setting the position and the operating temperature of the windows to avoid condensation of the residual gases. The condensed gas can dramatically change the RF properties of the window leading to enhanced dissipations or change in the electron secondary coefficient of the material. The occurrence of such events can very easily induce the failure of the window either by local heating and burnout (due to the increased the RF dissipation in some spot of the window), or by electron resonant discharges (multipacting) sustained by a secondary emission coefficient greater than unity. The enhanced secondary yield can produces a strong electron multiplication avalanche leading to puncturing of the window.

The thermal insulation task is accomplished trough a quite complex and sophisticated combination of low thermal conductivity low RF losses materials, counter-flow refrigeration of the coupling loop or antenna, choice of intermediate temperature intercept point.

2. MAIN COUPLERS: COAXIAL OR WAVEGUIDE?

Of all the possible geometries for coupling to superconducting cavities, two main choices have been adopted: coaxial and waveguide coupling.

Not being limited by a cutoff frequency coaxial couplers are in general more compact, especially for low frequency systems, and a variety of geometry and window arrangements are available to adapt to the specific need of the system; only power density considerations and suppression of multipacting levels play a role in determining the size of coaxial coupling systems. The design (both mechanical and electromagnetic) and the construction are somewhat complex and tricky due to the large number of components and the sudden changes in the electric impedance.

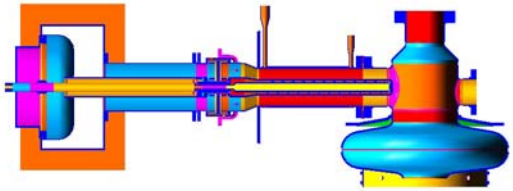


Fig. 1 The SNS coupler is based on a modification of the LEPII [15] coupler. The window matching design has been applied in the past to room-temperature systems. The SNS coupler developments has greatly benefited from the experience and the collaborations of several laboratories and industries around the world and has reached 2 MW peak in high-power tests

Waveguide coupling is conceptually simpler, since it does not require a transition between the waveguide, usually carrying the power from the RF sources, and the cavity interface. This solution, dating back to early stages of superconducting RF designs [10], was adopted in two accelerators now in operation (Cornell (CESR)/CEBAF, CESR-B, Figure 2. Due to the existence of a cutoff frequency in waveguides, the size of the coupler is generally larger at a given operating frequency than for the coaxial case. Because of the larger cross section of the coupling line, the contribution to the infrared heat transfer to the cryogenic environment is usually larger [11].

Fig. 2 The CESR-B single cell cavity makes use of a waveguide coupler. This is the highest power waveguide coupler in operation, having reached close to 300 kW CW in operation with beam. [12]

The CEBAF upgrade project [13] will use waveguide coupling, and new design options for TESLA have been considered which involve waveguide coupling as well [14]..

2.1 Coaxial Couplers

The ancestor of the Coaxial Coupler used to feed RF power to the beam in superconducting accelerator is the Doris coupler developed in Kernforschungszentrum Karlsruhe (KFK) in the early '80, figure 3.[1]

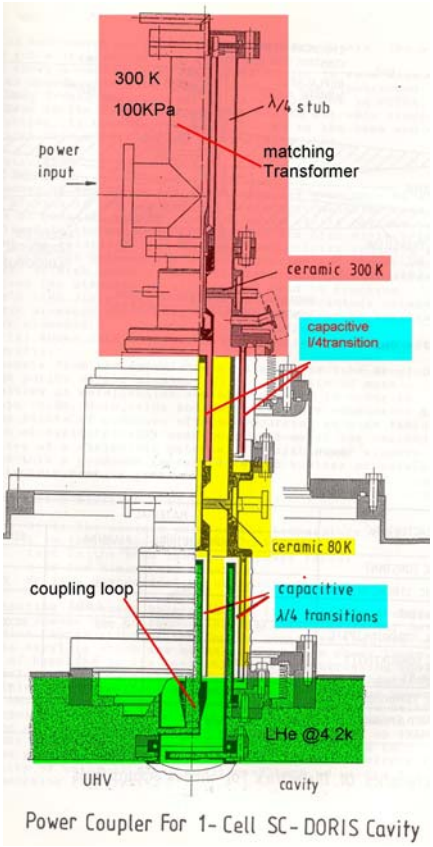


Fig. 3 Doris S/C cavity Coupler (1980 circa) 50KW @ 500 MHz rated powerThe coupler was developed in the Joint effort CERN-KFK for the proof of feasibility of a Superconducting RF Accelerating System to be used to boost the energy of LEP up to 100 GeV per beam.

Despite the early stage of the application of superconductivity to accelerating structures, the design tackle and solves all the key problems in such a device. The Doris coupler was used in 1980 at Desy in the first successful test run of a superconducting cavity used in an e^+e^- storage ring

Proceeding inside to outside the cryostat, we first find the Low Temperature (4.2K) section of the coupler. In this coupler, (as in all the early coax-couplers of the time) the coupling is obtained through the Magnetic Field. For this reason the coaxial line ends with a superconducting loop and a rather complex field transformer (Helium cooled) used to obtain a pure magnetic field coupling. The loop field transformer section was (together the windows) the trickiest part of the design. Due to dimensional constraints, the design team was forced to insert a demountable flange between the Loop and the field transformer. The RF current in the joint was quite large and produced some heating despite the use of a clever all-Niobium superconducting joint. This "Crocodile Joint" was able to scratch the thin oxide layer (10-15 nm) growing on the fresh niobium surface.

The low temperature (4.2K) section was thermally insulated by the intermediate section of the coupler working at 80K.

The RF transmission was guaranteed by a capacitive coupling (λ/4 at the operation frequency) among the inner and outer coaxial parts. The resonant length of the line gives a first order cancellation of the reflected power at the electrical discontinuity, forcing an electric short at the conductor gaps. The capacitive coupling allows thermal conduction between the two temperatures only by radiation, ruled by the Stefan Boltzman law.

$$Q = \sigma(T_1^{\$} - T_2^4) \tag{4}$$

σ=6x10^8 [watt]/[Kelvin]/[m^2]

The use of quite small surfaces and the intermediate temperature of 80 K, kept at a reasonable the heat flow from 80 to 4.2 Kelvin. Reasonable means, in this context, lower than the static heat load @ 4.2K coming from: the helium tank suspensions, the radiation from the helium tank and the 80K radiation shield, plus the dynamic load from the RF dissipation on the cavity surface. The 80K section is symmetric through the center plane, where the low temperature window is placed.

The window is a flat low loss ceramic brazed to the inner and outer conductors of the coaxial line. The outer conductor is in thermal contact with the 80K intermediate shield of the cryostat: the inner conductor is conduction cooled through the ceramic.

The 80K ceramic play in this case a triple role of:

Mechanical fastener of the inner conductor,

Thermal link: between the 80K thermostat and the inner conductor of the intermediate temperature section.

Vacuum Barrier between the UHV region in the coupler and the Cryogenic vacuum inside the cavity.

The 300 K section is electrically coupled to the 80k section in the same non-contacting way using the resonant capacitive coupling (λ/4 at the operation Frequency) used at the 4.2k side of the coupler. A room temperature coax ceramic window, similar to the 80k one, is used as a separation between the standard RF air coaxial line at atmospheric pressure, and the UHV room temperature line coupled to the 80K section. The λ/4 stub at room temperature is used to compensate the residual impedance mismatch introduced on the line by the window and the discontinuities produced by the capacitive coupling and the 90 Degrees bend.

Due to the rather modest (50-60KW) power level of operation, the KFK-CERN Doris Coupler was fully coaxial, connected to the RF Amplifier via a 3 1/8” standard EIA air coaxial line at room temperature. At higher power (greater than 100 KW), exceeding the air coax line rated power, as in Lep II and TRISTAN, the RF power distribution system at room temperature must use wave-guides.

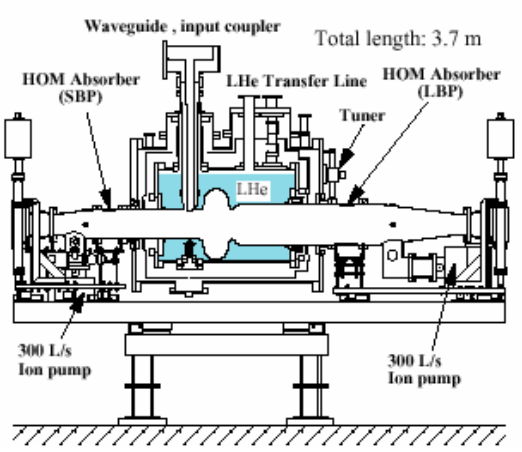


Fig. 4 The KEK-B coupler is the coaxial counterpart of the CESR-B coupler. The design makes use of a single planar coaxial ceramic window at room temperature. Operation close to 400 kW has been achieved. [16], [17]

The design for the main couplers for TRISTAN, LEPII [15] and HERA was mainly the evolution of the coaxial coupler used in the copper cavities of LEPI. The only relevant difference was the coupling to the accelerating cavity trough E-field at the beam tube. This feature is peculiar for the superconducting cavities; due to the high quality factor ($Q0 \sim 10^9$-10^{10}), the field strength of the evanescent wave in the beam tube is high enough to allow a coupling strength giving a loaded quality factor Q_L in the 10^5-10^7 range, and a optimum power transfer (critical coupling) at full beam. Modifications of this kind of coax coupler are currently used in KEK-B, up to 400KW continuous wave, and up to 2MW peak in the SNS cavities.

Coaxial couplers offer one more advantage: the impedance of the coaxial line can be chosen to be different from the standard 50 Ω, without modifying the coupler's outer dimensions in order to modify the power levels at which multipacting can occur [18].

The coupling strength depends on the longitudinal location and the size of the coupling port, but in the case of electric coupling, a large range of coupling values can be achieved by proper insertion of the center conductor into the line. Therefore, proper matching can be easily obtained by changing only one parameter and variable coupling can be achieved with proper (if not simple) adjustment of the inner conductor. This last feature can be achieved in accelerators having a too large (to be dealt just by RF controls) variation of beam loading during operation.

When the beam loading variation exceeds over an order of magnitude the coupling factor (and Q_L), variable coupling become mandatory, as in the LHC cavity coupler.

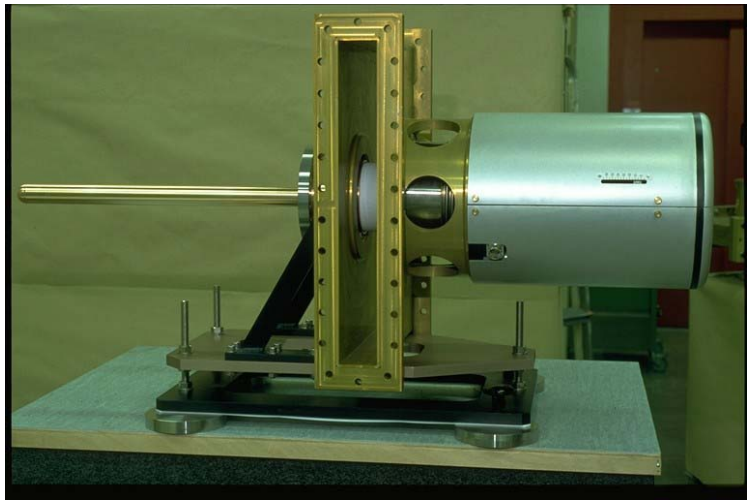


Fig. 5 Variable coaxial coupler for the LHC. The variable insertion mechanism allows a change in coupling factor by over an order of magnitude. This coupler has a cylindrical window in the waveguide. [19]

2.2 Waveguide couplers

A few important features of the waveguide coupling systems deserve some mention here. A good review of some of these issues is found in [9]. Because of the waveguide geometry,(a large hollow tube) windows for waveguide couplers are generally more difficult to manufacture and more prone to break. Multiple windows within the waveguide's cross-section have been used [19]. The coupling strength can be adjusted in three basic ways: 1) by the size of the coupling iris, 2) by the longitudinal location of the waveguide with respect to the cavity's end cell, and 3) by the location of the terminating short of the waveguide itself, as in the case of CEBAF's cavities.

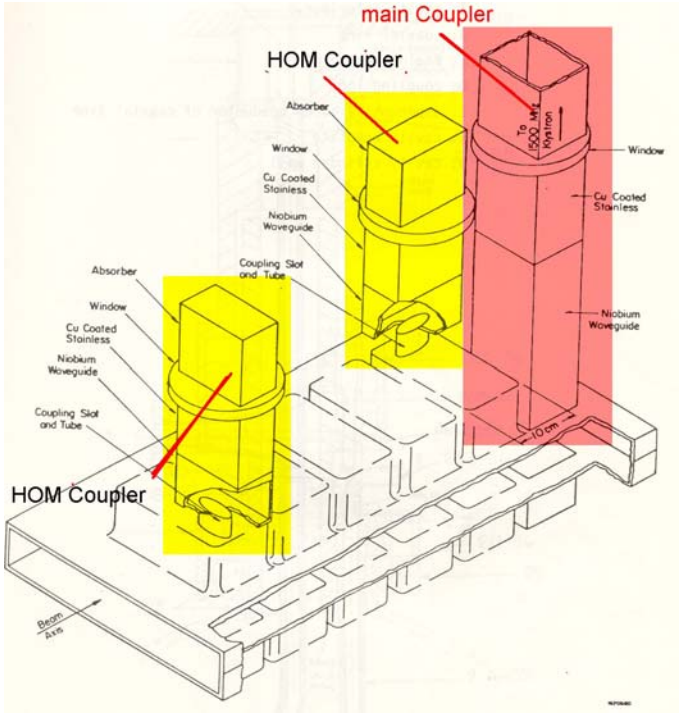


Fig. 6 Muffin Tin Cavity Waveguide coupler. Developed in CORNELL for use in a L-band cavity for e^+e^- Storage rinfg, in the late '70, early 80, contains all the features of the Waveguide couplers used in CEBAF and CESR.

The first example of waveguide coupler was the coupler for the muffin thin cavity (figure 6) developed in CORNELL in the late ’70 early ’80. [23] The coupling system was designed to handle 60 KW at 1500 MHz: The cavity was formed essentially of eleven cells. The top of the first cell was replaced with a superconducting waveguide, acting as the continuation of the cup, propagating the fundamental TE_{10}.The waveguide field couples with the electric field components of the accelerating structure. The bottom of the cell, opposite to the coupling port, was used as a shorting stub to adjust the coupling between the waveguide and the accelerating cavity.

A cold ceramic window operating at 80K was used. To keep to a minimum the heat flow from room temperature to the helium bath, the coupler used low conductivity stainless steel wave-guides copper plated. Two section of such waveguide provided breaks in the heath path between 4.2 and 80 K, and between 80 and 300K.

3. WINDOWS

Windows are designed to separate the vacuum of the superconducting cavity from the atmospheric pressure of the transmission line. As electromagnetic interfaces, they must satisfy strict matching requirements, so that power is reflected and dissipated only in minute quantities. Since dielectric materials are used for the transmission of electromagnetic power [24], the manufacturing techniques usually involve complicated interfaces of conductors, dielectrics and brazing metals.

In addition, electronic phenomena at the windows can complicate the design. Multipacting at the windows can be particularly dangerous, as large amounts of power can be deposited in small areas of the dielectric, potentially leading to failure. Careful choice of geometry and coating with low secondary electron emission coefficient materials can mitigate this phenomenon [3].

Exposure to radiation can also lead to charging phenomena at the window surface [25], [26], [27], 28] leading to flashover of the accumulated charge and to damage of the window. Geometrical protection [27] as well as metallic films of proper thickness can be used to decrease the incidence of this problem. As in the case of multipacting coating, it is essential that the appropriate thickness be carefully achieved (10-15 Å); otherwise excessive RF losses will occur and the subsequent excessive heating will lead to window failure [29]. In some cases, multiple windows in series are used (CEBAF: one window at 300 K and one at 2 K, the latter used for sealing cavity pairs as early as possible in the assembly process; TESLA: one room-temperature window and one at 70 K [30]; APT: two redundant room temperature windows for protection against failure) [8]. In spite of the added protection and some beneficial features, multiple windows tend to complicate the design of the couplers, add cost and increase the number of critical components that can fail.

As mentioned above, windows for waveguide systems are usually planar and can occupy a large fraction of the waveguide cross-section, either in a single piece, or in multiple pieces [25], [31], [32], [33].Coaxial windows are usually planar [8], [22], [34], [35] cylindrical [9], [36], [37] or conical [38].

Active pumping near the windows is desirable [6], [25] to avoid discharge problems during outgassing events associated with varying power levels, but in most cases design complications make this solution impossible and pumping is achieved only through the cavity itself. In this case, more careful initial conditioning and close attention to operational interlocks become even more necessary.

4. SIMULATIONS

Over the past several years, an ever-increasing activity has been noticeable in the area of coupler design connected with simulations of various aspects of the coupler’s performance. Thanks to better software tools, a larger and larger fraction of the design of couplers can be made well ahead of the actual construction, thus removing part of the uncertainty of the coupler’s performance and avoiding

the lengthy, tedious and expensive work of cut and try, which is particularly demanding for systems connected to superconducting cavities.

4.1 Electromagnetic calculations

4.1.1 Field distribution

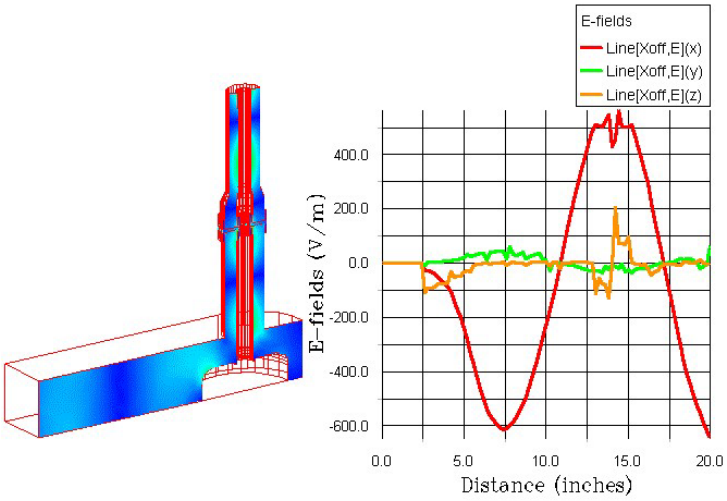


Fig. 8 Electromagnetic field simulations allow better designs of coupler components: here the SNS waveguide-to-coaxial transition and the window matching section are shown with the relevant electric field strengths. [46]

Programs such as HFSS™ [63] have been used by several groups to evaluate the field distribution in couplers and transitions and to improve the matching at windows and at waveguide/coaxial transitions. Such calculations have been carried out, for example, for couplers designed at Saclay [40], [41], at LANL [42], [34], [43], [44], [45], for Cornell windows [32] and for the SNS coupler [46] (Figure 8)

4.1.2 External Q calculations

The coupling strength of the coupler (external Q_{ext}) is a critical quantity, which needs to be set for each specific application. Calculations are also now routinely performed to determine the Q_{ext} in advance by matching cavity field calculations to the coupler's geometry field simulations. This has been done for the APT cavities [47], [48] and for the SNS [46]. In both cases, bench measurements give extremely good agreement with the simulations.

4.1.3 Multipacting calculations

Along with better understanding of field distribution in couplers and with the improvement in tracking programs for multipacting in accelerating structures, a great improvement has been effected in understanding electronic activities inside the couplers' structures and in estimating location and magnitude of multipacting phenomena [49].

Such efforts have been carried out at Cornell for the waveguide geometry [19], [20], and at Saclay for various window and coaxial geometries [40], [41], [50], [51], [45]. Activities in Finland in collaboration with TESLA and other laboratories have led to the study of the multipacting characteristics of several coupler geometries [52], [53], [54], [55]. From these studies a great deal of information has emerged which points to the fact that multipacting is generally unavoidable in couplers, as the amplitudes and phases of the forward and reverse wave change along various parts of the structure. Figure 9 and figure 10 give an example of the output of the multipacting simulations. The final result of the simulations is that electromagnetic design alone is insufficient to avoid multipacting. Materials and surfaces must be carefully controlled and conditioning must be implemented in order to decrease the negative impact of this phenomenon. As a side result of the

multipacting simulations, it is now possible to design a proper biasing method in order to disrupt the multipacting orbits and their effects. Calculations with bias can be performed with the present multipacting modeling tools [56].

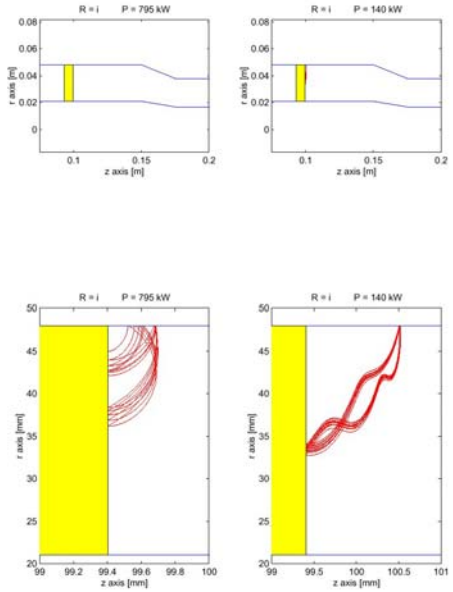


Fig. 9 Multipacting simulations at the SNS coupler's window. The ability of predicting dangerous phenomena, like multipacting at specific locations in the couplers, leads to better designs. [54]

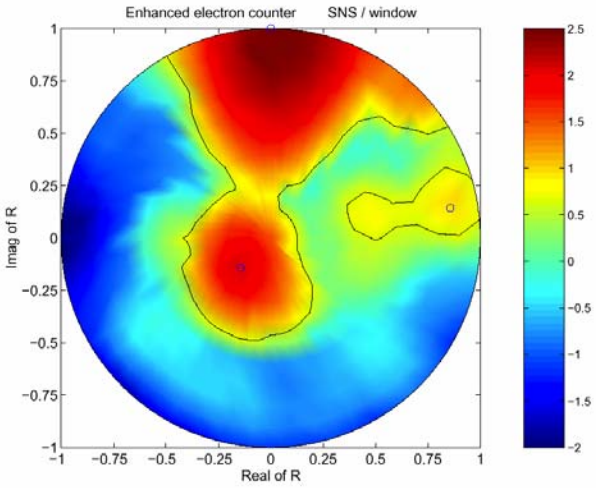


Fig. 10 The results of multipacting simulations in couplers can be summarized by graphs like the one shown above: the red areas show regions of the complex reflection coefficient where multipacting can occur [54].

4.2 Thermal calculations

Another area, where numerical calculations help in designing the coupler characteristics, is the determination of the thermal properties. For coaxial geometry, this has been done at the APT [57], [58] and at the SNS [35]. Both center conductor and outer conductor thermal profiles can be determined in this way under RF loading conditions. Similar calculations have been done for the waveguide geometry by other authors, taking into account the optimal length and thermal groundings to minimize cryogenic losses [15].

An area that requires additional attention is the modeling of the thermal profile of the coupler/cavity interface. Here the RF losses are small, but if the temperature is not properly stabilized,

the superconductor's critical temperature could be exceeded due to the highly nonlinear losses caused by the thermal loading from the coupler.

4.3 Mechanical stress calculations

Since some of the coupler designs rely on very delicate ceramic-to-metal brazing, it is important to evaluate the mechanical properties of the couplers to prevent costly mechanical failures. Whereas in most cases couplers are assembled on the accelerator premises and a failure of a coupler in transfer only affects the coupler itself, in the case of the SNS the assembly is done elsewhere from the installation point. A failure of a coupler during transport would have very costly consequences. In the future, this construction mode will become more and more frequent. Wilson [35] has evaluated the mechanical stresses on the SNS coupler, and the results indicate that the design should withstand the accelerations and stresses associated with the transport from one laboratory to the other.

5. DEVELOPMENT OF HOM COUPLERS FOR SUPERCONDUCTING CAVITIES

A single mode Accelerating cavity (as an ideal Lumped LC resonator) is the dream of any accelerator designer. Real cavities have many resonant modes corresponding to the solutions of the Maxwell equation inside the cavity volume. Particles traveling along an accelerator send information about position and speed to the following particles through the radiated electromagnetic field. This cross talk is enhanced at the occurrence of the resonance of the high modes of the cavities. If you are unlucky, (and you are always) particles will interact together via the radiated fields, giving rise to beam instabilities limiting the maximum current the accelerator can handle. In S/C cavities, due to the long decay time of the fields, the deposited energy need to be dumped on a load, lowering the threshold for the beam instabilities.

Coherent instabilities limit the beam current in circular machines (coupled bunch instabilities) or in linear accelerators (regenerative, cumulative and multipass Beam Break Up). These instabilities are mainly caused by the long-range wake fields excited by the beam bunches going through the cavities. The best way to increase the threshold is to fasten the decay of the excited fields, in other words to damp the higher order modes of the cavities. Typical Q values under 10^5 or 10^4 for the most dangerous parasitic modes are required in present designs of accelerators.

By chance the accelerating mode is the fundamental mode of a cavity, the mode at the lowest frequency. A device having an high pass frequency response, will allow to extract the RF power radiated by the particle beam, leaving unaffected the RF power fed to the cavity (and to the beam) by the main coupler; in this way the overall cavity response approach the ideal response of a lumped LC circuit.

Various devices and especially designed for superconducting cavities have been developed during the last years. The main differences lie in the type of coupling -hole, probe or loop used to couple out from the cavity the unwanted modes and in the way of keep to a minimum the coupling with the accelerating mode. Nevertheless, all the last versions follow the same rule: they couple to the cavity fields through the beam tube and not directly in the cells. The risk of multipacting and of thermal breakdown is thus avoided. However, care must be taken in the design of cavities with the so-called trapped modes, resonating at frequencies above the cut-off frequency of the beam tube, but having the field going rapidly to zero outside the cavity.

5.1 Estimation of the extracted power by HOM coupler

The cavity codes like Superfish [59] or Urmel [60] OSCAR2D [61] help in designing couplers that will be mounted on the beam tube. Once the electromagnetic fields at the foreseen location of the coupler known; assuming:

The coupling port does not changes too much the field pattern

The RF coupler behaves like an ideal filter without reflection (terminated on the characteristic impedance).

We can easily deduce an approximate value of the real damping that will be effectively measured on a copper prototype.

5.1.1 HOM waveguide cou1pling

The predictions can be hazardous because the theory generally developed concerns only small coupling holes (see for instance [62]). We could use instead 3D cavity codes as HFSS™ [63] or MAFIA [64] combined with one of the two later to evaluate the power flow through the waveguide terminated on a RF load. Otherwise, the waveguide coupler and the matching waveguide stubs are designed by cut and try.

5.1.2 Lumped coaxial filters coupling

The estimations, deduced from field level calculations with Urmel or Superfish, agree very well with the measurements.

For a probe coupler the electric flux on the probe tip furnishes the current induced by a cavity mode:

$$I = \omega \varepsilon E S \tag{5}$$

Where E is the electric field from a mode averaged over the probe tip and S is the antenna area.

The power dissipation on a matched resistive load R is given by

$$P = \frac{I^2 R}{2} \tag{6}$$

The external Q of this simple coupler terminated on a resistive load R for a mode with stored energy U is computed from the external Quality Factor definition (equation 1)

$$Q_{ext} = \frac{\omega U}{P} = \frac{2U}{R\omega\varepsilon^2 S^2 E^2} \tag{7}$$

In the same way for a loop coupler, the magnetic flux going through the loop furnishes the voltage induced in the loop by a cavity mode:

$$V = \omega \mu H S \tag{8}$$

Where H is the mode magnetic field at the loop location, averaged over the loop area S; the power dissipation on a matched resistive load R is given by

$$P = \frac{V^2}{2R} \tag{9}$$

The external Q of this simple coupler terminated on a resistive load R for a mode with stored energy U is

$$Q_{ext} = \frac{\omega U}{P} = \frac{2RU}{\omega \mu^2 S^2 H^2} \qquad (10)$$

5.1.3 The coupling via beam tube and its limits

The external Quality Factor evaluation is valid for any choice of the HOM position. Usually the location of the HOM coupler (as for the main Couplers) is on the beam tube, to avoid introducing (as already happened in the CERN-DESY-KFK cavity) strong modification to the cavity geometry in the high field region of the resonator. This choice helps to reduce to a minimum the possibility of producing field limitations either by increased RF losses due to local field enhancements or by enhanced field emission

The most efficient HOM coupler will be inadequate whenever the field level will be vanishing at the location of the coupling port. Measurements on a single cavity [65] showed the existence of modes above cut-off of the beam tube, which couple poorly to the propagating waveguide modes of the beam tube itself.. The situation is even worse in the case of a multicell cavity .as these trapped modes appear at a frequency well below three times the fundamental mode frequency [66](the usual cut off frequency of the beam tube.

The cavity codes predict these troublesome modes. For example, figure 11 shows the plots of the electrical field of two modes above cut- off for a 5-cell cavity with large iris aperture (ϕ/λ=O.35). One mode (top) belonging to the fifth dipole pass band remains confined inside the cavity (no field in the beam tube) in contrast with a lower frequency dipole mode (bottom), which couples clearly to the propagating TE_{11} mode of the beam tube. We expect for these trapped modes a very poor damping by the HOM couplers but we could hope to evacuate the power by propagating waves through the beam pipes.

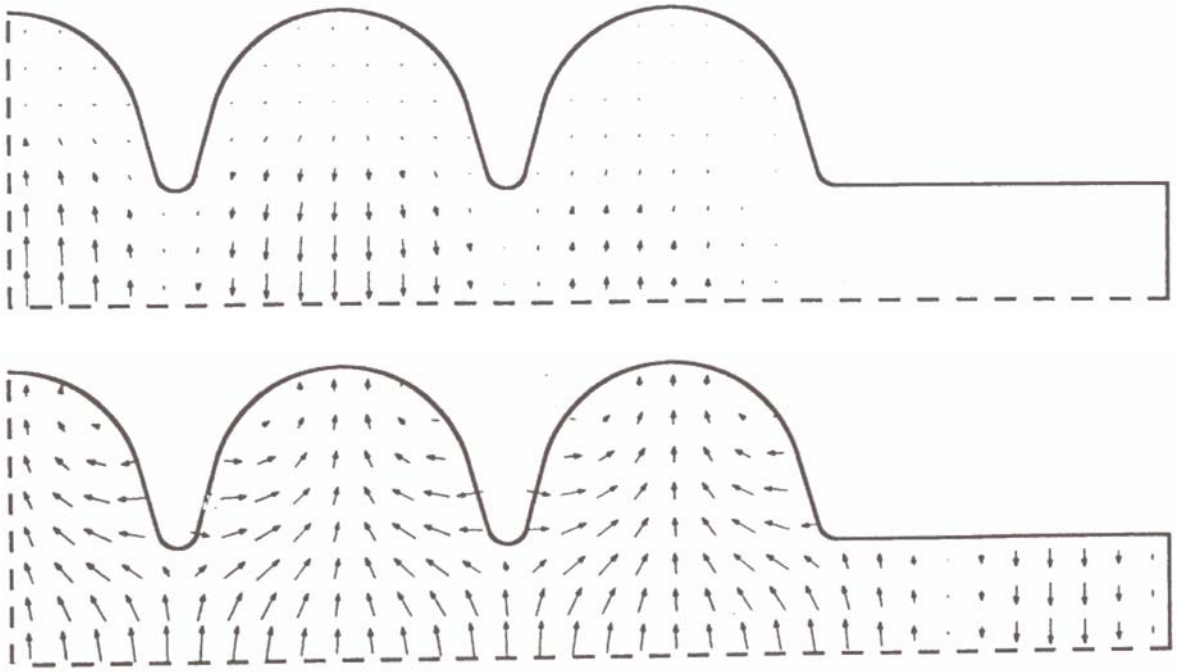


Fig. 11 Plots of the electrical field of two dipole modes, both above the cut-off frequency of the beam pipe .Due to the field configuration the first mode is trapped in the center cells, with a minimum propagation down the beam tube.

5.2 What is required of real life HOM coupler

5.2.1 The RF requirements

In addition to the damping of all dangerous modes as much as needed, we require from a HOM coupler that the coupling with the fundamental mode must be kept as small as possible. In the case of the waveguide coupler, this requirement is naturally obtained by a correct choice of the cut-off frequency and length of the waveguide.

Due to the large size of waveguides, coaxial couplers combined with a Superconducting notch-filter, for suppression of the fundamental mode, are preferred for low frequency cavities as in the case of the LEPII RF system.

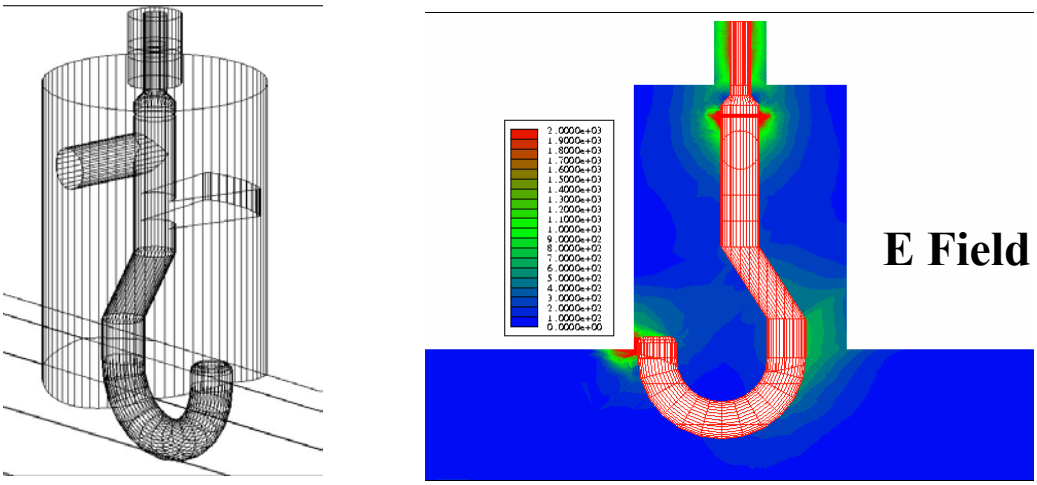


Fig. 12 Lumped beam tube HOM coupler, including a fundamental mode notch filter. The right side picture shows the E field distribution In the coupler.

On the other hand, great deal of care must be used in tuning the notch filter to avoid coupling a significant amount of RF power from the accelerating mode.

This effect will reduce the Q_o value of the accelerating mode, impairing the transfer of power to the beam, and increasing beyond any affordable limit the power dissipated on the HOM RF load.

Cryogenics also play a strong role, mainly in the design and on the cooling of the inner parts, to avoid excessive heating and quench of the coupler. In the few gigahertz range, both types of HOM couplers have been developed in different laboratories.

5.2.2 The "engineering" requirements

The RF properties of the couplers are the reason why laboratories try hard to simplify the design and reduce the cost as much as possible. Often the HOM coupler designer faces conflicting "engineering" requirements:

- **Low cost** (the number of HOM coupler in a cavity is quite large)
- **Effective Cooling** of the SC parts exposed to high fields for preventing from thermal quench
- **Compact structure** and minimal size to fit in the cryostat
- **Demount ability** to allow surface treatments or thin film deposition of the cavity alone
- **Avoid large RF windows** (if possible) in contact with helium to avoid damage to the cavity in case of vacuum failure

5.3 Coaxial HOMcouplers development

Designing coaxial HOM couplers we need first to choose the way of suppressing the fundamental mode and then to control the RF elements to design a fairly flat high pass transfer function to properly transfer to the RF Load the energy deposited by the beam on the cavity modes resonating at a frequency higher then the one of accelerating mode.

Coaxial lines do not have a low frequency cut of; but start to carry RF power from zero frequency; we need in some way to break the path for low frequency signals; a capacitor along the inner conductor does the wanted job of transmitting unaffected only the RF signals above the fundamental. Usually this is not enough to keep to a reasonable value the transmission of the fundamental mode.

For an antenna or loop coupler we must compensate the stray capacitance between the probe tip and cavity walls or the self-inductance of the loop. Numerical methods are used to optimize the RF circuit before the final step of measuring the transmission curve of a coupler prototype.

The design starts with a first rough trial design used to set up a 3D mechanical model of the coupler. The mechanical model is used to perform a 3D FEM RF analysis (usually using HFSS ™ or MAFIA) of the coupler computing the RF fields, cut off frequencies and Scattering parameters on the frequency range foreseen for the coupler.

Once the wanted frequency response obtained, a full analysis of the Cavity-Main Coupler-HOM coupler transfer function is mandatory, to check for unwanted and unforeseen effects, in the global RF system, produced by the mutual interaction between the different components of the RF system.

Last a full-scale mock-up is built to verify the correctness of the numerical simulation and perform a sensitivity check of the Coupler response against the build up of mechanical errors produced by the build up of the mechanical tolerances.

Figure 13 shows the measured and the calculated response curves for the HOM coupler shown in figure 12

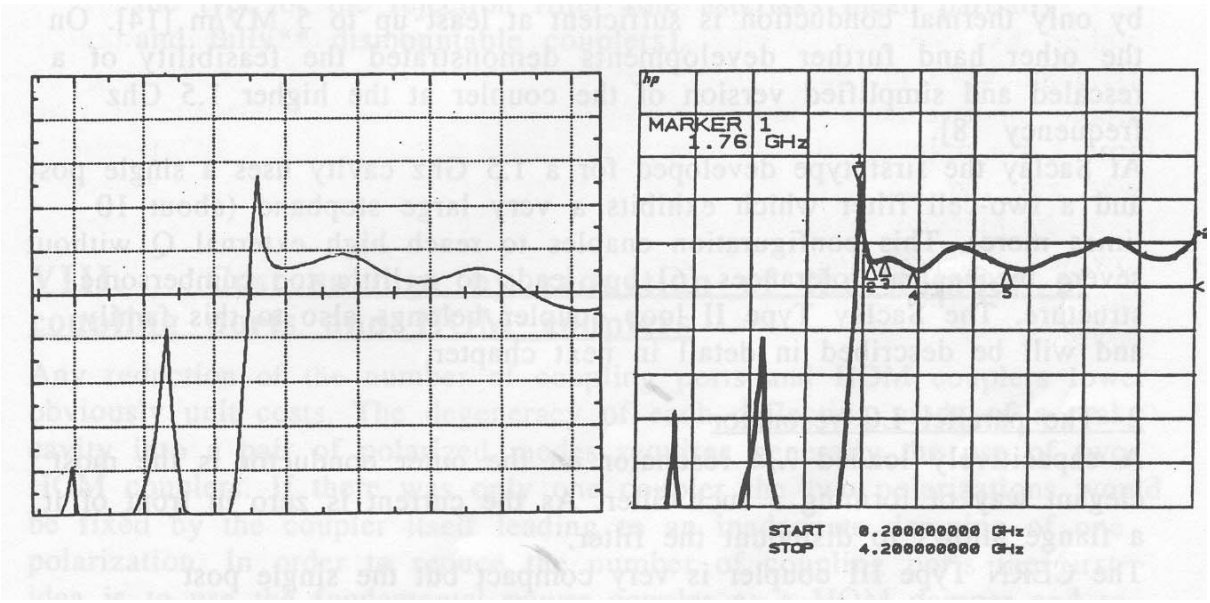


Fig. 13 Computed (left) and measured (right) Transfer function of a Prototype coaxial HOM Coupler. Developed in Genoa for a reduced beta L-Band cavity at 1.4 GHz Trasco project.

Coaxial Couplers mainly evolved from the design developed in the early ’80 for the Storage Ring RF systems as TRISTAN [68] and LEP[69].

The High pass frequency response is given by a capacitive gap in the central conductor; the rejection of the fundamental mode is obtained either through a series LC-resonator across the coaxial line or a parallel LC-resonator in series with the central conductor and the load.

A further adopted scheme was a distribute filter built using stepped impedance compact waveguides to perform the rejection task.

5.3.1 The series LC resonator

The first example of this type of HOM is the CERN Type I coupler [69]. It is fully dismountable but requires three small sapphire rods for a precise centering within the coupling port.

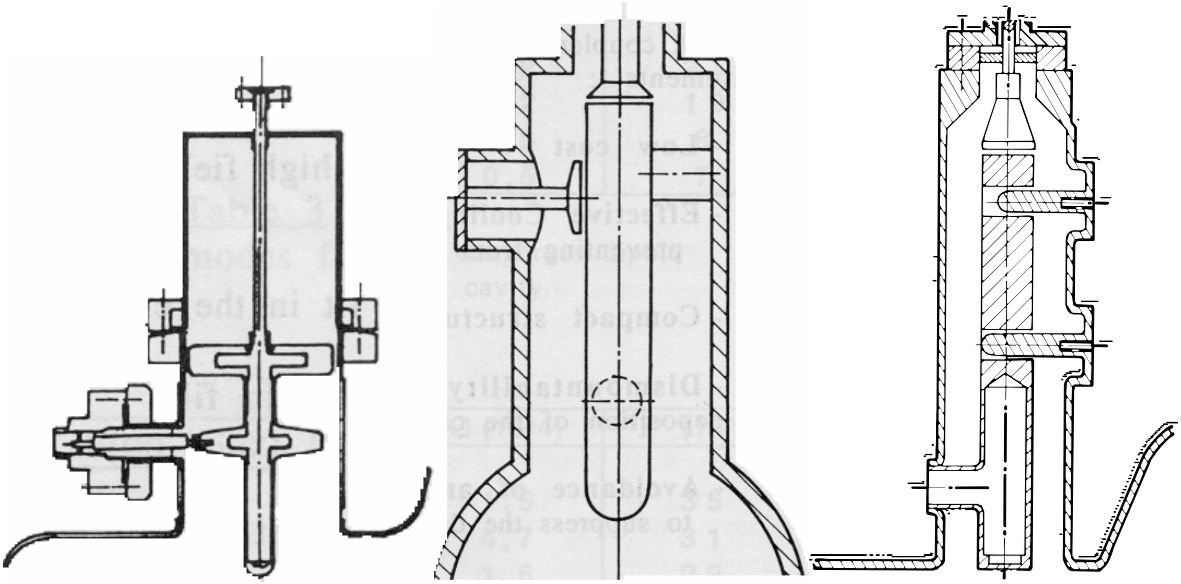


Fig. 14 Series LC resonator HOM Couplers: LEP I (left) DESY center and CEA-Saclay (Rigth)

The couplers developed at DESY for the 500 Mhz HERA cavities uses 2 inductive posts which are welded to the cavity walls [70]

At Saclay the first type developed for a 1.5 Ghz cavity uses a single post and a two-cell filter which exhibits a very large stopband This configuration enables to reach high external Q without severe mechanical tolerances but leads to a little too complex mechanical structure. The Saclay Type II loop coupler belongs also to this family.

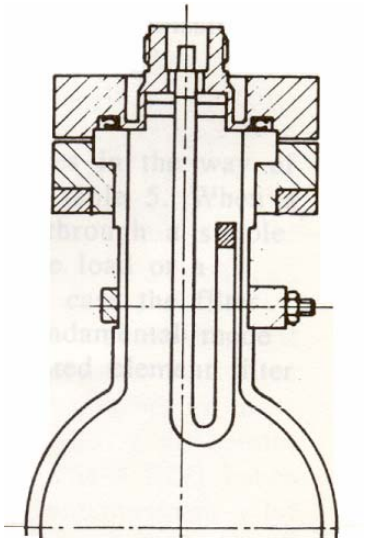


Fig. 15 CEA-Saclay loop coupler: The rejection of the fundamental mode is obtained by the resonance of the Inductance of the loop with the stray capacitance of the loop itself against the outer conductor of the coupler.

5.3.2 2- The parallel LC resonator

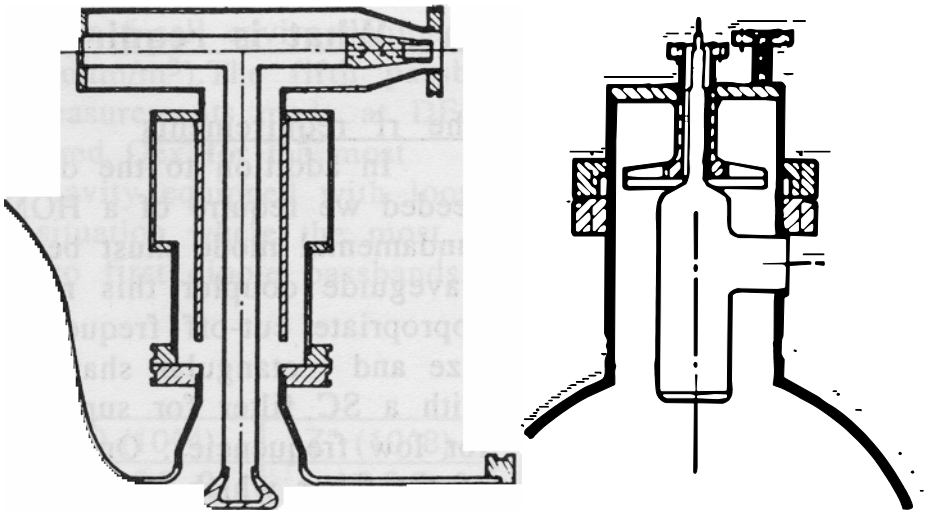


Fig. 16 KEK type HOM (left) and CERN III HOM right.

A capacitive loaded λ/4 resonator on the outer conductor is the most elegant way of forming the Fundamental mode rejection filter. As the current is zero at the λ/4 end, a flange at this position allows to dismount the filter.

The CERN Type III coupler is very compact but the single post supporting the inner conductor and used for cooling is welded to the cavity [69].

The KEK coupler is fully dismountable and includes a T stub at the exit allowing the cooling of the inner conductor

5.3.3 3- The distributed element filter

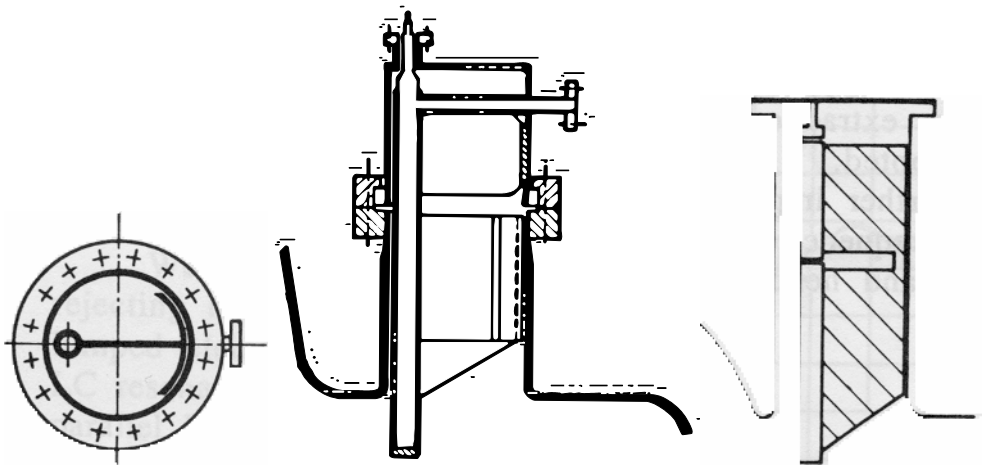


Fig. 17 Distributed element HOM filter Cern Type II a(left) and Type IV (right).The Coupler behave like an annular wave guide with a cut of Frequency.

This family (figure 16) is derived from the lunar guide, which has a cut-off frequency like a waveguide. To shorten the needed length for a sufficient attenuation either a distributed condenser along the sheet close by the outer conductor (CERN Type II) or a local LC resonator formed by a little aperture in the lunar guide (CERN Type IV) has been added [69]. While the second is welded to the cavity walls, the first is fully dismountable but with a filter only cooled by heat conduction with the risk of thermal quench.

The design of the HOM Couplers evolved from the three basic types by further refining and by crossed implementation of the more attracting features of the different Types.

The most successful design was the HOM coupler used In the LEP II cavity shown In Figure17

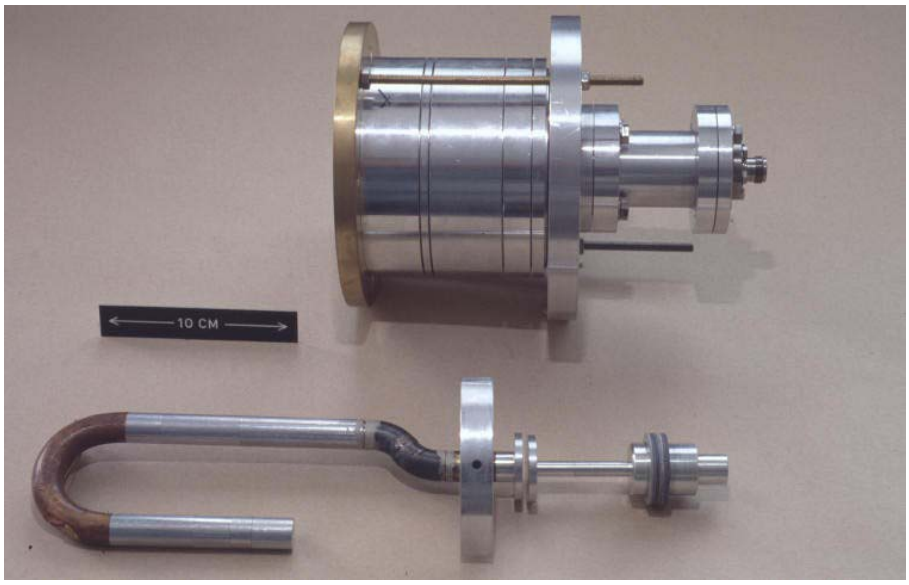


Fig. 18 LEP II cavity HOM coupler.

The design of the last generation of HOM couplers for the LHC cavities is even simpler: HOM couplers clearly show the evolution toward a lumped element circuit.

The E-Field coupled HOM uses a lumped capacitor to produce a series resonance at the fundamental mode frequency across the coaxial line; the H-field coupled HOM uses the stray

capacitance of the loop to produce the same effect. In Both couplers the two poles High Pass transfer function is forced by the lumped capacitance at the end of the central conductor and the inductance of the Coaxial line.

The Couplers are fully demountable, Helium cooled, and need only one weld between the central and the outer conductor (figure 18).

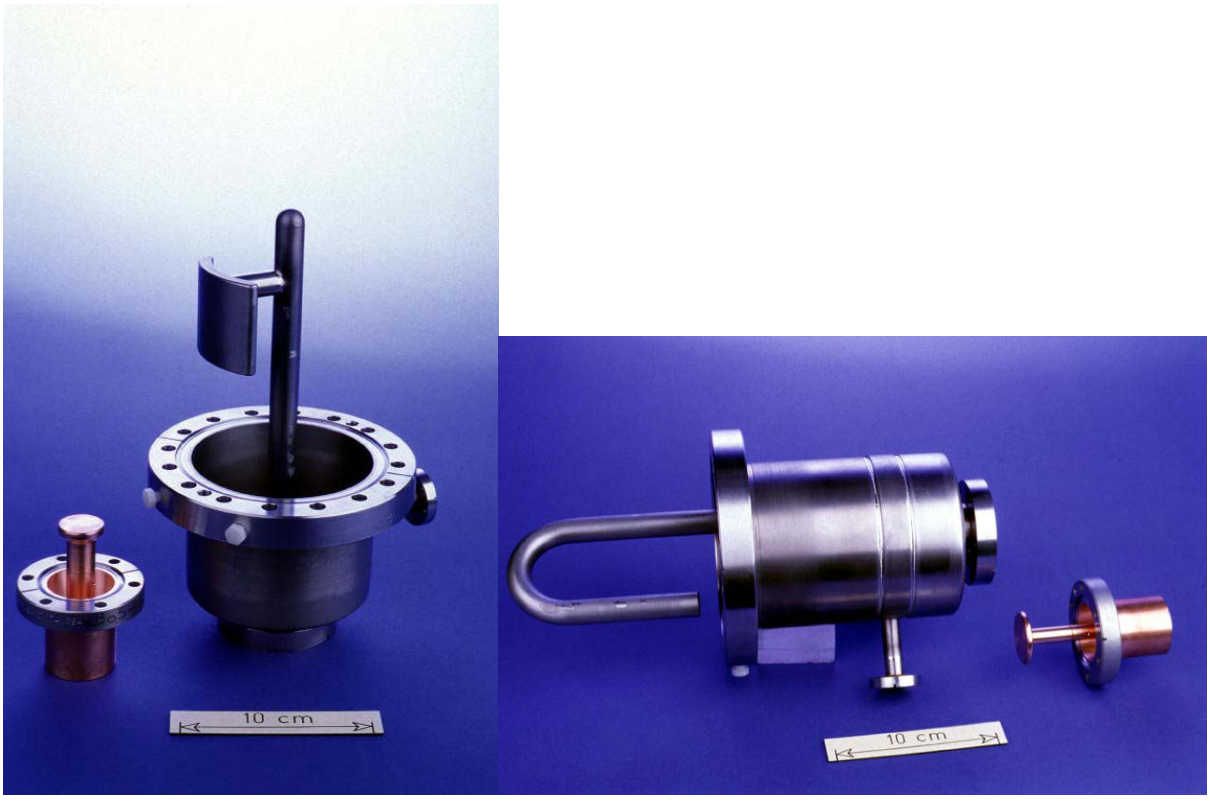


Fig. 19 The LHC Cavity HOM Couplersare the natural evolution of The LEPII HOM Couplers. The HOM use either E-Filed coupling (left) or H-Filed Loop coupling (right) to damp the Q value of the most dangerous modes of the cavity.

5.4 Waveguide HOM coupler development

The main advantage of the waveguide coupler is its high pass feature without the need of any tuning. The evanescent wave in the guide excited by the accelerating mode of the cavity must be sufficiently attenuated before arriving to the rf load. However, a guide length of a few attenuation length (at the fundamental mode frequency) is needed because the cut-off frequency is generally very close to the fundamental mode frequency. The first waveguide joint flange needs to be far enough from the beam tube for a negligible dissipated power.

In addition, a matching stub is required on the beam tube at the opposite side of the HOM coupler.

In the CEBAF/Cornell cavity [71] as the cut-off frequency is slightly higher than the first HOM's frequency, the fundamental coupler plus one extra stub is used for the the extraction. In order to avoid a pinning of the dipole mode polarizations by the main coupler, the Y configuration (two arms for extraction and one arm as matching stub) for the HOM couplers was adopted (Figure 19).

Further improvements at Cornell [72] allowed to couple out the lowest frequency modes through a modified coupling port of the HOM coupler and hence to eliminate the large stub on the fundamental power coupler.

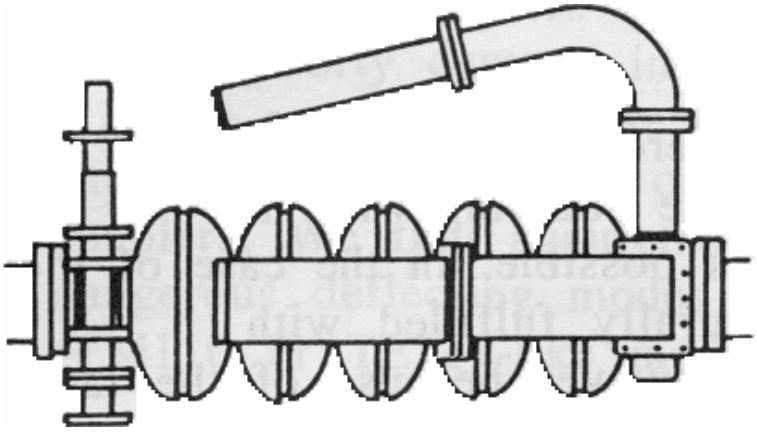


Fig. 20 The CEBAF module with the Input Coupler Waveguide (left) and the Two HOM waveguide couplers on the right.

The HOM coupler scheme is quite similar to the one used In the Cornell Muffin Tin cavity [9]; the extracted RF power is dissipated on a specially developed RF absorbers housed in the last part of the Waveguide section.

ACKNOWLEDGEMENTS

We wish to thank I.E.Campisi, from Jefferson Lab, for the very helpful and exhaustive paper on the Main Couplers presented at the RF/SC 2001 Conference held in Tsukuba (Japan). The paper was an invaluable source of information, references and pictures.

REFERENCES

[0] R.Lengeler "Superconducting cavities" (parts3 and 4) CERN Accelerator ScoolSuperconductivity in Particle Accelerators 1989 p.197

[1] R. M. Sundelin, "Joints, Couplers, and Tuners", Proc. First Workshop on RF Superconductivity,Karlsruhe, Germany, July 1980, P. 246.

[2] B. Dwersteg, "High Power Couplers", 4th Workshop on RF Superconductivity, p. 351.

[3] H. L. Phillips, "An Update on Windows, Couplers, Higher-Order-Mode Damping, and Interlocks", 6th Workshop on RF superconductivity, p. 267.

[4] E. Haebel, "Couplers for cavities," CERN Accelerator School: Superconductivity in Particle Accelerators, Hamburg, Germany, May 1995, p. 231.

[5] M. Champion, "RF Input Couplers and Windows; Performances, Limitations, and Recent Developments", 7th Workshop on RF Superconductivity, p. 195.

[6] E. Haebel, "Power Couplers: Sources and Consequences of Mismatches", 8th Workshop on RF Superconductivity, p. 725.

[7] H. Matsumoto, "High Power Coupler Issues in Normal Conducting and superconducting Accelerator Applications", PAC99, p. 536.

[8] I.E. Campisi "Fundamental Power Couplers For Superconducting Cavities" Proceedings Of the 10th Workshop On RF Superconductivity Tsukuba Japan September,6-11 2001.

[9] L. R. Doolittle, "Strategies for Waveguide Coupling for SRF Cavities", Linac98, p. 246

[10] W. B. Herrmannsfeldt (Ed.), "Feasibility Study for a Two-Mile Superconducting Accelerator", SLAC, 1969, Sect. V.B-4

[11] N. Jacobsen and E. Chojnacki, "Infra-Red Propagation Through Various Waveguide Inner Surface Geometries" Cornell University, SRF 990301-01, 1999

[12] S. Belomestnykh and H. Padamsee, "Performance of the CESR superconducting RF System and Future Plans", Proceedings Of the 10th Workshop On RF Superconductivity Tsukuba Japan September,6-11 2001

[13] J. R. Delayen, L.R. Doolittle, T. Hiatt, J. Hogan, J. Mammosser, L. Phillips, J. Preble, W.J. Schneider, G. Wu, "An R.F. Input Coupler System for the CEBAF Energy Upgrade Cryomodule", PAC99, p. 1462.

[14] B. Dwersteg, D. Kostin and W.-D. Moeller, "TESLA RF Couplers Development at DESY",Proceedings of the SRF 2001 Tsukuba Japan September,6-11 2001

[15] H.-P. Kindermann and M. Stirbet. "RF power tests of LEP2 couplers on a single cell superconducting cavity" 8th Workshop on RF Superconductivity, p. 732

[16] S. Mitsunobu, T. Furuya, T. Tajima, T. Kijima, T. Tanaka, "High Power Input Coupler for KEKB SC Cavity," 9th Workshop on RF Superconductivity, p. 505

[17] Y. Kijima, S. Mitsunobu, T. Furuya, T. Tajima, T. Tanaka, "Input Coupler of Superconducting Cavities for KEK-B", EPAC 2000, p. 2040

[18] E. Somersalo, P. Ylä-Oijala, D. Proch, "Analysis of Multipacting in Coaxial Lines", PAC95, p. 1500

[19] H. Kindermann, M. Stirbet, "The Variable Power Coupler for the LHC Superconducting Cavity", 9th Workshop on RF Superconductivity, p. 566

[20] D. Metzger, P. Barnes, A. Helser, J. Kirchgessner, H. Padamsee, "Test results and Design Considerations for a 500 MHz, 500 kW Vacuum Window for CESR-B" PAC93, p. 1399

[21] E. Chojnacki, S. Belomestnykh, "RF Power Coupler Performance at CESR and Study of a Multipactor Inhibited Coupler", 9th Workshop on RF Superconductivity, p. 560

[22] R L. Geng, H. Padamsee, "Exploring Multipacting Characteristics do a Rectangular Waveguide", PAC99, p. 429.

[23] J.Kirchgessner, P.Kneisel,H. Padamsee, J.Peters, D.Proch, R.Sundelin and M.Tigner, "Design Studies For A 1500 Mhz Superconducting Accelerator Cavity For Use In An e^+e^- Storage Ring", 11thInternational Conference on High-Energy Accelerators p.886.

[24] B. Dwersteg and Y. Zhenze, "Measuring Complex Dielectric Constant of Aluminum Oxide Window Ceramics for TTF Power Couplers at DESY"

[25] L. Phillips, J. Mammosser, and V. Nguyen, "New Window Design Options for CEBAF Energy Upgrade" PAC97, p. 3102

[26] L. Phillips, C. Reece, T. Powers, V. Nguyen-Tuong, "Some Operational Characteristics of CEBAF RF Windows at 2 K", PAC93, p. 1007

[27] S. Chel, M. Desmons, C. Travier, T. Garvey, P. Lepercq, R. Panvier, "Coaxial Disk Windows for High Power Superconducting Cavities Input Coupler", PAC99**.** p. 916

[28] C. Travier, S. Chel, M. Desmons, G. Devanz, P. Lepercq, T. Garvey, "Design and Test of a 1.3 GHz Travelling Wave Window", 9th Workshop on RF Superconductivity, p. 427

[29] J. Lorkiewicz, B. Dwersteg, A. Brinkmann, .W. -D. Moeller, D. Kostin, M. Layalan, "Surface TiN Coating of TESLA Couplers at DESY as an Antimultipactor Remedy", Proceedings Of the 10th Workshop On RF Superconductivity Tsukuba Japan September,6-11 2001

[30] W. -D. Moeller, "High-Power Coupler for TESLA Test Facility", 9th Workshop on RF Superconductivity, p. 577.

[31] V. Nguyen, H. L. Phillips, and J. Preble, "Development of a 50 kW CW L-band rectangular Window for the Jefferson Lab FEL Cryomodule", PAC99, p. 1459

[32] E. Chojnacki, T. Hays, J. Kirchgessner, H. Padamsee, M. Cole and T. Schultheiss, "Design of ea High Average power Waveguide Window", PAC97, p. 3177.

[33] E. Chojnacki, P. Barnes, S. Belomestnykh, R. Kaplan, J. Kirchgessner, H. Padamsee, P. Quigley, J. Reilly, and J. Sears, "Tests and Designs of High-Power Waveguide Vacuum Windows at Cornell", 8th Workshop on RF Superconductivity, p. 753

[34] F. Krawczyk J. Gioia, B. Haynes, R. Lujan, B. Rusnak, B. Smith, "The Power Coupler Design for the APT Superconducting Accelerator," 8th Workshop on RF Superconductivity, p. 762.

[35] I. E. Campisi, E. F. Daly, J. E. Henry, P. Kneisel, W. J. Schneider, M. Stirbet, K. M. Wilson, "The Fundamental Power Coupler Prototype for the Spallation Neutron Source (SNS) Superconducting Cavities," PAC 2001, Chicago, USA, June 18-22, 2001.

[36] M. Champion, "Input Coupler and Windows for TESLA", 6th Workshop on RF Superconductivity, p. 406

[37] H.P. Kindermann, E. Haebel, M. Stirbet, V. Veshcherevich, "Status of RF Power Couplers for Superconducting Cavities at CERN", EPAC 96, p. 2091

[38] M. Champion, D. Peterson, T. Peterson, C. Reid, M. Ruschman, "TESLA Input Coupler Development", PAC 93, p. 809.

[40] X. Hanus, A. Mosnier, "Coaxial TW window for power couplers multipactor considerations", 7th Workshop on RF Superconductivity, p. 701

[41] S. Chel, M. Desmons, C. Dupery, X. Hanus, A. Mosnier, G. Bienvenu, J. C. Bourdon, T. Garvey, J. Le Duff, J. Marini, G. Mace, R. Panvier, N. Solyak; A. le Goff, S. Maïssa, T. Junquera, "Power coupler development for superconducting cavities", EPAC96, p. 2088

[42] E. N. Schmierer, K. C. D. Chan, R. C. Gentzlinger, W. B. Haynes, F. L. Krawczyk, D. I. Montoya, P. L. Roybal, D. L. Schrage, T. Tajima, "Design of the ADTF Spoke Cavity ED&D Input Coupler", Proceedings Of the 10th Workshop On RF Superconductivity Tsukuba Japan September,6-11 2001

[43] G. Spalek and J. Kuzminski" High Power Variable Couplers for Ladder and Spoke Type Resonators", PAC 2001, Chicago, USA, June 18-22, 2001.

[44] E. Schmierer, R. E. Lujan, B. Rusnak, B.G. Smith W. B. Haynes, D. C. Gautier, J. Waynert, F. L. Krawczyk, J. G. Gioia, "Development of the SCRF Power Coupler for the APT Accelerator", PAC99, p. 977

[45] G. Devanz, H. Safa, C. Travier, **"**Preliminary Design of a 704 MHz Power Coupler for a High Intensity Linear Accelerator", EPAC 2000, p. 2031

[46] Y. Kang, S. Kim, M. Doleans, I. E. Campisi, M. Stirbet, P. Kneisel, G. Ciovati, G. .Wu, P. Ylä-Oijala, "Electromagnetic Simulations and Properties of the Fundamental Power Couplers for the SNS Superconducting Cavities," PAC 2001, Chicago, USA, June 18-22, 2001

[47] P. Balleyguier, "A straightforward method for cavity external Q computation", Particle Accelerators, Vol. 57, p 113, 1997

[48] P. Balleyguier, "External Q Studies for APT SC-Cavity Couplers", Linac98, p. 133 [49] F. L. Krawczyk, " Status of Multipacting Simulation Capabilities for SCRF

[50] G. Devanz, “Multipactor simulations in superconducting cavities and power couplers”, Phys. Rev. Special Topics- Accelerators and Beams, Vol. 4, 012001 (2001)

[51] G. Devanz, “A 2D Multipactor Simulation Code for RF Components and Accelerating Cavities”, EPAC 2000

[52] E. Somersalo, P. Ylä-Oijala, D. Proch, “Electron multipacting in RF structures”, DESY Print TESLA 94-‘4, July 94

[53] P. Ylä-Oijala, “Analysis of Electron Multipacting in Coaxial Lines with Traveling and Mixed Waves”, TESLA Report 97-20, 1997

[54] P. Ylä-Oijala and M. Ukkola, “Multipacting Simulations on the Coaxial SNS Coupler”, University of Helsinki Report, November 2000

[55] P. Ylä-Oijala and D. Proch, “MultiPac-Multipacting simulation package with 2D FEM Field Solver”, Proceedings Of the 10th Workshop On RF Superconductivity Tsukuba Japan September,6-11 2001

[56] P. Ylä-Oijala, “Suppressing Electron multipacting in Coaxial Lines by DC Voltage”, TESLA Report 97-21, 1997

[57] R. Bourque and G. Laughon, ”Thermal Analysis of a Refined APT Power Coupler and Thermal Shield and the Effect on the Cryoplant”, PAC 2001, Chicago, USA, June 18-22, 2001

[58] J. A. Waynert and F. C. Prenger, “A thermal analysis and Optimization of the APT 210 kW Power Coupler”, Linac98, p. 950.

[59] K.Halbach and R.F.Holsinger,Particle Accelerator 7, 213 (1976)

[60] T.Weiland, Nuclear Instruments and Methods 216, 329 (1983).

[61] P.Fernandes, R.Parodi

[62] “Theory of diffraction by small holes”, H.A.Bethe, Phys. Rev.,1944.

[63] HFSS, 3D High-Frequency electromagnetic simulation software, is a registered Trade Mark of ANSOFT corporation, Four Station Square, Suite 200,Pittsburgh, PA,15219-1119 USA.

[64] T.Weiland et al., DESY report M-86-07, 1986

[65] “Experimental investigation of low frequency modes of a single cell rf cavity”, Fornaca et al.,IEEE Part. Acc. Conf., Washington, 1987

[66] “First results on HOM measurements with a 5-cell copper cavity, A.Mosnier et al., DPHN/AL-87-295, Saclay (1987).

[67] “On the calculation of beam tube propagation of cavity modes from Superfish or Urmel”, E.Haebe1, P.Marchand and I.Tuckmantel, CERN/EF/RF 84-1.

[68] “Status of Tristan Superconducting RF Program”, S.Noguchi et al, 3rd Workshop on RF superconductivity, Argonne, 1987

[69] “Coupler developments at CERN”,E.Haebel et al, 3rd Workshop on RF superconductivity, Argonne,1987

[70] “Higher Order Modes Coupler Studies at DESY”, E.Haebel, I.Sekutowicz, internal report, DESY M-86-06

[71] “Development of HOM couplers at Cornell”,3rd Workshop on RF superconductivity ,Argonne, 1987

[72] “Higher Order Mode RF Power Extraction From Polarized Cavities with a Single Output Coupler”,I.Kirchgessner et al., Part. Acc. Conf., Chicago,1989

A SURVEY OF BASIC THERMODYNAMICS

W. F. Vinen
School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK

Abstract
A brief summary is given of the thermodynamics that underlies the use of superconducting and cryogenic techniques in particle accelerators and detectors. It is intended to remind participants at the School of basic principles, and it is not an introduction to the subject. A rigorous treatment is not possible within the limited space available.

1. INTRODUCTION: MACROSTATES AND MICROSTATES

Thermodynamics is concerned with macroscopic systems, with the concept of temperature, and with the way in which the behaviour of such systems is constrained by very general principles, called the Laws of Thermodynamics, which originate from a recognition that the random motion of particles in the system is governed by general statistical principles. It describes macroscopic systems by macroscopic parameters, such as pressure, p, volume, V, magnetization, $\tilde{M}$, applied magnetic field, $\tilde{H}$, and total internal energy, E, all of which have obvious meanings, and it introduces new and more subtle macroscopic parameters, such the temperature, T, and entropy, S. The constraints are expressed in terms of allowed relationships between these parameters. The so-called extensive quantities, $V, \tilde{M}, E$ and S, are proportional to the mass of material present; we shall generally assume that each relates to one mole.

The various macroscopic parameters define a **macrostate** of the system. We can also think about a microscopic state, or **microstate**, of the system, defined, for example, by specifying the energy or position of each individual atom or molecule. It is crucial to understand that a very large number of different microstates correspond to the same macrostate. For example, the same total internal energy, E, can come about from a very large number of different ways of distributing this energy among the individual atoms or molecules. We denote this large number by W. Other things being equal, the probability of the system being found in one microstate is the same as that of finding it in another microstate. Thus the probability that the system is in a given macrostate must be proportional to the corresponding number, W, of microstates. The number of microstates, W, is often called the *statistical weight*.

2. INTERNAL ENERGY AND THE FIRST LAW OF THERMODYNAMICS

In essence the **First Law** states that, if the macroscopic parameters such as pressure, volume, magnetization and applied magnetic field are specified, then so also is the internal energy, E. For a non-magnetic fluid system, the important parameters are usually p and V. Then the First Law states that E is a properly defined function of p and V: $E = E(p, V)$. If you take a fluid system through some cycle and return it to its original p and V, its internal energy returns to its original value. We say that E is a **function of state**. Obvious? No, it depends on the fact that the individual molecules are very quickly randomized among all the **accessible microstates**, the accessible microstates being those that are consistent with the given p and V. The previous history is irrelevant.

3. ENTROPY AND THE SECOND LAW OF THERMODYNAMICS

The number of microstates, W, is a very useful parameter because it relates to the probability of finding a given macrostate. But W is usually an inconveniently large number, so it is better to work in terms of $\ln W$. We define the **entropy** as $S = k_B \ln W$, where k_B is a constant, the value of which will be fixed

later. Note that if we have two systems, A and B, the number of microstates of the combined system is $W_A W_B$, so that the combined entropy is additive; it is similarly proportional to the mass of the system.

The **Second Law of Thermodynamics** states in part that S is a function of state: $S = S(p, V)$. Again, not obvious, but again depends on the fact that a system quickly randomizes over the accessible microstates (consistent with the given p and V), independently of the previous history.

4. THE PRINCIPLE OF INCREASE IN ENTROPY

This Principle completes our statement of the Second Law.

Suppose that a system is isolated from its surroundings, so that any change in the system can have no effect on the surroundings and *vice versa*. Then, if the system can be imagined to exist in two macroscopic states, corresponding to two different values of W, it is more likely to be found in the one with the larger W. In practice, one value of W is usually enormously larger than the other, so that the system will almost certainly be found in the one with the larger W. In other words, in the one with the larger entropy. For example, consider a gas in a box. Suppose that at some initial time all the gas were in the left hand side of the box, corresponding to $W = W_1$. If the gas were allowed to fill the whole box, $W = W_2$. It turns out that $W_2 >>>> W_1$. So the gas must expand to fill the whole box, in the sense that it must have only a negligible probability of being found in one side of the box only. In this sense **entropy tends to a maximum value**.

Note that the principle that a system tends to maximize its entropy applies only to an isolated system. Otherwise one would need to worry about changes in W or S for the surroundings. Often systems are not isolated, and therefore in due course we shall look for principles that are more convenient to use than the Principle of Increase in Entropy.

5. THE KELVIN TEMPERATURE

We define the **Kelvin temperature** by the relation: $T = (\partial E/\partial S)_V$. Why? And does this definition accord with our everyday experience? Suppose that we have two systems, A and B, between which energy can flow, while each is maintained at constant volume (this flow of energy is actually a flow of "heat"). The two systems are not isolated from each other, and therefore we can apply the principle of increase in entropy only to the combined system. If we transfer energy δE from one system to the other we have

$$\delta S = \left(\frac{\partial S_A}{\partial E_A}\right)_V \delta E - \left(\frac{\partial S_B}{\partial E_B}\right)_V \delta E. \tag{1}$$

(Remember that S is additive.) Therefore maximizing the entropy means that $T_A = T_B$. Therefore equilibrium requires equality of Kelvin temperature, in accord with our experience. (Note that we have in fact proved that the concept of temperature is meaningful; this is related to what is sometimes called the **Zeroth Law of Thermodynamics**.) If initially $T_A > T_B$, (heat) energy must flow from A to B (hot to cold), if S is to increase, which again accords with our expectations.

6. THE BOLTZMANN DISTRIBUTION, AND THE MEANING OF $T = 0$

Imagine a volume of gas, at constant volume, V, and constant internal energy, E_0. Focus on one particular molecule of the gas. If its energy is ϵ, the rest of the gas has energy $E_0 - \epsilon$. Therefore the probability that the molecule has this energy is the probability that the rest of the system has energy $E_0 - \epsilon$. We might try writing for this probability

$$W(E_0 - \epsilon) = W(E_0) - \epsilon \frac{\partial W}{\partial E}, \tag{2}$$

but W varies too rapidly with E to allow us to use this formula (Taylor expansion). Therefore work with $\ln W$, instead of W, and hence write

$$\ln W(E_0 - \epsilon) = \ln W(E_0) - \epsilon \left[\frac{\partial}{\partial E} \ln W \right]_{\epsilon=0} = \ln W(E_0) - \frac{\epsilon}{k_B T}. \tag{3}$$

Therefore the probability that a molecule has energy ϵ is proportional to $\exp(-\epsilon/k_B T)$. Note the meaning of $T = 0$. At this temperature no molecule can be in an excited state, and therefore the whole gas must be in its state of minimum energy.

7. THE EQUATION OF STATE, AND THE RELATIONSHIP BETWEEN THE KELVIN TEMPERATURE AND THE IDEAL GAS TEMPERATURE

The relation between p, V and T is called the **equation of state**. Given the Boltzmann distribution we can calculate the equation of state for an ideal gas. We find $pV = N_A k_B T$, where N_A is Avogadro's number. Therefore the Kelvin temperature is the same as the ideal gas temperature if we take $k_B = R/N_A$, where R is the gas constant. k_B is then called *Boltzmann's constant* (1.38×10^{-23} JK^{-1}).

8. WORK AND HEAT

Energy can be added to a system in the form of either **work** or **heat**. For example, work is done on a gas if it is compressed in a cylinder fitted with a piston. Heat can be added to the system by thermal conduction through the walls of the containing vessel. The total increase in energy is the sum of the heat added and the work done.

$$dE = \overline{d}q + \overline{d}w. \tag{4}$$

Note the convention of signs: energy added to a system is positive. q and w are not functions of state (you cannot associate a particular amount of heat with a particular state), and therefore $\overline{d}q$ and $\overline{d}w$ are not proper differentials (hence the bars over the ds).

9. REVERSIBLE PROCESSES

If the piston compresses the gas slowly and if there is no friction, the process is said to be **reversible**, because a small change in conditions (force on the piston) can reverse the direction of the process. The work done can then be written

$$\overline{d}w_{rev} = -p dV. \tag{5}$$

If the compression is not slow the gas swirls around and the pressure is not defined. If the gas expands into a vacuum through a valve (an extremely irreversible process), no work is done. Heat can also flow reversibly, if the temperature difference is very small. Generally we can write for a reversible process

$$dE = \overline{d}q_{rev} - p dV. \tag{6}$$

10. ENTROPY AND REVERSIBLE HEAT

Since E and S are functions of state, we can regard E as a function of S and V: $E = E(S, V)$. Therefore we can write down the following mathematical relation

$$dE = \left(\frac{\partial E}{\partial S} \right)_V dS + \left(\frac{\partial E}{\partial V} \right)_S dV; \tag{7}$$

i.e.

$$dE = TdS + \left(\frac{\partial E}{\partial V}\right)_S dV. \tag{8}$$

Arguments can be advanced that allow us to identify the two terms on the right hand side of this equation with those in Eq. (6), so that we can write

$$dS = \frac{\overline{d}q_{rev}}{T}; \qquad p = -\left(\frac{\partial E}{\partial V}\right)_S. \tag{9}$$

Changes in entropy can therefore be calculated from thermal data with the equation

$$\Delta S = \int \frac{\overline{d}q_{rev}}{T}. \tag{10}$$

11. A USEFUL EQUATION IN THERMODYNAMICS

From Eq. (8) we obtain the following useful relation involving only functions of state

$$dE = TdS - pdV. \tag{11}$$

12. ADIABATIC PROCESSES

A reversible change in which there is no heat flow is called an **adiabatic process**. Clearly, it occurs at constant entropy.

13. THE THIRD LAW OF THERMODYNAMICS

At the absolute zero of temperature any system in thermal equilibrium must exist in its lowest possible energy state; i.e. in a single unique quantum state, with $W = 1$. Therefore $S \to 0$ as $T \to 0$. This allows the calculation of *absolute* entropies from thermal data. The Law imposes important restrictions on thermodynamic behaviour near $T = 0$.

14. HEAT CAPACITIES

The amount of heat that must be added to a system reversibly to change its temperature defines its **heat capacity** or **specific heat**: $C = \overline{d}q_{rev}/dT$. The conditions under which the heat is supplied must be specified; usually constant pressure or constant volume, giving C_p or C_V. Therefore the entropy of a substance at temperature T and pressure p is given by

$$S(T, p) = \int_0^T \frac{C_p}{T} dT. \tag{12}$$

Note the following useful formulae.

$$C_p = T\left(\frac{\partial S}{\partial T}\right)_p; \qquad C_V = T\left(\frac{\partial S}{\partial T}\right)_V. \tag{13}$$

It can be proved that

$$\gamma \equiv \frac{C_p}{C_V} = \left(\frac{\partial V}{\partial p}\right)_T \left(\frac{\partial V}{\partial p}\right)_S^{-1}; \tag{14}$$

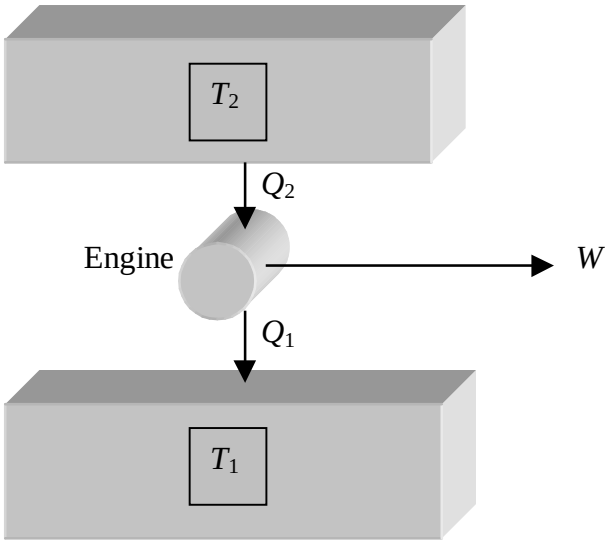


Fig. 1: Schematic heat engine operating between two fixed temperatures

i.e. that γ is the ratio of the isothermal to the adiabatic compressibilities, and that

$$C_p - C_V = T\left(\frac{\partial p}{\partial T}\right)_V \left(\frac{\partial V}{\partial T}\right)_p, \tag{15}$$

where the right hand side can be evaluated from the equation of state. For an ideal gas, Eq. (15) reduces to

$$C_p - C_V = R. \tag{16}$$

The proof of Eq. (15) is an interesting exercise in the use of the Maxwell thermodynamic relations (see section 16).

15. HEAT ENGINES, REFRIGERATORS AND HEAT PUMPS

A heat engine is a cyclic device that can continuously produce work from heat.

Consider two thermal reservoirs (systems of large heat capacity) held at fixed temperatures T_2 and T_1, where $T_2 > T_1$ (Fig. 1). The cylindrical box is an engine, working in a cycle. For each cycle, it extracts heat Q_2 from the high-temperature reservoir, does work W, and puts heat Q_1 into the low-temperature reservoir.

After each cycle the entropy of the two reservoirs cannot have decreased. Therefore

$$\frac{Q_1}{T_1} \geq \frac{Q_2}{T_2}. \tag{17}$$

At the same time conservation of energy (no friction) requires that $W = Q_2 - Q_1$. Therefore the efficiency, defined as $\eta = W/Q_2$, has a maximum value given by

$$\eta_{max} = 1 - \frac{T_1}{T_2}. \tag{18}$$

This maximum value obtains when the engine operates reversibly, with no frictional dissipation.

An engine that operates reversibly in this way between two fixed temperatures is called a **Carnot engine**. Cycles for a Carnot engine with an ideal gas as a working substance are shown in Fig.2.

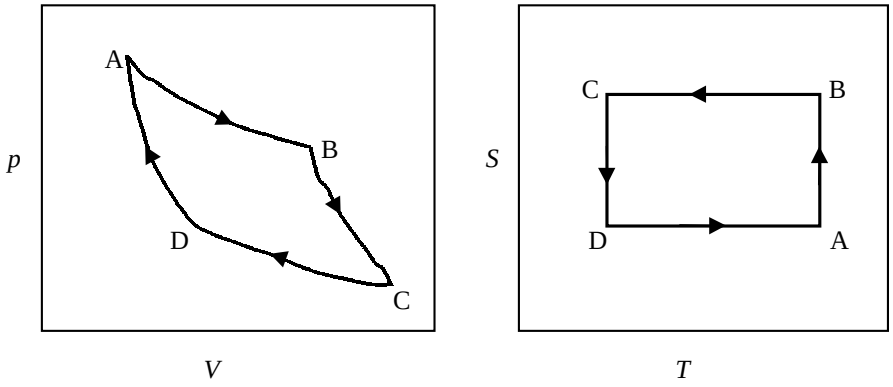


Fig. 2: Carnot cycles

The elements of the cycle are as follows.

A→B	Isothermal expansion of gas, removing heat Q_2 from reservoir at T_2. Entropy of working substance increases.
B→C	Adiabatic expansion, isolated from reservoirs; temperature falling to T_1.
C→D	Isothermal compression, transmitting heat Q_1 to reservoir at T_1. Entropy of working substance decreases.
D→A	Adiabatic compression, isolated from reservoirs, temperature rising to T_2.

Note that the work done, W, is the area within the $p - V$ cycle.

If the engine operates backwards, it becomes a **refrigerator**. Work is put in with an electric motor, and heat is transferred from the reservoir at the low temperature T_1 to that at the high temperature T_2. The reservoir at T_1 is refrigerated. The coefficient of performance is given by

$$\frac{Q_1}{W} = \frac{T_1}{T_2 - T_1}. \tag{19}$$

An engine working backwards can also act as a heating device, called a **heat pump**. The reservoir at T_2 is a house; that at T_1 is a neighbouring river. The work W derived from an electric motor causes heat Q_2 to be discharged into the house, where

$$\frac{Q_2}{W} = \frac{T_2}{T_2 - T_1}. \tag{20}$$

Note that Q_2 can be much larger than W, so that the heating is much more effective (in theory) than can be obtained by dissipating the energy W in a resistive heater. The river is cooled.

16. ENTHALPY, (HELMHOLTZ) FREE ENERGY, AND GIBBS FREE ENERGY

It is convenient to introduce three other functions of state.

Enthalpy:

$$H = E + pV \tag{21}$$

(Helmholtz) free energy:

$$F = E - TS \tag{22}$$

Gibbs free energy:

$$G = E - TS + pV \tag{23}$$

Using Eq. (11), i.e.

$$dE = TdS - pdV \tag{24}$$

we find that

$$dH = TdS + Vdp; \tag{25}$$

$$dF = -pdV - SdT; \tag{26}$$

$$dG = -SdT + Vdp. \tag{27}$$

If Z, P and Q are functions of x and y, and if

$$dZ = Pdx + Qdy, \tag{28}$$

then we have the mathematical result

$$\left(\frac{\partial P}{\partial y}\right)_x = \left(\frac{\partial Q}{\partial x}\right)_y. \tag{29}$$

Applying this result to dE, dH, dF and dG, we find

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial p}{\partial S}\right)_V; \quad \left(\frac{\partial T}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p; \quad \left(\frac{\partial p}{\partial T}\right)_V = \left(\frac{\partial S}{\partial V}\right)_T; \quad \left(\frac{\partial S}{\partial p}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_p \tag{30}$$

These very useful relations are called the **Maxwell Thermodynamic Relations**. Note that they depend on E and S being functions of state, so that they are direct consequencies of the Laws of Thermodynamics. Often they are the form in which these laws are used. For example, suppose we compress a gas isothermally and ask by how much the entropy changes. We need to know $(\partial S/\partial p)_T$. The 4th Maxwell relation tells us that this derivative is equal to $-(\partial V/\partial T)_p$, which can be calculated from the equation of state. For an ideal gas it is simply equal to $-R/p$, so that $\Delta S = -R\ln(p_{final}/p_{initial})$.

We note for future reference that

$$C_p = \left(\frac{\partial H}{\partial T}\right)_p, \tag{31}$$

as is easily proved from Eqs. (11), (13) and (21).

17. DIRECTIONS OF SPONTANEOUS CHANGE

We have seen that for a system that is thermally isolated a spontaneous change must lead to an increase in entropy, so that in equilibrium the entropy must be a maximum. In practice physical systems are not usually thermally isolated. We show now how to use the free energies F and G to formulate more useful equilibrium conditions.

Consider a physical system (S) that is in continuous thermal contact with its surroundings (R), which we regard as a system with a very large heat capacity at a constant temperature T_0. The system S

is held at constant volume. We can imagine that the combined system R+S is thermally isolated, so that we can say that for this combined system the entropy must increase in any spontaneous process; i.e.

$$\delta S = \delta S_S + \delta S_R \geq 0. \tag{32}$$

During this process heat δQ may flow from R to S. Because R has a very large thermal capacity its temperature remains essentially constant, and therefore $\delta S_R = -\delta Q/T_0$. But the flow of heat into S increases the internal energy of S by an amount $\delta E_S = \delta Q$ (remember that S is held at constant volume). Therefore Eq. (32) becomes

$$\delta \left(E_S - T_0 S_S \right) \leq 0. \tag{33}$$

The spontaneous change must therefore be accompanied by a reduction in $E_S - T_0 S_S$. This quantity must therefore tend to a minimum in equilibrium. But in equilibrium the temperature T_S of the system S must be equal to the temperature T_0 of the surroundings. **Therefore in equilibrium the free energy $F_S = E_S - T_S S_S$ of the system must tend to a minimum**.

Another physically interesting situation is one in which the system S is held in continous thermal contact with its surroundings and at the same time is held at constant pressure. A straightforward modification of the above argument then leads to the result that the **Gibbs free energy must tend to a minimum**.

18. CHANGES IN FREE ENERGY AND MAXIMUM AVAILABLE WORK

Suppose a system undergoes a reversible change at constant temperature T_S. The work done *by the system* is then given by

$$-\delta W = -\delta E_S + T_S \delta S_S = -\delta F_S; \tag{34}$$

i.e the work done is equal to the reduction in free energy. This work done, achieved under reversible conditions, is the maximum work that can be extracted from the system for the given change. Hence the name “free energy”.

A similar argument shows that the maximum amount of work, other than that related to a pushing back of the surroundings against a constant pressure, is equal to the reduction of Gibbs free energy. For example, the maximum electrical work that can be extracted from a battery is equal to the reduction in Gibbs free energy of the contents of the battery.

19. PHASE EQUILIBRIA

Substances can exist in the form of different phases. For a simple pure fluid there can be solid, liquid and gas phases, depending usually on the pressure and temperature, in a way that is summarized in a **phase diagram**, such as that shown schematically in Fig. 3.

Along each solid line two phases can coexist in equilibrium. At the triple point T all three phases can coexist. To the right of the critical point C there is no distinction between the liquid and gas phases.

Suppose that p and T are on the line separating the solid and gas phases. The solid and gas phases can then coexist in equilibrium, as shown in Fig. 4, where the gas + solid are maintained at constant temperature and constant pressure.

l The equilibrium condition is that $G \rightarrow$minimum. Let us transfer a small number δN of moles from the solid to the gas phase. The change in total Gibbs free energy is then $(G_{gas} - G_{solid})\delta N$, where the Gs refer to one mole. Therefore in equilibrium we must have

$$G_{gas} = G_{solid}. \tag{35}$$

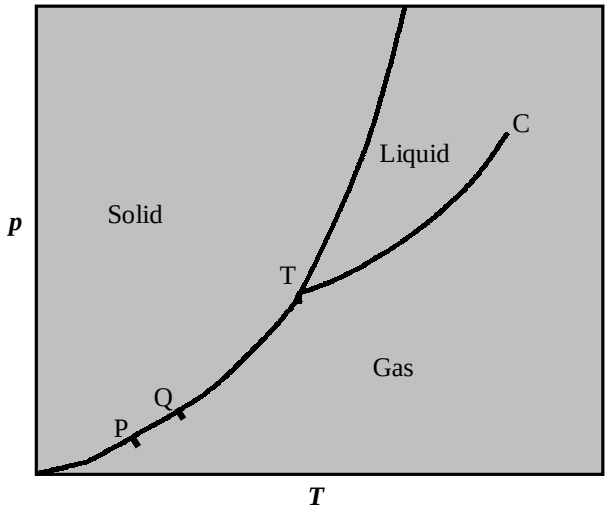


Fig. 3: Schematic phase diagram of a typical pure substance

If we plot G for each phase in a three-dimensional plot as functions of p and T, we get three surfaces. The lines in the phase diagram can then be regarded as projections of the intersections of these three surfaces on to the $p-T$ plane. (*Exercise for the student*: how do these surfaces look near the triple point and near the critical point?)

Let P and Q be neighbouring points on the line separating the solid and gas phases in the phase diagram (see Fig. 3). Then in an obvious notation

$$G_{gas}(Q) = G_{gas}(P) + \left(\frac{\partial G_{gas}}{\partial p}\right)_T dp + \left(\frac{\partial G_{gas}}{\partial T}\right)_p dT = G_{gas}(P) + V_{gas}dp - S_{gas}dT, \qquad (36)$$

where all quantities refer to one mole. Therefore the slope of the $p-T$ line on the phase diagram is given by

$$\frac{dp}{dT} = \frac{S_{gas} - S_{solid}}{V_{gas} - V_{solid}} = \frac{L}{T(V_{gas} - V_{solid})}, \qquad (37)$$

where L is the molar latent heat for the solid→gas transition. (L is the heat required to convert one mole from one phase to another, at constant temperature.) This is the **Clausius-Clapeyron equation**. It applies to a **first-order phase transition**, which is characterized by a latent heat (discontinuity in the entropy). There exist also **higher order transitions**, which will be mentioned in the next section.

20. THE THERMODYNAMICS OF MAGNETIC MATERIALS

Thermodyanmics can be valuable in discussing the behaviour of magnetic materials, including superconductors. Consider a piece of magnetic material placed inside a coil with zero electrical resistance (we assume for simplicity that the material has a shape such that its demagnetizing coefficient is zero). The material can be magnetized by slowly establishing a current $\tilde{I}$ in the coil. While the current is being established there is a back-emf $\tilde{E}$ in the coil, equal to the time rate of change of magnetic flux linkage, so that energy equal to $\int \tilde{I}\tilde{E}dt$ is fed into the system by the source of current. It is easy to show that this energy is equal to $\int \tilde{H}d\tilde{M}$, where $\tilde{H}$ is the magnetic field applied to the magnetic material and $\tilde{M}$ is its magnetic moment (we have subtracted the magnetic field energy in the absence of the material). We can regard this energy as reversible work done on the magnetic material, and therefore we can write

$$\overline{d}w_{rev} = \tilde{H}d\tilde{M}, \qquad (38)$$

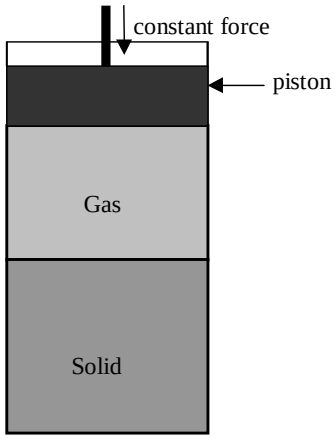


Fig. 4: A solid phase in equilibrium with the gas phase at constant p and T

where from now on we shall regard $\tilde{M}$ as relating to one mole of the material. Eq. (38) is analogous to the relation $\overline{d}w_{rev} = -pdV$ for work done in compressing a fluid, if we take $\tilde{H} \equiv -p$ and $\tilde{M} \equiv V$.

Let us apply this idea to the magnetization of a superconductor. As long as the superconductor remains in the Meissner state, the magnetic induction ($\tilde{B} = \mu_0\tilde{H} + \tilde{M}$) within it is always zero. Therefore $\tilde{M} = -\mu_0\tilde{H}$, and $\overline{d}w_{rev} = -\mu_0\tilde{H}d\tilde{H}$.

Application of a sufficiently large magnetic field to a type I superconductor leads to a phase transition to the normal state. We can treat this phase transition thermodynamically, by analogy with the treatment of the gas-solid transition in section 19. The phase diagram is shown in Fig.5 (a). Along the phase boundary the equilibrium condition (constant $\tilde{H}$ and T) is that the magnetic Gibbs free energy, $\tilde{G} = E - TS - \tilde{H}\tilde{M}$, be equal in the two phases (cf Eq. (35)). Note that $d\tilde{G} = -SdT - \tilde{M}d\tilde{H}$ (in analogy with $dG = -SdT + Vdp$). Therefore the magnetic Gibbs free energy in the Meissner phase can be written

$$\tilde{G}_s(\tilde{H}, T) = \tilde{G}_s(0, T) - \int_0^{\tilde{H}} \tilde{M}d\tilde{H} = \tilde{G}_s(0, T) + \frac{1}{2}\mu_0\tilde{H}^2. \tag{39}$$

The Gibbs free energy , $\tilde{G}_n(T)$, of the normal metal, which is almost non-magnetic, is practically independent of $\tilde{H}$. Therefore the critical field, $\tilde{H}_c$, of the superconductor is given by (Fig. 5(b))

$$\frac{1}{2}\mu_0\tilde{H}_c^2 = \tilde{G}_n(T) - \tilde{G}_s(0, T). \tag{40}$$

Note that this analysis applies only to type I superconductors; in type II superconductors a "mixed" state formed by adding quantized flux lines has the smallest Gibbs free energy over a range of fields between $\tilde{H}_{c1}$ and $\tilde{H}_{c2}$ ($\tilde{H}_{c1} < \tilde{H}_c < \tilde{H}_{c2}$).

The analogue of the Clausius-Clapeyron equation can be written down.

$$\frac{d\tilde{H}_c}{dT} = \frac{L}{T\tilde{M}_c} = -\frac{L}{\mu_0 T\tilde{H}_c}. \tag{41}$$

As we see from Fig.5, $d\tilde{H}_c/dT$ tends to a finite value as $\tilde{H}_c \to 0$. It follows from the Clausius-Clapeyron equation that $L \to 0$ as $\tilde{H}_c \to 0$; in other words the superconducting to normal transition has associated with it no latent heat in zero applied magnetic field. This makes it different in character from

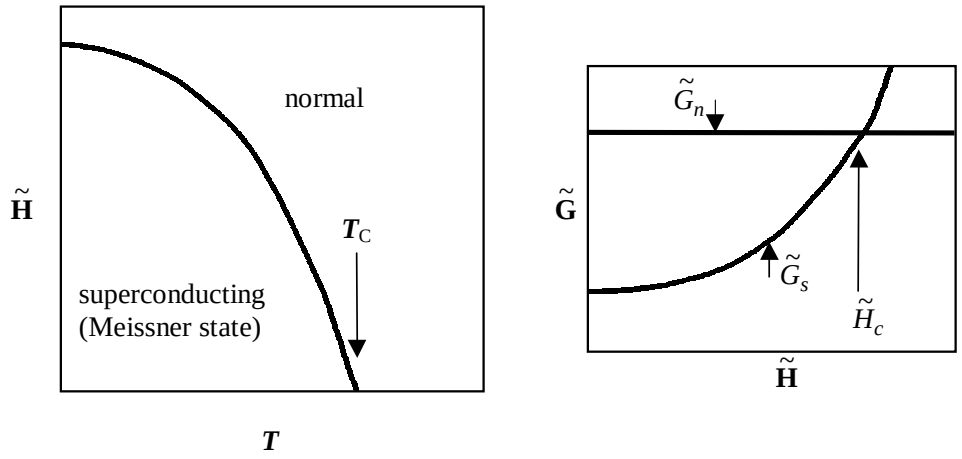


Fig. 5: (a) The phase diagram of a type I superconductor; (b) the variation of magnetic Gibbs free energy as a magnetic fi eld is applied to a type I superconductor.

more usual phase transitions. It is what we call a **second order transition**. A normal, first-order, phase transition is characterized by a discontinuity in the **first** derivatives of the Gibbs free energy, such as $S = -(\partial G/\partial T)_p$. A second-order phase transition has no discontinuity in these derivatives, but does have discontinuities in the second derivatives, such as the heat capacity $T(\partial S/\partial T)_p = -T(\partial^2 G/\partial T^2)_p$. Analogues of the Clausius-Clapeyron equation can be derived for these higher-order transitions. Unfortunately, genuinely high-order transitions are rare, since the higher derivatives of the Gibbs free energy usually exhibit divergences at the transition temperature, instead of simple discontinuities. This is the case for the superfluid phase transition in liquid helium, discussed in another lecture, for which $C_p \to \infty$ at the transition.

21. REFRIGERATION

Here we summarize very briefly the thermodynamics underlying various methods of low-temperature refrigeration.

21.1 Abiabatic expansion of a gas

The gas in compressed at room temperature, thereby reducing its entropy (Fig. 2), and then allowed to expand adiabatically, and as far as possible reversibly, through a suitable reciprocating engine or turbine. In essence we have a Carnot cycle. The cooling during adiabatic expansion is described by the partial derivative $(\partial T/\partial p)_S$, which can be evaluated in terms of the heat capacity and the equation of state of the gas as follows.

$$\left(\frac{\partial T}{\partial p}\right)_S = -\left(\frac{\partial T}{\partial S}\right)_p \left(\frac{\partial S}{\partial p}\right)_T = \frac{T}{C_p}\left(\frac{\partial V}{\partial T}\right)_p , \tag{42}$$

where we have made use of the mathematical identity $(\partial z/\partial y)_x = -(\partial z/\partial x)_y(\partial x/\partial y)_z$ and one of the Maxwell relations (Eq. (30)). Adiabatic expansion always leads to a cooling. In order to obtain a large fall in temperature the expanded gas (or some of it) is used to cool the incoming high-pressure gas in a **heat exchanger**; a cascade of engines may also be used. The design of efficient heat exchangers is often crucial.

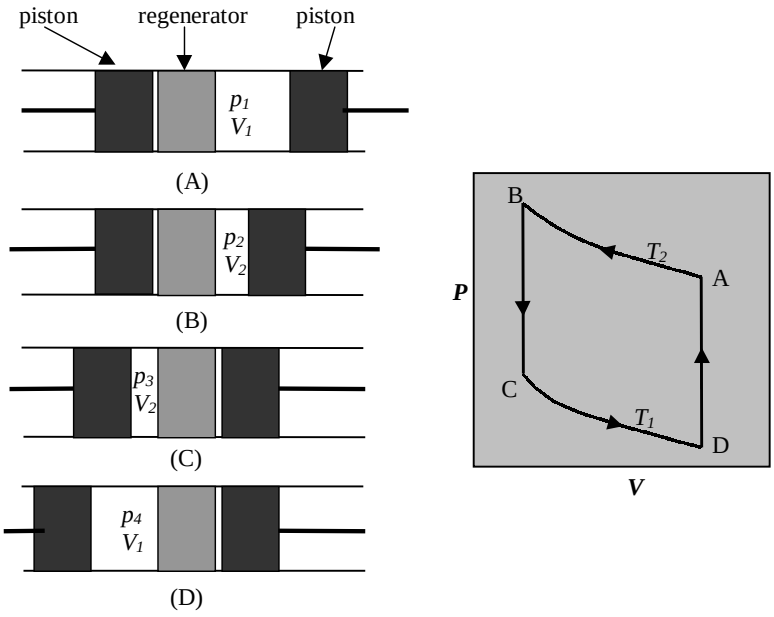


Fig. 6: The Stirling cycle

21.2 The Stirling cycle refrigerator

Air liquefiers are commonly based not on what is essentially a Carnot cycle, but rather on a Stirling cycle, in which the adiabatic expansion and compression are replaced by cooling or heating at constant volume with a **regenerator**, as shown schematically in Fig. 6. The regenerator allows gas to pass through it. It has a large heat capacity and large surface area, but ideally does not conduct heat along its length. The temperature at its right hand side is maintained at a "high" temperature T_2 (room temperature) and that at its left hand side at a low temperature T_1. A gas passing through it, from say right to left, is therefore cooled from temperature T_2 to temperature T_1.

The cycle of operation is as follows.

A→B	Gas to the right of the regenerator is compressed isothermally at T_2 by moving the right hand piston. The entropy of the gas is reduced. Heat is removed through a heat exchanger.
B→C	Both pistons move together, transferring gas through the regenerator, at constant volume. The gas cools to temperature T_1, again reducing its entropy (the regenerator warms slightly).
C→D	The left hand piston moves to the left, causing isothermal expansion of the gas. The entropy of the gas increases, and heat is removed through a heat exchanger from the system that is being cooled (eg air being liquefied).
D→A	Both pistons move to the right. The temperature of the gas rises to T_2 as it passes through the regenerator.

Ideally the cycle can be reversible (with Carnot efficiency), but the extent to which it is really reversible depends very much on the effectiveness of the regenerator.

21.3 Joule-Kelvin (Joule-Thomson) expansion

The gas is allowed to expand through a throttling valve from a fixed high pressure to a fixed low pressure, the whole system being thermally isolated (Fig. 7).

Suppose that one mole of gas, volume V_1, pressure p_1 passes through the value, emerging as a

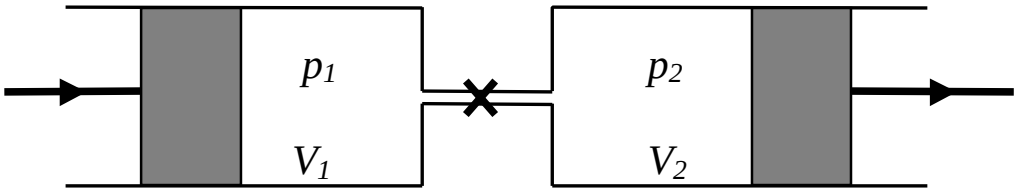


Fig. 7: Joule-Kelvin expansion

volume V_2, at pressure p_2. Owing to the thermal isolation we can write

$$E_1 + p_1 V_1 = E_2 + p_2 V_2, \tag{43}$$

where E_1 and E_2 are the molar internal energies before and after expansion. Therefore the initial and final enthalpies are equal

$$H_1 = H_2. \tag{44}$$

The temperature change can therefore be obtained either by integrating

$$\left(\frac{\partial T}{\partial p}\right)_H = -\left(\frac{\partial T}{\partial H}\right)_p \left(\frac{\partial H}{\partial p}\right)_T = -\frac{1}{C_p}\left[T\left(\frac{\partial S}{\partial p}\right)_T + V\right] = \frac{1}{C_p}\left[T\left(\frac{\partial V}{\partial T}\right)_p - V\right] \tag{45}$$

or from lines of constant enthalpy on a $T - p$ diagram (Fig. 8).

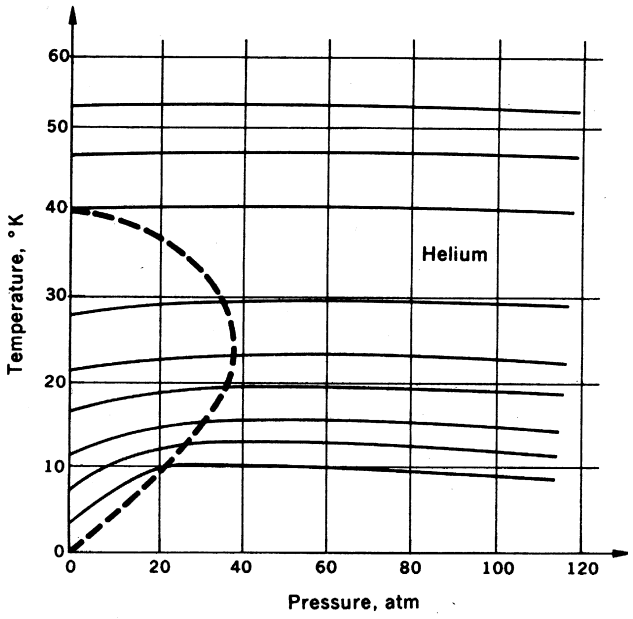


Fig. 8: Full lines: constant H; broken line: inversion temperature

This throttling process produces no change in temperature for an ideal gas. For real gases there is cooling only if the temperature is less than the **inversion temperature**. For helium the maximum inversion temperature is 40K. In helium liquefiers the gas is usually cooled below the inversion temperature by adiabatic expansion, and the final liquefaction is then achieved by throttling. This arrangement

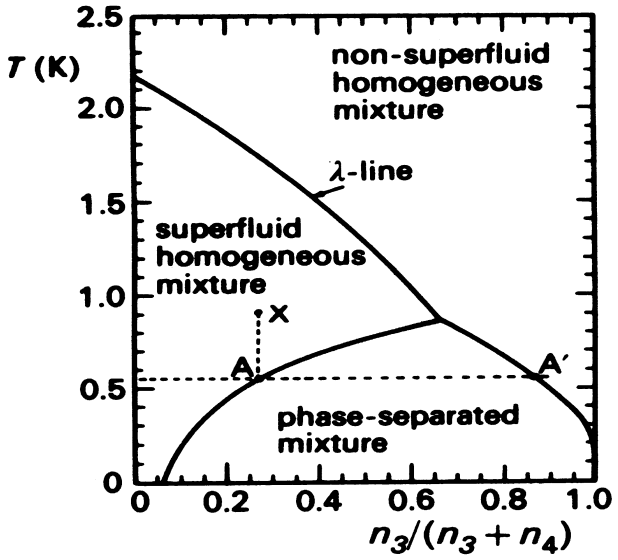


Fig. 9: The phase diagram for a liquid mixture of ^{3}He and ^{4}He.

avoids the production of liquid inside an engine. Again heat exchangers play an important role in the cooling process. Note that a Joule-Kelvin expansion is inherently irreversible, so that it is likely to be less efficient than an adiabatic expansion.

21.4 Evaporative cooling

Liquid helium boils at about 4.2K under atmospheric pressure. Temperatures down to about 1K can be obtained by evaporative cooling. Evaporative cooling of liquid ^{3}He can be used to obtain temperatures down to about 0.3K.

21.5 ^{3}He-^{4}He dilution refrigerators

A liquid mixture of ^{3}He and ^{4}He undergoes phase separation at low temperatures, as shown in the phase diagram in Fig. 9. We see that at the lowest temperatures two phases can coexist, one of which is almost pure ^{3}He (the concentrated phase), and the other is superfluid ^{4}He containing about 6 percent ^{3}He (the dilute phase). Since the dilute phase is superfluid, the ^{3}He can move through it as though it were in a vacuum. Thus the two-phase system is analogous to a liquid (the pure ^{3}He in the concentrated phase) in equilibrium with its vapour (the ^{3}He in the dilute phase). If the ^{3}He "vapour" can be continuously removed, by evaporating it from the dilute phase at a remote point (the "still") where the temperature is higher, there will be "evaporative cooling" at the interface between the concentrated and dilute phases.

The resulting device, shown schematically in Fig. 10, is called a **^{3}He-^{4}He dilution refrigerator**, and it is universally used to obtain temperatures in the range from 100mK to 2mK.

21.6 Magnetic cooling

Before the invention of the dilution refrigerator temperatures below about 300mK had to be obtained by magnetic cooling (adiabatic demagnetization), with a suitable paramagnetic salt containing electron magnetic moments. Magnetic cooling is still extensively used for temperatures below about 2mK, but it must then be based on nuclear magnetic moments.

The principle of magnetic cooling is illustrated in Fig. 11, which shows how the entropy of an assembly of magnetic dipoles (each with spin $1/2$ and magnetic moment μ) embedded in a solid varies with temperature at two different applied magnetic fields: $H_2 > H_1$.

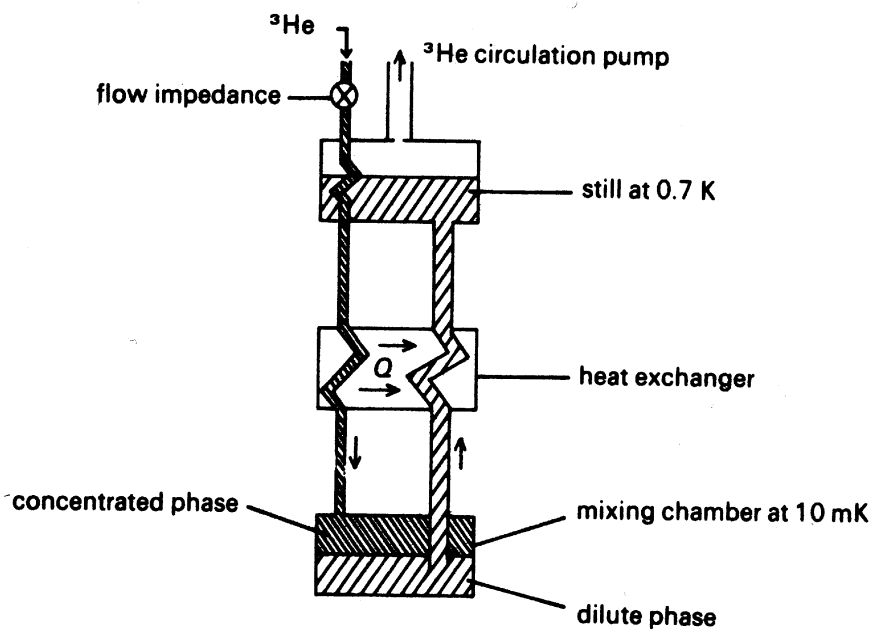


Fig. 10: Schematic dilution refrigerator

At high temperatures and small magnetic fields ($k_B T \gg \mu H$) the dipoles are arranged randomly in direction and have entropy $R \ln 2$ per mole. In the opposite limit the dipoles are all aligned parallel to the field, and the entropy is practically zero. The change in entropy occurs when $k_B T \sim \mu H$. The magnetic cooling process is shown by the arrowed lines. The assembly of dipoles is kept initially in thermal contact with a "high" temperature reservoir, and the magnetic field is raised from a low value to a high value. The entropy of the dipole system decreases, so the entropy of the reservoir must rise: heat is transferred to the reservoir (compare with an isothermal compression of a gas). Thermal contact with the reservoir is then broken, and the magnetic field is reduced slowly and adiabatically: i.e. at constant entropy. The dipole system cools as indicated in Fig. 11. In practice the dipole system must be used to cool some other system, but at low temperatures this other system is likely to have a heat capacity much smaller than that of the dipole system, so cooling of it presents no problem if thermal contact between the dipoles and this other system can be maintained.

In practice any dipole system will become ordered (lose its entropy) at a sufficiently low temperature even in the absence of an applied magnetic field; this is caused by residual interactions within the dipole system or between the dipoles and the host solid. If the dipole system is an electron paramagnet, these residual interactions limit the final temperature to a value not less than a few mK. However, if the dipole system is a nuclear paramagnet, the residual interactions are much smaller, and the final temperature can be much lower. This is the basis of the nuclear cooling technique for obtaining temperatures less than 1mK, starting from a "high" temperature of about 2mK, obtained with a dilution refrigerator. The nuclear paramagnet is commonly copper metal, in the form of fine wires to avoid eddy current heating during demagnetization.

(*Exercise for the student*: Use curves such as those in Fig. 11, together with the Third Law, to demonstrate the unattainability of the absolute zero of temperature.)

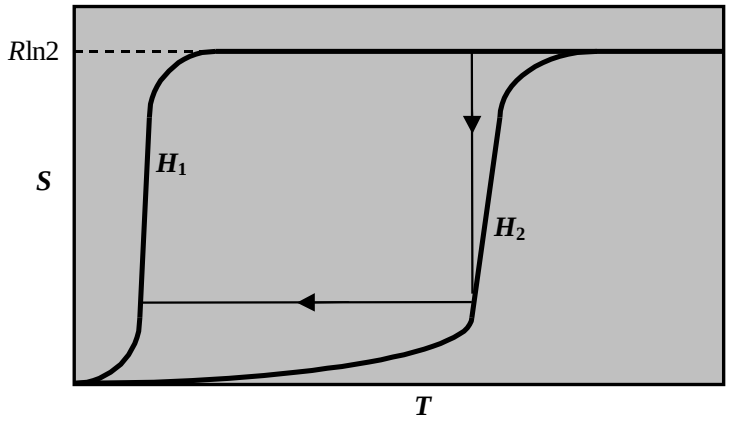


Fig. 11: Illustrating the principle of magnetic cooling

22. STATES OF METASTABLE EQUILIBRIUM

Strictly speaking thermodynamic arguments relate to physical systems in equilibrium, although directions of irreversible change can be discussed. For a system held at constant temperature and volume, the state of equilibrium is the one for which the free energy is an absolute minimum. It is only for a system in equilibrium that the entropy tends to zero as $T \to 0$. Often a system will reach this state of equilibrium if it is left undisturbed for a reasonably long time. But this is not always the case: systems can reach states of metastable equilibrium, in which the free energy is a local minimum but not an absolute minimum. An example is a solid containing frozen-in defects; if the temperature is so low that the defects cease to be mobile, there is a local minimum in the free energy out of which the system cannot be thermally excited in any reasonable time, but the entropy does not vanish as $T \to 0$. And the internal energy may well depend on the particular configuration of defects that is frozen in. These states of metastable equilibrium can still be treated thermodynamically, provided we recognize that, for example, the free energy is not an absolute minimum, but is only a minimum subject to certain constraints. These constraints may relate to defects that have ceased to be mobile on any reasonable timescale. Or they may be more subtle, as in the case of a ferromagnetic material. Such a material has a permanent magnetic moment, but in the absence of any applied magnetic field the moment is equally likely to point in one direction as in the opposite direction, the two states having the same free energy. In a state of strict equilibrium the moment would occasionally flip from one direction to the other, leading to zero magnetic moment on the average. But the probability of such a flip is usually extremely small and can be neglected. We can apply thermodynamic reasoning to the magnet, but it will be subject to the constraint that the magnetic moment points in a specified direction; or equivalently that the accessible microstates are restricted to those corresponding to this specified direction of magnetization. I am grateful to some of the students at the School who raised these matters with me in discussion.

23. EPILOGUE

It is clear that entropy plays a crucial role in thermodynamics. Refrigeration often involves in essence the transfer of entropy from one system to another and the changes in temperature that accompany isentropic processes. Our approach has been to introduce entropy as a measure of the number of microstates or statistical weight of a system, and this serves to provide an immediate physical feel for the concept. Other approaches are possible. The Second Law can be stated in terms of processes that heat engines are not allowed to perform, and this approach leads usually to Eq. (10) as a definition of entropy. This latter approach is more directly related to many of the applications that we have had in mind, but it leaves one

with little feel for the meaning of entropy and no means of calculating entropy from the microscopics of the system of interest. In a sense our approach is *not pure phenomenological thermodynamics*, but rather thermodynamics based on some understanding of the underlying *statistical mechanics*.

REFRIGERATION

U. Wagner
CERN, Geneva, Switzerland

Abstract
This introduction to cryogenic refrigeration is written in the frame of the CERN Accelerator School (CAS) Course on Superconductivity and Cryogenics. It consequently concerns refrigeration for cooling of superconductive devices with emphasis of liquid helium refrigeration. Following a general reminder of the thermodynamics involved, i.e. principally the first and second laws, the different thermodynamic cycles used to produce low temperatures are presented. Refrigerators working according to the different cycles are explained starting with the low capacity range covered by cryocoolers and then concentrating on Claude-cycle refrigerators. The different thermodynamic and technological elements of a Claude-cycle refrigerator are discussed and an example for a large refrigerator used in an accelerator equipped with superconducting magnets is presented.

1. INTRODUCTION

The cryogenic temperature range has been defined as from -150 ˚C (-123 K) down to absolute zero (-273 ˚C or 0 K), the temperature at which molecular motion comes as close as theoretically possible to ceasing completely [1].

The application of cryogenics can roughly be separated into five major domains:

i) Liquefaction and separation of gases

ii) Storage and transport of gases

iii) Altering material and fluid properties by reduced temperature

iv) Biological and medical applications

v) Superconductivity

With the exception of the quest to get closer and closer to the absolute zero, in all the cases listed above cryogenics is a utility enabling the desired application or function. Since the first liquefaction of helium by Heike Kammerlingh Onnes in 1908, cryogenics has evolved in an engineering science combining different technologies in a "best compromise" to achieve a given goal. This should be remembered when reading the following paragraphs as several of the presented technological solutions are governed not only by the optimum achievable, in terms of e.g. efficiency or performance but as well by cost of investment and operation.

For particle accelerators, cryogenics is an ancillary technology for enabling the superconducting operation of accelerating or beam-positioning devices. Thus the choice of a cryogenic system is lagely determined by the superconductors used.

1.1 Critical and operating temperatures of superconductors

In order to operate superconducting devices it is evident that the cooling agent used must have a temperature significantly below the temperature of superconductive transition, or critical temperature of the superconductor used. Table 1 lists the critical temperatures of some commercially available materials.

Table 1
Critical temperature of superconducting materials

Material	T_c [K]
Pb	7.2
Nb	9.2
Nb-Ti alloys	~ 9.6
Nb_3Sn	18.1
Nb_3Ge	23.2
$Y_1Ba_2Cu_3O_{7-x}$ (YBCO)	90
$BI_2Sr_2Ca_1Cu_2O$ (Bi2212)	80
$BI_2Sr_2Ca_2Cu_3O$ (Bi2223)	125

1.2 Properties of cryogenic fluids

The cryogenic fluid is both the working and cooling agent of the cryogenic system. Only fluids having a triple point below 100 K are considered "cryogenic", i.e. they are still in either liquid or gaseous form below this temperature. Table 2 shows some properties of cryogenic fluids. The critical temperature of the fluid refers to the upper limit at which the liquid phase may appear.

Since the disappearance of liquid hydrogen bubble chambers, hydrogen is very little used as cryogenic cooling fluid and oxygen is never used due to the serious hazards implied. Neon though being an inert gas is very expensive and has therefore never been considered up to now. Consequently the only fluids used are helium and nitrogen. The emergence of the high-Tc superconducting materials has opened the way to use nitrogen as cooling fluid. As the range of application using nitrogen is limited by its triple point still today the prevailing cryogen is helium.

Table 2
Property data of cryogenic fluids

Fluid		^4He	H_2	Ne	N_2	O_2
Normal boiling point	[K]	4.22	20.4	27.2	77.4	90.2
Critical temperature	[K]	5.20	33.2	44.4	126.	155.
Critical pressure	[MPa]	0.23	1.32	2.72	3.39	5.08
Triple point temperature	[K]	2.18*	14.0	24.6	63.1	54.4
Triple point pressure	[kPa]	5.04*	7.20	43.2	12.8	0.15
Liquid density at ambient pressure	[kg/m^3]	125.	70.8	1204.	808	1140
Vapour density at ambient pressure	[kg/m^3]	16.7	1.33	9.46	4.59	4.75
Normal density	[kg/m^3]	0.18	0.09	0.90	1.25	1.43

*: Lambda point

Helium shows the particularity that it has no triple point, it may solidify only at pressures above 2.5 MPa. The commonly given lambda point refers to the transition from normal to superfluid helium.

2. BASIC ENGINEERING THERMODYNAMICS

Cooling in a steady-state process is done by letting a working fluid undergo a series of cyclic thermodynamic transformations, called a thermodynamic cycle. The working fluid absorbs heat at low temperatures and rejects it at higher temperatures.

2.1 The Carnot factor

The Carnot factor is the predominant estimator for all low temperature process considerations. It is named after Sadi Carnot, who in his "Réflexions sur la puissance motrice du feu" introduced first the concept that the work that can be extracted from a heat engine depends on the temperatures between

which this engine works. For a heat engine, the Carnot factor defines the maximum amount of work that can be extracted from a process operating between two temperature levels. For a heat pump, or a refrigerator, it defines the minimum amount of work necessary to extract heat at a low temperature and reject it at a higher one. In fact, the Carnot factor is a direct consequence of combining the first and second law of thermodynamics. Consider an open process removing heat from a given (cold) temperature level and rejecting it at a higher level as sketched in Figure 1.

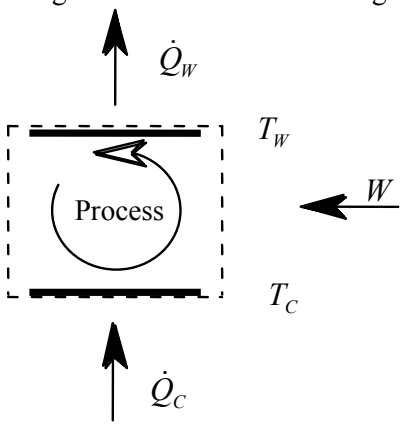


Figure 1: Basic open process transporting thermal energy from any temperature to a higher temperature level

The first law of thermodynamics concerns the conservation of energy.

$$\dot{Q}_w + \dot{Q}_c + W = 0 \tag{1}$$

The second law of thermodynamics defines that in any process the entropy is either constant (reversible process) or increased (irreversible process) but never diminished.

$$\frac{\dot{Q}_w}{T_w} + \frac{\dot{Q}_c}{T_c} \leq 0 \tag{2}$$

Combining equation (1) and (2) one can define the necessary power input into the process.

$$W \geq \dot{Q}_c \left(\frac{T_w}{T_c} - 1 \right) \tag{3}$$

The Carnot factor is defined as:

$$\left(\frac{T_w}{T_c} - 1 \right)$$

By convention heat or work is considered positive when entering the system and negative when leaving the system.

Figure 2 shows the Carnot factor as function of the cold temperature with 300 K as warm reference temperature. It is quickly evident that the energy necessary to extract heat below 20 K increases rapidly with decreasing temperature. It is further not astonishing that for T -> 0 K the Carnot factor goes to infinity. Besides defining the minimum work necessary in a reversible process to extract heat, the Carnot factor clearly demonstrates why in any cryogenic system heat entering the low temperature level should be limited to the necessary minimum. The refrigeration work in real systems is always above the limit given by the Carnot factor due to the inevitable entropy losses.

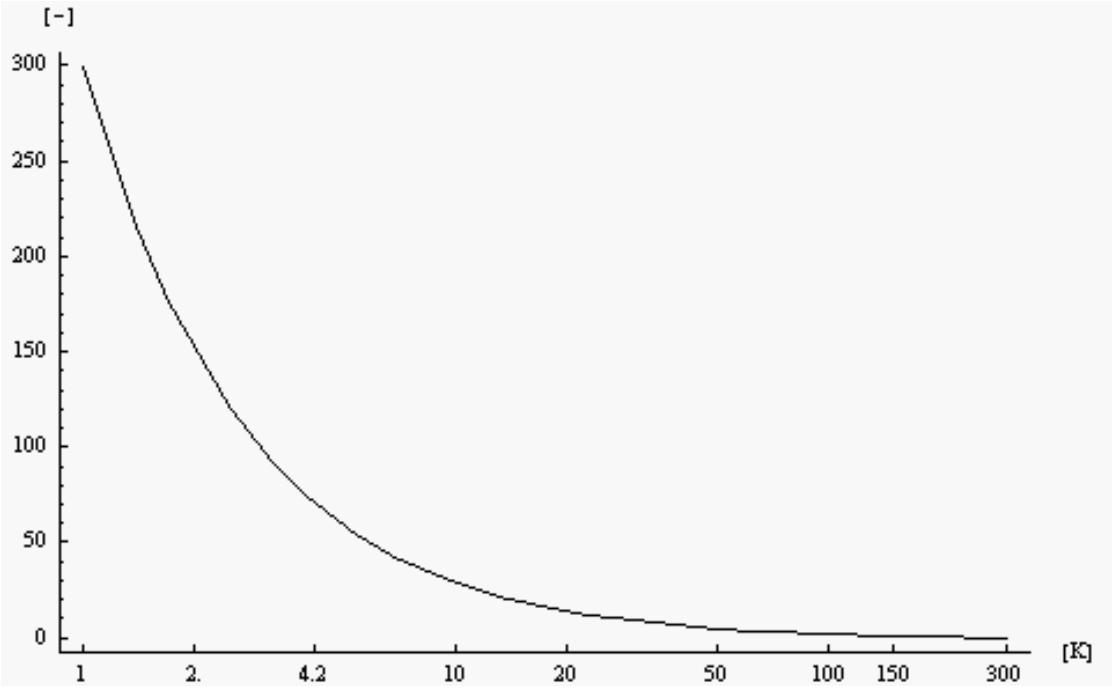


Figure 2: Carnot factor as function of cold temperature for T_W = 300 K

2.2 The diagram

Within the number of different phase diagram for fluids, the temperature-entropy [T-s] diagram is the most practical to illustrate the different cycles used for the closed cryogenic processes.

In a T-s diagram the reversibly exchanged heat for any change of state is represented by the area under the path for the change of state. This allows to compare the heat exchanged for different process changes. Figure 3 shows the commonly used T–s diagram issued by the National Bureau of Standards (NBS, now National Institute of Standards and Technology, NIST).

For low temperature application it may be convenient to use a T-s diagram with a logarithmic temperature scale. In this case for any ideal gas the isochors, isobars, and isentalphs are straight lines and the isenthalps additionally coincide with the isotherms. To go one step further, if the lines for constant enthalpy, pressure or volume are straight in a logarithmic T-s diagram for a real gas, on can deduce that in these areas the gas behaves like an ideal gas. Taking the logarithmic T-s diagram for helium, we quickly can establish that ideal gas laws are sufficient to calculate a process above an entropy of 13 J/g K and below a pressure of 2.0 MPa covering a large part of any helium cycle. Close to the two-phase area and below 2.2 K helium becomes anything but ideal [2], detailed calculation of any process needs to relate of the real gas data. Another advantage of the logarithmic T-s diagram is that the cold area, i.e. the one of interest, is best represented. The disadvantage of the logarithmic temperature scale is that the area under the path for a change of state does not represent the reversibly exchanged heat. Figure 4 shows a logarithmic T-s diagram in order to illustrate the statements above. The author agrees that this document, dated from 1941, is difficult to read and from a time prior to SI units. It is nevertheless the only such document he could lay hands on.

The world reference for the helium properties is NIST Technical note 1334 (1989) [3]. The property equations serving as basis for this note have been compiled into a program by its authors, and is commercially available [4].

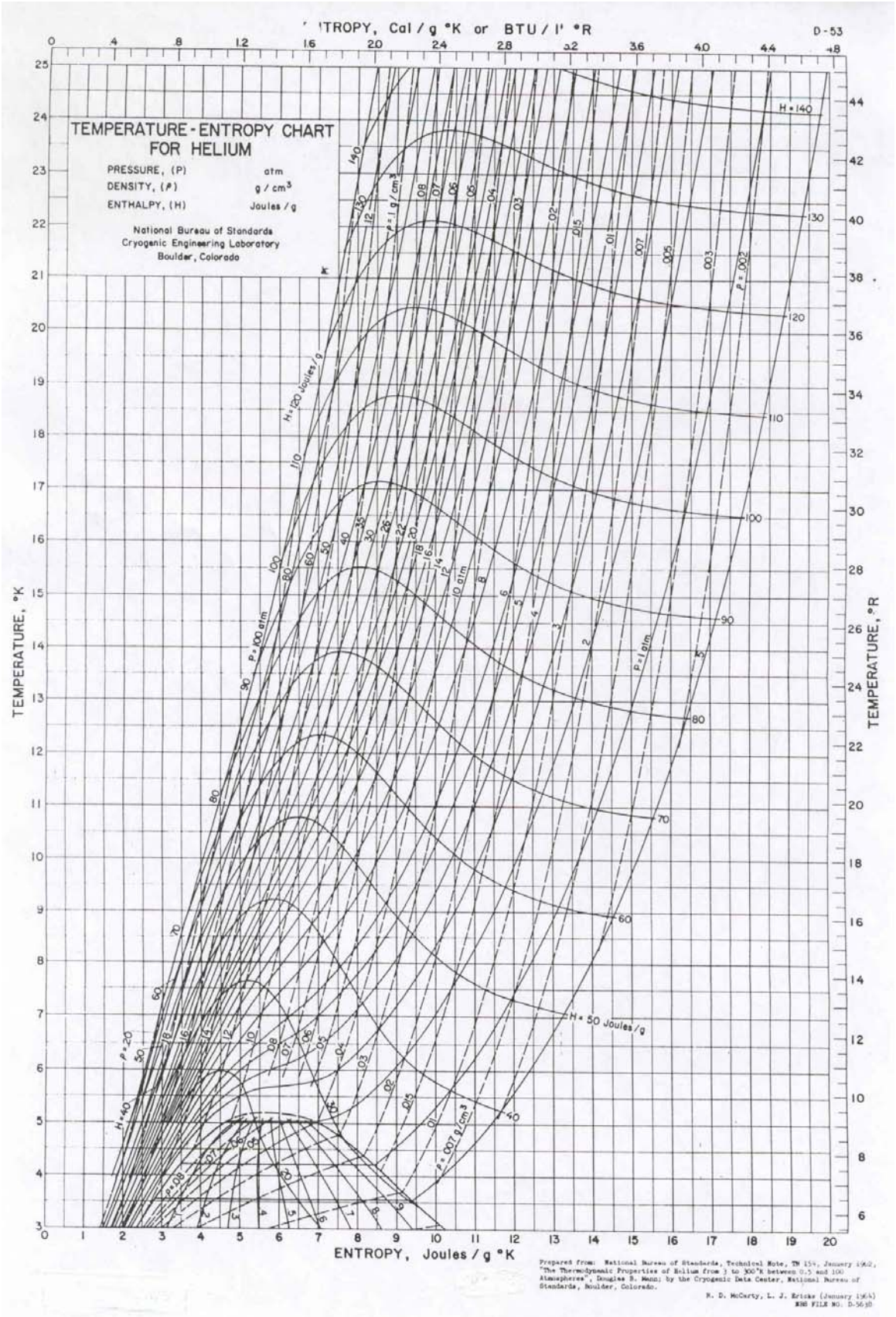


Figure 3: T-s Diagram for Helium (issued by NBS)

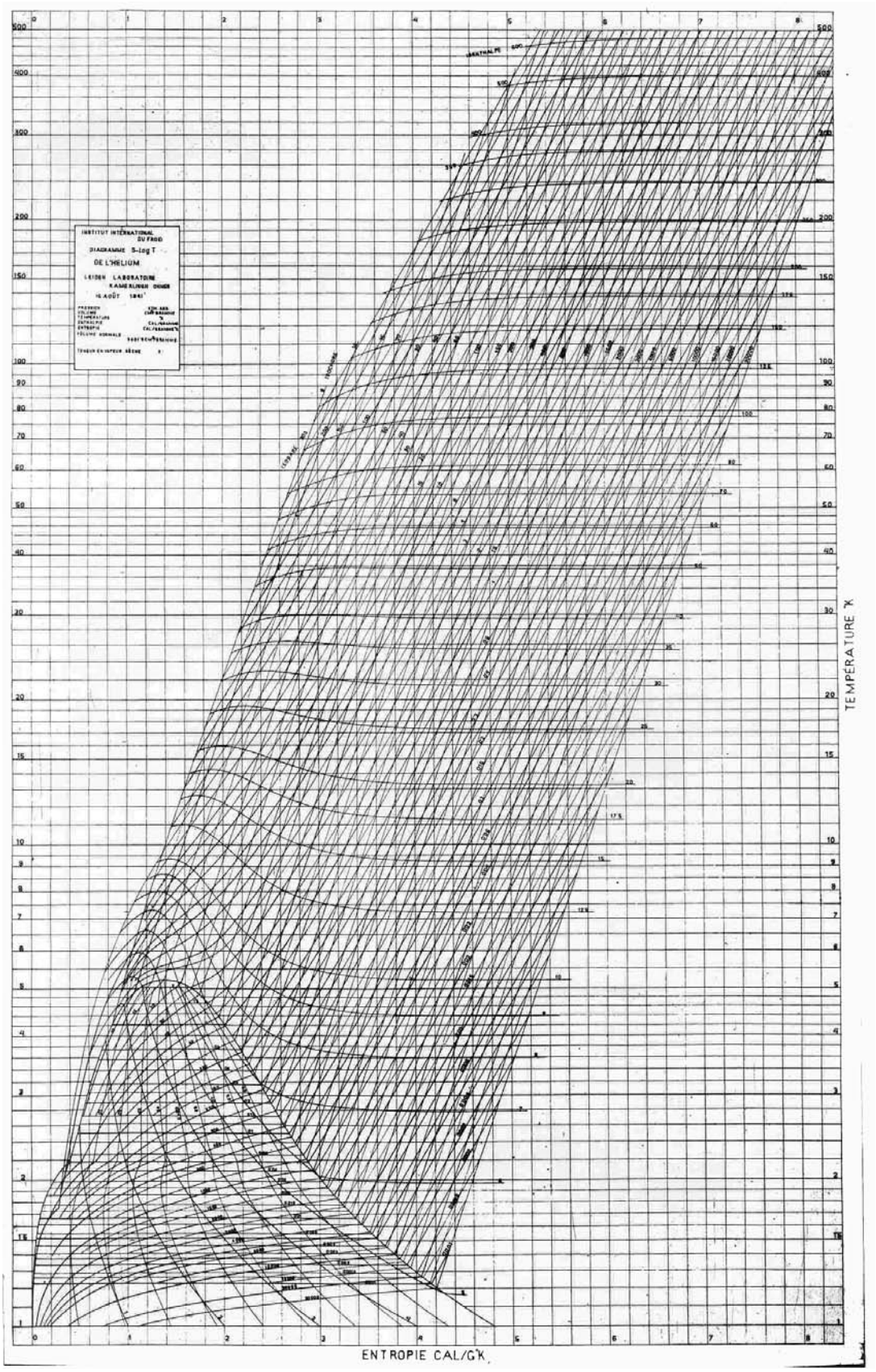


Figure 4: lgT-s diagram for helium (issued by: Institut International du Froid)

The use of T-s diagrams for designing any process has significantly decreased since the widespread use of calculation software in combination with the property data. It should nevertheless be mentioned that the representation of a process in a T-s diagram considerably eases its understanding.

2.3 The basic cryogenic refrigeration cycles

All cycles listed below are "ideal" cycles, meaning they do follow ideal transformations that in reality can only be approximated. Some of these cycles are easier to realise with existing technology than others as it is possible to achieve nearly isobaric heat exchange, whereas it is very difficult to achieve any isentropic change of state. Of the multitude of possible thermodynamic cycles only those with frequent application in cryogenics are listed. Please note that some of these cycles may as well appear under a different name in literature; the identification used in this work corresponds to the most spread in the cryogenic community as far as experienced by the author.

The Carnot cycle as the ideal cycle represents the reference to which any other process is measured. The process follows a sequence of isothermal expansion, isentropic compression, isothermal compression and isentropic expansion.

The Stirling cycle follows a sequence of isothermal expansion, isochoric compression, isothermal compression and isochoric expansion.

The Ericsson cycle follows a sequence of isothermal expansion, isobaric heating, isothermal compression and isobaric cooling.

Figure 5 shows the these three cycles in a logarithmic T-s diagram for ideal gas. The necessary work input for all three cycles would in this case (ideal gas!) be equal, in other words all three cycles would have the some efficiency.

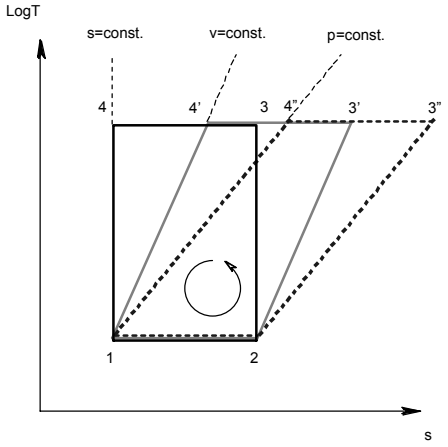


Figure 5: Carnot cycle (1,2,3,4), Stirling cycle (1,2,3',4') and Ericsson cycle (1,2,3",4")

The Brayton cycle, follows a sequence of isentropic expansion, isobaric heating, isentropic compression and isobaric cooling.

Figure 6 shows the Brayton cycle in comparison to the Ericsson cycle. In this comparison, the cooling capacity of the Brayton cycle is inferior to that of the Ericsson cycle though it is available at a lower temperature, the work input into the Brayton cycle is higher. These two aspects immediately qualify the Brayton cycle as less efficient. Though by definition less efficient then the three cycles shown in Figure 5 , this cycle combines thermodynamic transformations that are easier to approximate in real systems which made the Brayton cycle one of the references for cryogenic refrigeration.

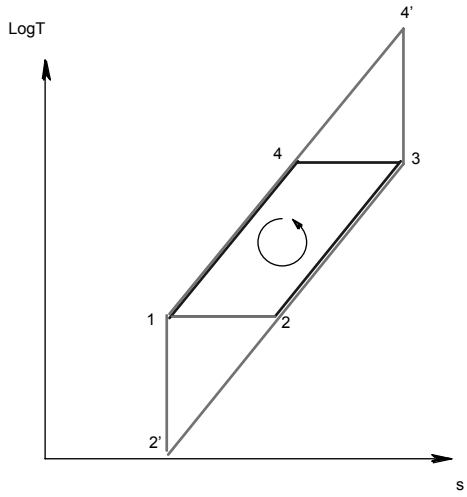


Figure 6: Ericsson cycle (1,2,3,4) and Brayton cycle (1,2',3,4')

2.4 Real gas aspects

A real gas is characterised by the fact that pV≠RT . In this case isenthalpic expansion, e.g. in a valve can produce cooling. This effect is called the Joule-Thomson effect and happens in a property region limited by the Joule-Thomson inversion curve that is dependent on temperature and pressure [5]. The two cycles listed below make use of this effect.

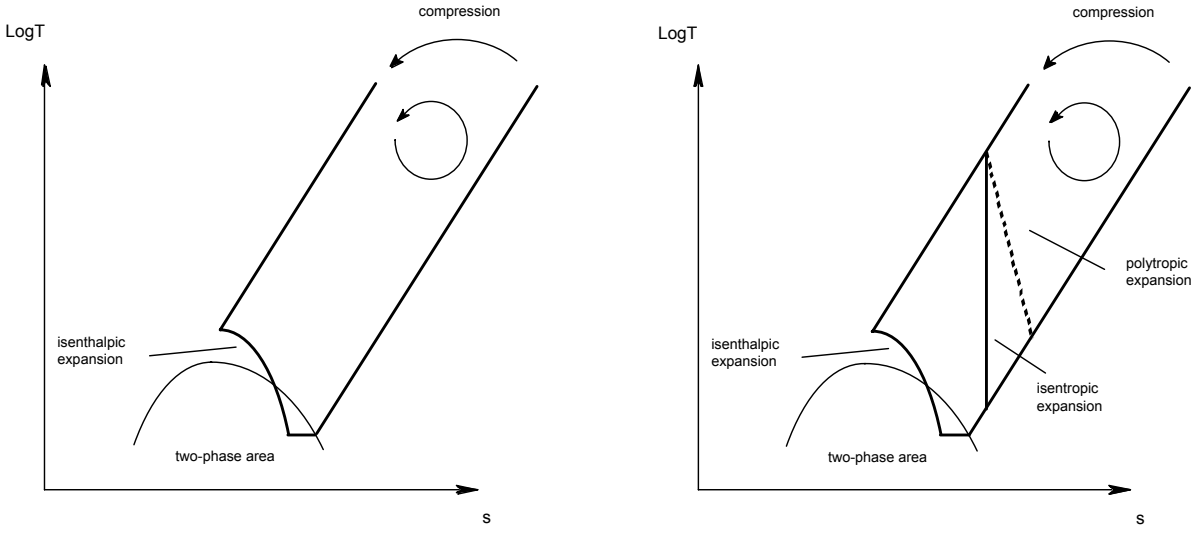


Figure 7: Joule-Thomson cycle (left) and Claude cycle (right) for a real gas

The Joule-Thomson (J-T) cycle combines isobaric cooling and heating with isothermal compression and isenthalpic expansion. For an ideal case the expansion would be isotherm, leading back to the Ericsson cycle.

The Claude cycle is in fact a combination of one (or more) Brayton cycles and a Joule–Thomson cycle at the cold end.

Figure 7 shows these two cycles in an simplified logarithmic T-s diagram for real gas. The compression is not further qualified concentrating only on the cycle part below atmospheric temperature. In the Joule-Thomson cycle the total amount of gas is cooled isobaric to a point close to

the two-phase zone. The isenthalpic expansion then results in the creation of a two-phase fluid. This cycle was first applied by Linde for the industrial liquefaction of air. In the Claude cycle part of the gas is expanded at an intermediate temperature in an expansion machine that ideally would realise an isenthalpic expansion in the real case the expansion is polytropic as sketched in Figure 7. This cycle was used and patented by Claude for the liquefaction of air, leading to a considerably higher total efficiency as the Joule-Thomson cycle.

A process based on a pure Joule-Thomson cycle and starting from ambient temperature can only be used to liquefy a gas that has a Joule-Thomson inversion temperature above ambient. This is the case for nitrogen, oxygen and air. Helium, hydrogen and neon cannot be liquefied in a Joule-Thomson process as sketched.

In a Claude cycle any gas can be liquefied, it is today the prevailing process arrangement for the liquefaction of helium. Though thermodynamically less efficient then the cycles listed in section 2.3, it is best adapted for technical realisation.

2.5 Cycle efficiencies

The efficiency of a cycle to produce cooling at liquid helium temperature determines the cost of operation for the refrigerator. As concerns the total cycle efficiency, two ways of calculation are commonly used, the Carnot efficiency and the exergetic efficiency. The latter can be directly derived from the Carnot efficiency and has the advantage of additionally providing a means to analyse any part of the process concerning its contribution to losses.

2.5.1 Carnot efficiency

The Carnot factor as presented in section 2.1, defines the theoretical limit for an amount of energy to be injected in a cooling process. Consequently the efficiency of any process is expressed as efficiency relative to the Carnot factor or, in short, Carnot efficiency.

$$\eta_{carnot} = \frac{W_{carnot}}{W_{effective}} \cdot 100\,[\%] \tag{4}$$

2.5.2 Exergetic efficiency

The concept of "énergie utilisable" was invented by Gouy in 1889 [6],the term "exergy" was coined by Rant in 1956 [7]. Exergy takes into consideration both the energy and its thermodynamic grade i.e. the Carnot factor based on T_W equal to the ambient temperature. Applying the equation under (3) for a non-isothermal process between the two cold temperatures T_{c1} and T_{c2} and assuming no irreversibility, i.e. defining the minimum necessary work it would be changed to:

$$W_{\min} = \int_{T_{C1}}^{T_{C2}} d\dot{Q}_C \left(\frac{T_W}{T} - 1 \right) \tag{5}$$

Assuming further that we have a working fluid to which the heat is transferred at the relevant temperature level, again without irreversibility i.e. no temperature difference and no pressure losses, we can write for this fluid:

$$W_{\min} = -T_W \int_{T_{C1}}^{T_{C2}} \frac{d\dot{Q}_C}{T} + \int_{T_{C1}}^{T_{C2}} d\dot{Q}_C \tag{6}$$

Respectively:

$$W_{\min} = -T_W \int_{T_{C1}}^{T_{C2}} dS + \int_{T_{C1}}^{T_{C2}} T dS \tag{7}$$

For the case of a steady state fluid process, that is often considered for a process definition, and by using the specific properties of state at the inlet and outlet of the considered cooling process one can transform equation (7) to:

$$W_{\min} = -T_W \dot{m}(s_{T_{C2}} - s_{T_{C1}}) + \dot{m}(h_{T_{C2}} - h_{T_{C1}}) \tag{8}$$

Equation (8) defines the exergetic capacity the fluid absorbs, i.e. the exergetic capacity required to cool the process considered. T_W in this case represents the reference temperature on which finally all heat and work absorbed by the process will be rejected as heat flow. For our earthbound system this is the ambient temperature, usually set to 300 K.

In other words the exergetic capacity describes the necessary effort to change in any thermal process the temperature as compared to the natural equilibrium temperature given by our earthbound environment.

The specific exergy , defined as:

$$e = -T_{ambient} s + h \tag{9}$$

is a combined property of state once $T_{ambient}$ is defined.

It is very convenient to calculate the exergetic capacity of any cooling process using the inlet and outlet conditions of the cooling fluid taking the heat. It can be directly compared to the Carnot capacity. Exergetic efficiency can be calculated for any part of a process by comparing the exergy flow balance on the input and output boundaries.

$$\eta_{exergetic} = \left| \frac{\sum_{output\ boundary} \dot{m}\left(-T_{ambient} s + h\right)}{\sum_{input\ boundary} \dot{m}\left(-T_{ambient} s + h\right)} \right| 100\,[\%] \tag{10}$$

The absolute value is necessary as exergy, like energy flowing into the boundary is, is defined positive, flowing out negative.

As electrical power is defined as pure exergy, it has the same "value" independent of the temperature where it is used, the exergetic efficiency for a total process can therefore be written as:

$$\eta_{exergetic} = \left| \frac{\sum_{cooling\ boundary} \dot{m}\left(-T_{ambient} s + h\right)}{P_{electric\ input}} \right| 100\,[\%] \tag{11}$$

An exergy analysis determining the exergetic losses for a process and its different parts is an excellent tool for cycle analysis. The sum of the losses for the different parts equals the total loss and the contribution of the individual losses to the total can therefore be quantified.

2.5.3 Achieved efficiencies in cryogenic equipment

Figure 8 shows a large number of realised cryogenic refrigeration systems plotted with their Carnot efficiency versus refrigeration capacity at 4.5 K. It is significant to note, that the efficiency increases with refrigeration capacity and seems not dependent on the technical solution.

The reason for this behaviour may be explained by several combining factors, both technologically and commercially. Technologically because larger equipment has a higher the individual efficiency like e.g. rotating equipment like turbo expanders. Commercially as the incentive to spend investment cost (equipment) in order to save on operation cost (efficiency) increases with the absolute value of the power consumption. The relative increase in investment cost for a more efficient machine is further smaller in case of a large refrigerator.

The plot shown in Figure 8 must be understood under these two aspects, it does not represent a law of physics. Experience nevertheless shows that the tendency shown in the graph is universal.

It should be mentioned at this point that any helium refrigerator represents a complex system combining different technologies in a "best compromise" following the taste of the designer and the requirements of the client for whom it was built. Both of which may influence the efficiency of the installation in one way or other.

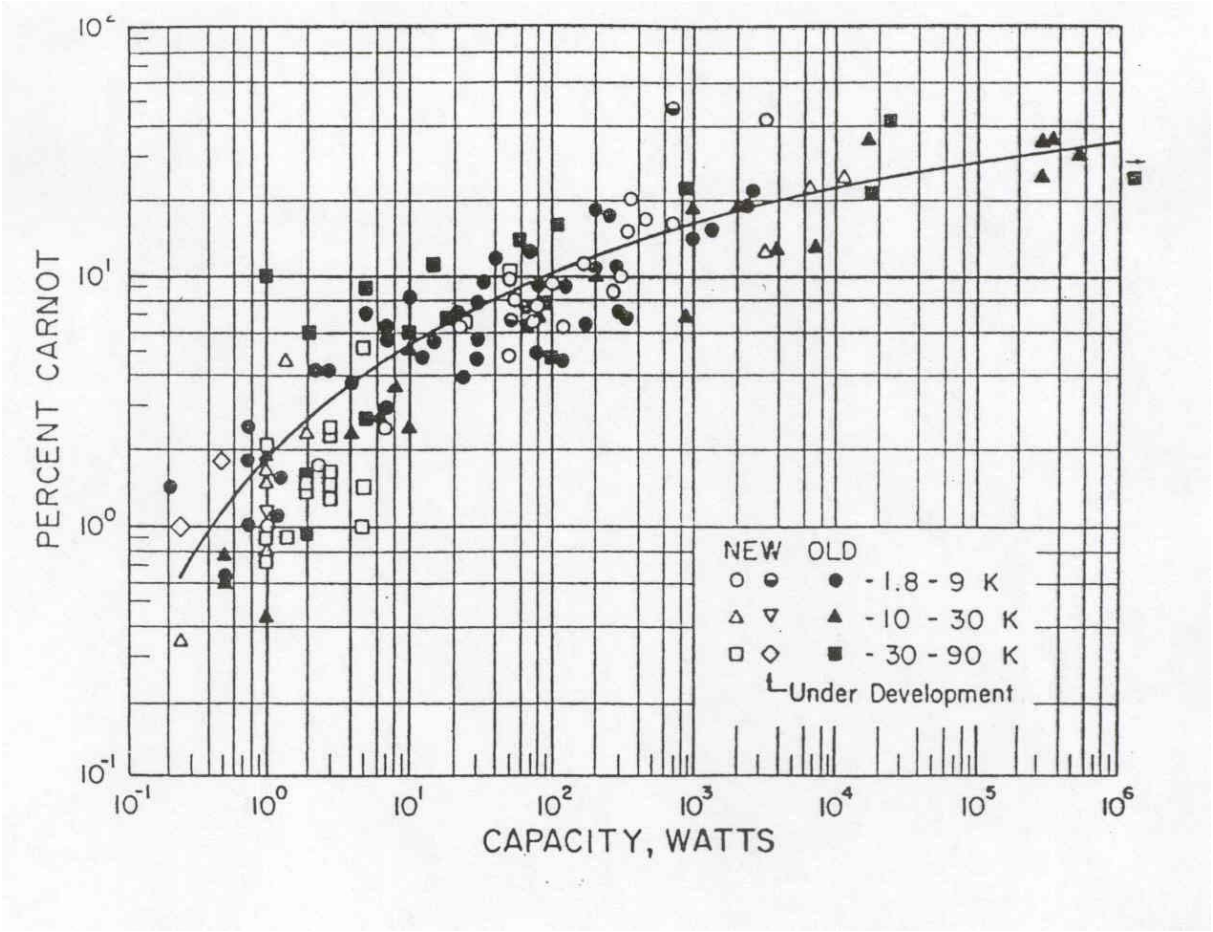


Figure 8: Efficiency of cryogenic refrigeration systems (from [8])

2.6 Liquefaction and refrigeration

The two terms of liquefier and refrigerator are often used for the same type of cryogenic machinery. The two processes of liquefaction and refrigeration are in fact basically different. A "pure" refrigerator delivers only isothermal cooling capacity, at the liquid temperature of the working fluid. A liquefier uses its cooling capacity to cool a part of the working fluid from ambient to its boiling point and supply this in liquid form as the final product of the process. The two process variants are sketched in Figure 9.

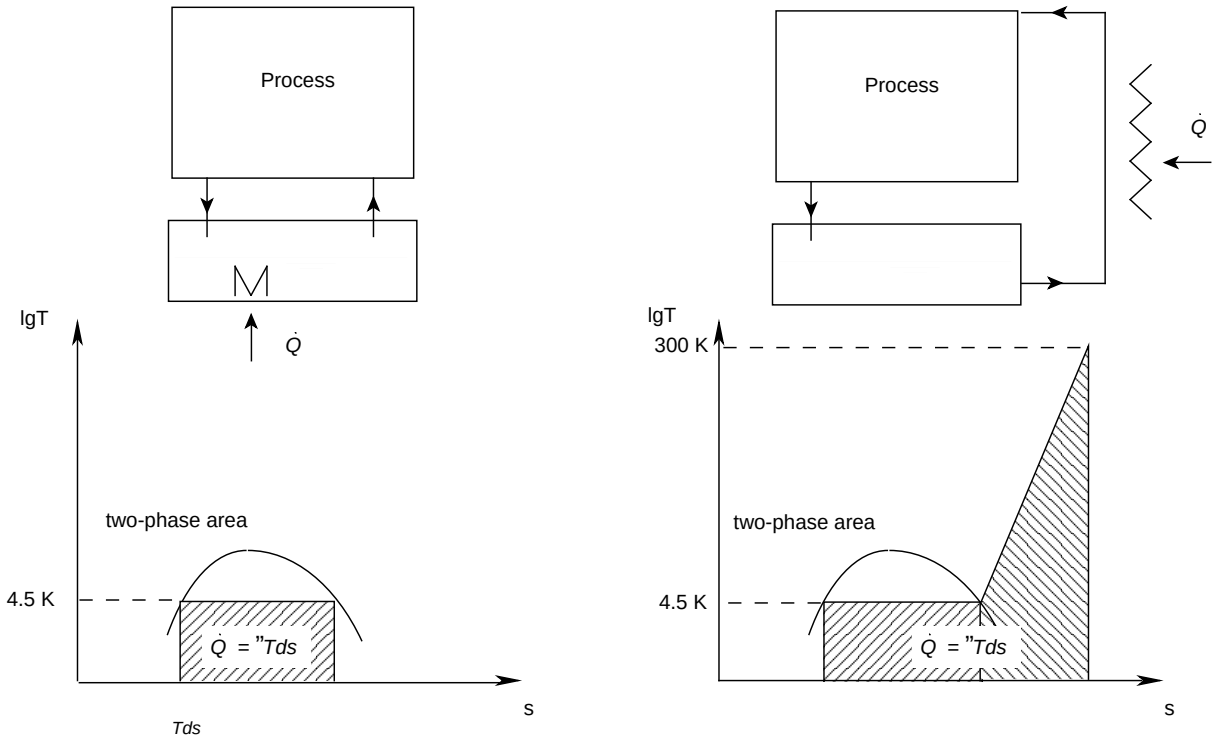


Figure 9: Comparison of cryogenic refrigeration (left) and liquefaction (right)

Remembering the remark from section 2.2 we may identify that the reversibly exchanged heat for a given mass flow is considerably higher for a liquefier than for a refrigerator.

2.7 Equivalent isothermal capacity

Following the Carnot factor it is evident that the effort necessary to extract heat at low temperatures is strongly temperature dependent. For pure isothermal cooling in a two-phase bath it is evident to relate cooling capacity and cooling temperature. For non-isothermal cooling, like for liquefaction, this becomes more difficult. In order to compare capacities of cryogenic machines the equivalent capacity was introduced expressing the total cooling capacity as equivalent isothermal refrigeration at 4.5 K.

For practical reasons for pure liquefiers one refers to the capacity rather by defining the liquefaction flow either in volume flow or mass flow.

The equivalent capacity represents an exergy calculation for the different cooling capacities referred to the exergetic capacity of pure refrigeration at 4.5 K. If we can compare the exergetic capacity necessary to liquefy one g/s of helium from 300 K to 4.5. K with the same exergetic capacity for pure refrigeration at 4.5 K.

Table 3 shows the property data for helium at 4.5 K vapour, 4.5 K liquid and ambient conditions. The exergetic capacity for the liquefaction of 1 g/s helium can be calculated to: 6754 W.

For the same exergetic capacity we may evaporate 5.47 g/s of liquid helium which would have an isothermal cooling of 103 W. One may therefore roughly say that 1 g/s of liquefied helium needs an equivalent effort as isothermal cooling of 100 W.

Table 3
Property data for helium to define the equivalent capacity of liquefaction and refrigeration

Temperature [K]	Pressure [kPa]	Entropy [kJ/kg K]	Enthalpy [kJ/kg]	Exergy [kJ/kg]
300	100	31.61	1574	-7910
4.5 (vapour)	100	8.067	30.47	-2390
4.5 (liquid)	100	3.893	11.64	-1156

3. COOLING METHODS

For an experimentalist having e.g. a small sample to cool to low temperatures for his work, the choices for supplying the cooling can be expressed in the following cases.

i) Buying liquid helium use it to cool the sample and let it evaporate to atmosphere.

ii) Buying liquid helium, recover the evaporated gas and return this in compressed form to the vendor.

iii) Buying gaseous helium and a refrigerator producing cooling capacity in closed cycle

These three simple cases change the necessary infrastructure for cryogenics considerably. In the first case one may install one big central dewar storing the helium from which one fills smaller, transportable dewars that supply the experiment cryostat. In the second case this equipment is increased by a recovery system, usually using balloons, a purification unit and a compression system with high pressure gas tanks. In the last case one may find a whole range of systems from compact cryocoolers supplying cooling to a small experiment over an installation with central liquefier, storage dewar, transport dewars, recovery and purification system, to a dedicated refrigerator for any large experiment including cryogenic distribution system with liquid helium transfer lines, recovery systems, purification systems and gas storage.

Cryogenic installations may become specialised systems of their own requiring considerable investment and operating cost. For anybody approaching this problem for the first time it is therefore recommended to ask for advice at institutions where such an infrastructure already exists.

The following chapters describe the principle machinery of cryogenic cooling equipment emphasising on helium refrigeration systems down to 4.2 K. As concerns systems using nitrogen as cooling agent, Stirling refrigerators are very common for a midrange capacity of about 20 kilowatt at 80 K. Otherwise liquid nitrogen is a comparably cheap and widely available fluid generally used in an open circuit, i.e. returning the evaporated gas to atmosphere where it originated.

4. CRYOCOOLERS

Cryocooler is the common name for refrigerators with small capacity ($\leq$ 10 W at 4.5 K). These refrigerators are very compact and avoid the need of extensive systems. A typical example for commercial use are the magnetic resonance imaging (MRI) systems where a cryocooler covers the heat losses of the liquid helium cryostat. A wide range of cryocoolers has been developed. The most common on the market are still Gifford-McMahon-type refrigerators. In recent years the development of pulse tube cryocooler made impressive progress and this equipment will become more and more frequently used.

This course is held within the frame of the CERN Accelerator School. For accelerator machines cryocoolers are in general not applicable as they lack the necessary capacity, especially concerning a liquefaction load. The selection below is therefore mainly intended to demonstrate the application of different cycles and in order to remind that for the small capacity range this equipment may be chosen with advantage.

4.1 Stirling cryocoolers

Though these machines are not widely used as cryocoolers the realised performance ranges from less than one watt to some tens of kilowatts at a temperature rang of 20 K to 80 K.

The Stirling cycle is realised following the simplified process in Figure 10. The numbering of the different process stages is done following the one for the Stirling cycle in Figure 5.

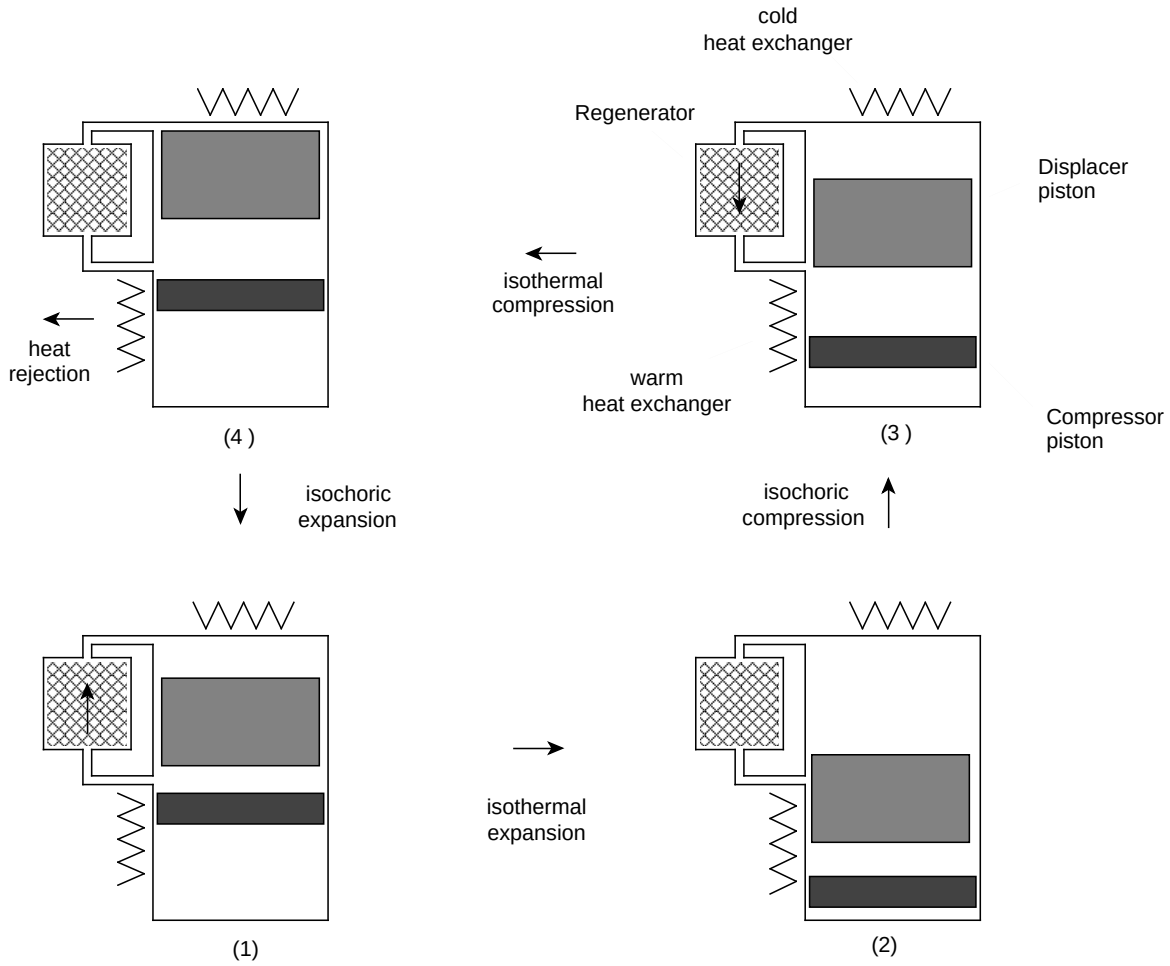


Figure 10: Schematic sketch of a Stirling cycle refrigerator

Between stage (1) and (2) the displacer piston and the compressor piston move down cooling the gas volume above the displacer piston and allowing heat to be absorbed and the cold heat exchanger. From stage (2) to stage (3) the displacer piston moves up pushing cold gas through the regenerator and warming the gas up while cooling the regenerator material. From stage (3) to stage (4) the displacer piston and the compressor piston move up heating the gas volume between the two pistons and allowing the heat rejection at the warm heat exchanger. Finally from stage (4) to stage (1) the displacer piston moves down, pushing the warm gas between the pistons back through the regenerator and cooling the gas while heating the regenerator material.

One may notice that the movement of the two pistons can be achieved by a phase angle of 90 degrees. This can be realised either by the geometry of the crank shaft or by separating the cylinder in two volumes arranged with orthogonal axes.

4.2 Gifford-McMahon

Gifford-McMahon cryocoolers are designed as two stage Ericsson refrigerators. Their useful operation temperature is at about 10 K for the second stage and 50 K for the first stage. The different steps of the process can be followed according Figure 11. The numbering of the different process stages is done following the one for the Ericsson cycle in Figure 5.

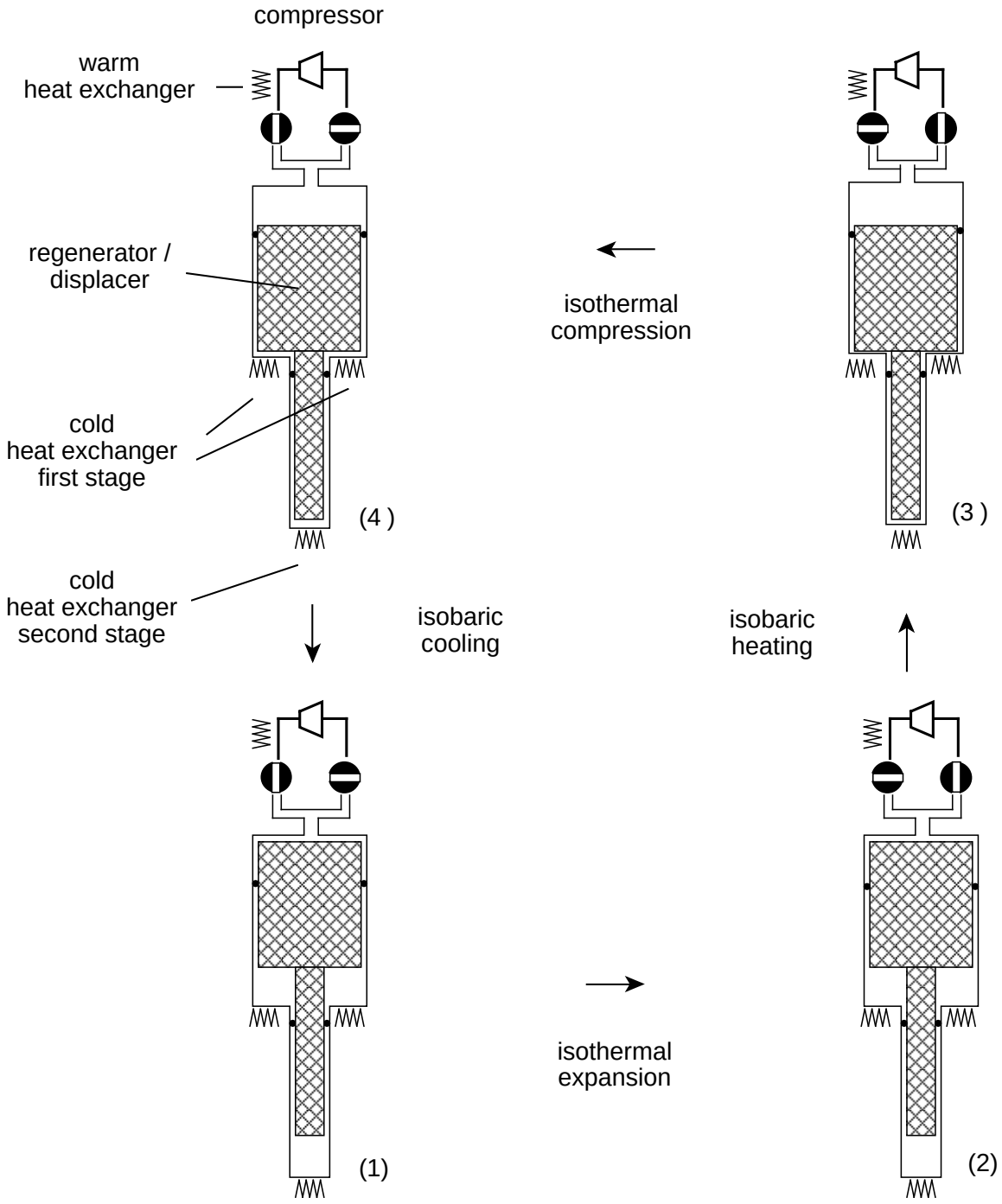


Figure 11: Schematic sketch of a Gifford-McMahon refrigerator

Between stage (1) and (2) the regenerator is at the upper position and the gas pressure is lowered from high pressure to low pressure by opening the low-pressure valve. Heat can be absorbed on the two temperature level heat exchangers. From stage (2) to stage (3") the regenerator is lowered, displacing and heating the gas from bottom to top and cooling the regenerator material. From stage (3") to stage (4") the whole volume is pressurised by opening the high-pressure valve. Finally between stage (4") and stage (1) the regenerator is moved up, displacing the gas down and cooling it, while heating the regenerator material. The heat rejection is done at ambient temperature at the compressor exhaust.

4.3 Pulse-tube refrigerators

Pulse tube refrigerators have seen an impressive development during the last ten years. With temperature records reached in this kind of refrigerators going down to 2.2 K . Nowadays it is possible to buy compact pulse tube systems delivering about one watt at 4.5 K and about 20 W at 50 K.

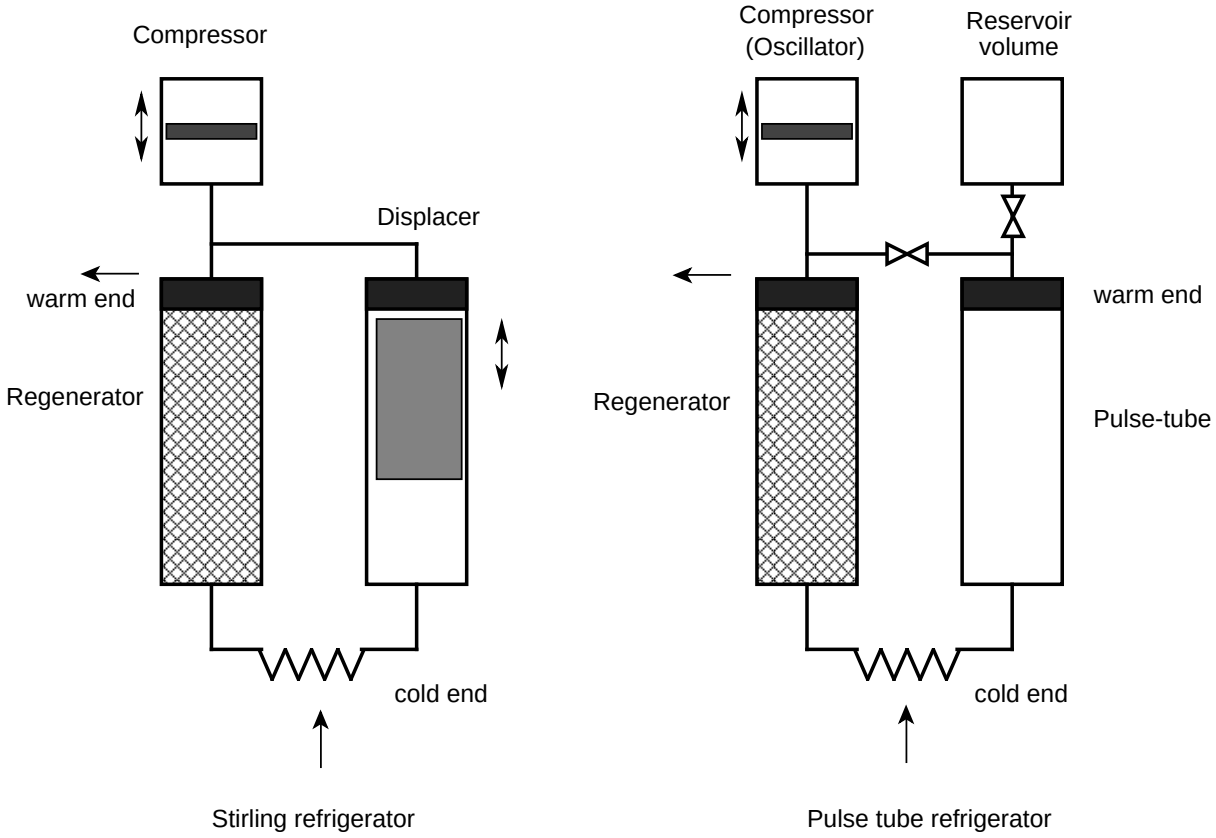


Figure 12: Pulse-tube refrigerator compared to Stirling refrigerator

The process can be compared to a Stirling cycle according to Figure 12. The function of the displacer is replaced by the oscillating gas volume in the pulse tube by correctly calibration the orifices that connect the pulse-tube to the reservoir volume and the oscillating source a phase angle of 90 degree between the pressure wave of the oscillator and in the pulse-tube can be reached.
Pulse-tube refrigerators have a distinctive advantage as they need no moving parts in the cold area. For small capacity applications down to 4.5 K these devices will certainly be increasingly used in near future.

5. CLAUDE-CYCLE REFRIGERATORS

All refrigerators using helium as working fluid are based on the Claude cycle if their range of capacity exceeds that furnished by the cryocoolers. Up to a wide range of capacity one usually finds two Brayton stages combined with one J-T stage. Figure 13 shows the typical process for such a refrigerator in a simplified LogT-s Diagram on the left side and the arrangement of the main process elements on the right side. The basic thermodynamic elements of the cycle and some rules for calculation are listed in the sections below.

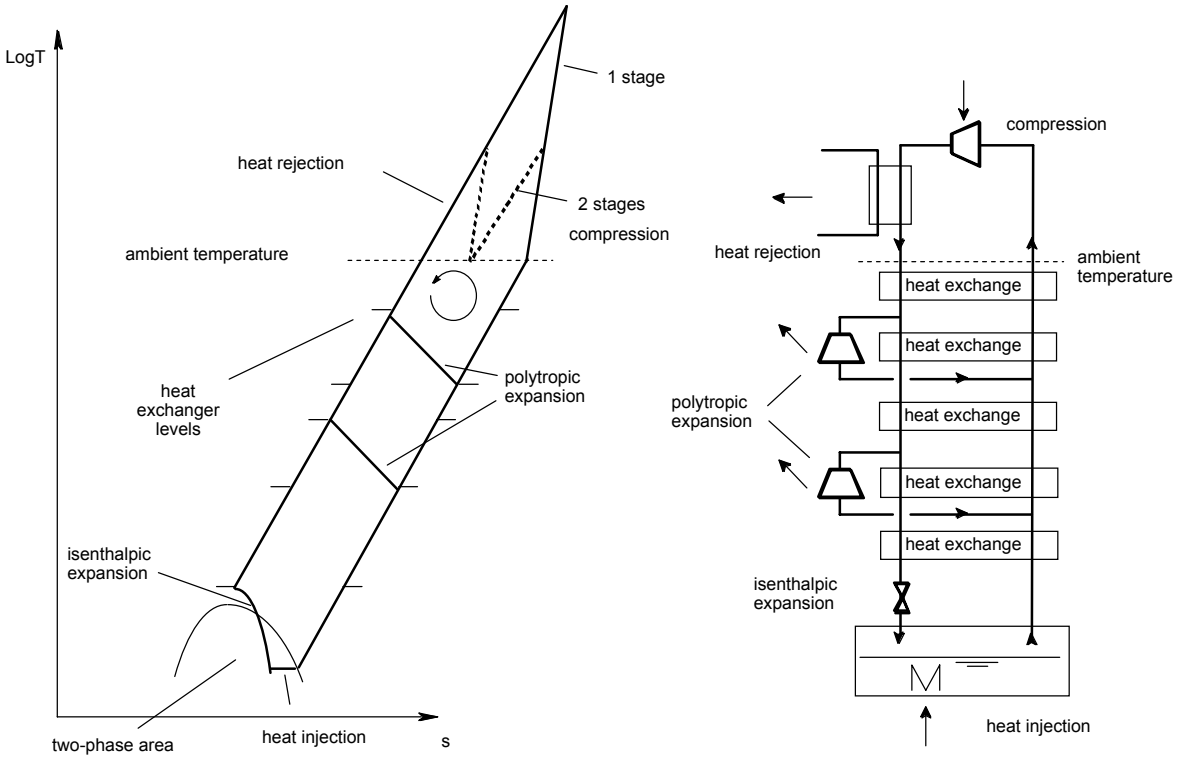


Figure 13: Typical helium refrigerator cycle combining two Brayton stages and one Joule-Thomson stage

5.1 Compression at ambient suction temperature.

The work input necessary for the refrigeration process is all injected during the compression of the helium gas. Compression above ambient temperature would ideally be isothermal. Figure 13 shows a polytropic compression, as can be achieved by a compressor without integrated cooling. Comparing the different area in the T-s diagram covered by single-stage and two-stage compression, the advantage of multi-stage compression is clearly illustrated. An increasing number of compression stages would gradually approach the isothermal limit.

Comparing the necessary compression work with the one for the isothermal compression leads to the isothermal efficiency of the compression stage. This expresses the "quality" of the compression.

The isothermal work input is defined by:

$$T \cdot \Delta S \text{ or } \dot{m} \cdot T \cdot \Delta s \tag{12}$$

In the case of an ideal gas as helium at ambient temperature, this can be written as:

$$m \cdot T \cdot R \cdot \ln \frac{p_{out}}{p_{in}} \tag{13}$$

with p_{out} and p_{in} being the discharge and suction pressure of the compression stage.

The isothermal efficiency is consequently defined as:

$$\eta_{isothermal} = \frac{\dot{m} \cdot T_{suction} \cdot R \cdot \ln p_{out} / p_{in}}{P_{el}} \tag{14}$$

5.2 Heat rejection to ambient

This heat rejection which re-cools the compressed fluid to ambient temperature, in general via the intermediate cooling medium water, represents the rejection of the total heat entering the cycle plus the work necessary for compression.

5.3 Heat exchange

Besides the heat exchange for the heat rejection above, the isobaric cooling and heating is approximated in the process heat exchangers exchanging the heat between the two fluid streams. In order to limit the entropy generation i.e. irreversibility losses, any heat exchange should ideally be done without temperature difference and pressure loss. Both goals are impossible to reach in the real case. One may "classify" the heat exchange process by calculating the exergetic efficiency of the heat exchanger balance following equation (10), or one may calculate the exergetic capacity created by summing up the individual exergy flows, and by this calculate the losses created by the heat exchange. The process of heat exchange is thermodynamically rather complex. Several characteristic figures for the classification and calculation of heat exchangers exist, like efficiency of the heat exchanger, the overall heat transfer coefficient times area (UA-value) and the Number of Transfer Units (NTU value). All these characteristic values only consider the thermal exchange performance but no losses created by friction.

The author considers the NTU value as most informative for cryogenic applications and therefore only this is explained below.

Figure 14 shows the temperature profile along the heat exchanger length for a balanced flow of two ideal gas streams. Following this sketch one can write for the heat transferred:

$$\dot{Q} = \dot{m}_W \cdot c_{pW} \cdot (T_{W1} - T_{W2}) \tag{15}$$

$$\dot{Q} = \dot{m}_C \cdot c_{pC} \cdot (T_{C1} - T_{C2}) \tag{16}$$

Defining the heat capacity flow as:

$$C = \dot{m} \cdot c_p \tag{17}$$

equations (15) and (16) can be reformulated to:

$$\dot{Q} = C_W \cdot (T_{W1} - T_{W2}) \tag{18}$$

$$\dot{Q} = C_C \cdot (T_{C1} - T_{C2}) \tag{19}$$

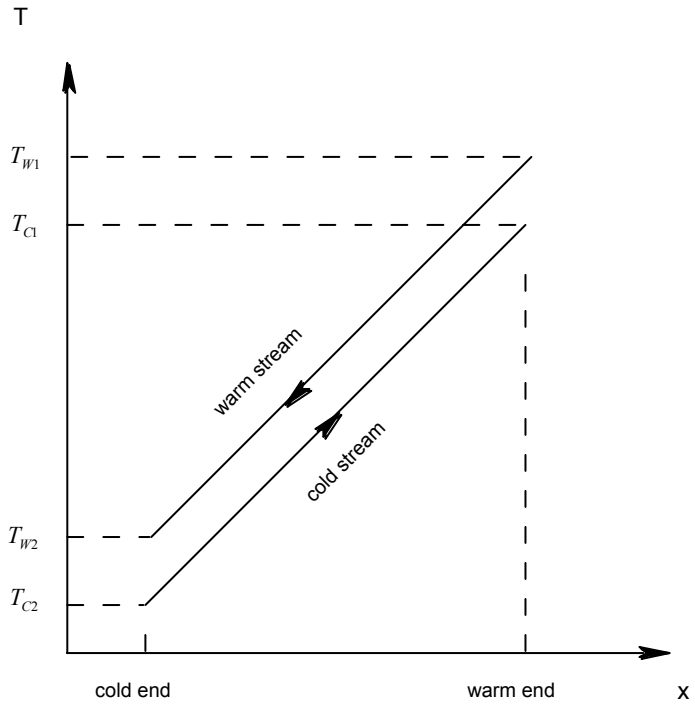


Figure 14: Heat exchange of two equal flows of ideal gas

This implies that the heat exchanger is thermally insulated to the environment which is sufficiently well achieved by placing it inside a vacuum insulated cold box. The transferred heat may also be expressed by using the overall heat transfer coefficient U, the exchanger area to which this coefficient is referred, A and the mean logarithmic temperature difference, which in this case equals the temperature difference of the two streams in any cross section of the exchanger.

$$\dot{Q} = U \cdot A \cdot \overline{\delta T} \tag{20}$$

With C_{min} being the smaller of the two values C_C and C_W, the NTU is a dimensionless value defined as:

$$\mathrm{NTU} = \frac{U \cdot A}{C_{\min}} \tag{21}$$

Or in general form as:

$$\mathrm{NTU} = \int \frac{U(T) \cdot dA}{C_{\min}(T)} \tag{22}$$

For an ideal gas case with c_p constant, we can following the equations above write:

$$\mathrm{NTU} = \frac{\Delta T_{Stream}}{\overline{\delta T}} \tag{23}$$

Where ΔT_{stream} represents the temperature difference of the stream with the smaller mass flow. In the case of a balanced heat exchanger, i.e. the mass flows of both streams are equal, the temperature differences are equal as well.

The NTU gives therefore a ratio between the temperature difference between the two streams and the one of which each stream can be cooled or heated. A standard NTU for the warm-end heat exchanger in a modern refrigerator would be around 25. In other words with a temperature of 5 K between the two streams one may cool or heat either stream for a temperature difference of 125 K. The inverse calculation can be done defining the characteristic NTU of a heat exchanger for which the performance is known. It should be remembered that ideal gas condition can be assumed in a helium refrigerator down to a temperature of 20 K for the common pressure range of up to 2.0 MPa.

5.4 Expansion machines

In the example of Figure 13 two stages of expansion machines are used, a typical compromise between efficiencies and cost for a common refrigerator for capacities up to 2 kW at 4.5 K.

Ideally the adiabatic expansion would also be isentropic; in real machines isentropic transformations can only be approximated. In analogy to what has been written for the compression, one defines an isentropic efficiency for the expanders in order to define their quality of thermodynamic performance.

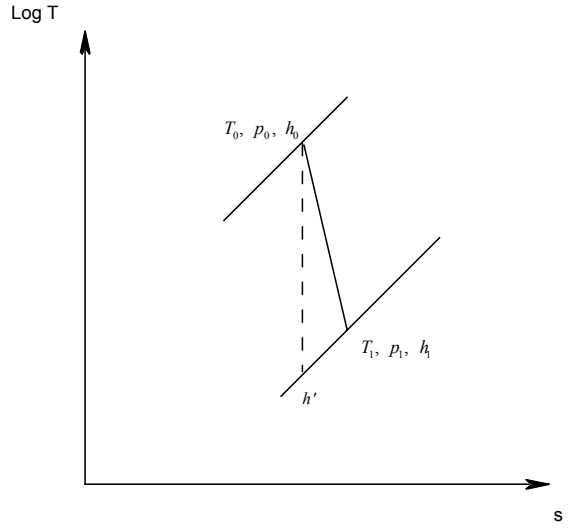


Figure 15: Polytropic expansion compared to isentropic expansion

Following the nomenclature of Figure 15, the isentropic efficiency is defined as:

$$\eta_{isentropic} = \frac{h_0 - h_1}{h_0 - h'} \cdot 100\,[\%] \tag{24}$$

It compares such the maximum possible isentropic enthalpy drop for a given pressure difference to the actual achieved enthalpy drop. Using ideal gas equations we can write for the isentropic efficiency:

$$\eta_{isentropic} = \frac{T_0 - T_1}{T_0\left(1 - \frac{p_1}{p_0}^{\frac{\gamma-1}{\gamma}}\right)} \cdot 100\,[\%] \tag{25}$$

The capacity extracted during the expansion process equals:

$$\dot{Q}_{\exp} = \dot{m} \cdot (h_0 - h_1) \tag{26}$$

or, again considering ideal gas and the isentropic efficiency:

$$\dot{Q}_{\exp} = \dot{m} \cdot c_p \frac{\eta_{isentropic}}{100} \cdot T_0 (1 - \frac{p_1}{p_0}^{\frac{\gamma-1}{\gamma}}) \tag{27}$$

The work extracted by the expanders is dissipated as heat at ambient temperature resulting in a pure loss for the refrigerator process. Ideally this work should be used to compress the gas. Considering that even in big helium refrigerators using more than two expansion stages the work dissipated from the expansion of the gas amounts to only about 2% of the work input into the compressors, the technical effort to recover this is not considered worth the possible gain in efficiency. For Brayton cycles working for air or nitrogen the work from the gas expansion is usually recovered as compression work above ambient temperature.

5.5 Isenthalpic (Joule-Thomson) expansion

The isenthalpic expansion can be generated in any thermally isolated restriction. In general this is represented by a valve. If this expansion takes place in the property region where the Joule-Thomson effect produces cooling, the valve is called Joule-Thomson (J-T) valve. Thermodynamically this expansion is a pure loss. It would be more efficient to use an expansion machine instead. Due to it's simplicity the J-T valve stays a common process element. Figure 16 shows a sketch of a process combining a Joule-Thompson expansion and a heat exchanger.

A simple energy balance immediately shows that the heat that can be absorbed in the helium bath is equal to:

$$\dot{Q} = \dot{m} \cdot \Delta h \tag{28}$$

With Q and Δh as indicated in Figure 16.

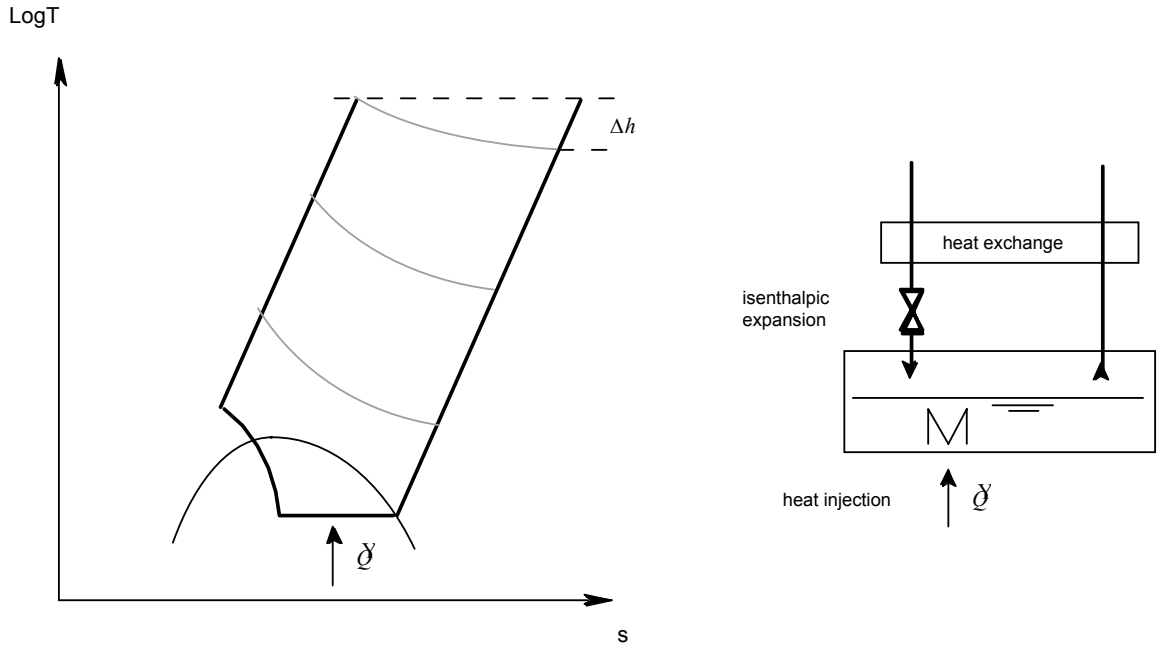


Figure 16: Joule-Thomson stage

Figure 16 further sketches the insenthalpic lines of the non-ideal gas in the property region where Joule-Thomson effect happens. On can imagine the whole system at the same upper temperature and

then cooling down following the sketched isenthalpic lines. This may illustrate that no gas that has ideal behaviour at room temperature can be cooled, or liquefied by a pure Joule-Thomson cycle as already mentioned in section 2.4. Table 4 lists the maximum inversion temperatures for some gases. It should be noted that close to the maximum inversion temperature this effect is very small. For helium it becomes exploitable below 20 K.

Table 4
Maximum Joule-Thomson inversion temperature for some gases

Gas	Maximum Joule-Thompson inversion temperature [K]
Helium	43
Hydrogen	202
Neon	260
Air	603
Nitrogen	623
Oxygen	761

5.6 The phase separator

A two-phase vessel allowing the flow of liquid helium to separate is integral part of any refrigerator. It may be incorporated in the refrigerator cold box, it may be a storage or transport dewar or it may be formed by the cryostat housing the load to be cooled. Due to the small difference in specific weight between liquid and vapour of helium (see Table 2) the separation of the two phases is not trivial. The design of the phase separator should avoid returning liquid droplets back into the process heat exchangers, as this would decrease their performance and consequently the performance of the refrigerator or liquefier.

6. BASIC COMPONENTS OF A REFRIGERATOR

Whereas in chapter 5 the different process elements were presented concerning their thermodynamic characteristic, this chapter will introduce the basic technical and mechanical aspects of the elements used for Claude-cycle refrigerators.

6.1 Cycle compressors

Cycle compressors for helium work from about ambient temperature and pressure conditions to pressures between 1.0 MPa and 2.0 MPa. The suction pressure is selected to be just above ambient level in order to avoid the intake of air impurities in case of small leaks in the system. The discharge pressure is generally not defined by the process in the refrigerator, that could profit from higher pressure ratios, but by the mechanical limits of the equipment in the refrigerator or the compressor itself. Common pressure ratios for Claude-cycle refrigerators are 16 to 20.

Due to the properties of helium its compression is accompanied by a high temperature rise. A pressure ratio of 16 would in an ideal isentropic compression process generate a temperature ratio of 3, i.e. the helium temperature would rise up to 900 K. Any deviation from the ideal increases this temperature difference. This must be considered when designing compressors for helium refrigeration.

Two types of compressors are generally used in helium refrigerators, piston compressors or screw compressors.

6.1.1 Piston compressors

Piston compressors for helium refrigeration are "dry" compressors, i.e. no lubricating agent is used for the piston-cylinder in order to prevent contamination of the process gas.

Small units may use synthetic (PTFE) piston rings that need regular replacement, bigger compressors usually use a labyrinth seal. Due to the high temperatures generated by the compression, a compressor has several pistons on one crankshaft, compressing the gas in series with a pressure ration of about 2 to 3. After each compression stage the gas is re-cooled in a gas-to-water heat exchanger. Due to the change in volumetric flow accompanied with the increasing pressure for each stage, the piston diameter decreases from low-pressure to high pressure stage.

The isentropic efficiency of the compression process is between 60 % and 70 %. The total isothermal efficiency of these machines is usually around 55%.

For Claude-cycle refrigerators piston compressors have been replaced by screw compressors mainly because these are more compact.

6.1.2 Oil-lubricated screw compressors

The compression process in a screw compressor is shown in Figure 17. The gas volume is enclosed in the screw pitch and continuously compressed until it is expelled. Though not evident at first sight, a screw compressor is a volumetric machine. Oil is injected at a pressure above discharge pressure and is mainly needed to ensure the gas-tightness for the compression process. The injection of oil has a secondary effect as some of the compression heat is exchanged between oil and gas so that the discharge temperatures are considerably lower than for a piston compressor thus getting closer to isothermal. A single screw may therefore easily reach compression ratios of 10 or higher.

The disadvantage of these machines is the presence of oil in contact with the helium as this creates an undesired contamination of the gas. The oil must therefore be removed by a series of coalescing filters and a final adsorber with activated charcoal filling. Only synthetic oil with low vapour pressure may be used in order to achieve an oil content in the range of parts per billion after the oil removal system. Though the cooling by oil during the compression process can decrease the compression work, the mechanical losses of the machine result in typical isothermal efficiencies of about 50%, comparable of those of a piston compressor.

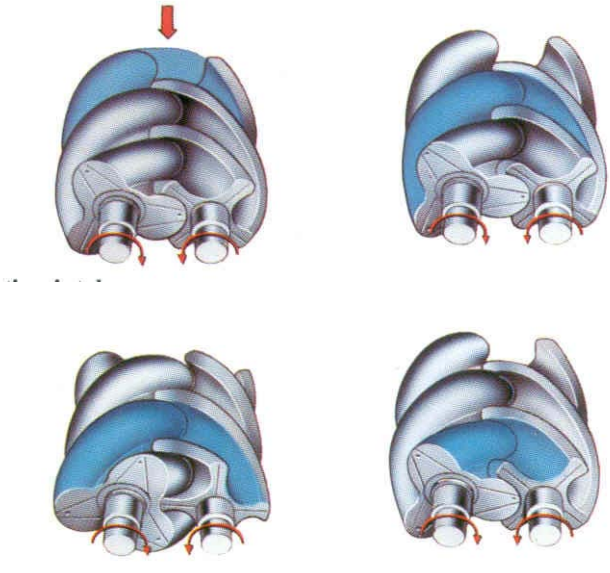


Figure 17: Compression process in a screw compressor (courtesy of AERZENER MASCHINENFABRIK)

6.2 Process heat exchangers

As this chapter concentrates on Claude-cycle refrigerators, only recuperative heat exchangers are considered.

As process heat exchangers one commonly define those that exchange heat between the different helium flow in the process below ambient temperature. The heat exchangers serving to reject the heat and work to ambient temperature are not considered.

A multitude of engineering solutions exist for the process heat exchangers especially for small capacity refrigerators. The general interest is reaching high thermal performance, i.e. large NTU-values and in the same time staying comparably compact in order to limit the size of the refrigerator cold box. This leads to heat exchangers having a large specific exchange surface, i.e. A large exchange surface per volume unit. This again consequently leads to the distribution of the different flow streams on several channels inside a heat exchanger. The heat exchanger design has to carefully consider an equal distribution of the flow over the different channels as any imbalance would degrade the thermal performance.

For refrigerators with capacities above 100 W at 4.5 K commonly only two types of heat exchangers are used, Coiled fin-tube exchangers and aluminium-plate-fin heat exchangers.

6.2.1 Coiled-fin tube heat exchanger

It consists of copper tubes with copper fins on the outside of the tube. The fins may be rolled on the tube or soldered on it. The finned tube is then wound around a mandrel or pipe in one or several layers. The whole assembly is then fitted into another tube of appropriate diameter.

The leads to a two-flow heat exchanger where the warm high-pressure stream flows inside the copper tube and the cold low pressure flows around the finned outer space.

These heat exchangers are by their geometry limited to small volume flows which confines them to smaller refrigerators or to the low temperature end of the refrigerator. It is further difficult to use them for more than two streams at different pressure, a feature that might be required in big refrigerators (> 1 kW at 4.5 K).

6.2.2 Aluminium-plate-fin heat exchanger

Aluminium plate-fin heat exchangers are composed by stacking layers of corrugated sheets, forming the fins, between flat sheets thus forming one stream channel. The core of this kind of heat exchanger then consists of a multitude of these channels. The whole arrangement is brazed under vacuum. Figure 18 shows a typical arrangement of this kind of heat exchanger.

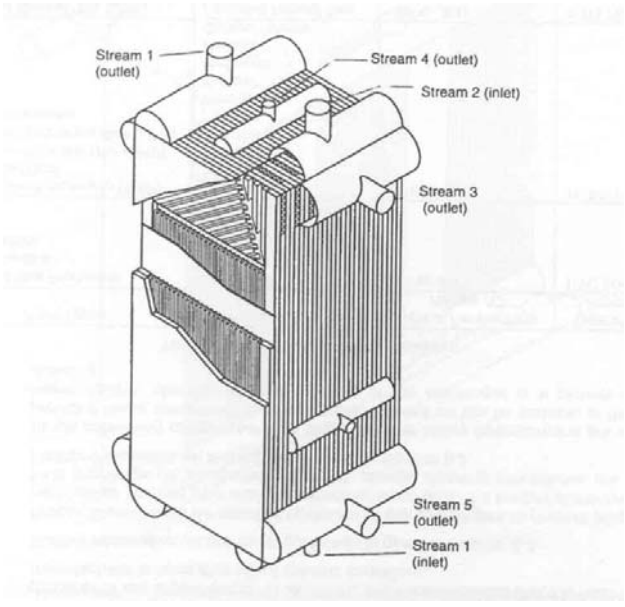


Figure 18: Plate-fin heat exchanger (courtesy of Linde AG)

The construction allows to incorporate a large number of streams at different pressures in one block. These units can operate at pressures up to 6 MPa and temperatures between 340 K and 2 K. With a specific heat exchanger surface of up 1000 m2/m3 these exchangers are very compact and efficient.

6.3 Expansion machines

6.3.1 Piston expanders

The first expansion machines used for helium refrigeration were reciprocating piston expanders that are still used today in some applications.

The isentropic efficiency that can be reached by these expanders typically ranges from 75% to 82%. Figure 19 shows a principle scheme of such expanders. Their use is limited by the maximum volume flow they can treat. The sealing packing of piston expanders require regular replacement in order to limit the leak rate losses. The energy extracted form the gas is recovered in an electric generator working as a brake.

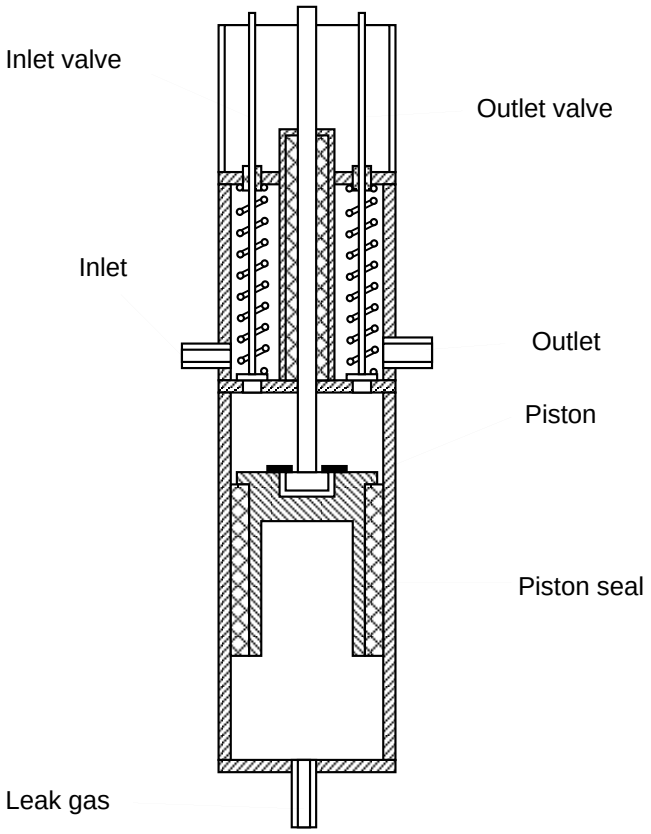


Figure 19: Schematic view of a piston expander

6.3.2 Turbine expanders

In turbine expanders the recovery of the work from the expansion of gas is done via a gas compressor mounted on the same shaft as the turbine wheel. The characteristic figure for the hydrodynamic reaction process is the blade-jet-speed ratio, comparing the speed of the turbine blades at the gas inlet with the theoretical maximum speed to which the gas can be accelerated by isentropic expansion. An example of a turbine expander is shown in Figure 20.

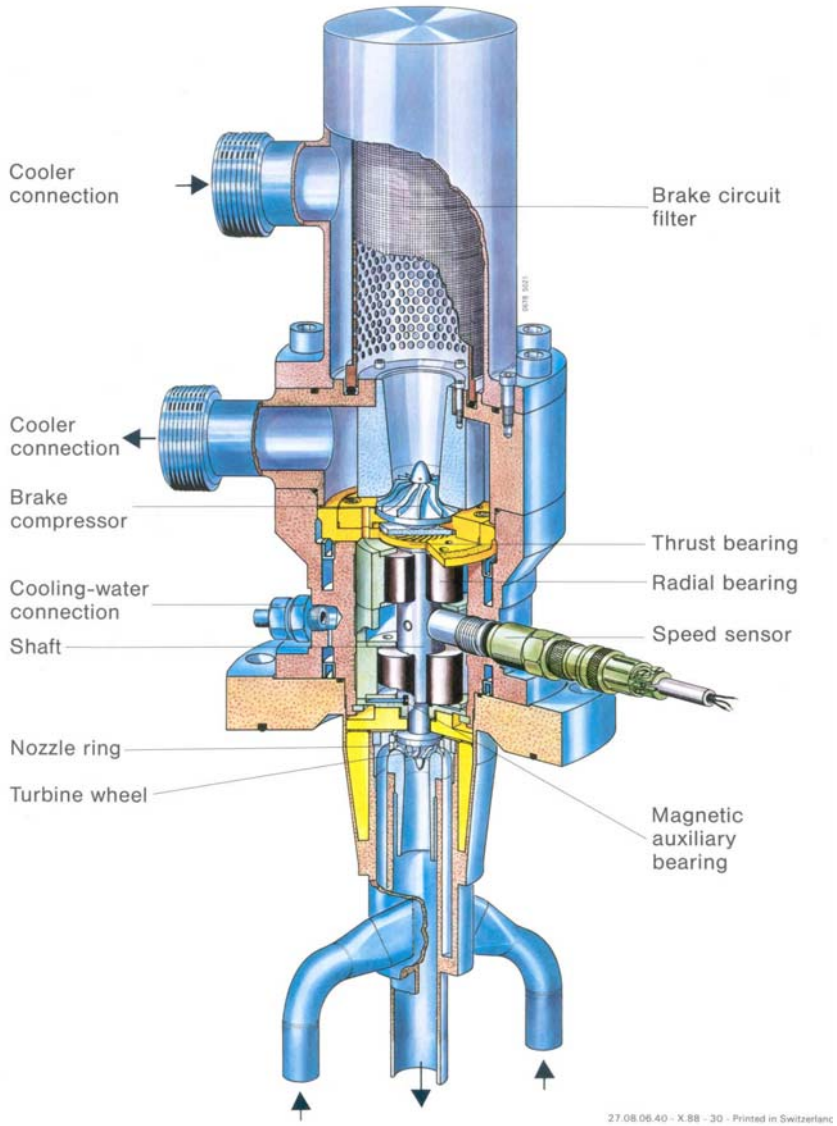


Figure 20: Dynamic gas bearing turbine (courtesy of Linde Kryotechnik AG)

The turbines have a maximum efficiency for a blade-jet-speed ratio of 0.55 to 0.7 depending on the wheel geometry.

The blade speed is defined by:

$$u = D\pi f \tag{29}$$

The jet speed by:

$$c = \sqrt{2(h_0 - h')} \tag{30}$$

For a "typical" turbine having a wheel diameter of 50 mm and running with inlet temperature at 80 K, inlet pressure 1.6 MPa and outlet pressure of 0.4 MPa, the optimum rotation speed would be 2300 Hz. Theses speed ranges cannot be covered by conventional bearings. The expansion turbines therefore use gas or oil bearings. As oil bearings have the inherent risk to contaminate the cold process parts only turbine expanders with helium gas bearings are nowadays used.

The obtainable isentropic efficiency of these expanders depends on its size i.e. the volume flow treated and typically ranges from 65 % for small expanders to 80 % for larger units.

7. A LARGE CLAUDE-CYCLE REFRIGERATOR: EXAMPLE HERA

Chapter 5 explained a standard Claude-cycle refrigerator with two expanders and a J-T valve For the range of equivalent capacity of 50 W to 1000 W at 4.5 K these cycles are commonly applied as being a good compromise between investment cost (equipment cost) and operation cost (efficiency) for higher capacities it is of advantage to design more complicated Claude-cycles with more expanders and a more sophisticated flow distribution in the heat exchangers.

As an example for a refrigerator designed for an accelerator using superconducting magnets the refrigerators for the HERA machine are selected as their process design has been well presented in several publications [9, 10] and especially in [11] a detailed exergy analysis was published,

Table 5
Design capacity for one HERA refrigerator

Cryogenic load		Process design capacity
Refrigeration at 4.3 K	[W]	6875
Vapor superheating at 0.11 MPa	[W]	275
Liquefaction at 4.3 K	[g/s]	22.5
Refrigeration at 40 K to 80 K	[W]	22000

The designed capacity for each HERA refrigerator is given in Table 5 showing a typical mixed duty including isothermal refrigeration and liquefaction at liquid helium temperature, in this case 4.3 K, and a non-isothermal refrigeration load for the cooling of thermal shields below 80 K. In order to calculate the equivalent capacity it is necessary to know that for refrigeration and liquefaction at 4.3 K subcooled helium at 4.5 K and 0.4 MPa is supplied and the vapour is returned at 0.11 MPa. The refrigeration at 40 K is supplied at a pressure of 1.8 MPa and returned at 1.7 MPa. Using the helium property data, one calculates a total exergetic capacity of 799 kW, equivalent to 12.2 kW isothermal refrigeration at 4.5 K.

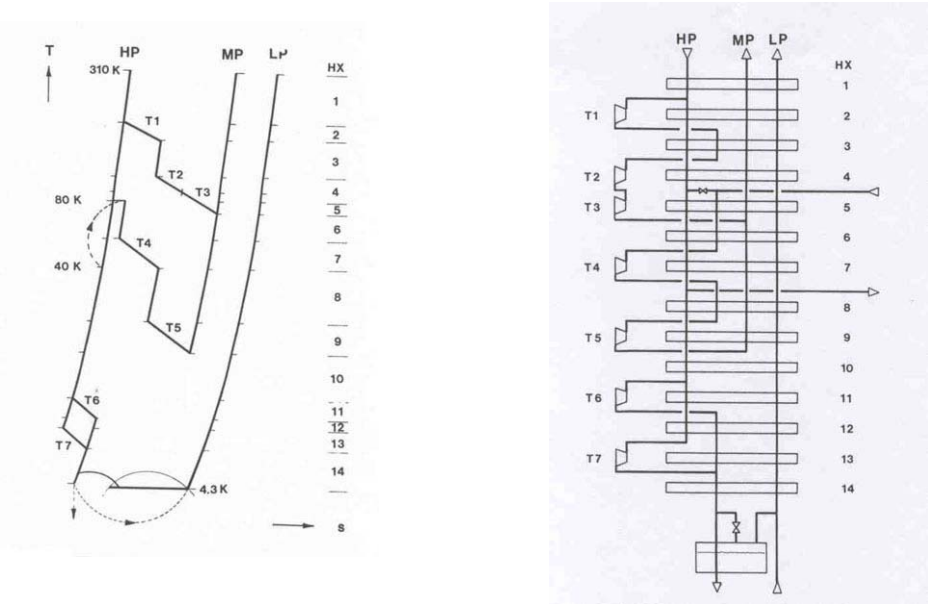


Figure 21: T-s diagram and simplified process flow diagram of the HERA refrigerators [from 11]

Figure 21 shows the T-s diagram and the simplified process flow diagram for one HERA refrigerator. The Claude-cycle process may be compared with the one sketched in Figure 13 it seems much more complicated, though it can quickly be reduced to a cycle with two Brayton stages including turbines T1, T2 and T3 in series for the first stage and turbines T4 and T5 for the second. Instead of expansion through a J-T valve the HERA cycle uses expansion in two turbines, T6 and T7 operating in paralles on slightly different temperatur levels. The basic features of the Claude cycle from Figure 13 are kept in the HERA design, though the arrangement allows a higher efficiency of the refrigerastion process.

In [11] a detailed exergy analysis of the process is presented, identifying the loss contribution of the individual process elements. The result is given in form of exergy flow diagrams that are shown in Figure 22 and Figure 23. One notices the important contribution to the overall loss by the helium compressors, and of the compression system in general where only about 50% of the exergy is passed to the cryogenic part of the process. This is given by the machines used for the helium compression and represents a figure that is generaly valid for any Claude-cycle process. As this figure could not be improved the designer of the refrigeration cycle made an effort to ameliorate the cryogenic part of the process which is reflected in the T-s Diagram. The cold box efficiency in this case is about 56% which can be considered as a very good value obtained in large refrigerators. Limit values that have been reached in comparable machines are about 60% exergetic efficiency.

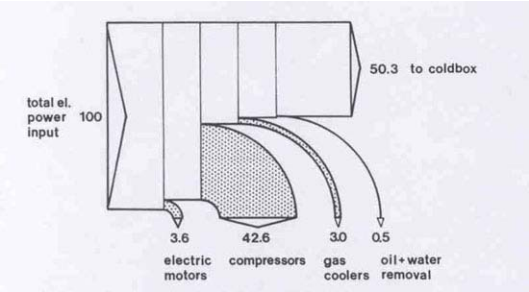


Figure 22: Exergy flow diagram of the compressor station for the HERA refrigerators [from 11]

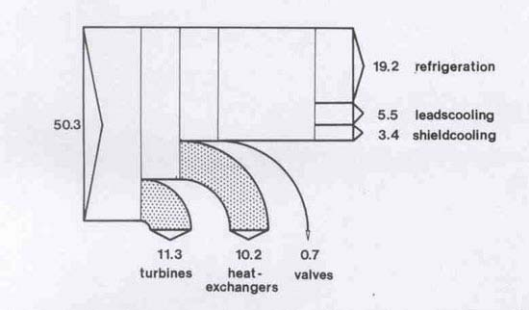


Figure 23: Exergy flow diagram of the cold box for the HERA refrigerators [from 11]

The final exergetic efficiency delivered at the cold box boundary is given to be 28.1% corresponding in a electrical power input of 2.84 MW into the system in the design case.

8. CRYOGENIC SYSTEMS

In the previous chapters helium refrigerators with various sizes and process cycles were presented stressing the emphasis on Claude-cycle refrigerators and their process elements typical for accelerator requirements. Such refrigerators are no stand-alone elements but are part of the cryogenic system that comprises:

i) The refrigerator itself including a compressor station with one or several compressors and an oil removal system in case oil lubricated compressors are used, and a cold box housing all the equipment of the process.

ii) A system to distribute the cold cryogen to the different users comprising vacuum insulated transfer lines and distribution valve-boxes. Depending on the extent of such a system it may be as costly as the refrigerator.

iii) Cryostats housing the cold load to be cooled.

iv) Instrumentation and process control for each sub unit and for the overall system .

This usually leads to a cryogenic system of extended complexity and hence cost, both in investment as in operation and maintenance [12].

9. CONCLUSIONS

The aim of the present article is to give a modest introduction into the physical background and technical realisation of helium refrigerators and liquefiers. Though the thermodynamics involved are now well established since a century, the realisation of especially larger units with their complex and often mainly empirically understood machinery can still today represent a considerable challenge. Like in many other fields of engineering cryogenic refrigeration requires a combination of practical experience and know-how with thorough understanding of the physics involved in order to successfully design cryogenic systems.

Mainly used as a service technology the basic functioning of cryogenic refrigeration should as well be understood by a user of cryogenic services in order to comprehend reactions, response times and particularities proper to such a system.

REFERENCES

[1] "cryogenics" Encyclopædia Britannica http://search.eb.com/eb/article?eu=28520 [Accessed July 2, 2002].

[2] W. Vinen, "Physics on Superfluid He", in these procedings

[3] V. Arp and R.D. McCarthy, "Thermophysical properties of helium4 from 0.8 to 1500 K with pressures to 2000 MPa", NIST technical note 1334 (1989(

[4] HEPAK (© copyright Cryodata Inc.) Helium property program written by V. Arp, R.D. McCarty and B.A. Hands

[5] W. Vinen, "Survey of Basic Thermodynamics", in these procedings

[6] G. Gouy, "Sur les transformations et l'equilibre en thermodynamique", C. R. Acad. Sc., Gauthiers-Villars, Paris 1889

[7] Z. Rant, "Exergie, ein neues Wort fur technische Arbeitsfahigkeit", Forsch.-Ing.-Wes. Nr.22, 1956

[8] R. A. Byrns and M. A. Green "An Update on estimating the cost of Cryogenic Refrigeration", Adv. Cryo. Eng. 43B: 1661 (1998)

[9] G. Horlitz, Refrigeration of a 6.4 km Circumference 4.5 Tesla Superconducting Magnet Ring System for the Electron-proton Colllider HERA, Proc. ICEC 10, 377, Butterworth (1984)

[10] H.R. Barton et al., "Central refrigeration System for the Superconducting HERA Proton Magnet Ring, Proc. Worklshop on Superconducting Magnets and Cryogenics, Brookhaven National Laboratory (1984)

[11] B. O. Ziegler, "Second Law Analysis of the Helium Refrigerators for the HERA Proton Magnet Ring, Adv. Cryo. Eng. 31, 693 (1986)

[12] Claudet, S; Gayet, P; Lebrun, P; Tavian, L and Wagner, U, "Economics of Large Helium Cryogenic Systems : experience from Recent Projects at CERN" Adv. Cryog. Eng. 46A (2000)

BIBLIOGRAPHY

B.A. Hands, Cryogenic Engineering, Academic Press (1986)

K.D. Timmerhaus and Th. F. Flynn, Cryogenic Process Engineering, Plenum Press (1989)

S. W. van Sciver, Helium Cryogenics, Plenum Press (1986)

J. G. Weisend II, Handbook of Cryogenic Engineerring, Taylor and Francis (1998)

HEAT TRANSFER

G. Vandoni
CERN, European Organization for Nuclear Research, Geneva, Switzerland

Abstract
Distribution of refrigeration power and limitation of heat leaks at every temperature stage, required by thermodynamic effectiveness, involve a careful analysis of heat transfer in a cryogenic environment. We review here the fundamental laws of the three modes of heat transfer focusing onto cryogenic operation conditions, and give useful data and examples of practical application for cryostat design.

1. INTRODUCTION

Thermodynamics' first and second laws introduce us to the concepts of heat Q [J] and temperature T [K]. Heat transfer is the energy interaction between two systems driven by their respective temperature difference. The objective of the study of heat transfer is to describe precisely how the temperature difference governs the rate of energy transfer, but it is again thermodynamics which identifies two classes of objectives in cryogenic engineering practice. Consider the efficiency of a thermodynamic engine, or the quantity of work W required to extract a quantity of heat Q produced at temperature T and rejected at a temperature T_a:

$$W \geq Q\left(\frac{T - T_a}{T}\right) \tag{1}$$

with the equality sign applying to the Carnot cycle. As a consequence of (1), both the capital and the operational cost of a cold installation will be determined by the heat loads on the system, with an increasingly high thermodynamic "value" for decreasing heat sink temperature. In other words, the lower the temperature, the smaller the heat loads on a cold system have to be maintained. One objective of thermal design will then be to work out the most effective **thermal insulation**, to reduce the heat transfer rate $\dot{q}$ between two temperature extremes T_A and T_B , which generally are fixed.

On the other hand, in a large system, to extract heat at a long distance from the refrigeration plant, refrigeration power must always be produced at a temperature lower than that of the furthest location. The additional temperature drop ΔT implies an increase ΔW in refrigeration power, or an increase in entropy generation. In the design of heat exchangers and sinks one will therefore aim at **enhancing heat transfer**, to evacuate the heat flux $\dot{q}$, this time usually a fixed quantity, across a minimized temperature difference T_A-T_B.

An extended treatment of heat transfer is beyond the scope of these lectures: specialized literature can be found in the bibliography. Limiting us to an introduction to cryogenic heat transfer, we shall review the fundamental laws and present materials' properties and examples of practical use for application to cryostat design.

There are three basic mechanisms for transferring heat: **conduction**, **convection** and **radiation**. In conduction, heat is transported inside solids or fluids at rest by atomic scale processes. In convection, macroscopic movement of a fluid produces the heat transfer. Finally, in radiative heat transfer, energy is transported by electromagnetic radiation, emitted by any surface at T>0 K, without need of matter's support. Fig.1 shows the temperature dependence of the three mechanisms (conduction and convection only for fluids and gases) below ambient temperature.

The basic laws for the flow of heat $\dot{Q}$ through an area A or onto a surface A in the three modes are:

Fourier's law for conduction $$\dot{Q} = -kA\, grad\, T \qquad (2)$$

General convection law, free or forced $$\dot{Q} = hA(T_w - T_f) \qquad (3)$$

Stefan-Boltzmann's law for radiation $$\dot{Q} = \sigma \varepsilon A(T_h^4 - T_c^4) \qquad (4)$$

where k is the thermal conductivity of the medium, h is the heat exchange coefficient, T_w the wall and T_f the fluid temperature, σ=5.67 10^{-8} W m^{-2} K^{-4} Stefan-Boltzmann's constant, ε an effective emissivity, T_h the warm surface and the T_c the cold surface temperature.

For a small ΔT, linearisation of the three laws permit to develop the analogy with an electrical circuit, where ΔT is the equivalent of the electrical potential difference and $\dot{Q}$ the equivalent of the electrical current. The thermal impedance R_{th} defined by this analogy, $R_{th} = \Delta T / \dot{Q}$, allows to describe complex systems for numerisation or to simplify them for qualitative evaluation.

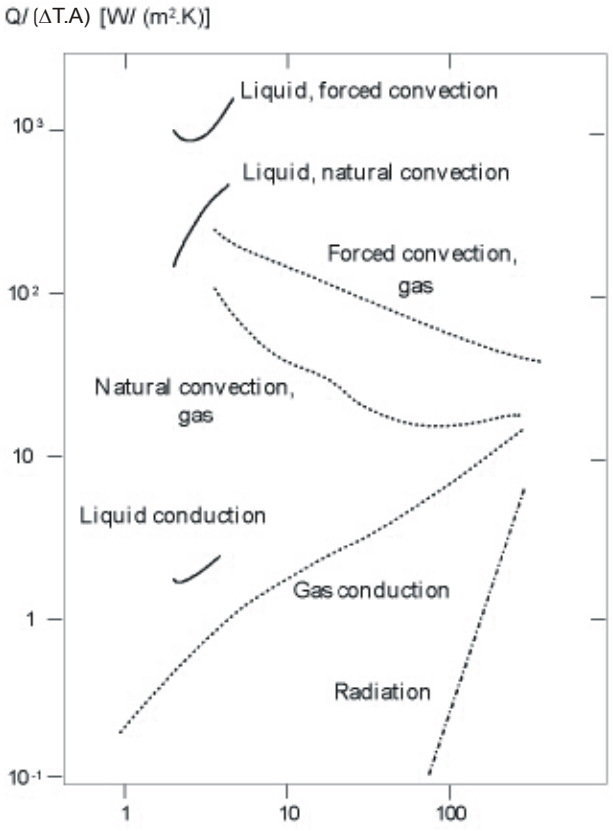


Fig. 1 Typical heat transfer rate per unit surface and temperature interval at temperature T, exchanged by **radiation**, by **conduction** through helium between two walls separated by 1cm, or by **convection** in helium around a horizontal stick of 1cm diameter.

2. CONDUCTION

2.1 General law

In most cases, the thermal conductivity k depends on temperature, resulting in Fourier's law being non-linear. In the one-dimensional case of time-independent heat flow through a cross section $A(x)$ over a length L, between T_h and T_c, integration of (2) yields:

$$\dot{Q} = \int_{T_c}^{T_h} k\, dT \Big/ \int_0^L \frac{dx}{A} = \frac{A}{L}\int_{T_c}^{T_h} k\, dT \tag{5}$$

Thermal conductivities and their integrals can be found in literature for most structural materials. We give some useful ones in table 1. They clarify the important gain in thermal losses which can be obtained by heat interception at intermediate temperature: depending on the material, the thermal conductivity integral between liquid nitrogen temperature and liquid helium temperature is a factor three to twenty smaller than the integral between ambient and liquid helium temperature.

Energy conservation in an infinitesimal volume, together with Fourier's law, permits us to write the one-dimensional differential equation of general conduction for time-dependent heat flow with heat source $\dot{Q}_h$:

$$\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + \dot{Q}_h = \rho C \frac{\partial T}{\partial t} \tag{6}$$

with ρ density and C specific heat of the solid. The first term accounts for longitudinal conduction, the second, for internal heat sources, and the third, for thermal inertia. We define the heat diffusivity $D=k/\rho C$, which allows to evaluate the characteristic propagation time τ of a thermal perturbation. Being r_o a characteristic dimension of the body, for instance the diameter of a sphere cooled through its surface, then the propagation time τ is $\tau \approx r_o^2/D$. We can now distinguish between an "early" regime, $t << \tau$, when a large temperature gradient exists in the external layers of the object while the internal temperature remains unperturbed, and a "late" regime, $t >> \tau$, when the internal body temperature is equal to the surface temperature and T evolves only with time. The advantage of this approach is that the late regime can be described by a simple exponential variation of the temperature with time,

$$T(t) - T(\infty) = (T_o - T_\infty)\exp(-\beta t) \tag{7}$$

with time constant $\beta = 1/R_{th}C_{th}$, R_{th} being the thermal impedance of the surface, for instance $R_{th}=1/hS$ for convective cooling, and $C_{th}=\rho CV$ the thermal inertia of the body.

Under some circumstances, not only the late regime, but the whole temperature history of the body may be described by (7). This happens when the internal thermal resistance of the body is much smaller than that at the surface, i.e. when the temperature difference established inside the body is always much smaller than the temperature difference between the surface of the body and its volume. Equating the surface heat flux

$$\frac{\dot{Q}_{surf}}{A} = h\Delta T_S \tag{8}$$

and the internal heat flux

$$\frac{\dot{Q}_{vol}}{A} = \frac{k\Delta T_{vol}}{r_o} \tag{9}$$

we see that the condition of $\Delta T_{vol} << \Delta T_{surf}$ leads to $Bi = hr_o/k << 1$. The number Bi is called Biot number. The quantity of heat released by the body upon cooling, for instance transmitted to a cooling fluid, is equal to the decrease in internal energy of the body. Equation 7 constitutes a useful approximation of the thermal equilibrium for any body.

2.2 Conduction in solids

Heat carriers in solids are delocalized electrons and phonons, the elementary excitations of the atomic lattice. At low temperature, the thermal conductivity associated with phonons varies with T^3, whereas for electrons, it varies linearly with T. Generally, good electrical conductors are also good thermal conductors, although the best thermal conductors, diamond and sapphire, are electrical insulators. Heat transmission at low temperature is hindered by electron-defect and phonon-defect scattering, limiting the mean free path of the heat carriers: as a result, a pure metal is a better conductor than its alloys. Treatments modifying the defect's content of a metal affect its thermal conductivity: cold work reduces it, annealing enhances it. At higher temperature, it is electron-phonon and phonon-phonon scattering which limit the mean free path, such that the differences between pure metals and dilute ones vanish.

Wiedemann-Franz-Lorenz's law yields an acceptable estimate for the thermal conductivity of pure metals at low temperature, based on the free electron approximation:

$$\rho_{el} k = LT \tag{10}$$

where L is the Lorenz constant, $L = 2.445 \cdot 10^{-8}$ $W \cdot \Omega \cdot K^{-2}$. For some metals, a more precise description of thermal conductivity over a broad temperature range has been developed by NBS [18,19]. Here, the thermal conductivity $k(T)$ is parametrized by the *RRR* value, the ratio between the electrical resistivity at ambient temperature and at 4K. The *RRR* parameter accounts both for chemical impurities and for extended lattice defects. Since the measurement of the *RRR* is more straightforward than that of ρ, this parametrisation proves very useful for practice.

Table 1
Thermal conductivity integrals for some structural materials

T2 [K]	Copper ETP [W cm^{-1}]	Aluminium 1100 [W cm^{-1}]	Austenitic st.steel [W cm^{-1}]	Glass [mW cm^{-1}]	PTFE [mW cm^{-1}]	G10 Composite [mW cm^{-1}]	Graphite [W cm^{-1}]
10	33.2	6.1	0.0293	6.81	4.4	0.31	0.025
20	140	27.6	0.163	20.2	16.4	1.56	0.15
30	278	59.2	0.424	36.8	32.3	4.23	0.48
40	406	96.2	0.824	58.6	50.8	9.04	1.1
50	508	134	1.35	84.6	71.6	17.1	2.9
60	587	170	1.98	115	93.6	29.3	4.6
70	651	202	2.70	151	116	48.4	6.9
80	707	232	3.49	194	139	67.5	9.8
90	756	258	4.36	240	163		13
100	802	284	5.28	292	187	130	22
120	891	330	7.26	408	237		32
140	976	376	9.39	542	287		45
160	1060	420	11.7	694	338		59
180	1140	464	14.1	858	390		74.
200	1220	508	16.6	1030	442	900	99
250	1420	618	23.4	1500	572		133
300	1620	728	30.6	1990	702	2650	251

For superconductors, the thermal conductivity in the normal state is essentially electronic, and hence varies linearly with temperature. Below the transition temperature, however, the thermal conductivity is determined entirely by phonons, varying with T^3. A ratio of $k_{normal}/k_{supra}=45/T^2$ is obtained for Pb and of $1/T^2$, for In.

Figure 2 displays the diffusivity for some structural materials. Since the specific heat usually decreases faster than the thermal conductivity, at low temperature the time to reach thermal equilibrium decreases markedly. The diffusivity of copper, equal to $D=1.12 \cdot 10^{-4}$ m^2/s at ambient temperature, increases to $D=2.2$ m^2/s at low temperature, that of stainless steel increases from $4 \cdot 10^{-6}$ m^2/s at ambient temperature to $1.4 \cdot 10^{-5}$ m^2/s at low temperature.

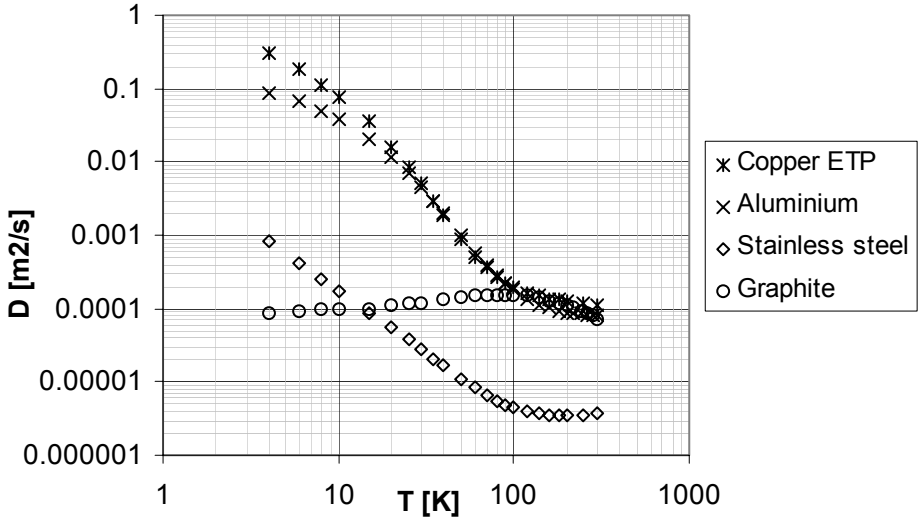


Fig. 2 Thermal diffusivity of selected materials.

The specific heat, needed for the calculation of the thermal diffusivity, can often be approximated by the Debye function, which describes the heat capacity per kilogram mole C_v in terms of Debye temperature θ_D.

$$C_V = 9R\left(\frac{T}{\theta_D}\right)^3 \int_0^{\theta_D/T} \frac{x^4 e^x}{(e^x-1)^2}\,dx \tag{11}$$

Values of the Debye temperature for several materials are given in table 2.

Table 2
Debye temperature for some elements and materials.

Material	θ_D [K]	Material	θ_D [K]
Aluminium	385	Iron	460
Carbon (graphite)	760	Nickel	440
Copper	310	Niobium	265
Gold	180	Titanium	355
Indium	105	Quartz	255

2.3 Conduction in gases and fluids

Heat transmission between two surfaces separated by gas obeys two different regimes according to the ratio between the mean free path of gas molecules λ and the distance L between the two surfaces. For $\lambda << L$, the *viscous* regime, heat transmission is described in terms of thermal conduction k, which is independent of pressure. The heat flux is therefore inversely proportional to wall distance. Decreasing the residual gas pressure, the *molecular* regime is reached as $\lambda >> L$. The molecules travel undisturbed from the warm to the cold surface and heat transfer becomes proportional to residual gas pressure and independent of wall distance.

Kinetic gas theory predicts the mean free path at temperature T [K], pressure p [Pa] and viscosity μ [Pa.s], for a gas of molar mass M [g/mole] as being equal to:

$$\lambda = 115 \cdot \frac{\mu}{p}\sqrt{\frac{T}{M}} \tag{12}$$

Table 3
Mean free path at 300K for different gases, in [cm]

[Pa]	$1x10^{-6}$	$1x10^{-4}$	$1x10^{-2}$	1	100
Ar	6300	63	0.63	$6.3\ x10^{-3}$	$6.3\ x10^{-5}$
Air	6100	61	0.61	$6.1\ x10^{-3}$	$6.1\ x10^{-5}$
N2	6000	60	0.60	$6.0\ x10^{-3}$	$6.0\ x10^{-5}$
He	$1.77\ x10^{4}$	177	1.77	$1.77\ x10^{-2}$	$1.77\ x10^{-4}$

In the viscous regime, the thermal conductivity coefficient, given in table 4 for different gases under atmospheric pressure, is

$$k = \frac{1}{3}\rho\left(\frac{8RT}{\pi M}\right)^{1/2}\lambda C_V \tag{13}$$

Table 4
Thermal conductivity k [W cm^{-1} K^{-1}] as a function of temperature under atmospheric pressure for several gases.

T [K]	^4He	H_2	O_2	N_2	Ar
300	$1.56\ 10^{-3}$	$1.92\ 10^{-3}$	$2.63\ \ 10^{-4}$	$2.60\ \ 10^{-4}$	$1.8\ \ 10^{-4}$
80	$0.64\ 10^{-3}$	$0.60\ 10^{-3}$		$0.76\ \ 10^{-4}$	$0.6\ \ 10^{-4}$(90 K)
20	$0.26\ 10^{-3}$	$0.16\ 10^{-3}$			
5	$0.10\ 10^{-3}$				

In the molecular regime, Kennard's law permits describes the heat transfer between two surfaces at temperature T_1 and T_2:

$$\dot{Q} = A_1\alpha\left(\frac{\gamma+1}{\gamma-1}\right)\left(\frac{R}{8\pi}\right)^{1/2}\frac{p}{\sqrt{MT}}(T_2 - T_1) \tag{14}$$

with $\gamma=C_p/C_v$, R the ideal gas constant, A_1 the surface receiving the heat flow and α an accommodation coefficient. The latter parameter, ranging between 0 and 1, defines the degree of thermal equilibrium between the gas molecules and the walls, and for simple geometries it takes a form analogous to the effective emissivity of radiation heat transfer:

$$\alpha = \frac{\alpha_1 \alpha_2}{\alpha_2 + \alpha_1 (1 - \alpha_2) \frac{A_1}{A_2}} \tag{15}$$

Close to condensation temperature, α=1, a value which can be used for practical purposes for most common gases. Light gases, however, may have much lower α values, as shown in table 5. In particular, the low accommodation coefficient of helium is responsible for the smaller heat transfer for helium with respect to hydrogen, in spite of its higher intrinsic thermal conductivity.

Table 5
Accommodation coefficients for some gases

T [K]	Helium	Hydrogen	Air
300	0.3	0.3	0.8-0.9
76	0.4	0.5	1
20	0.6	1	-

In general, liquids are bad thermal conductors, with the important exception of superfluid helium. Conductivities of some liquids are listed in table 6.

Table 6
Thermal conductivity for some common cooling fluids [mW/cm.K].

Water at 300K	6
Oil at 300K	1.3
Nitrogen at 77K	1.38
Helium at 4.2K	0.27

2.4 Contact resistance

When the apparent contact surface A_C between two solids is traversed by a heat flow $\dot{Q}$, a temperature discontinuity ΔT appears. The thermal boundary resistance R_H per unit area is defined as:

$$R_H = \frac{\Delta T}{\dot{Q} / A_C} \tag{16}$$

Two reasons explain why a contact resistance appears. On one hand, due to microscopic roughness, the two surfaces only touch at a limited number or spot-like points. On the other hand, phonon scattering occurs at the interface due to acoustic mismatch between dissimilar materials. Obviously, the contact resistance depends strongly on surface preparation, and can be decreased by applying an appropriate filler, for instance indium, vacuum grease, solder, or simply a thin plate of copper, between the two surfaces.

Some general trends are observable in contact heat resistance [21]:

a) *Contact resistance is inversely proportional to applied force and not to local pressure.* The contact conductance is actually almost independent of the apparent contact area, because the average contact spot size is approximately constant. Thus, increasing the load only changes the number of contact spots. In a first-order approximation, the slope of the contact conductance versus applied force is linear; saturation is reached at about 30 N at 300 K for Cu, Al, brass and stainless steel.

However, cycling loading experiments always result in hysteresis, due to plastic deformation of the local contact spots.

b) For contacts made in electrically conducting materials, the thermal conductance can be inferred from the electrical resistance.

c) Contact conductance is approximately proportional to T^n (n=3 for insulators, 1<n<2 for metals) at liquid helium temperature. With increasing temperature, $K_H(T)$ becomes rather linear with T and above 200K tends to a temperature independent value.

d) The thermal contact resistance is, in a crude approximation, proportional to the ratio of thermal conductivity to the microhardness.

Very complete reviews of contact resistance measurements are given in [20] and in [21].

2.5 Thermal Switches

Applications requiring a rapid cooldown but necessitating thermal isolation at operating conditions may make use of thermal switches, connecting the body with a refrigeration source. Some possible implementation techniques are given below:

a) Exchange gas: A gas which is non-condensable at the operation temperature is introduced, then pumped away once the operation temperature is reached. Since the volumes to be pumped may be large, this technique is rather lengthy.

b) Heat exchanger: The thermal contact is given by a small pipe in insulating material, filled with gas. Once equilibrium is reached, the pipe is evacuated. Small volumes reduce pumping times.

c) Superconducting switch: The thermal conductivity of some superconductors (Pb, In) is much larger in the normal state than in the superconducting state. Switching between the normal and the superconducting state is obtained by applying a magnetic field larger than the critical field to a superconductor strip inserted in a coil.

d) The thermal conductivity of polycristalline graphite varies with T^3 up to 100K, where it attains a value of 100W/m.K, a strong variation which can be exploited for thermal switching.

3. RADIATION

3.1 General laws

Any surface at finite temperature absorbs and emits energy in the form of electromagnetic radiation, depending on wavelength and direction. For an incident light flux P, a fraction *a* is absorbed, a fraction *t* (transmissivity) traverses the body, and a fraction *r* (reflectivity) is reflected back. Energy conservation requires the sum $a+t+r$ to be equal to 1.

A body is called black when its surface absorbs all of the incident flux, i.e. no energy crosses the body or is reflected back. Per definition, the absorptivity *a* is then equal to 1. The best real-life approximation of a black-body is a hole on the surface of a hollow cavity: radiation reaching the hole is trapped in the cavity and eventually absorbed after successive reflections on its walls.

The electromagnetic energy flux [W/cm^3] emitted per unit wavelength and surface by a black-body in the hemisphere as a function of wavelength is described by Planck's law:

$$E_{b,\lambda} = \frac{C_1 \lambda^{-5}}{\exp(C_2 / \lambda T) - 1} \tag{17}$$

The energy flux spectrum features a maximum for a wavelength which depends on temperature according to Wien's law:

$$\lambda_{\max} = \frac{2898}{T} \ [\mu\text{m/K}] \tag{18}$$

At T=300 K, the maximum energy flux is emitted at a wavelength of 10 μm, situating the radiation spectrum relevant for cryogenics in the far-infrared domain. For comparison, the sun emitting in the visible light range has a surface "black-body" temperature of above 5500 K.

Integrating Planck's law over all wavelengths, we obtain Stefan-Boltzmann's law, describing the total hemispherical emissive power, emitted in all directions by a black-body at temperature T:

$$E_b = \sigma T^4 \tag{19}$$

with σ=5.67 10^{-8} $Wm^{-2}K^{-4}$.

The net heat transfer rate between two black surfaces is obtained by computing the radiation emitted by the first surface and intercepted by the second one, and substracting the radiation emitted by the second surface and intercepted by the first. We get:

$$\dot{Q} = \sigma A_1 F_{12} (T_2^4 - T_1^4) \tag{20}$$

where F_{12} , the geometric view factor, is the ratio between the radiation leaving A_1 and being intercepted by A_2 and the radiation leaving A_1 in all directions. By definition, F_{12} is equal to the integral of the solid angle under which surface 1 sees surface 2. Values for several geometries are available in literature [4].

The intensity of radiation emitted at temperature T by a real surface is only a fraction of the corresponding intensity for a black-body. For example, relative to the energy flux $E_{b,\lambda}(\lambda,T)$ emitted by a black-body, a real surface emits directional, monochromatic radiation $E_\lambda(\lambda,T,\phi,\theta)=\varepsilon(\lambda,T,\theta)\, E_{b,\lambda}(\lambda,T)$ where $\varepsilon(\lambda,T,\theta) = 1$ is termed monochromatic directional emissivity. The total hemispherical emissivity ε of a real surface is defined by

$$\dot{q} = \sigma \varepsilon T^4 \tag{21}$$

The complicated function of direction and wavelength which usually is $\varepsilon(\lambda,T,\theta)$ can be approximated by considering it as being independent from wavelength (*grey-body*) and from direction (*diffuse body*). For a diffuse-grey body, Kirchhoff's law applies, stating that the absorptivity $\alpha(T)$ and the emissivity $\varepsilon(T)$ are equal. In practice, this law is used to estimate α from ε, provided that the incident radiation and the absorbing surface have the same temperature.

Within the diffuse-grey body approximation, we have now all the ingredients to write the expression for the radiative heat exchanged between two real enclosed surfaces A_1 and A_2 at temperature T_1 and T_2 and emissivity ε_1 and ε_2 , facing each other with a geometrical view factor F_{12} :

$$\dot{q} = \frac{\sigma (T_2^4 - T_1^4)}{\frac{1-\varepsilon_1}{A_1 \varepsilon_1} + \frac{1}{A_1 F_{12}} + \frac{1-\varepsilon_2}{A_2 \varepsilon_2}} \tag{22}$$

We can interpret this expression considering the blackbody emissive power difference $\sigma(T_2^4 - T_1^4)$ to act as a potential difference and the heat flux as a net current. The total resistance in the denominator is constituted by three resistances in series, the first and last being surface resistances to black-body radiation emission, and the intermediate being the purely geometrical resistance between two black bodies. Eq. 22 may be rewritten as

$$\dot{q} = \sigma \varepsilon_{12} (T_2^4 - T_1^4) \tag{23}$$

with ε_{12} an effective emissivity accounting for the emissivities of the two surfaces and their mutual view factor. Table 7 displays the value of ε for some common geometries.

Table 7
Effective emissivity for particular two-surface enclosures: cylinders and spheres may not be concentric.

Parallel plates ($A_1 = A_2 = A$)	$\dfrac{\varepsilon_1 \varepsilon_2}{\varepsilon_2 + (1-\varepsilon_2)\varepsilon_1}$
Annular space between long cylinders or spheres ($A_1 < A_2$)	$\dfrac{\varepsilon_1 \varepsilon_2}{\varepsilon_2 + \dfrac{A_1}{A_2}(1-\varepsilon_2)\varepsilon_1}$

Notice that if $A_1 << A_2$, or if $\varepsilon_2 \sim 1$, the effective emissivity reduces to ε_1. In this particular case, black-body radiation "fills" the cavity between the two surfaces and is collected by A_1 proportionally to its emissivity. Values of ε_{12} for particular surface conditions are given in [24] in this book.

3.2 Emissivity and materials

For real materials, emissivity depends on wavelength and direction, two facts neglected in the diffuse-grey body approximation. As an example, the emissivity of metals increases sharply above 60^o angle of incidence. However, many real surfaces satisfy reasonably well the diffuse-grey approximation and hence Kirchhoff's law. The hemispherical emissivities of some structural materials are given in table 8. We observe that clean, well-polished metallic surfaces have small emissivities, whereas non-metallic surfaces have high emissivities. In general, the best reflectors are also good electrical conductors. Actually, the emissivity and the electrical resistance of an ideal metal are related by Drude's law:

$$\varepsilon(\lambda, T) = 0.365 \sqrt{\frac{\rho(T)}{\lambda}} \tag{24}$$

The emissivity varies also with temperature: for metals, it decreases with decreasing temperature and is almost proportional to T at cryogenic temperature. For non-metals, the emissivity may decrease or increase with temperature, and in particular it increases with decreasing temperature for organic materials.

The emissivity of a coating is related to its surface resistance and not to the bulk resistance and it is worth noticing that the latter is strongly determined by the thickness of the film[1]. Reflectivity close to 1 is attained at cryogenic temperatures only for a film thickness above 40nm. Anodized aluminum has an emissivity close to 0.8. The emissivity of a superconductor is close to the one it has in its normal state. Finally, let's observe that the brilliant aspect a surface might display for visible radiation is not necessarily indicative of a low emissivity in the far-infrared region.

[1] Actually, the electron mean-free path is then largely determined by the distance between the two boundary surfaces, sort of macroscopic defects of the system.

Table 8

Emissivity of various materials as a function of temperature [9] [16].

Temperature [K]	4	20	80	300
Copper mechanically polished	0.02		0.06	0.1
Copper black oxidized				0.8
Gold			0.01	0.02
Silver	0.005		0.01	0.02
Aluminium electropolished	0.04		0.08	0.15
Aluminium mechanically polished	0.06		0.1	0.2
Aluminium with 7μm oxide				0.75
Magnesium				0.07
Chromium			0.08	0.08
Nickel			0.022	0.04
Rhodium			0.08	
Lead	0.012		0.036	0.05
Tin	0.012		0.013	0.05
Zinc			0.026	0.05
Brass, polished	0.018		0.029	0.035
St.steel 18-8	0.2		0.12	0.2
Glass				0.94
Ice				0.96
Oil paints any color				0.92-0.96
Silver plate on copper		0.013	0.017	
Aluminium film 400A on Mylar			0.009	0.025
Aluminium film 200A on Mylar			0.015	0.035
Nickel coating on copper		0.027	0.033	

3.3 Thermal radiation insulation: multi-layer insulation

"Floating", i.e. solely radiation cooled, reflective screens, interposed between a warm and a cold surface, effectively reduce radiative heat flow between the two surfaces. Assuming the same emissivity ε for the two juxtaposed surfaces and the n screens, it is easy to demonstrate that the heat flux is reduced by a factor $(n+1)$. A larger reduction is obtained with one single screen cooled at sufficiently low temperature, for instance by the vapor escaping from a cold bath. The concept of floating radiative screens is implemented in multi-layer insulation (MLI), an assembly of reflective films (usually aluminium or aluminized polyester film) separated by insulating spacers (polyester, glass-fiber nets or paper), operated under vacuum. Heat transmission through MLI can be seen as the parallel acting of solid conduction and radiation, resulting in an apparent conductivity:

$$k_{eff} = k_1 T + k_2 T^3 \tag{25}$$

The actual performance of real MLI depends on layer density, with a minimum at 10-20 cm^{-1}. Figure 8 shows the heat flux as a function of number of layers for a layer density of 15 cm^{-1}. Residual gas pressure affects the performance, as shown in figure 9. In practice, with some caution heat fluxes as low as 0.5-1 $W.m^{-2}$ are achieved in vacuum for boundary temperature 300K and 80K and 25-100 $mW.m^{-2}$ for boundary temperature 80K and 4K. Actual performance depends on field installation, where care has to be exerted to ensure low packing and avoid holes and bad joints which strongly compromise the efficiency of MLI. For complete application data and empirical formulas, see [22].

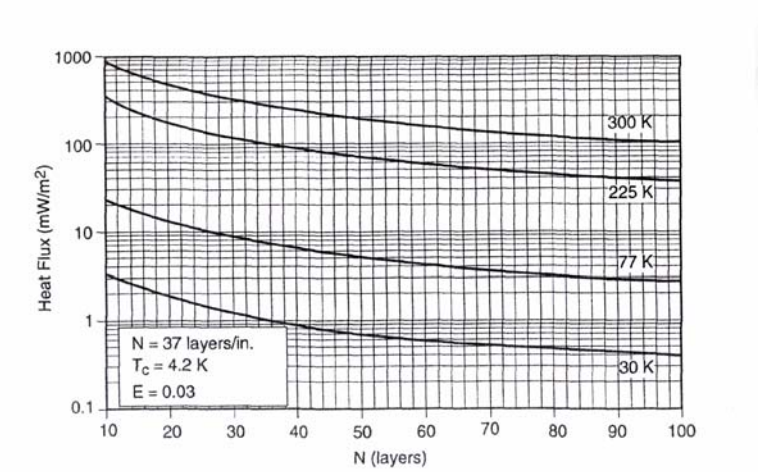


Fig. 8

Heat flux versus number of layers for different warm boundary temperatures, cold boundary 4.2 K, packing density 15 cm^{-1} [22].

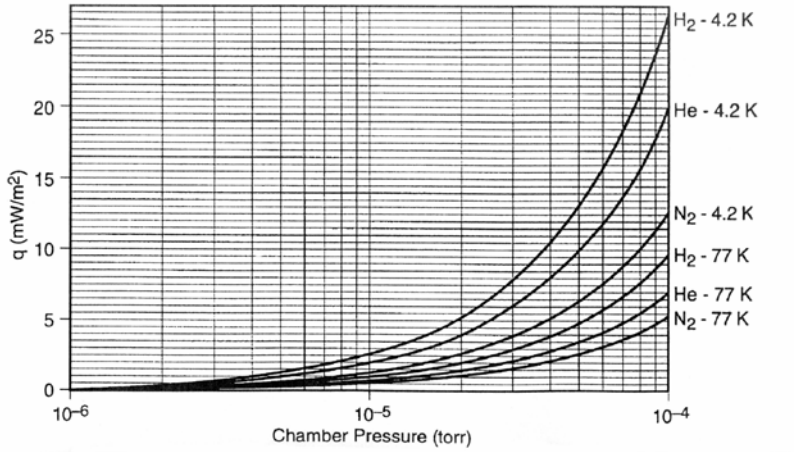


Fig. 9

Gas conduction versus chamber pressure for various interstitial gases and cold boundaries, hot boundary 300K, 37 layers, 15 cm^{-1} [22].

3.4 Screening of holes

A discontinuity in an intermediate radiation screen behaves as a blackbody at the temperature of the warm surface behind the screen; even a small hole has then a large impact on the cold losses. Depending on the respective areas and temperatures of hole and screen, it may be interesting to blacken the screen, for example by anodisation, to absorb radiation from the hole instead of redirecting it onto the cold parts. When low losses are sought, care should be applied to pipes, e.g. pumping lines, connecting directly the cold core to ambient: to reduce the waveguide effect for 300K radiation to the cold parts, black radiation screens are inserted in the tube in correspondence of a heat sink.

3.5 Passive cooling by radiators

Electromagnetic radiation towards a cold screen is exploited to obtain contact-less passive cooling of devices having a sufficiently large surface-to-volume ratio, provided the surface emissivity of devices and screen can be made close to 1. Black silicone paints compatible with high vacuum are available from space industry, and were for instance exploited for purely radiative cooling of mobile pick-up electrodes for the CERN antiproton collector [23]. Similar techniques are applied for cooling in space applications, where the cold source is the cosmic background radiation at 2.7K. Simultaneous screening of the solar radiation has then to be ensured.

4. CONVECTION

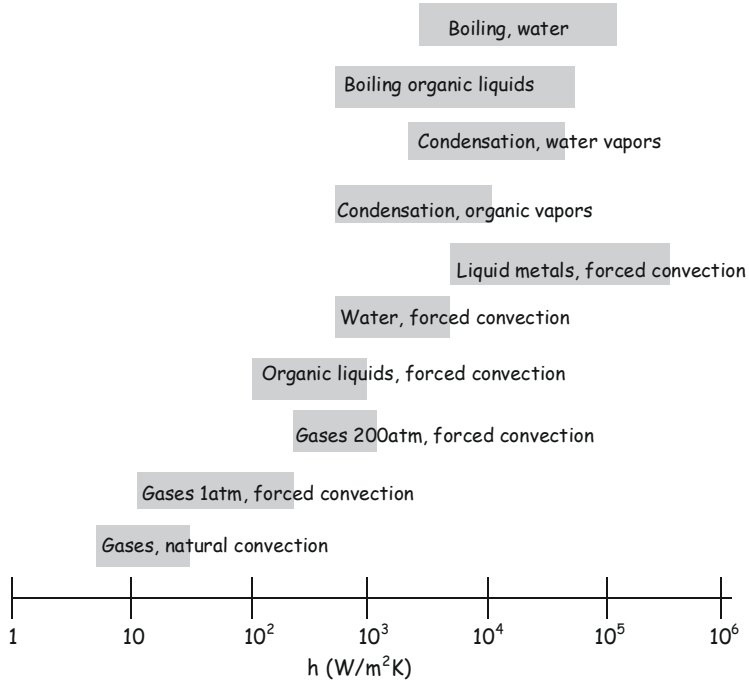


Fig. 11

Effect of fluid type and flow regime on the convection heat transfer coefficient.

Owing to the variety and complexity of problems encountered in convection, it is impossible to treat the subject here in detail. We will restrain us to how convection problems are solved through dimensionless groups describing fluid and flow properties, and give useful relations valid in the most common cryogenic cases.

The general law of convection heat transfer has the simple form:

$$\dot{q} = h(T_s - T_f) \tag{26}$$

where T_f is the temperature in the bulk of the fluid (internal flow) or at infinite distance from the wall (external flow) and T_s the temperature of the surface in contact with it. The engineering problem consists in determining the exchange coefficient h. Figure 11 displays ranges of h for fluid types and flows.

4.1 Dimensionless numbers and their physical interpretation

Convection exchange is characterized by the Nusselt number Nu, which compares h with that of purely conductive exchange; a characteristic dimension is chosen as an estimate of the thickness of the perturbed fluid layer.

The Reynolds number, comparing destabilizing inertia forces with stabilizing viscous forces, characterizes the flow regime: below a critical value, the flow is laminar, above, it becomes turbulent. Although the value of the critical Re separating laminar from turbulent flow depends also on the roughness of the tube, for most purposes a value of Re=2000 is taken as the upper limit for laminar flow, and Re=10^4 as the lower limit for turbulent flow. The heat transfer is small in the laminar regime, because of the small layer conduction and large in the turbulent one, due to the reduced dimension of the boundary layer. The Prandtl number, comparing the mass diffusion in the fluid with the heat diffusivity, is a characteristics of the fluid. Typically, Pr=0.7 for a gas, but also for normal liquid helium; further, Pr=7 for water at 20°C, Pr=2.5 for nitrogen at 77 K. The Grashof number, which appears in the description of natural convection, compares the buoyancy, due to gravity, with the viscous forces. Temperature differences along a vessel wall cause local flows due to density differences, which in turn enhance the heat transfer. The magnitude of Gr indicates wether these local

fluid flows are laminar or turbulent, with the associated low or high heat transfer rate. In cryogenic problems, large *Gr* and hence large heat transfer rates are to be expected, because of the large volume expansivity β.

Table 9
Dimensionless groups and their physical interpretation.

Group	Name	Definition	Physical interpretation
Re	Reynolds	$\frac{\rho V d}{\mu}$	$\frac{inertia\ force}{viscous\ force}$
Pr	Prandtl	$\frac{\mu C_p}{k}$	$\frac{momentum\ diffusivity}{thermal\ diffusivity}$
Nu	Nusselt	$\frac{hd}{k}$	$\frac{convection\ exchange}{conduction\ exchange}$
Gr	Grashof	$\frac{\beta g \Delta T d^3 \rho^2}{\mu^2}$	$\frac{buoyancy\ force}{viscous\ force}$
Ra	Rayleigh	*Gr Pr*	

d = characteristic dimension, ex. tube diameter or hydraulic diameter, μ = dynamic viscosity, β = coefficient of volume expansion, k = thermal conductivity, ΔT = temperature difference, ρ = density, C_p = specific heat at constant pressure, h = heat transfer coefficient

4.2 Natural convection

Natural convection occurs when the fluid movement has its origin in the expansion and ascension of heated sections of fluid. The useful dimensionless groups are *Nu*, *Pr* and *Gr*. A relation of the type *Nu=f(Gr,Pr)* permits to calculate the exchange coefficient *h*. Empirical forms

$$Nu = a\,(GrPr)^n = a\,Ra^n \tag{27}$$

where the Rayleigh number *Ra=Gr Pr* appears, have been developed for several typical configurations, some of which are displayed in table 10.

Table 10
Empirical relations for natural convection in some typical configurations

configuration	regime	limits	a	n
vertical, free surface	laminar	$5.10^3<Ra<10^9$	0.59	¼
	turbulent	$10^9<Ra<10^{13}$	0.13	1/3
horizontal, free surface		$Ra<10^3$	1.18	1/8
	laminar	$10^3<Ra<2.10^7$	0.54	¼
	turbulent	$2.10^7<Ra<10^{13}$	0.14	1/3

The characteristic dimension *d* to use in the calculation is:

the diameter	for a horizontal cylinder
the height	for vertical plates or cylinders
the smallest exchange dimension	for horizontal planes
the distance between walls	for enclosures

In the case of a common gas, it is easy to see that the exchange factor h is proportional to the square root of pressure, $h \sim p^{1/2}$, and inversely proportional to the square root of temperature, $h \sim T^{-1/2}$. Useful formulas valid for air close to ambient conditions, with d in [m], are given below:

for vertical plates $$h = 1.4\left(\frac{\Delta T}{d}\right)^{1/4} \quad [\text{Watt m}^{-2}\ \text{K}^{-1}] \tag{28}$$

for horizontal plates $$h = 1.3\left(\frac{\Delta T}{d}\right)^{1/4} \quad [\text{Watt m}^{-2}\ \text{K}^{-1}] \tag{29}$$

For free convection in cold gaseous helium at 1 bar, a Nusselt number ~3.65 can be assumed.

4.3 Monophase forced convection

Three dimensionless numbers are applied in problems of forced convection: *Nu, Re* and *Pr*. A relation of the type *Nu=f(Re,Pr)* permits to infer *h*, the exchange coefficient. Empirically, a variety of relations of the type

$$Nu = f(Re, Pr) = a\, F\, Re^m Pr^n \tag{30}$$

have been established for the most usual cases. We give in table 11 those which are more often used in cryogenic problems, valid for flow parallel to the duct.

Table 11
Empirical relations for mono-phase forced convection in some typical configurations.

Configuration	Regime	limit	a	m	n	F
horizontal plate	laminar	$10^3 < Re < 10^5$	0.66	½	1/3	1
	turbulent	$3 \cdot 10^5 < Re$	0.036	0.8	1/3	1
horizontal tube or annular space	laminar	$10^3 < Re < 2.1 \cdot 10^3$ $Re\ Pr\ d/L > 10$	1.86	1/3	1/3	$(d/L)^{1/3}$
	turbulent	$10^4 < Re < 1.2 \cdot 10^5$ $RePr\ d/L > 2.4 \cdot 10^5$	0.023	0.8	0.33	$1+(d/L)^{0.7}$

The relation describing exchange in *laminar* flow inside a horizontal tube takes the name of **Sieder & Tate** formula, the one valid for *turbulent* flow is the well-known **Colburn** formula.

The pressure drop along a segment L of a tube is determined by the friction force:

$$\frac{dp}{dx} = \frac{\rho v^2}{2}\frac{4f}{d} \tag{31}$$

with v flow velocity, f Fanning friction factor, and d hydraulic diameter, defined as $d=4A/p$, where A is the cross section of the duct and p its wetted perimeter. The relation between Re and f depends on the flow regime, on the geometry of the cross section of the duct, but also on the roughness of the surface.

Three useful relations for flow inside a duct of circular or annular cross section and smooth surface, are

laminar	Re<2000	$f = \frac{16}{Re}$	Poiseuille law	(32)
turbulent	$2 \cdot 10^3 < Re < 2 \cdot 10^4$	$f = 0.079\, Re^{-0.25}$	Blasius law	(33)
	$2 \cdot 10^4 < Re < 10^6$	$f = 0.046\, Re^{-0.2}$		(34)

An exchange problem in forced convection is solved by first determining the flow character from the calculation of *Re*, then *Pr* from the fluid characteristics, and then choosing the appropriate correlation formula for *Nu*. In case of doubt about the importance of natural convection, the calculation of the *Gr* or *Ra* might be helpful.

4.4 Convection with change of phase

4.4.1 Boiling heat transfer or pool boiling

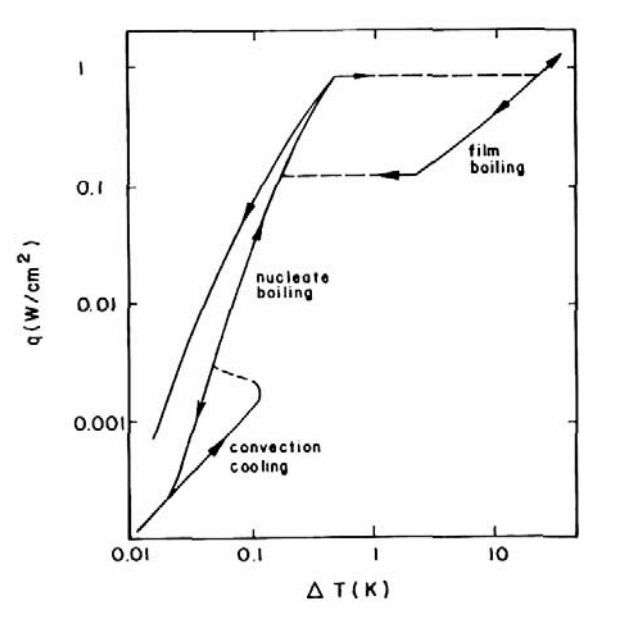


Fig. 12

Boiling heat transfer to liquid helium [8]

Boiling heat transfer occurs when the temperature of a solid surface is sufficiently higher than the saturation temperature of the liquid in contact with it. The heat flux versus temperature difference features a strongly non-linear behavior with hysteretic features. With increasing temperature difference, after a first regime where natural convection increases heat transfer, nucleate boiling appears, in which vapor bubbles are generated locally on the surface. The ascending currents promote heat transfer. A maximum is reached at the peak-nucleate boiling flux (PNBF), above which individual bubbles merge into a continuous film, which strongly limits the heat transfer. If the temperature difference is decreased from above the film-boiling crisis, recovery to nucleate boiling does not occur on the same portion of the curve. Figure 12 illustrates boiling heat transfer to liquid helium. The PNBF at the normal boiling point amounts to 1 Wcm^{-2} for helium and 10 Wcm^{-2} for nitrogen, with respective excess temperatures of 1 K and 10 K. For water, a heat flux of 100 Wcm^{-2} is obtained at 30 K excess temperature. In saturated superfluid helium, in spite of an exceptionally high heat conduction, owing to the small density of the liquid and the slope of the vapor pressure curve, the heat transfer switches to film boiling at heat flux values not much higher than those of normally boiling helium. A positive

consequence of the existence of a maximum heat flux in pool boiling is the limit it sets on the heat release from a warm object, like a quenching magnet, or in the cooling of a nitrogen spill on human skin.

4.4.2 Two-phase convection and instabilities

In two-phase convection, both natural and forced, the actual heat transfer rate is due to a combination of bubble formation and motion near the walls and the direct sweeping of the heated surface by the fluid. Due to significant density changes in the two-phase regime, or to pressure-drop change, instabilities and oscillations may occur. A guiding map to instabilities can be found in [6]. The frequency of oscillations may vary from 1/20 Hz to some hundreds of Hz. In the near-critical phase-space region, density-wave type oscillations may appear. Pressure waves increase locally the heat transfer rate, create an expansion of the fluid, decrease thereby its conductivity and hence the heat transfer rate. Oscillations may be avoided by applying the following practical rules:

- always maintain a low vapor quality throughout the flow
- avoid large differences in elevation along the circuit, which enhance natural buoyancy
- avoid downstreams flow restrictions, which are destabilizing, and introduce upstreams flow resrictions, which are stabilizing both on density-wave oscillations and on pressure-drop oscillations.

4.5 Refrigeration properties of cryogens

Table 12
Refrigeration properties of helium, nitrogen and water.

		He	N_2	H_2O
Normal boiling point	[K]	4.2	77	373
Critical temperature	[K]	5.2	126	647
Critical pressure	[bar]	2.3	34	221
Liquid density/ Vapor density*		7.4	175	1600
Heat of vaporisation*	[Jg^{-1}]	20.4	199	2260
Liquid viscosity*	[μPa s]	3.2	152	283
Enthalpy increase between T_1 and T_2 [Jg^{-1}]	T_1 = 4.2 K T_2 = 77 K	384	-	-
	T_1 = 77 K T_2 = 300 K	1157	228	-

*at normal boiling point

In particle accelerators and detectors, the most common cryogens are nitrogen and helium, whose properties are listed in table 12. They are characterized by the small gap in thermodynamic phase-space separating relevant operation domains. Consequently, the properties of the liquid and vapor phase in cryogenic systems are much closer to each other than is the case, for instance, for water. The latent heats of vaporization are low, the liquid and vapor density are similar, the liquid viscosity is low. A large enthalpy can be absorbed in cold vapor between the normal boiling point and ambient for helium, and to a lesser extent, for nitrogen. This refrigeration capacity proves very advantageous to cool structural elements and thermal shields.

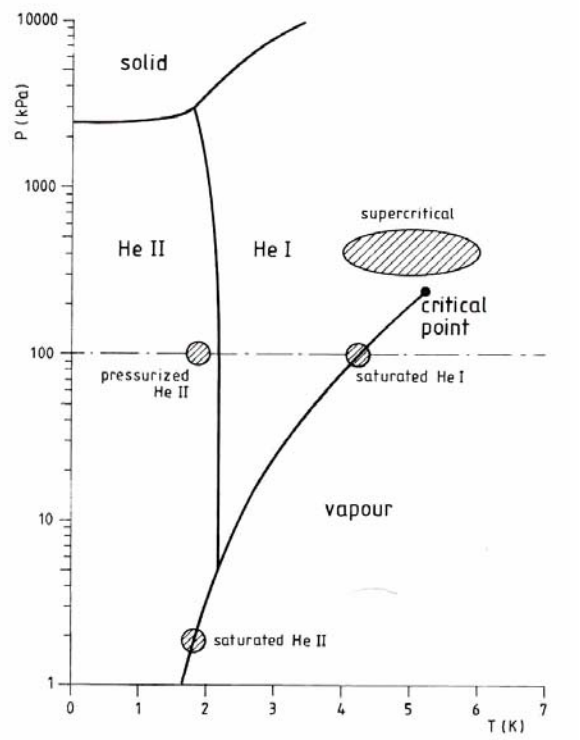


Fig. 13

Phase-diagram of helium; usual operating domains are shown hatched.

The phase diagram of helium is shown on figure 13. Instead of a triple point, a phase transition into a superfluid state appears. The transition leads to a macroscopic quantum state in which the liquid features unusual transport properties: an exceptionally high heat conductivity and a very low viscosity. The transition is a second-order one, i.e., no latent heat is associated to it. A peak appearing in the specific heat across the phase transition has given the transition line the name of lambda-line. Superfluid helium physics is presented in [8] and in this school, ref.[14], while superfluid helium technology is illustrated in [14bis].

Table 13
Comparison of operation domains of helium refrigeration.

Phase-diagram domain	Advantages	Drawbacks
Saturated normal helium bath	Monovariant: constant T independent of heat load	Large inventory Nucleate boiling crisis
Forced flow of saturated helium	Good heat transfer Small liquid inventory Joule-Thomson cooling	Flow instabilities Limited working range in T and p
Forced flow of supercritical helium	Easy modelling Small inventory No instabilities	Bivariant High pressure operation Joule-Thomson heating
Saturated superfluid helium bath	Low temperature operation Excellent heat transfer	Dielectric breakdown Sub-atmospheric operation Large gas volume
Pressurized superfluid helium bath	Low temperature operation Excellent heat transfer Large heat capacity	Cost of refrigeration Subatmospheric piping

Typical working domains of helium, highlighted on figure 13, are summarized in table 13. *Saturated normal helium* is a monovariant system, where the pressure above the bath determines the equilibrium temperature. Operating in *supercritical helium* permits to avoid two-phase problems like instabilities in the circuits, however, as a bivariant system, its temperature will depend on applied heat load. The use of *saturated superfluid helium* is undermined by the risk of dielectric breakdown and operation at subatmospheric pressure; *pressurized superfluid helium*, usually at atmospheric pressure or slightly above, is more advisable for cooling devices.

4.6 Thermal shielding potential of cold vapor

Taking advantage of the large enthalpy stored in the gases evaporating from a bath, the heat inleak is reduced by vapor-cooling cryostat necks, supports, current leads. In the case of a support or neck of length L and cross section A, assuming perfect exchange between it and the escaping gas (i.e., identical temperature of gas and wall), the heat balance equation yields

$$k(T)A\frac{dT}{dx} = \dot{Q} + \dot{m}C_p(T - T_1) \tag{35}$$

where Q is the residual heat inleak from the element into the liquid, at temperature T_1. If the regime is self-sustained, i.e., the evaporation of the liquid is entirely due to residual heat inleak, then

$$\dot{Q} = L_v\dot{m} \tag{36}$$

Integration of eq.35 yields

$$\dot{Q} = \frac{A}{L}\int_{T_1}^{T_2}\frac{k(T)dT}{1 + (T - T_1)C_p / L_v} \tag{37}$$

which, compared with the conduction integral eq.5, shows that vapor cooling brings to a substantial reduction in heat inleak. Table 14 illustrates the reduction of heat flow obtained by self-sustained helium vapor cooling for selected technical materials.

Table 14
Reduction of heat conduction obtained by self-sustained helium vapor cooling for technical materials.

	Thermal conductivity integral [W cm^{-1}]	Effective thermal conductivity integral [W cm^{-1}]
ETP copper	1620	128
OFHC copper	1520	110
Aluminium 1100	728	39.9
AISI 300 st.steel	30.6	0.92

Comparing the refrigeration characteristics of He and N_2, presented in table 8, it appears that vapor cooling by nitrogen is less convenient than by helium.

4.7 Neck cooling

We shall now discuss the calculation of the heat inleak from the vapor-cooled neck of an open cryostat. We call *A* the cross section of the neck, *l* its length.

The global losses on the cryostat are partitioned between losses from the neck, $\dot{Q}_1$, and from the rest of the environment, $\dot{Q}_R$. Liquid is evaporated by these losses with rate $\dot{m}$. Let us postulate perfect exchange between the walls of the neck and the escaping vapors. Applying Fourier's law between T_1 and T_2 and assuming a linear form for the thermal conductivity of neck's material,

$$k = a + bT \tag{38}$$

valid for instance for stainless steel, we obtain the relation (31) between the form factor of the neck, $Z=A/l$, and the ratio $\Phi = \dot{Q}_1 /(\dot{Q}_1 + \dot{Q}_R)$.

$$\dot{Q}_1 = \frac{Z\Phi}{C_p \beta}\left\{\left[a + b(T_1 - \frac{\Phi}{\beta C_p})\right] L\left[1 + (T_2 - T_1)\frac{\beta C_p}{\Phi}\right] + b(T_2 - T_1)\right\} \tag{39}$$

To account for the vapor volume remaining in the cryostat and replacing the evaporated liquid volume, we have applied a correction factor β,

$$\beta = (1 - \frac{\rho_{vap}}{\rho_{liq}}) / L \tag{40}$$

This correction may safely be neglected for most cryogens but helium (~15%), for which the difference between the liquid and vapor densities is small.

By letting Φ vary between 0 and 1, we can calculate $\dot{Q}_1$ and $\dot{Q}_R$ and plot $\dot{Q}_1 + \dot{Q}_R$ as a function of $\dot{Q}_R$. The utility of this approach is best understood with the help of fig. 14. For a particular value of $Z=A/l$, the graph can be used to evaluate if it is useful or not to reduce the residual losses by improving the overall insulation of the cryostat. It also shows that for each *Z* there is a critical value of $\dot{Q}_R$ such that the gases evaporated by $\dot{Q}_R$ exactly compensate the losses from the neck.

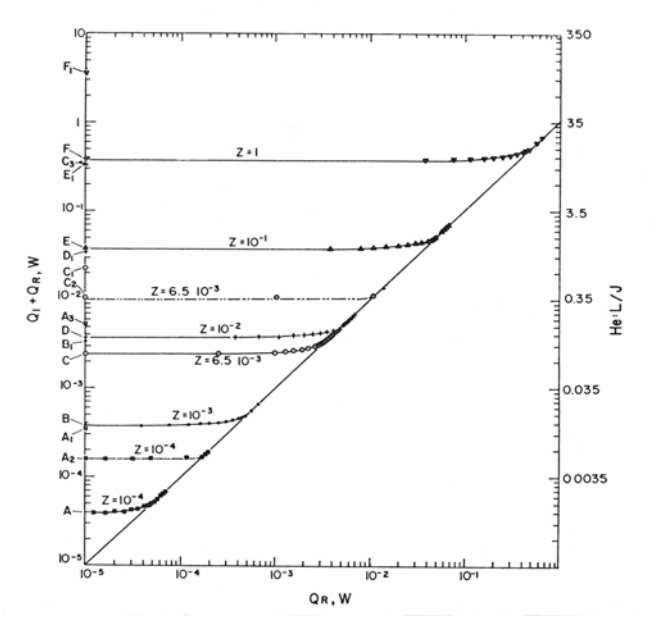


Fig. 14

Losses in a helium vessel with stainless steel neck as a function of $\dot{Q}_R$, for different values of $Z=A/l$ (cm). Solid lines, temperature difference between 77 and 4.2 K. Dashed lines: temperature difference between 300 and 4.2 K. From [11].

4.8 Current lead cooling

Vapor cooling also considerably reduces heat inleaks from electrical conductors feeding high currents into a cryogenic environment. Ensuring a good exchange between the current lead and the escaping vapor, and with appropriate optimization with current I of the aspect ratio l/A, both the heat leaks due to Joule heating and thermal conduction can be minimized.

For this type of calculations, it is always convenient to define the space variable increasing in the direction of heat flow. Heat balance in the elementary volume $A\,dz$ for a current lead of aspect ratio l/A yields:

$$\dot{Q}(z+dz) - \dot{Q}(z) - \frac{\rho I^2}{A} dz = 0 \tag{41}$$

the third term representing Joule heating in the volume. Integrating in $\dot{Q}$ between z and l under the hypothesis of constant ρ, we obtain

$$\dot{Q}(l) - \dot{Q}(z) = \frac{\rho I^2 (l-z)}{A} \tag{42}$$

On the other hand, from Fourier's law,

$$\dot{Q}(z) = -kA \frac{dT}{dz}$$

which we can plug in eq. 42, integrate a second time between 0 and l and T_c and T_w, to obtain:

$$\dot{Q}(l) = kA \frac{(T_w - T_c)}{l} + \frac{1}{2} \frac{\rho I^2 l}{A} \tag{43}$$

This expression has a minimum as a function of l/A,

$$(\dot{Q}_l)_{\min} = I\sqrt{2\,\rho k (T_w - T_c)} \tag{44}$$

$$\text{for } (l/A)_{\min} = \sqrt{2k(T_w - T_c)/\rho I^2} \tag{45}$$

Relaxing the hypothesis of constant ρ and applying Wiedemann-Franz-Lorenz law, $\rho k = L_{Lorenz} T$, we obtain a minimum heat flux $(\dot{Q}_l)_{\min} \approx 1$ mW/A almost independently of the conductor's material, for a vapor-cooled current lead operating between 300K and 4K. For comparison, a purely conduction cooled current lead would dissipate ~47 mW/A.

5. CONCLUSIONS

Lecture [24] in this same course shall illustrate the way for efficient thermal design of a cryostat for accelerator devices. In high-energy accelerators, however, the cryostat heat inleaks represent only a fraction of the total heat load budget. Table 15 shows the origin of heat loads in superconducting accelerator devices. Large dynamic loads stem from the beam, through synchrotron radiation, beam image currents, beam-gas inelastic scattering and beam spills. Synchrotron radiation is an important issue in leptonic accelerators. In the existing hadronic machines, usually the energy and the luminosity are such that static losses, i.e. the cryostat heat inleaks and the resistive dissipation, dominate over beam-induced losses. In the LHC, however, owing to the high energy and high luminosity of the beam, dynamic losses will largely override the heat inleak from the cryostat itself.

Table 15
Origin of heat inleaks in high-energy accelerators.

Cryostat heat inleak	Resistive dissipation	Beam-induced losses
radiation to cold surface	superconductor splices	synchrotron radiation
cold mass supports	wall resistance	beam-image currents
warm-to-cold feedthroughs	instrumentation	beam-gas inelastic scattering
		beam losses

6. BIBLIOGRAPHY

6.1 Heat transfer

[1] W.H.McAdams, Heat Transmission, (McGraw-Hill Int. Editions, 1958)

[2] W. Frost, Heat Transfer at Low Temperature, (Plenum Press, New York, 1975)

[3] L.Weil, Elements des echanges thermiques, (Gauthier-Villars, Paris, 1965)

[4] A.Bejan, Heat Transfer, (J.Wiley & Sons, New York 1993)

[5] L.C.Thomas, Heat Transfer Professional version, (Prentice Hall, Englewood Cliffs, 1993)

6.2 Cryogenics

[6] B.A.Hands, Cryogenic Engineering, (Academic Press, London, 1986)

[7] Handbook of Cryogenic Engineering, J.P.Weisend ed., (Taylor&Francis, Philadelphia 1998)

[8] S.Van Sciver, Helium Cryogenics, (Plenum Press, New York, 1986)

[9] R.B.Scott, Cryogenic Engineering, (Van Nostrand, New York, 1959)

[10] G.G.Haselden, Cryogenic Fundamentals, (Academic Press, London, 1971)

[11] R.Conte, Elements de Cryogenie, (Masson & Cie, Paris, 1970)

[12] Ph.Lebrun, Cryogenics, in CAS School Superconductivity in Particle Accelerators, CERN Yellow Report 89-04,

[13] F.Pobell, Matter and Methods at Low Temperatures, (Springer, Berlin, 1996)

[14] T.Vinen, The physics of superfluid helium, these proceedings

[14bis] Ph.Lebrun, L.Tavian, The technology of superfluid helium, these proceedings

[15] R.T.Jacobsen, S.G.Penoncello, E.W.Lemmon, Thermodynamic Properties of Cryogenic Fluids, (Plenum Press, New York, 1997)

6.3 Material's properties

[16] Y.S.Touloukian, Thermophysical Properties of Matter, (Plenum Press, New York, 1995)

[17] G.Hartwig, Polymer Properties at Room and Cryogenic Temperatures (Plenum Press, New York, 1994)

[18] N.J.Simon, E.S.Drexel, R.P.Reed, Properties of copper and copper alloys at cryogenic temperatures, NIST Monograph 177, (1992)

[19] V.D.Arp, Electrical and thermal conductivities of elemental metals below 300K, Thirteenth Symposium on the Thermophysical Properties, June 22- 27, 1997, Boulder Colorado, U.S.A.

6.4 References

[20] L.J.Salerno and P.Kittel, in Handbook of Cryogenic Engineering, J.P.Weisend ed., (Taylor&Francis, Philadelphia 1998), 164

[21] E. Gmelin, M.Asen-Palmer, M.Reuther, R.Villar, J.Phys D: Appl.Phys. 32 (1999) R19-R43

[22] T.Nast, in Handbook of Cryogenic Engineering, J.P.Weisend ed., (Taylor&Francis, Philadelphia 1998), 186

[23] Ph.Lebrun, S.Milner, A.Poncet, Adv. Cryog. Eng. 31 (1986) 543

[24] P.Lebrun, Design of a cryostat for superconductoing accelerator magnet: the LHC main dipole case, these proceedings.

DESIGN OF A CRYOSTAT FOR SUPERCONDUCTING ACCELERATOR MAGNET: THE LHC MAIN DIPOLE CASE

Ph. Lebrun
CERN, European Organization for Nuclear Research, Geneva, Switzerland

Abstract
As an application of the lectures on heat transfer, this tutorial presents the simplified thermal design of a cryostat for a superconducting accelerator magnet. Starting from functional requirements and design constraints, and using standard formulae and typical data for heat transfer at low temperatures, we sketch an engineering solution for the cryostat and evaluate its thermal performance. We then present how this design is implemented in a real cryostat, industrially produced in series of 1250 units.

1. INTRODUCTION

Although an ancillary component, the cryostat housing a superconducting accelerator magnet fulfils several important roles. First of all, it provides adequate thermal insulation for low-temperature operation of the magnet, and in some cases of the associated cryogenic pipework and beam vacuum system. Equally important, the cryostat constitutes the mechanical interface of the superconducting magnet to the outside world. It must therefore rigidly support and precisely position the cold mass with respect to the external alignment fiducials, in a reproducible fashion throughout steady and transient regimes of operation. The cryostat may also perform other functions, such as magnetic screening of the stray field or radiation shielding of the beam halo. It is therefore clear that cryostat design and optimisation must integrate all boundary conditions and constraints set by superconducting magnet, cryogenic and other relevant accelerator systems.

2. FUNCTIONAL REQUIREMENTS AND DESIGN CONSTRAINTS

For the purpose of this tutorial, the basic thermal design and configuration of the LHC main dipole cryostat can be derived from the following simplified set of requirements and constraints.

The cryostat must house a superconducting magnet "cold mass" operating in superfluid helium at 1.9 K.

The cold mass is contained inside a stainless-steel helium vessel with an external diameter of 0.6 m, length 15 m.

The cold mass is 30,000 kg, i.e. 2000 kg/m length.

Due to external space constraints (installation in accelerator tunnel), the outer vessel of the cryostat must have an external diameter not exceeding 1.0 m.

The cold mass must be supported inside the cryostat in a stable and stiff fashion.

The maximum heat load budget at 1.9 K in normal operation is 3 W, i.e. 0.2 W/m length.

In view of possible degradation of the insulation vacuum, a design solution which limits the increase of heat inleak is favoured.

3. HEAT TRANSFER DATA

The laws and formulae of heat transfer by solid conduction, residual gas conduction and radiation are given in the lectures on "Heat transfer" [1].

Typical values for thermal conductivity integrals, emissivity of technical materials and heat fluxes across different insulation systems at low temperatures are recalled respectively in Tables 1, 2 and 3. These Tables should primarily be used to orient design choices; they however have no claim to exhaustivity, and thermal insulation solutions based on other materials, surface finishes and temperature levels may also be considered.

Table 1
Thermal conductivity integrals of selected materials [W/m]

From vanishingly low temperature up to	20 K	80 K	290 K
OFHC copper	11000	60600	152000
DHP copper	395	5890	46100
1100 aluminium	2740	23300	72100
2024 aluminium alloy	160	2420	22900
AISI 304 stainless steel	16.3	349	3060
G-10 glass-fiber/epoxy composite	2	18	153

Table 2
Emissivity of technical materials at low temperatures

	Radiation from 290 K Surface at 77 K	Radiation from 77 K Surface at 4.2 K
Stainless steel, as found	0.34	0.12
Stainless steel, mech. polished	0.12	0.07
Stainless steel, electropolished	0.10	0.07
Stainless steel + Al foil	0.05	0.01
Aluminium, as found	0.12	0.07
Aluminium, mech. polished	0.10	0.06
Aluminium, electropolished	0.08	0.04
Copper, as found	0.12	0.06
Copper, mech. polished	0.06	0.02

Table 3
Typical heat fluxes at vanishingly low temperature between flat plates [W/m^2]

Black-body radiation from 290 K	401
Black-body radiation from 80 K	2.3
Gas conduction (100 mPa He) from 290 K	19
Gas conduction (1 mPa He) from 290 K	0.19
Gas conduction (100 mPa He) from 80 K	6.8
Gas conduction (1 mPa He) from 80 K	0.07
MLI (30 layers) from 290 K, pressure below 1 mPa	1-1.5
MLI (10 layers) from 80 K, pressure below 1 mPa	0.05
MLI (10 layers) from 80 K, pressure 100 mPa	1-2

4. A DESIGN SOLUTION

We present a simplified version of the reasoning which has permitted to roughly size the heat load budget of the LHC [2] and led to the construction of the first superfluid-helium cooled accelerator magnet cryostat [3], thus establishing the cryogenic feasibility of the project.

4.1 A two-dimensional problem

All reasoning will be done per unit length of cryomagnet, which reduces to a two-dimensional problem in the transverse plane. The lumped heat loads, e.g. through the supports, are then calculated as distributed along the length of the cryostat.

4.2 Basic configuration

The mechanical members supporting the 1.9 K cold mass from the room-temperature outer vessel should be such that the heat conduction path is made as long as possible, while still preserving sufficient rigidity, stability and positioning accuracy.

Two basic configurations are considered:

- support posts in compression (Fig. 1.a),
- tie rods in tension (Fig. 1.b)

at regular intervals along the length of the cryomagnet.

The thermal profile along the support members develops over 0.2 m in the first case, and 0.5 m in the second one.

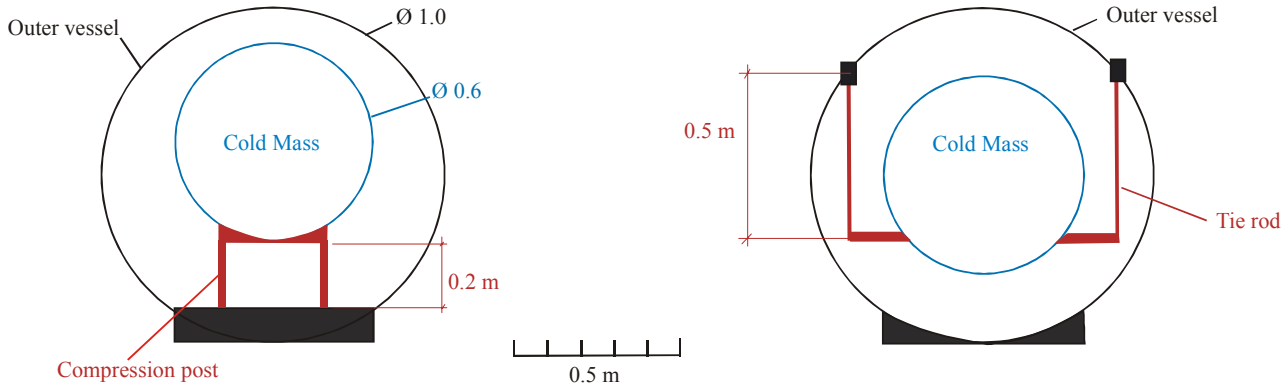


Fig. 1 Basic transverse configurations of cryomagnet, using compression support posts (a) or tension tie rods (b).

4.3 Does one need intermediate shielding?

Per unit length of cold mass, the lateral cold surface area is

$$\pi \times 0.6 = 1.88 \text{ m}^2$$

Consider only multilayer insulation in good vacuum, for example 30 layers wrapped around the cold mass; the resulting heat load would then be $1.5 \times 1.88 = 2.8$ W, i.e. an order of magnitude too high with respect to the budget of 0.2 W.

$\Rightarrow$ The cryostat will feature a thermal shield actively cooled at intermediate temperature, say 80 K which can be provided by liquid nitrogen or gaseous helium. A quasi-cylindrical shield with a

diameter of 0.8 m can be fitted in-between cold mass and vacuum vessel, leaving sufficient radial space for thermal insulation and mechanical tolerances.

4.4 Does one still need insulation around the cold mass, in presence of the 80 K shield?

The cold mass surface, made of stainless steel, has an emissivity $\varepsilon_{cm} = 0.12$ (Table 2).

The inner surface of the thermal shield, made of aluminium alloy sheet as found, has an emissivity ε_{ts} of 0.12 (Table 2).

The emissivity factor E for long nested cylinders in the case of diffuse reflection is

$$E = \frac{\varepsilon_{cm} \cdot \varepsilon_{ts}}{\varepsilon_{ts} + \frac{A_{cm}}{A_{ts}} \cdot (1 - \varepsilon_{ts}) \cdot \varepsilon_{cm}}$$

where A_{cm} and A_{ts} are the respective surface areas of cold mass and thermal shield, of emissivity ε_{cm} and ε_{ts}

$$E = \frac{0.12 \times 0.12}{0.12 + 0.75 \times 0.88 \times 0.12} \simeq 0.07$$

The heat inleak by radiation is then:

$$\begin{aligned} \dot{Q}_{rad} &= E\sigma A_{cm}\left(T_{ts}^4 - T_{cm}^4\right) \\ &= 0.07 \times 5.67 \times 10^{-8} \times 1.88 \times 80^4 \\ &= 0.31\ \text{W} \end{aligned}$$

which exceeds the thermal budget.

Consider now an improvement of the emissivity of the cold mass by wrapping one layer of aluminium foil, $\varepsilon_{cm} = 0.01$.

Then $\dot{Q}_{rad} = 0.04$ W within budget.

4.5 What about residual gas conduction?

Let's assume that the residual gas is pure helium, since all other species are condensed at low temperature, with vanishingly low vapour pressure.

Consider only the molecular conduction regime, which prevails at low pressure. In this regime, heat transfer is proportional to residual gas pressure, according to Kennard's law

$$\dot{Q}_{res} = A_{cm}\, \alpha\, \Omega P\left(T_{ts} - T_{cm}\right)$$

with $\Omega = 2.13\ \text{W/m}^2\,.\,\text{Pa}\,.\,\text{K}$ for helium and the overall accommodation coefficient α given by:

$$\alpha = \frac{\alpha_{cm} \cdot \alpha_{ts}}{\alpha_{ts} + \alpha_{cm}\left(1 - \alpha_{ts}\right)\frac{A_{cm}}{A_{ts}}}.$$

With a 0.8 m diameter thermal shield,

$$\frac{A_{cm}}{A_{ts}} = 0.75$$

From literature, for helium at 2 K $\alpha_{cm} = 1$
at 80 K $\alpha_{ts} = 0.4$

This yields a value $\alpha = 0.47$.

Then, for 1 mPa residual helium pressure,

$$\dot{Q}_{res} = 0.15 \text{ W}$$

so that the total radiative and gas conductive heat inleak from the 80 K shield to the cold mass wrapped with one layer aluminium foil is

$$\dot{Q}_{rad} + \dot{Q}_{res} = 0.04 + 0.15 = 0.19 \text{ W} \text{ within budget.}$$

Consider, however, a 100-fold increase in residual helium pressure (100 mPa). Then

$$\dot{Q}_{res} = 15 \text{ W} \text{ which is unacceptable.}$$

A better solution is to wrap the cold mass with a multilayer system, say 10 layers of multilayer insulation. The total heat inleak in good vacuum is then, from Table 3,

$$\dot{Q} = 1.88 \times 0.05 = 0.09 \text{ W.}$$

Under degraded vacuum conditions (100 mPa helium), it still increases significantly

$$\dot{Q} = 1.88 \times 2 = 3.8 \text{ W}$$

however, a factor 4 lower than with the previous insulation system.

$\Rightarrow$ We therefore retain the latter solution of 10 layers multilayer insulation around the cold mass, yielding a distributed heat load of $\boxed{0.09 \text{ W}}$.

4.6 Which type of supporting system? of which material?

The heat load by solid conduction along the supports is given by

$$\dot{Q}_{con} = \int_{T_1}^{T_2} k(T)dT \times \frac{S}{L}.$$

Consider first a low-conductivity metallic material, i.e. austenitic stainless steel, for which a reasonable design stress in compression or tension is 100 MPa. To take the 20 kN weight of the 2000 kg mass, one needs a cross section $S = 2 \times 10^{-4}$ m^2.

For austenitic stainless steel, from Table 1,

$$\int_{1.9}^{290} k(T)dT = 3060 \text{ W/m}$$
$$\int_{1.9}^{80} k(T)dT = 349 \text{ W/m.}$$

In the case of compression posts, $L = 0.2$ m between the vacuum vessel and the cold mass, so that

$$\dot{Q}_{con} = 3060 \times \frac{2 \times 10^{-4}}{0.2} = 3.06 \text{ W}$$

which is well above budget.

Consider introducing heat interception at 80 K and midlength of the post; then

$$\dot{Q}_{con} = 349 \times \frac{2 \times 10^{-4}}{0.1} = 0.7 \text{ W}.$$

The situation has improved, but insufficiently.

In the case of tension tie-rods, $L = 0.5$ m. In the same fashion as previously

- without heat interception $\dot{Q}_{con} = 1.22$ W
- with 80 K heat interception at mid-length $\dot{Q}_{con} = 0.28$ W.

The latter value, although significantly lower, still remains above budget.

The only practical issue is to make use of an alternative material with lower heat conductivity: consider G-10 glass-fiber/epoxy composite, for which an acceptable design stress in compression is 40 MPa. To take the 20 kN weight, the cross section of G-10 required is thus $S = 5 \times 10^{-4}$ m^2.

For G-10 glass-fiber/epoxy composite, from Table 1,

$$\int_{1.9}^{290} k(T)dT = 153 \text{ W/m}$$
$$\int_{1.9}^{80} k(T)dT = 18 \text{ W/m}.$$

For a G-10 post without heat interception

$$\dot{Q}_{con} = 153 \times \frac{5 \times 10^{-4}}{0.2} = 0.38 \text{ W}$$

which is still too high.

Introducing 80 K heat interception at midpoint reduces the conductive heat inleak to

$$\dot{Q}_{con} = 18 \times \frac{5 \times 10^{-4}}{0.1} = \boxed{0.09 \text{ W}}$$

which is finally acceptable.

The heat load from 290 K intercepted at 80 K is then

$$135 \times \frac{5 \times 10^{-4}}{0.1} = 0.68 \text{ W}.$$

5. THERMAL DESIGN SUMMARY

⇒ Use a 0.8 m diameter actively-cooled thermal shield at 80 K, wrapped with 30 layers multilayer insulation. Its lateral surface of 2.51 m^2 will receive 3.76 W @ 80 K from the outer vessel.

⇒ Wrap the cold mass with 10 layers multilayer insulation: in good insulation vacuum, it will then receive 0.09 W @ 1.9 K.

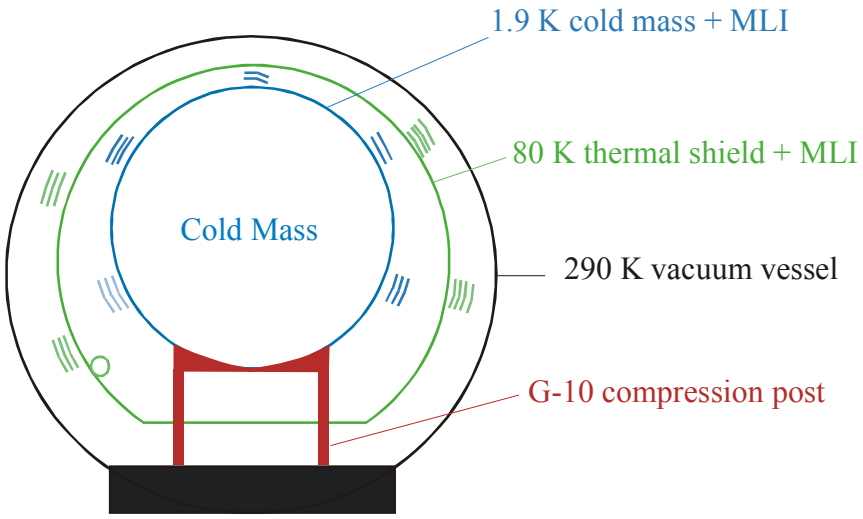
⇒ Use compression support posts made of G-10 glass-fiber/epoxy composite, with 80 K heat intercept at midpoint. Solid conduction then yields 0.09 W @ 1.9 K, and 0.68 W @ 80 K.

The transverse cross section of the cryostat which we have designed is sketched in Fig. 2. The corresponding thermal flow scheme is given in Fig. 3. The total heat load budget per meter length in nominal conditions is then:

@ 80 K 3.76 – 0.09 + 0.68 – 0.09 = **4.26 W**

@ 1.9 K 0.09 + 0.09 = **0.18 W**

Fig. 2 Sketch of cryostat transverse cross section



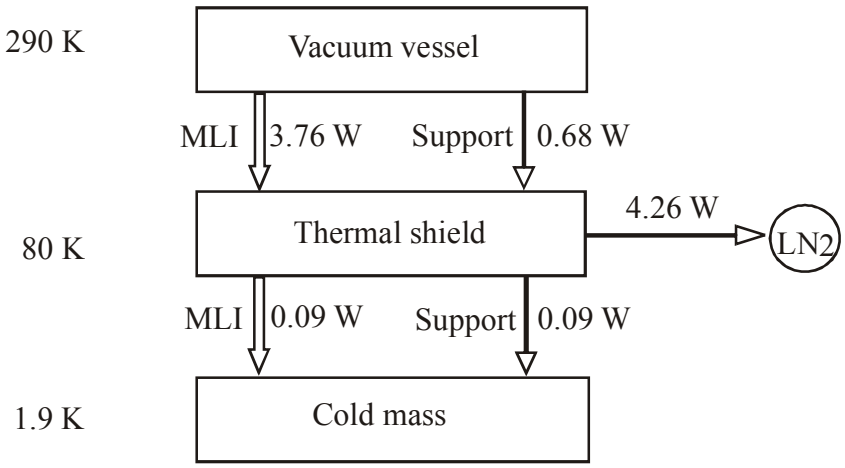


Fig. 3 Thermal flow scheme of cryostat (heat loads per m length)

6. FROM DESIGN TO CONSTRUCTION

The preceding study has permitted to sketch a design solution for the transverse cross-section of the cryostat and estimate its heat load budget. The final LHC dipole cryostat [4] exhibits the main features of the above solution: non-metallic composite support posts, an intermediate temperature thermal

shield providing heat interception at the support posts, multilayer reflective insulation around the thermal shield and cold mass (Fig. 4). A difference with the sketched solution resides in the second level of heat interception, using an available 5 K supercritical helium line, on the support posts. This feature enables to reduce further the heat conducted to the 1.9 K plate, from 0.09 W to 0.01 W (per m length of cryomagnet). The LHC support post described in [5], is shown in Fig. 5. Another important difference, which however does not appear on the cross-section, is the use of a forced flow of gaseous helium between 50 and 75 K, instead of liquid nitrogen, to cool the thermal shield. This choice is primarily driven by reasons of safety, to avoid the potential release of large amounts of nitrogen in the underground tunnel, in case of accident, and the resulting oxygen-deficiency hazard.

We shall now address some of the practical issues for constructing such a cryostat.

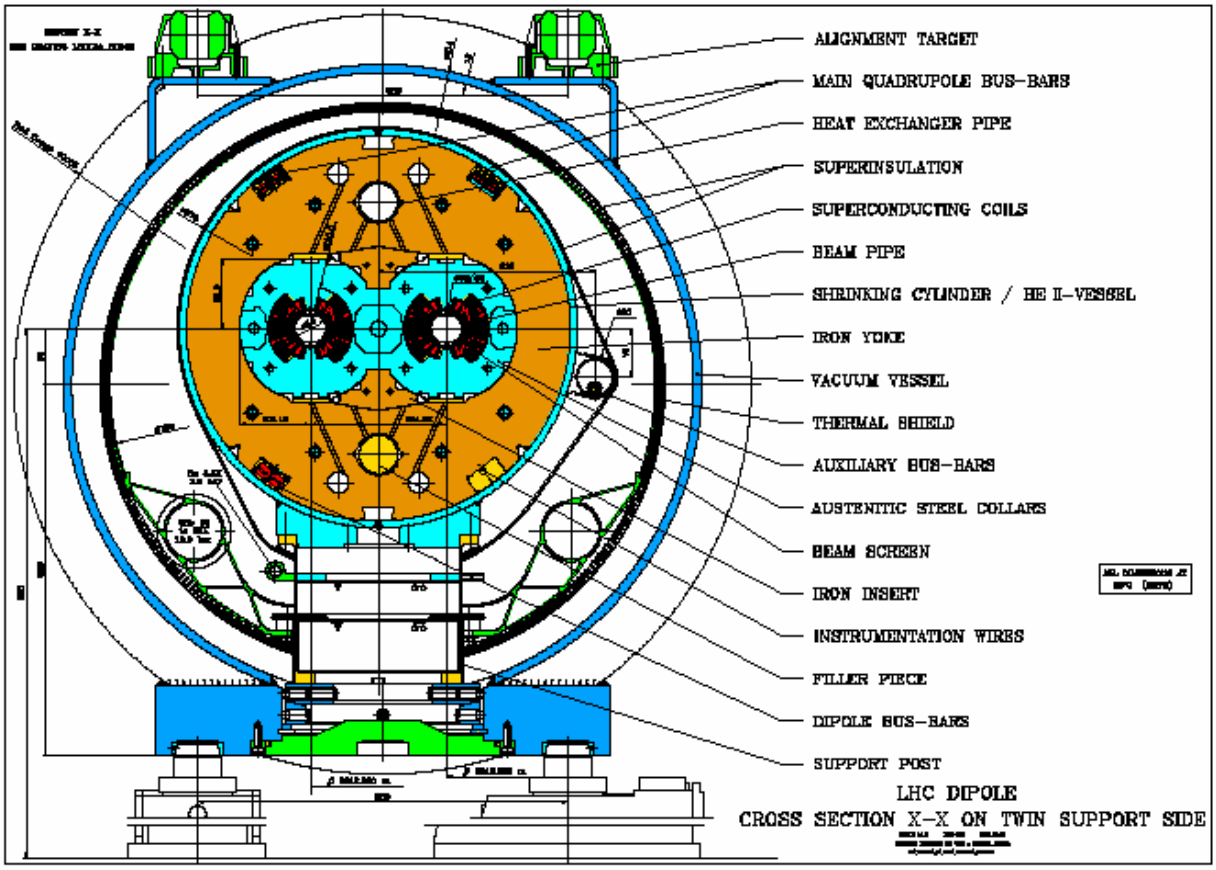


Fig. 4 Transverse cross section of final LHC dipole cryostat

Fig. 5 Non-metallic composite support post for LHC cryomagnet

6.1 The third dimension

The cryomagnet is a 3-dimensional object (Fig. 6): the supporting of the cold mass is performed at precise locations along the length. The maximum spacing, and hence the minimum number of support posts, is given primarily by the flexural rigidity of the cold mass and the allowed sagitta. For the 15-m long LHC dipole, the allowed sagitta of a few tenths of mm imposes to have three support posts. The system composed of the cold mass supported on three points from the vacuum vessel is then hyperstatic. For evident reasons of symmetry, there is a support post at mid-length. The longitudinal positions of the other two are determined by a compromise between considerations of stress (evenly-shared loading) and deformation (displacement and angle of the overhang at the ends of the cold mass).

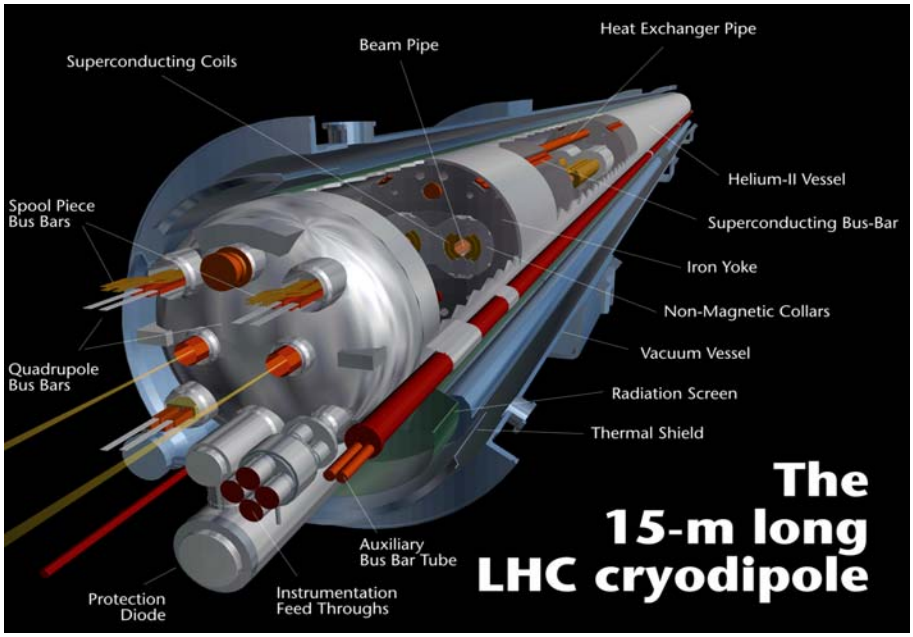


Fig. 6 Artist's perspective of LHC cryodipole

Another issue coming from the longitudinal dimension is that of thermal contraction. Assimilating the cold mass to an austenitic stainless-steel object, its integrated contraction upon cooldown to operating temperature is about 3×10^{-3}, i.e. 45 mm. Two constructional features are needed to cope with this.

- Bellows with sufficient travel, flexibility and resistance must be used to compensate for the contraction between cold masses at the cryomagnet interconnects (Fig. 7). The design optimisation, construction and installation of these several thousands bellows constitute an engineering feat in itself [6].
- To avoid the build-up of shear stresses on the lateral support posts, longitudinal displacement must be allowed between them and the vacuum vessel, thermal shield and/or cold mass. This displacement can be achieved by rolling or sliding, on the room-temperature or the cold end of the support posts. In the LHC cryomagnets, the support posts slide on low-friction surfaces on the vacuum vessel, i.e. under vacuum but at room temperature. Actually, as the cold mass is curved along the particle beam trajectory, the central post, which is fixed longitudinally, can however slide transversely to accommodate the cooldown/warmup displacements.

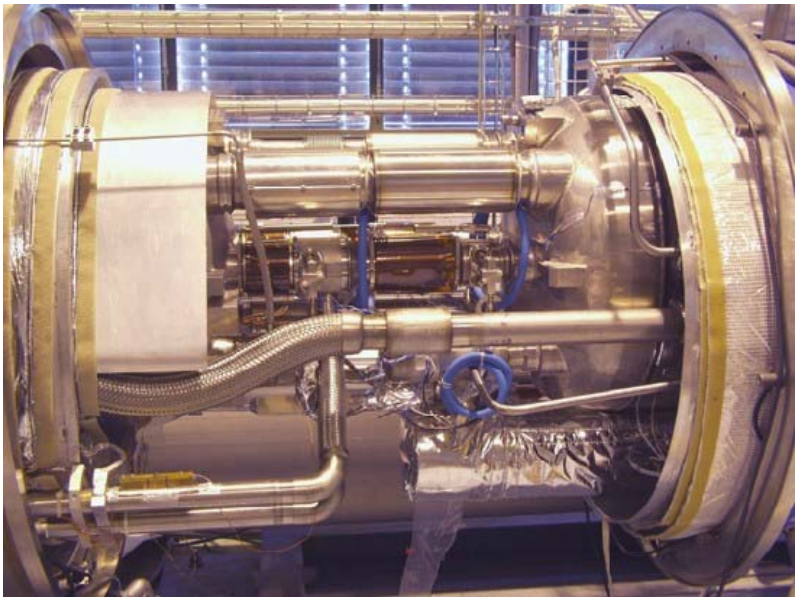


Fig. 7 Cryomagnet interconnect

6.2 Temperature homogeneity of the thermal shield

The thermal shield is a cylindrical shell cooled from one generatrix, along which the cooling tube runs. Its proper operation as a quasi-isothermal horizon surrounding the cold mass imposes that azimuthal thermal gradients developed by heat conduction in the shell remain below a limiting value, say 5 K. This requires selecting a material with high thermal conductivity and proper sizing of the shell.

Assuming a constant heat flux $\dot{q}$ received by the shield, and constant thermal conductivity k in the limited temperature range considered, the temperature profile along the developed circumference is parabolic, and the maximum azimuthal temperature difference is:

$$\Delta T_{azimuth} = \frac{\pi^2}{8} \frac{D^2}{k\,e} \dot{q}$$

with D diameter of thermal shield
e thickness of thermal shield
k thermal conductivity of thermal shield.

Let us consider first a 2-mm thick sheet of 2024 aluminium alloy, with an average thermal conductivity in the 50-75 K range of 46 W/m . K

Then, for $D = 0.8$ m
$\dot{q} = 1.5$ W/m^2

one finds $$\Delta T_{aximuth} = \frac{\pi^2}{8} \frac{0.8^2 \times 1.5}{46 \times 2 \times 10^{-3}} = 12.9 \text{ K}$$

This clearly exceeds the requirement. To stay within the 5 K limit, a sheet thickness of more than 5 mm would be required.

An alternative is to use a 2-mm thick sheet of commercially pure aluminium (1100), which shows an average thermal conductivity of 330 W/m . K

Then, $$\Delta T_{aximuth} = \frac{\pi^2}{8} \frac{0.8^2 \times 1.5}{330 \times 2 \times 10^{-3}} = 1.8 \text{ K}$$

which is acceptable.

With respect to 2024 aluminium alloy, commercially pure aluminium exhibits much weaker mechanical strength: this is however of secondary importance for this application driven by thermal conductivity, provided the rigidity is given by an extruded “bottom tray” which also integrates the cooling channel (Figs 8 and 9).

Fig. 8 Thermal shield extruded “bottom tray” with integrated cooling channel

Fig. 9 Section of aluminium thermal shield around dipole cold mass

6.3 Achieving efficient heat interception

As the thermal flow scheme in Fig. 3 shows, the very low residual heat load at the 1.9 K level rests on the ability to construct heat intercepts which operate as calculated. In practice, these heat intercepts often involve solid-to-solid contacts under vacuum at low temperature, which are known to be difficult to build in an efficient and reproducible manner, preserving their low thermal impedance over time and thermal cycles.

Industrial-type solutions for these intercepts, implemented in the LHC cryodipole construction, include the following.

- The contact between the shell and bottom tray of the thermal shield is ensured by interrupted weld seams covering at least 25% of the length.
- The heat intercept plates on the support posts are made of 10-mm thick aluminium, which ensures their shrink fitting onto the G-10 column, and glued to it.
- The lower heat intercept plate (50-75 K) is connected to the thermal shield bottom tray via flexible aluminium strips, in an all-welded construction (Fig. 5).
- The upper intercept plate (5 K) integrates a 1-mm wall-thickness stainless steel cooling tube, fitted through it at the time of casting (Fig. 5).

6.4 Installing multilayer insulation

The solution retained for the LHC cryomagnets is that of prefabricated multilayer blankets incorporating reflective film, spacer and outer protection net, factory-cut and assembled with the precise geometry, and equipped with VELCRO® fasteners before packaging. Upon assembly of the cryostat, the blankets are easily installed around the cold mass and thermal shield (Fig. 10), and fixed with the VELCRO® fasteners (Fig. 11). It must be noted that the latter show a high emissivity (close to that of a blackbody), and must therefore be covered with aluminium tape, when exposed, or concealed below the internal force of the blanket.

Fig. 10 Prefabricated multilayer insulation blankets upon installation

6.5 The outer vessel

The outer vessel, which constitutes the interface of the cryomagnet with the outside world, combines several functions. First of all, it constitutes the leak-tight enclosure for the insulation vacuum of the cryostat. Equally important, it participates to the supporting and precise positioning of the cold mass, a function specific to accelerator magnet cryostats, which the standard tolerances of sheet-metal work may have difficulty to match. It also performs auxiliary duties, such as magnetic shielding of the stray flux. Finally, in view of its dimensions and mass, it represents the single largest component of the cryostat in terms of cost, and must therefore be optimised also from this point of view.

Fig. 11 VELCRO® fasteners for fixing multilayer insulation blankets

The vacuum vessels of the LHC dipole cryostats are essentially welded steel tubes, with reinforcing rings and precision-machined surfaces at the three locations of the support posts (Fig. 12).

In this fashion, the expensive high-precision manufacturing operations are confined to these locations, while the rest of the vessel can be constructed with the tolerances of sheet-metal work. For reasons of cost and magnetic shielding, the vessels are made of low-carbon steel, with stainless steel flanges for interconnects at the ends. They must therefore be painted outside for corrosion protection and sandblasted inside for low outgassing. Locating the external supports at the same longitudinal positions as the support posts suppresses bending moments and corresponding deformations. The vessel wall can then be sized from the standard construction codes applicable for pressure and vacuum vessels.

Fig. 12 Outer vacuum vessels awaiting assembly

7. CONCLUSION

This tutorial will have reached its primary goal if it has convinced the student that cryostats meeting complex and demanding requirements can be conceptually designed from the basic laws of heat transfer and the thermophysical properties of construction materials at low temperature, with the guidance of second-law thermodynamics and – as a consequence – the virtues of temperature staging and heat interception. We hope it will also have demonstrated some of the difficulties of implementing the conceptual design into an engineering product, and illustrated technological solutions to achieve reproducible performance while limiting complexity and cost.

ACKNOWLEDGEMENTS

I wish to thank A. Poncet and his colleagues of the LHC Cryostats and Integration group for briefing me on the latest updates of cryostat construction and providing the corresponding pictures. Thanks are also due to E. Delucinge for carefully preparing this writeup.

REFERENCES

[1] G. Vandoni, Heat transfer, in this report.

[2] G. Claudet, F. Disdier, A. Gauthier, Ph. Lebrun, J. Schmid, Conceptual study of the superfluid helium cryogenic system for the CERN Large Hadron Collider, Proc. ICEC12, Butterworths, Guildford (1988) 497-504.

[3] 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.

[4] J.C. Brunet, V. Parma, G. Peón, A. Poncet, P. Rohmig, B. Skoczen and L.R. Williams, Design of the second series of LHC prototype dipole magnet cryostats, Adv. Cryo. Eng. 43A (1998) 435-441.

[5] M. Castoldi, M. Pangallo, V. Parma and G. Vandoni, Thermal performance of the supporting system for the LHC superconducting magnets, Adv. Cryo. Eng. 45A (2000) 795-802.

[6] A. Jacquemod, A. Poncet, B. Skoczen, J.-Ph. Tock, The interconnections of the LHC cryomagnets, Proc. PAC2001, IEEE, Piscataway (2001) 616-618.

THE PHYSICS OF SUPERFLUID HELIUM

W. F. Vinen
School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK

Abstract
The paper contains a brief account of the physics of superfluid ^{4}He, with emphasis on the underlying physical principles; it uses the minimum of mathematics, and there is some emphasis on aspects that relate to practical applications.

1. INTRODUCTION AND HISTORY

Helium was first liquefied by Kammerlingh Onnes in Leiden in 1908. During the late 1920s and early 1930s it was noticed that the liquid had some strange properties, but it was not until 1938 that it was discovered independently by Allen and Misener and by Kapitza that it exhibited frictionless flow and was what we now call a superfluid. Shortly afterwards Fritz London suggested that superfluidity could have some connection with Bose-Einstein condensation, which was known as a theoretical possibility in an ideal Bose gas. London also realized that there might be a strong connection with superconductivity, which had been discovered many years before and which could be seen as superfluidity in the electron gas in a metal. With impressive intuition he also suggested that both superfluidity and superconductivity were "quantum mechanisms on a macroscopic scale", although the significance of this idea did not become really clear until the late 1950s or early 1960s.

Shortly after London produced these seminal ideas he and Tisza suggested that the superfluid phase of the liquid could be described by a two-fluid model, the condensed and non-condensed atoms being identified respectively with the superfluid and normal components. In 1941 Landau wrote a remarkable paper in which he suggested that superfluidity can be understood in terms of the special nature of the thermally excited states of the liquid: the well-known phonons and rotons. This idea led Landau also to the idea of a two-fluid model, but with a microscopic interpretation that was different from that of London and Tisza. Indeed, Landau expressed the view that superfluidity has no obvious connection with Bose condensation, although, as we shall see, this view was certainly wrong. Nevertheless, the basic ideas in Landau's paper were correct, and his interpretation of the two-fluid model showed brilliant intuition.

After the second world war the two-fluid model was placed on a firm experimental basis, especially with the experiment of Andronikashvili and the discovery of second sound. At the same time the properties of the normal fluid (the gas of phonons and rotons) were explored in great theoretical detail by Khalatnikov, with parallel confirmatory experiments.

A theoretical proof that Bose condensation does occur in a liquid such as superfluid helium was provided by Onsager and Penrose. Feynman wrote a number of important papers in the 1950s, exploring how the properties of liquid helium were strongly related to the fact that the atoms obey Bose statistics. The quantization of superfluid circulation and the existence of free quantized vortices were proposed theoretically and independently by Onsager and Feynman, and the first experimental confirmation came from the work of Hall and Vinen with the discovery of mutual friction in rotating helium and with the direct observation in a macroscopic experiment of the quantization of circulation. This work led to an appreciation for the first time of the full significance of London's "quantum mechanism on a macorscopic scale", and of the underlying importance of Bose condensation in superfluidity.

In 1957 Bardeen, Cooper and Schrieffer wrote their famous paper on the theory of superconductivity. In due course this led to a better appreciation of the connection between superfluidity and superconductivity, and the discovery of the quantization of flux and of free flux lines in type II superconductors demonstrated clearly the analogies between the two systems. As far as we know all superfluids

and superconductors have one basic feature in common: their properties derive from the existence within them of some type of Bose condensation, involving atoms or pairs of atoms or pairs of electrons.

Liquid ^{3}He exhibits no superfluid behaviour at the relatively high temperatures involved in superfluid ^{4}He, thus confirming the importance of particle statistics in this behaviour. The discovery of superfluidity in liquid ^{3}He by Osheroff, Richardson and Lee in 1973 at a temperature of about 2mK completed the story, showing that BCS pairing can occur in an uncharged Fermi liquid; the pairs are now pairs of atoms, but the pairing is unconventional in that it involves relative p-states rather than the s-states of the conventional BCS theory. Unconventional pairing is now known to occur in exotic superconductors, such as the heavy-fermion metals and the high-temperature materials.

In these brief notes we shall focus our attention on superfluidity in liquid ^{4}He, emphasizing the underlying physical principles, including those associated with macroscopic quantum phenomena, and we shall place some emphasis on aspects that relate to practical applications.

The following references contain useful introductory reading [1], [2], [3].

2. THE PHASE DIAGRAM OF ^{4}He

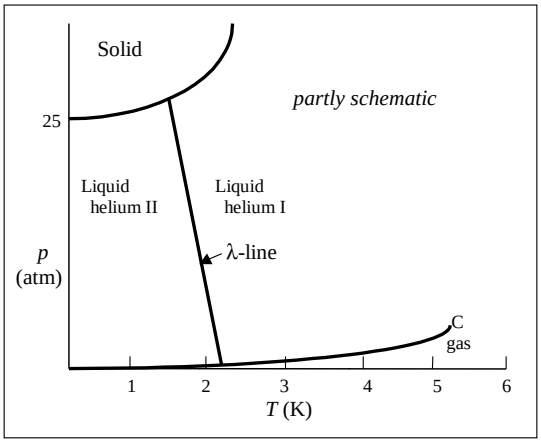


Fig. 1: The phase diagram of ^{4}He.

We see that the phase diagram in Fig. 1 exhibits two anomalous features. The liquid phase exists over a range of pressure up to about 25 atm even at the absolute zero of temperature; and there are two liquid phases, helium I, which is conventional in its properties, and helium II, which is superfluid.

The existence of a liquid over a range of pressures at $T = 0$ must be a quantum effect. It arises from quantum mechanical zero point energy: the fact that a confined particle must have kinetic energy, this energy increasing as the particle is more strongly confined. In the absence of a high pressure, the atoms cannot become sufficiently closely spaced to allow the formation of an ordered crystal, without the penalty of too large a zero point energy.

The Third Law of Thermodynamics requires that the entropy of a system in equilibrium should vanish at $T = 0$. Therefore the liquid must be in some sense completely ordered at $T = 0$. This ordering must be quantum mechanical in origin, as in the ordering of particles among quantum mechanical energy levels rather than in position. It seems reasonable to suppose that superfluidity is a consequence of this ordering.

3. THE HEAT CAPACITY

The heat capacity, C, is shown in Fig. 2, for the case when the helium is under its own vapour pressure. We see that the transition to superfluidity is accompanied by a large peak in the heat capacity. There is no latent heat, but the heat capacity tends to infinity at the transition, so that the transition cannot be classified as strictly second-order. The shape of the heat capacity near the transition is like a greek letter λ: hence the term $\lambda-$point to describe the transition. The type of anomaly depicted in Fig. 2 is

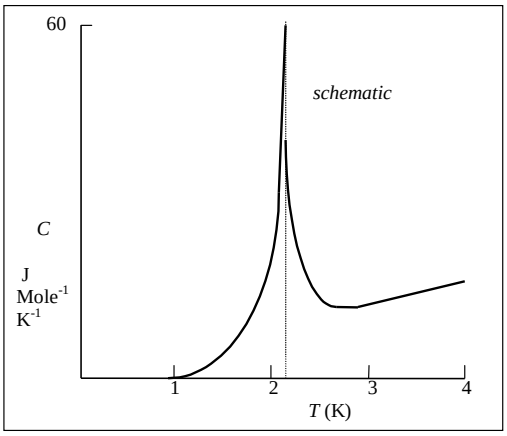


Fig. 2: The heat capacity of liquid ^{4}He.

quite common in nature, and it is characteristic of a system that exhibits an order-disorder transition; an example is the ferromagnetic transition. We see clear confirmation that superfluidity must be associated with a (quantum mechanical) ordering in the liquid. A similar anomaly in the heat capacity appears at the transition temperature of a superconductor, although in this case it has more nearly the character of a strictly second order transition.

We note that, although the heat capacity becomes rather small at low temperatures, it is quite large just below the $\lambda-$point; for example at 1.8K. This feature can be useful in applications.

4. THE OBSERVED PROPERTIES OF SUPERFLUID ^{4}He: THE TWO-FLUID MODEL

At first sight these properties present a confusing picture, but they make sense in terms of the two-fluid model, regarded as a purely phenomenological description. We describe the essential features of this model, and then give examples of properties that can be described in terms of it. The superfluid phase can be regarded as a mixture of two fluids, which can support different velocity fields. The **normal fluid**, with density ρ_n, flow velocity field $\mathbf{v}_n$ and conventional viscosity η_n, carries all the thermal energy and entropy in the system. The **superfluid component**, with density ρ_s and flow velocity field $\mathbf{v}_s$, can flow without friction and carries no thermal energy. The densities, ρ_n and ρ_s vary with temperature in the way shown in Fig. 3.

A pressure gradient will tend to drive both fluids in the same direction. An increase in temperature increases ρ_n but decreases ρ_s, so a temperature gradient tends to drive the superfluid component in one direction (towards to high temperature) and the normal fluid in the opposite direction.

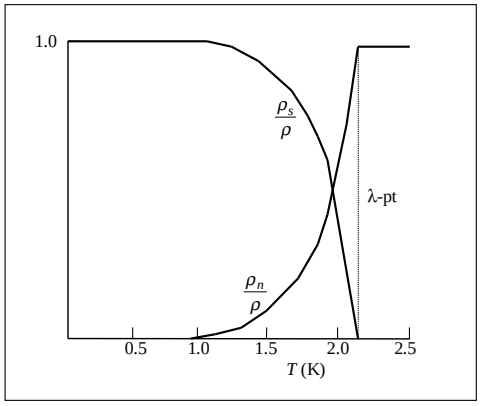


Fig. 3: The observed dependence of ρ_n and ρ_s on temperature.

5. EXAMPLES OF "TWO-FLUID" BEHAVIOUR

The superfluid component can flow without friction through even very narrow channels, so narrow that the normal fluid is rendered completely immobile by its viscosity. A striking example is provided by "film flow". Any solid surface in contact with the liquid is covered by a film of liquid, about 30 nm in thickness, as a result of van der Waals attraction between the helium atoms and the substrate. This is true in principle for any liquid, but in helium flow of the superfluid component through the very thin film becomes possible, with the result illustrated in Fig. 4.

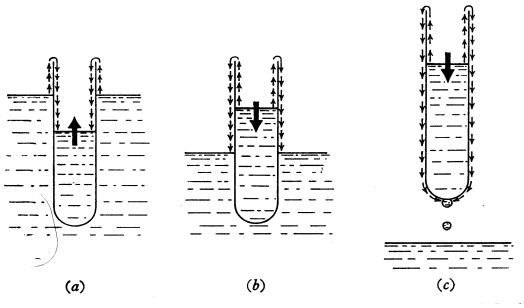


Fig. 4: Film fbw

A famous experiment was performed by Andronikashvili. He constructed a pile of discs, which he suspended in helium by a torsion fibre, as shown in Fig. 5. He measured the period of torsional oscillation as a function of temperature. The spacing between the discs was such that at the period of oscillation the normal fluid was completely coupled to the disc system. However, the superfluid component was not coupled, so that only the normal fluid contributed to the moment of inertia of the disc system. These measurements provided the first evidence for the dependence of normal fluid density on temperature shown in Fig. 3.

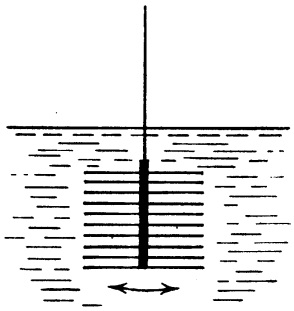


Fig. 5: The Andronikashvili experiment

Heat transport in superfluid helium takes place by counterflow of the two fluids, the superfluid component moving towards the source of heat and the normal fluid away from it, as shown in Fig. 6. Only the normal fluid carries thermal energy, at a rate per unit area, $Q = \rho S T v_n$, where S is the entropy of the helium per unit mass. This leads to very effective thermal transport, at a rate limited only by the small viscosity of the normal fluid. In practice the thermal transport is not quite as effective as is suggested by this idea, as will be explained later.

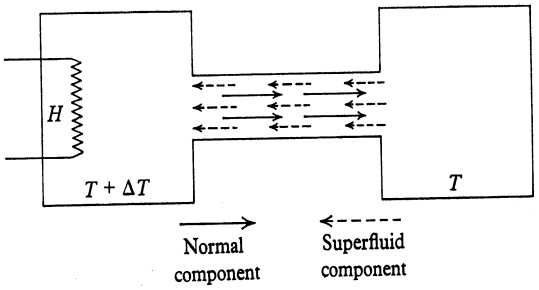


Fig. 6: Illustrating thermal transport by counterfow

The existence of two fluids allows two modes of longitudinal wave propagation. The two fluids can oscillate in phase, giving rise to **first sound**; or thay can oscillate in antiphase, giving rise to **second sound**. First sound is an isentropic pressure or density wave, analogous to ordinary sound in a fluid; it propagates at a speed of $c_1 = (\partial p/\partial \rho)_S^{1/2} \approx 240$ ms^{-1}. Second sound involves to a good approximation no change in density, but only a change in the proportions of the two fluids; it is therefore a temperature wave, but one that obeys the wave equation rather than the diffusion equation. The speed of second sound is given by $c_2^2 = TS^2\rho_s/C\rho_n$, and its value is roughly 20 ms^{-1} over the temperature range from 1K to 2K. Transient thermal effects in superfluid helium can therefore be very different from those in a conventional fluid, and discussion of them must allow for the existence of second sound.

The examples of two-fluid behaviour that we have described apply in their simplest form only if the flow velocities do not exceed certain critical values, which are often quite small ($\sim$few mm s^{-1}). We shall discuss the reason later, after we have explained the theoretical basis of the two-fluid model.

The two-fluid model applies also to superconductors. The resistive loss that occurs in a rapidly oscillating electric field is due to motion of the normal fluid.

6. WHY IS HELIUM DESCRIBED BY A TWO-FLUID MODEL?

Part of the answer to this question was given by Landau in his famous 1941 paper. In effect he focussed his attention on the nature of the normal fluid. He considered the form of the thermally excited states in the liquid at a low temperature. He argued with great insight (but less rigour) that they would consist of quantized sound waves, which are called **phonons**, and elementary forms of rotational motion called by him **rotons**. His ideas were placed on a firmer theoretical basis by Feynman, who was able to be more precise about the nature of a roton, which he showed to be in essence a free atom moving through the liquid, with a backflow formed from motion of the other atoms. Suppose that we set all these excitations into motion with a drift velocity $\mathbf{v}$, leaving the fluid otherwise at rest. Given the properties of the excitations (in particular their energy-momentum relationship, which can now be determined experimentally by neutron scattering) Landau calculated the momentum density, $\mathbf{J}_e$, associated with the drifting excitations. He found that

$$\mathbf{J}_e = \rho_e \mathbf{v} < \rho \mathbf{v}, \tag{1}$$

where the inequality holds at sufficiently low temperatures, which turn out to be temperatures below the $\lambda-$point. Thus the drifting excitations do not cause the whole fluid to drift, in the sense that they carry an effective density that is less than the total density of the helium.

We identify the gas of excitations with the normal fluid. Then $\rho_e = \rho_n$ can be calculated, and it can be shown to be equal to the observed normal-fluid density.

The superfluid component in Landau's picture is what is left over after the thermal excitations have been taken into account. Landau also considered what would happen if this background were to move. He showed that it could not slow up by creating or scattering excitations if its velocity were less that a critical value, which is about 60 ms^{-1}. This picture of the superfluid component is not wholly satisfying, and it is certainly not the whole story, not least because observed critical velocities are typically very much less than 60 ms^{-1}. We shall now examine the nature of the superfluid component in more detail, and we shall demonstrate its connection with Bose condensation.

7. THE NATURE OF THE SUPERFLUID COMPONENT

To understand the real nature of the superfluid component we must start by looking at the phenomenon of Bose-Einstein condensation. Bose condensation plays a crucial role in superfluidity, contrary to Landau's original opinion.

Consider an ideal gas formed from Bose particles: i.e. particles such as ^{4}He atoms that are quantum-mechanically indistinguishable, but are not subject to the exclusion principle (i.e. there can be any number of particles in one quantum state). If we calculate the way in which the particles of the gas are distributed over the quantum states determined by the shape and size of the containing vessel, we find an interesting result: below a critical temperature, T_0, a finite fraction of the particles are "condensed" into the lowest quantum state. The way in which this fraction varies with temperature is shown in Fig. 7(a), and the calculated heat capacity is shown in Fig. 7(b). The heat capacity reflects the ordering of the particles into a single quantum state below the temperature T_0. Very recently, such Bose condensation has been observed directly in weakly-interacting gases formed from alkali-metal atoms levitated magnetically and trapped in a vacuum, the gas being cooled below the temperature T_0 (typically in the range 0.1-1 μK) by a combination of laser and evaporative cooling [4].

For an ideal hypothetical gas of non-interacting helium atoms with the same density as liquid helium the condensation temperature $T_0 \sim 3$K. An obvious question is whether a similar type of ordering occurs in real liquid helium, albeit modified in some way by the strong interactions between the helium atoms.

The answer is that it does, as shown first by Penrose and Onsager. The fraction of condensed particles is smaller than in the ideal gas; even at $T = 0$ it is only about 10 percent. But it remains the

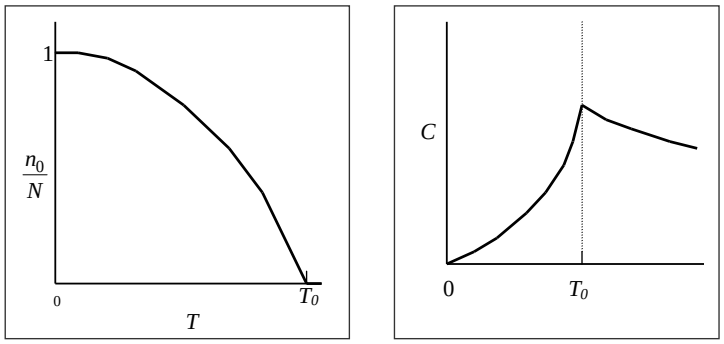


Fig. 7: The predicted behaviour of an ideal Bose gas. (a) The temperature dependence of the condensed fraction of particles; (b) the predicted heat capacity.

case that a macroscopic fraction, and a very large absolute number, of the atoms does condense into what is effectively a single quantum state, and it turns out that at $T = 0$ the non-condensed atoms are effectively locked to the condensed atoms.

We now understand that this is indeed the ordering process taking place below the $\lambda-$point, and that ultimately it is this ordering that is responsible for superfluidity. It is a remarkable process, because it is closely analogous to the formation of a coherent electromagnetic wave in a laser, which can be viewed as a condensation of photons into a single quantum state. In helium there is a coherent matter wave. A similar process occurs in a superconductor, except that the coherent wave is formed from Cooper pairs. A coherent matter wave lies at the heart of both superfluidity and superconductivity.

The assembly of condensed atoms is called the condensate, and the associated wavefunction is called the condensate wave function (CWF). If the condensed atoms are at rest the CWF is just a constant Ψ_0, where Ψ_0^2 is a measure of the number of condensed atoms. If they are moving, each with momentum $m_4 v$ along the x-axis, the CWF becomes

$$\Psi = \Psi_0 \exp\left(\frac{i m_4 v x}{\hbar}\right). \tag{2}$$

For a more general motion of the condensate we can write

$$\Psi = \Psi_0 \exp\left(iS(\mathbf{r})\right), \tag{3}$$

where the local velocity of the condensed atoms is equal to $(\hbar/m_4)\,\nabla S$. We identify this velocity with the velocity of the superfluid component

$$\mathbf{v}_s = \left(\frac{\hbar}{m_4}\right)\nabla S. \tag{4}$$

We can ask how this view of superfluidity relates to that proposed by Landau, which was very successful in accounting for two-fluid behaviour. We now know that the two approaches are intimately connected, in the sense that the *form* of the spectrum of the thermal excitations, which underlies Landau's calculation showing that $\rho_n/\rho < 1$ below the $\lambda-$point, is intimately connected with the existence of the condensate. Without the condensate the spectrum would have the wrong form. Note especially that we now have a clear view of the meaning of the velocity of the superfluid component, which was not provided by Landau.

A condensate exists also in a superconductor, formed from the Cooper pairs. The mass m_4 is replaced by $2m$, where m is the electron mass.

8. QUANTUM RESTRICTIONS ON SUPERFLUID FLOW

As we shall now demonstrate, the macroscopic occupation of a single quantum state in the Bose-condensed helium gives rise to macroscopic quantum effects, as London had foreseen.

It follows from Eq. (4) for the superfluid velocity that

$$\text{curl}\mathbf{v}_s = 0. \tag{5}$$

This means that there can be no local rotational motion of the superfluid component. This is really a consequence of the quantization of angular momentum, as we see more clearly in a moment. But there *can* be a finite **hydrodynamic circulation**, defined as

$$\kappa = \oint_C \mathbf{v}_s \cdot d\mathbf{r}, \tag{6}$$

round any cicuit that cannot shrink to nothing while remaining in the fluid; for example, a circuit round a solid cylinder passing through the fluid (Fig. 8). However, the circulation cannot take any value. If we

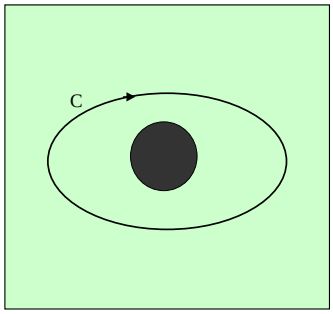


Fig. 8: Illustrating a circuit round which there can be a fi nite superfluid circulation.

substitute from Eq. (4) into Eq. (6) we obtain

$$\kappa = \frac{\hbar}{m_4} \oint_C \nabla S \cdot d\mathbf{r} = n\frac{2\pi\hbar}{m_4}, \tag{7}$$

where n must be an integer in order to satisfy the condition that the CWF be single-valued. This means that the **superfluid circulation must be quantized in units of** $2\pi\hbar/m_4$. This circulation is macroscopically large (it can be measured in a macroscopic mechanical experiment), and this fact provides the clearest evidence that superfluidity is indeed a "quantum mechanism on a macroscopic scale". It arises from the quantization of angular momentum, combined with the fact that all the particles in the condensate must have the *same* angular momentum. In the absence of any quantized circulation there can be no local angular momentum, as we have seen in connection with Eq. (5). The quantization of circulation is has its analogue in superconductivity, where it is observed as the quantization of trapped flux.

9. WHY CAN THE SUPERFLUID FLOW WITHOUT FRICTION?

As we have mentioned, Landau showed that the flowing superfluid component cannot decay into excitations unless the velocity is very large. With the idea of the condensate we can gain greater insight into this frictionless flow.

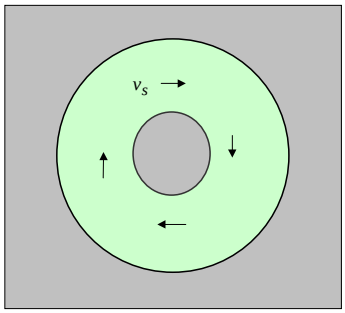


Fig. 9: Persistent superflow round a torus

Suppose that there is a persistent superflow round a torus, as shown in Fig. 9. This flow can be only metastable, because a state with no flow has a smaller (free) energy. Why is it metastable? The condensate contains a macroscopic number of atoms. Interaction of these atoms with the walls of the torus will cause scattering, and some atoms may as a result be knocked out of the condensate. This will reduce the amplitude of the CWF, but it will not alter its coherent phase. Therefore the superfluid velocity does not change, although the superfluid density may decay a little, which would correspond to the creation of more normal fluid in the form of excitations. Putting it in another way, we can say that the destruction of superflow would require a transition that takes a macroscopic number of atoms from one state to another simultaneously, and such a process has very low probability.

But superflow *can* decay through a mechanism that we have not yet considered: the creation of free vortex lines, to which we now turn our attention.

10. QUANTIZED VORTEX LINES IN SUPERFLUID HELIUM

We have seen that a quantized superfluid circulation can exist round a solid cylinder running through the helium. A free quantized vortex line in the superfluid component is a quantum of circulation round a tiny cylindrical hole in the helium. Such a line always has one quantum of circulation, and the hole then has a natural size, determined by a balance between the kinetic energy of flow and the surface energy of the hole, that is less than an interatomic spacing.

Such vortex lines can exist in superfluid helium, and, as we shall show, they play an important role in its behaviour. Most obviously, perhaps, they allow the superfluid component to rotate if the helium is placed in a rotating vessel; otherwise such rotation would be forbidden by Eq. (5). A parallel array of

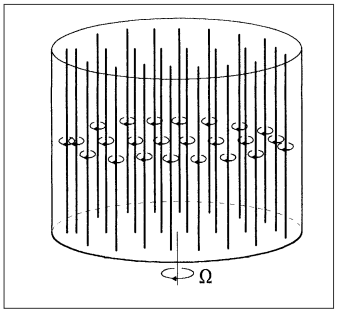


Fig. 10: Vortex lines in the uniformly rotating superfluid component.

lines, as shown in Fig. 10, gives rise to a flow field that mimics uniform rotation on length scales larger

than the line spacing, which is about 0.2 mm at $\Omega = 1$ s^{-1}. This array is analogous to the array of flux lines in the mixed state of a type II superconductor.

Vortex lines scatter the excitations that constitute the normal fluid, and therefore they give rise to a frictional force between the two fluids, called **mutual friction**. This is observed as an attenuation of second sound when it propagates in the uniformly rotating helium. The observation of this attenuation provided the first experimental evidence for the existence of vortex lines.

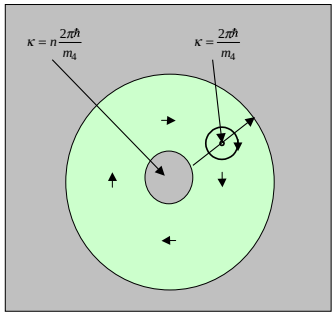


Fig. 11: Decay of a persistent current by vortex motion.

Vortex lines provide a new mechanism by which a persistent superflow can decay (Fig. 11). Consider again a persistent superflow in a torus. Let the persistent current consist of n quanta of circulation. If a free vortex, with the appropriate sign, crosses the channel, this value of n falls to $n - 1$. Does this mean that the current simply decays? It does not, because the movement of the free vortex across the channel is opposed by a potential barrier. This barrier arises because a vortex is attracted to a solid boundary by its image. The barrier is quite large in cases of practical interest, and it can be overcome only at high velocities ($>\sim$1-10 ms^{-1}), either thermally or by quantum tunnelling. **Without this barrier there would be no superflow.** The barrier exists only because a vortex has a finite quantized circulation, so it is quantum in origin. The barrier exists also in a superconductor, where it is usually called the Bean-Livingston barrier.

In practice frictionless superflow usually breaks down at velocities much less than 1 ms^{-1}. This is due to a few **remnant vortices**, which can expand and multiply, and then cross the channel (cf remnant dislocations in a solid allowing the solid to deform much more easily than might have been expected). Remnant vortices seem always to be created when the helium is cooled through the $\lambda-$point.

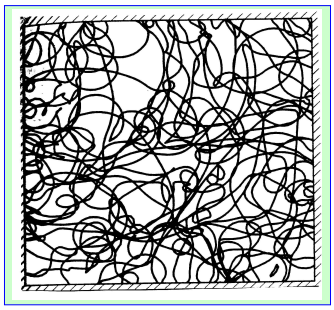


Fig. 12: A turbulent tangle of vortex lines.

This expansion and multiplication leads to a type of turbulence in the superfluid component: a kind of tangle of vortex lines (Fig. 12). Superfluid turbulence is very common. It seems always to be

generated when the flow velocity exceeds a critical value that depends on channel size and is often as small as 1 mm s^{-1}.

11. PRACTICAL CONSEQUENCES OF SUPERFLUID TURBULENCE

Superfluid turbulence plays an important role in limiting heat transport in superfluid helium by counterflow. The counterflowing fluids cause remnant vortices to multiply (through the action of mutual friction), and this leads to a self-sustaining regime of homogeneous turbulence. The vortices thus generated lead to a steady average force of mutual friction per unit volume between the two fluids, given by

$$F_{sn} = A\rho_s\rho_n|v_s - v_n|^3, \tag{8}$$

which limits the heat transport rate, Q per unit area, in a way that is generally much more important than normal-fluid viscosity. The parameter A is about 800 m s kg^{-1} at 1.8K. Q becomes a non- linear function of the temperature gradient, which is given by

$$\nabla T = \frac{A\rho_n}{\rho_s^3 S^4 T^3} Q^3, \tag{9}$$

where S is again the entropy per unit mass of the helium. Although mutual friction becomes the dominant dissipative process limiting the heat flow, the effective thermal conductivity remains generally very high.

Superfluid helium can be forced to flow down a tube or past an obstacle, just as can any conventional fluid. Except at very small velocities or in very narrow channels both the superfluid component and the normal component become turbulent. It turns out that this turbulence is surprisingly similar to that in a conventional fluid at high Reynolds number. The reasons are complicated, but they seem to be connected with two facts: on a scale large compared with the spacing between the vortex lines even the superfluid component looks like a classical fluid flowing at high Reynolds number; and the mutual friction associated with the vortex lines serves to lock the two velocity fields together. Thus the flow of the superfluid phase of liquid helium at high velocities in situations having a classical analogue is described quite well by classical formulae describing the flow of a conventional fluid, with density equal to the total helium density and viscosity similar to that of the normal fluid (for a recent extensive review of quantum turbulence see reference [5]).

12. THE KAPITZA THERMAL BOUNDARY RESISTANCE

As we have seen the effective thermal conductivity of superfluid helium is very high, but often it is necessary to transfer heat out of a solid body into the helium, or *vice versa*. We must then take account of a high thermal boundary resistance between the solid and the helium (the Kapitza resistance). This resistance arises from the fact that it is generally difficult for a thermal excitation in the solid to convert to one in the helium. This can be seen most easily when both the excitations are quantized sound waves or phonons. When a sound wave approaches a change of medium, some is transmitted and some is reflected, the relative amounts being determined by the characteristic impedances ($Z = \rho c$) of the two media. For liquid helium Z has a value that is much smaller than for any solid, and the resulting serious acoustic mismatch at the boundary leads to the high thermal boundary resistance. Its value is typically of order 2×10^{-4} K W^{-1} m^2.

13. SUMMARY AND CONCLUSIONS

The superfluid phase of liquid ^{4}He behaves in strange ways, which can be summarized as follows. It shows "two-fluid" behaviour; a normal fluid coexisting with a superfluid component. The superfluid component can exhibit frictionless flow at low velocities and in narrow channels. Rotational motion in

the superfluid component is severely restricted by quantum effects, associated with the quantization of circulation (essentially the quantization of angular momentum). This unconventional behaviour has its origin in quantum effects and especially in the formation of a coherent matter field within the liquid, associated with the phenomenon of Bose-Einstein condensation. At high flow velocities ideal superfluid behaviour, involving frictionless flow, breaks down through the generation of a form of quantum turbulence, which leads to a frictional interaction between the superfluid and normal components. Quantum turbulence is likely to be important in many situations of practical importance.

References

[1] D. R. Tilley and J. Tilley, *Superfluidity and Superconductivity*, 3rd ed. (Adam Hilger, Bristol, 1990).

[2] R. P Feynman, *Progress in Low Temperature Physics*, vol. 1, ch. 2, ed. C. J. Gorter (North Holland Publishing Co., Amsterdam, 1955).

[3] R. J. Donnelly, *Quantized Vortices in Helium II* (Cambridge University Press, 1991).

[4] See, e.g., *Bose-Einstein condensation in atomic gases, Proc. Int. School of Physics Enrico Fermi*, eds. M. Inguscio, S. Stringari and C. Wieman (IOS Press, Amsterdam, 1999).

[5] W. F. Vinen and J. J. Niemela, *J. Low Temperature Physics*, in press.

THE TECHNOLOGY OF SUPERFLUID HELIUM

Ph. Lebrun and L. Tavian
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, are discussed in terms of heat transfer capability and limitations in conductive and convective modes. Large-capacity refrigeration techniques below 2 K are reviewed, as concerns 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 also 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 classes of reasons call for the use of superfluid helium as a coolant for 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 Figure 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 [3-5], nuclear magnetic resonance [6, 7], magnetic confinement fusion [8, 9] and circular particle accelerators and colliders [10-12]. In the case of high-frequency superconducting devices, such as acceleration cavities [13], 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 [14, 15] and future high-energy lepton colliders [16-18] operate in the temperature range which minimizes capital costs and overall energy consumption. This issue is schematized in Figure 2.

The technical heat transfer characteristics of superfluid helium basically derive from peculiar transport properties [19, 20]. Its low bulk viscosity enables superfluid helium to permeate to the heart of magnet windings, while its very large specific heat (typically 10^5 times that of the conductor per unit mass, 2 x 10^3 per unit volume), combined with excellent heat conductivity at moderate heat flux (10^3 times that of cryogenic-grade OFHC copper) can produce powerful stabilization against thermal disturbances. In order to fully exploit these properties, however, both in steady-state and transient regimes, e.g. for power heat transport over macroscopic distances as well as intimate stabilization of superconductors, an elaborate thermo-hydraulic design of the cooling circuits, conductor, insulation and coil assemblies is required. 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, which some authors prefer to keep the qualificative "superfluid" for.

In the following, we will only address the specific issues of cryogenic technology pertaining to the use of superfluid helium as a technical coolant, namely different cooling methods as well as processes and machinery for sub-lambda temperature refrigeration [21]. Reference is made to companion lectures for cryogenic techniques which – however important in system design - are not superfluid-helium specific, such as thermal insulation and cryostat design [22, 23].

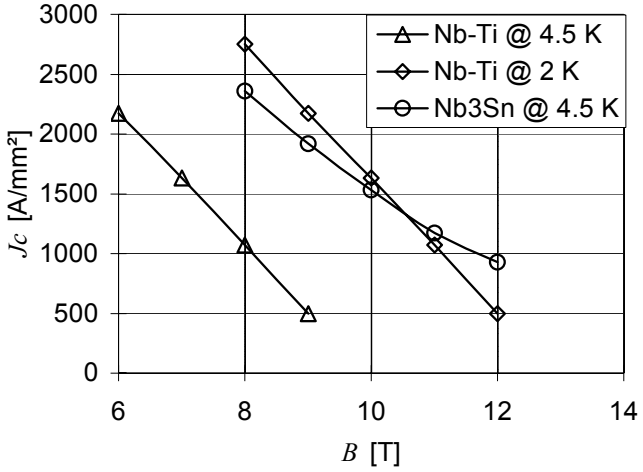


Fig. 1 Critical current density of technical superconductors.

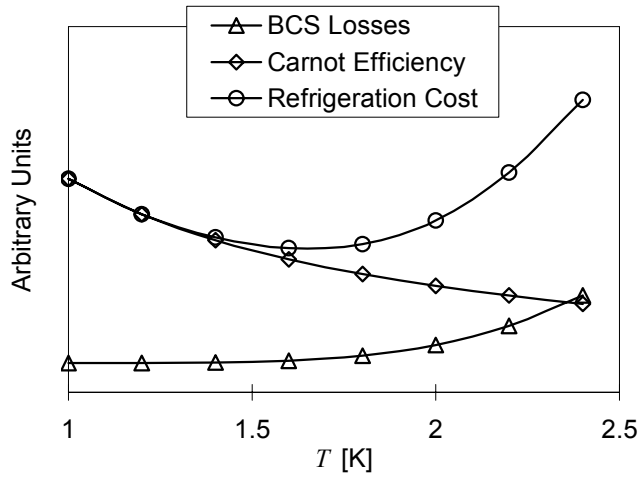


Fig. 2 Optimal operating temperature of RF superconducting cavities.

2. DIFFERENT COOLING METHODS

2.1 Pressurized versus saturated superfluid helium

A look at the phase diagram of helium (Figure 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 (about 100 kPa).

Although requiring one more level of heat transfer and additional process equipment - in particular a pressurized-to-saturated helium II heat exchanger - pressurized helium II cooling brings

several important technical advantages [24]. 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 [25], in the vicinity of the minimum of the Paschen curve (Figure 4) [26], brings the additional risk of electrical breakdown at fairly low voltage. Operating in pressurized helium II avoids this kind of problem.

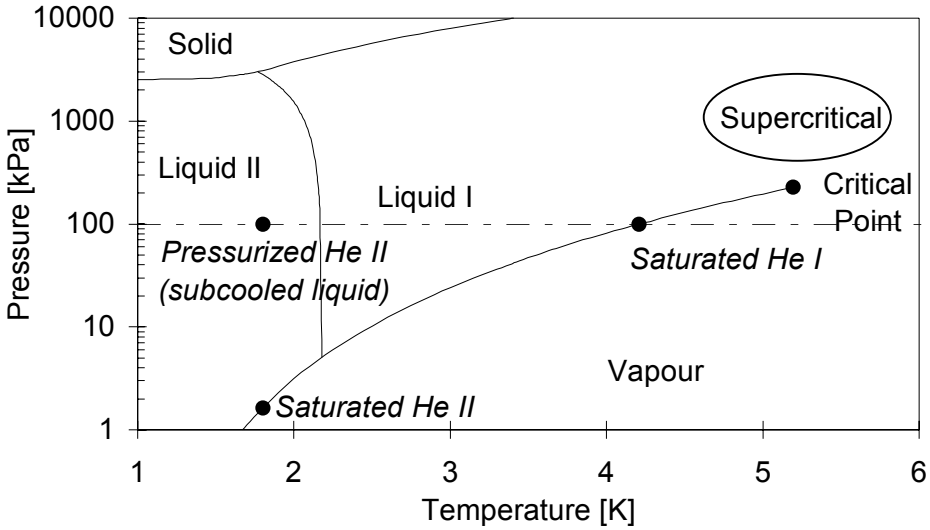


Fig. 3 Phase diagram of helium.

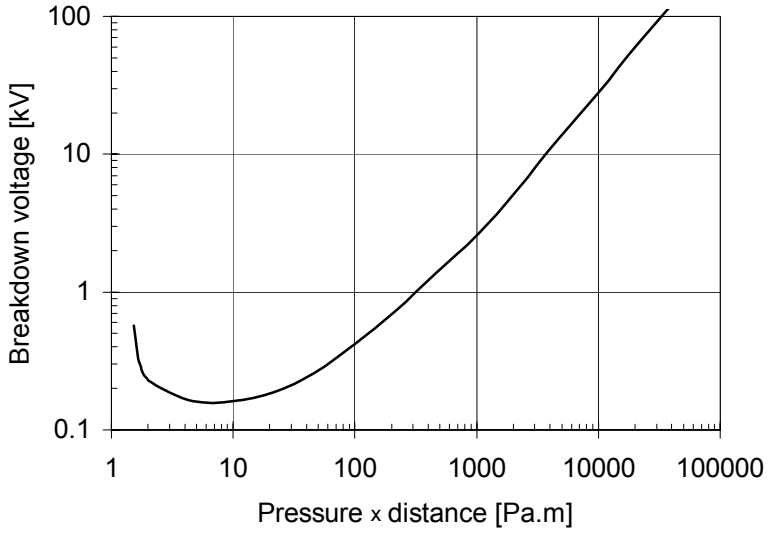


Fig. 4 Paschen curve for helium at 300 K.

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 only 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 [27], typically yields an order of magnitude better performance in favour of pressurized helium II.

2.2 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 [28]. 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 [29] theory*), practical data useful for engineering design has been established in a number of experiments [30-35].

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 $\dot{q}$ is given by:

$$\dot{q}^n \cdot L = X(T_C) - X(T_W) \tag{1}$$

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 [30]. A plot of this function reveals that the apparent thermal conductivity of helium II goes through a maximum at around 1.9 K (Figure 5).

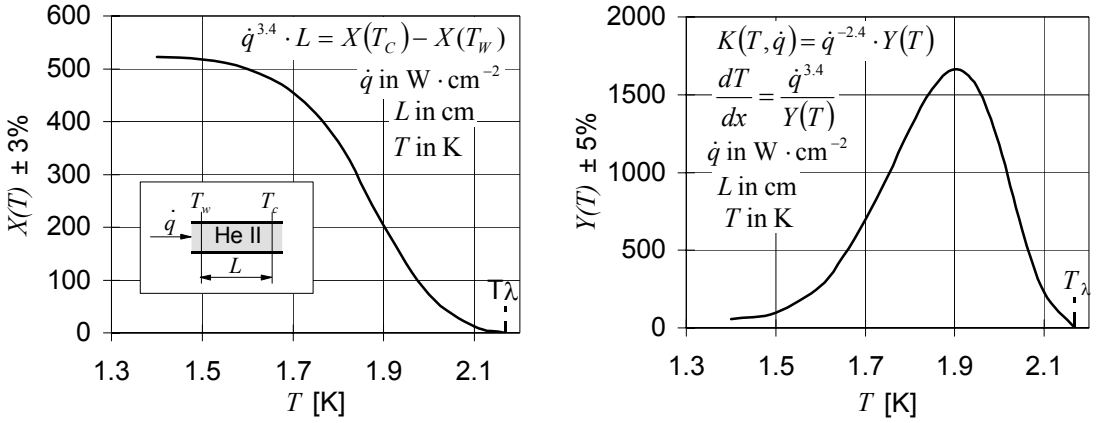


Fig. 5 Thermal conductivity integral and apparent thermal conductivity of pressurized superfluid helium [22].

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 what would be 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. Figure 6 shows the steady-state conduction $\dot{Q}$ in superfluid helium between 1.9 and 1.8 K versus the static column length L for different equivalent nominal diameters of the column. This abacus clearly shows that 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 %.

*) 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.

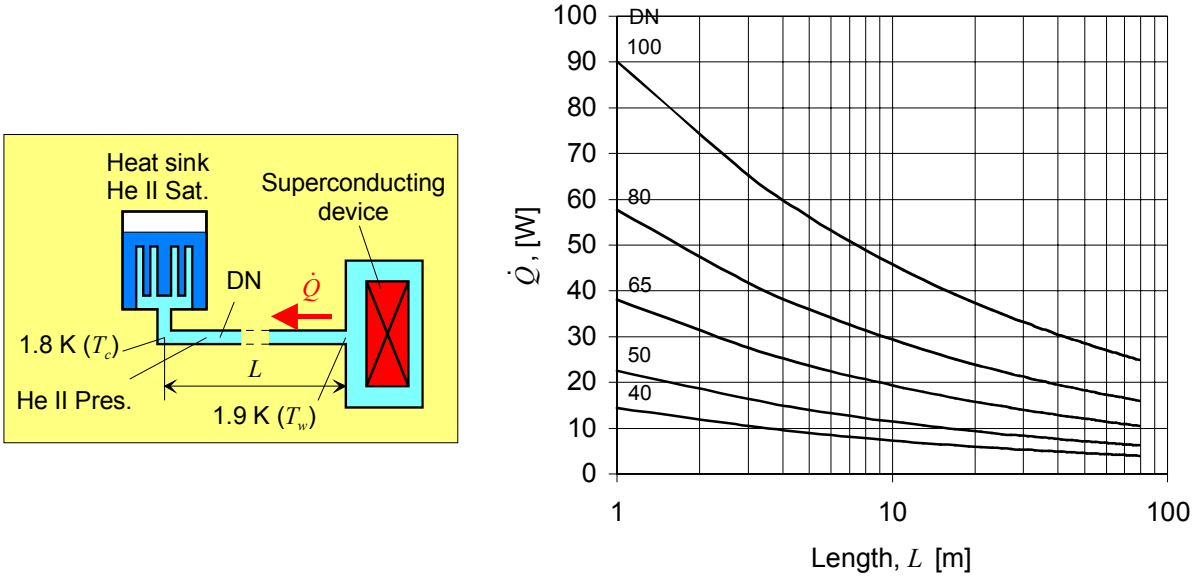


Fig. 6 Steady-state conduction in pressurized superfluid helium.

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:

$$\dot{q}_{total}{}^{n} \cdot L = (n+1) \cdot \left[X(T_C) - X(T_W) \right] \tag{2}$$

where $\dot{q}_{total}$ is the total heat flux flowing through the section at temperature T_C, near the heat sink. Figure 7 shows the steady-state conduction $\dot{Q}_{tot}$ in superfluid helium of a cryomagnet string with linear heating ξ between 1.9 K (temperature of the warmest magnet) and 1.8 K (temperature at the heat sink). As an example, cooling by conduction a 50-m long cryomagnet string, with a uniform linear thermal load of 1 W.m^{-1}, 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 considered for the LHC project [36] was later abandoned.

Conduction through static pressurized superfluid helium however remains the basic process for extraction and local transport of heat from the LHC magnet windings, across their polyimide-wrap electrical insulation. Although the polyimide tape, which constitutes the insulation of the superconducting cable, is wrapped in two layers with half overlap (Figure 8), in order to achieve sufficient mechanical toughness and dielectric strength, this still preserves sufficient percolation paths for helium II conduction to significantly improve the heat transfer, well above the solid conduction across the sole polyimide [37].

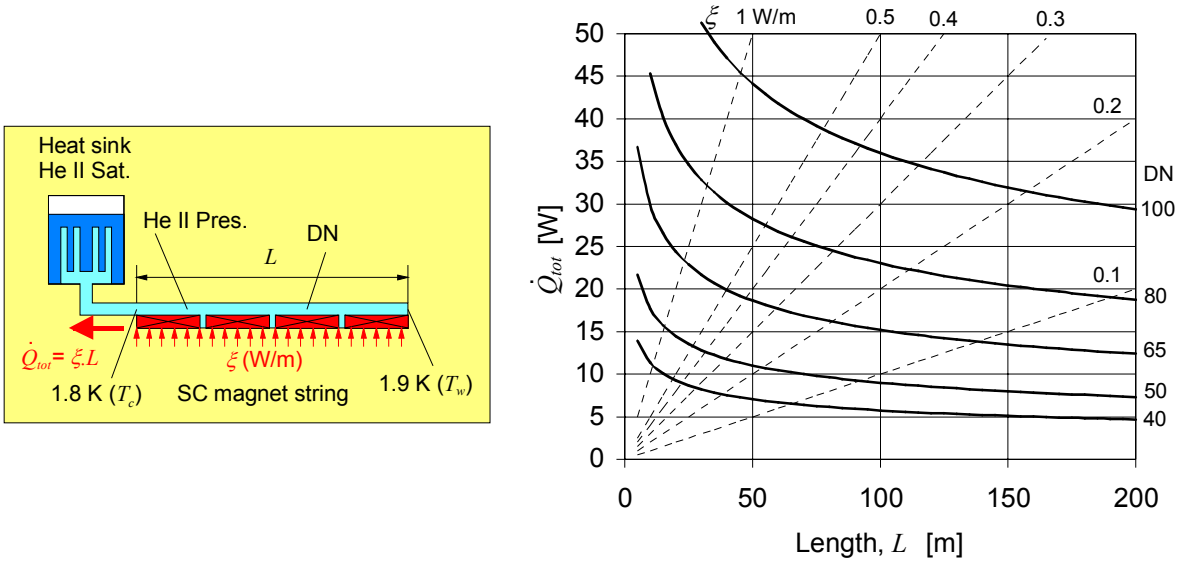


Fig. 7 Steady-state conduction cooling of cryomagnet string with linear applied heat load.

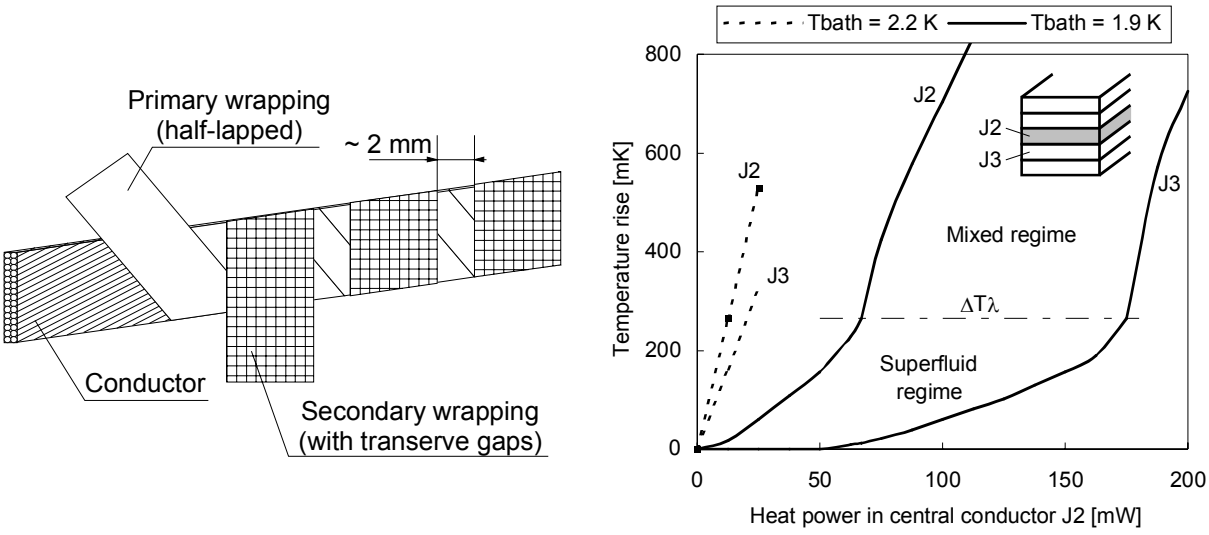


Fig. 8 Heat transfer across polyimide-wrap insulation of superconducting cable.

The high thermal conduction in helium II can also 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 permits to assess enthalpy changes of the system, and thus to perform calorimetric measurements. This technique proves very convenient for measuring minute heat inleaks [38] or substantial energy dissipation [39] such as produced by ramping losses or resistive transitions in superconducting magnets.

2.3 Forced-flow 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 of 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 a few mm. 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 convective heat transport $\dot{Q}$ between two points 1 and 2 of the cooling loop is then given by the difference in enthalpy H of the fluid flowing with a mass flow-rate $\dot{m}$:

$$\dot{Q} = \dot{m} \cdot (H_2 - H_1) \tag{3}$$

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 paragraph 2.2 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 (Figure 9) reveals a positive Joule-Thomson effect [40]: 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 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 [41].

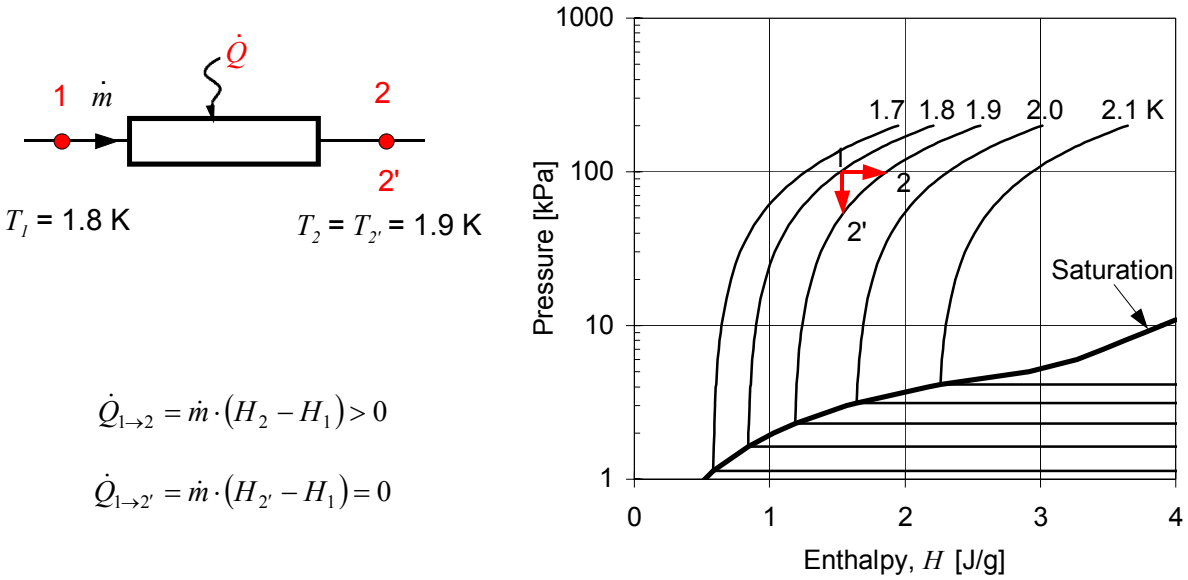


Fig. 9 Pressure-enthalpy diagram and forced-flow convection in superfluid helium.

Following early work [42, 43], several experimental programs have investigated heated flow of pressurized helium II in pipes and piping components [44, 45], culminating with the 230-m long test loop in Grenoble [46, 47] 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 [48, 49], 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 [50], the thermomechanical effect, specific of the superfluid, may also be used for driving cooling loops by means of fountain-effect pumps [51-54]. In spite of their low thermodynamic efficiency [55], 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 [56]. At higher heat loads, they have been considered [57] and tested [58] 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 [59].

2.4 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, schematized in Figure 10: 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 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 vaporizes [11].

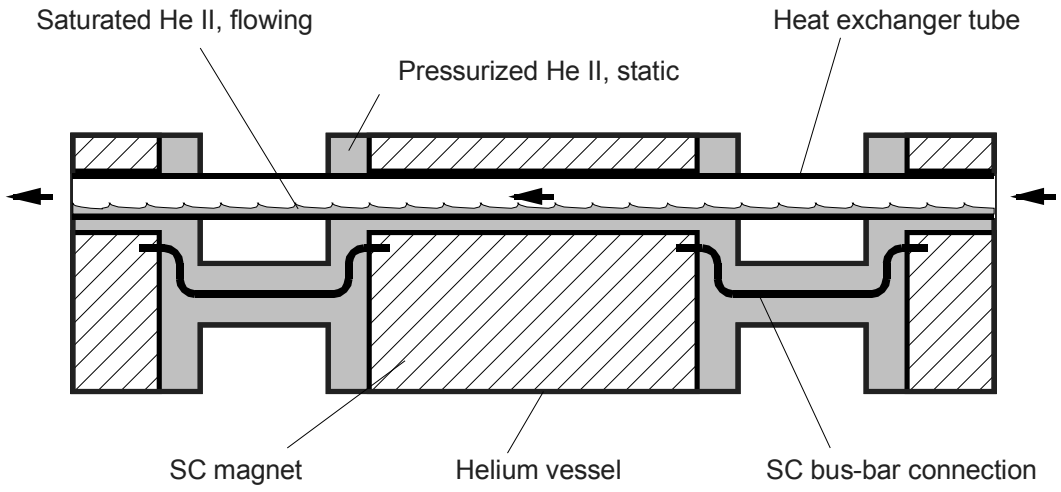


Fig. 10 Principle of the LHC superfluid helium cooling scheme.

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 originally available on flowing saturated helium II, and very little for other cryogenic fluids in this configuration. Following first exploratory tests [60] which demonstrated the validity of the concept on a reduced geometry, a full-scale thermohydraulic loop [61] 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 DN40 heat exchanger tube, made of industrial-grade deoxidized phosphorus copper.

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 (Figure 11). 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, usually dominates below 2 K [62].

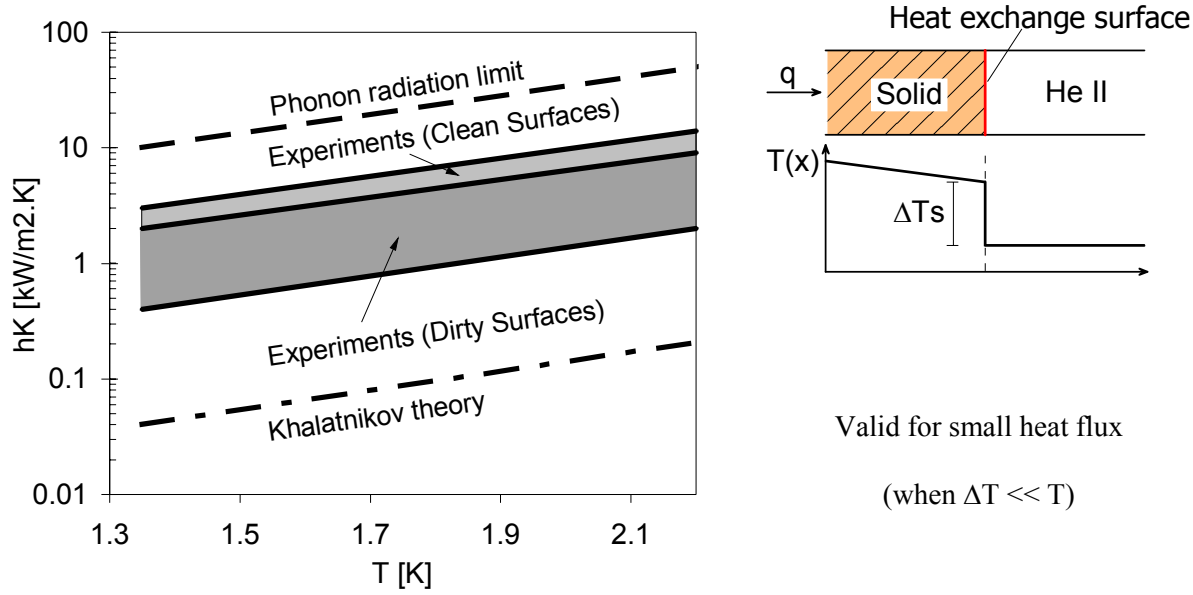


Fig. 11 Kapitza conductance at copper-helium II interface.

The final validation of the two-phase helium II flow cooling scheme for LHC has been performed successfully on a 100-m long test string, equipped with full-scale prototype cryomagnets, operated and powered in nominal conditions [63, 64]. 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 [65, 66]. In complement of that applied work, more fundamental experimental studies have been conducted on specially instrumented test loops at CEN-Grenoble, comprehensively equipped with diagnostics and a transparent section for visual observation and interpretation of the flow patterns [67-69]. As long as the vapor velocity remains sufficiently low to maintain stratified flow (up to a few m/s), engineering design of such a cooling scheme rests on a few simple sizing rules [70]. At higher vapor velocity, entrainment and atomization effects, still under investigation, complicate the flow pattern and impact on the heat transfer [71].

This type of cooling scheme may also be used for extracting much higher linear heat loads, typically about 10 $W.m^{-1}$, as present in the low-beta quadrupoles in the high-luminosity insertions of

the LHC [72, 73], at the expense of a larger-diameter heat exchanger tube to limit the saturated vapour velocity and thus preserve flow stratification.

3. REFRIGERATION CYCLES AND EQUIPMENT

The properties of helium at saturation (see Figure 3) impose to maintain 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 4.5 K. Figure 12 shows the basic scheme for refrigeration below 2 K. A conventional refrigerator produces liquid helium at 4.5 K, later expanded down to 1.6 kPa in a Joule-Thomson expansion stage. The gaseous helium resulting from liquid vaporization is compressed above the atmospheric pressure and eventually recovered by the 4.5 K refrigerator. We will therefore start by presenting the Joule-Thomson expansion stage.

Three types of cycles, sketched in Figure 12, can be considered [74, 75] for producing refrigeration below 2 K:

- the “warm” compression cycle based on warm sub-atmospheric compressors,
- the “cold” compression cycle based on multistage cold compressors,
- the “mixed” compression cycle based on a combination of cold compressors in series with warm sub-atmospheric compressors.

We will then proceed to discuss thermodynamics and machinery for these three types of cycle.

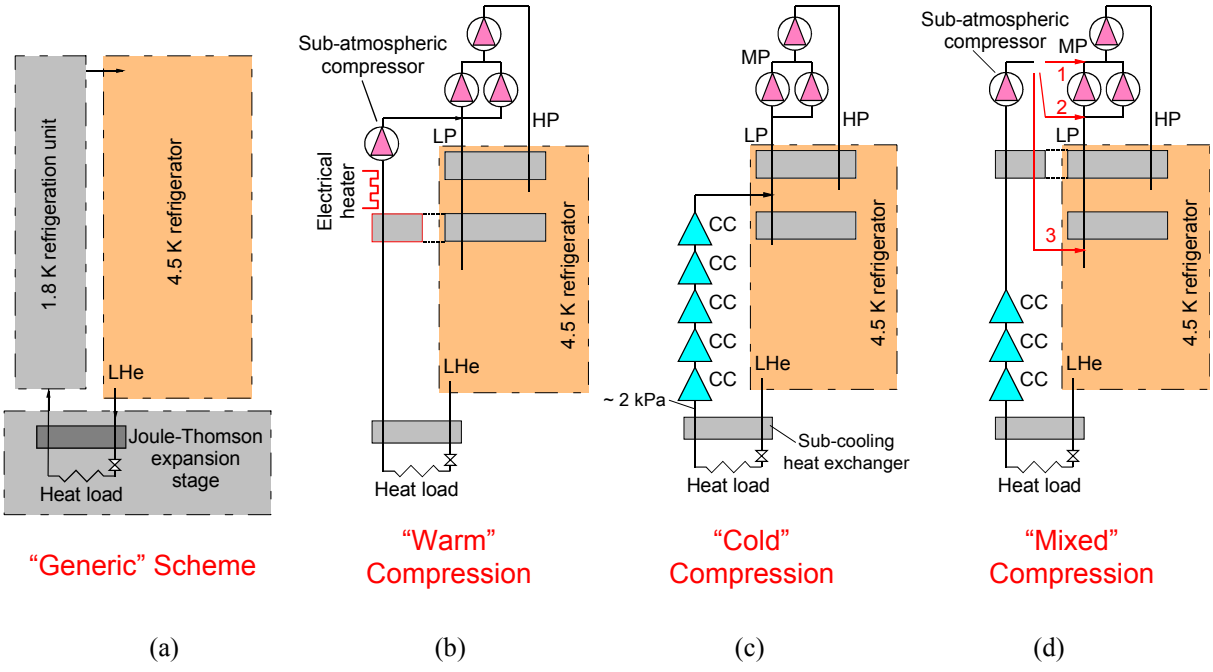


Fig. 12 Generic process cycles for refrigeration below 2 K.

3.1 Joule-Thomson expansion stage

The efficiency of the Joule-Thomson expansion of liquid helium, say from 0.13 MPa and 4.5 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 (Figure 13). This is performed in a counter-flow heat exchanger, subcooling the incoming liquid down to 2.2 K by enthalpy exchange with the very-low-pressure saturated vapour. This heat exchanger has to produce limited pressure drop, particularly in the very-low-pressure stream. A maximum pressure drop of 100 Pa is generally acceptable, corresponding to a few per cent

of the absolute saturation pressure. The design of such heat exchangers for large flow-rate [76] is not straightforward, and their qualification impractical. As a consequence, the LHC cryogenic system features several hundred small-size (5 to 20 g/s) heat exchangers, distributed around the ring. This also avoids transporting subcooled helium over long distances, saving one header in the ring distribution line. Following prototyping, technical validation of different solutions [77, 78] and commercial selection, these heat exchangers are now series-produced by industry.

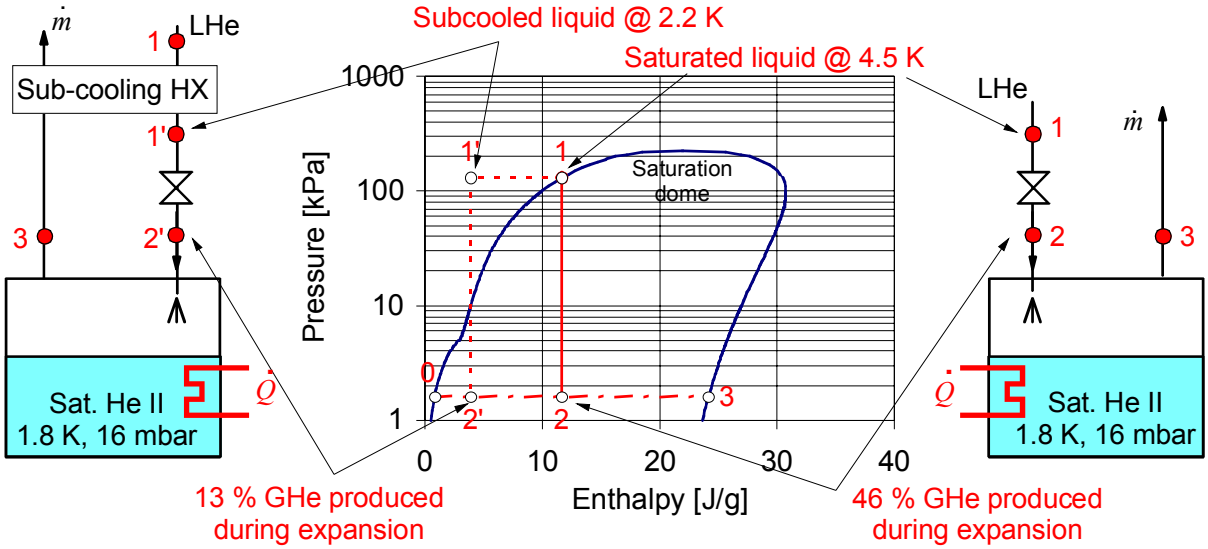


Fig. 13 Efficiency of Joule-Thomson expansion.

3.2 "Warm" compression cycle

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 through a heat exchanger and/or an electrical heater. 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 4.5 K cycle [79], 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.

All these compressors are positive-displacement machines having volumetric characteristics. Screw compressors are routinely used in helium refrigeration and their implementation in a 1.8 K cycle therefore follows from current practice. Special attention however has to be paid to the protection against air inleaks: in particular the motor shaft and its rotary sealing must be located on the discharge side to operate above atmospheric pressure.

A first limit to the use of subatmospheric screw compressors stems from volumetric flow requirements: the biggest available machines have a swept volume of about 4600 m^3/h, so that higher flow-rates require parallel arrays. Moreover, the isothermal efficiency - defined as the ratio of isothermal compression work to the effective compression work of the machine - decreases markedly with the suction pressure as shown in Figure 15, thus precluding their use at very low pressure in efficient process cycles.

(a) (b)

Fig. 14 Subatmospheric compressors.

(a) combination of Roots and rotary-vane vacuum pumps (b) compound screw.

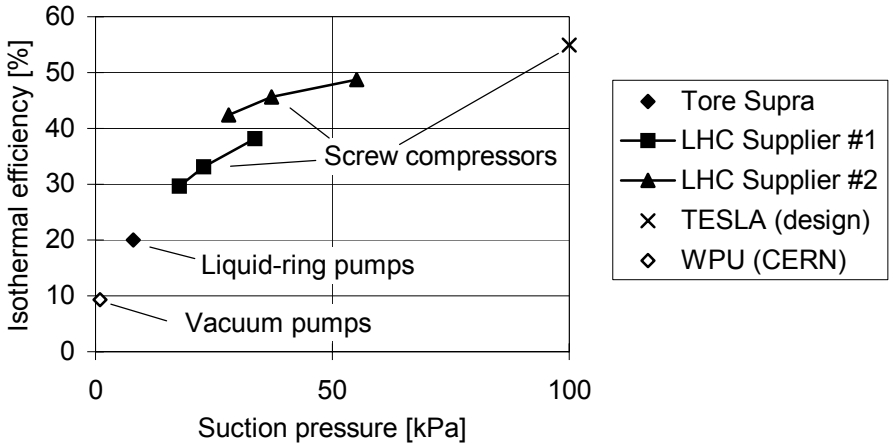


Fig. 15 Isothermal efficiency of warm sub-atmospheric compressors.

3.3 "Cold" compression cycle

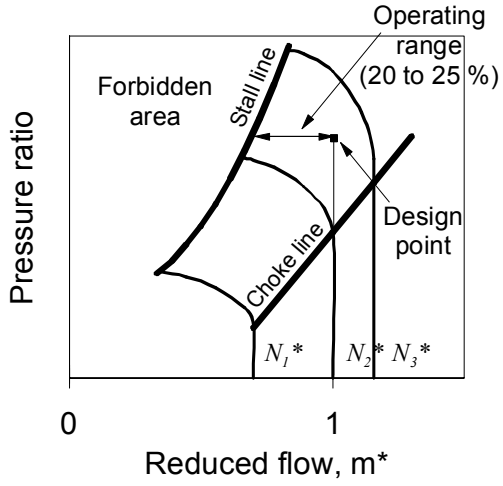
The alternative process is to perform compression of the vapour at low temperature, i.e. at its highest density. The pumps and recovery heat exchangers get smaller in size and less expensive, but the work of compression is then injected in the cycle at low temperature, so that the inevitable irreversibilities 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 [80]. Their pressure ratio limited to 2 to 3.5 per stage however imposes to arrange them in multistage configurations [81, 82], thus narrowing the operational range of the system, in particular for startup or off-design modes.

Depending on the operating temperature (2 K or 1.8 K), the "cold" compression cycle requires at least 4 or 5 stages in series in order to perform the overall pressure ratio of 45 to 80. The compressed helium is directly returned to the cold low-pressure (LP) stream of the 4.5 K refrigerator.

The main drawback of this cycle concerns turndown capability. The cold compressor set has to guarantee the same pressure ratio for any load. A typical operating field for hydrodynamic

compressors (Figure 16) displays the pressure ratio as a function of the reduced flow m^* and the reduced speed N^*. The working area is limited on the left side by the stall line, on the right by the choke line and on top by the maximum rotational speed of the drive. At constant pressure ratio, the compressor can handle a flow reduction of only about 20 % before reaching the stall line. Below 80 % of nominal, additional vapour generation by electrical heating must be used to compensate for the load reduction. Such a cycle is therefore not very compliant to turndown, and its operating cost is not optimised for part-load operation.

This led CERN to conduct, in view of the LHC project, a R&D programme on cold compressors, procuring from specialised industry three prototype hydrodynamic compressors of different designs [83-86] to investigate critical issues such as drive and bearing technology, impeller and diffuser hydrodynamics, mechanical and thermal design, as well as their impact on overall efficiency [87]. The choices eventually retained for the LHC series machines [88-90] are 3-phase electrical induction motor drives working at room temperature with rotational speed varying from 200 to 700 Hz, active magnetic bearings working at room temperature, axial-centrifugal (three-dimensional) impellers and fixed-vane diffusers (Figure 17 & 18).



$$m^* = \frac{\dot{m}}{\dot{m}_0} \cdot \sqrt{\frac{Tin}{Tin_0}} \cdot \frac{Pin_0}{Pin}$$

and

$$N^* = \frac{N}{N_0} \cdot \sqrt{\frac{Tin_0}{Tin}}$$

with: $\dot{m}$: mass-flow
Tin: inlet temperature
Pin: inlet pressure
N: rotational speed
Subscript 0: design contition

Fig. 16 Typical operating field of hydrodynamic compressor.

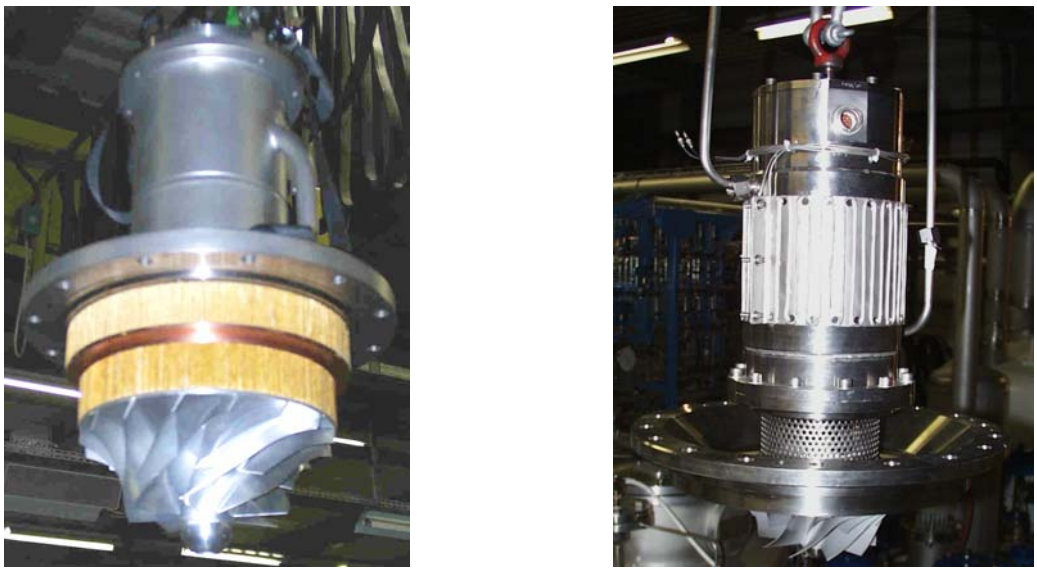


Fig. 17 Axial-centrifugal cold compressor cartridges for the LHC.

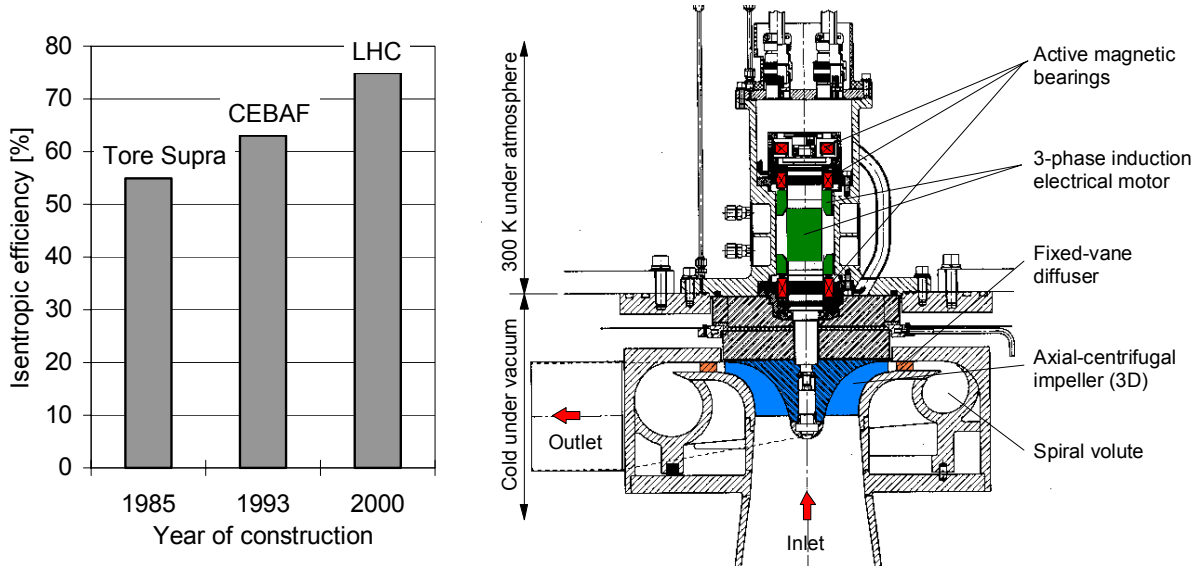


Fig. 18 Isentropic efficiency and typical cross-section of cold compressors.

Thermodynamic efficiency of cold compressors, determined by hydrodynamic design as well as by limitation of heat inleaks along the drive shaft, has significantly improved (Figure 18). Here the relevant estimator is isentropic efficiency, defined as the ratio of compression work in the adiabatic, reversible case, to the real one. Recent machines can be expected to reach 75 % isentropic efficiency at their design point.

3.4 "Mixed" compression cycles

For large systems, oil liquid-ring pumps or lubricated screw compressors may be used in series with cold compressors, in "mixed" compression cycles. Cold compressors are well-suited for the lower stages, while the presence of volumetric machines in the upper stages permits independent adjustment of flow-rate or wheel inlet conditions, thus improving load adaptation [91, 92].

In "mixed" compression cycles, the number of cold compressor stages can be reduced to 3, depending on the swept volume and number of warm sub-atmospheric machines. The compressed helium can be returned to the 4.5 K refrigerator at different levels:

- at the warm medium-pressure (MP) side (connection #1 on Figure 12 d). This requires the use of screw compressors having a sufficient built-in pressure ratio. In this case, the enthalpy of the gas at the outlet of the cold compressors has to be recovered by the heat exchangers of the 4.5 K refrigerator. The main advantage of this solution is that the same oil-removal and final cleaning systems can be used for the warm sub-atmospheric compressors and for the booster stages of the 4.5 K refrigerators, thus minimising the investment cost of the system.

- at the warm low-pressure (LP) side (connection #2 on Figure 12 d). This solution is compatible with the use of either screw compressors or liquid ring pumps. The enthalpy of the cold gas at the outlet of the cold compressors also has to be recovered by the heat exchangers of the 4.5 K refrigerator. In this case, the warm sub-atmospheric stage requires its own oil-removal system.

- at the cold low-pressure side (connection #3 on Figure 12 d). This is required when the enthalpy of the cold gas at the outlet of the cold compressors cannot be recovered by the heat exchangers of the 4.5 K refrigerators (LHC case) [88]. In this case, the warm sub-atmospheric stage requires its own oil-removal and final cleaning system (coalescers and charcoal adsorbers), increasing the investment cost.

The main advantage of the "mixed" cycle resides in its turndown capability. With sub-atmospheric compressors having volumetric characteristics, the pressure at the outlet of the cold compressors decreases linearly with the flow-rate, i.e. if the temperature and rotational speed do not change, the reduced flow-rate m* stays constant, thus keeping the working point fixed in the operating field. Such a cycle can then handle a large dynamic range, e.g. 3 for the LHC, without any additional electrical heating. Moreover, the total pressure ratio of the cold compressor train is lowered and the speed of some machines can then be reduced, thus decreasing the total compression power and operating cost.

Another operational advantage concerns the possibility of maintaining the load in cold standby with the cold compressors freewheeling and all compression performed, though at much reduced flow, by the warm machines. This mode allows repair or exchange of a cold-compressor cartridge without helium emptying of the system. In addition, the load adaptation provided by the warm volumetric machines proves very useful during transient modes like cool-down and pump-down, in which the cold compressors operate far from their design conditions.

The only drawback of this cycle concerns the risk of air inleaks due to the presence of sub-atmospheric circuits in air. Helium guards are recommended to prevent pollution of the process helium [93].

3.5 Application range of low-pressure helium compression techniques

The practical ranges of application of the different techniques appear in Figure 19, setting a *de facto* limit for warm compression above 20'000 $m^3.h^{-1}$, or typically 300 W at 1.8 K. The diagram also illustrates the large span of refrigeration power and diversity of projects using superfluid helium. Investment and operating costs of large superfluid helium refrigeration systems can be assessed from basic thermodynamics and practical scaling laws derived from recent experience [94], thus providing input for technical-economical optimisation of such systems.

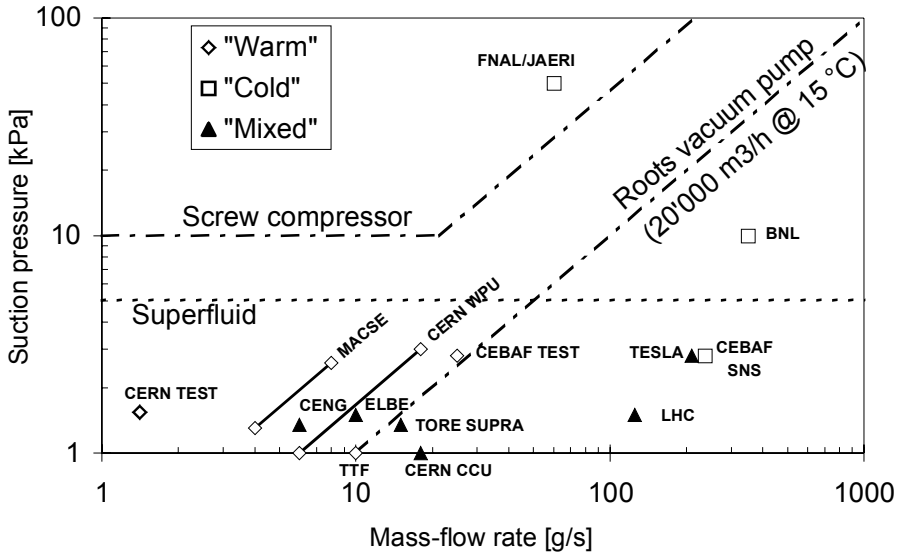


Fig. 19 Range of application of low-pressure helium compression techniques.

4. CONCLUSION

Operating superconducting devices at 1.8 K, using superfluid helium as a technical coolant, has now become state-of-the-art. The specific aspects of superfluid helium technology addressed in this article can be combined with standard cryogenic practice to design, build and operate complete industrial-

type helium II systems. The particle accelerator projects in construction or in study, however represent major challenges and opportunities for further progress, in view of their large size, complexity and quest for reliability and efficiency.

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] W.D. Markiewicz, I.R. Dixon, C.A. Swenson, W.S. Marshall, T.A. Painter, S.T. Bole, T. Cosmus, M. Parizh, M. King, G. Ciancetta, 900 MHz wide-bore NMR spectrometer magnet at NHMFL, IEEE Trans. Appl. Superconductivity 10 (2000) 728-731

[7] Ultrastabilized™ NMR magnets by Bruker Biospin Corporation, http//www.bruker-biospin.com/nmr/products/ustab.html

[8] 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

[9] G. Claudet, G. Bon Mardion, B. Jager & G. Gistau, Design of the cryogenic system for the Tore Supra tokamak, Cryogenics 26 (1986) 443-449

[10] L.R. Evans, LHC accelerator physics and technology challenges, Proc. PAC99, A. Luccio & W.W. MacKay editors, IEEE Piscataway, New Jersey, USA (1999) 21-25

[11] Ph. Lebrun, Superfluid helium cryogenics for the Large Hadron Collider project at CERN, Cryogenics 34 ICEC Supplement (1994) 1-8

[12] Ph. Lebrun, Cryogenics for the Large Hadron Collider, IEEE Trans. Appl. Superconductivity, 10 (2000) 1500-1506

[13] H. Padamsee, Superconducting RF cavity design, these proceedings

[14] C.H. Rode, CEBAF cryogenic system, Proc. PAC95, Am. Phys. Soc. & IEEE Piscataway, New Jersey, USA (1995) 1994-1998

[15] C. Rode & the JLab SNS team, The SNS superconducting linac system, Proc. PAC2001, P. Lucas & S. Webber editors, IEEE Piscataway, New Jersey, USA (2001) 619-623

[16] F. Richard, J.R. Schneider, D. Trines & A. Wagner editors, TESLA Technical Design Report, DESY 2001-011 and ECFA 2001-209 (2001)

[17] 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

[18] G. Horlitz, The cryogenic system for the superconducting e+ e- linear collider TESLA, Proc. Symposium on cryogenic systems for large-scale superconducting applications, NIFS-PROC-28, NIFS, Toki, Japan (1996) 85-89

[19] 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

[20] F. Vinen, Physics of superfluid helium, these proceedings

[21] L. Tavian, Large cryogenic systems at 1.8 K, Proc. EPAC 2000, J.L. Laclare, W. Mitaroff, Ch. Petit-Jean-Genaz, J. Poole & M. Regler editors, Austrian Academy of Sciences Press, Vienna, Austria (2000) 212-216

[22] G. Vandoni, Heat transfer, these proceedings.

[23] Ph. Lebrun, Design of a cryostat for superconducting accelerator magnet: the LHC main dipole case, these proceedings.

[24] 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

[25] B. Fallou, J. Galand & B. Bouvier, Dielectric breakdown of gaseous helium at very low temperature, Cryogenics 10 (1970) 142-146

[26] 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

[27] B. Rousset & F. Viargues, An alternative cooling scheme for the TeV superconducting linear accelerator project, Cryogenics 34 ICEC Supplement (1994) 91-94

[28] V. Arp, Heat transport through helium II, Cryogenics 10 (1970) 96-105

[29] C.J. Gorter & J.H. Mellink, On the irreversible processes in liquid helium II, Physica 15 (1949) 285-304

[30] 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

[31] S.W. van Sciver, Heat transfer in superfluid helium II, Proc. ICEC8, C. Rizzuto editor, IPC Science & Technology Press, Guildford, UK (1980) 228-237

[32] 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

[33] S.W. van Sciver, Developments in He II heat transfer and applications to superconducting magnets, Adv. Cryo. Eng. 27 (1982) 375-398

[34] G. Claudet & P. Seyfert, Bath cooling with subcooled superfluid helium, Adv. Cryo. Eng. 27 (1982) 441-449

[35] P. Seyfert, Results on heat transfer to He II for use in superconducting magnet technology, Proc. ICEC9, K. Yasukochi & H. Nagano editors, Butterworth, Guildford, UK (1982) 263-268

[36] 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, Brookhaven National Laboratory, USA (1986) 270-275

[37] C. Meuris, B. Baudouy, D. Leroy & B. Szeless, Heat transfer in electrical insulation of LHC cables cooled with superfluid helium, Cryogenics 39 (1999) 921-931

[38] 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

[39] S. Caspi, The use of calorimetry in superfluid He II to measure losses in superconducting magnets, Proc. ICEC9, K. Yasukochi & H. Nagano editors, Butterworth, Guildford, UK (1982) 347-350

[40] P.L. Walstrom, Joule Thomson effect and internal convection heat transfer in turbulent helium II flow, Cryogenics 28 (1988) 151-156

[41] 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, R.G. Scurlock & C.A. Bailey editors, Butterworths, Guildford, UK (1988) 497-504

[42] A. Kashani & S.W. van Sciver, Steady-state forced convection heat transfer in He II, Adv. Cryo. Eng. 31 (1986) 489-498

[43] A. Kashani & S.W. van Sciver, Transient forced convection heat transfer in He II, Proc. ICEC11, G. & I. Klipping editors, Butterworths, Guildford, UK (1986) 654-658

[44] 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

[45] 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

[46] 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

[47] 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

[48] H.A. Snyder & A.J. Mord, Flow of superfluid helium in tubes with heated walls, Adv. Cryo. Eng. 37A (1991) 81-88

[49] A.J. Mord, H.A. Snyder & D.A. Newell, End-to-end modelling of helium II flow systems, Cryogenics 32 (1992) 291-299

[50] G. Morpurgo, Design and construction of a pump for liquid helium, Cryogenics 17 (1977) 91-93

[51] 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, G. & I. Klipping editors, Butterworth, Guildford, UK (1986) 312-316

[52] S.W.K. Yuan & T.C. Nast, The design of fountain-effect pumps, Adv. Cryo. Eng. 33 (1988) 457-464

[53] P. Kittel, Operating characteristics of isocaloric fountain-effect pumps, Adv. Cryo. Eng. 33 (1988) 465-470

[54] 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

[55] P. Kittel, Losses in fountain-effect pumps, Proc. ICEC11, G. & I. Klipping editors, Butterworth, Guildford, UK (1986) 317-322

[56] G. Klipping & H.D. Denner, The thermomechanical effect of superfluid helium and its technical application, Adv. Cryo. Eng. 35A (1990) 81-93

[57] 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

[58] 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

[59] 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

[60] 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

[61] 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

[62] D. Camacho, S. Chevassus, C. Policella, J.M. Rieubland, G. Vandoni & R. van Weelderen, Thermal characterization of the He II LHC heat exchanger tube, Proc. ICEC17, D. Dew-Hughes, R.G. Scurlock & J.P. Watson editors, IoP Publishing (1998) 647-650

[63] 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, Adv. Cryo. Eng. 41A (1996) 777-784

[64] A. Bézaguet, J. Casas-Cubillos, H. Guinaudeau, B. Hilbert, Ph. Lebrun, L. Serio, A. Suraci, L. Tavian & R. van Weelderen, Cryogenic operation and testing of the extended LHC prototype magnet string, Proc. ICEC16/ICMC, T Haruyama, T. Mitsui & K. Yamafuji editors, Elsevier Science (1997) 91-94

[65] 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)

[66] B. Flemsaeter, E. Blanco, J. Casas-Cubillos, C. de Prada & S. Saelid, Applying advanced control techniques for temperature regulation of the LHC superconducting magnets, Proc. ICEC17, D. Dew-Hughes, R.G. Scurlock & J.P. Watson editors, IoP Publishing (1998) 631-634

[67] B. Rousset, A. Gauthier, L. Grimaud, A. Bézaguet & R. van Weelderen, Thermohydraulic behaviour of He II in stratified co-current two-phase flow, Proc. ICEC16/ICMC, T Haruyama, T. Mitsui & K. Yamafuji editors, Elsevier Science (1997) 519-522

[68] B. Rousset, A. Gauthier & A. Grimaud, Stratified two-phase superfluid helium flow : part 1, Cryogenics 37 (1997) 733-737

[69] L. Grimaud, A. Gauthier, B. Rousset & J.M. Delhaye, Stratified two-phase superfluid helium flow : part 2, Cryogenics 37 (1997) 739-744

[70] Ph. Lebrun, L. Serio, L. Tavian & R. van Weelderen, Cooling strings of superconducting devices below 2 K: the helium II bayonet heat exchanger, Adv. Cryo. Eng. 43A (1998) 419-426

[71] B. Rousset, B. Jager, E. di Muoio, L. Puech, P. Thibaud, R. Vallcorba, R. van Weelderen & P.E. Wolf, He II co-current two-phase flow at high vapor velocities, Adv. Cryo. Eng. 47B (2002) 1311-1318

[72] Y. Huang, J. Kerby, T. Nicol & T. Peterson, Cryogenic system and cryostat design for the LHC IR quadrupole magnets, Adv. Cryo. Eng. 43A (1998) 403-410

[73] R. Byrns, Y. Huang, J. Kerby, Ph. Lebrun, L. Morrison, T. Nicol, T. Peterson, R. Trant, R. van Weelderen & J. Zbasnik, The cryogenics of the LHC interaction region final-focus superconducting magnets, Proc. ICEC17, D. Dew-Hughes, R.G. Scurlock & J.P. Watson editors, IoP Publishing (1998) 743-746

[74] G. Gistau-Baguer, High-power refrigeration at temperatures around 2 K, Proc. ICEC16/ICMC, T Haruyama, T. Mitsui & K. Yamafuji editors, Elsevier Science (1997) 189-194

[75] F. Millet, P. Roussel, L. Tavian & U. Wagner, A possible 1.8 K refrigeration cycle for the Large Hadron Collider, Adv. Cryo. Eng. 43A (1998) 387-394

[76] E. Kashtanov, V. Pleskatch, K. Polkovnikov, A. Shembel, V. Sytnik, S. Zintchenko, B. Krakovsky, O. Krasnikova, O. Popov, H. Burmeister, B. Petersen & H.O. Roggenbuck, Large low-pressure heat exchanger for the TTF cryogenic system, Proc. ICEC18, K.G. Narayankhedkar editor, Narosa, New Delhi, India (2000) 315-318

[77] 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

[78] P. Roussel, A. Bézaguet, H. Bieri, R. Devidal, B. Jager, R. Moracchioli, P. Seyfert & L. Tavian, Performance tests of industrial prototype subcooling helium heat exchangers for the Large Hadron Collider, Adv. Cryo. Eng. 47B (2002) 1311-1318

[79] G.M. Gistau & G. Claudet, The design of the helium refrigerator for Tore Supra, Proc. ICEC10, H. Collan, P. Berglund & M. Krusius editors, Butterworth, Guildford, UK (1984) 288-291

[80] G.M. Gistau, J.C. Villard & F. Turcat, Application range of cryogenic centrifugal compressors, Adv. Cryo. Eng. 35B (1990) 1031-1037

[81] 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

[82] C.H. Rode, D. Arenius, W.C. Chronis, D. Kashy & M. Keesee, 2 K CEBAF cryogenics, Adv. Cryo. Eng. 35A (1990) 275-286

[83] L. Decker, K. Löhlein, P. Schustr, M. Vins, I. Brunovski, L. Tucek, Ph. Lebrun & L. Tavian, A cryogenic axial-centrifugal compressor for superfluid helium refrigeration, Proc. ICEC16/ICMC, T Haruyama, T. Mitsui & K. Yamafuji editors, Elsevier Science (1997) 195-198

[84] L. Decker, A. Kündig, K. Löhlein, W. Pürtschert, B. Ziegler, Ph. Lebrun, L. Tavian, I. Brunovsy & L. Tucek, Operational experience with a cryogenic axial-centrifugal compressor, Adv. Cryo. Eng. 43A (1998) 637-642

[85] M. Bonneton, L. Tavian, G. Gistau-Baguer, F. Turcat & P. Viennot, A high-reliability gas-driven helium cryogenic centrifugal compressor, Adv. Cryo. Eng. 43A (1998) 643-650

[86] N. Saji, H. Asakura, S. Yoshinaga, K. Itoh, T. Nogaku, A. Bézaguet, J. Casas, Ph. Lebrun & L. Tavian, A 1 kPa centrifugal cold compressor for the 1.8 K helium refrigeration system,

Proc. ICEC17, D. Dew-Hughes, R.G. Scurlock & J.P. Watson editors, IoP Publishing (1998) 295-298

[87] A. Bézaguet, Ph. Lebrun & L. Tavian, Performance assessment of industrial prototype cryogenic helium compressors for the Large Hadron Collider, Proc. ICEC17, D. Dew-Hughes, R.G. Scurlock & J.P. Watson editors, IoP Publishing (1998) 145-148

[88] S. Claudet, P. Gayet, B. Jager, F. Millet, P. Roussel, L. Tavian & U. Wagner, Specification of eight 2400 W @ 1.8 K refrigeration units for the LHC, Proc. ICEC18, K.G. Narayankhedkar editor, Narosa, New Delhi, India (2000) 207-210

[89] B. Hilbert, G.M. Gistau-Baguer & F. Dagut, 2.4 kW @ 1.8 K refrigeration units for the LHC project : the Air Liquide system, Proc. ICEC18, K.G. Narayankhedkar editor, Narosa, New Delhi, India (2000) 211-214

[90] H. Asakura, J. Boesel, T. Honda, A. Kuendig, K. Kurtcuoglu, A. Meier, M. Mori, A.E. Senn & S. Yoshinaga, Four 2400W/1.8 K refrigeration units for CERN-LHC: the IHI/Linde system, Proc. ICEC18, K.G. Narayankhedkar editor, Narosa, New Delhi, India (2000) 215-218

[91] 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)

[92] M. Kauschke, C. Haberstroh & H. Quack, Safe and efficient operation of multistage cold compressor systems, Adv. Cryo. Eng. 41A (1996) 931-926

[93] F. Minot, B. Gravil, A. Bocquillon, C. Dugas, D. Henry, J.L. Maréchal, S. Nicollet, P. Prochet, J.P. Serries & R. Simon, Problems encountered with subatmospheric circuits on the Tore Supra cryogenic system, Proc. ICEC17, D. Dew-Hughes, R.G. Scurlock & J.P. Watson editors, IoP Publishing (1998) 609-612

[94] S. Claudet, Ph. Lebrun & L. Tavian, Towards cost-to-performance optimisation of large superfluid helium refrigeration systems, Proc. ICEC18, K.G. Narayankhedkar editor, Narosa, New Delhi, India (2000) 203-206

ELEMENTS OF 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

Proceedings of the CERN Accelerator School “Superconductivity in particle accelerators”, CERN Report 96-03 (1996)

VACUUM AND CRYOPUMPING

Oswald Gröbner
Schmiedgasse 5, Innsbruck, A-6020 Austria
e-mail : Oswald.Groebner@chello.at

Abstract
Cold surfaces can provide large pumping speeds combined with a large pumping capacity and low ultimate pressures for all condensable gases. The pumping of hydrogen at low pressures and in large quantities represents a challenge for the vacuum engineer and requires particular attention. To have good performance cryopumps must be shielded from parasitic desorption which could be caused by synchrotron radiation, electron bombardment or even by room temperature thermal radiation. The LHC vacuum system with its actively cooled beam screen is given as an example for a system, which relies almost entirely on cryopumping on the cold bore surface.

1. INTRODUCTION

1.1 Ideal gas law

The ideal gas law is given by

$$P\ V = \frac{N}{No}\ R\ T \tag{1}$$

with : P pressure, V volume, T temperature, N number of molecules,

R gas constant = 8.31 kJ kmol^{-1} K^{-1}, No = 6.02 10^{26} molecules kmol^{-1}.

The molecular density $$n = N/V \tag{2}$$

Pressure and gas density are related $$P = n\ k\ T \tag{3}$$

Boltzmann constant k = 1.38 10^{-23} J/K and is related to the gas constant through $R = No\ k$.

Useful introductions to the principles of the kinetic theory of gases and can be found in Ref[1, 2].

It is important to note that in nearly all cases where uhv conditions are required, it is the gas density rather than the pressure, which really matters and which is the quantity of interest. Therefore, when quoting pressures in a cryogenic vacuum system, it is necessary to specify the temperature as well.

1.2 Molecular velocities

Molecular velocities in a system in thermal equilibrium follows the distribution of Maxwell Boltzmann. For vacuum calculations it is common practice to use the average molecular velocity, which is given by

$$\bar{v} = \sqrt{\frac{8kT}{\pi M m_o}} \text{ and numerically } \bar{v} \sim 150\sqrt{\frac{T}{M}}\ (m/s). \tag{4}$$

Here m_o is the unit mass (1.6 10^{-26} kg) and M is the molecular mass. Table 1 gives mean molecular velocities for some common gas molecules.

Table 1
Mean molecular velocities at 20°C (m/s)

H_2	N_2	Air	A	Kr
1754	470	464	393	272

1.3 Wall collisions

The frequency with which molecules collide with the walls of the system is directly proportional to the mean molecular velocity

$$\nu = \frac{1}{4} n \bar{v} \tag{5}$$

and to the molecular density. The pressure to the walls of the system is due to the momentum transfer of the molecules.

1.4 Mean free path

The mean free path of molecules

$$l = \frac{1}{\sqrt{2}\pi D^2 n} \tag{6}$$

with the molecular diameter D, which is of the order of ~3 10^{-8} m. One finds for air at room temperature (R.T.) and for a pressure of 1 Pa, a mean free path of about 1 mm. The behavior of a vacuum system depends very much on whether or not the mean free path is less or larger than the typical dimensions of the system. For l >> dimensions of then vacuum system one has molecular flow conditions. A consequence of the molecular flow regime, which is of interest for a vacuum system where parts are at different temperatures, is the effect of 'thermal transpiration'.

1.5 Thermal transpiration

In a vacuum system where different parts have different temperatures, the pressures and the gas densities for molecular flow are given by the relations

$$n_1\sqrt{T_1} = n_2\sqrt{T_2} \quad \text{or} \quad P_1/\sqrt{T_1} = P_2/\sqrt{T_2}\,. \tag{7}$$

To arrive at these relations it is convenient to consider a system with two different temperatures as shown in Fig. 1. Assuming that there is no net flow of gas between the two parts of the system, the rate with which molecules traverse the orifice in the separating wall from either side must be equal [3]. This equilibrium condition is satisfied when $\nu_1=\nu_2$. Using the expressions for the molecular incident rate and the mean molecular velocity these relations for thermal transpiration can be derived.

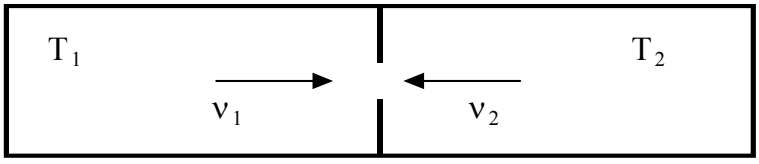


Fig. 1 Thermal transpiration.

Taking as an example the LHC vacuum system, a density of 10^{15} H_2 molecules/m^3 inside the beam screen, which has an average temperature of 15 K, would give a room temperature gauge reading (commonly expressed as a nitrogen equivalent pressure) of ~5 10^{-9} mbar.

1.6 Surface pumping

Molecules strike a surface with a rate given by Eq. (5). Since a molecule has a finite probability to stick to the surface (sticking probability s) this will result in a pumping speed which is proportional to the surface area, F, and which can be expressed as

$$S = s \cdot S_o = \frac{1}{4} \bar{v}\, s F\, n \tag{8}$$

S_o is the theoretical surface pumping speed. At R.T. S_o = 120 $m^3/s/m^2$ for nitrogen. Obviously, to construct a useful pump, it is necessary to ensure that the sticking probability, s, is sufficiently large. On this principle, a variety of uhv pumps have been developed [4]. Conventional pumps operating at R.T. use chemisorption as the physical process for trapping gas molecules. Since nearly any clean surface exhibits chemical affinity to most of the common gas molecules present in a vacuum system, it is sufficient to produce such a 'clean surface' by, e.g. sublimating or evaporating under vacuum a suitable material onto a substrate. Since the substrate can be the wall of the vacuum system, a compact and cost effective pump can be obtained. In this category of pumps one finds the titanium sublimation pumps and non-evaporable getter pumps (using Ti, Zr, V, Ba and other elements).

In presence of cold surfaces, it is possible to use the much weaker van der Waal's forces to adsorb gas molecules by physisorption and condensation.

1.7 Monolayer capacity, monolayer time

An important parameter for surface pumps is their capacity, since further gas molecules are no longer adsorbed when a monomolecular layer (monolayer, Θ_o) has been accumulated -> the pumping speed vanishes. Fortunately, the amount of gas present in a uhv system is in nearly all cases much less than a monolayer.

A rough estimate can illustrate this fact:

If one assumes an evacuated, closed sphere of 1 m^3 which has on its wall a gas coverage corresponding to one monolayer, this amounts to

$$N_{ads} = 4\pi r^2 \Theta_o \,. \tag{9}$$

In case this layer would be desorbed, it would produce a volume gas density

$$n = \frac{N_{ads}}{V} = \frac{3\Theta_o}{r}. \tag{10}$$

Using as a typical value for a monolayer of 3.10^{19} molecules/m^2 one arrives at a pressure of ~0.4 Pa, which is definitely not uhv. Conversely, the 'volume gas' in a uhv system is usually very small compared to the reservoir of the surface. It is also important to be aware of the fact that in a uhv system, the surface gas is a determining factor for the quality of the vacuum.

The time required to form a monolayer, i.e. the approximate useful operating time of a surface pump, can be estimated from

$$t = \frac{\Theta_o}{\frac{1}{4}\bar{v}\, s\, n}\,. \tag{11}$$

For a pressure of 10^{-6} Pa and s = 1 one calculated a time of 0.3 hours.

This result can be interpreted in two ways:

If one would like to study an atomically clean sample in a surface analysis system which takes several hours to prepare, it will be necessary to achieve pressures in the 10^{-9} Pa range.

Alternatively, in a cold system with a monolayer condensed on the surface, the displacement of the adsorbed gas will require about one hour at a pressure of 10^{-6} Pa (external pumps have been neglected).

1.8 Thermal desorption

The specific desorption rate, q(Pa $m^3/s/m^2$) from a surface as a function of temperature can be described in first approximation as

$$q = const \cdot e^{-\frac{E}{kT}} \tag{12}$$

where E represents the binding energy of the molecule to the surface. The constant factor is proportional to the surface concentration of gas. As a function of temperature, the desorption rate is given by

$$\log[q] = A - \frac{B}{T} \tag{13}$$

with two constants A and B. The binding energy for chemisorbed molecules is in the range of 80 kJ/mole or more, while it is less than 0.4 kJ/mole for physisorbed molecules. Adsorbed molecules oscillate in the potential well of the surface potential and have a finite probability to leave the surface and to enter the volume.

The average molecular residence time on the surface is

$$\tau = \frac{1}{\nu_o} \cdot e^{\frac{E}{kT}} \tag{14}$$

with the vibration frequency $\nu_o \sim 10^{13}$ s^{-1}. Using for the purpose of an illustration a binding energy of 30 kJ/mole, the average residence time of a molecule would amount to 10^{-7} s at R.T. but it would increase to 50 years at 80 K. This example shows that by cooling down a system the thermal outgassing rate will be reduced. The strong temperature dependence of the sojourn time of molecules is used routinely in a vacuum bakeout to reduce the outgassing rate by many orders of magnitude.

1.9 Adsorption-desorption processes

1.9.1 Langmuir isotherm

In a practical system, adsorption and desorption of molecules occur at the same time and there will be an equilibrium between the rate of adsorption and the rate of desorption. In the model of Langmuir, it is assumed that all molecules are adsorbed with a constant probability on all free adsorption sites. Under this assumption, the molecular adsorption rate

$$j_{ads} = s(1 - \frac{\Theta}{\Theta_o})\frac{n\bar{v}}{4} \tag{15}$$

and the molecular desorption rate $$j_{des} = \frac{\Theta}{\tau}\,. \tag{16}$$

In equilibrium the two rates must be equal, hence

$$\Theta = \frac{L\ p}{1 + L\ p} \quad \text{where} \quad L = \frac{s\tau\bar{v}}{4\,k\,T}. \tag{17}$$

For very small coverage and low pressures (*Henry isotherm*) the surface coverage depends linearly on the pressure.

1.9.2 Multilayer adsorption

According to the Langmuir isotherm, at most one single layer can exist. However, multiple layers accumulate and condense on cold surfaces. This process has been described by a multiplayer model, where the subsequent layers are adsorbed with an energy corresponding to the heat of evaporation. The BET isotherm (called after Brunauer, Emmett and Teller) can be expressed as

$$\Theta_{BET} = \frac{pL_{BET}}{(p_s - p)(1 + (L_{BET} - 1)\,{}^{p}\!/_{p_s})} \tag{18}$$

here p_s is the saturated vapour pressure of the adsorbate and L_{BET} is the ratio of the mean sojourn times for the first and for the subsequent adsorbed layers.

2. CRYOPUMPING

Cryopumping occurs when molecules are adsorbed on a cold surface so that the resulting sojourn time is long [[5]]. For practical reasons three mechanisms of cryopumping can be distinguished:

2.1.1 Sorption

Adsorption of molecules on a substrate with sub-monolayer coverage. To increase the pumping capacity, the effective surface area of the cryosorber should be increased with a coating with a large specific surface area e.g. activated charcoal or molecular sieve. As long as the coverage is well below a monolayer, the redesorption of gas, i.e. the vapour pressure of the adsorbed molecules, remains small and thus, sorption pumps are able to provide very low limit pressures for ultra high vacuum applications.

2.1.2 Condensation

Adsorption in multi-layers on the same molecular species. Since the heat of evaporation is less than the adsorption energy on a heterogeneous substrate, a limitation will occur due to the vapour pressure of the condensed gas. Nevertheless, below 20 K only Ne, H_2 and He have a significant saturated vapour pressure and will require special precaution.

2.1.3 Cryotrapping

Cryo-sorption of a gas mixture, e.g. H_2 or He with a high vapour pressure in the presence of an easily condensable carrier gas e.g. Ar. The 'volatile' species are effectively trapped in a thick layer of the condensate of the carrier gas.

2.2 Characteristics of cryopumping

According to Eq. 8 large pumping speeds can be obtained, proportional to the cryosorbing surface, *F*. The pumping speed decreases when the saturated vapour pressure of the adsorbed gas layer, p_s, is comparable to the system pressure.

$$S = S_o(1 - {}^{p}\!/_{p_s}) \tag{19}$$

The pumping speed S_o will be close to the theoretical limit since s~1 for condensation. At 20 K all gases with the exception of He, H_2 and Ne can be condensed in large quantities compatible with uhv conditions. At a temperature of 4.5 K pumping of large quantities of H_2 requires cryosorbing substrates with large specific surface area, to stay well away from the saturated vapour pressure regime. Pumping of He, even in small quantities, is difficult and hence it is essential to avoid helium leaks. Figure 2 shows vapour pressure curves for some gases, in particular those, which are common in a uhv system.

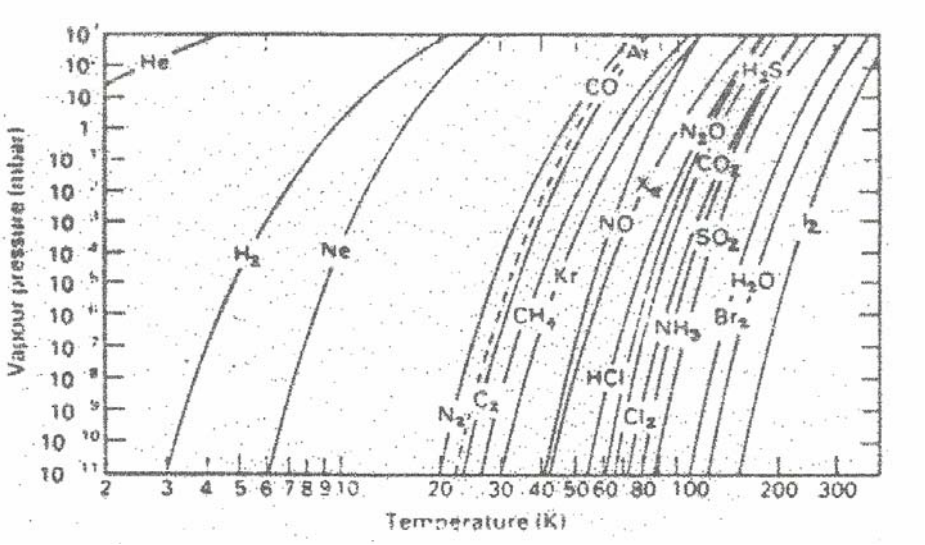


Fig. 2 Vapour pressures of common gases.

The capacity of a cryopump depends on the substrate onto which the molecules are adsorbed primarily on the roughness factor of the surface. For technical surfaces this roughness factor can be of the order of 10. The effect of cryosorption onto different substrates is illustrated in Fig. 3, which shows that for porous, anodised aluminium the largest capacity is obtained [6].

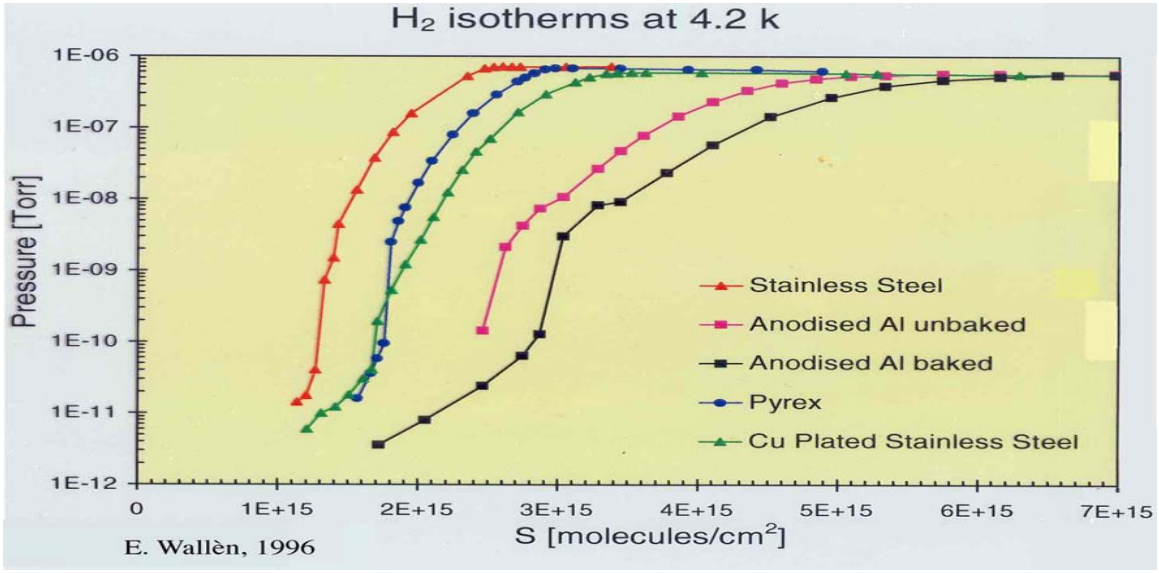


Fig. 3 Hydrogen isotherms at 4.2 K for different substrates.

2.3 Hydrogen vapour pressure below 4.7 K

To be able to pump hydrogen at very low pressure with a condensation cryopump, it is important to ensure that the saturated vapour pressure remains low and, e.g. compatible with the operating pressure in the LHC machine. Figure 4 shows the results of measurements at CERN prior to the LHC project [7], which have provided valuable input for the design of the LHC. It can be seen that at a temperature below ~3 K, the saturated vapour pressure of H_2 is well within the uhv range.

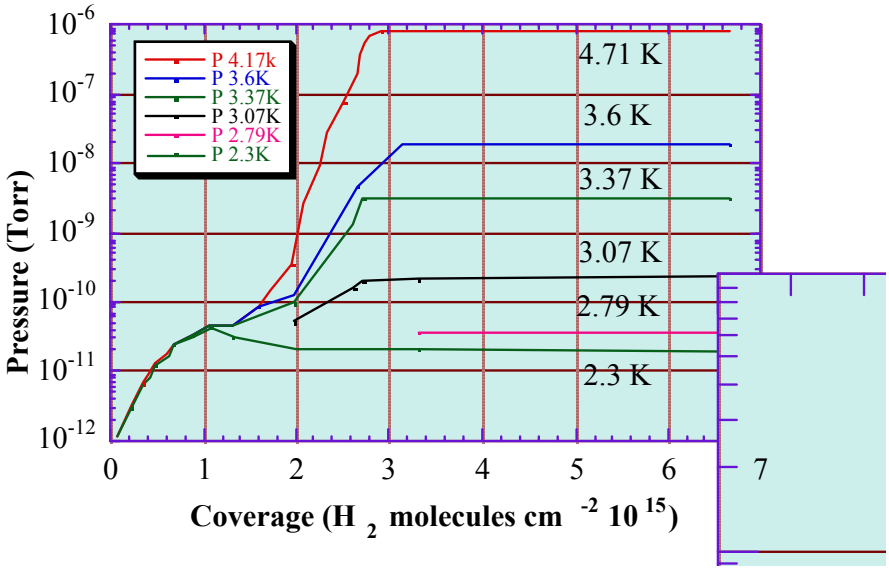


Fig. 4 Hydrogen isotherms for temperatures between 2 and 4.7 K.

On the contrary, at 4.5 K, this pressure would increase to an unacceptable value of 10^{-6} mbar. At the cold bore temperature of 1.9 K, the hydrogen vapour pressure in the LHC becomes negligible. Nevertheless, when the cold bore warms up during a magnet quench, the cryosorbed hydrogen will be released. Since one has a virtually closed system, the possibility to pump the released gas longitudinally by external pumps is very limited. It is expected, that the gas will either be transferred to the adjacent cold parts or will be readsorbed locally when the system is cooled down again.

The validity of the curves in Fig. 4 is subject to the condition that the cold surface is completely shielded from external room temperature radiation. It has been found that the absorption of a few mW/cm^2 of room temperature thermal radiation is sufficient to cause an abnormal increase of the vapour pressure for hydrogen [8].

2.4 Real surfaces and practical situations

Adsorption sites have a range of adsorption energies, which will modify the simple description of the isotherm. Consequently, the substrate and the surface topology (roughness, porosity) have an important influence, as has been shown in Fig. 3.

Porous substrate: adsorption sites inside narrow pores have a larger binding energy and hence result in a lower vapour pressure [9].

Adsorption of most diatomic molecules is dissociative and thus requires two free adsorption sites. Surface mobility of ad-atoms to sites with high binding energies can be an important process leading to clustering of molecules. This effect has been observed with the evolution of the CO_2 vapour pressure as a function of time after the exposure to the gas [10].

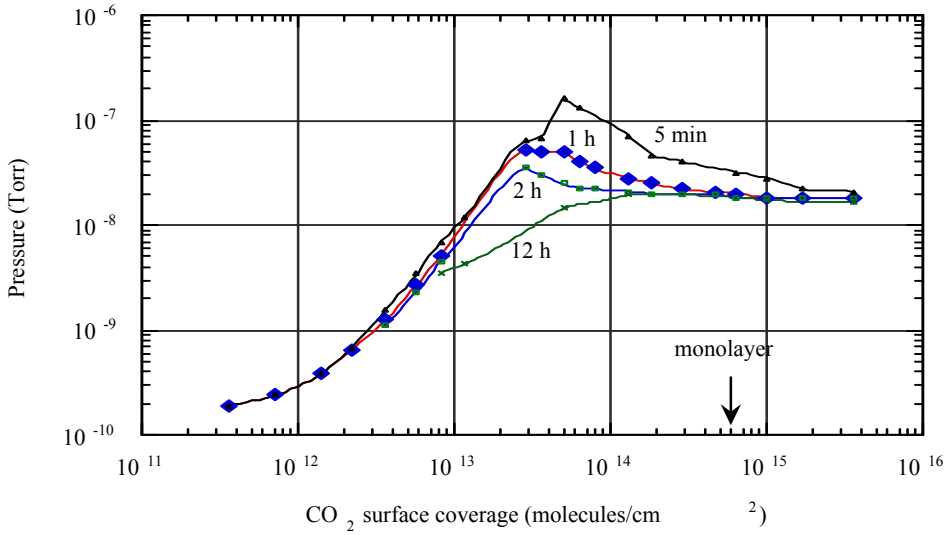


Fig. 5 Vapour pressure of cryosorbed CO_2 as a function of surface coverage and time.

Finally, in most practical vacuum systems more than one gas species is present and has to be pumped. In this case the effect of cryotrapping of one species by an other may occur. Figure 6 shows the effect of co-adsorption of CO_2 on the vapour pressure of H_2 [11].

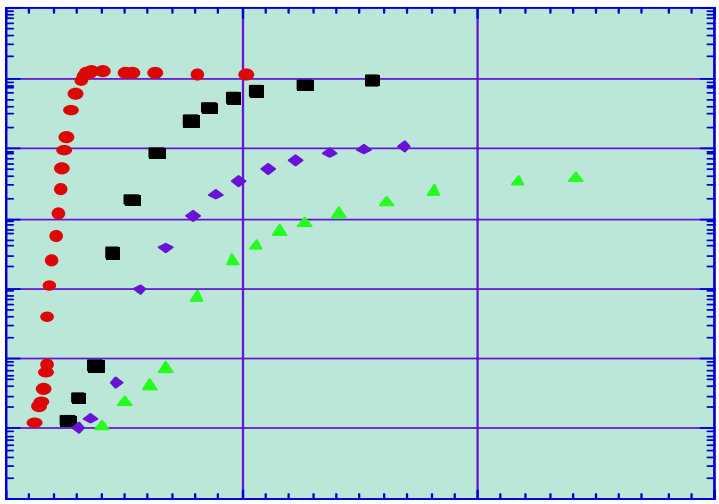


Fig. 6 Lowering of the H_2 vapour pressure with the co-adsorption of CO_2.

Even though this method may not be practical for an accelerator vacuum system, it could be used in special applications to lower the vapour pressure of cryotrapped H_2 molecules by several orders of magnitude.

2.5 Cryopumping in an accelerator environment

In combination with superconducting magnets or accelerating cavities, at little (or no) extra cost very effective integrated cryopumps can be obtained in otherwise strongly conductance limited vacuum systems. The vacuum engineer has indeed a large freedom in the design of cryopumps since the cold walls of the vacuum system act as pumps (LHC). Nevertheless, it must not be overlooked that there are limitations due to the exposure to environmental room temperature radiation and due to the bombardment by beam induced energetic particles (photons, electrons, ions), which must be avoided.

For stand-alone pumps this imposes the use of (liquid nitrogen) cooled baffles and for the LHC an actively cooled beam screen, which is shown in Fig. 7. For the vacuum system the requirement for screening arises not only for heat load reasons but mainly to ensure efficient pumping and to avoid re-desorption of molecules [12].

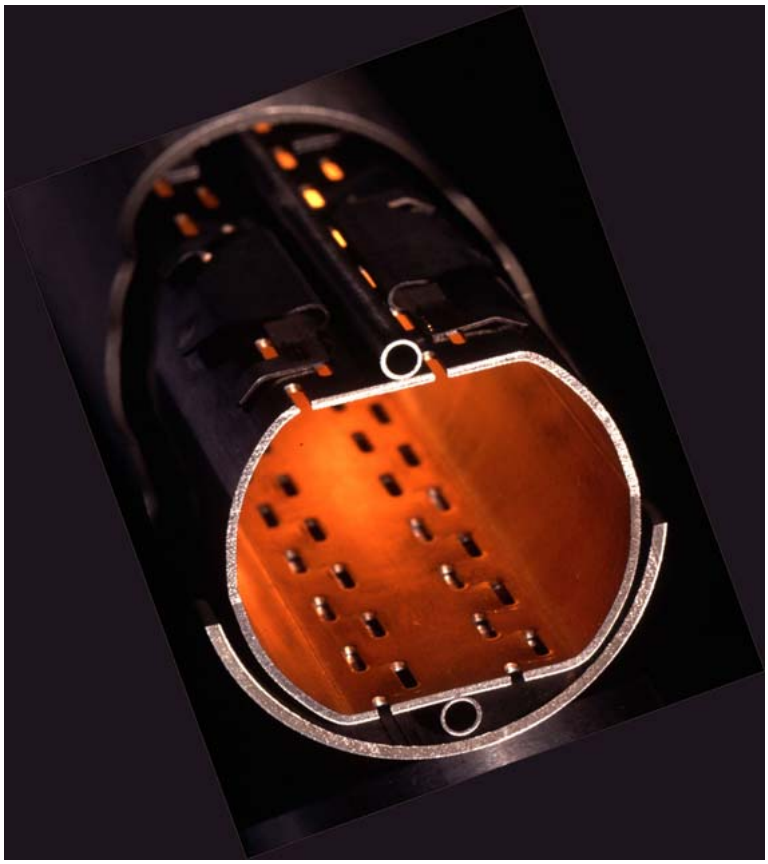


Fig. 7 LHC beam screen with cooling capillaries inside the magnet cold bore.

2.5.1 Cryogenic heat load from nuclear scattering

For very high energy beams, the heat load to the cryogenic system due to the power deposition by inelastically scattered particles, i.e. protons in the LHC, will become a significant effect. Since a fraction of this beam loss cannot be eliminated by collimation systems, it is distributed around the machine and deposits heat in the coils of magnets and hence in the 1.9 K system. The heat load, P_{ns} can be expressed in terms of the nuclear scattering beam lifetime τ, the beam current, I and the beam energy, E. Here c is the particle velocity.

$$P_{ns} = \frac{I\,E}{c\,\tau}\,, \tag{20}$$

and numerically $$P_{ns}(W/m) = 0.93 \frac{I(A)\, E(TeV)}{\tau(h)} \,. \qquad (21)$$

For the LHC, a cryogenic budget of 0.1 W/m has been allocated, which puts a direct requirement for the beam lifetime of more than 100 h. With the nuclear scattering cross section for high energy protons, (for 7 TeV protons on hydrogen atoms ~5×10^{-30} m^2) this lifetime can be converted in a corresponding gas density, called 'lifetime limit density', amounting to 10^{15} H_2/m^3. For heavier gas species a correspondingly smaller density is required.

2.5.2 The LHC cold bore : a cryopump exposed to synchrotron radiation

Any surface exposed to synchrotron radiation, Γ (photons/s), generates desorption of strongly bound, chemisorbed gas species, which increases the outgassing rate and hence the pressure in the vacuum system [13]. This effect is characterised by a 'primary' molecular desortpion yield (η molecules/photon). The photon stimulated desorption rate is proportional to $\Gamma\eta$. Effectively, it constitutes a source of 'new' gas molecules, which has to be removed by the vacuum pumps. With its critical photon energy of 45 eV the synchrotron radiation in the LHC is very efficient to desorb large amounts of gas. Figure 8 shows the primary desorption yield for the main desorbed gas species H_2, CH_4, CO and CO_2 for an OFHC copper surface as a function of the photon dose. In this figure is also shown the effect of a prolonged exposure to photons and the gradual 'clean-up' of the vacuum chamber wall with photon dose which improves the vacuum conditions during operation. It has been estimated that about one order of magnitude improvement will be requred to operate the LHC with ultimate beam currents.

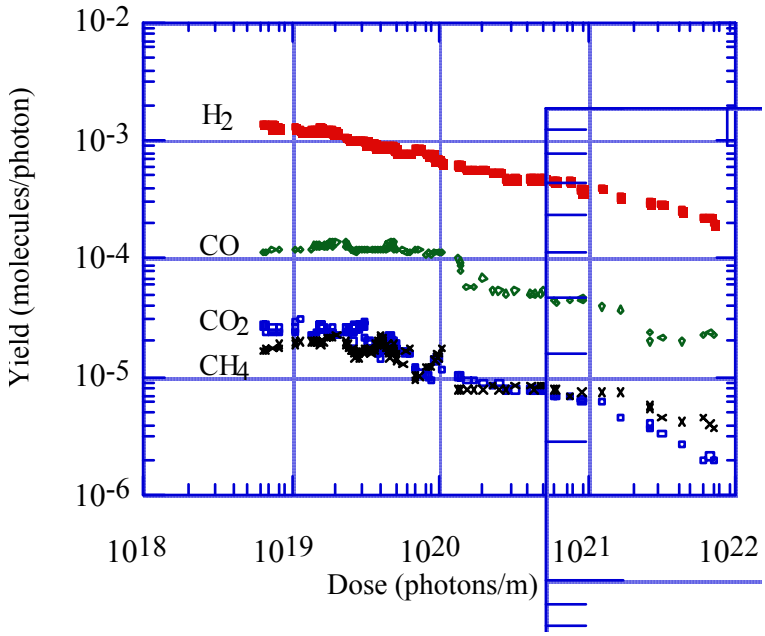


Fig. 8 Primary desorption yield for OFHC copper as a function of the photon dose.

When a cryosorbing surface is exposed to synchrotron radiation or even to the low energetic thermal radiation from the surrounding room temperature walls, the physisorbed, weakly bound molecules on the surface will be desorbed very effectively. As a result the initial balance between adsorption and thermal desorption will be affected and an additional desorption term has to be included :

$$j_{des} = \frac{\Theta}{\tau} + \kappa\Gamma \,. \qquad (22)$$

κ is a desorption yield (molecules/photon) from the physisorbed gas layer. This quantity has been called 'recycling yield' since it constitutes not a source of 'new' gas but rather the release of existing molecules, which have been pumped on the cold surface. Depending on the relative importance of these effects, this additional term in Eq. 22 may dominate the pumping mechanism. Effectively, this process shortens the mean sojourn time of molecules and hence shifts the balance of the isotherm to low surface coverage. Figure 9 shows measured recycling yields for a copper substrate as a function of the condensed gas layer. It can be seen that the desorption yield of weakly bound, physisorbed molecules increases with surface coverage and reaches a maximum value at very high coverage [14].

It can be seen that hydrogen is the most critical gas for the LHC since its recycling coefficient increases to a value close to unity at a monolayer coverage already. Therefore, contrary to the other gas species, the recycling coefficient for H_2 is significantly larger than the primary desorption yield.

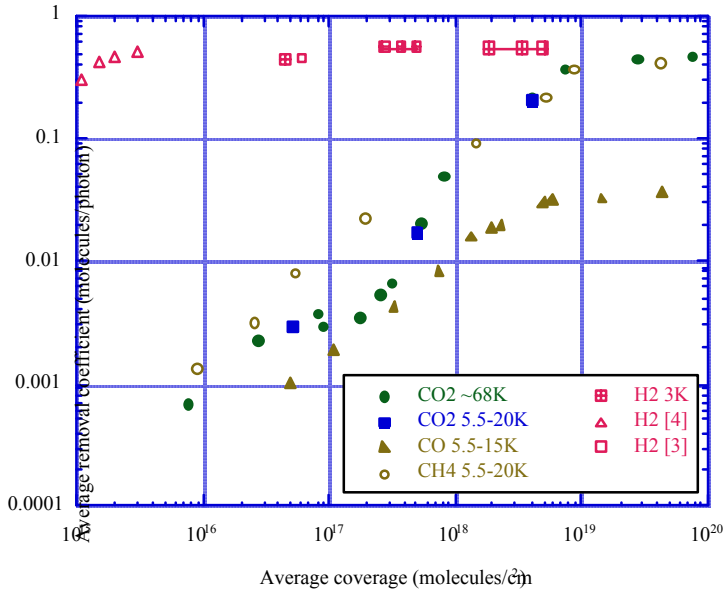


Fig. 9 Recycling yields of cryosorbed H_2, CH_4, CO and CO_2 molecules.

In the LHC vacuum system with its large flux of synchrotron radiation, (about 10^{17} photons m^{-1} s^{-1}) this recycling effect of physisorbed gas dominates the wall pumping of H_2 to such an extent, that any directly exposed surface looses all useful pumping efficiency.

In the early design of the LHC, a beam screen has been proposed for cryogenic reasons to intercept the synchrotron radiation power of ~ 0.2 W/m [15]. It has been recognised only later that a vital purpose of the beam screen for the vacuum system is to shield the cryosorbing surface of the 1.9 K cold bore from synchrotron radiation photons - and for that matter also from any other beam produced effects like photoelectrons and ions which can all contribute to the desorption of the physisorbed gas.

2.6 The dynamic pressure in the LHC beam screen

The gas density inside the beam screen of the LHC is described by the interaction between the volume density, $n(m^{-3})$ and the surface density, $\theta(m^{-2})$ of adsorbed molecules [16]. Since the system must be considered as conductance limited no longitudinal flow of gas occurs and therefore, longitudinal diffusion can be neglected. The volume and the surface gas can be described by a set of equations:

$$V \frac{dn}{dt} = q - a\, n + b\, \theta \tag{23}$$

and

$$F\frac{d\theta}{dt} = c\ n - b\ \theta \tag{24}$$

Here, V is the volume and F is the net wall area per unit length of the beam screen excluding the area of the pumping holes.

q represents a source of gas, here the photon induced primary desorption rate $q = \eta\ \Gamma$ (molecules s^{-1} m^{-1}).

a represents the total pumping on the surface of the wall and through the holes in the beam screen.

$$a = \frac{1}{4}\ \bar{v}\ F\ (s + f) \tag{25}$$

s is the sticking probability of the molecules on the beam screen and f is the fraction of the total surface of the beam screen with pumping holes, f<<1.

b represent the gas source originating from the adsorbed surface phase, θ, by thermal and by photon induced desorption.

$$b = F\ \nu_O\ e^{-\frac{E}{kT}} + \kappa\Gamma \tag{26}$$

c is the rate of adsorption of gas molecules on the surface of the beam screen only. This expression contains the sticking probability of the molecules, s,

$$c = \frac{1}{4}\ \bar{v}\ s\ F. \tag{27}$$

This model contains a small number of parameters and in its simplest form implies a linear adsorption isotherm, i.e. Henry's law for low pressures. To a first approximation, one may assume constant parameters; more generally, this will not be valid since in particular, the desorption yield and the sticking coefficient depend on coverage.

The equilibrium between surface and volume gas is

$$\theta = \frac{c}{b}\ n = \frac{1/4\ \bar{v}\ s\ F}{F\nu_O\ e^{-E/kT} + \kappa\Gamma}\ n \tag{28}$$

This result illustrates that the equilibrium coverage θ(n) depends on the re-cycling cross-section as well as on the photon flux and therefore, on the specific conditions and on the beam current in the LHC.

2.7 Dynamic vacuum evolution during synchrotron radiation exposure

Since any scattered/reflected photons re-desorb (i.e. '*recycle*') molecules at a rate increasing with coverage this leads to an increasing gas density. More and more molecules escape through the pumping holes and are permanently adsorbed on the shielded 1.9 K cold bore. Ultimately, this effect stabilises the gas density in the beam pipe at a value which should be safe for machine operation, i.e. to $<10^{15}$ molecules/m^3. Figure 10 shows the gas density (pressure rise) during photon exposure of an LHC test section (COLDEX) in an external beam line at the EPA storage ring at CERN [17]. The initial pressure rise and the subsequent stabilisation at an equilibrium value can clearly be seen.

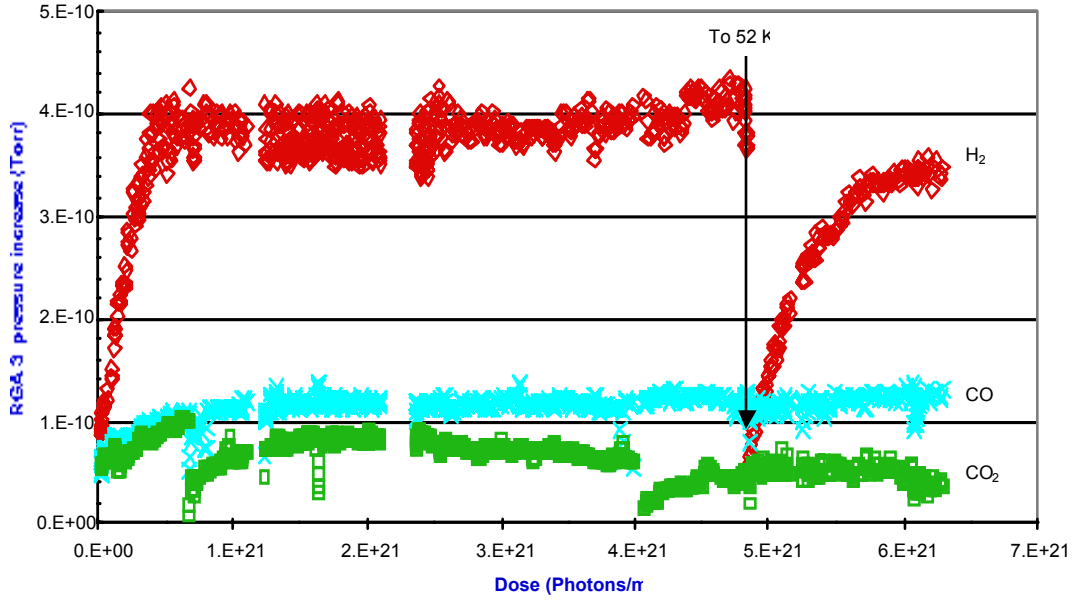


Fig. 10 Dynamic pressure rise in the COLDEX system at EPA.

2.8 Recycling coefficient of cryosorbed molecules

The transient pressure response when exposing a beam screen covered with an 'excess' amount of hydrogen to synchrotron radiation is shown in Fig. 11. The transient pressure spike is provoked by the recycling process of hydrogen ultimately leading to a new equilibrium pressure after several hours of exposure.

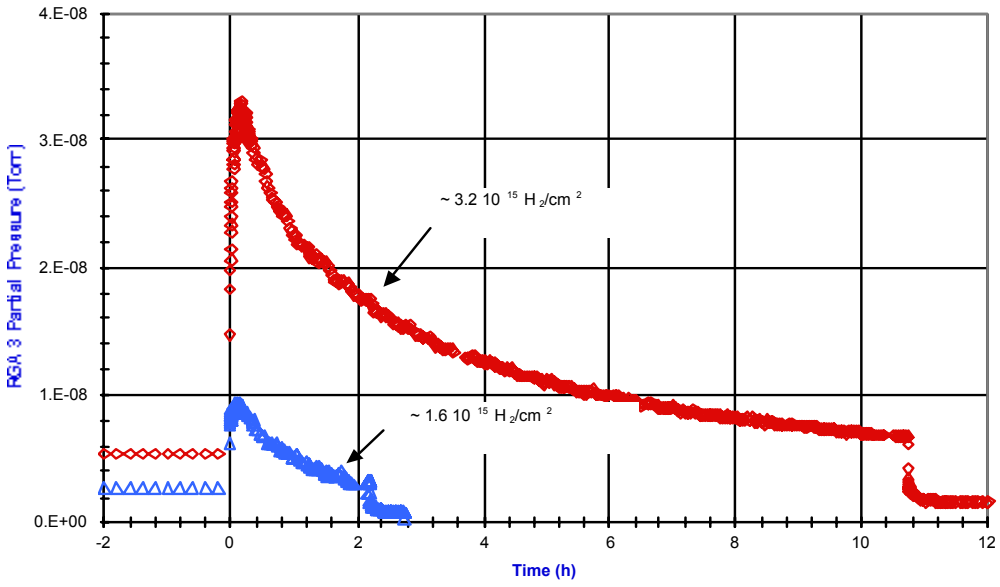


Fig. 11 Recycling of an excessive coverage of H_2 leads to a transient pressure spike.

The vertical scale in Fig. 11 gives the H_2 partial pressure in Torr. The new equilibrium state between volume and surface gas densities is obtained under the given conditions of photon stimulated desorption after approximately 10 hours. Inspection of the equilibrium condition in Eq. 28 shows that an unbalance between surface and volume gas densities, with a subsequent transient pressure response in the system, may be provoked not only by photon stimulated desorption, but equally by a change of the wall temperature.

3. SUMMARY

Adsorption/desorption phenomena dominate the vacuum behaviour at cryogenic temperature. Thermal desorption is largely absent, limited primarily by the vapour pressure of the adsorbate. Temperature stability of the cryosorbing surface is critical to avoid the release of gas and pressure spikes.

Cryopumps offer very large pumping speeds for nearly all gas species and produce a clean, hydrocarbon free vacuum. In combination with superconducting elements, cryopumps offer attractive solutions to achieve uhv in a conductance limited vacuum system. In the LHC, the pumping of H_2 relies entirely on crypo-pumping on the 1.9 K magnet bore. The cold bore offers a practically unlimited capacity for all species except for helium.

Absence of thermal desorption at cryogenic temperature and large pumping speed of the walls implies that extremely low pressures can be achieved. It is worth mentioning that a pressure of 10^{-17} Torr has been claimed for the Penning trap experiment at CERN as deduced from the storage time of antiprotons [18].

ACKNOWLEDGEMENTS

The material presented in this review is based on a large number of contributions from my former colleagues in the LHC vacuum group. In particular, I would like to acknowledge the extensive work by Drs. V. Baglin, R. Calder, I.R. Collins, G. Mullard, E. Wallèn and by the late A. G. Mathewson. Discussions with Dr. C. Benvenuti concerning the influence of thermal radiation on the vapour pressure of condensed H_2 are gratefully acknowledged.

REFERENCES

[1] Foundations of Vacuum Science and Technology, Edited J.M. Lafferty, John Wiley & Sons, 1998

[2] Handbuch Vakuumtechnik, M. Wutz, H. Adam, W. Walcher, K. Jousten, Vieweg, 2000

[3] The Physical Basis of Ultrahigh Vacuum, P.A. Redhead, J.P. Hobson, E.V. Kornelsen, American Vacuum Society Classics, American Institute of Physics, 1993

[4] C. Benvenuti, Molecular surface pumping: the getter pumps, CERN Accelerator School, Vacuum Technology, CERN 99-05, 1999

[5] R. A. Haefer, Cryopumping Theory and Practice, Clarendon Press, Oxford, 1989

[6] E. Wallèn, J. Vac. Sci. Technol. A 15(2) Mar/Apr. 1997

[7] C. Benvenuti, R.S. Calder, G. Passardi, J. Vac. Sci. Technol. Vol 13, No 6, Nov,/Dec. 1976

[8] C. Benvenuti, Molecular surface pumping: cryopumping, CERN Accelerator School, Vacuum Technology, CERN 99-05, 1999

[9] G. Moullard, B. Jenninger, Y. Saito, Vacuum Vol 60 pp 43-50, 2001

[10] V.V. Anashin, O.B. Malyshev, R. Calder, O. Gröbner, A. Mathewson, Vacuum Vol 48, No 7-9, pp 785-788, 1997

[11] B. Angerth, F. Bertinelli, J-C. Brunet, R. Calder, J. Gomez-Goñi, O. Gröbner, A. Mathewson, A. Poncet, C. Reymermier, E. Wallèn, EPAC 94, London, 26 June-1 July 1994

[12] O. Gröbner, Vacuum, 60 pp 25-34, 2001

[13] O. Gröbner, Dynamic outgassing, CERN Accelerator School, Vacuum Technology, CERN 99-05, 1999

[14] V.V. Anashin, O.B. Malyshev, R. Calder, O. Gröbner, Vacuum, 53, pp 269-272, 1999

[15] The LHC Study Group, The Large Hadron Collider Accelerator Project, CERN/AC/93-03 (LHC), 1993

[16] R. Calder et. al., J. Vac. Sci. Technol. A 14(4), Jul/Aug 1996

[17] V. Baglin, I.R. Collins, C. Grünhagel, O. Gröbner, B. Jenninger, EPAC, Vienna, 2001

[18] G. Gabrielse et. al., CERN-PPE/90-98, 1 July 1990

PARTICIPANTS

AFONSO, M.	Babcock Noell Magnettechnik GmbH, Zeitz, Germany
ALLITT, M.P	CERN, Geneva, Switzerland
AMBROSIO, G.	Fermilab, Batavia, IL, USA
ANTOINE, C.	CEA Centre de Saclay, Paris, France
AXENSALVA, J.	CERN, Geneva, Switzerland
BAJKO, M.	CERN, Geneva, Switzerland
BARBERO SOTO, E.	CERN, Geneva, Switzerland
BATE, R.	Daresbury Laboratory, Warrington, UK
BENEDICO MORA, E.	CERN, Geneva, Switzerland
BLAQUIERE, M.	CEA Centre de Saclay, Paris, France
BOUTBOUL, T.	CERN, Geneva, Switzerland
BOZZINI, D.	CERN, Geneva, Switzerland
BRINKMANN, A.	DESY, Hamburg, Germany
BUTIN, F.	CERN, Geneva, Switzerland
BUZIO, M.	CERN, Geneva, Switzerland
CALVI, M.	CERN, Geneva, Switzerland
CAMPI, D.	CERN, Geneva, Switzerland
CAPATINA, O.	CERN, Geneva, Switzerland
CASAGRANDE, F.	ORNL, Oakridge, TN, USA
CECCARELLE, G.	INFN-LNF, Frascati, Italy
CECCARELLI, R.	INFN-LNF, Frascati, Italy
CHAMIZO, R.	CERN, Geneva, Switzerland
CHEN, E-S.	SRRC, Hsinchu, Taiwan
CHIARELLI, R.	Ecole Centrale Paris, France
CHOHAN, V.	CERN, Geneva, Switzerland
COCCOLI, M.	LBL, Berkeley, CA, USA
CRAIEVICH, P.	Sincrotrone Trieste, Italy
DE GIORGI, M.	INFN-LNF, Frascati, Italy
DELLE MONACHE, G.	INFN-LNF, Frascati, Italy
DENARIE, C.-H.	CERN, Geneva, Switzerland
DENIAU, L.	CERN, Geneva, Switzerland
DEY, J.	Fermilab, Batavia, IL, USA
DI GENNARO, E.	University of Naples 'Federico II', Naples, Italy
DI GIACOMO, M.	GANIL, Caen, France
EICHHORN, R.	GSI, Darmstadt, Germany

FABRE, C.	CERN, Geneva, Switzerland
FABRIS, A.	Sincrotrone Trieste, Italy
FAZILLEAU, P.	CEA Centre de Saclay, Paris, France
FESSIA, P.	CERN, Geneva, Switzerland
FRATINI, M.	University of Pisa, Pisa, Italy
GASTINEAU, B.	CEA Centre de Saclay, Paris, France
GERWIG, H.	CERN, Geneva, Switzerland
GINOCCHIO, S.	Edison Termoelettrica SpA, Trofarello, Italy
GOMES, P.	CERN, Geneva, Switzerland
GONZALES, J.	CIEMAT, Madrid, Spain
GRANATA, V.	CERN, Geneva, Switzerland
GRUEHAGEN, H.	CERN, Geneva, Switzerland
GUBELLO, G.	CERN, Geneva, Switzerland
HEIKKILA, A.	Helsinki Institute of Physics, Helsinki, Finland
HELZEL, W.	DESY, Hamburg, Germany
HOFFMANN, M.	Bonn University, Bonn, Germany
IRISO ARIZ, U.	CERN, Geneva, Switzerland
JUSTER, F-P.	CEA Centre de Saclay, Paris, France
KARPPINEN, M.	CERN, Geneva, Switzerland
KAUSCHKE, M.	GSI, Darmstadt, Germany
KIRBY, G.	CERN, Geneva, Switzerland
KNABBE, E-A.	DESY, Hamburg, Germany
KOMOROWSKI, P.	CERN, Geneva, Switzerland
KOSTIN, D.	DESY, Hamburg, Germany
KUIJPER, A.	CERN, Geneva, Switzerland
KUTZSCHBACH, A.	Lehrstuhl für Kälte- und Kryotechnik, Dresden, Germany
LA CHINA, M.	CERN, Geneva, Switzerland
LANZA, C.	CERN, Geneva, Switzerland
LE NAOUR, S.	CERN, Geneva, Switzerland
LIGI, C.	INFN-LNF, Frascati, Italy
LIN, M-C.	SRRC, Hsinchu, Taiwan
MAGGIORE, M.	INFN-LNS, Catania, Italy
MAHNER, E.	CERN, Geneva, Switzerland
MARDOR, I.	SOREQ NRC, Yavne, Israel
MARINI, J.	LAL, Orsay, France
MARTIN, M.	BESSY, Berlin, Germany

METRAL, E.	CERN, Geneva, Switzerland
MIELE, P.	CERN, Geneva, Switzerland
MILCENT, H.	CERN, Geneva, Switzerland
MILES, J.	CERN, Geneva, Switzerland
MITRA, A.	TRIUMF, Vancouver, Canada
MUEHLE, C.	GSI, Darmstadt, Germany
NAGLER, A.	SOREQ NRC, Yavne, Israel
NAKAMOTO, T.	KEK, Ibaraki, Japan
NARDUZZO, A.	University of Oxford, Oxford, UK
OLRY, G.	IPN, Novosibirsk, Russia
PAULETTA, S.	CERN, Geneva, Switzerland
PAYN, A.	CEA Centre de Saclay, Paris, France
PENCO, G.	Sincrotrone Trieste, Italy
PIROTTE, O.	CERN, Geneva, Switzerland
PISCHALNIKOV, Y.	FNAL, Batavia, IL, USA
POJER, M.	CERN, Geneva, Switzerland
PRIN, H.	CERN, Geneva, Switzerland
QUINTIERI, L.	INFN-LNF, Frascati, Italy
RAINERI, S.	Edison Termoelettrica Spa, Trofarello, Italy
RANSOM, J.	University of Saskatchewan, Saskatoon, Canada
RICCI, R.	INFN-LNF, Frascati, Italy
ROGERS, J.	Daresbury Laboratory, Warrington, UK
SANMARTI, M.	CERN, Geneva, Switzerland
SAUGNAC, H.	IPN, Novosibirsk, Russia
SCHIAPPAPIETRA, A.	CERN, Geneva, Switzerland
SCHILD, T.	CEA Centre de Saclay, Paris, France
SCHROEDER, C.	GSI, Darmstadt, Germany
SILZER, R.M.	University of Saskatchewan, Saskatoon, Canada
STEFFEN, B.	DESY, Hamburg, Germany
STOUT, D.	ORNL, Oakridge, TN, USA
SUGITA, K.	Graduate Univ. f. Adv. Studies, Ibaraki, Japan
TODESCO, E.	CERN, Geneva, Switzerland
TOMASSINI, S.	INFN-LNF, Frascati, Italy
VENTURINI, W.	CERN, Geneva, Switzerland
VINCENT-VIRY, O.	CERN, Geneva, Switzerland
WANG, X.	BNL, Upton, NY, USA

WEI-HSIANG, W-H.	SRRC, Hsinchu, Taiwan
WELSCH, C.	Johann-Wolfgang-Goethe Univ. Frankfurt, Germany
WILDNER, E.	CERN, Geneva, Switzerland
ZANGRANDO, D.	Sincrotrone Trieste, Italy
ZERLAUTH, M.	CERN, Geneva, Switzerland