

DISSERTATION

Conceptual Design of Superconducting Magnets for the LHC Using Genetic Optimization Algorithms

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften unter der Leitung von

O. Univ.-Prof. Dipl.-Ing. Dr.techn. Adalbert Pechtl
Institut für Grundlagen und Theorie der Elektrotechnik, E351

Dr.-Ing. Stephan Russenschuck
CERN, LHC-ICP, Genf, Schweiz

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik

von

Dipl.-Ing. Suitbert Ramberger

Wien, im April 1999

TO MY PARENTS FOR THEIR
SUPPORT AND ENCOURAGEMENT

Acknowledgements

The work presented in this thesis has been carried out in the LHC-ICP group at CERN, in the framework of the Doctoral Student Programme at CERN.

I am most grateful to my supervisor at CERN, Dr. Stephan Russenschuck, who initiated this work, for his excellent support and the many fruitful discussions.

I wish to express my gratitude to my supervisor Professor Dr. Adalbert Prechtel, head of the Institute for Fundamentals and Theory of Electrical Engineering, Vienna University of Technology, for his support and encouragement from the inception of this project.

I would like to thank Professor Dr. Wolfgang Rucker, head of the Institute for Theory of Electrical Engineering, University of Stuttgart, for the great interest he has shown for the thesis.

I am indebted to Julio Lucas for the invaluable discussions, which led to many new ideas regarding the work and its representation. Furthermore, I appreciate the help of Gary Mathers in the many textual reviews, and Anne Doyle in a painstakingly thorough correction of my English.

In addition, I would like to thank the members of the ICP group for their help and pleasant working environment.

Finally, I would like to thank my family and friends for their help and encouragement during my stay at CERN.

Geneva, April 1999

Suibert Ramberger

Contents

1	Introduction	1
1.1	Outline of the Thesis	2
2	The Large Hadron Collider	5
2.1	Design Challenges	6
2.2	Layout of the LHC	7
2.3	The Main Dipoles	8
2.4	Superconducting Coils and Cables	11
2.5	Effects of Superconductivity	12
2.6	The Short Dipole Model Program	14
2.7	The Software Package ROXIE	14
3	Field Computation of Superconducting Magnets	17
3.1	The Multipole Expansion	17
3.2	Field Calculation	20
3.3	Magnetic Field of a Line Current	21
3.4	Symmetry Considerations	22
3.5	Pure Multipole Fields	23
3.6	Segment Approximation	24
3.7	Persistent Currents	25
3.8	Sensitivity	27
3.9	Iron Cross-Section	28
4	Mathematical Optimization	31
4.1	Optimization Formulation	31
4.2	Pareto Optimality	32
4.3	Objective Weighting	33
4.4	Utopia Solution	34
4.4.1	Circular Cylindrical Conductor	35
4.4.2	Sinusoidal Current Distribution	37
4.4.3	Intersecting Elliptical Conductors	39
4.5	Optimization Techniques	42

5	Genetic Algorithms	45
5.1	Introduction to Genetic Algorithms	45
5.2	Nomenclature in Evolutionary Computation	46
5.3	Royal Road Genetic Algorithm	47
5.4	Parameter Encoding and Quantization	48
5.5	Gray-coding	49
5.6	Genetic Operators	50
5.6.1	The Generation Operator	50
5.6.2	The Crossover Operator	50
5.6.3	The Mutation Operator	52
5.6.4	The Selection Operator	53
5.7	Niching Methods	56
5.7.1	Fitness Sharing	57
5.7.2	Crowding Method	57
5.8	The Implementation	58
5.8.1	The Algorithm	58
5.8.2	The Parameters	60
5.9	Methods of Description	63
5.9.1	The Schema Theorem	63
5.9.2	Dynamics and Convergence	65
5.9.3	No Free Lunch Theorems	66
6	Conceptual Coil Design	67
6.1	Conceptual Design Phase	67
6.2	Coil Design Phase	69
6.3	Design Variables	70
6.3.1	Constraints on Design Variables	72
6.4	Objectives for the Coil Optimization	73
6.5	Objective Function	74
6.6	Parameters of the Genetic Algorithm	75
6.7	Coil Optimization Results	76
6.8	The LHC main dipole coils	77
7	Material Distribution Problems	83
7.1	Introduction to Yoke Optimization	83
7.2	Design Variables	85
7.3	The FEM Modeling	86
7.4	Optimization Objectives	86
7.4.1	The Problem of Jagged Structures	88
7.5	Parameters of the Genetic Algorithm	89
7.6	Circular Grid Structures	89

7.7	Brick Structures	91
7.8	Abstraction of Feasible Structures	92
7.9	Comparison with the Yoke Design MBP2	97
8	Conclusions	99

Chapter 1

Introduction

The Large Hadron Collider (LHC) will be the next particle accelerator installed at the European Laboratory for Particle Physics, CERN, Switzerland. The machine will provide physicists with proton and lead-ion beams of high energy at an unprecedented rate of interaction, in a consequent step to continue the search for ever heavier particles. Approved by the CERN Council in 1994, the LHC shall start operation in 2005 in the 26.8 km tunnel of the current Large Electron Positron (LEP) accelerator.

Apart from the experiments, the LHC project involves the biggest program for research and development at CERN, with major areas comprising the field of cryogenics and magnet development. To reach the design energy of the LHC, high magnetic fields of 8.4 T have to be achieved to guide the particle beams along the tunnel. Consequently the magnet system must operate in the superconducting state. For the LHC the operating temperature will be below 2 K, at which helium is superfluid. 1232 main dipoles, 386 main quadrupoles and some twenty different types of corrector and insertion magnets will have to be built, adding up to more than 8000 superconducting magnets.

A major concern of the magnet design regards the field quality which is necessary to ensure the stability and quality of the particle beams supplied to the experiments. Conductor shapes, deformations due to cable manufacturing, collaring, and magnetic forces have to be modeled accurately. Only a tenth of a millimeter of deviation from the nominal design may deteriorate the field quality severely. However, mechanical precision is not the only consideration, also dynamic effects have to be taken into account. Field quality has to be maintained over the full magnet cycle, starting with the injection of particles into the ring, accelerating the beams for about twenty minutes, and maintaining the nominal field level for several hours of data acquisition in the collision experiments. Persistent eddy currents in the superconducting coils and the non-linearity of the iron yoke have to be considered in the design stage.

In order to reach the required accuracy in the electromagnetic field calculation, and to model the geometry of conductors and yokes precisely, the ROXIE program package was developed at CERN in collaboration with the Technical University Graz, Austria and the University of Stuttgart, Germany. The package not only comprises routines for the modeling of coils and yokes as cross-sections and in three dimensions, but enhances them by several optimization routines, CAD/CAM interfaces, and visualization tools in an approach towards

an integrated design of superconducting magnets. In order to support the engineer already in the initial phase of a magnet design project, a method is sought that is capable of providing several conceptual designs and ideas from which the best may be chosen with respect to manufacturability and overall performance.

The purpose of the thesis is to enlarge upon the conceptual design phase and to introduce niching genetic algorithms as an appropriate means to automate this design process. The key characteristic and salient property of genetic algorithms is their ability to globally optimize multimodal objective functions in a discrete or continuous design variable space. As such the importance of the proper choice of design variables and objective functions must not be underestimated and is highlighted in the respective chapters. In the current implementation, the introduction of a niching technique supplies the designer with a number of solutions. The method is complemented by the rational choice of the engineer obeying further non-formalized design criteria. Design examples including their evaluation underline the success of the approach towards a more creative design tool.

1.1 Outline of the Thesis

The design of the Large Hadron Collider with its main bending dipoles in the arcs of the accelerator tunnel constitutes a major undertaking. With the main magnets being a key element in the design, research efforts concentrate on their optimization. Stringent requirements in field strength and quality have to be met, and effects of superconductivity have to be taken into account. In order to match the required precision in the calculation, multipole field distributions are used as a design criterion. The estimation of ideal current distribution with pure multipole fields give an idea of the attainable field strength.

The low interference of coil and yoke design lead to the standard approach of separating the two aspects. Both parts were therefore investigated individually, and a different approach was taken for each of them. The main parameters of the coil design are the number of conductors, and the positioning and inclination angles of each coil block. The optimization problem therefore has to treat discrete and continuous non-linear variables concurrently. Since a new conceptual design is sought in the complete design space, global optimization algorithms have to be employed. Genetic algorithms are found appropriate under these circumstances. Niching, the concurrent maintenance of multiple potential solutions in the optimization, introduces a major improvement to the procedure. The correct formulation of the problem, including persistent current effects due to superconductivity, is shown to be essential for the application. Objectives and weights for the objective function, are selected adequately. This method is applied to the re-evaluation of the main dipole coils and leads to new coil configurations for the LHC. A major improvement is found in the new designs compared to an already existing design. Test magnets built confirm the predicted performance gain.

The yoke design was approached separately from the coil design. The yoke optimization is treated as a material distribution problem, rather than a shape optimization. A finite element method is used for the evaluation of the potential designs. A number of regions, consisting of finite elements, is defined to be mutable from magnetic iron to non-magnetic material. The

fact that this implementation of genetic algorithms is working on bit-strings, makes their application straightforward. Jagged structures, which appear in first optimization runs, are overcome by the evaluation of the design at several current excitation levels. In order to gain more creativity in the procedure, a rectangular mesh is set up. Local saturation effects are treated by subdividing the grid into a brickwall-like structure with overlapping elementary regions. The aim of the optimization is the deduction of ideas for the yoke design. These features are assembled in feasible structures and the influence of each of them is evaluated. Two alternative designs for the existing yoke structure are proposed, confirming the relevance of this approach.

Chapter 2

The Large Hadron Collider

Luminosity, or rate of interaction per unit cross-section, a main design criterion for accelerators, will reach the next order of magnitude with the LHC as compared to the most powerful current accelerators, LEP at CERN, Tevatron at Fermilab, or HERA at DESY. This will allow experiments to further penetrate into the elementary structure of matter. Key questions are sought to be answered. The origin of mass is thought to be connected to the Higgs mechanism used in the standard model of physics, which in turn implies the existence of a heavy particle called the Higgs Boson. The high energies of the new accelerator should be sufficient to make this particle detectable.

The grounds for a grand unified theory (GUT) of the elementary forces, electromagnetic force, weak force combined with the strong force will be investigated. If supersymmetry, abbreviated SUSY, one of the concepts of such theories is right, supersymmetric particles should be detected at the LHC. Consequences of this theory are far reaching as it will predict the life-time of protons, compound particles where no decay was ever measured. Insight into these questions will shed light on the state of matter prevailing shortly after the widely assumed birth of the universe in a big-bang.

Though the collider will start as a proton accelerator, it can later be upgraded to provide for heavier lead-ion experiments. Another state in the development of the universe even closer to the big-bang, the so-called quark-gluon plasma, will be created in the collision experiments, and can then be studied in more detail. The physics of the biggest structure which is our universe, is thus directly connected to the physics of the smallest particles, which is to be scrutinized by the most powerful accelerator ever built.

In the following the structure of the LHC is discussed focusing on the magnet design. The peculiar requirements of superconducting accelerators are introduced, starting at the conditions of physics that lead to the current design. The structure of the dipole magnets is elaborated upon to provide grounds for the subsequent treatment of the subject. The chapter is completed by a brief outline of the software package ROXIE, and its importance as a result of the special needs of accelerator magnet design.

2.1 Design Challenges

The Large Electron Positron (LEP) collider, with a circumference of 26.8 km in its last stage of upgrade at CERN, currently provides two particle beams of about 94.5 GeV of opposite charged particles: electrons and their antiparticles – positrons. Each beam counter-rotates in a single elliptically shaped beam tube. Due to its bending radius the beams lose synchrotron radiation of about 16 MW which is about 3.5% of the power needed to accelerate the beams. Since the amount of radiation is proportional to E^4 , with E the beam energy, it is impossible to upgrade the LEP to higher energies. Heavier protons were thus chosen as collision particles to achieve the design goal for the LHC of a beam energy of about 7 TeV for a single beam or 14 TeV center of mass energy in a collision. This decision will allow the reuse of the current accelerator tunnels and the adaption of the smaller proton accelerators PSB, PS and SPS as an injector chain (fig. 2.1).

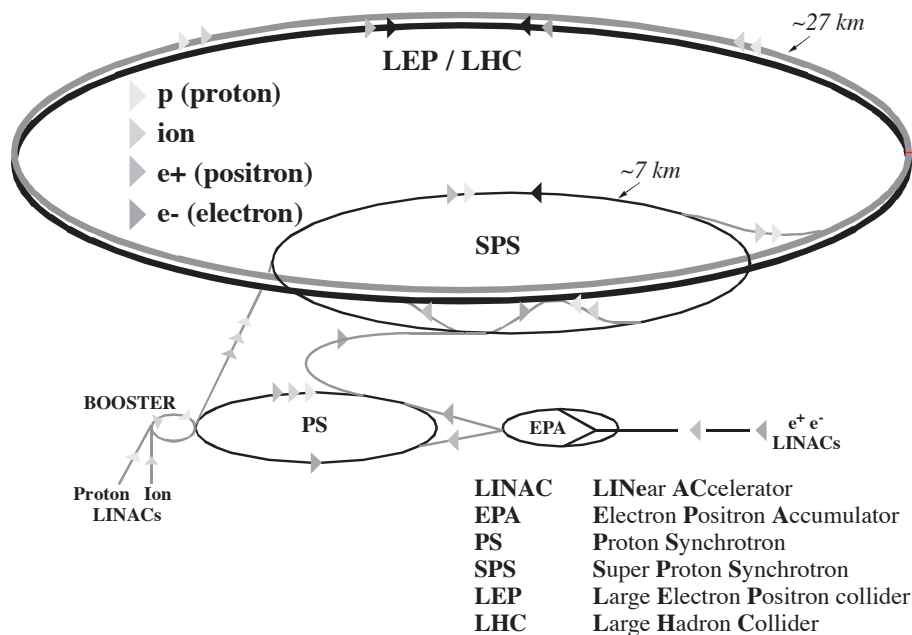


Figure 2.1: Schematic layout of accelerators at CERN

The seemingly obvious concept of colliding particles with their associated antimatter particles as in the case of LEP and for protons at SPS, was discarded for the LHC since the necessary number of anti-protons can only be reached with great difficulty. Consequently, two separate beam tubes with counter-oriented magnetic fields are necessary to accelerate the two beams in opposite directions.

The higher energy E of the particles in the LHC compared to its predecessors reduces their interaction cross-section proportional to $1/E^2$. This relationship is due to a decrease with $1/E$ of the De Broglie wavelength associated with a particle. The LHC subsequently requires a 100 times higher rate of interactions per unit cross-section or luminosity and should reach unprecedented $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ after a certain time of operation. This luminosity

L can be calculated by the approximate formula

$$L = \frac{N^2 f n_b}{4\pi \sigma_x \sigma_y} \quad (2.1)$$

assuming a beam consisting of $n_b = 2835$ bunches of $N = 10^{11}$ particles, with a beam size of $\sigma_x = \sigma_y = 16 \mu\text{m}$ at collision, and a frequency of revolution $f = 11240 \text{ Hz}$ as given by the circumference of the accelerator. These 2835 bunches are distributed on 3564 equally spaced positions along the ring according to an injection scheme. The remaining empty places are needed for the rise time of injection magnets guiding the beam into the next accelerator stage. As a result the minimum space between bunches is 25 ns or 7.5 m.

The most important luminosity limitation stems from the beam-beam interaction when two bunches of particles cross in the interaction region of the detector. Only a tiny fraction of the particles will collide head-on to produce the desired events. As a byproduct, the motion of the other non-collided particles will be disturbed. In addition, long-range interactions occur on either side of the experimental region, where the two beams run close to each other. The effects add up and result in an increased tune spread, a dispersion of the transversal oscillation frequency, which must not exceed tight limits.

To safeguard against tune spread, high requirements on the field quality of the superconducting magnets and their alignment are imposed, since the beam has to be kept stable for up to 10 hours corresponding to more than 10^8 revolutions.

2.2 Layout of the LHC

Based on the geometry of the LEP machine as shown in figure 2.2, the LHC is organized in 8 octants. It consists of a succession of arcs and straight sections. 4 interaction regions in straight sections of the octants are reserved for the collision experiments ATLAS, CMS, ALICE, and LHC-B. The remaining straight sections each 528 m in length, are used for the beam support, as there are acceleration by radio-frequency cavities, beam cleaning by collimation systems for a high quality stable beam, and dumping at the end of an experimental cycle. The radio-frequency insertion in octant 4 may also be adapted as an additional experimental area at a later stage.

The accelerator is filled by bunch trains from the injection system consisting of the linear accelerator, the booster (PSB), the proton synchrotron (PS), and the super proton synchrotron (SPS). The operation is synchronized and allows for filling of the LHC rings in about 7 minutes. Once both rings are filled the beams are accelerated to nominal collision energy in about 20 minutes providing experimental times of several hours. Further time will be necessary for an adjustment of the beam system so that an estimate for a good performance should be a cycle-time of about 9 hours with a physics data-taking time of about 7 hours.

While the straight sections serve special purposes, the arcs have to provide the necessary bending forces by a homogeneous magnetic field on the beam. All the arcs employ the same structure of 23 arc cells where each cell consists of two identical half-cells.

Each half-cell is made up of three 14.3m long twin-aperture dipole bending magnets (MB)

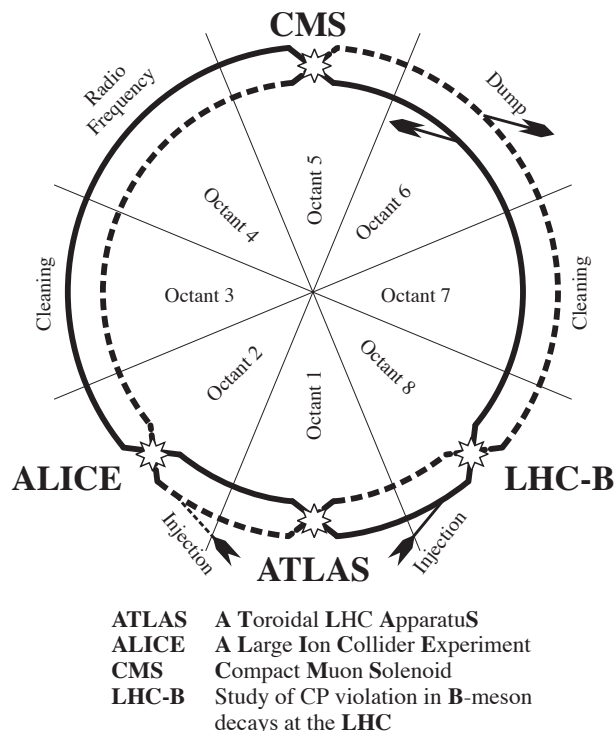


Figure 2.2: Schematic layout of the LHC showing collision experiments and insertions for the beam support

and a short straight section housing a 3.1 m long twin-aperture quadrupole magnet (MQ), a combined sextupole/dipole corrector, either an octupole (MO) or a skew quadrupole or a trim quadrupole depending on the position, and a beam position monitor (BPM). Small sextupole and decapole correctors are attached to the ends of the dipole magnets (fig. 2.3).

The magnetic field gradient in the quadrupoles leads to focusing of the beam in one transversal direction while defocusing in the other. Alternating focusing and defocusing quadrupoles along the lattice leads to a net focusing effect, as particles oscillating within this envelope will always tend to be further off axis in focusing quadrupoles than in defocusing quadrupoles and will sense the higher magnetic field there. Such a setup keeping the particles on track constitutes a so-called FODO cell. Further corrector magnets compensate for higher order errors in the dipoles and quadrupoles.

The number of necessary main dipoles adds up to 1232. In addition 386 main quadrupoles have to be built. Numbers of small correctors in the same order of magnitude will be employed. Additionally some strong insertion and extraction magnets are needed. Altogether more than 8000 magnets have to be manufactured.

2.3 The Main Dipoles

Letting particles of an energy of 7 TeV, accelerated to near light speed, travel on a bending radius in the arcs of about 2.7 km, requires a homogeneous magnetic field B of almost 8.4 T.

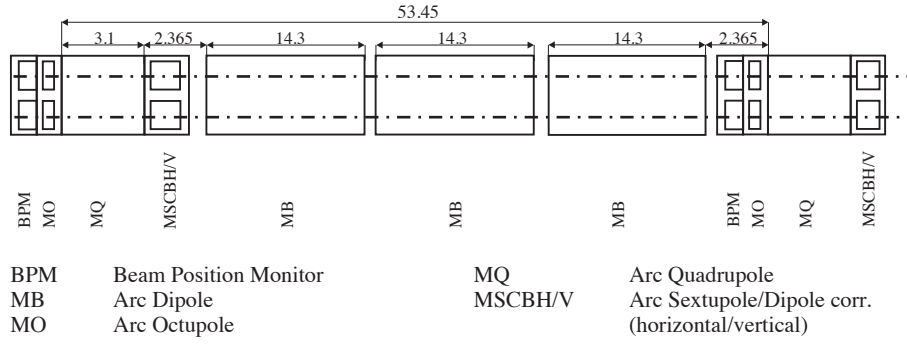


Figure 2.3: Layout of an arc half-cell mainly consisting of 3 bending dipole magnets and a focusing or defocusing quadrupole

This value can be calculated by the formula of motion

$$B = \frac{p}{e\rho} \quad (2.2)$$

for a certain particle momentum p and a bending radius ρ , where e is the electric charge of a proton.

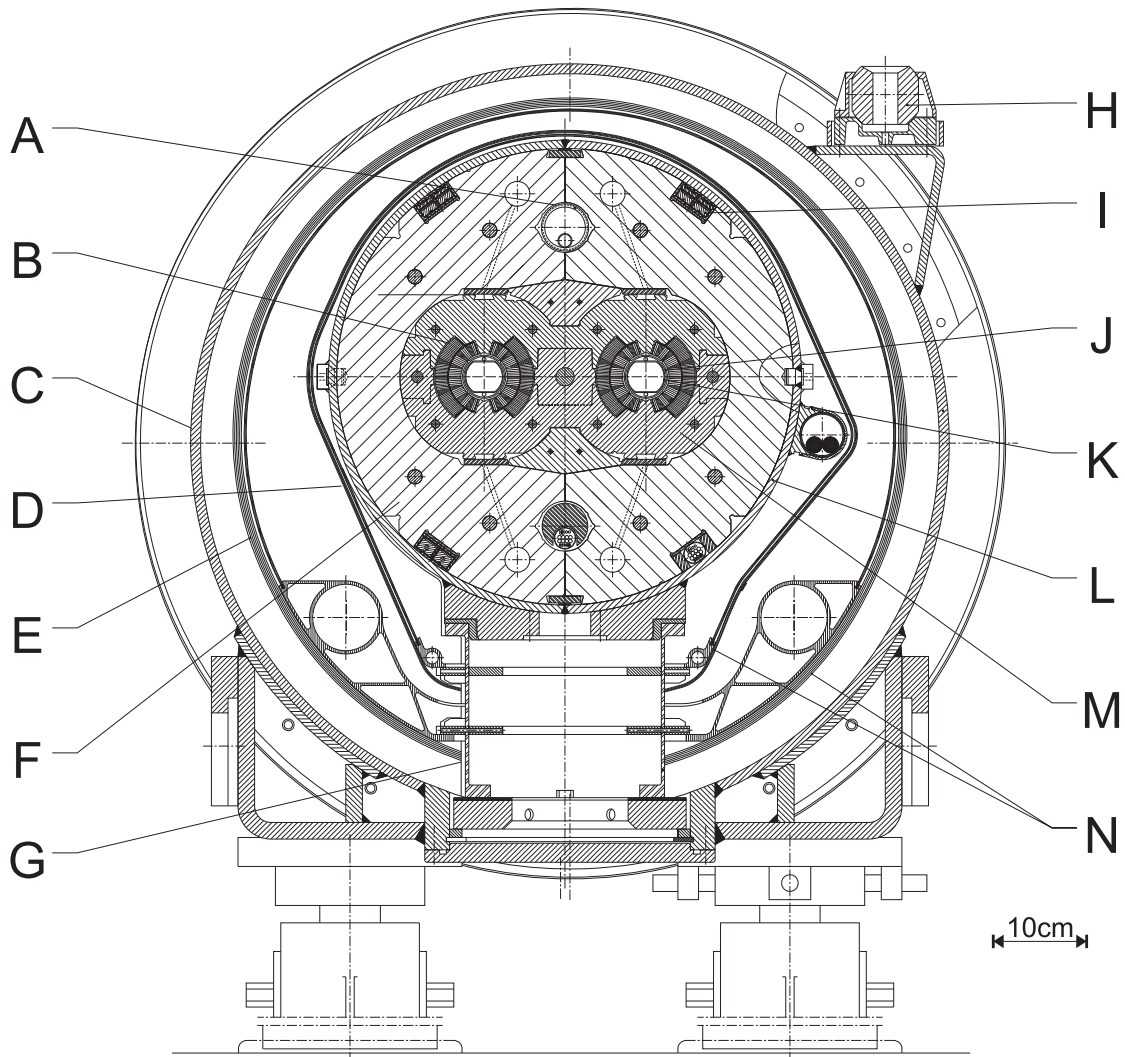
Such high fields can in turn only be produced by electromagnets with high current densities which cannot be achieved by normal conducting wires. Superconducting technology must be employed, which thereby also reduces the size of the magnets and the overall necessary energy. The magnet structure must therefore provide for the cooling and insulation needed.

As shown in figure 2.4 the dipole structure consists of three main parts:

- The cold mass comprising the superconducting coils, surrounded by non-magnetic stainless steel collars, held together by an iron yoke and compressed by a shrinking cylinder of 57 cm diameter.
- The thermal insulation shields consisting of a multi-layer superinsulation against heat inleaks by radiation. The cooling system with stages at 60 K and 4.6 K is contained within.
- The vacuum vessel of 1.05 m diameter to provide the insulation vacuum of 10^{-6} mbar to reduce heat inleaks by convection.

Of main importance for electromagnetic calculations, is the cold mass where the coils and most of the magnetic material is located. The two beam tubes separated by 194 mm are contained in a common yoke structure also known as a two-in-one design, making the structure more compact, and fitting it into the space available in the existing tunnel. Nevertheless the cross-talk between the two coil systems has to be taken into account.

The superconducting coils are clamped together by the force containment structure of collars, yoke and shrinking cylinder, to ensure the proper control of the Lorentz forces. It



A	Heat Exchanger Pipe	H	Alignment Target
B	Superconducting Coils	I	SC Bus-Bars
C	Vacuum Vessel	J	Beam Pipe
D	Radiation Screen	K	Beam Screen
E	Thermal Shield	L	Shrinking Cylinder
F	Iron Yoke	M	Non-magnetic Collars
G	Support Post	N	Superinsulation

Figure 2.4: Transverse cross-section of the main dipole including cryogenics and vacuum tank

produces the necessary azimuthal pre-stress in the coils to prevent the conductors from movements under the electromagnetic forces in the superconducting state.

The laminated iron yoke helps to increase the magnetic field in the tube and to reduce the fringing fields by guiding the flux. Cut into the yoke are the bus bars that transport the supplied current in superconducting cables from magnet to magnet. One of the center holes

is used to provide coolant by a corrugated heat exchanger tube. The shrinking cylinder not only performs its mechanical function but serves as a helium tank at the same time, keeping the cold mass immersed in superfluid helium.

During assembly the horizontally split collar laminations are pressed around the pre-manufactured coils. Locking rods align the collars and maintain the pre-compression of the coils. The yoke laminations are pushed on the collars and the cold mass is bent to a 2.7 km radius of curvature. The two halves of the pre-bent shrinking cylinder are welded together around the iron yoke exerting further pre-stress on the structure.

The two beam pipes in the center consist of a cold bore of stainless steel which serves as an inner confinement of the helium tank, and a beam screen which intercepts the synchrotron radiation power emitted by the beam. This beam screen is kept at a temperature between 5 K and 20 K by little cooling tubes. Little holes in the beam screen make sure that the vacuum can be maintained during operation.

The size of the mechanical aperture of the beam screen of 44 mm diameter is chosen to be just big enough to get a beam through, taking mechanical tolerances and deviations of the beam orbit into account. The beam diameter is about 14 mm at injection measured by its dynamic aperture, which is the maximum distance of stable particles from the beam center. Particles outside the dynamic aperture, lost in the cold bore, must not be too many since the dissipated energy could provoke a quench, a sudden irreversible breakdown of superconductivity, by taking conductors locally over the critical temperature.

2.4 Superconducting Coils and Cables

The magnet fields are produced by an upper and a lower coil with two layers of superconducting cable each. The inner layer is made of 4 blocks of cable separated by copper wedges, whereas the outer layer employs only two blocks. The layers and coils are connected in series. In order to achieve a higher current density in the outer layer where the magnetic field is lower, different cable types are used.

In the production of the coils the cable is wound along a mandrel and guided back around spacers in the end, keeping the beam pipe clear. These end-spacers are designed to confine the conductors to a consistent shape, close to a constant perimeter profile, outer and inner edge of the cable having the same length for each individual turn. Each coil is polymerized in a mold, and a perforated glass-epoxy spacer is inserted between the layers allowing coolant to diffuse. The two coils are joined together in the top of the layers by a splice.

For magnetic fields of up to 8.4 T only low temperature superconductors as available from industrial production, like niobium-titanium and niobium-tin alloys can be considered. Since niobium-tin is brittle and more expensive, preference was given to niobium-titanium, at the cost of a more complex cooling technology for an operating temperature of 1.9 K instead of standard 4.2 K absolute temperature (fig. 2.5) [1].

The main challenge is the comparatively small specific heat of the super-conducting cable so that a quench may occur due to microscopic movements of the cable under the enormous electro-magnetic forces. These movements create heat through friction, locally taking the material above its critical temperature and provoking a quench. This disadvantage has to

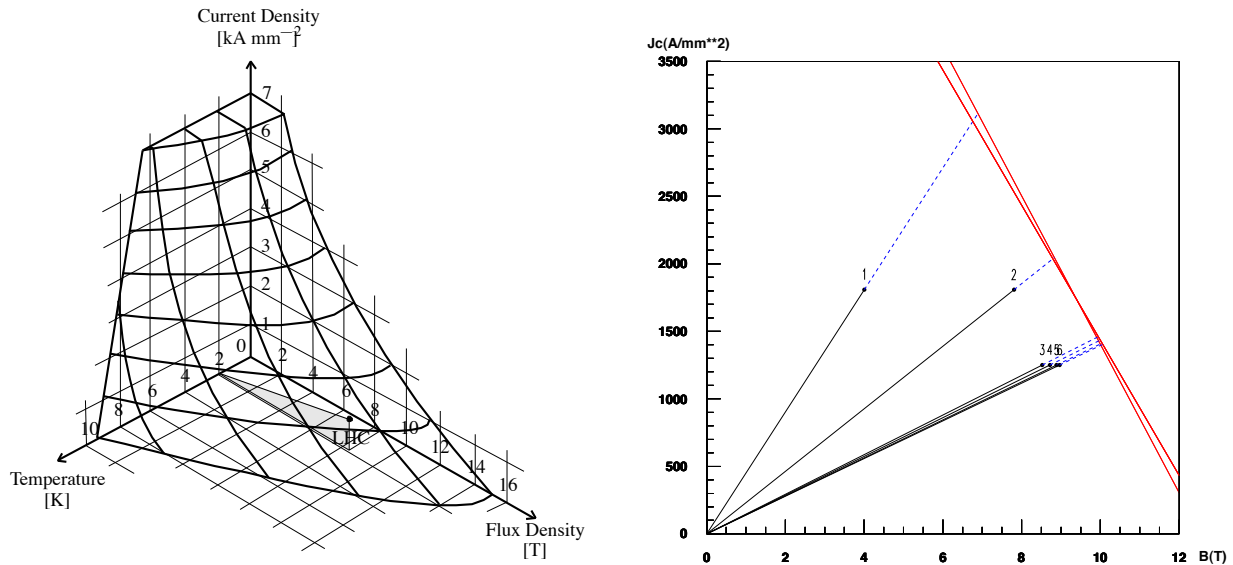


Figure 2.5: Left: Critical phase transition surface of niobium-titanium. Right: Critical current - field diagram at 1.9 K for the most critical strand of each block used in the field computation program ROXIE.

be overcome by the special properties of the coolant, superfluid helium. Its comparatively large specific heat together with its high thermal conductivity can stabilize the conductor. It is therefore important to get the helium to permeate the coil and to transport heat away.

To reach the theoretical limits of niobium-titanium, and to withstand the forces in the magnet and possible quenches, the cable has to be designed for the desired properties. The superconducting material is inserted into a copper matrix for thermal stabilization and drawn into a wire of about 1 mm. For the inner dipole cable 28 and for the outer 36 conductors are twisted together along their length and pressed into a so-called key-stoned Rutherford cable (fig. 2.6).

The cable is wrapped in 3 overlapping layers of a porous polyamide tape for insulation. The porosity allows the superfluid helium to penetrate into the conductor, still the insulation must safely withstand a turn-to-turn of about 50 V at a quench. Further polyamide layers around the coils guarantee insulation against other coils and the collars. A coil protection sheet between collars and coils safeguards against damage from the collar lamination. As a further means of protection, quench heaters are installed around the outer coils. In case of a quench, the quench must be detected, the heaters must be switched on to protect the quenching spot by making the whole coil resistive, and the stored energy must be extracted.

2.5 Effects of Superconductivity

With niobium-titanium and niobium-tin being type II superconductors, high critical field and critical current densities can be achieved. In type I superconductors the induced magnetic field at the surface shields the bulk material from any currents. Due to the higher

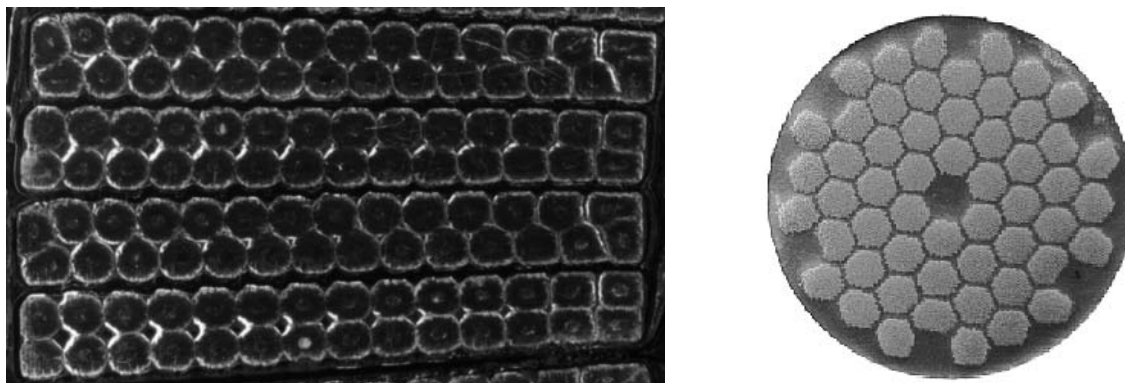


Figure 2.6: Cross-section of a block of key-stoned Rutherford cables and one of its 28 strands as used in the inner layer of the main LHC dipoles

density of charge carriers in the surface region, type II superconductors reach a new lower thermodynamic equilibrium allowing the magnetic flux to penetrate the material. The flux inside the material is not distributed uniformly, but rather concentrated in flux tubes which form a triangular pattern separated by about 20 nm.

These flux tubes are not fixed per se but can move due to the Lorentz force of applied currents and fields. Since this movement of flux tubes is resistive, it generates heat. Because a temperature rise would deteriorate the critical properties of the material, this movement must be prohibited. The remedy lies in the generation of flux pinning centres, which are separated by about the distance of the flux tubes, to fix their location. The pinning centres are introduced by cold working and thermal treatment of the material, causing non-uniformities in the microstructure. The energy of the flux tubes become position dependent, giving rise to minimum energy configurations. Superconducting materials of this type are called hard superconductors. Unfortunately there is a trade-off, in that these conductors exhibit a strong magnetic hysteresis, creating undesired persistent-current multipoles.

Superconductivity has to be treated by thermodynamic calculations for a possible cable-design, cooling, and quench protection. Magnet design, however, from the field optimization standpoint mainly considers the shape of the conductors, their critical field and current properties, and persistent currents only.

The main limitation is imposed by the critical current density, which must not surpass a certain level defined by the local magnetic field density. In figure 2.5 the relationship between critical current density and magnetic field is shown. The maximally reachable field is best estimated from the load-line which is only dependent on the design. In the region of interest the critical properties may be approximated by a line. The percentage left on the load line between operating point and critical point is termed quench-margin.

A second consideration regards persistent currents which are induced by the changing fields in the magnets. Their effects are non-linear and dependent on the history of previous current cycles. Their influence on the field quality has to be compensated [2]. In the operation of the accelerator, superconducting magnets have to be pre-cycled to reach a reproducible operating point.

2.6 The Short Dipole Model Program

In order to evaluate several different designs and to optimize the assembly and collaring parameters of the coils with a short turnaround, a short dipole model program was started in 1995. So far 3 double aperture and 20 single aperture 1 m long models have been made. Several of these units were reassembled into new variants, totalling thus 30 short models tested in a cold state at a rate of about one per month [3].

3 prototypes of 14.3 m length, built by European industry, allow the comparison with the short dipole model program. Valuable experience with this program is gained and conclusions for long dipoles can be drawn. Performance improvements coming from refinements in design, components, and assembly are studied and statistics based on a small-scale production are accumulated.

The short dipole magnet tests are performed in a vertical cryostat with high temperature stability for the test environment. A rotating measurement coil device is inserted into the beam tube. Three 20 cm long pick-up coils are employed to monitor the longitudinal variation of the field quality in the shaft. Two additional outer coils are used to measure the fields in the magnet ends.

The test setup allows the control of excitation current and temperature with high accuracy, and magnet cycles as used in accelerators can be simulated. 50 voltage taps per model help detect and locate a quench, and analyze the quench propagation. Numerous strain gauges monitor the coil stress distribution in order to gain insight into the behaviour of the mechanical structure.

Latest models reached initial quenching fields above 9 T and came close to the conductor limit exceeding 10 T after several more quenches. This behaviour of increasing the quench level with a rising number of quenches is termed training, and is attributed to slight movements of the structure further stabilizing with each quench. A comparison of optimization results to measurements at the short dipole model test facility is given in chapter 6.

2.7 The Software Package ROXIE

For the support of accelerator magnet calculations necessary for the LHC, the software package ROXIE (Routine for the Optimization of magnet X-sections, Inverse field calculation and coil End design) has been developed at CERN [5]. The special requirements in accelerator design regarding precision required the development of a custom-tailored application to support the developers' need. The importance of exact coil modeling and calculations can be seen in figure 2.7 where the movement of one block of only 0.03 mm causes a drastic reduction in the field quality. Only the darkest region is within the tolerances of the beam-parameters.

The ROXIE package includes a coil modeler for 2d cross-sections and 3d coil-ends. Approximating the influence of the yoke by a mirroring technique, fast coil computations can be done including peak field calculations and persistent current estimation. A yoke modeler allows for parametric discretization of the iron material for finite element calculations. Two finite element packages were included; FEM2D developed at the IGTE, Technical University of Graz, Austria and BEM-FEM developed at the ITE, University of Stuttgart,

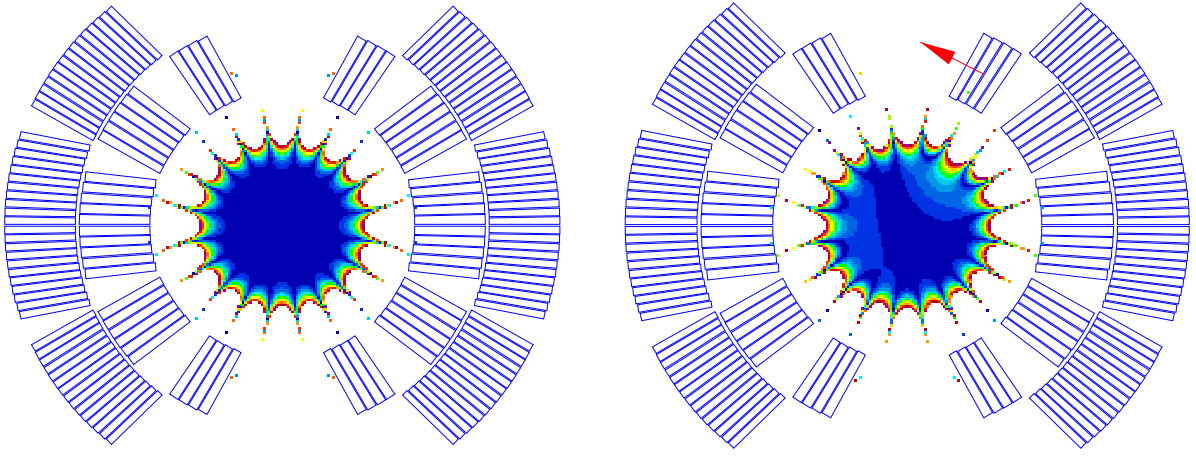


Figure 2.7: Disturbance of the field quality due to a azimuthal block movement of 0.03 mm of the inner uppermost block, shown at the coil geometry as described in the “yellow book” [4]. The relative field error outside the darkest region is more than 10^{-5}

Germany. Both avoid the discretization of the coil, thereby vastly improving the accuracy of the results. The method of coupled boundary/finite elements as provided by the BEM-FEM program adds another degree of freedom as it supports detached regions of magnetic material.

Both FEM packages were chosen as their formalism provides for the high precision necessary in accelerator magnet design. This is achieved by the separation of coil and yoke modeling. The excitation of the iron yoke structure is found from the accurate coil field calculation by ROXIE. The coils do not have to be represented by finite elements. The current implementation of FEM2D requires a quadrilateral finite element grid. The input files are generated by a preprocessor which is enhanced by a macro-language. Several macros were defined that allow the meshing of regular regions. In addition, macros exist for the definition of holes and half-holes frequently encountered in standard yoke structures (fig. 2.8) [5].

Optimization of coil and user specified iron parameters is handled by a suite of optimization algorithms. The tools support local optimization of a known structure, sensitivity calculation and the tracking of manufacturing errors by inverse calculations. Support of the initial design phase by global optimization with genetic algorithms is added by this thesis. Graphical output and several interfaces including CNC machining output completes the package as a computer aided manufacturing (CAM) tool [6].

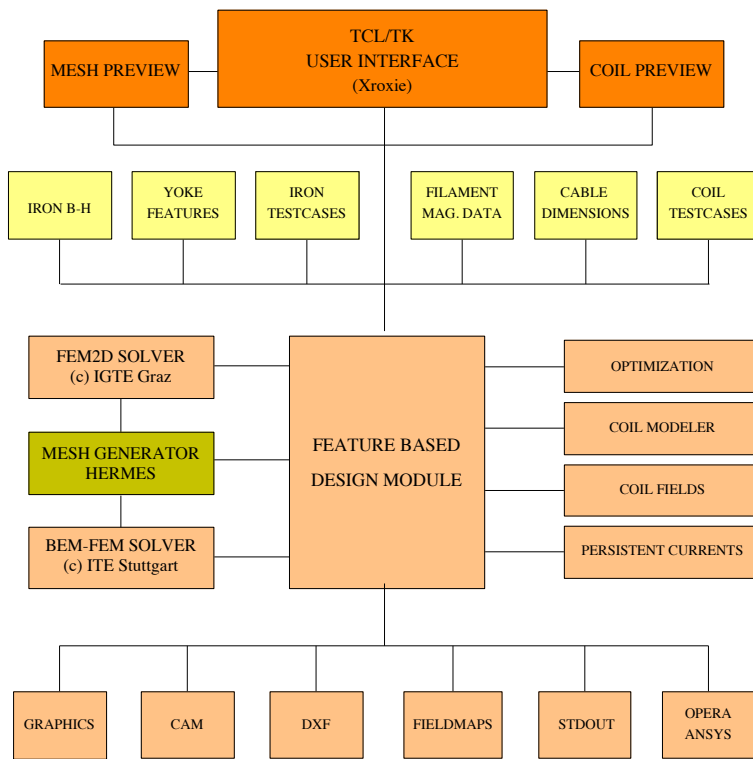


Figure 2.8: Program structure of ROXIE

Chapter 3

Field Computation of Superconducting Magnets

The special requirements of accelerator physics treating the beam properties call for a cylindrical coordinate system. Naturally an orthogonal field expansion known as multipole expansion appears to be useful, solving Maxwell's equations. Applying this multipole expansion in complex coordinates to the fields of a conductor distribution results in relatively simple expressions to be used in the following. Employing geometric symmetry leads to the definition of multipole magnets used for the correction of field errors. Two current distributions known to be ideal with respect to geometrical multipole errors are derived later in the chapter. These current distributions with pure multipole fields serve as models for technically feasible designs. Such designs can be approximated by arc segments. By this means field calculation is carried out analytically. Later in this chapter, the calculation of persistent currents is shown and the relationship to sensitivity evaluation is given. A short introduction into finite element calculation as used in this thesis and its advantages concludes this chapter on field computation in superconducting accelerator magnets.

3.1 The Multipole Expansion

The multipole expansion deduced below is found as part of the fundamental field solution in a current free and non-magnetic region. The orthogonality of the expansion also allows separate physical effects of which the most important are the bending of the beam by the dipole component and the focusing and defocusing by the quadrupole.

The stationary magnetic field in non-magnetic ($\mu = \mu_0$) and current free ($J = 0$) regions is defined by the special form of the Maxwell equations

$$\begin{aligned}\nabla \times \vec{B} &= 0 \\ \nabla \cdot \vec{B} &= 0\end{aligned}\tag{3.1}$$

which are Cauchy-Riemann differential equations. For two-dimensional problems in rectan-

gular coordinates they are reduced to

$$\begin{aligned}\frac{\partial B_x}{\partial y} &= \frac{\partial B_y}{\partial x} \\ \frac{\partial B_x}{\partial x} &= -\frac{\partial B_y}{\partial y}\end{aligned}\tag{3.2}$$

The system (3.2) is equivalent to the Laplace equation of a potential A_z as defined in coherence to the z -component of the vector potential in the three dimensional case

$$\Delta A_z = 0\tag{3.3}$$

The magnetic fields are found as $B_x = \frac{\partial A_z}{\partial y}$ and $B_y = -\frac{\partial A_z}{\partial x}$. Since the field quality is defined in a circular region around the beam center the equation is solved in polar coordinates

$$\frac{\partial^2 A_z}{\partial r^2} + \frac{1}{r} \frac{\partial A_z}{\partial r} + \frac{1}{r^2} \frac{\partial^2 A_z}{\partial \varphi^2} = 0\tag{3.4}$$

Using a separation ansatz, constraining the field to finite values in $r = 0$, and disregarding a constant contribution leads to

$$A_z(r, \varphi) = \sum_{n=1}^{\infty} r^n (G_n \cos(n\varphi) + H_n \sin(n\varphi))\tag{3.5}$$

The radial field component of the magnetic field is then found by

$$B_r(r, \varphi) = \frac{1}{r} \frac{\partial}{\partial \varphi} A_z(r, \varphi)\tag{3.6}$$

yielding

$$B_r(r, \varphi) = \sum_{n=1}^{\infty} r^{n-1} (B_n \sin(n\varphi) + A_n \cos(n\varphi))\tag{3.7}$$

with $B_n = -nG_n$ and $A_n = nH_n$. The azimuthal component is found as

$$B_\varphi(r, \varphi) = -\frac{\partial}{\partial r} A_z(r, \varphi) = \sum_{n=1}^{\infty} r^{n-1} (B_n \cos(n\varphi) - A_n \sin(n\varphi))\tag{3.8}$$

Setting the radius to a reference $r = r_{\text{ref}}$ where the field is evaluated, the fundamental solution (3.7) appears as a Fourier series and the multipole coefficients can be calculated as Fourier integrals. The implementation in ROXIE evaluates the magnetic field in a number of points on the reference radius and utilizes a discrete Fourier transform [5].

In the following the equivalent complex notation is derived which results in simpler formulas. The formalism, however, is only valid for two-dimensional cases. Since the Cauchy-Riemann differential equations (3.2) are fulfilled for the components of the two dimensional magnetic field, the notation $B = B_y + jB_x$ can be introduced.

Rewriting (3.5) yields

$$A_z = \operatorname{Re} \left\{ \sum_{n=1}^{\infty} (G_n - jH_n) r^n e^{jn\varphi} \right\} \quad (3.9)$$

By analytic continuation, the complex potential can be calculated as

$$\Phi = A_z + jV = - \sum_{n=1}^{\infty} \frac{1}{n} C_n z^n \quad (3.10)$$

where V is the scalar magnetic potential, the complex multipole components $C_n = B_n + jA_n$, and the complex coordinate $z = x + jy = re^{j\varphi}$. This result is not surprising as it is the general representation of an analytic function in a region around $z = 0$ conforming to the requirement of the Cauchy-Riemann differential equations (3.2).

As known from function theory the field F is found as

$$F(z) = B_y + jB_x = -\frac{d\Phi}{dz} = \sum_{n=1}^{\infty} C_n z^{n-1} \quad (3.11)$$

This is a Taylor series valid around $z = 0$. It is in fact the standard definition of the multipole coefficients as used at CERN [7]. American laboratories usually use

$$B_y + jB_x = \sum_{n=0}^{\infty} C_n^{(\text{USA})} z^n \quad (3.12)$$

so that $C_n = C_{n-1}^{(\text{USA})}$. This way it looks similar to a z-transform and the inverse z-transform can be directly applied to retrieve the multipole coefficients.

Most times the coefficients are given with respect to a reference radius r_{ref} and (3.11) is recast as

$$B_y + jB_x = \sum_{n=1}^{\infty} c_n \left(\frac{z}{r_{\text{ref}}} \right)^{n-1} = \sum_{n=1}^{\infty} (b_n + ja_n) \left(\frac{z}{r_{\text{ref}}} \right)^{n-1} \quad (3.13)$$

The multipole components c_n are therefore found as $c_n = C_n r_{\text{ref}}^{n-1}$. The coefficients b_n are then called the normal components and a_n the skew components, and are usually given with respect to the main field component B_M in units of 10^{-4} relative to a certain radius (here $r_{\text{ref}} = 10 \text{ mm}$ if not stated otherwise).

For the magnetic field outside a current distribution a Laurent expansion around $z = \infty$ can be employed

$$F(z) = \sum_{n=0}^{\infty} C_{-n} z^{-n-1} \quad (3.14)$$

which uses the missing coefficients of the expansion (3.11). The general expansion with current distributions inside and outside the field region is therefore found by a complete

Laurent expansion using negative and positive exponents. The Fourier integral for such expansions reads

$$C_{n+1} = \frac{r_{\text{ref}}^{-n}}{2\pi} \int_{-\pi}^{\pi} F(r_{\text{ref}}^n e^{jn\varphi}) e^{-jn\varphi} d\varphi \quad (3.15)$$

for a field distribution F on a reference radius r_{ref} .

3.2 Field Calculation

The multipole fields as discussed in the last section are excited by a current distribution which is subject to optimization. For 3-dimensional problems the field distribution is found from the current distribution by Biot-Savart's law

$$\vec{B} = \frac{\mu_0}{4\pi} \int_V \frac{\vec{J} \times \vec{r}}{r^3} dV \quad (3.16)$$

where \vec{J} is the current density, \vec{r} is the vector from the point of integration of the current to the point of field measurement and r is its modulo.

For two dimensional problems with currents perpendicular to the calculation plane the integrals are reduced to

$$\vec{B} = \frac{\mu_0}{2\pi} \int_A \frac{J \vec{n}}{r} dA \quad (3.17)$$

where \vec{n} is a unit vector perpendicular to \vec{r} (right-hand associated). Rephrasing this formula for the complex field $F(z)$ as introduced in (3.11) yields

$$F(z) = B_y + jB_x = \frac{\mu_0}{2\pi} \int_A \frac{J(z_c)}{z - z_c} dA_c \quad (3.18)$$

where z is the coordinate of the field point $F(z)$ and z_c the coordinate of the differential current density element $J(z_c)$.

$$z = re^{j\varphi} \quad \text{and} \quad z_c = r_c e^{j\varphi_c} \quad (3.19)$$

Using the Taylor expansion for $|z| < |z_c|$, (the current distribution is outside a certain circle, whereas the field point is inside)

$$\frac{1}{z - z_c} = - \sum_{n=1}^{\infty} \frac{z^{n-1}}{z_c^n} \quad (3.20)$$

The multipole components are found from (3.18) by (3.11) as

$$C_n = -\frac{\mu_0}{2\pi} \int_A \frac{J(z_c)}{z_c^n} dA_c \quad (3.21)$$

For the case where the field outside of the current distribution is sought ($|z| > |z_c|$), (3.20) has to be rewritten as

$$\frac{1}{z - z_c} = \sum_{n=0}^{\infty} \frac{z_c^n}{z^{n+1}} \quad (3.22)$$

resulting in

$$C_n = \frac{\mu_0}{2\pi} \int_A J(z_c) z_c^n dA_c \quad (3.23)$$

3.3 Magnetic Field of a Line Current

Superconducting cables as described in the last chapter consist of a number of strands. Each of those strands is small compared to the conductor size and can be therefore simulated by an ideal line current. Integrals over current distributions consequently resolve into sums of line currents

$$J(z) = \sum_k I_k \delta(z_k) = \sum_k I_k \delta(x_k) \delta(y_k) \quad (3.24)$$

and the integral in (3.18) yields

$$F(z) = B_y + jB_x = \frac{\mu_0}{2\pi} \sum_k \frac{I_k}{z - z_k} \quad (3.25)$$

where z is the position of the field point and z_k is the position of the conductor k .

As above the multipole components of a current filament distribution as defined in (3.11) can be found from (3.25) using the expansion (3.20)

$$C_n = -\frac{\mu_0}{2\pi} \sum_k \frac{I_k}{z_k^n} \quad (3.26)$$

The same formulation can be used for the mirror image of a metal screen of $\mu_r = \infty$. The multipole components due to a screen centered at $z = 0$ are found by inverting all the conjugates of the distances z_k at the radius of the screen R .

$$\overline{C}_n = -\frac{\mu_0}{2\pi} \sum_k \frac{I_k}{\left(\frac{R^2}{z_k^*}\right)^n} = -\frac{\mu_0}{2\pi} \frac{1}{R^{2n}} \sum_k I_k z_k^{*n} \quad (3.27)$$

In case the metal screen consists of magnetic material with constant finite permeability ($\mu = \mu_0 \mu_r$) the mirror image can still be calculated by (3.27) modified by a reflection factor

$$\overline{C}_n = -\frac{\mu_0}{2\pi} \frac{\mu_r - 1}{\mu_r + 1} \frac{1}{R^{2n}} \sum_k I_k z_k^{*n} \quad (3.28)$$

Comparing (3.26) to (3.28) the influence of the iron yoke can be estimated. For a single line current in a highly permeable cylindrical yoke ($\mu_r = \infty$) the relative gain with respect to a current line without yoke is found from

$$\frac{\overline{C}_n}{C_n} = \frac{\frac{z_c^{*n}}{R^{2n}}}{\frac{1}{z_c^n}} = \left(\frac{|z_c|}{R} \right)^{2n} \quad (3.29)$$

For the dipole magnets with $R = 98$ mm and a mean value of conductor positions $|z_c| = 42$ mm, (3.29) results in a contribution of 18 % to B_1 but only 0.6 % to B_3 and diminishes for higher orders. Assuming that the yoke influence on the field quality can be kept as low as found in the cylindrical case, this result is important in that it allows separate coil development from yoke design. An advantage of this procedure is taken in the conceptual magnet design and will be employed in the actual optimization examples.

3.4 Symmetry Considerations

Several different types of magnets like sextupoles, octupoles, and decapoles are used to correct higher order multipoles as described in the introduction to the layout of the LHC (sec. 2.2). Azimuthal symmetry in the coil layout is employed to suppress lower order multipoles.

Assuming a multipole symmetry of a conductor distribution the actual multipole components are subject to destructive and constructive interference. Multipolar symmetry is found by rotating the conductor distribution $2M$ times by an angle of $m\pi/M$ with $m \in \{0 \dots 2M - 1\}$ and superposing the images. For each odd m the negative current has to be taken.

$$C_n^M = -\frac{\mu_0}{2\pi} \sum_k \frac{I_k}{z_k^n} \sum_{m=0}^{2M-1} \frac{(-1)^m}{(e^{j\frac{m}{M}\pi})^n} = C_n \sum_{m=0}^{2M-1} e^{j\frac{m}{M}(M-n)\pi} \quad (3.30)$$

The last sum in (3.30) represents a geometric progression which can be evaluated by

$$\sum_{n=0}^{N-1} \alpha^n = \begin{cases} \frac{1-\alpha^N}{1-\alpha} & \text{if } \alpha \neq 1 \\ N & \text{if } \alpha = 1 \end{cases} \quad (3.31)$$

where the solution for $\alpha = 1$ is the continuation of the general case and can be deduced by the rule of De L'Hôpital. Using (3.31), C_n^M is found to be zero except when $e^{j\frac{1}{M}(M-n)\pi} = 1$ leading to the condition $\frac{1}{M}(M-n)\pi = -2\pi p$, where p is a non-negative integer. For the possible multipole components C_n^M follows

$$C_n^M = 2MC_n \quad \forall n = (2p+1)M, p \in \{0, 1, \dots, \infty\} \quad (3.32)$$

As is the case in magnet design, each block is also mirrored in angular position reducing the multipole components to their normal part B_n or skew part A_n , their real or imaginary

component of C_n respectively. Assuming currents I_k at position z_k and their mirror images at z_k^* , the multipole fields C_n are reals consisting of the normal part B_n only.

$$C_n = B_n = -\frac{\mu_0}{2\pi} \sum_k \left[\frac{I_k}{z_k^n} + \frac{I_k}{z_k^{*n}} \right] = -\frac{\mu_0}{2\pi} 2 \sum_k I_k \operatorname{Re} \left\{ \frac{1}{z_k^n} \right\} \quad (3.33)$$

In the special case of two-in-one magnets with two apertures in one yoke, vertical symmetry is maintained, but the horizontal symmetry is not met anymore because of the non-symmetric fields of one aperture acting on the other, so that all B_n values appear, but A_n values are still zero.

3.5 Pure Multipole Fields

A homogeneous field generated by dipoles is needed to keep particles on track along the accelerator tunnel. Quadrupoles are used for focusing and defocusing, keeping the beam stable. Higher order multipolar corrector magnets correct beam properties to achieve the necessary beam quality and life-time.

As has been shown in section 2.7, the multipole components of a magnet design are crucial for its physical performance. Because of manufacturing restrictions and tolerances, a magnet design can never achieve absolutely pure multipole fields. Nevertheless such theoretical configurations exist from which feasible designs can be derived. Pure multipole fields can be obtained by sinusoidal current distributions on a constant radius, and by superposition of elliptical conductors [8]. For any kind of (integer) multipole field, a so-called cosine current distribution is found by varying the current along a constant radius proportional to the desired magnetic field (fig. 3.1). The superposition of elliptical current regions results in

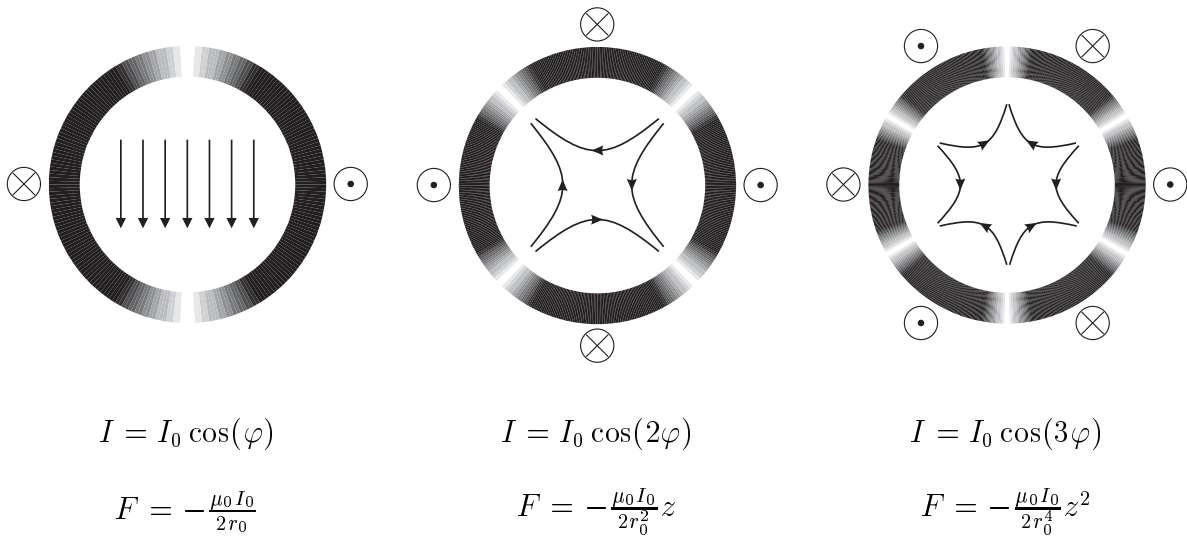


Figure 3.1: Cosine current density distributions. The resulting magnetic fields are ideal dipole, quadrupole, and sextupole fields.

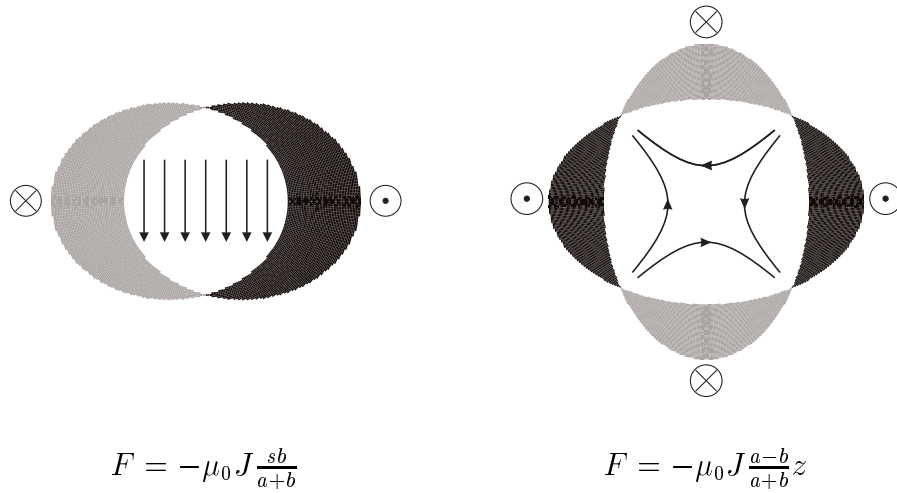


Figure 3.2: Elliptical current distributions with constant current density. Such structures are possible for ideal dipoles and quadrupoles. The ellipses are defined by their half-axes a and b . The ellipse centers are offset by the distance s .

pure dipole or quadrupole fields, depending on the setup. The current density is constant in the whole region of the conductor. In the region of overlap of the ellipses the fields are pure (fig. 3.2). Apart from any costs and constraints, these structures are ideal. The only criterium, field quality, is absolutely reached, although they are not technically feasible. Introducing constraints, however, leads to utopian fitness values which are reasonably close to a feasible solution.

The constraint of critical current and critical field introduced below is necessary to maintain superconductivity in the conductors. The critical current field relationship is approximated as a linear dependence as shown in figure 2.5. The mathematical relationship for the magnetic field B at the position r reads

$$B(r) = B_s \left(1 - \frac{J(r)}{J_s}\right) \quad (3.34)$$

where B_s is the critical field at $J = 0$ and J_s is the critical current at $B = 0$. These limits are found by material tests on sample conductors.

3.6 Segment Approximation

Ideal current distributions are not technically feasible for high-field magnets, because of the special requirements of superconductivity. The cables have to be specifically designed to reduce persistent currents, and to withstand the high electromagnetic forces. The number of turns must be limited because of requirements for the quench protection. As a compromise, manufacturable approximations have to be considered. In first step to a feasible magnet design, the current distribution can be approximated by circular segments. Writing down the equations, it can be seen that certain orders of multipoles may be reduced by solving a system of nonlinear equations.

Assuming constant current in a circular segment and inserting into (3.21) leads to

$$C_n = -\frac{\mu_0}{2\pi} J_0 \int_{r_1}^{r_2} r^{-n} r dr \int_{\varphi=\varphi_1}^{\varphi_2} e^{-jn\varphi} d\varphi \quad (3.35)$$

Doing the integration results in

$$C_n = -\frac{\mu_0}{2\pi} J_0 \frac{j}{(2-n)n} (r_2^{2-n} - r_1^{2-n}) (e^{-jn\varphi_2} - e^{-jn\varphi_1}) \quad (3.36)$$

for all $n \neq 2$. For $n = 2$ the special solution is found as

$$C_2 = -\frac{\mu_0}{2\pi} J_0 \frac{j}{2} \ln \left(\frac{r_2}{r_1} \right) (e^{-j2\varphi_2} - e^{-j2\varphi_1}) \quad (3.37)$$

The number of concurrently optimizable multipole components is dependent on the number of design parameters. A higher number of blocks therefore results in more flexibility in the design. For a higher accuracy of simulation, it is necessary to calculate the correct position of the conductors from the technical data, taking the manufacturing procedure into account, since the sensitivity to tolerances is high.

In many cases such multi-block distributions are used as a starting point for a magnet design. Basic parameters as the obtainable field strength and the amount of necessary superconductor volume can be estimated. Before high quality tools for conceptual design were available, feasible designs were found starting with simple coil blocks, which are susceptible to an analytical investigation. Refining the design by subdividing these major blocks, allowed for the compensation of higher order multipoles in ideal block structures. A resulting coil block structure was then used as starting solution in numerical simulations.

3.7 Persistent Currents

Section 2.5 about superconductivity introduced the strong magnetic hysteresis of hard superconductors. The effect is explained by flux tubes that are screened by little current loops creating a magnetization. The superposition of these current loops is known as persistent current. A macroscopic model of persistent currents in superconducting filaments was first developed by Bean [9].

The magnetization process consists of two phases, the penetration phase and the fully penetrated phase. In the penetration phase only the surface of the superconductor is penetrated as is the case for all types of superconductors. Above a threshold B_c superconductivity in a type I superconductor breaks down, whereas in a type II superconductor flux tubes are generated, which allow the magnetic field to penetrate the superconductor.

The filament magnetization can be approximated by adjusting an analytical model to a measured magnetization-field relationship. The penetration phase is modeled by a parabola whereas the fully penetrated phase is fitted as $M(B) = \alpha B^{-n}$. The actual magnetization is dependent on the history of excitation. In figure 3.3 the magnetization curves used in the

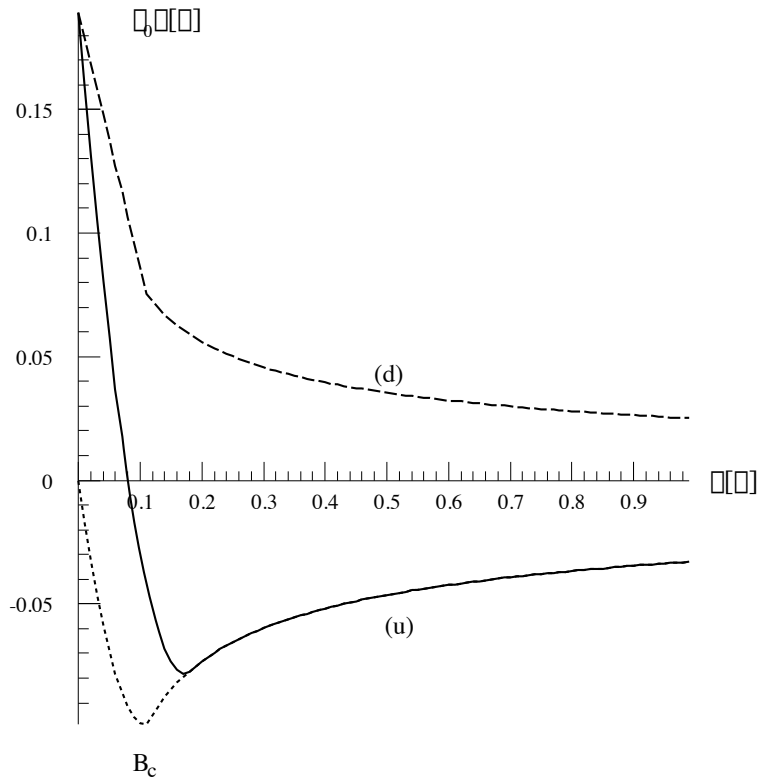


Figure 3.3: Approximation of the filament magnetization as used in REM. Two phases: $B < B_c$ penetration phase, $B > B_c$ fully penetrated phase. u ... up-ramping, d ... down-ramping

program REM as incorporated into ROXIE for the main dipole inner cable is shown [10]. The agreement to measurements for the LHC dipole models at the short model test facility is investigated in [11].

Using this magnetization curve the modulus of the magnetization is found. The direction of the magnetization is assumed to be the same as of the magnetic excitation field. This is exactly the case if the yoke material is assumed to be linear.

In the following the filament magnetization is introduced as a current dipole of two line currents of equal size I and opposite sign and a displacement vector δ pointing from the negative to the positive current line. The current dipoles are equivalent to a magnetization $M = M_x + jM_y$ integrated over the filament area A

$$I\delta = jMA \quad (3.38)$$

The influence of the persistent currents on the magnetic field can now be calculated similarly to (3.25). The magnetic field of a number of current dipoles is evaluated as

$$B_y + jB_x = \frac{\mu_0}{2\pi} \sum_k \left[\frac{I_k}{z - (z_k + \frac{\delta_k}{2})} - \frac{I_k}{z - (z_k - \frac{\delta_k}{2})} \right] \quad (3.39)$$

where δ_k is the complex displacement in the filament k with current I_k .

Rewriting this relation yields

$$F(z) = B_y + jB_x = \frac{\mu_0}{2\pi} \sum_k \frac{I_k \delta_k}{(z - z_k)^2 + \left(\frac{\delta_k}{2}\right)^2} \approx \frac{\mu_0}{2\pi} \sum_k \frac{I_k \delta_k}{(z - z_k)^2} \quad (3.40)$$

where the small term $\left(\frac{\delta_k}{2}\right)^2 \ll (z - z_k)^2$ can be ignored. Deriving the multipole components as in (3.26) leads to

$$C_n = n \frac{\mu_0}{2\pi} \sum_k \frac{I_k \delta_k}{z_k^{n+1}} \quad (3.41)$$

The influence of an iron yoke can now be calculated as for single line currents in (3.28).

3.8 Sensitivity

During the conceptual design phase it is particularly important to take sensitivities to manufacturing tolerances into account. Systematic errors due to manufacture have to be traced and compensated by design modifications, while random errors can only be compensated by sophisticated magnet sorting schemes before installation in the tunnel [12, 13]. Low sensitivity can safeguard against both sources of error during design.

Practically the sensitivity to errors may be estimated by the differential quotient around the nominal design. In the following an analytical approach is used, leading to further insight. The sensitivity of the multipole components of a current distribution to displacements is calculated directly from the multipole components themselves.

Deriving (3.11) yields

$$\frac{dF}{dz} = \frac{d}{dz} \sum_{n=1}^{\infty} C_n z^{n-1} = \sum_{n=2}^{\infty} (n-1) C_n z^{n-2} = \sum_{n=1}^{\infty} n C_{n+1} z^{n-1} \quad (3.42)$$

A movement of the field calculation point z in a direction dz is equivalent to a movement of the whole conductor distribution in the opposite direction $-dz_k$.

$$\frac{dF}{dz_k} = \sum_{n=1}^{\infty} \frac{dC_n}{dz_k} z^{n-1} \quad (3.43)$$

From the theory of analytical functions it is known that two Laurent expansions are equal if the analytical functions described are equal, from which follows that

$$\frac{dC_n}{dz_k} = -n C_{n+1} \quad (3.44)$$

resulting in an approximate error of ΔC_n due to a shift δ

$$\Delta C_n = -n C_{n+1} \delta \quad (3.45)$$

For a conductor distribution with (3.26) therefore follows

$$\Delta C_n = n \frac{\mu_0}{2\pi} \sum_k \frac{I_k}{z_k^{n+1}} \delta \quad (3.46)$$

which is similar to (3.41) as can be anticipated from the idea that a conductor movement is equivalent to a current dipole of a current in the new position and another current of opposite sign at the old position compensating the current there.

As an example, change in field quality due to the movement displayed in fig. 2.7 is estimated. The block of four conductors can be represented approximately by a line current in its center at about $z = (1.55 + j 3.13)$ cm. The azimuthal shift of about $\delta = 0.03$ mm is a non-symmetric change of the current distribution. All normal and skew components now appear. The movement leads to a new quadrupole component of $\Delta c_2 = (-0.147 + j 0.112)$ units and a change of the sextupole component of $\Delta c_3 = (0.0147 + j 0.0760)$ units with respect to a reference radius of $r_{ref} = 10$ mm. These values are within few percents of the more exact calculation of ROXIE. The influence of a mirror current due to the iron yoke with an inner radius of $R = 84.75$ mm is taken into account.

Relation (3.45) directly explains another source of error, the effect of a misalignment of two magnets in the lattice. Slight shifts of the magnet axis result in multipole components which are proportional to the multipole component of the next higher order. This effect is termed feed-down in accelerator design.

Assuming again a coil as shown in figure 2.7 with a normal sextupole component of $b_3 = 0.030$ units, a horizontal shift of $\delta = 0.03$ mm results in a normal quadrupole of $\Delta b_2 = 0.00018$ units. A vertical shift of the same amount leads to a equally large skew component, as found with (3.45). Comparing the shift of the whole magnet due to an alignment error, to the single block shift discussed before, it might be surprising that the change in quadrupole is much smaller. This difference is explained by the fact that the sextupole c_3 of the whole magnet is much smaller than the sextupole of each block, since the design strives to compensate this error.

3.9 Iron Cross-Section

The FEM method used throughout this thesis is FEM2D from the IGTE (Institut für Grundlagen und Theorie der Elektrotechnik) Technical University Graz. The software employs the so-called reduced vector potential method [14, 15]. The basic considerations that lead to this method are described below.

The main advantage of the reduced vector potential method is that the coil does not need to be meshed. The current distribution in the coil can be modeled accurately and the resulting fields are calculated analytically by Biot-Savart's law. Another benefit of this method is the higher precision of the finite element calculations, as only the comparably low iron induced fields are treated numerically (sec. 3.3). At low fields cancellation errors are avoided which are encountered in highly permeable material regions at the injection (low-field) level in a full vector potential formulation. Of further advantage is the low reduced field

in the far field boundary. The effect of inaccurately set boundary conditions is consequently low.

This characteristic is achieved by separating the coil field \vec{B}_s from the field due to the yoke excitation \vec{B}_r

$$\vec{B} = \vec{B}_s + \vec{B}_r = \mu_0 \vec{H}_s + \nabla \times \vec{A}_r \quad (3.47)$$

where H_s is the Biot-Savart field from the coils and A_r is the vector potential representing the excitation of the yoke. The equation automatically satisfies the Maxwell equation $\nabla \cdot \vec{B} = 0$ since the coil field is divergence free. Inserting (3.47) in the Maxwell equation $\nabla \times \vec{H} = \vec{J}$, yields

$$\nabla \times \frac{1}{\mu} \vec{B}_r = \vec{J} - \nabla \times \frac{1}{\mu} \vec{B}_s \quad (3.48)$$

Considering two different material regions of magnetic iron Ω_i and air Ω_a , (3.48) can be simplified for each region. In the air region Ω_a the magnetization is zero ($\mu = \mu_0$) and the right-hand side vanishes. In the iron region the current is zero ($J = 0$).

$$\begin{aligned} \nabla \times \frac{1}{\mu} \vec{B}_r &= \vec{0} \quad \text{in } \Omega_a \\ \nabla \times \frac{1}{\mu} \vec{B}_r &= -\nabla \times \frac{1}{\mu} \vec{B}_s \quad \text{in } \Omega_i \end{aligned} \quad (3.49)$$

Consequently the current is removed from the formulation, avoiding a special treatment.

At the finite element boundaries Γ_{12} between two regions Ω_1 and Ω_2 , boundary conditions must be satisfied. Neglecting any surface currents or fictitious magnetic surface charge density, the conditions read

$$\left. \begin{aligned} \vec{n} \cdot (\vec{B}_2 - \vec{B}_1) &= 0 \\ \vec{n} \times (\vec{H}_2 - \vec{H}_1) &= \vec{0} \end{aligned} \right\} \text{on } \Gamma_{12} \quad (3.50)$$

with \vec{n} a normal vector on the boundary. In addition, boundary conditions on the outer boundaries of the problem are needed. For each element either a Dirichlet boundary condition Γ_B or a Neumann boundary condition Γ_H has to be defined. Neglecting any magnetic surface currents or charges, the conditions read

$$\vec{n} \cdot \vec{B}_r = -\vec{n} \cdot \vec{B}_s \quad \text{on } \Gamma_B \quad (3.51)$$

$$\vec{n} \times \frac{1}{\mu} \vec{B}_r = -\vec{n} \times \frac{1}{\mu} \vec{B}_s \quad \text{on } \Gamma_H \quad (3.52)$$

Equation (3.48) together with the boundary conditions defines the complete boundary value problem. With a vector weighting function \vec{w} that satisfies additional homogeneous Dirichlet boundary conditions, a weak formulation can be derived

$$\int_{\Omega} (\nabla \times \vec{w}) \cdot \frac{1}{\mu} (\nabla \times \vec{A}_r) d\Omega = \int_{\Omega} (\nabla \times \vec{w}) \cdot \left(\vec{H}_s - \frac{\mu_0}{\mu} \vec{H}_s \right) d\Omega \quad (3.53)$$

with the reduced vector potential \vec{A}_r and the source field \vec{H}_s of (3.47). The weak formulation (3.53) is discretized into finite elements in the whole region Ω . By Galerkin's method, the same shape functions f_k are used for the reduced vector potential \vec{A}_r and the vector weighting functions \vec{w} . In the 2d-implementation iso-parametric 9-noded quadrilateral finite elements are used

$$\vec{A}_r \approx \vec{A}_D + \sum_{k=1}^K \vec{A}_{r_k} f_k \quad (3.54)$$

where \vec{A}_D satisfies the Dirichlet boundary condition (3.51). All equations together constitute a non-linear system of equations, which is solved iteratively, until the permeability distribution converges.

The high precision together with the accurate modeling of the coil renders the reduced vector method favourable for accelerator magnet design. The convergence of the magnetic permeability is achieved in few iterations. The integration in the ROXIE environment allows the optimization of parameterized models. The adaption of genetic algorithms for the solving of material distribution problems added a new perspective to its application. In chapter 7, this finite element method is employed in the conceptual design. In the course of the chapter, several finite element optimizations are carried out.

Chapter 4

Mathematical Optimization

The aim of magnet optimization is to find feasible designs with multipole field errors as low as possible and a quench margin as large as possible. Since the notion “as good as possible” is not sufficiently precise for a numerical treatment, the frame for the optimization must be set. Real world problems normally involve multiple conflicting objectives which cannot all be met at the same time. Crucial to the success is the correct formulation of the optimization problem. Objectives which cannot be directly quantified, or which are too complicated for the particular design phase, must be considered at a later stage. The remaining objectives of the vector optimization problem are treated by appropriate techniques. Constraints have to be met either by the optimization algorithm or by a transformation of the constrained problem into an unconstrained. Particular constraints are incorporated into the objective function. Other constraints have to be checked against the solutions.

A goal programming method is employed to formalize the problem. Weights and residuals are defined by analytical considerations or iterative adjustments. A stopping criterion must be found and a priori estimates help measure the quality of the reached results. Both can be achieved by utopian objective values gained from analytical or numerical evaluations. The two analytical solutions for pure multipole fields shown in the last chapter fulfill these requirements in coil optimizations. These current distributions are neither technically optimal, nor do they take persistent currents into account as occur in real designs of superconducting magnets. The main-field, however, is high and unwanted multipoles are absolute zero, unreachable in real-world designs, which justifies their use as utopian solutions. Finally, the choice of optimization algorithm is important by its type and robustness. The difference in speed of several methods is only secondary since the final optimization result is more important.

4.1 Optimization Formulation

Constrained nonlinear optimization problems are typically formalized by a set of n parameters $\{x_1, \dots, x_n\}$, also denoted as vector $\vec{x} = (x_1, \dots, x_n)^T$ in the n -dimensional space R^n , and a nonlinear objective function F . The general optimization problem can be written as

$$\min_x F(\vec{x}) \quad F : R^n \longrightarrow R \quad (4.1)$$

subject to nonlinear constraints

$$\begin{aligned} g_i(\vec{x}) &\leq 0 & i = 1, 2, \dots, l \\ h_j(\vec{x}) &= 0 & j = 1, 2, \dots, m \end{aligned} \quad (4.2)$$

and bounds for the design variables

$$x_{kl} \leq x_k \leq x_{ku} \quad k = 1, 2, \dots, n \quad (4.3)$$

which define the feasible domain $D \subset R^n$. Without restriction of generality all optimization problems can be handled as minimization problems since a maximization of $F(\vec{x})$ is equivalent to a minimization of $-F(\vec{x})$.

A function $F(\vec{x})$ in the domain $D \subset R^n$ is said to have a global minimum at \vec{x}_0 if

$$F(\vec{x}_0) \leq F(\vec{x}) \quad \forall \vec{x} \in D \quad (4.4)$$

For continuous unconstrained problems a necessary condition for optimality is defined by the gradient vector

$$\nabla F(\vec{x}_0) = 0 \quad (4.5)$$

A local minimum is defined by the sufficient condition that the Hessian H

$$H = \left[\frac{\partial^2 F}{\partial x_i \partial x_j} \right] \quad (4.6)$$

is positive definite in a point \vec{x}_0 .

4.2 Pareto Optimality

Real-world optimization problems usually require the simultaneous optimization of multiple q conflicting objectives $F_p(\vec{x})$ with $p = 1, 2, \dots, q$ defining a vector optimization problem $\text{Min } \vec{F}(\vec{x})$. The operator “Min” is not a standard minimization, as a multiple of objectives have to be minimized concurrently. It is probable to find a solution which is best with respect to one objective, and a second which is best with respect to another objective. Therefore the scalar concept of optimality does not apply directly in the multiobjective setting. A practical replacement is the notion of Pareto optimality. A solution \vec{x}_0 is said to be (globally) Pareto optimal, if there is *no* solution, for which

$$F_p(\vec{x}) \leq F_p(\vec{x}_0) \quad \forall \vec{x} \in D \wedge p \in \{1, 2, \dots, q\} \quad (4.7)$$

with one strict inequality $F_p(\vec{x}) < F_p(\vec{x}_0)$. The Pareto optimal set consequently consists of all solutions \vec{x}_0 , where the improvement of any of the objectives leads to a degradation in at least one other objective. In figure 4.1 two such regions, $[x_1, x_2]$ and $[x_4, x_5]$, are shown. An additional region $[x_3, x_4]$ exists with locally Pareto optimal solutions. For points in this region (4.7) applies, with respect to points x in a neighbourhood of \vec{x}_0 . Ideally an optimization algorithm should report the full set of Pareto-optimal solutions. In practice however, it is impossible to generate all the solutions of a numerical problem in a reasonable time.

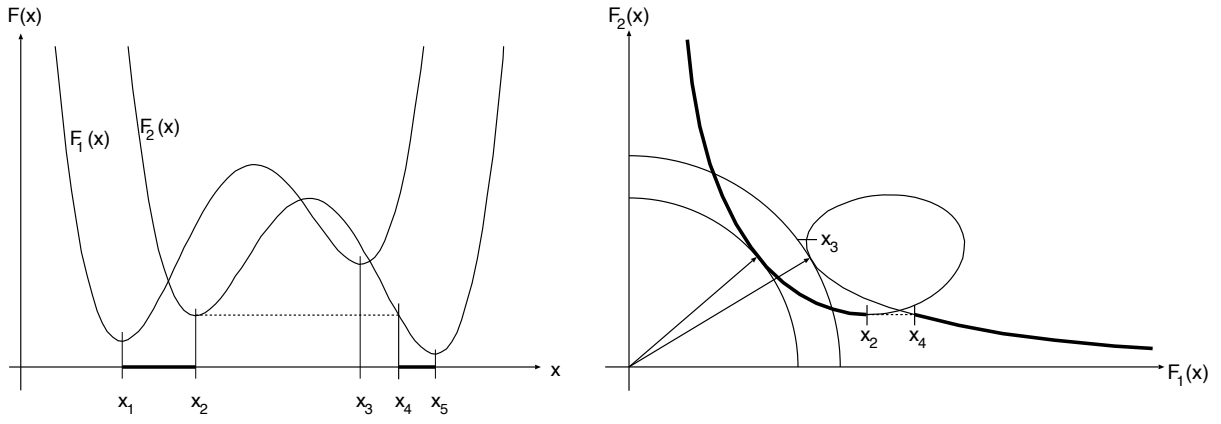


Figure 4.1: The optimization problem of two independent objective functions has two regions of (globally) Pareto optimal solutions $[x_1, x_2]$ and $[x_4, x_5]$. A further region $[x_3, x_4]$ is locally Pareto optimal. The two objective functions (left) are shown as parametric plot (right). A weighted distance function, combining the two objectives F_1 and F_2 , results in a circle where the radius is to be minimized.

4.3 Objective Weighting

One of the classical methods of multiobjective optimization is objective weighting and is used in the following. Several objectives F_p are combined into one overall objective function F by means of weighting factors $w_p > 0$.

$$F(\vec{x}) = \sum_p w_p F_p(\vec{x}) \quad (4.8)$$

For each weighting vector, a different Pareto-optimal solution is found. This objective weighting method can be combined with the method of distance functions, with the residuals y_p

$$F(\vec{x}) = \left(\sum_p w_p |F_p(\vec{x}) - y_p|^r \right)^{\frac{1}{r}} \quad (4.9)$$

In case that $F_p(\vec{x}) > y_p$ for all p , the whole Pareto optimal set can be constructed by optimization of F for all combinations of w_p or all possible r . Usually only $r = 1$ and $r = 2$, the absolute and the Euclidean metric are used.

Since the determination of the Pareto-optimal set is costly, and in many cases a certain importance can be attributed to each objective, properly chosen weights w_p will narrow down the search in the feasible domain. These weights are either found by analytical considerations or have to be adjusted by an iterative procedure. The residuals y_p are often clear from the problem definition or have to be found by the evaluation of individually best parameter values or assumptions. In figure 4.1 right, the weighting function with $y_p = 0$ and $r = 2$ is drawn. The two weights are equal in this example, resulting in concentric circles. A Pareto optimum is found, where the smallest circle just osculates the function.

In case that all objectives and restrictions are included in the vector optimization problem, the Pareto optimal set usually consists of all useful solutions. In many optimization problems, however, not all of the objectives can be included, because of several reasons:

- The objectives are fuzzy rather than clearly defined.
- A necessary restriction compromises the convergence of the optimization algorithm.
- The objectives are dependent on a physical relation which can not be cast into a formula, or which would unnecessarily complicate the problem formulation or computability.

It is therefore necessary to consider local optima as well. Further restrictions may move the focus from infeasible global optimization results to possibly feasible local Pareto solutions.

Another reason why locally optimal solutions are of importance is the ambiguity in the choice of weights. The weighting factors represent the user's preference of the objectives. This preference, however, has to be adjusted to the sensitivity of the objectives. This sensitivity represents the difficulty with which an objective may be reached. Since the sensitivity is different for any local optimum, the weights cannot properly account for the user's preference in global optimization. As a consequence, chosen weights will usually not result in optimal solutions with regard to the design goals. It is therefore normally necessary to locally post-process solutions found by a global optimization.

In the example of figure 4.1 another point should be mentioned. A local Pareto optimum is shown in the region $[x_3, x_4]$. An adaption of the weights in a post-processing by local optimization, would result in an elliptically shaped weighting function. In case that the local optimum moved via x_4 , the local optimization would lead to a global optimum within the prior constraints. This solution would be missed by a pure local optimization starting from a point in the region $[x_1, x_2]$.

4.4 Utopia Solution

Though the optimization should be restricted as little as possible, a frame for the optimization still has to be set. Essential considerations are necessary to determine degrees of freedom and absolute bounds to the design space, where they are not obvious. In order to get insight into the problem, standard optimization procedures are appropriate, as shown in the considerations below.

Engineering design often starts from ideal solutions which are found analytically. Simple solutions allow the a priori estimation of design goals. This process is described for magnet cross-sections, below. The calculations will give an estimate for the maximally reachable field at perfect field quality. Such a utopian value may serve as possible residual to the method of distance functions. Beyond that, utopian quality measures are important to define a stopping criterium in stochastic optimization. Additionally it may serve as valuable information to the designer, checking the quality of the results gained.

The use of payoff tables allows measurement of the performance of a local solution a posteriori and helps define possible objectives in a further local optimization step. In order

to create such a table, individual optimization problems are solved, to find the best solution for each of the q objectives. The individual optima together define a utopian solution which may be used in a distance function formulation. In addition to the optimal values, the price for each objective is defined by the values of the other parameters. These numbers can be used as an estimate for the difficulty in reaching the objective. An example of payoff tables is shown in chapter 7.

4.4.1 Circular Cylindrical Conductor

In order to find the maximally reachable field in a conductor distribution, first a circular cylindrical conductor is investigated. The simple cylinder symmetric structure allows one to calculate and optimize the field distribution analytically. Imposing the constraints of superconductivity, the reachable field in the material region is limited by the current in that position. Assuming constant current in the whole material cylinder leads to a limiting field on the surface of the conductor. Constant current, however, does not produce the maximally achievable field. A radially varying current distribution is therefore derived which produces higher fields. Consequently the conductor is maintained at the critical field level in every point.

Starting with Maxwell's equation of the magnetic field in integral form for the stationary case (Ampère's law)

$$\int_A (\nabla \times \vec{H}) \cdot d\vec{A} = \int_s \vec{H} \cdot d\vec{s} = \int_A \vec{J} \cdot d\vec{A} \quad (4.10)$$

For a circular symmetric conductor this integral formulation can be simplified into

$$B(r) = \frac{\mu_0}{r} \int_0^r J(\rho) \rho d\rho \quad (4.11)$$

Assuming that the circular conductor consists of a large number of filaments, any current distribution can be realized. Maintaining cylindrical symmetry, the current distribution along the radius is optimized, so that the resulting field at the surface of the conductor is maximal. As a limitation to the field level, the relation (3.34) is used. Substituting (3.34) for $B(r)$ in (4.11) and differentiating with respect to r , a linear differential equation is found

$$J(r) + \frac{J_s}{B_s} \mu_0 r J(r) + r \frac{dJ(r)}{dr} = J_s \quad (4.12)$$

Solution of this equation results in

$$J(r) = C \frac{1}{r} e^{-\frac{J_s}{B_s} \mu_0 r} + \frac{B_s}{\mu_0 r} \quad (4.13)$$

where C is an integration constant. The boundary value of $J(0)$ must observe the critical current limit $J(0) \leq J_s$ and therefore be finite, resulting in

$$C = -\frac{B_s}{\mu_0} \quad \text{and} \quad J(0) = J_s \quad (4.14)$$

as the only suitable solution. This determines

$$J(r) = \frac{B_s}{\mu_0 r} (1 - e^{-\frac{J_s}{B_s} \mu_0 r}) \quad \text{and} \quad B(r) = B_s - \frac{B_s^2}{J_s \mu_0 r} (1 - e^{-\frac{J_s}{B_s} \mu_0 r}) \quad (4.15)$$

for the current distribution and for the field distribution respectively.

Gain over Constant Current

For a comparison, a circular conductor is assumed with the same radius as the optimized structure, carrying a constant current density $J(r) = J_0$ with respect to the radial position. Also in this case the conductor is operated at the critical current limit which is only reached at the surface. As is well known and can be derived from (4.10) the magnetic field increases linearly with the radius

$$B(r) = \frac{\mu_0 r}{2} J_0 \quad (4.16)$$

Using the critical current relation (3.34), the equation (4.16) can be reformulated as a function of the conductor radius r_c

$$B_{cc}(r_c) = \frac{B_s}{1 + \frac{2B_s}{J_s \mu_0 r_c}} \quad (4.17)$$

Comparing these two results a gain may be defined $G = \frac{B_{opt}}{B_{cc}} - 1$ with a maximum of 13.94% at $r_{max} = 2.688 \frac{B_s}{\mu_0 J_s}$ (fig. 4.2).

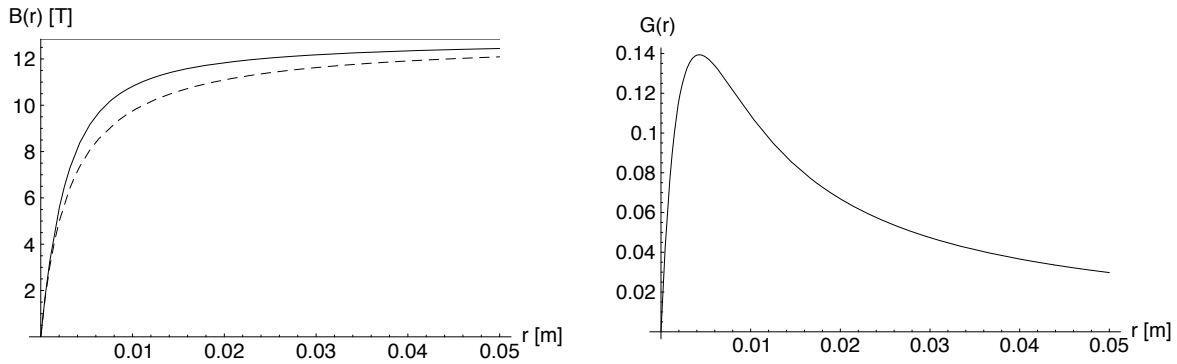


Figure 4.2: Maximum field and gain over the radius of a circular cylindrical conductor of a material as used in the 6 block design ($J_c = 6434 \text{ Amm}^{-2}$ and $B_c = 12.86 \text{ T}$). The optimized current distribution (continuous) is compared to a constant current (dashed), resulting in a gain (right).

Consequently it is possible to reach an almost 14% higher magnetic field with a special current distribution compared to a constant current under the constraint of the critical properties of the conductor. This result is especially interesting when applied to feasible current distribution. As a further step in this direction the magnetic field gain for ideal sinusoidal structures is evaluated.

4.4.2 Sinusoidal Current Distribution

The calculation of the field gain is repeated here for sinusoidal current distributions. While the azimuthal distribution is defined as shown in figure 3.1, the radial distribution may still be chosen independently. While the fields from a constant current distribution may still be calculated analytically, the optimized current distribution is found by numerical optimization with a deterministic algorithm. The simplest way to calculate stationary magnetic fields from sinusoidal current distributions is via the multipole expansions (3.11) with (3.21). Using a sinusoidal current distribution

$$J(z) = J(x, y) = J(r, \varphi) = J(r)\cos(\varphi) = J(r)\frac{e^{j\varphi} + e^{-j\varphi}}{2} \quad (4.18)$$

and inserting (4.18) into (3.21) yields

$$C_n = -\frac{\mu_0}{2\pi} \int_{r_1}^{r_2} J(r)r^{-n} r dr \int_{\varphi=0}^{2\pi} \frac{e^{j\varphi} + e^{-j\varphi}}{2} e^{-jn\varphi} d\varphi \quad (4.19)$$

Evaluating the coefficients C_n , only one term of the expansion for $n = 1$ remains

$$F_i(z) = C_1 \quad (4.20)$$

where

$$C_1 = -\frac{\mu_0}{2} \int_{r_1}^{r_2} J(r)dr \quad (4.21)$$

In the case where the current distribution is inside a circle and the measurement is done outside $|z| > |z_c|$, inserting the sinusoidal current distribution (4.18) into (3.23) as above, the only nonzero component is found to be

$$C_{-1} = \frac{\mu_0}{2} \int_{r_1}^{r_2} J(r)r^2 dr \quad (4.22)$$

and the expansion (3.14) reduces to

$$F_o(z) = C_{-1}z^{-2} \quad (4.23)$$

The complete field solution for the sinusoidal current distribution, inside the current region is thereby found as the sum of F_i and F_o

$$F(z) = F_i + F_o = -\frac{\mu_0}{2} \int_r^{r_2} J(\rho)d\rho + \frac{\mu_0}{2} \frac{1}{z^2} \int_{r_1}^r J(\rho)\rho^2 d\rho \quad (4.24)$$

Since the optimization of the radial current distribution has to take the critical field properties into account at any point azimuthally, the optimization is not as simple as for the

circular cylindric conductor. The solutions are therefore found numerically for a discretized current distribution, using the ROXIE program. Analytically the field can be calculated for a radially constant current density $J(r) = J_0$ for $r_1 < r < r_2$, resulting in

$$F(z) = -\frac{\mu_0}{2} J_0 \left[r_2 + \frac{1}{3} \frac{r_1^3}{r^2} e^{-2j\varphi} - r \left(1 + \frac{1}{3} e^{-2j\varphi} \right) \right] \quad (4.25)$$

Numerical Investigation

What was shown for a circular conductor, is true for a sinusoidal current distribution as well: The maximum field in the aperture can be increased by optimizing the radial current distribution in the conductors. The limit that must not be exceeded is again defined by the critical current - critical field relationship.

$$|F(z)| = B_s \left(1 - \frac{J(r)}{J_s} \right) \quad (4.26)$$

Numerical optimization was used to investigate the possible field gain in such a structure using the cable parameters and iron yoke diameter as defined for the main dipole magnets (sections 2.3, 2.4). The current distribution was optimized for 32 different cable heights and compared to the maximum field for radially constant current. The resulting curves

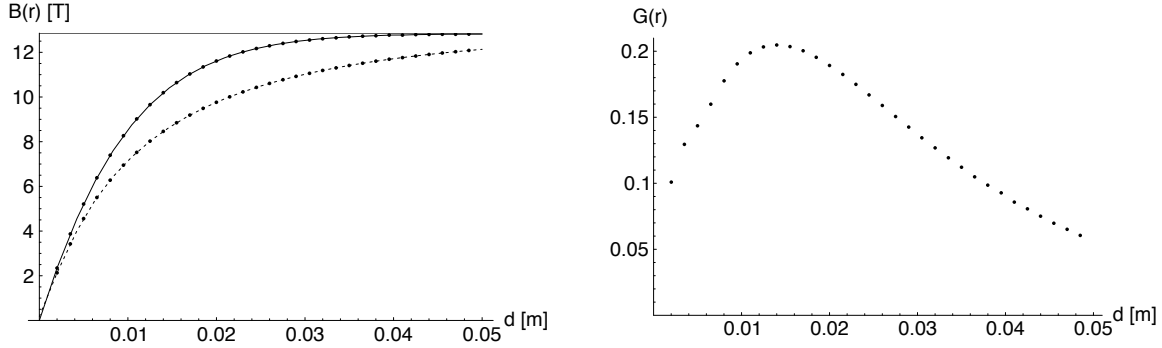


Figure 4.3: Maximum field and gain of a cosine distribution of a standard conductor with a height d of material as used in the 6 block design. The radially optimized current distribution (continuous) is compared to a radially constant current (dashed) resulting in a gain (right).

are similar to those of the cylindrical conductor in figure 4.2. Using the gain definition $G = \frac{B_{opt}}{B_{cc}} - 1$ as above, a maximum of almost 20.5% at about $d_{max} = 14$ mm is found for a conductor type like the inner cable of the 6 block design.

The radial current distribution in the conductor is shown in figure 4.4 and supersedes the constant current situation by a factor of more than 1.6. Nevertheless the higher field in the aperture can only be reached by a lower current density at the inner edge. The current density is shown for the optimal conductor height which is close to the actual height of 15mm in the dipole design.

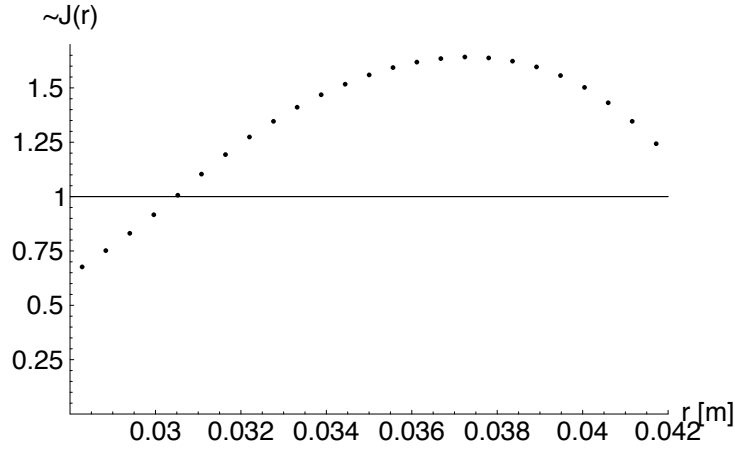


Figure 4.4: Current distribution factor of the current density over the radius of the magnet for an optimal conductor of a height of 14 mm compared to an optimal constant current.

Concluding this section the optimal current distribution is shown in figures 4.5 and 4.6. The radially optimized cosine current distribution (left) leads to a minimum quench margin (right) at a certain point in each structural layer. This has to be compared to the standard radially constant current distribution as approximately existent in feasible magnet designs.

Though the results suggest that an optimal radial current distribution can increase the maximum field considerably, application to magnet design is limited. In the main dipole design of the LHC, different conductors carrying the same current are used for the inner and the outer layer, resulting in different current densities.

4.4.3 Intersecting Elliptical Conductors

As discussed above, elliptical conductors are an alternative to the cosine current distribution with pure multipole fields. The fact that the current density is constant over the whole cross-section of the conductors renders this analytically treatable structure useful for the calculation of a utopian solution.

For a numerical analysis with ROXIE including persistent currents, segment-shaped conductors were cut in their height to achieve this particular shape. Specifying the aperture radius r_a and the conductor height h , the semi-axes a and b are given by the assumption that the ellipse is shaped around the aperture defining r_a as the radius of the osculating circle of the ellipse.

$$a = r_a + \frac{h}{2} \quad (4.27)$$

$$b = \sqrt{a r_a} = \sqrt{r_a \left(r_a + \frac{h}{2}\right)} \quad (4.28)$$

A plot of the resulting field distribution is shown in figure 4.7.

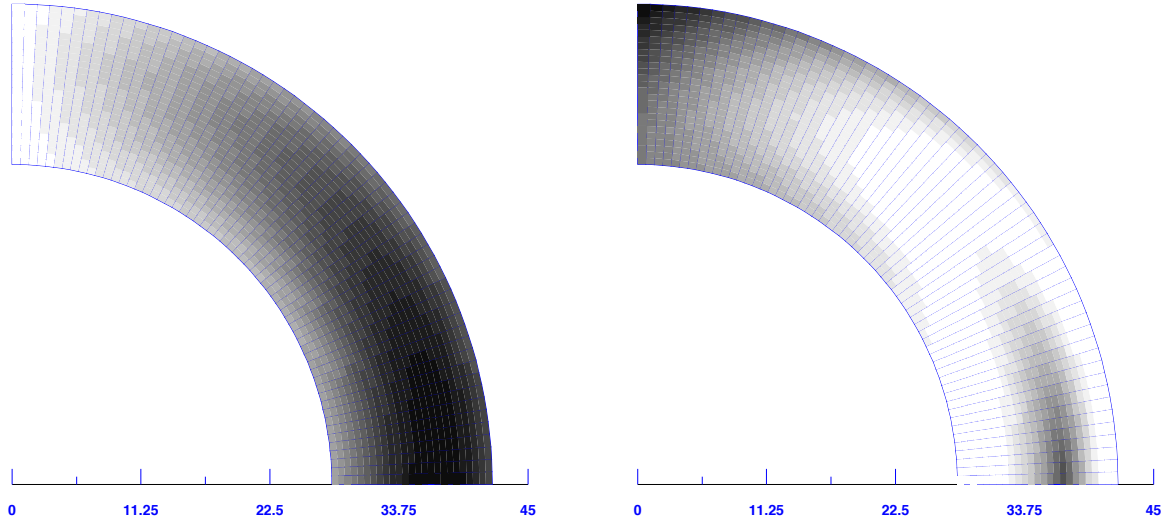


Figure 4.5: Current distribution (white in plot 0 A/mm^2 – black 1404 A/mm^2) and quench margin (white 0% – black 47%) over the radius (in mm) for the radially optimized current distribution. Each radial layer reaches 0% quench margin. (The radius is given in mm.)

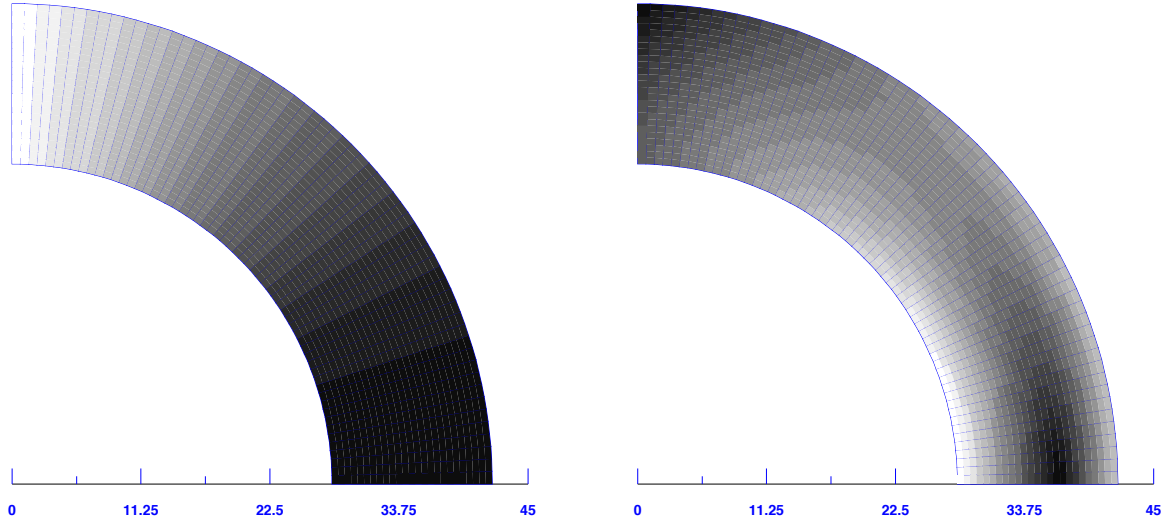


Figure 4.6: Current distribution (white in plot 0 A/mm^2 – black 855 A/mm^2) and quench margin (white 0% – black 51%) over the radius (in mm) for the radially constant current distribution. The innermost layer limits the current density in the whole magnet. (Radius in mm.)

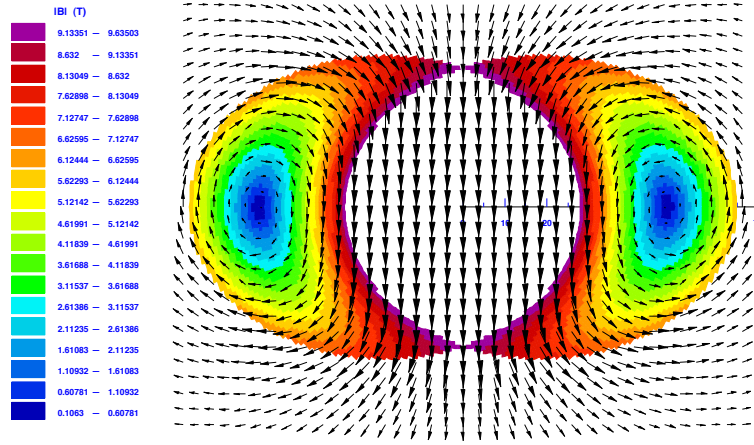


Figure 4.7: Field distribution of the intersecting ellipses design for comparisons with the main dipole cross-section of LHC (V6-1).

For constant current density (using the inner conductor) the intersecting ellipses are set up to yield the same field at the same current density resulting in $r = 28$ mm $h = 35.9$ mm. A geometrical comparison is shown in figure 4.8. In table 4.1 the main parameters of the two structures are compared.

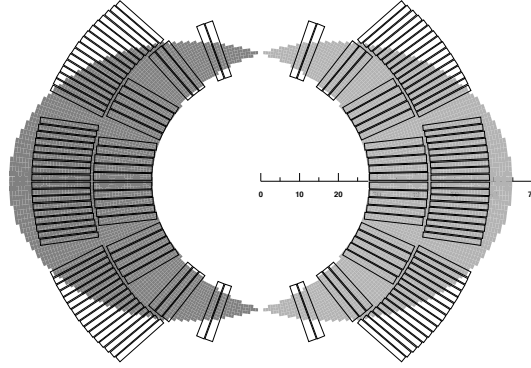


Figure 4.8: Geometrical comparison of the standard main dipole cross-section with intersecting ellipses. The current distribution of the ellipse design is less effective but reaches a higher field level.

Though the special cosine and elliptical structures are perfect with regard to direct magnet field errors, persistent currents as described in section 2.5 lead to field distortions. The amount of persistent current multipoles in the case of intersecting ellipses is even higher than for the standard design. The maximally reachable magnetic field, however, is more than 8% higher than in a representative technically feasible case. This is not surprising as the field of the standard cross-section is not constant over the whole aperture, but a superposition of the multipole components. Therefore the peak field is usually a few percents higher than the main field. Assuming an optimized current distribution, an even higher field level should

	V6-1	ellipse
I (A)	13460	13460
$B(I)$ (T)	9.76	9.76
B_{ss} (T)	9.76	10.6
NI/B (A/T)	110346	105729
SC (mm ²)	1283	1522
b_3 (pers)	-3.67	-5.81
b_5 (pers)	0.15	0.11
b_3 (geom)	1.41	0.00
b_5 (geom)	-0.1055	0.00
b_7 (geom)	0.0255	0.00

Table 4.1: Comparison of an elliptical cross-section to the standard main dipole cross-section of LHC. The ellipse is set up to provide the same field at the same current density. B_{ss} ... maximally reachable field (the quench margin is zero), SC ... total area of superconducting material in the cross-section, pers ... errors due to persistent currents at injection, geom ... geometrical field errors.

be attainable. Estimations may take the results of the preceding section into account and therefore assume about 10% improvement according to the gain value for a 37 mm thick cable (fig. 4.3). A utopian field level should therefore be found at about 11 T for an ideal structure of a conductor as used in the inner layer of the standard LHC cross-section, if all resources for field strength are exploited.

4.5 Optimization Techniques

With the aim of doing conceptual design, optimization methods have to be reviewed. A quick introduction to the different methods is given and the particular needs of the conceptual design phase are discussed.

All optimization algorithms can be assigned to either of two groups:

- Deterministic Optimization Methods
- Stochastic Optimization Methods

Deterministic methods show mostly local behaviour. They usually operate on continuous function spaces. Depending on their reliance on derivatives

- Search Methods, and
- Gradient Methods

are distinguished. While gradient methods need to know the derivatives of the objective function with respect to the parameters, search methods can be used without this knowledge. Pseudo-gradient methods are special gradient methods which try to combine the

advantages of both methods and estimate the gradient from prior solutions. Examples for search methods are the simplex algorithm [16] and EXTREM [17]. The simplex method is based on the idea of comparing the values of the objective function at the $n + 1$ vertices of a simplex (polytope) in n -dimensional space, moving the simplex gradually towards the minimum. EXTREM uses one-dimensional minimizations extrapolating in a main search direction and an conjugate direction evaluated by Gram-Schmidt orthogonalization. Important gradient methods are the Levenberg-Marquardt algorithm, quasi-Newton methods, and conjugate gradient methods. Gradient methods usually employ a quadratic approximation of the objective function and estimate the location of the minimum. The Levenberg-Marquardt algorithm simply performs a line search along the gradient direction, modified by an estimate for the Hessian. A weighting factor adjusts the estimate according to the success of the last optimization step. Quasi-Newton methods also estimate the Hessian, but use the results of more previous optimization steps. Normally the Broyden-Fletcher-Goldfarb-Shanno update procedure is used. Convergence of quasi-Newton methods is less dependent on the optimization problem than of Levenberg-Marquardt algorithms. Hybrid methods were therefore developed that combine the advantages of both methods. Conjugate gradient algorithms use conjugate directions for updates in subsequent steps. They combine the ideas of conjugate search methods and quasi-Newton methods. Their performance is found to be superior for large scale problems. A special method for global optimization are Branch and Bound algorithms. These algorithms rely on subdividing the parameter space into intervals and run a local optimizer in those intervals. An overview on deterministic optimization methods is given, e.g., in [18].

Stochastic optimization methods rely on knowledge gained from randomized tests on the objective function. The methods fall in one of several classes:

- Bayesian Methods
- Simulated Annealing
- Evolutionary Methods

Bayesian Methods use stochastic measures like mean values and correlations. Conditional probabilities decide on new samples. Simulated Annealing relies on thermodynamic models of material cool-down. The algorithm starts from a highly excited material phase and reduces the temperature continuously, until a final state is reached. Evolutionary methods draw their principles from biology. Several methods are known in this field.

- Evolutionary Programming
- Evolution Strategies
- Genetic Algorithms

All of these methods work on a number of parameter-sets, and employ a mutation operator and a selection mechanism. The actual implementation of the operators distinguishes the different methods. Evolutionary Programming uses several mutation operators and tournament selection on the generated parameter-sets. Evolution Strategies employ a normally

distributed mutation operator. The selection is implemented as a tournament between the new and the old parameter string. Genetic algorithms usually operate on bit-strings and employ an operator, known as crossover, which is modeled after the genetic crossover of chromosomes. The next chapter will discuss this method and its specific vocabulary in more detail.

The inherent demand of approaching the diverse issues limits the choice of possible algorithms. By definition, the optimization has to succeed on incomplete formulations, as too stringent restrictions would hamper the invention of new solutions in the conceptual design phase. Since problems are multimodal, multiple local solutions have to be generated. Moreover, it is crucial to be able to proceed in a mixed continuous and discrete design space.

Considering the multiplicity of optimization algorithms, several methods have to be discarded a priori:

- Gradient methods are inapplicable because of missing derivatives. Pseudo-gradient techniques do not help as problems are discrete and therefore non-continuous.
- Other local optimization algorithms are applicable in combination with random start values. The optimization however is non-effective as thousands of start values have to be run over at least one hundred optimization steps. The issue of non-continuity is to be taken into account as well.
- Global deterministic optimization methods such as, for instance, Branch and Bound, rely on assumptions on or knowledge of the optimization problem. This information is often not available and the algorithm is therefore not robust. The convergence into a single minimum is inappropriate.
- Stochastic methods as simulated annealing, evolution strategies and genetic algorithms, usually converge into a single global solution. They are not applicable as such.

Stochastic methods, however, seem to be most favourable for robust global optimization in complex, non-continuous function spaces. The adaption of genetic algorithms to multimodal problems, in particular, is shown to be straightforward and is therefore discussed in the next chapter.

Chapter 5

Genetic Algorithms

The first steps to the development of genetic algorithms were set in the 1950s when models for biological processes were sought. Although at that time not apt for function optimization, the potential of adaption was recognized and evaluated. Driven by these results, Holland put the models on a firm basis [19]. Since then the research on those procedures, called genetic algorithms, has greatly expanded [20, 21]. Increasing computing capabilities allow for complex problem solving and global optimization in multidimensional function spaces.

In this chapter, genetic algorithms for the solving of real world optimization problems are introduced and the basic operators and mechanisms are discussed. The standard implementation is reviewed and its drawbacks for technical design problems is outlined. Operators which are adapted to the special requirements in technical optimization are described and a feasible method is assembled. Especially the principle of niching is introduced in order to evolve multiple solutions concurrently. Comparisons show the power of this approach. Since research on the underlying principles and their mathematical foundations is ongoing, mere indications for the optimization mechanisms conclude the chapter.

5.1 Introduction to Genetic Algorithms

Although genetic algorithms were originally designed for the simulation of biological processes, their power of adaption has soon been recognized as a potential optimization mechanism. In cases where standard optimization algorithms such as gradient methods ceased to perform acceptably, genetic algorithms were identified as powerful enough to cope with the complexity involved.

NP (nondeterministic polynomial) complex problems were successfully treated, for which no algorithm is known to solve them in polynomial time. However, by definition solutions to problems of class *NP* may be verified by a polynomial algorithm. Polynomial in this sense means that for any of n possible instances of the algorithm at most $p(n)$ algorithmic steps are necessary to find the solution, where $p(n)$ is a polynomial of n [22]. This $p(n)$ steps may be distributed in time and space. Assuming that only one step at each moment can be executed, the number of time steps is non-polynomial. Despite the success of genetic algorithm, its convergence in finite time cannot be guaranteed, as will be discussed in a later

section.

Since each instance of an *NP* problem can be verified in polynomial time, the solution could be found in polynomial time too, if all n possible instances were tested in parallel. But as there is no such thing as a free lunch, in such a case the resulting space complexity would become non-polynomial. This consideration, nevertheless, leads to the idea of parallelizing the optimization problem by evolving several parameter-sets concurrently. Consequently statistical measures may be employed on the parameters and probability considerations may guide the evolution. Such techniques are known by the notion of stochastic optimization.

Among such methods are Monte Carlo algorithms, simulated annealing, Bayesian methods, evolution strategies, and genetic algorithms. The main reasons for the choice of genetic algorithms for an optimization problem are their ability to deal with multimodal objective functions and to work in discrete as well as continuous parameter spaces. Their robustness in the choice of optimization parameters is an additional advantage. Necessary enhancements for the proper treatment of multiobjective cases and adaption to engineering problems are shown below.

The optimization examples discussed in the following chapters will emphasize the correct formulation of an optimization problem, which usually turns out to be more difficult than the proper choice of the optimization algorithm. It is clear that more complex algorithms exist which are better optimizers for a special problem, if certain properties of the objective function are exploited. Wrong problem formulation, however, might lead to convergence to useless results. The thesis is therefore not going into detailed comparisons of different algorithms, but tries to comment on the decisions that were taken to provide the necessary robustness. More time is spent on the optimization examples as this was identified as the key factor in the success of stochastic optimization methods.

At this point the computational cost should be considered. The time spent in the optimization program code is particularly low. The underlying concepts are surprisingly simple though effective. Especially for the examples treated, the overall calculation time was almost solely determined by the number of function evaluation and its average time. Although the convergence rate should be kept low, in many cases decisions were taken in favour of robustness. A trade-off between convergence and quality of the results has to be considered, as will later be discussed.

5.2 Nomenclature in Evolutionary Computation

Because the development of genetic algorithms has its roots in research into biological evolution, some biological vocabulary was adopted for algorithm design. Parameter-sets of a vector optimization problem are encoded as *chromosomes*, with the information carrying parameters represented by *genes*. It is important to distinguish between the *phenotype*, which are the parameters themselves and the *genotype*, their encoded representation. Since the genotype and the phenotype carry the same information, references to them are often imprecise. In this sense also the terms *individual*, *string* and *structure* are used for the genotypic representation.

Individuals are organized in a *population* which is a kind of data-base with respect to

the parameter-sets. In a standard implementation, this population is iterated generation by generation employing the operators *crossover* and *mutation*, reminiscent of processes which occur in the bio-genetic reproduction. The *environment* in which the population of chromosomes lives is evaluated by the design simulation. A *fitness* is generated from the resulting objective function value, defining the quality of the design by a probability for *reproduction* in the next generation.

The standard encoding as used in the actual implementation of this thesis is operating on the bit-level. Bits, referred to as *alleles* in biology, interfere in the simulation. A non-linear interdependence between different bits is comparable to *epistasis* in genetics. Further equivalents can be drawn from biology. Special research in bio-dynamics tried to model nature's mechanisms precisely. The artificial implementation of genetics, however, is not the aim of applied genetic algorithm research. The number of generations and population sizes in biological evolution alone exceeds any useful measures in complex function optimization. Therefore any similarities to natural processes remain vague.

Genetics	Algorithms
Chromosome	Parameter-set
Gene	Parameter
Allele	Bit
Genotype	Bit encoding
Phenotype	Design
Epistasis	Nonlinearity
Population	Data Base
Generation	Iteration

Table 5.1: Nomenclature of genetic algorithms and their technical implementation

5.3 Royal Road Genetic Algorithm

In order to gain insight into the optimization processes going on in genetic algorithms, the standard implementation referred to as Royal Road Genetic Algorithm is discussed. The advantage of this setup is that the resulting reproduction probabilities can be estimated. A population of chromosomes is iteratively manipulated by operators as shown in figure 5.1. Out of a population or gene pool, chromosomes are taken in order to create new offspring by single-point crossover and mutation to be discussed later.

The chromosomes are evaluated by means of an objective function. Each parameter-set is assigned a fitness value according to its objective function. This fitness value then guides the selection process defining the new population for the next iteration. The initial population of chromosomes is produced by an additional generation operator. From there the algorithm proceeds for a number of iterations usually in the thousands. The optimization is stopped either by a criterium, testing the objective function, or after a certain iteration is reached. The results are found in the decoded parameter-sets of the population ranked by the fitness function.

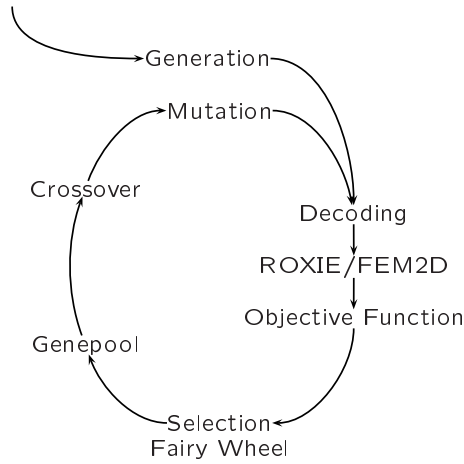


Figure 5.1: Structure of the Royal Road Genetic Algorithm

Below all stages of the algorithm are discussed in more detail outlining the dynamics of the procedure. Several adaptations to the standard method were found to be necessary in order to make the procedure as robust as possible.

5.4 Parameter Encoding and Quantization

Since standard genetic algorithms work on the bit-level, an encoding for the parameters is necessary. Real variables have to be quantized into integers. Then all integers are Gray-encoded to provide for better behaviour of the genetic algorithm. Put on this basis, genetic algorithms are not longer concerned with the properties of the parameters anymore and can directly operate on their encoding as bit-strings.



Figure 5.2: Parameter encoding and decoding

Quantization is achieved by linear sampling of the parameter range. The parameter mapping is defined by giving lower x_{kl} and upper bounds x_{ku} , as well as the number of bits of the integer encoding n .

$$x_k = x_{kl} + \frac{x_{ku} - x_{kl}}{2^n - 1} i_k \quad (5.1)$$

The k^{th} parameter x_k corresponds to the integer value i_k . In the current approach the full integer range of n bits with 2^n values is used. The quantization has to be fine enough to limit the quantization errors. The chromosomes are then found by Gray-encoding these integers and packing them together into a bit-string where the least number of necessary bits is used.

5.5 Gray-coding

Although the integers as found by the quantization could immediately be used as a parameter representation, a further coding stage was introduced to improve the quality of the optimization process. According to the following table (shown in fig. 5.3 for only 4 bits) each bit combination is uniquely transformed into another. This scheme is called Gray code, after Frank Gray, who patented its use for shaft encoders in 1953 [23].

	binary		Gray		binary		Gray
0	0000	→	0000	8	1000	→	1100
1	0001		0001	9	1001		1101
2	0010		0011	10	1010		1111
3	0011		0010	11	1011		1110
4	0100		0110	12	1100		1010
5	0101		0111	13	1101		1011
6	0110		0101	14	1110		1001
7	0111		0100	15	1111		1000

Figure 5.3: Gray encoding table.

The coding is achieved by a bit-shift right and exclusive-or \oplus of the unshifted and the shifted version of the binary code. The binary string is truncated on the right and padded with zeros on the left. Decoding proceeds by multiply shifting right and exclusive-or operations on the shifted versions as shown in figure 5.4. Below the term Hamming-distance is

Example: decimal 13:	encoding		decoding
	binary 1101		Gray 1011
			0101 \oplus
	<u>0110</u> \oplus		0010 \oplus
	Gray 1011		<u>0001</u> \oplus
			binary 1101

Figure 5.4: Algorithm for Gray encoding and decoding

used which is the number of different bits of two bit-strings. The Hamming distance can be written formally as $HD = \sum_i u_i \oplus v_i$, a sum over all i bits generated by an exclusive-or of two bit-strings, where the sign \oplus again denotes the exclusive-or operator.

The advantage of Gray-codes stems from the fact that chromosomes close to each other in parameter space (having a small Hamming-distance) show similar patterns with higher

probability. For example the codes of adjacent decimals 3 and 4, 011 and 100 in a binary representation have a Hamming-distance 3, but 010 and 110 as Gray-encoded show a Hamming-distance of only 1. This means that mutation produces nearby strings with higher probability.

In Gray-codes both strings that are close to another code in parameter space, are also close in their bit representation. In binary codes, by contrast, only one of the bit representations of the strings is close. Consequently, the percentage of close strings after a one bit change is $2/n$ for Gray-codes dependent on the overall number of bits n compared to $1/n$ for binary codes. Therefore the influence of Gray coding becomes smaller for longer bit-strings. All the rest of bit changes however is further away, as for example 0000 and 1000 codes 0 and 8 in binary but 0 and 15 in Gray-codes. For single bit parameters, Gray codes are not different from binary codes and the notion of coding is rendered void.

The question of whether Gray codes are advantageous to the optimization process is dependent on the objective function to be optimized. It was shown that the optimization is improved for a number of continuous functions, e.g. De Jong's test functions F1 or F2 [24]. The improvement naturally becomes smaller with longer bit-strings. Gray-coding usually is helpful, in particular, when the hamming distance is used as a distance measure in the genetic algorithms. This is the case for niching methods that are discussed later.

5.6 Genetic Operators

Genetic algorithms are driven by 3 main operators: crossover, mutation, and selection. The crossover operator, in particular, is unique to genetic algorithms. It is thought to be the driving force in the optimization. The mutation operator helps in spreading the search and overcoming local minima. Finally the selection operator guarantees convergence to an optimum by keeping the better chromosomes and discarding the less fit ones. Looping through these operators some thousand times, diversity is reduced iteration by iteration thus leading to a single solution. An additional operator, generation, is used for initialization. In the following sections, the operators are reviewed in detail.

5.6.1 The Generation Operator

Generation is the simplest operator. It randomly initializes a bit-string. The operator is independent of any existing population. This operator is usually used for the creation of the start population. It can also be applied, if the randomness in the population should be increased. The rate of generated bit-strings, however, is usually kept low.

5.6.2 The Crossover Operator

Crossover is a recombination of bit strings of two chromosomes by swapping the strings at random points. It is the major force in genetic algorithms directing the search process to good solutions. A recombination of good parameters of any chromosome with good parameters of another should lead to a better set [19]. The type of crossover which is usually considered

is single-point crossover. Two chromosomes are taken out of the population. A certain splitting point is determined. Each chromosome is cut at this point and the first parts of each chromosome is recombined with the second part of the other chromosome.

$$\begin{array}{ccccccc}
 chr1 : 1101010010 & | & 0101101 & & 1101010010 & 1011010 & & 1101010010 & | & 1011010 \\
 & & \Rightarrow & & \times & & \Rightarrow & & & \\
 chr2 : 1010110100 & | & 1011010 & & 1010110100 & 0101101 & & 1010110100 & | & 0101101
 \end{array}$$

Figure 5.5: Single-point crossover of two chromosomes. The chromosomes are cut in a random position and the parts are recombined cross-wise with each other.

As shown in figure 5.5, the bits on the left-hand side of chromosome 1 are connected to the bits on the right-hand side of chromosome 2 and vice versa. This exchange results in two new chromosomes which resemble two parameter-sets that are a combination of two different earlier parameter-sets. Taking a closer look at the effects of the operator on the parameter-sets itself reveals that cutting the chromosomes introduces some randomness regarding the very parameter that may be split by the operation. This effect is owing to the bit representation of the parameter-sets. Implementation on a parameter level can avoid this random behaviour, but since randomness is introduced by other operators (especially mutation) on purpose, there is not much reason to do so.

Illustrating single point crossover in a pseudo 4-dimensional lattice (fig. 5.6) shows that such an operator produces offspring on one of two trajectories. It is therefore clear that crossover is not a completely random operation but a search operator in a sub-space of the parameter domain. Some genetic algorithm implementations favour two point crossover which covers additional points (here: 0000 and 1011). Nevertheless the larger space of possible offspring due to 2 crossover points can be equally produced by two successive iterations using single point crossover.

Improvement of the Crossover Operator

Although there are two possible offsprings that a crossover might have, in many implementations only one of the chromosomes is kept. In certain cases this might be a shortcoming of the algorithm, since the two offsprings are not independent of each other. To clarify this issue let us assume independency of the parameters and allow splittings only on the parameter level. The fitness f of a parameter-set ps may then be calculated by combination of the fitness of the two parts of the parameter-set:

$$f(ps'_1) \star f(ps'_2) = f' \quad (5.2)$$

$$f(ps''_1) \star f(ps''_2) = f'' \quad (5.3)$$

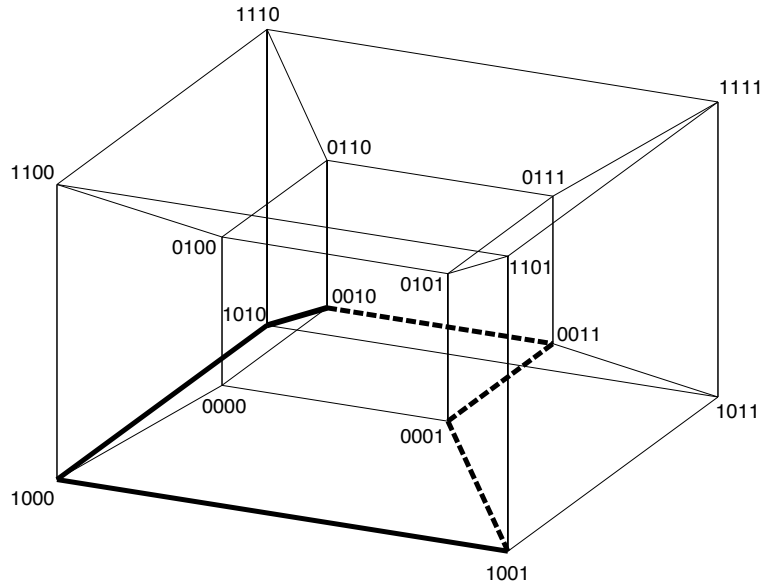


Figure 5.6: Hypercube with transition patterns for single point crossover of 0010 and 1001. The possible offspring is found in the subspace of the two trajectories. Crossover is therefore not a completely random process.

where \star is a monotonic operator. A crossover which is nothing else than an exchange of parameters results in two offsprings with a certain fitness:

$$f(ps'_1) \star f(ps''_2) = f'_n \quad (5.4)$$

$$f(ps''_1) \star f(ps'_2) = f''_n \quad (5.5)$$

For one offspring, the fitness will be worse than any of the parents $f'_n < f'$ and $f'_n < f''$ with a certain probability. With the assumptions of independence the fitness of the second offspring must be better than its parents $f''_n > f'$ and $f''_n > f''$. Discarding the second offspring therefore means neglecting a safe hit for a better fitness. But real-world problems are not as simple as the aforementioned model and complete independency is normally not present. A good test is therefore to measure the probability of a safe hit for the second offspring in a test case depending on the quality of the first offspring.

For a standard 2d optimization problem as described in Chapter 6 the objective function was calculated for a test set of 7000 pairs of parents. It was found that the conditional probability $P(f''_n > f', f''_n > f'' | f'_n < f', f'_n < f'') = 0.426$ is considerably higher than the absolute probability $P(f''_n > f', f''_n > f'') = 0.276$. One may therefore conclude that there is an advantage for a starting population to generate the second offspring when the fitness of the first was worse than the fitness of its parents. This advantage vanishes, however, for parents close to an optimum. There the probability for better offspring approaches zero.

5.6.3 The Mutation Operator

After crossover, the offspring is mutated. In the standard implementation, single bit mutation is used. Each time the operator is applied, a single bit is changed from 0 to 1 or 1 to 0.

The effect of the mutation operator is twofold. First it avoids preliminary convergence of the entire population towards a local minimum and secondly it improves local convergence by a hill-climbing-like mechanism. Although both mechanisms seem to be contradictory, they result from the different significance of bits in the bit-string as discussed for Gray coding above. Depending on the parameter encoding, the effect may result in small or big changes

	10	7	2	1	1	0	Phenotype
	1111	0100	11	01	1	0	Genotype
a)	10	8	2	1	1	0	Phenotype
	1111	1100	11	01	1	0	Genotype
b)	13	7	2	1	1	0	Phenotype
	1011	0100	11	01	1	0	Genotype

Figure 5.7: Influence of mutation depending on the encoding. Single bit changes have different influence on the phenotype. a) The mutation of bit 5 results in a phenotype difference of 1, whereas b) The mutation of bit 2 leads to a change in the phenotype by 3.

of a single parameter. As shown in figure 5.7 a change of the fifth bit (a) results in a one step change of the second parameter from 7 to 8 whereas a change in the second bit (b) results in a change of the first parameter from 10 to 13.

Small changes may help locally optimize the objective function whereas big effects help in the global search. While crossover introduces a kind of reasoning, mutation mainly enhances the random behaviour and is normally kept low. The investigation of the search space solely by mutation is slow, since bit changes in all bits are needed to cover the complete space. However the whole search space may be reached.

5.6.4 The Selection Operator

Selection is the most important operator in that it has to ensure convergence. Its algorithmic implementations vary in a wide range. Convergence alone, however, must not be seen as separate from the quality of the solution reached. The specific method used influences the robustness of the search as well.

Selection methods can be divided into two principally different groups:

- **Fitness Proportionate Selection:**

Chromosomes get selected proportionally to a fitness value derived from the objective function. The choice of fitness function is problem dependent.

- **Tournament and Rank Selection:**

The relationship of objective function values determines the survival of chromosomes. No fitness function is needed.

In the following sections, the methods of fair wheel selection and rank selection are reviewed, and advantages and disadvantages are discussed.

Fitness Proportionate Selection

In fitness proportionate selection, also known as fair wheel selection, the selection operator chooses chromosomes proportional to their assigned fitness value. The fitness value therefore represents a probability density. The mapping of a fitness value $f(\vec{x}_i)$ to an objective function value $F(\vec{x}_i)$ is achieved by means of a fitness function $f(F)$. A standard procedure for fitness evaluation is normalizing the objective function values by their sum over all individual parameter-sets \vec{x}_j .

$$f(\vec{x}_i) = \frac{F(\vec{x}_i)}{\sum_j F(\vec{x}_j)} \quad (5.6)$$

Consequently a cumulative distribution function can be derived from this fitness function. This distribution function may now be used for a selection of chromosomes transforming an equally distributed random variable into a fitness proportional selection for the operations of the next generation.

Since fitness values are derived from the actual value of the objective function, and these objective functions may differ in their characteristics, the scaling of fitness functions has to be adjusted to the objective function under investigation. This is a major disadvantage, since the user has to know about the dynamics of genetic algorithms and the objective function to be adapted. Last but not least it must be noted that with the function (5.6) the objective function F is maximized and negative objective values may not appear. Ranking may give better results for objective functions with several orders of magnitude or at least unknown limits.

Tournament and Rank Selection

In order to avoid fitness functions, and to simplify the selection process, ranking methods were developed, the major methods of which are

- Rank Selection:
The chromosomes are sorted by their objective function and a fitness is assigned according to their rank.
- Tournament Selection:
A group of individuals is randomly picked out of the population and the chromosome with the best objective function value is selected.
- Elitist Selection:
Additionally to any of the other selection methods a copy of the best individual is kept, in order not to lose the best solution found so far.

Since the objective function values are only used for sorting the chromosomes, no fitness function is needed to map the objective function into a probability distribution. This prevents scaling problems and allows better adjustment of the convergence behaviour. On the other hand, valuable information from the actual objective function values is ignored. A

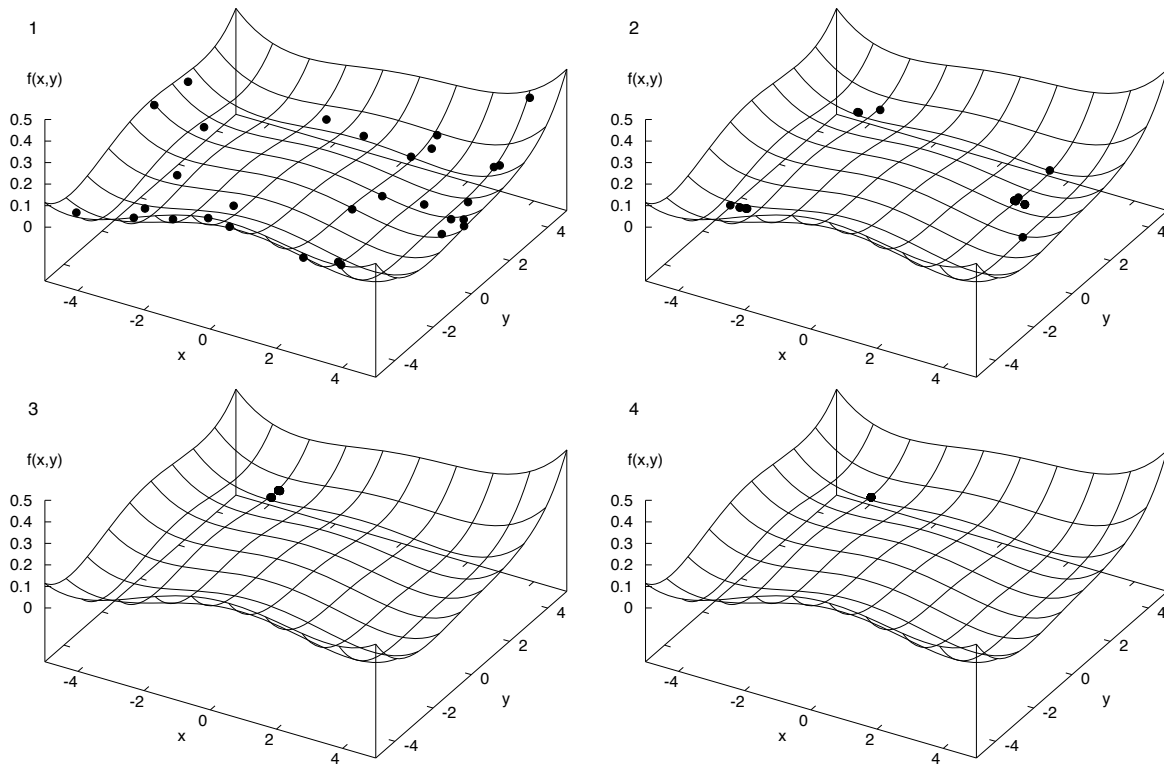


Figure 5.8: Genetic algorithm with crossover and fitness ranking only. The development of a population of 30 individuals is shown over 500 function evaluations. All of the 30 individuals degenerate into a single stable solution.

comparison between different mechanisms, however, reveals that tournament selection is not inferior to fitness proportionate selection, giving a good compromise between quality and speed of convergence [25].

Limitations of Standard Selection Operators

Theoretical investigations often assume an infinite population size and an infinite number of function evaluations, whereas practical applications have to deal with both limited population sizes and number of iterations. As a result, optima are frequently missed because of premature convergence. The cause of this process is identified as genetic drift, also known from biology. A major aspect of genetic drift is the reduction of diversity due to the random propagation of alleles, or bits, from parents to their offspring. The search space of the crossover operator as shown in figure 5.6 is reduced by every lost bit and can only be regained randomly by mutation. In cases where a better solution is missed accidentally, the genetic algorithm does not converge into the global optimum anymore.

A further problem are multimodal objective functions where multiple solutions cannot be found simultaneously with standard genetic algorithms. Figure 5.8 shows the development of a population of 30 individuals over 500 function evaluations using a genetic algorithm with crossover and fitness ranking. The test function to be optimized is known as the Himmelblau

function [26].

$$f_1(x, y) = \frac{(x^2 + y - 11)^2 + (x + y^2 - 7)^2}{2186} \quad (5.7)$$

The function has 4 global minima of function value $f_1 = 0$ at the positions (3, 2), (3.58443, -1.84813), (-2.80512, 3.13131), and (-3.77931, -3.28319). The optimization converges relatively quickly to only one of the optima. Knowledge about the function close to other solutions is lost in the course of optimization. All the individuals drift into a single solution.

Although most shortcomings of deterministic algorithms for the global function optimization are overcome, standard ranking genetic algorithms still need improvement regarding genetic drift. Viewed from a magnet designer's standpoint, there are important reasons to ask for more than one optimum in the search space, even though they might not be global optimizers:

- Because of the different sensitivity of the multiple objectives in the different sub-domains of the feasible domain the solution might not represent the designer's preference.
- Not all of the constraints of the conceptual design problem can be taken into account in a numerical optimization run either because of their complexity or their unknown influence at the time of optimization.

In order to overcome these problems niching methods are introduced in the following.

5.7 Niching Methods

The fact that many real world optimization problems are multimodal, when all dependencies are taken into account, has recently produced considerable interest in niching methods. While a standard genetic algorithm aims at finding a single global optimum, the application of the concept of niching provides the designer with a set of solutions rather than only one solution. The main advantage is that these solutions can then be "post-processed" by the designer incorporating additional objectives and constraints, and that good solutions can be selected from this set.

The purpose of niching on the algorithmic level is to categorize solutions by their design variables into groups, known as niches, and to evolve them in the optimization run. In order to do so, a measure discriminating between such groups is needed. Imagining the design variables in their multidimensional optimization space, niching methods employ a distance function $d(x_i, x_j)$ either using the phenotype with the parameter values, or the genotype on the parameter encoding level. Close and therefore related parameter-sets are distinguished by a low distance and as such establish the niche.

Several niching methods were developed with many variations. The algorithms which are employed most frequently are:

- Fitness Sharing:
The fitness value of each individual is modified by the distance to other individuals. Nearby parameter-sets reduce the fitness.

- Crowding Method:

Replacement of old individuals with new offspring favours chromosomes which are close in parameter or code space. A tournament between close solutions decides about the survival of the individuals.

A short review of fitness sharing is followed by a description of the actual implementation of a crowding method.

5.7.1 Fitness Sharing

A standard method of niching is fitness sharing as introduced in [20]. A sharing function $s(d)$ is defined which determines the neighbourhood for each chromosome in the population. The simplest type of sharing functions assigns one to chromosomes which are close in parameter space and zero to individuals which are further away than a certain dissimilarity threshold. More complicated versions define a continuous monotonic function from one to zero according to the distance. Accumulating the total share, a new fitness value is defined as

$$f_s(x_i) = \frac{f(x_i)}{\sum_{j=1}^n s(d(x_i, x_j))} \quad (5.8)$$

The method therefore fits nicely with the standard genetic algorithm approach of fitness proportional selection. The fitness value of individuals with a high number of neighbours is derated, so that highly populated niches have a lower probability of reproduction.

5.7.2 Crowding Method

Crowding methods employ a replacement mechanism for the offspring generated from individuals of a population. Several principles are used. The new offspring may either replace one of its parents, or another chromosome in the population, that is closest in parameter space. A very important aspect is the interference with the selection operator used. In first implementations of crowding, fitness proportional selection was used [27]. Ranking methods, however, lend themselves better to an implementation. This is comprehensible, as selection is only applied to the few individuals which are close to each other.

Mahfoud [28] compares several different methods of crowding and develops two criteria for algorithm performance. The criteria are the number of peaks an algorithm maintains and the number of replacement errors it makes. A peak is accepted to be maintained, if a solution is found around the peak within 0.2 of its height. A replacement error is defined by a case where a replacement of a parent by an offspring leads to a change in the peak. It is further shown that both criteria are closely related and that the improvement of the first is achieved by a reduction of the second. In the course of his tests he develops a method, termed deterministic crowding. In this crowding method, two parents are selected from the population without replacement. Crossover and mutation are applied, giving new offspring. Each of the two offspring enters into a tournament with the closer parent. Depending on the fitness found, the parents are replaced by the offspring. The replacement error is shown to

be relatively low. For genetic algorithms, working on bit-strings, Gray encoding may further reduce the replacement errors. A special advantage of the algorithm is its low algorithmic complexity.

5.8 The Implementation

In the implementation chosen in this thesis a crowding mechanism is used. Algorithmic complexity is not important at all, since the evaluation of the objective function takes orders of magnitude more time than the execution of the genetic algorithm. The algorithm therefore uses the simpler scheme of replacing the closest individual in the population in case of better fitness. This principle should in fact further reduce the number of replacement errors. Replacement errors, however, might also be favourable whenever the newly found solution falls on a newly discovered peak or niche. Important for overall execution time is that offspring which already exists in the population, is not evaluated for its fitness again.

Some compromises had to be found, to make the routines more robust for everyday use. Robustness is achieved by the following considerations:

- Since the method is global, parameter ranges can be wide. No special knowledge is needed about the region of interest. Multiple solutions are found, from which the designer may choose.
- No derivatives are needed. Non-continuous problems can easily be tackled. Integer variables can be encoded as such.
- Operations deal with chromosomes on the bit level only. The algorithm therefore does not have to check for the encoding of the parameters. Distances, however, can only be calculated as Hamming-distance and not on the parameter level. Gray-coding corrects this situation for close strings.
- Only few parameters have to be chosen. Most of them are non-sensitive. Only the population size and the number of bits for the encoding is critical, but may be adjusted automatically. Due to the crowding method with tournament selection, neither a dissimilarity threshold, nor a fitness function, nor crowding factors have to be adjusted.

Still, it has to be stated that global optimization algorithms are normally bad local optimization methods. Therefore a first stage of global optimization typically has to be followed by local optimization. The reasons are due partly to the stochastic nature of the algorithms, partly to the parameter encoding. In addition, the parameters used in the goal finding method for global optimization are usually different from local optimization parameters.

5.8.1 The Algorithm

The implementation of the genetic algorithm follows the modified pattern as shown in figure 5.9. The principle is displayed in more detail in figure 5.10. The algorithm starts with

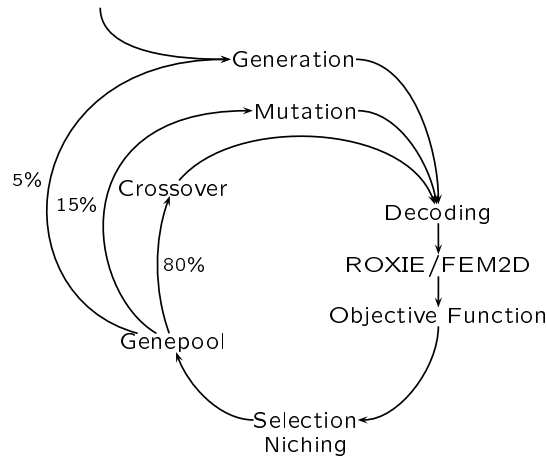


Figure 5.9: Structure of the the actual implementation of the genetic algorithm with niching.

a population of randomly generated chromosomes. An operator is selected which is subsequently applied to chromosomes of the population creating a new offspring. After a new offspring is generated by one of the operators, the chromosome with the smallest Hamming-distance is located and replaced, if its fitness is worse than that of the offspring.

To increase the diversity in the beginning of the optimization, a fourth operator is added, i.e. generating new chromosomes in a random process. Normally this operator is only used for initialization of the genetic algorithm with random chromosomes. After each application of an operator, the offspring is evaluated and selected strings are introduced into the chromosome pool for immediate participation. Distinct generations no longer exist. Although this method might seem unorthodox in the genetic algorithm design (tab. 5.1), it resembles nature more closely since mating is possible beyond generations. Every new chromosome profits from the improvements by the latest operations. A parallelization of the process, however, could easily be introduced.

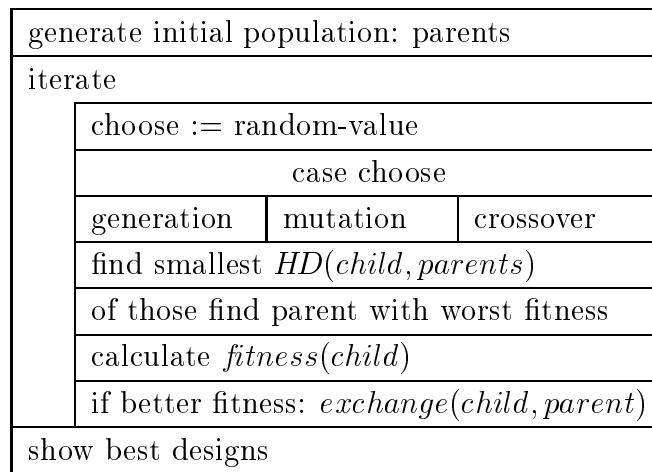


Figure 5.10: Structogram of the actual implementation of the genetic algorithm with niching

With the aim of getting an idea of how the niching genetic algorithm works, the optimization problem presented in figure 5.8 is solved with niching selection as shown in figure 5.11. Although the optimization needs considerably more time, it converges into a number of optima concurrently, instead of only a single solution without niching. All four minima are found in about 1000 function evaluations.

A second example using the same algorithm, is shown in figure 5.12. The function

$$f_2(x, y) = \frac{\cos(3x + 2y) + \cos(3y + 2x)}{4} \quad (5.9)$$

has 14 minima inside the design region at regular intervals. In the example run, 12 out of the 14 minima of function value $f_2 = -0.5$ are found. In both examples the discretization is 10 bits for each variable.

5.8.2 The Parameters

As thoroughly discussed by Michalewicz, higher diversity as achieved by niching decreases the selective pressure [21]. As too high selective pressure may lead to premature convergence, weak selective pressure may lead to an ineffective search process. Therefore the main parameters, number of iterations and population size, still have to be chosen properly. Systematic investigations were done by Mahfoud [29], which suggest that the number of individuals in a population should be kept at around ten times the number of peaks. This result, however, cannot be applied to real-world design problems directly, where the number of solutions is not known a priori.

In the examples shown in the next chapters, the population sizes and number of iterations were therefore found heuristically. Mahfoud's work suggests that the number of different solutions or niches desired, should be proportional to the number of individuals in the population. Since the discretization of the design parameters should be adapted to the influence of the parameters on the objective function, and consequently to the number of optima, a proportionality to the number of bits in the chromosomes should exist as well. In the current optimizations a factor between 1 and 3 was effective.

In order to stop the algorithm at a certain point a termination criterion is necessary. As long as no technically feasible optimal solution is known, a utopian solution gained by analytical reasoning is helpful. The optimization algorithm might be halted when the solutions have come acceptably close to an optimum and convergence becomes slow. The number of function evaluations or iterations was chosen in a range of 100 to 200. As discussed before, the global optimization phase should be followed by a local optimization run for each niche solution. Assuming that this is done, the optimization can be stopped when niches become apparent, even though the solutions could still have been improved by the genetic algorithm.

The parameters of crossover, mutation, and generation rate, were chosen after a number of tests on the coil optimization problem of chapter 6. The population size was adjusted to give good results with acceptable convergence. A representative convergence graph is shown in figure 5.13a. After increasing mutation to 60% as shown in figure 5.13b, the convergence is lower than for our standard genetic algorithm in the beginning. After a

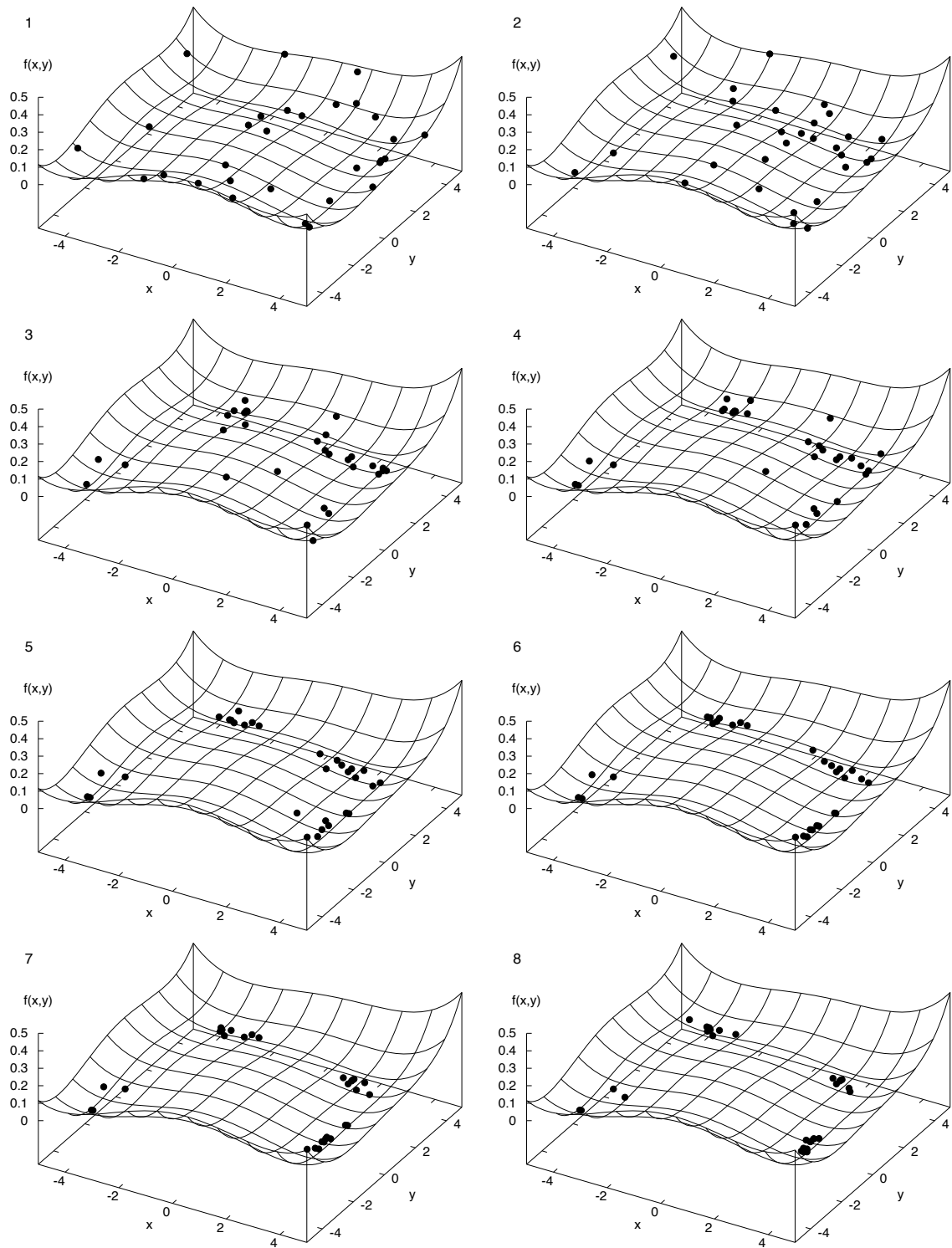


Figure 5.11: Development of a niching genetic algorithm with crossover and crowding method over 1000 function evaluations. All four minima are found. Niching is therefore crucial for the conceptual design phase.

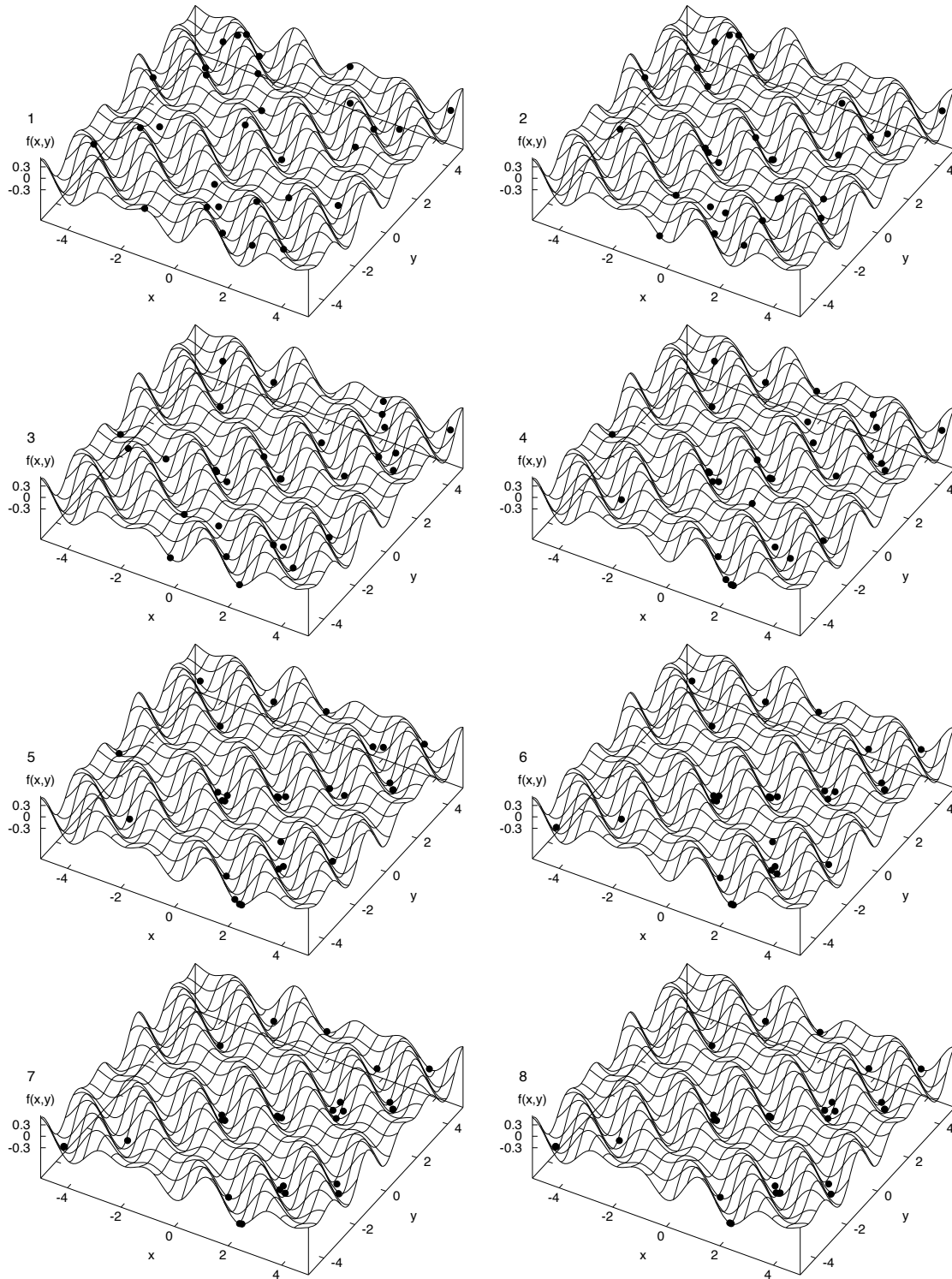


Figure 5.12: Development of a niching genetic algorithm with crossover and crowding method over 1000 function evaluations. The function has 14 minima in the optimization region of which 12 are found. In practical applications, the designer has to further evaluate the solutions.

few thousand iterations when the population becomes stable, higher mutation rate allows for further improvement since changes of single bits often correspond to small changes in parameters. Results for single and 2-point crossover show no significant differences.

In a further test, the population size was changed to 150 individuals, thereby increasing the diversity in the chromosome pool from which offspring is created (fig. 5.13c). The number of necessary function evaluations is increased, resulting in better refined final solutions. Although higher quality is achieved, there is no real advantage, since in both cases each solution has to be further optimized by local methods.

5.9 Methods of Description

Several notable studies were carried out on the behaviour of genetic algorithms. A number of ideas were developed to explain the dynamics, but only lately rigorous proofs on convergence of simpler implementations and for special cases were conducted using Markov chains. The following sections may serve to explain why the subject is complicated, and no proper model for niching genetic algorithms exists. The robustness of many genetic schemes, however, make the application feasible, as will be shown in later chapters.

5.9.1 The Schema Theorem

Holland developed the idea of schemata being processed by genetic algorithms. The fact that adaptive processes may evaluate schemata in parallel led to the notion of implicit parallelism, which will be explained later [19]. Estimating the probability of reproduction, he developed the schema theorem which is explained below as rephrased by Goldberg [20].

A subset of individual strings is described by a schema consisting of strings composed by a three-letter alphabet $A = \{0, 1, \star\}$ where 0 and 1 denote fixed bits, whereas the asterisk \star is a wild card symbol matching 0 or 1. The two important properties of shemata are order and defining length. The order of a schema is the number of fixed positions in the schema string. The order of string S_1 and S_2 is therefore 4 and 2 respectively. The defining length is the distance from the first to the last fixed position in the string. For our examples 5 and 1 is found respectively.

Supposing that at a given iteration t , the population consists of $m = m(H, t)$ strings incorporating a particular schema H . In the fairy wheel selection, a string is selected with probability $p_i = f_i / \sum f_j$. For a population of n individuals this means that the population at the iteration $t + 1$ consists of

$$m(H, t + 1) = m(H, t) n \frac{f(H)}{\sum_j f_j} = m(H, t) \frac{f(H)}{\bar{f}} \quad (5.10)$$

samples. All schemata thereby grow or decay in the population proportional to their fitness.

Since the reproduction does not only proceed by selection but also by crossover and mutation, this theorem has to be modified by a probability for the reproduction $p_o(H)$ due

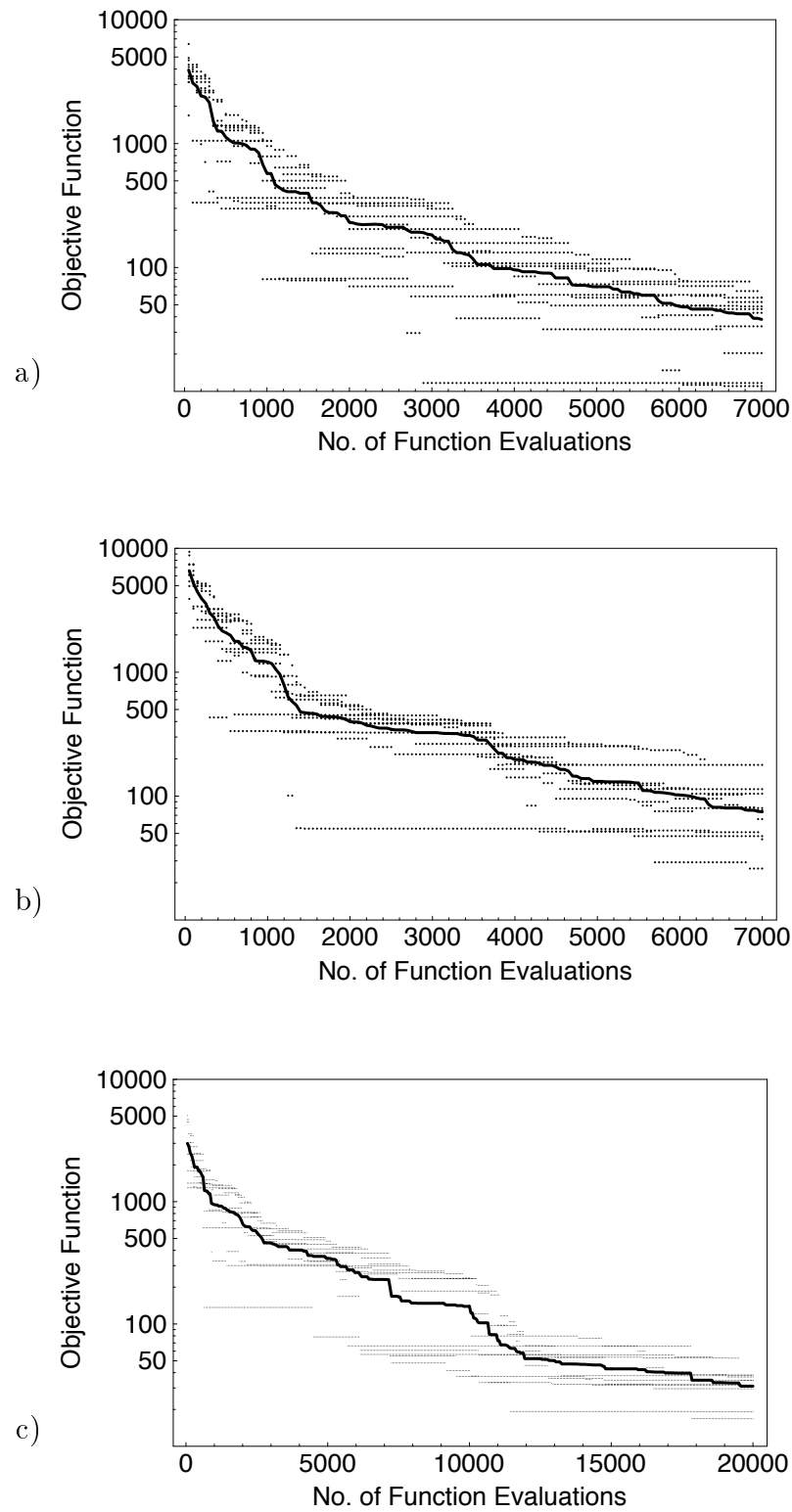


Figure 5.13: Convergence test on the coil optimization example. Objective function of the 10 best chromosomes and its average (continuous line), a) Convergence for standard parameters: crossover rate 0.8, mutation rate 0.15, generation rate 0.05, b) Convergence for mutation rate 0.60, crossover rate 0.35 c) Increased population size of 150 individuals.

$$\begin{aligned}
C_j &= 1101010001 \\
S_1 &= \star\star\star\star 010\star\star 1 \\
S_2 &= \star\star 01\star\star\star\star\star
\end{aligned}$$

Figure 5.14: Sample chromosome with two possible shemata. 0 and 1 denote fixed bits, the asterisk \star is a wild card symbol matching 0 or 1.

to crossover and mutation.

$$m(H, t+1) = m(H, t) \frac{f(H)}{\bar{f}} p_o(H) \quad (5.11)$$

This reproduction probability is again a function of the schema H . Since short schemata are less likely to be cut by crossover operations, their probability for reproduction is higher. Mutation schemata with many fixed positions are more likely to be changed than those with only few such positions. Reproduction according to its fitness will increase or decrease the number of strings incorporating a certain schema by a constant factor which is equivalent to exponential increase or decrease. This was formulated by Holland in the well-known schema theorem: Short, low-order, above-average schemata receive exponentially increasing trials in subsequent generations.

The number of possible schemata incorporated in a population is proportional to the number of possible bit combinations 3^l including wild-cards with l the number of bits in the string and n the number of strings in a population. This allows a string to represent at maximum 2^l different possible schemata at the same time (depending on which positions are included in the schema and which are not). The genetic algorithm evaluates those schemata inherently in parallel. This is called implicit parallelism and is seen as a major advantage of genetic algorithms.

The schema theorem alone is not enough to explain the driving force of genetic algorithms, since the interdependence of parameters vastly reduces the actual number of schemata, and reproduction is not a way to improve solutions but only to prevent good solutions from being destroyed. Holland therefore created the building block hypothesis: A genetic algorithm seeks near-optimal performance through the juxtaposition of short, low-order, high-performance schemata, called the building blocks. Goldberg and his group based the development of a specialized algorithm, known as messy Genetic Algorithm, on this hypothesis claiming improved performance [30].

5.9.2 Dynamics and Convergence

For the standard genetic algorithm, also known as Royal Road Genetic Algorithm, further investigations into the dynamics were done by Altenberg [31], Qi and Palmieri [32], Vose and Liepins [33], and Rudolph [34].

Altenberg and Qi work directly on populations of infinite size describing the dynamics of genetic algorithms. Both developed formulas for the mean fitness increase by each iteration.

Qi uses continuous variables, whereas Altenberg investigates binary genetic algorithms. Both prove convergence of their implementations, but results on infinite size populations do not scale properly to finite size populations where convergence is dependent on the starting population.

More adapted in this regard is the work of Vose and Liepins who formalize the algorithmic dynamics for finite size populations using Markov chains. They show that the standard genetic algorithm converges into one stable global population at the optimum, traversing local optima. Rudolph nevertheless shows that the absolute optimum is not necessarily found in the final population, if the selection operator does not preserve the best solution from iteration to iteration.

Although the convergence is proven, no criterion yet exists about the duration of the search. For a practical optimization run, a stopping criterion has to be found. Investigations into ranking methods were done by Mahfoud [35] using order statistics. Niching methods however are highly nonlinear and no appropriate models exist to date.

5.9.3 No Free Lunch Theorems

To end this chapter, a reflection on optimization in a whole is appropriate. Global optimization algorithms in general are considered ineffective. It is true that convergence and exactness of the achieved solution or number of solutions is comparably lower than for solutions found by many local optimization methods. The comparison is of limited view, however, and is not fair as is formalized by Wolpert and Macready in a discussion on “no free lunch” theorems [36]. Global optimization algorithms are designed to search the whole parameter space, whereas local optimization methods may trust in the uniqueness of the solution and may therefore use considerably more knowledge from earlier iterations. Consequently the optimization is dependent on the problem under investigation. Any effective optimization algorithm is therefore adapted to one class of problems and there is no free lunch to be gained for any other class.

The fact that genetic algorithms are able to test huge design spaces with multiple minima, discrete parameters, and non-linearity makes them appropriate for the conceptual design phase. Whenever the search is narrowed down to few continuous parameters, local deterministic optimization methods are better applicable. Robustness is traded for speed of convergence, and the quality of the final result for the ability to work on global problems.

Chapter 6

Conceptual Coil Design

Although many authors focus on the choice of optimization algorithm, the major problem in magnet design is seen in the correct definition of design variables and the objective function. This importance is reflected in the design example of the main dipole coils for the LHC, that will be discussed in this chapter. At this stage the key features of the design are already defined by the physics requirements. However, the setup of conductors in blocks and layers is subject to optimization, defining the performance that can be achieved.

The first section introduces general considerations necessary in the conceptual design phase. The process of variable identification and parameter evaluation is discussed in detail. The results of the dipole coil optimizations are shown and their performance characteristics are discussed in order to further elaborate the treatment of different objectives.

6.1 Conceptual Design Phase

With respect to accelerator magnets developed at CERN, the conceptual design phase will be outlined and basic considerations regarding the setup of parameters and objectives will be discussed. The magnet design from the standpoint of field computation can be divided into the following stages:

- The coil design:
The magnetic field due to the conductor layout is calculated as a current distribution problem. The correct position of the conductors is taken into account. The iron yoke is simulated by the imaging method as a cylinder, reflecting with a constant magnetic permeability. A value for the magnetization can be estimated by measurements on prototype magnets or calculated from FEM simulations using an approximated coil distribution. The quench margin can already be taken into account if the peak fields in the conductors are evaluated. The problem scales linearly with the current excitation. Nevertheless optimization for optimal main field constrained by the maximum peak field is a non-linear problem.
- Persistent current compensation:
The persistent currents are calculated by a non-linear estimation from the magnetic

field in the strands. The magnetic field errors due to the persistent currents are added to the systematic multipole components. A compensation of these errors renders a re-iteration of the coil design necessary. Since persistent current effects are a magnitude smaller than the systematic multipole components, the changes to the coil geometry are comparatively smaller.

- The yoke design:

The forces on the conductors by the excited magnetic field has to be taken into account and limit the possible structure of the yoke. The interaction of collars, yoke, and shrinking cylinder constitute the force retaining structure. The collars are made from non-magnetic material. Consequently, the structure of the surrounding yoke is the object of electromagnetic optimization. Nevertheless the contour of this yoke allows for various alterations. Holes in the yoke itself may help to redirect flux symmetrically.

- The coil end design:

As the coil has to be wound by a machine, and the conductors have to be cycled back on the opposite side of the magnet, a setup of the conductors in the ends of the magnet has to be designed. The degrees of freedom are rather limited and the structure is mainly defined by the minimum bending radius of the cable. The spacers, defining the shape of the cable in the ends, have to guarantee tight positioning, to avoid any movement leading to a quench. Since the possible variations in the end design are rather small, conventional optimization methods can be applied. This part of the coil design is therefore not considered any further.

This approach is reasonable since the interdependencies of the various stages are usually small. The field errors due to persistent currents, for instance, are dependent on the magnetic field in the strands. The filament magnetization is much lower at high field levels than at low field levels as was shown in figure 3.3. Regions with major persistent currents are therefore found in the lower coil blocks, where the magnetic field is low. These blocks are normally not subject to changes. An iterative approach is thus appropriate for high field magnets. However, persistent current evaluation with the method described in section 3.7 is not too time consuming and can be included right from the beginning. Persistent current effects have to be taken into account in the first iterations for the design of correction magnets, in particular, where the main operation area is around zero field level.

The coil and yoke design processes can be regarded separately, since higher order multipoles are mainly defined by the conductor distribution, whereas the iron mostly influences lower order multipoles. The quadrupole, like other even multipole components in dipole design, is solely caused by the non-symmetry with respect to the beam apertures of the two-in-one iron yoke. A bare dipole coil does not produce any quadrupole component. By contrast, a change in the coil design of a two-in-one structure influences the quadrupole component, though the magnitude of the effect is usually much smaller than on the odd multipole components.

6.2 Coil Design Phase

In the following, the initial coil design stage is discussed using the example of the main dipoles of the LHC. At the time when this investigation was started, a feasible solution for the coil already existed and model magnets had been built and tested. Several re-iterations had been done to adjust for the changes in the cable size and parameters. A point was reached where the design became inflexible and further adjustments could not be made. At this stage the whole design had to be reconsidered as shown below.

Generally, as a starting point for a new coil design, specifications can be derived from parameters of an existing design. Scaling laws can give a good estimation of a possible structure and its tolerances. Basic parameters can be found by analytical considerations as shown in chapter 4. Based on these calculations the feasibility can be estimated and the frame for a new design can be defined. Using this procedure, major decisions are already taken before the first simulations are carried out. In the case of the LHC dipoles, the necessary field level is attainable only by magnets in which the field distribution is dominated by the coil. The dominant non-linearities of so-called super-ferrie magnets, where the iron pole shape determines the field pattern, do not create the desired field quality when saturation of the iron poles is reached. The electromagnetic forces on the coil structure required a two layered design of non-magnetic collars and a magnetic iron yoke surround held together by a shrinking cylinder (section 2.3). Although the positioning of the coils is to be defined by electromagnetic field calculations, the assembly method is predefined and causes certain restrictions. For example, in a standard collaring procedure, the outer coil must not exceed the angular position of the highest inner block.

At this stage, still to be defined is the coil layout. The design parameters may vary in a large range and the dependencies of the objectives on the design variables are non-linear. Multiple possible parameter-sets exist that reach the design specifications. The aim of the conceptual design stage is therefore to find these parameter-sets and to choose the best alternative. The optimization has to take into account that several design parameters are inherently discrete (e.g. the number of conductors per coil block) whereas others are continuous (e.g. inclination angles of coil blocks).

As shown in figure 6.1 changes of integers may introduce big jumps in the objective function which cannot be optimized by methods assuming continuity. As an example, a possible 6 block design is modified by one conductor shifting from one block to the next adjacent one. Although those shifts should not influence the structure of the rest of the coil, the resulting changes in the multipole components are rather big. Most notably, a shift of one conductor from block 4 to block 5, results in an enormous change of 9 units on the b_3 component and 0.2 units on the b_7 component. Therefore an algorithm has to be employed which is capable of optimizing non-continuous design variables. Moreover, most continuous parameters exhibit a non-linear influence on the objective function resulting in multiple minima.

One problem remains though: Not all of the objectives can be included in a single objective function and several constraints are either not easy to implement, or would influence the convergence badly. In order to overcome this problem, the optimization algorithm was adapted to provide multiple solutions from which the designer may choose. This ensemble is

	Number of conductors						Multipole components			
Case	1	2	3	4	5	6	B1 (T)	b3	b5	b7
v6-1	9	16	5	5	3	2	8.359	1.4193	-0.1053	0.0255
2 to 1	10	15	5	5	3	2	8.369	3.7043	0.02	0.0285
3 to 4	9	16	4	6	3	2	8.351	-4.5274	-1.0601	-0.0425
4 to 5	9	16	5	4	4	2	8.33	-7.6596	0.7452	0.1996
5 to 6	9	16	5	5	2	3	8.316	2.3245	1.3293	-0.1831

Table 6.1: Influence of single conductor shifts on the field quality. In each case, a conductors is shifted to an adjacent block. The number of conductors in each block and the resulting field quality is shown. Shifts in the higher inner blocks, in particular, result in big changes in the multipole components. Relative components are given in units of 10^{-4} at a radius of 10 mm.

generated according to his or her formulation of objectives. Still, the final evaluation of each parameter-set has to be carried out by the engineer, checking for constraints and comparing the solutions with each other as well as to analytically found quality measures.

Genetic algorithms as discussed in chapter 5 were found to fit with all of the requirements. They are

- global optimizers,
- insensitive to weights,
- able to handle integer and real parameters,
- adjustable to produce a set of solutions.

Nevertheless, the disadvantages also have to be considered by the designer in all optimization runs:

- Global optimizers have to be followed by further optimization stages.
- The choice of global weights is difficult. Analytical considerations may help.
- Genetic Algorithms need a high number of function evaluations.

6.3 Design Variables

Variables which are still unknown at this stage are the number of blocks in each layer, the number of conductors in each block, and the positioning and inclination angles (fig. 6.1). In certain cases, the parameters of the cables and the positioning radii may be subject to optimization in a predefined range.

Regarding the number of blocks, it is clear that with each additional block, two design variables, a positioning angle and an inclination angle, are added. This results in an increase in degree of freedom by two, making the design more flexible and allowing for a better

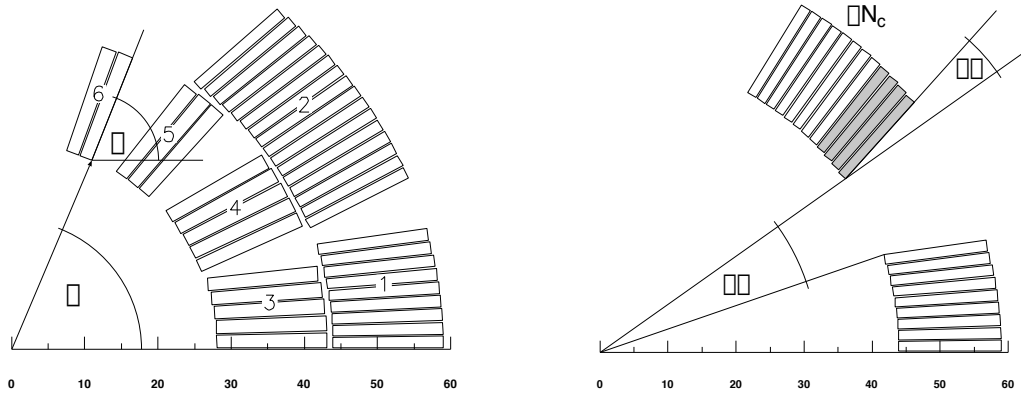


Figure 6.1: Definition of block angles and block numbers (left). Design variables are the number of additional conductors ΔN_c , the relative positioning angle $\Delta\varphi$, and the relative inclination angle $\Delta\alpha$ of each coil block (right).

compensation of multipoles. It is clear that with a multi-layered design where the bottom blocks are fixed in their position by the insulating mid-plane, the number of design variables is $2 \times \text{blocks} - \text{layers}$. It is desired to use as few blocks as possible, since each new block needs an additional wedge-spacer in between to maintain the correct position of the block. It was found in trials that a 5 block design is feasible, though a 6 block design yields better results. The reason for this improvement will be discussed later. Further blocks do not improve the results any further. In all genetic optimizations, the number of blocks was therefore kept constant.

The number of conductors is limited by the available space per layer and some mechanical and winding considerations. Since the key-stoning of the cables is not sufficient to create an exact arc segment at the positioning radius, the more conductors are stacked in a block, the more the inclination angle of the conductors decreases (e.g. block 2 in fig. 6.1). This angle in turn is limited by the winding procedure. Regarding the inner blocks the peak field has to be taken into account. Especially for the uppermost blocks, the local field enhancement by the self-field of the excited conductors has to be considered, and favours smaller upper blocks. Secondly, the Lorentz forces on upper blocks are mainly oriented radially, pointing out of the aperture. The cable has to be well supported by the structure which is mechanically easier for small blocks and an inclination angle as radial as possible.

For the optimization with genetic algorithms, all continuous variables have to be discretized and encoded into bit-strings, according to the implementation of the algorithm as was described in chapter 5. This is done by linearly matching the integer numbers, represented by a certain number of bits, with the real numbers of the design. The number of bits used for such a representation has to be high enough to correctly represent the parameter in the optimization run, but at the same time should be as low as possible to limit the overall length of the bit-strings. Discretization intervals that are too large may cause minima to be missed. Nevertheless the shorter the strings, the smaller the degree of freedom of the optimization and the fewer function evaluations necessary. The number of bits used for each design variable is shown in table 6.2. The overall number is therefore 57.

	Number of Conductors ΔN_c						Rel. Cond. Position $\Delta\varphi$ [deg]				Rel. Cond. Inclin. $\Delta\alpha$ [deg]			
Block	1	2	3	4	5	6	2	4	5	6	2	4	5	6
Minimum	5	5	3	2	2	2	1	1	1	1	-5	-5	-5	-5
Maximum	20	20	10	9	5	3	8	5	10	10	15	15	15	15
Bits	4	4	3	3	2	1	5	5	5	5	5	5	5	5

Table 6.2: Minimum and maximum values of the design variables and the number of bits in the encoding

6.3.1 Constraints on Design Variables

The correct choice of design variables is dependent on the manufacturability of the related design. As the coil structure is defined by the winding procedure, the conductors get stacked one above the other and the position of each conductor is defined by the conductors below. It is therefore consequent to use relative positioning angles $\Delta\varphi$ for the block definition rather than to define absolute angles and to check for missed constraints (fig. 6.1). Overlapping blocks then are not possible by definition, and the size of the wedges can be kept in a pre-defined range. For inclination angles, the method of choice is to supply a deviation from the absolute positioning angle $\Delta\alpha$. Using the absolute positioning angle as inclination angle sets the first conductor upright on the radius. Any deviation is defined with respect to this position, and guarantees that the block inclination neither becomes too flat nor too steep with respect to the inner edge of the coil.

The range of the design variables should be chosen to provide for the complete set of possible solutions. This means that extremal cases for each single variable have to be included. Certain variable combinations may therefore lead to geometrically infeasible structures. To avoid such cases the plausibility of the parameters would have to be checked before any calculation is done. Since such a check is problem dependent and ROXIE itself does not allow for such tests, a slightly different approach was taken: It has been found for the coil optimization that geometrically infeasible designs also lead to bad field quality. Therefore the optimization automatically avoids such designs. When care is taken that only few such cases exist in the parameter space, the computation time wasted is low.

In addition, it is necessary to include the excitation current in the design variables to guarantee that a maximum field level may be reached when the quench margin, as defined in section 2.5, approaches zero. Although the field level is a linear function of the current, the peak field in the coil is dependent on the conductor distribution and position. Optimization for maximum main field with the constraint of a positive quench margin is therefore a non-linear problem. On the other hand, the linearity of the current-field dependency may be used to extrapolate the maximum reachable field level from the quench margin directly. By this means, indirectly including the current in the objective function may reduce the number of design variables and speed up the optimization process. The resulting objective function will be shown below.

6.4 Objectives for the Coil Optimization

The definition of the optimization objectives is a crucial step of the optimization process. In particular, the conceptual design phase has to deal with objectives that are not clearly defined. This “fuzziness” as well as the fact that some objectives are contradictory, has to be taken into account.

The main objectives for the coil design are high main field and low multipole components. As was shown in section 3.5 only ideal current distributions reach this goal absolutely. For a technically feasible block design the complete reduction of one multipole component deteriorates others. Maximal deviations from these optimal conditions are defined by accelerator physicists as found from beam tracking simulations. However, these values are not absolute as they depend on the chosen lattice of dipoles, quadrupoles and correctors (section 2.2). Only evaluation in a beam simulation of a complete accelerator design may show its qualities and deficiencies.

- Regarding the magnet structure itself, the enormous electro-magnetic forces (about 4000 kN/m radially) need to be considered. The local force distribution in the collars has to be optimized, but this would require computations of coupled electromagnetic mechanical problems. Currently, this is only considered by a first order approximation. Lorentz forces are evaluated by ROXIE and then introduced in mechanical calculations. Possible deformations have to stay within manufacturing tolerances. Fortunately small adaptations of the design may be made in a later phase.
- The protection of the superconducting magnets in case of a quench imposes further requirements, as discussed in section 2.4. Not only must the cable be optimized as such, but also the complete coil design influences the quench properties by its inductance and stored magnetic energy, – two contradictory parameters. Fewer turns reduce the inductance, but a higher current has to compensate for that. Detailed evaluation of the quench propagation leads to complex thermodynamic calculations and network analysis [37].
- As the manufacturing will introduce systematic and random errors due to the applied tooling, these factors have to be already taken into account in the design phase. Designs with low sensitivity to such errors are favoured. As was shown in section 3.8, the sensitivity of the multipole components to block or conductor movement may be calculated from the components per block. Including sensitivities, however, would increase the number of objectives considerably. Since the field quality seems more important, preference is given to the overall coil performance.
- Moreover, a tunability margin has to be considered, to allow for re-optimization of the block positions after the pre-series construction of the magnets. Random errors may be compensated by a sorting procedure as described in [13].

Although all of these objectives must be pursued in the design phase, most of them either cannot be handled by ROXIE alone or lead to an ill-posed problem. As such they can not be included in the objective function. In order to overcome this dilemma, the optimization

algorithm was set up to provide the user with a number of proposals (local minima) which can then be “post-processed” using other methods and tools.

6.5 Objective Function

As was discussed before, a goal programming method is used to define an objective function. The objectives are weighted and superposed in a least mean square sense

$$F = \sum_p w_p (F_p - y_p)^2 \quad (6.1)$$

where F_p is the objective, y_p is the residual, and w_p is the weight given by the user.

Regarding the main field, its linearity with respect to the excitation current is exploited. The main field is rescaled to the value of quench margin zero. All the higher order multipole components are already scaled to the main field value and are therefore independent of the excitation level. More difficult is the proper choice of weights as they directly influence the optimization run. In local optimization problems these weights can be estimated from the sensitivity of the multipoles to changes in the design variables around the starting point. For global optimization, this approach would have to be adapted since the sensitivity varies considerably in the whole design variable space.

In the case of coil optimization a different approach is followed since the weights of the multipole components can be estimated analytically as scaling factors. The overall field criterium is found such as to limit the field errors in the mechanical aperture A (section 2.3). The mechanical aperture for the main dipoles is found to be 22 mm [4].

$$\min \{ \max(B(r, \varphi) - B_1)^2 \} \quad \text{in } A \quad (6.2)$$

The magnetic field $B(r, \varphi)$ can be derived from its multipole components b_n according to (3.11). Because of the orthogonality of the Fourier expansion, (6.2) can be replaced by

$$\min \left\{ \max \sum_{n=2}^{\infty} |C_n|^2 |z|^{2(n-1)} \right\} \quad \text{in } A \quad (6.3)$$

As the maximum of $|z|^{2(n-1)}$ in the aperture is found at the edge $|z| = r_A$, (6.3) reduces to

$$\min \sum_{n=2}^{\infty} |C_n|^2 r_A^{2(n-1)} \quad (6.4)$$

ROXIE calculates the multipole components with respect to a certain user-given radius r_{ref} . The weights therefore have to be set up to rescale the multipole components to the larger radius of the mechanical aperture r_A by scaling factors s_n

$$s_n = \left(\frac{r_A}{r_{\text{ref}}} \right)^{2(n-1)} \quad (6.5)$$

The actual weights of the multipole components w_p are setup to be these scaling factors $w_{\frac{n+1}{2}} = s_n$ without further adjustment.

The weight of the main field component must be adapted according to the designers wishes. The value has to be low if the main field may be sacrificed for lower multipole errors, and high if the main field is more important. The value is found either from existing designs or by iteration over several optimization runs. Based on the linear field current relationship, the maximum field B_{ss} can be calculated from the peak field in the conductors $B_{ss} = B_1 I_{ss} / I$. Using B_{ss} in T, and the multipole components b_n in units of 10^{-4} , the objective function thus reads

$$F = 500 \frac{1}{B_{ss}^2} + 3b_3^2 + 70b_5^2 + 1646b_7^2 + 38565b_9^2 + 903408b_{11}^2 \quad (6.6)$$

In cases where the weights on higher order multipoles are found to be more or less important than suggested from this approach, the radius r_A may simply be increased or decreased respectively. The parameters given here achieved good results.

The calculation employs the imaging method to simulate a linear iron-yoke. The permeability is estimated from measurements of the transfer function $B(I)$ and FEM calculations. The conductors are the standard ones as defined for the LHC. The REM routines as incorporated into ROXIE estimate the persistent current influence [10]. The optimization therefore compensates persistent current effects.

6.6 Parameters of the Genetic Algorithm

The results that will be shown below were achieved with parameter settings of 5% generation, 15% mutation, and 80% crossover, given per string and iteration (not with respect to the bits of the strings as in most publications). The number of iterations was set to 5700, one hundred times the number of bits. As was discussed in section 5.8.2 these parameters were found by tests. The optimization process appeared not to be sensitive to the exact values. The parameters are problem dependent, however, and might have to be adjusted.

The number of 5700 evaluation must be regarded as low. Assuming that all of the possible combinations should be tested, $2^{57} = 1.4 \cdot 10^{17}$ evaluations would be needed. Since for the continuous parameters a deterministic algorithm could be used, a lower estimate appears to be more reasonable. Testing all the possible conductor combinations (within the limits of tab. 6.2), still $2^{17} = 131072$ deterministic optimization runs with about 50 function evaluations would be needed. Even if it is considered that a coil design must use between 35 and 45 conductors because of the necessary field level, 64743 deterministic optimization algorithm runs, again with 50 evaluations, would have to be carried out. This number is found from a combinatorial consideration [38]. The problem of distributing a number of conductors on the 6 coil blocks is equivalent to the problem of distributing a number of objects on 6 boxes with $2^{k_i} - 1$ indistinguishable places, $k_i = 4, 4, 3, 3, 2, 1$ (according to the maximum number of conductors defined from the bit encoding tab. 6.2). Then, however, more a priori knowledge is introduced which is difficult for a procedure to be generally used in the design phase.

6.7 Coil Optimization Results

Using the parameters, design variables, and objective function as described so far, the results of several genetic algorithm runs are discussed. The genetic algorithm was set up to start with a pseudo random sample. This means in turn that each new run of the genetic algorithm will start with a new seed and will end in a final population after a finite number of iterations which is different each time. On the first view this might suggest that the optimization either did not work at all or did not run long enough to evolve into the global optimum. Closer investigation nevertheless shows that the results are rather similar in three out of five runs, and that the objective function values reached clearly prove that the optimization is successful. The optimization has to be stopped prematurely, since convergence slows down with the number of iterations. Improvements become smaller with the run-time. Absolute convergence of this stochastic method is only reached for an infinite number of iterations. Moreover, not a single global optimum is sought, but a reasonable number of close enough local optima is preferred.

In table 6.3 the results of 25 optimization runs are compiled. Since the number of conductors in each coil-block is the primary information which should be gained in the conceptual design phase, only these are displayed. More specifically, only the inner blocks are taken into account, as the objective function and therefore the coil-design is less sensitive to the number of conductors in the outer blocks. As can be seen from the table, more favourable solutions are found with a higher multiplicity.

Bl.	3	4	5	6
6×	10	2	2	2
6×	3	3		
2×	8			
3×	9			
2×	6	4		
9×	5	5		
2×	4	6		
1×	5			
1×	6	7		
19×	3	2	3	
1×	4			
6×	3	3		
1×	8			
1×	3	4		
1×	4			
3×	5			

Bl.	3	4	5	6
4×	6	4	3	2
2×	7			
6×	3	5		
9×	4			
2×	8			
1×	9			
5×	3	6		
4×	4			
4×	5			
5×	4	7		
6×	4	4	4	
18×	3	5		
15×	4			
2×	5			
2×	6			
1×	10			

Bl.	3	4	5	6
12×	3	6	4	2
19×	4			
1×	5			
7×	3	7		
1×	4			
2×	3	8		
1×	4			
2×	3	5	5	
1×	4			
1×	9			
25×	3	6		
1×	5			
22×	3	7		
6×	4			
1×	4	5	3	3
1×	5			

Bl.	3	4	5	6
1×	3	6	3	3
1×	7	7		
6×	3	4	4	
3×	4			
1×	10			
15×	3	5		
1×	3	6		
1×	5	3	5	
1×	10			
3×	3	4		
4×	4			
6×	10			
10×	3	5		
1×	9			
1×	3	6		
1×	4	7		

Table 6.3: Optimal designs for block 3 to 6 (inner layer) from 25 genetic algorithm runs with their multiplicity of occurrence. For readability, unchanged values are not repeated.

As the number of conductors in the inner layer increases, the number of conductors in the outer layer decreases to reach the same field level within the quench margin. Since the distance of the outer layer from the magnet center is bigger, the effectivity of current in the inner layer is higher, and designs with more turns in the inner layer certainly need less conductor material. The outer layer on the other hand may be used to gradually compensate

for multipole errors because of its lower sensitivity. The design of the two layers, however, is not independent of each other.

Some of the more favourable solutions have been post-optimized by EXTREM [17], a standard local optimizer included in ROXIE. Six representative examples are shown in figure 6.2. All of the six designs exhibit a similar field quality and would be suitable for the main dipoles of the LHC (tab. 7.4). Nevertheless further considerations have to test for force distributions, quench protection, manufacturability, sensitivity, and tunability.

The first design appears to be feasible as is. It achieves the highest main field of all the six cases, but needs the most conductor material. The second design suffers from an inclined block 6 which makes collaring difficult. Additional shimming would be necessary. Further adjustment might be able to cure this problem. Examples three and four need big copper wedges in the outer layer. When heating up the coils in case of a quench, wedges which are too big, lead to unnecessary delay. Secondly, the current density is high, creating problems with quench protection and current leads. Apart from other deficiencies, designs five and six have rather flat conductor blocks which lead to protruding conductors. Such spots are prone to quenches.

Design	1	2	3	4	5	6
Turns	41	40	39	38	40	40
I_{nom} (A) @8.36T	11199.	11553.	11814.	12032.	11538.	11515.
B_{ss} (T)	9.86	9.77	9.72	9.71	9.8	9.78
L (mH/m)	7.53858	7.1859	6.79842	6.38552	6.98956	6.95108
b_3 (geom+pers)	0.01337	0.01368	0.02328	-0.05162	-0.07218	-0.05946
b_5 (geom+pers)	0.00354	0.00519	0.00479	0.00010	-0.00362	0.00313
b_7 (geom+pers)	0.00029	-0.00045	0.00009	-0.00012	0.00045	0.00004
b_9 (geom+pers)	0.00005	-0.00015	0.00029	-0.00006	0	0.0001
b_{11} (geom+pers)	-0.00001	0	0.00009	0	0	0
NI/B (A/T)	109854.	110555.	110231.	109388.	110415.	110198.
F_p (N/m)	19472.	-1583.	20915.	3411.	3989.	-5467.

Table 6.4: Characteristic data for the 6 examples. F_p ... electro-magnetic force parallel to broad face in the uppermost conductor. b_n in units of 10^{-4} at a radius of $r = 10\text{mm}$. pers ... errors due to persistent currents at injection, geom ... geometrical field errors.

6.8 The LHC main dipole coils

The 5 block design as described in the “Yellow Book” has been subject to many changes since its conception (fig. 6.3) [4]. Most notably, a change of cable dimensions and a part-compensation of persistent current multipoles resulted in an inflexible design. A further compensation of the sextupole component if required would have been difficult, due to geometrical constraints. A further deficiency of this design in quench tests seemed to be due to an unfavourable force distribution in the top block of the inner layer. In order to gain more flexibility for adjustments, a re-evaluation of 6 block designs was considered [39, 40]. The

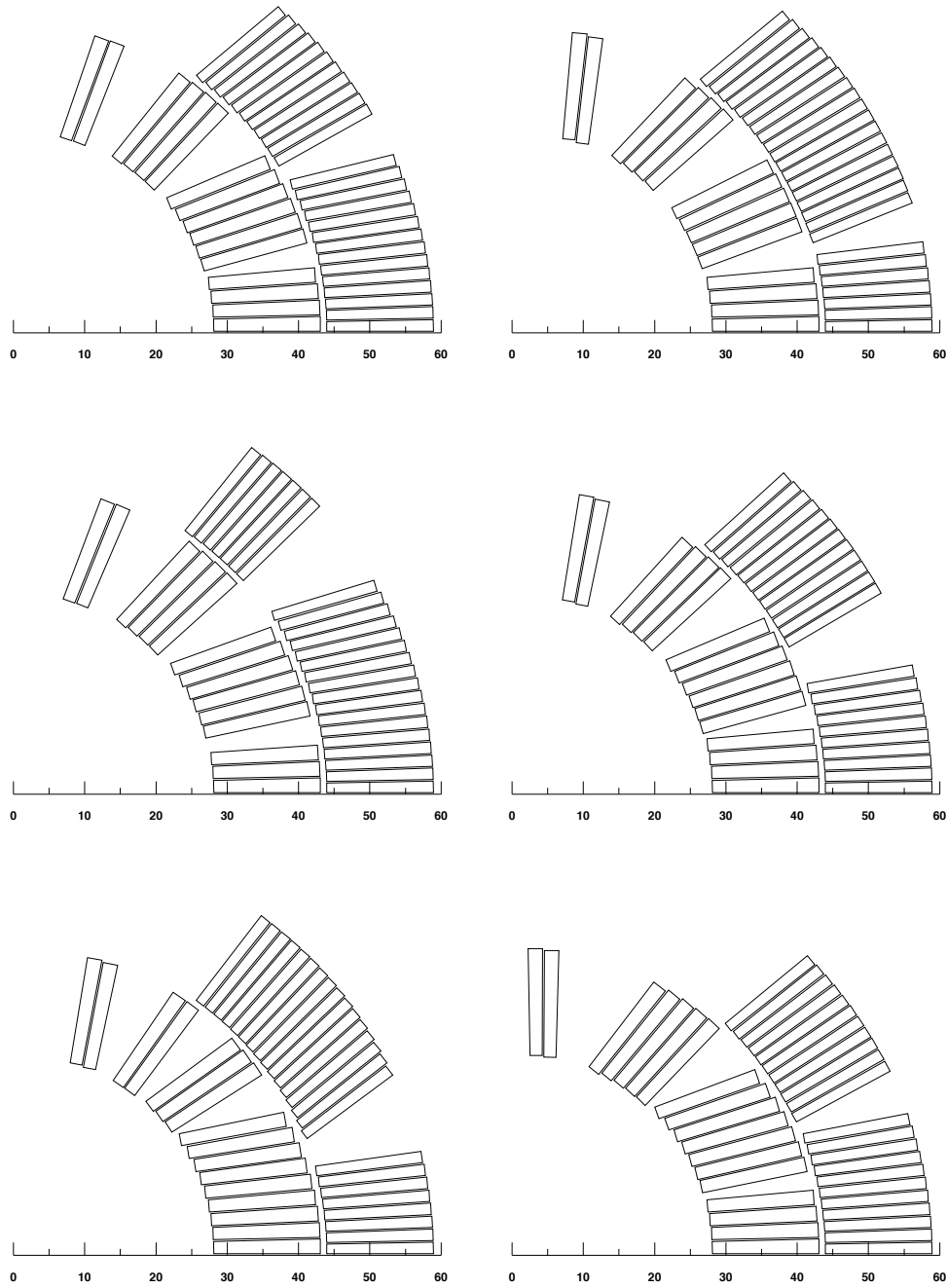


Figure 6.2: 6 block designs from multiple genetic algorithm runs with similar field quality. Nevertheless force distributions, quench protection, manufacturability, sensitivity, and tunability have to be checked. Thus the lower two designs have to be excluded immediately because of overlapping spacers and a too steep top block in the lower right design.

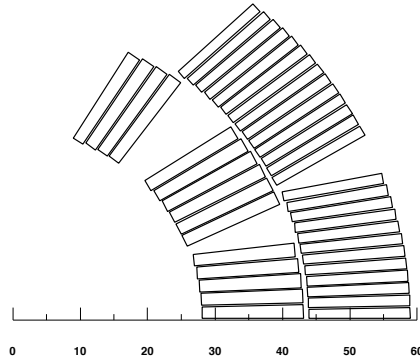


Figure 6.3: Classical 5 block design (LHC Yellow Book)

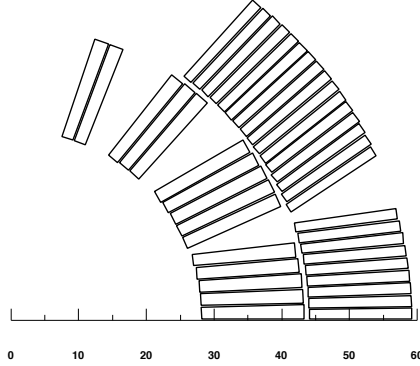


Figure 6.4: 6 block (40 turns) design (V6-1)

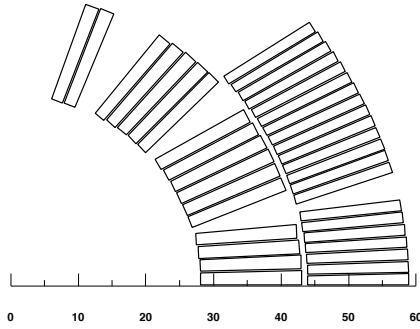


Figure 6.5: 6 block (38 turns) design (V6-3)

objectives of this optimization were to increase the quench margin, to lower the higher order multipole components (especially b_{11}), and to make the wedges symmetric to avoid torque due to electromagnetic forces.

Unlike the procedure outlined above, only a part-compensation of the persistent currents was desired because of further effects in superconducting cables and limits of the corrector scheme at high fields due to saturation. Two 6 block designs were found to be promising and were studied in detail (fig. 6.4, 6.5) [41]. Both designs reach a higher maximum field level B_{ss} than the “Yellow Book” design (tab. 6.5). This is remarkable as it is achieved with fewer turns. The reason is a lower peak-field to main-field ratio in the top block of the inner layer, because of fewer turns and a lower self-field. The down-side is a lower quench margin

	V6-3	V6-1	VY
Turns (coil)	38	40	41
% on LL outer	81.05	84.92	82.5
% on LL inner	86.15	85.64	86.5
PF / MF outer	0.83	0.89	0.87
PF / MF inner	1.03	1.03	1.052
I_{nom} (A) (8.36T)	11879.	11532.	11224.
B_{ss} (T)	9.70	9.76	9.65
L (mH/m)	6.64	7.17	7.47
b_3 (pers)	-4.11	-3.67	-4.17
b_5 (pers)	0.20	0.15	0.21
b_7 (pers)	-0.021	-0.022	-0.036
b_9 (pers)	0.003	0.0035	0.0073
b_3 (geom)	1.0	1.41	0.1
b_5 (geom)	-0.198	-0.1055	-0.19
b_7 (geom)	0.0122	0.0255	0.0342
b_9 (geom)	-0.0087	0.0014	-0.01
b_{11} (geom)	0.0037	0.0029	0.0088
Pole angle (deg)	70.5	70.99	57.4
Pole size (mm)	7.1	7.43	8.7
F_p (N/m)	16400.	17239.	33877.
σa_2	0.637	0.590	0.741
σb_2	0.591	0.572	0.796
σa_3	0.265	0.242	0.304
σb_3	0.239	0.235	0.318
ID-OD effect on b_3	0.05	0.07	0.3

Table 6.5: Characteristic data for the 5 block coil design and the two alternatives. F_p ... electromagnetic force parallel to broad face of cable no. 41 (VY), no. 40 (V6-1), no.38 (V6-3). PF/MF ... Peak-field (in the coil) to main-field (in the aperture) ratio. b_n in units of 10^{-4} at a radius of $r = 10\text{mm}$. pers ... errors due to persistent currents at injection, geom ... geometrical field errors.

in the outer layer and a few percentage increase in current density at the same field-level. The advantages, however, prevail.

Further advantages are lower radial forces and a reduced b_{11} component. The importance of a low b_{11} component was found only recently during beam tracking simulations. Moreover, problems of alignment of the coils on the mandrel are reduced. A representative measure is the ID-OD (inner diameter – outer diameter) difference in the b_3 component. Coils are wound on a mandrel with a certain inner diameter, but as assembly and curing may cause the coil to touch the outer diameter given by the collars, a certain difference to the calculated design is possible dependent on the inclination of the blocks α .

Important for a low probability of quench is also the force distribution in the conductors. The highest forces are naturally found in the region of highest field. There it is important to reduce the force component parallel to the broad face of the conductors (more or less radially). The uppermost conductors of the inner block are usually most sensible. The forces F_p

of this conductor are comparably lower for the two new designs. Further investigations into random errors due to manufacturing, tunability for later compensation, and cable tolerances were carried out [41]. The V6-1 design appears to be superior in most of the investigated points. The inductance values needed for quench calculations do not show any important differences.

Following the design, a number of model magnets (MBSMS) of the V6-1 type were built and tested in the Short Dipole Magnet Test Facility at CERN. The results of a comparison with the “Yellow Book” design confirmed the advantages of the new design (fig. 6.6 and 6.7). The average quench levels over the set of test magnets increase considerably as predicted from the calculations. Almost all of the models achieved the required nominal field of 8.36 T without a quench. Further increase in quality due to training, which is a stabilization of the structure by provoked quenches, resulted in unreached performance close to 10 T [3].

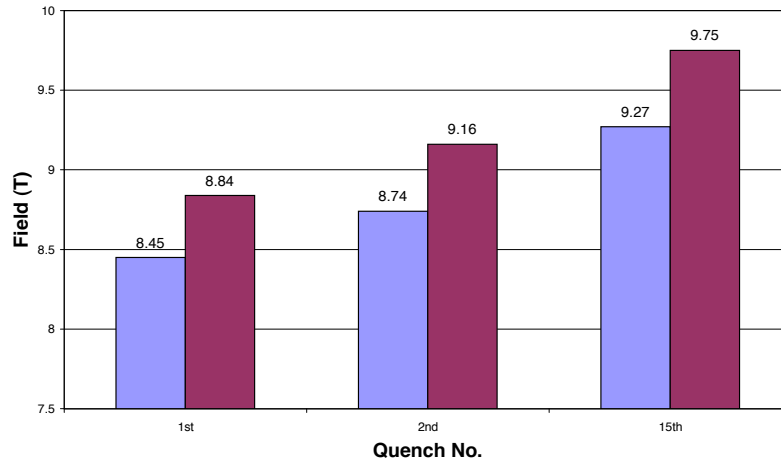


Figure 6.6: Average quenching field of MBSMS models. The new 6 block design (V6-1, dark grey) achieves higher quenching fields, than the 5 block design (VY, light grey).

The design examples confirm the usefulness of the conceptual design approach with genetic algorithms. The method accelerated and simplified the procedure of coil design. Multiple solutions are gained by a few optimization runs of less than half an hour, speeding up the process. More variants can be investigated automatically, resulting in a higher quality of the final design.

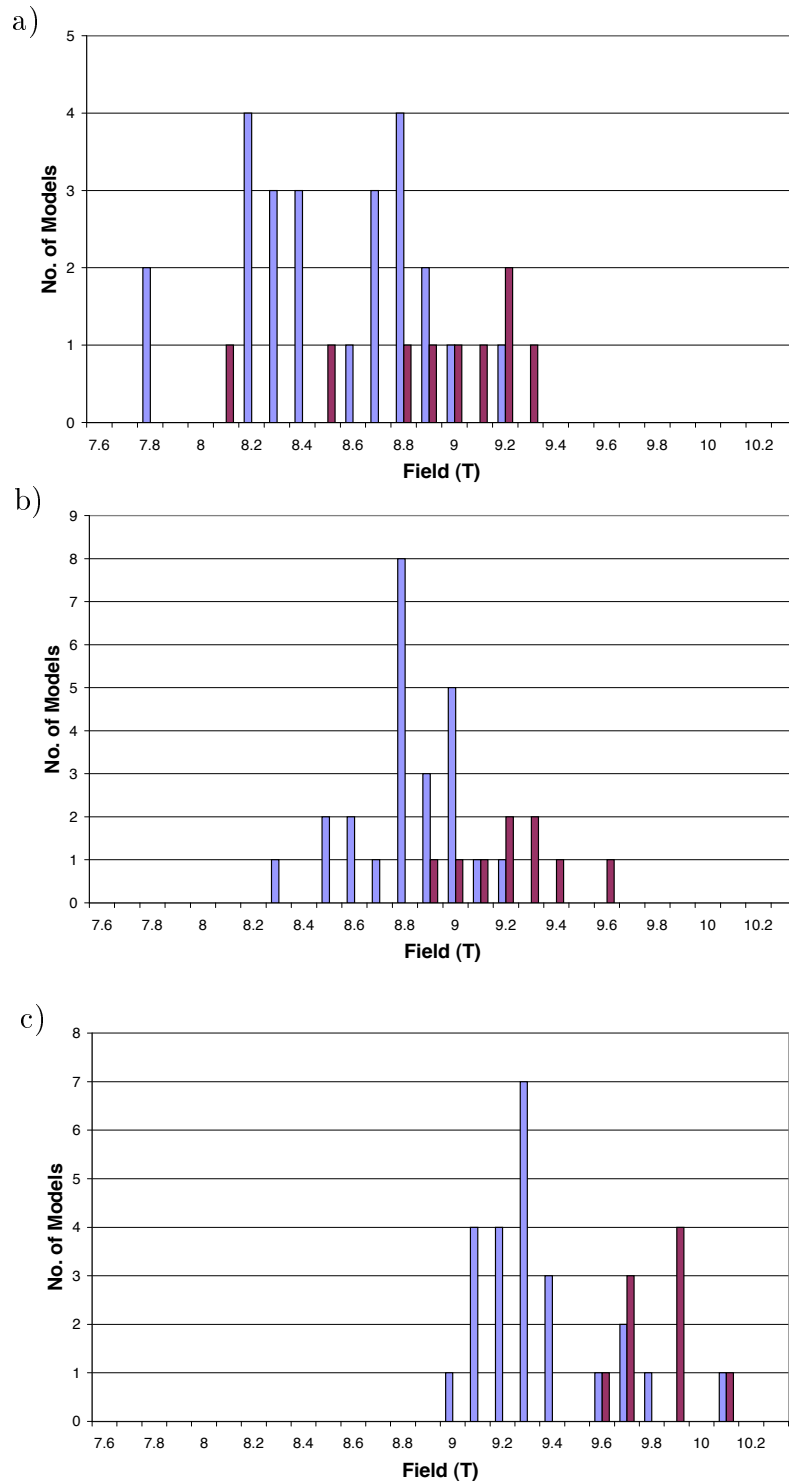


Figure 6.7: Number of model magnets over quench field level reached in the MBSMS series a) at first quench, b) at second quench, c) at 15th training quench. The distribution of quench levels improves in mean value and variance with the number of quenches. 5 block design (VY, light grey) and 6 block design (V6-1, dark grey)

Chapter 7

Material Distribution Problems

The successful development of coil cross-sections, as discussed in the last chapter, gave rise to the application of genetic algorithms to the optimization of yoke structures. The optimization of pole shapes with conventional algorithms as well as evolutionary methods have been successfully undertaken and described in the literature [42, 43, 44]. Considering magnet design as a material distribution problem, in contrast, opens up new degrees of freedom to the design process. Such a technique of structural design was successfully employed to magnet design in [45], using a conventional conjugate gradient method coupled with the Augmented Lagrange Multiplier method.

Harnessing the ideas of structural design, genetic algorithms are adapted to the yoke design of accelerator magnets in this chapter. The property of genetic algorithms to work on bit-strings, allows the direct setup of discretized material regions with mutable characteristics. In combination with finite element programs, this approach leads to a more creative design tool. The new freedom in the magnet structure allows the creation of new features in the design region. In a following abstraction phase, these features have to be detected and parametrized, and serve as concepts for a further optimization stage.

Material distribution problems have to be set up with care. Size and shape of material facets need to be well adjusted. Local effects due to the discretization of the design region must be avoided. The development of a magnet design with genetic algorithms is described below, proceeding through the major stages. The outcome of this procedure is compared to an actual design, and confirms the value of this approach.

7.1 Introduction to Yoke Optimization

The yoke and collar shape for the main dipoles of the LHC obtained so far (fig. 7.1), resulted from several iterations doing mechanical tests on 1 m long models and a few 15 m long prototypes (sec. 2.6). The designs evolved by checking heuristically gained design ideas and running local optimization algorithms over a limited number of parameters.

While collar and yoke have to fulfill their mechanical purposes, keeping the coils in place at all field levels and inhibiting micro-movements of the cables, their shape is designed to exhibit the desired electromagnetic properties. Position and size of the heat exchanger tube

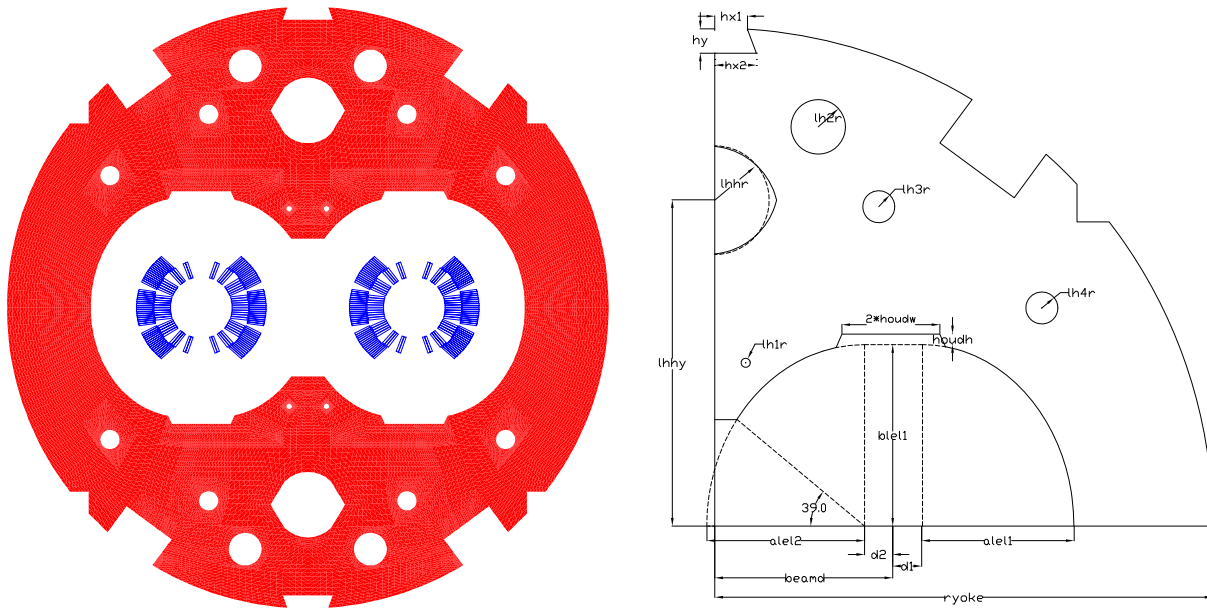


Figure 7.1: Actual design (MBP2) for the main dipoles of LHC found by conventional methods. Collars of aluminium and stainless steel where tested. Several design variables (right) were optimized.

on the yoke center line is fixed. Rectangular grooves must leave space for superconducting bus-bars conducting current from magnet to magnet. Small holes are needed for rods, tying the yoke laminates together. While the size of those grooves and holes is fixed, the position may be changed within a region. All additional holes are used for flux balancing. Vertical cuts on the outer yoke diameter are needed for assembly, to apply the necessary forces for a pre-compression of the coil. The yoke openings which surround the collars provide several parameters, which may be adjusted as well. A notch in the top of the opening introduces further parameters and allows for adjustments at a later stage. In the current design the shape of collars was chosen to be slightly elliptic. The minimal size, however, is restricted by mechanical requirements.

In prior iterations the collars were designed in aluminium. Mechanical tests with aluminium and alternative stainless steel pieces, however, showed that the higher rigidity of steel helps increase the maximal achievable field by preventing conductor movements, and facilitates the yoke design because of well matched contraction coefficients during cooldown. The decision in favour of stainless steel collars was therefore taken. The material change consequently opened up new design opportunities, which rendered a reconsideration of the yoke design appropriate. Most importantly, the rigidity of steel allows considerable reduction of the collar size. Genetic algorithms appear to be a valuable tool in generating ideas for re-evaluation of the design space. Checking these design proposals and locating sensitive regions, the designer is able to get a good overview of possible structures. Using standard deterministic methods in a second step, a feasible design can be produced quickly.

7.2 Design Variables

The magnetic importance of the yoke is not only to increase the field strength. As estimated in section 3.3 the amount of the increase in dipole component is maximally 18% at injection. Another purpose of the yoke is to shield the two apertures from each other, to control their crosstalk, and to shield the outside from the high magnetic fields. Conversely the influences of the cryostats and instrumentation on the field quality must be kept low. The yoke, however, introduces non-linearity into the current/field relationship. As the saturation effect proceeds into the yoke with higher excitation current, the shape of the outer yoke regions comes into play. Holes may compensate the field errors by balancing the magnetic flux.

In material distribution problems the design variables are material regions of the FEM mesh, which may change their properties. In the current implementation these material regions are defined by the FEM grid of the yoke region. As the FEM software, FEM2D, used for this purpose relies on quadrilateral elements, also the mutable regions are of quadrilateral shape. The correct choice of these regions is essential as it defines the range of possible solutions.

The slightly modified yoke structure shows the features of the actual yoke, including the bus-bar grooves and the central heat exchanger pipe and employs the same yoke radius and coil positioning. The mutable regions were selected to allow for the representation of the two extremal cases of separated and combined collars, as well as for a large number of intermediate structures. Inherently the grid is prepared for the definition of holes, grooves, cog wheel structures, magnetic inserts, a central post, or laminations.

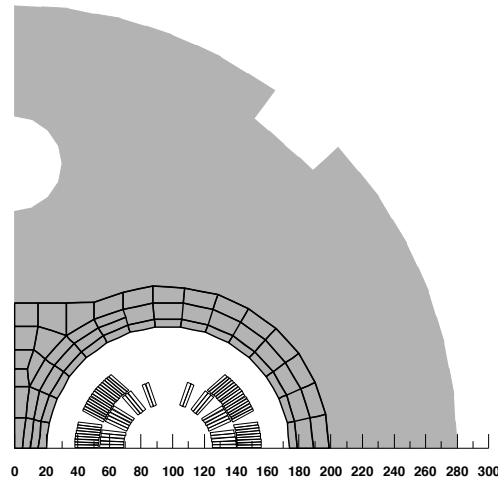


Figure 7.2: A quarter of the iron yoke with an area where quadrilateral facets are free to change their material properties between iron and non-magnetic material.

Altogether 65 macro elements were chosen. To keep the number of areas limited, in order to improve convergence and speed, some macro elements are assigned the same material as a neighbouring cell. This reduces the number of design variables to 56. Each of these design variables is encoded by a single bit for the material type. The variable takes on either of two states. The material type may therefore change between magnetic and non-magnetic. As genetic algorithms are designed to handle bit-strings of a high number of bits, they are well

adapted for this purpose. Since each design variable consists of only a single bit, no further encoding is necessary.

The influence of size and structure of the elementary areas has to be taken into account in the definition of the objective function. A material change in big areas can result in big changes in the objective function with the effect of strong non-linearities. The influence on higher order multipoles, in particular, has to be considered. Though genetic algorithms are able to cope with such non-linearities, the convergence rate is reduced. On the other hand the areas cannot be made too small as that would result in too many design variables, needing more iterations to converge to reasonable results.

7.3 The FEM Modeling

The correct modeling of the yoke structure in the FEM software is especially important for a good result. As the yoke is horizontally and vertically symmetric with respect to its diameters only a quarter of the structure has to be modeled. The horizontal and vertical symmetries are achieved by Neumann boundary conditions. Note that the two apertures as shown in figure 7.1 must produce magnetic fields of opposite sign to bend the counter-rotating beams of protons in the same direction.

Several points have to be obeyed in order to achieve a correct result. The elements have to be well shaped with regard to their angles, and must be neither too small nor too big. In the region outside the yoke, considerable space has to be meshed to correctly simulate the fringing fields. In addition, the center region in the aperture, where the harmonic analysis is carried out, has to be meshed regularly to reduce the sampling error. Special attention has to be paid to the mesh density. In order to reduce calculation time, as coarse a mesh as possible is sought. The necessary precision, in contrast, can only be reached by a sufficiently fine mesh. In table 7.1 the influence of the mesh density on the multipoles is shown.

The multipole components converge to stable values with sufficiently dense meshes. Since the vertically symmetric coil produces only odd multipoles, the even components can be attributed to the effect of the yoke. The dependence of the multipoles on the distance of the yoke from the evaluation points diminishes with higher orders. The values of higher odd components therefore approach the results of a bare coil, whereas the even components approach zero. A number of about 6000 elements is found to be fine enough to accurately model the multipoles up to the b_7 component. Further reduction in the number of elements can be achieved, adapting the mesh with the distance to the field evaluation points. In the genetic algorithm runs about 4600 elements are used.

7.4 Optimization Objectives

The yoke cross-section must be optimized for low variation of the multipole components. At low magnetic fields as encountered at injection, the variation of multipole components must be kept particularly low, in order to allow effective injection of the beam into the accelerator. At higher field levels the beam size becomes smaller and a limited variation of the parameters can be tolerated.

Nodes	1128	3217	6296	10365	15424
b_2	7.4849	7.3444	7.3216	7.2708	7.2516
b_3	0.9462	0.9139	0.9270	0.9260	0.9302
b_4	-0.5919	-0.5527	-0.5470	-0.5487	-0.5499
b_5	-0.8992	-0.9539	-0.9531	-0.9520	-0.9514
b_6	-0.0766	0.0056	0.0143	0.0131	0.0127
b_7	0.6040	0.5960	0.6010	0.5988	0.6000
b_8	0.0119	0.0003	0.0029	0.0005	-0.0001
b_9	-0.0692	0.0964	0.0993	0.0991	0.0996
b_{10}	0.0552	0.0096	0.0049	-0.0002	-0.0001
b_{11}	0.6784	0.5614	0.5676	0.5676	0.5696

Table 7.1: Relative multipole field errors at a radius of 17 mm as a function of the mesh-size. For an accurate representation of b_6 at least 6000 nodes are necessary.

Similar to the case of coil optimization, the objective function is defined by a weighted sum of the objectives. The choice of objectives is more complex for this type of optimization due to the significant non-linearity of the problem with respect to material areas and the current. Again the aim of the optimization is a reasonable field quality. The non-symmetry as introduced by the combined collar design leads to even multipole components b_2 , b_4 , etc. The influence on higher order multipoles, however, decreases with the order n of the multipole. As the coil structure is still horizontally symmetric, these multipole components can be solely attributed to the yoke structure. The reduction of these components is therefore of primary importance. Influences of the yoke on the multipole components b_7 and b_9 can usually be neglected with respect to the coil influence. Because of vertical symmetry the skew multipoles a_n are still zero. At low excitation levels with respect to the nominal dipole field, the magnetic flux is well screened by the yoke structure. The inner dimensions of the yoke exert most influence on the resulting field quality. At higher field levels, the outer dimensions of the yoke and the heat exchanger tube also come into play.

A further objective might be the manufacturability. Such an aim could be defined as the avoidance of non-connected structures, the size of holes or the minimum thickness of material regions. None of these objectives were included in the following optimizations in order to make sure not to restrict the optimization by heuristic reasoning. As already mentioned, the purpose of using genetic algorithms for the yoke optimization is seen in the development of new ideas. Directing the optimization too much by prior knowledge, could therefore result in a narrowed view of the problem.

As the optimization should seek solutions which comply with the limits equally well for all of the acceptable minima, the weights would have to be selected for this purpose. The uncertainty about the achievable results, nevertheless, calls for an iterative procedure for the adjustment of weights. This is done simply by inspection of the results, and increase of weights of insensitive objectives over easily optimized sensitive ones. About three runs were found to be sufficient in order to gain reasonable solutions.

The fact that only finite elements can change their properties in the approach described

here, is a limitation which has to be taken into account in the choice of the objective function. Too high constraints on objectives which are sensitive owing to the discretization into finite elements, results in a high number of pseudo minima. The optimization was started with a simple objective function at injection field level, with a radius of harmonic analysis of 20 mm. (Results, however, are given with respect to the standard radius of 17 mm, as defined for the LHC.) Using B_{ss} in T, and the multipole components b_n in units of 10^{-4} , the objective function reads

$$F = 20 \frac{1}{B_{ss}^2} + 10b_2^2 + 0.4(b_3 - 5.66)^2 + 20b_4^2 + 10(b_5 + 1.68)^2 + 10(b_7 - 1.63)^2 \quad (7.1)$$

7.4.1 The Problem of Jagged Structures

Jagged structures (fig. 7.3) were obtained from the objective function (7.1). Such yoke shapes are not only impossible to manufacture, an evaluation at a number of excitation levels also revealed a problem of strongly varying multipoles. None of the resulting structures exhibits the desired field quality except at a single excitation level for which the optimization was done. Such a behaviour is not acceptable for accelerator magnets.

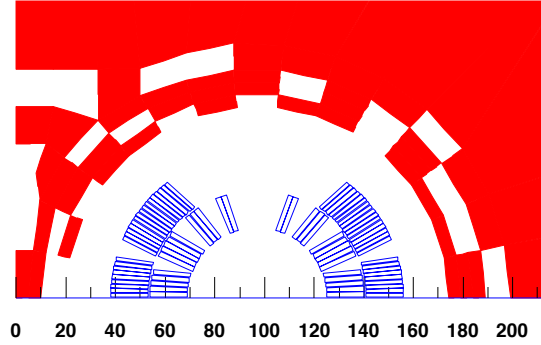


Figure 7.3: Jagged yoke structure resulting from an optimization performed for a single excitation. The structure is physically meaningless because of its sensitivity and cannot be built.

The cause for this deterioration in field quality was found in local saturation effects. To remedy these effects, the design must be evaluated at a number of excitation levels and the variation has to be included in the objective function. In the examples below, the variation of b_2 and b_3 was included. The modified objective function reads (B_{ss} in T, b_n in units of 10^{-4})

$$F = 20 \frac{1}{B_{ss}^2} + 10b_2^2 + 0.4(b_3 - 5.66)^2 + 20b_4^2 + 10(b_5 + 1.68)^2 + 10(b_7 - 1.63)^2 + 10\Delta b_2^2 + 10\Delta b_3^2 \quad (7.2)$$

7.5 Parameters of the Genetic Algorithm

The parameters of the genetic algorithm are similar to those of the coil cross-section problem. The values are 5% generation, 15% mutation, and 80% crossover per iteration. The number of bits in the bit-string for the discretization in figure 7.2 is 56, equal to the number of material elements which may change independently. With the aim of creating a higher diversity, the number of bit-strings was increased to 180. Consequently the number of iterations had to be increased similarly. In the actual examples 20000 evaluations were done.

Though the correct choice of the genetic algorithm parameters is attributed a high value in the literature and therefore thoroughly discussed for a large number of different algorithms and problems, the optimal values are not found to be that important in the present case. As long as the amount of crossover is kept considerably higher than the percentage of the other operators, the behaviour of the genetic algorithm is robust with regard to these parameters. In the case of FEM optimization especially, the correct setup of the optimization problem by the appropriate choice of the finite element mesh, the material regions, the objective functions and the excitation levels was found to be of higher importance. Investigations of different parameter values in connection with FEM calculations, would have resulted in several months of calculation time.

7.6 Circular Grid Structures

In the following examples the macro element mesh as shown in figure 7.2 was used for an optimization. First the excitation levels were chosen to be 1000 A and 4000 A, since the variation of the low field region is of primary importance to the beam stability.

A representative example is shown in figure 7.4. The result is well optimized for low fields,

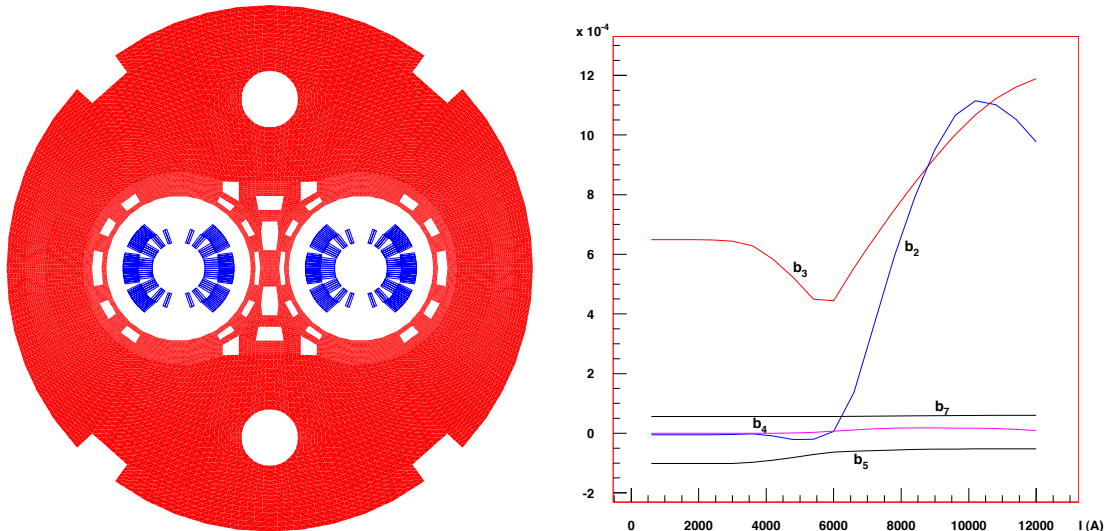


Figure 7.4: Cross-section and transfer function of a solution, where emphasis is put on the field quality at injection field level only.

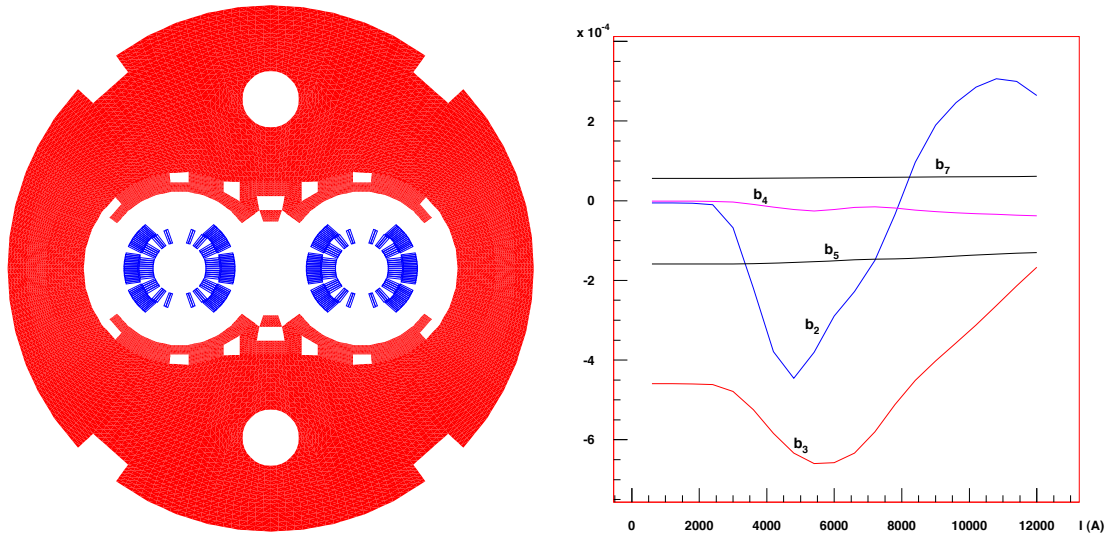


Figure 7.5: One solution when a minimal variation of the multipole content versus excitation is required.

but at high fields the variation of the quadrupole b_2 and the sextupole b_3 components is too big. The structure consists of two collars, which are separated by a perforated magnetic wall. A thin, but continuous, inner layer of magnetic material around the collars guarantees the low variation in the low field region. Layers of holes produce a shielding effect and avoid the closure of flux at regions where the multipole components b_3 would become unbalanced. The flux through the yoke is therefore mainly guided via the center part of the opening. Interestingly, a former work on the optimization of the LHC main dipoles resulted in a similar structure, employing little holes around the collars for fine-tuning of the field quality [46].

In a second test, the excitation levels were changed to 1000 A and 8000 A. The variation at these two values is effectively reduced, the variation outside, however, still amounts to some units. One exemplary result for this case, shown in figure 7.5, demonstrates a combined collar design, similar to the standard cross-section shown further above (fig. 7.1). The probability for an outcome of the combined collar type was generally found to be lower than for the separate collar type. The reason for this preference can be attributed to the higher sensitivity of the low multipole components to material changes in open structures. Important for a high main field is the relatively close placement of the material brace over the center of the coils, guiding the magnetic flux into the iron. Its size and shape mainly defines the amount of the b_3 component.

A special problem in the setup of material regions is found in this example: A single block in the center is not directly connected to the rest of the structure, but only touches it in corner points. Apart from not being manufacturable, local saturation effects produce a high variation in the multipole content. Additionally such structures are sensitive to variations in the exact position of such blocks.

7.7 Brick Structures

With the aim of giving the yoke structures more freedom in the choice of possible structures, a rectangular grid was set up (fig. 7.6). In order to get rid of local saturation effects, the material regions are interlaced creating a brickwall-like structure. Elementary regions cannot touch in one point anymore, but overlap by half an edge. The structure is set up symmetrical with respect to the coil center. The number of possible material regions is increased to 85. The finite element mesh was found to be reasonably fine with about 3600 nodes. It is clear, that the rectangular grid now prohibits smooth radial shapes. A higher variation of the multipoles must therefore be expected.

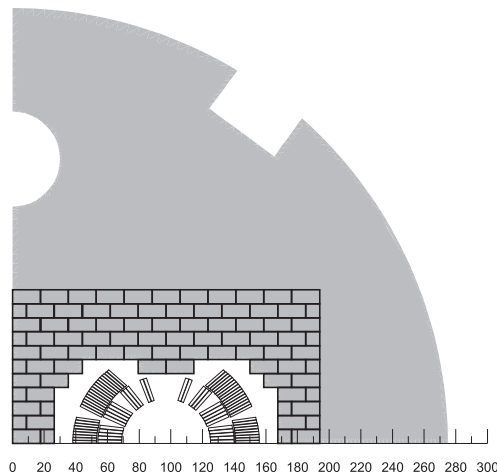


Figure 7.6: Using a brick structure with overlapping material regions, avoids local saturation effects at facet corners. Each of the rectangular facets is free to change material properties between iron and non-magnetic material.

Yoke shapes, that fit into the pattern of rectangular elementary regions, are not favourable for an engineering design as such. The resulting patterns, however, allow the deduction of ideas. Carefully investigating these ideas and combining them in a feasible structure, can lead to a new technical design. This approach is highlighted in two examples below.

The non-linear effect due to the inhomogeneous inner yoke edge is highest when the permeability of the iron changes in this region. This is especially the case at lower to medium excitations. It is therefore crucial to include these excitation levels. The optimization was then done at three current excitation levels, each increasing the computation time linearly. The run-time rose to about two weeks (on a 300MHz DEC Workstation).

The weight on the main field component was considerably increased, to force the optimization to fill the material region. The rectangular grid allows the definition of material regions, which are much closer to the coil than in the circular grid case. The high influence of such regions on the multipoles, led to a removal of the seven closest elements in almost all results.

The high variation of the b_3 component in the first structure (fig. 7.7) can be attributed to the coarse discretization of the inner yoke edge. The b_3 component is too sensitive to changes in single bricks, so that a reduction of the variation to an acceptable amount could

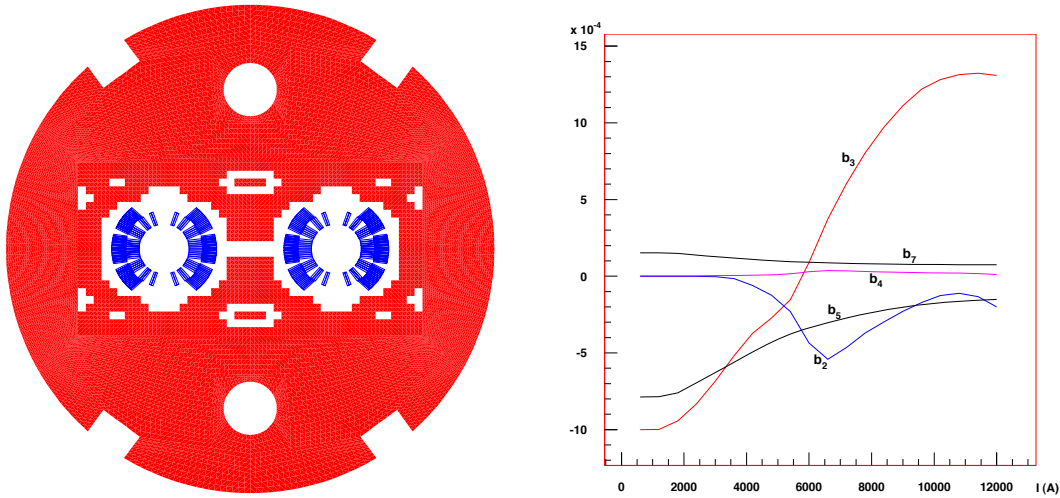


Figure 7.7: Brick structure solution with high weight on the main field component. The result are closed structures with small adjustments for field quality.

not be achieved. A big center hole, small outer holes, and a center gap are mainly used to adjust the b_2 component at higher fields.

Reducing the weight on the main field resulted in more open structures, as shown in figure 7.8. The lower weight on the main field simultaneously increased the importance of the other objectives. The variation of b_2 and b_3 is therefore comparably lower. In order to distribute the weight on the variation equally over the excitation function, five evaluation points were used: 1000 A, 2000 A, 3000 A, 5000 A, 8000 A. The opening reaches higher up in the collars, creating a notch that improves the b_3 variation. The center gap is increased and balanced by an outer notch.

A further remedy to the high variation in the multipoles are smaller brick facets. A necessary higher number of finite elements, nonetheless, leads to a higher calculation time. Increasing the number of elementary areas in x - and y -direction by a factor of two, for instance, would result in an increase in calculation time by a factor of about four and a similarly higher number of material regions. Moreover, a higher number of elementary regions leads to longer bit-strings, needing more iterations in the genetic optimization. Since the calculation time is already in the range of weeks, an increase in resolution appears to be infeasible.

7.8 Abstraction of Feasible Structures

Below, the next step in the design process is illustrated by two examples. The ideas found so far are combined into feasible structures with an effective yoke shape. The elements for such a synthesis are

- the possible shapes of the inner edge, ranging from circular to elliptical,
- the center gap, leading to separate or combined collars,

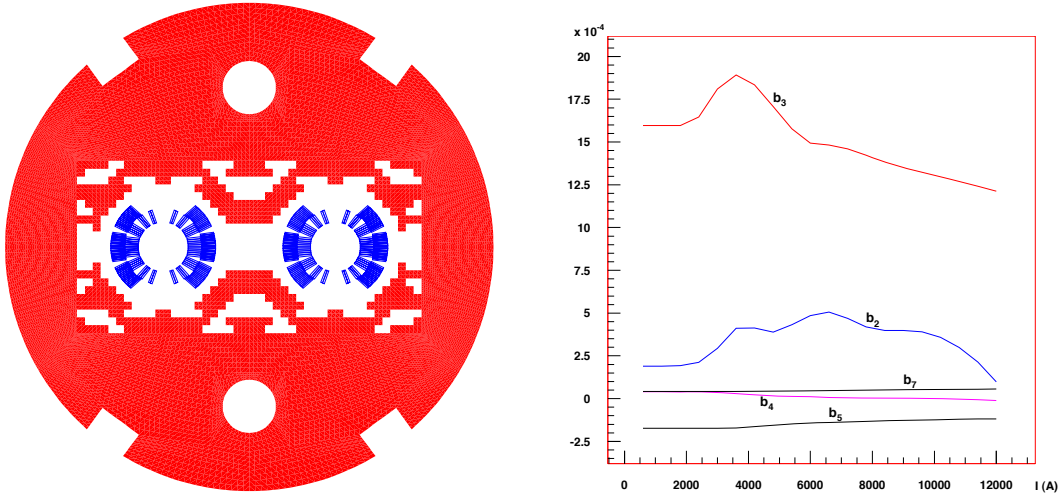


Figure 7.8: Brick structure solution with evaluation at five excitation levels. Less weight on the main component, produces more open structures, with lower variation of the multipoles. Several features can be deduced: A central gap, and several holes and notches influence the field quality.

- center holes of different sizes,
- balancing holes, which are symmetrical with the coil.

Each of these features can be defined with a number of continuous parameters. The features, however, are not independent.

The two yoke shapes are coarsely optimized. In order to evaluate several features, the optimized structure is modified. Removing each of the features separately, their influence is estimated. Though the effects of such changes are basically non-linear, this approach gives a good insight into the importance of the features.

In the first example (fig. 7.9, tab. 7.2), a cross-section with separate elliptical collars is investigated. As in the first brick structure (fig. 7.7), small holes are adjusted to balance the flux through the yoke. The resulting structure shows low variation of the multipole components at low field levels, where it is most important for the accelerator design. The b_3 component, however, is high because of the elliptical shape of the collars. Changing the collars to circular shape (a), the b_3 component is considerably reduced. The variation, on the other hand, is too big for a technical design. A little hole in the outer part of the yoke (b) is used to compensate the influence of the center hole (c) and the center gap (d). Apart from a small effect on the b_3 component a big variation of the b_2 component can be seen. The influence of the little hole is comparably big, because of its proximity to the opening.

The second design (fig. 7.10, tab. 7.3) was gained from an adaption of the rather big holes in the genetic optimization result of figure 7.8 in a symmetric collar geometry. Both geometries look rather similar on the first view. The behaviour of the two structures and its details, however, are essentially different. The space for the collars is still of the combined type. Nonetheless, two separate symmetric collars with side notches can be fitted into this space. The collaring procedure is simplified, since each of the coils can be processed

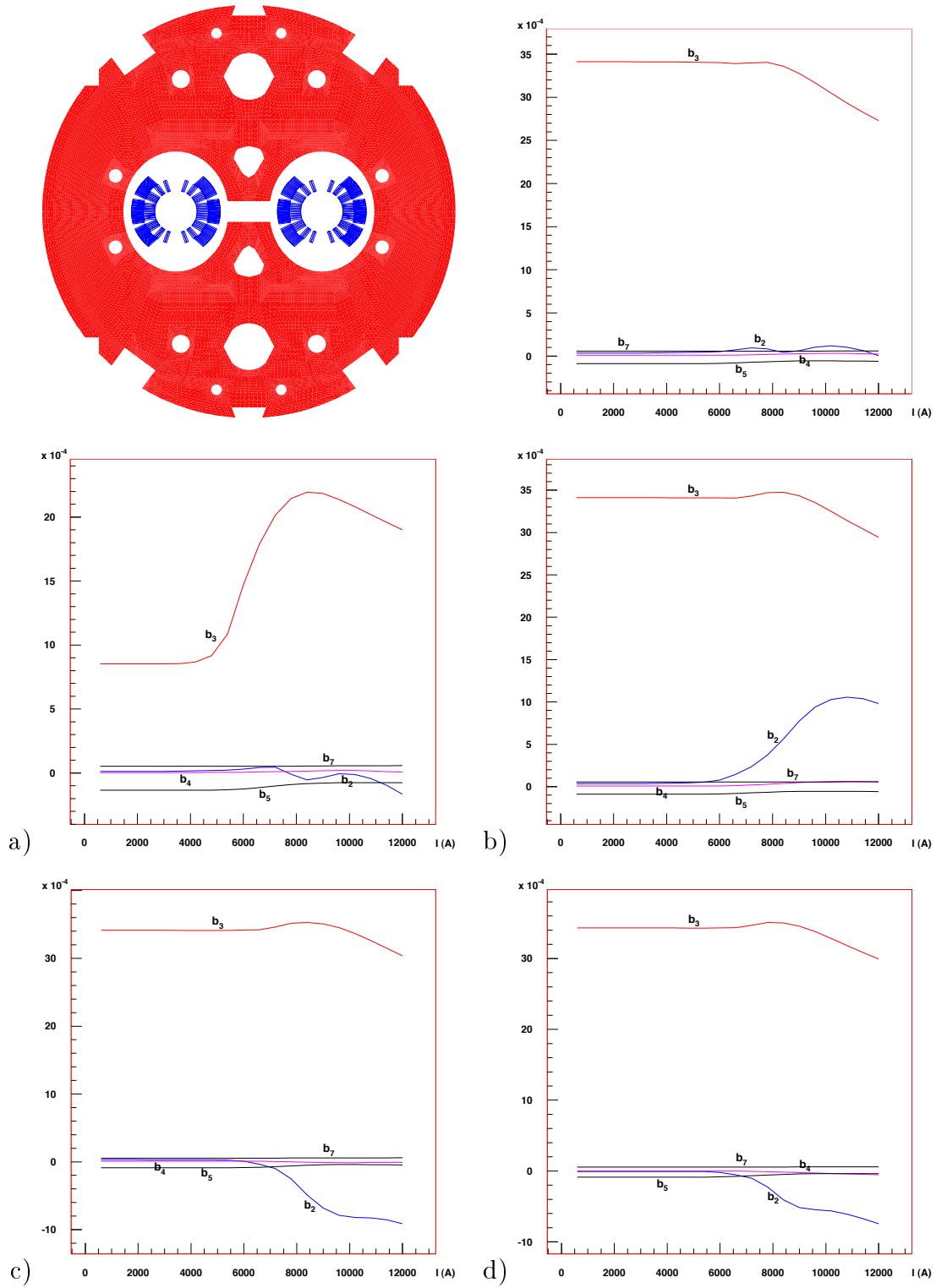


Figure 7.9: Variant 1: Combining the features found in the genetic algorithm run, variants are found. Evaluating those features one by one, the influence is found of a) the elliptical collars, b) the outer hole, c) the center hole, and d) the gap. The multipole components are displayed over the excitation current in units of 10^{-4} at a reference radius of 17 mm. Note the different scales of the images.

	Variant 1	circular	no o-hole	no c-hole	no gap
B_{12kA} (T)	9.10388	9.26467	9.12358	9.15321	9.11947
b_{2inj}	0.34527	0.12312	0.35649	0.31595	-0.06570
b_{3inj}	34.12452	8.53445	34.12586	34.12943	34.33277
b_{4inj}	0.08661	0.04222	0.08654	0.08611	0.00048
b_{5inj}	-0.88874	-1.34605	-0.88885	-0.88875	-0.85754
b_{7inj}	0.55376	0.53986	0.55375	0.55375	0.55689
Δb_2	1.13358	2.11540	10.20551	9.45923	7.39789
Δb_3	6.82818	13.40605	5.32568	4.90434	5.14254
Δb_4	0.23187	0.16616	0.54527	0.22575	0.54806
Δb_5	0.33300	0.59669	0.35366	0.49006	0.49900
Δb_7	0.03643	0.05087	0.03231	0.03756	0.04720

Table 7.2: Variant 1: The field quality of the variant of figure 7.9 and its modified structures. The cases of circular collars, no outer hole, no center hole, and no gap, correspond to the multipole transients.

separately. A bigger center gap, and much bigger and differently positioned holes, further modify the structure. Again the comparison between circular and elliptical collars is made, showing a similar effect as in the first example. The absolute result, however, is rather different. The center hole shows a big effect on the b_2 and a smaller effect on the b_3 component. These multipole components are compensated by the outer hole and the outer notch, which symmetrizes the influence of the center gap. A similar structure was proposed in [47].

	Variant 2	elliptical	no o-hole	no c-hole	no notch
B_{12kA} (T)	9.02284	8.95667	9.06917	9.11621	9.06010
b_{2inj}	-0.05541	-0.05792	-0.04000	-0.13895	2.59095
b_{3inj}	13.43214	29.79277	13.43193	13.44844	14.72905
b_{4inj}	0.00064	0.00065	-0.00051	-0.00105	0.49977
b_{5inj}	-1.41863	-1.43205	-1.41894	-1.41853	-1.25303
b_{7inj}	0.52487	0.51359	0.52488	0.52485	0.53772
Δb_2	2.93956	3.57486	22.23019	24.68521	11.73731
Δb_3	3.01193	14.48337	3.42388	4.86229	2.71633
Δb_4	1.19655	0.99129	0.71543	0.54035	1.92086
Δb_5	0.28857	0.25116	0.13216	0.44718	0.46731
Δb_7	0.04203	0.04899	0.04037	0.03800	0.04242

Table 7.3: Variant 2: The field quality of the variant of figure 7.10 and its modified structures. The cases of elliptical collars, no outer hole, no center hole, and no outer notch, correspond to the multipole transients.

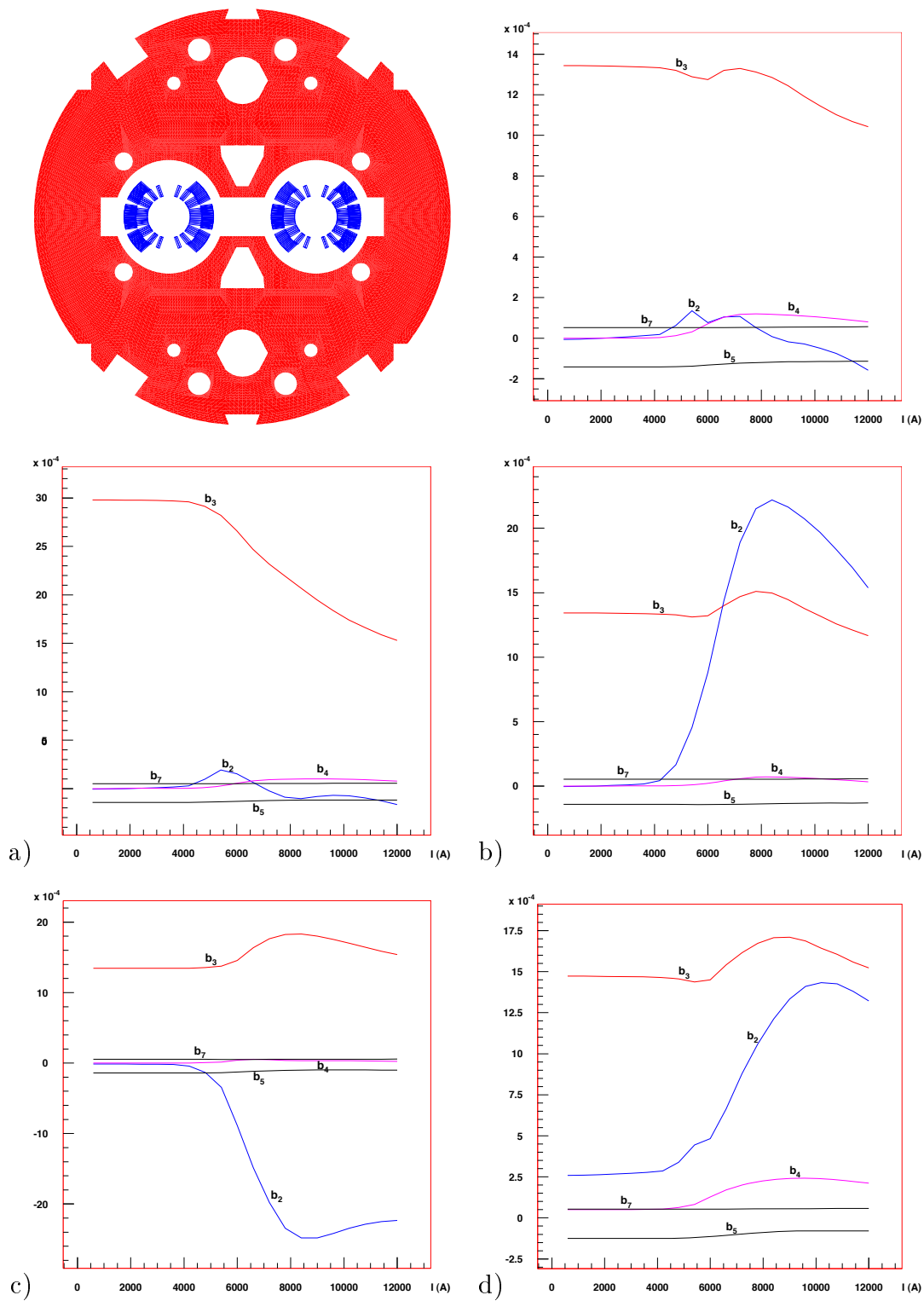


Figure 7.10: Variant 2: Modifying the features of this structure, the influence is evaluated of a) the circular collars, b) the outer hole, c) the center hole, and d) the outer notch. The multipole components are displayed over the excitation current in units of 10^{-4} at a reference radius of 17mm. Note the different scales of the images.

7.9 Comparison with the Yoke Design MBP2

Comparing the two alternative structures, several differences have to be discussed. The most important change to the actual geometry, as depicted in figure 7.1 is the much smaller collar radius in the examples. The fact that the iron is getting closer to the coils, increases the main field B_{12kA} at constant current or reduces the necessary current I_{nom} at constant field. Because of the lower current, a higher main field B_{ss} can be reached (tab. 7.4). A higher peak field results in a higher quench margin at constant current, reducing the danger of a quench. This improvement, however, comes with a price: The sextupole component is higher in both absolute value, and variation. A high absolute b_3 component can easily be counterbalanced by a similarly high sextupole component b_3 in the coil. This approach, nevertheless, would have to guarantee a precise positioning of the coil in the yoke, because of the high sensitivity of the quadrupole component b_2 due to a misplaced structure (sec. 3.8). All the other multipole components are well within tolerable limits.

	MBP2	Variant 1	Variant 2
I_{nom} (A)	11870	10890	11000
B_{ss} (T)	9.68089	9.88291	9.83344
B_{12kA} (T)	8.45456	9.10388	9.02284
b_{2inj}	0.66873	0.34527	-0.05541
b_{3inj}	5.77641	34.12452	13.43214
b_{4inj}	0.21824	0.08661	0.00064
b_{5inj}	-0.91453	-0.88874	-1.41863
b_{7inj}	0.63258	0.55376	0.52487
Δb_2	1.54156	1.13358	2.93956
Δb_3	0.41268	6.82818	3.01193
Δb_4	0.15851	0.23187	1.19655
Δb_5	0.00077	0.33300	0.28857
Δb_7	0.00294	0.03643	0.04203

Table 7.4: Comparison of the MBP2 design with the two variants found by genetic algorithms. The smaller collars of the variants result in a higher main field B_{ss} . The price to be paid is a higher multipole variation.

A local optimization continuing at this stage, has to proceed on estimates for the objectives. Such estimates can be gained from payoff tables. As discussed in section 4.4, each of the objectives is optimized separately and a utopian performance is found. With the values obtained, a constraint formulation for a local optimization can be set up. In table 7.5 this evaluation is done for the standard geometry (fig. 7.1) [48]. Subject to an optimization are only six independent geometrical parameters: The half axes of the collar ellipses $alel1$, $alel2$ and $blel1$, the radius of the upper hole $lh2r$, the starting angle of the center notch $leang$, and the height of the hood above the coils $houdh$. At the same time with the optimum, also the price for an optimized parameter can be retrieved from the table. For example, a reduction of the b_3 -variation Δb_3 results in a big increase of the b_2 -variation Δb_2 . Since the variation of b_3 is important in the design, minimization of the b_2 -variation is expensive. Both tables

show the payoff in a high b_3 component for a higher maximum achievable field B_{ss} , achieved by a closer yoke.

	B_{ss} ≥ 8.2	Δb_2 ≤ 4.0	Δb_3 ≤ 5.0	Δb_4 ≤ 1.0	b_{4inj} ≤ 0.1	alel1 (mm)	alel2 (mm)	blel1 (mm)	lh2r (mm)	leang (deg)	houdh (mm)
max B_{ss}	-9.73	0.758	1.464	0.086	0.103	75.84	76.08	95.0	7.22	22.14	4.04
min Δb_2	-9.66	0.036	0.551	0.077	-0.116	92.94	84.44	102.2	9.15	33.92	9.21
min Δb_3	-9.65	3.847	0.002	0.267	0.085	86.73	88.23	102.6	17.24	41.69	4.65
min Δb_4	-9.66	0.822	0.721	0.057	-0.047	86.39	95.94	105.0	7.03	34.26	11.99
min b_{4inj}	-9.65	1.265	0.306	0.114	0.000	88.52	89.02	103.5	13.43	39.75	6.91

Table 7.5: Payoff-table for the MBP2 magnet geometry. Six geometry parameters are evaluated for single objectives of the field quality. The combination of the bold-face numbers defines a utopian optimization aim.

In all optimization stages, the achievable objectives have to be seen in the context of all the design aims. While a higher main field is favourable for a higher beam energy, low and insensitive multipole components are indispensable for a working accelerator design. Two variants to the existing design were developed in this chapter, that reach a comparable field quality. Local optimization can further adjust the designs. These two examples highlight the conceptual design procedure. The ability to gain several ideas on the yoke shape in one genetic optimization run helps evaluate potential designs. The designer can combine these ideas and easily verify their quality. Objectively judging the results, several alternatives are derived. This process may therefore speed up and improve the magnet development.

Chapter 8

Conclusions

The Large Hadron Collider, the next particle accelerator at CERN, requires the design of some twenty different types of magnets, the biggest of which is the main bending dipole. The high accuracy of these magnets regarding the necessary main field and multipole field properties, justifies the intense research on these devices.

In order to improve the initial design of the superconducting magnets and to allow for quick re-evaluation of existing designs, genetic algorithms were adapted for the conceptual design phase. The ability to handle non-linear and discrete parameters in a global optimization space, favoured the use of this stochastic method. Niching, the concurrent maintenance of a number of solutions in the optimization run, proved to be a key factor in the algorithmic procedure. Crucial in this approach appeared to be the proper treatment of design variables and constraints, and the appropriate choice of objectives in the mathematical definition of the problem. The correct treatment of the vector optimization problem was found to be important. Certain objectives and constraints like manufacturability, force distribution, sensitivity, or quench behaviour are difficult to estimate in the mathematical procedure. The fuzziness of these objectives, their complexity in the evaluation, and their influence on the convergence, led to the neglect of such parameters in the problem formulation. Consequently, these parameters had to be checked after the optimization run.

The separation of coil and yoke design, because of their low interference, is a practical approach in magnet design. Both parts were therefore investigated individually, and a different technique was applied to each of them. The coil design, parametrized by the positioning and inclination angles, and the number of conductors of each block, defines the parameter space of this problem. Genetic algorithms appear to be appropriate for the optimization of such a non-linear, discrete, and global problem. The solutions have achieved a high main field while maintaining low multipole components. The application of this approach yields two alternative 6 block coil distributions. The two designs were compared to the 5 block coil and were shown to be superior. The decision to build model magnets with this new coil configuration was therefore taken by the LHC project management. The tests carried out on these models proved the quality of the design and achieved the predicted performance gain.

The other part of the magnet design, namely the yoke optimization, was treated as a material distribution problem rather than a shape optimization. The aim of such a formulation

is to deduce useful design ideas for the yoke structure. The region of interest was subdivided in small facets, consisting of finite elements, that may change their material properties from magnetic iron to non-magnetic material. The operation of genetic algorithms on bit-strings, made this application straightforward. The definition of the objective function appeared to be more complicated, however. In order to overcome jagged structures, which are neither manufacturable nor physically favourable, major modifications had to be made. The evaluation of the field parameters at multiple excitation levels, improved the material distribution. In an attempt to gain more design creativity, the yoke region was redefined as a rectangular shaped grid. The reduction of local saturation effects refined the results. From a number of designs, two representative examples were further investigated. The ideas, that could be retrieved, were assembled in feasible structures. Evaluation of the field quality showed the advantages of this approach to material distribution problems. High computation time of finite element methods and the high number of necessary function evaluations, however, currently constrain the discretization of the material regions. With the increase in computation speed, a further subdivision of the elementary regions should become possible, promoting the relevance of the method as a new approach to yoke design.

Nomenclature

A, A_c	Area, area for integration over conductor position	$[\text{m}^2]$
A_z	z-component of the magnetic vector potential	$[\text{Tm}]$
\vec{A}_r	Reduced vector potential	$[\text{Tm}]$
A_n	Skew field components	$[\text{Tm}^{1-n}]$
a_n	Skew relative field components at reference radius	$[10^{-4}]$
B_n	Normal field components	$[\text{Tm}^{1-n}]$
b_n	Normal relative field components at reference radius	$[10^{-4}]$
B, \vec{B}	Magnetic field (induction, flux density)	$[\text{T}]$
B_{ss}	Short sample field	$[\text{T}]$
B_1	Main dipole field	$[\text{T}]$
B_{cc}	Magnetic field at constant current	$[\text{T}]$
B_r	Reduced magnetic field	$[\text{T}]$
B_s	Magnetic source field	$[\text{T}]$
B_s	Critical field at zero current	$[\text{T}]$
C_n	Complex multipole components	$[\text{Tm}^{1-n}]$
$C_n^{(USA)}$	Complex multipole components, US notation	$[\text{Tm}^{-n}]$
\overline{C}_n	Complex multipole components of mirror currents	$[\text{Tm}^{1-n}]$
C_n^M	Complex multipole components of multipolar symmetric structure	$[\text{Tm}^{1-n}]$
$F(z)$	Complex flux density	$[\text{T}]$
$F(\vec{x})$	Objective function	
\vec{F}	Objective function vector	
F_p	Component p of objective function vector	
f_k	Shape function k	$[1]$
$f(\vec{x})$	Fitness function	
G_n, H_n	Field components of the magnetic vector potential	$[\text{Tm}^{1-n}]$
g_i	Inequality constraint i	
H, \vec{H}	Magnetic field (excitation)	$[\text{Am}^{-1}]$
H	Hessian matrix	
h_i	Equality constraint i	
I_k	Current of differentially small current line k	$[\text{A}]$
J, \vec{J}	Current density	$[\text{Am}^{-2}]$
J_0	Constant current density	$[\text{Am}^{-2}]$
J_s	Critical current density at zero magnetic field	$[\text{Am}^{-2}]$
J_{ss}	Short sample current	$[\text{A}]$
j	Imaginary unit	$[1]$

L	Self inductance	$[\text{VsA}^{-1}]$
M, \vec{M}	Magnetization	$[\text{Am}^{-1}]$
\vec{n}	Normal unit vector	$[\text{m}]$
R	Inner radius of iron yoke	$[\text{m}]$
r	Radius	$[\text{m}]$
r_{ref}	Reference radius	$[\text{m}]$
V	Scalar magnetic potential	$[\text{Tm}]$
\vec{w}	Weighting function	$[1]$
w_p	Weight p	
x, y, z	Coordinates	$[\text{m}]$
\vec{x}	Design variable vector	
x_k	Design variable k	
x_{kl}	Lower bound of design variable x_k	
x_{ku}	Upper bound of design variable x_k	
y_p	Residual p	
z	Complex position	$[\text{m}]$
z_c	Complex position of conductor filament	$[\text{m}]$
α	Inclination angle of coil block	$[\text{rad}]$
δ	Dirac Delta Function	$[1]$
Γ_{12}	Domain boundary	
Γ_B	Dirichlet domain boundary	
Γ_H	Neumann domain boundary	
μ_0	Magnetic permeability of vacuum	$[\text{VsA}^{-1}\text{m}^{-1}]$
μ_r	Relative magnetic permeability	$[1]$
μ	Magnetic permeability of material	$[\text{VsA}^{-1}\text{m}^{-1}]$
Ω	Region	
Ω_a	Air region	
Ω_i	Iron region	
Φ	Complex vector potential	$[\text{Tm}]$
φ	Positioning angle of coil block	$[\text{rad}]$

Bibliography

- [1] L. R. Evans. LHC accelerator physics and technology challenges. In *EPAC'98, Stockholm*, June 1998.
- [2] *CAS - Superconductivity in Particle Accelerators*, May 1996.
- [3] N. I. Andreev, K. Artoos, T. Kurtyka, D. Oberli, L. Perini, S. Russenschuck, N. Siegel, A. Siemko, D. Tommasini, I. Vanenkov, and L. Walckiers. State of the short dipole model program for the LHC. In *EPAC'98, Stockholm*, June 1998.
- [4] *The Large Hadron Collider, Conceptual Design*, CERN/AC/95-05. CERN, European Organization for Nuclear Research, 1995.
- [5] S. Russenschuck, editor. *ROXIE - Routine for the Optimization of Magnet X-Sections, Inverse Field Computation, and Coil End Design*. CERN, 1999.
- [6] S. Russenschuck, F. Calmon, M. Lewin, C. Paul, S. Ramberger, R. Rodriguez-Mateos, T. Tortschanoff, A. Verweij, and R. Wolf. Integrated design of superconducting accelerator magnets: a case study of the main quadrupole. *The European Physical Journal – Applied Physics*, 1(1):93–102, January 1998.
- [7] R. Wolf. Field error naming conventions for LHC magnets. Technical Report LHC-MMS/ES/01, CERN, June 1998.
- [8] R. Perin. Field, forces and mechanics of superconducting magnets. In *CAS - Superconductivity in Particle Accelerators*, May 1996.
- [9] C. P. Bean. Magnetization of hard superconductors. *Physical Review Letters*, 8(6):250–253, March 1962.
- [10] R. Wolf. Persistent currents in LHC magnets. *IEEE Trans. Magn*, 28(1):374–377, January 1992.
- [11] L. Bottura, M. Schneider, L. Walckiers, and R. Wolf. Cable magnetization effects in the LHC main dipole magnets. Technical Report LHC Project Report 169, CERN, July 1997.
- [12] V. Ziemann. Sorting the LHC dipoles using simulated annealing. In *4th European Particle Accelerator Conference: EPAC '94 London, UK*, June 1994.

- [13] R. Bartolini and W. Scandale. Sorting strategies for the LHC dipoles. In *1997 Particle Accelerator Conference: PAC '97 Vancouver, Canada*, May 1997.
- [14] K. Preis, I. Bardi, O. Biro, C. Magele, W. Renhart, K. R. Richter, and G. Vrisk. Numerical analysis of 3D magnetostatic fields. *IEEE Trans. Magn.*, 27(5):3798–3803, September 1991.
- [15] C. Paul, S. Russenschuck, N. Siegel, and K. Preis. Saturation induced field errors in the LHC main dipoles. In *MT-15, Beijing*, 1997.
- [16] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer J.*, 7:308–313, 1965.
- [17] H. G. Jacob. *Rechnergestützte Optimierung statischer und dynamischer Systeme*. Springer, 1982.
- [18] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Academic Press, 1981.
- [19] J. H. Holland. *Adaption in natural and artificial systems*. MIT Press, 1992.
- [20] D. E. Goldberg. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, 1989.
- [21] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag Berlin Heidelberg, 3rd, revised and extended edition, 1996.
- [22] R. Rojas. *Theorie der neuronalen Netze*. Springer, 1996.
- [23] F. Gray. Pulse code communication. Technical Report 2 632 058, U. S. Patent, March 1953.
- [24] T. Jones and S. Forrest. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In L. J. Eshelman, editor, *Proceedings of the 6th International Conference on Genetic Algorithms*, pages 184–192. Morgan Kaufmann Publishers, 1995.
- [25] D. E. Goldberg, K. Deb, and J. H. Clark. Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, 6(4):333–362, August 1992.
- [26] D. M. Himmelblau. *Applied Nonlinear Programming*. McGraw-Hill, 1972.
- [27] K. A. DeJong. *Analysis of the behavior of a class of genetic adaptive systems*. PhD thesis, Dept. Computer and Communication Sciences, Univ. of Michigan, 1975.
- [28] S. W. Mahfoud. *Niching Methods for Genetic Algorithms*. PhD thesis, University of Illinois, Urbana-Champaign, 1995.
- [29] S. W. Mahfoud. Population size and genetic drift in fitness sharing. In L. D. Whitley and M. D. Vose, editors, *Foundations of Genetic Algorithms*, volume 3, pages 185–223. Morgan Kaufmann, 1995.

- [30] D. E. Goldberg, K. Deb, H. Kargupta, and G Harik. Rapid, accurate optimization of difficult problems using fast messy genetic algorithms. In *ICGA 5*, pages 56–64, 1993.
- [31] L. Altenberg. The Schema Theorem and Price’s Theorem. In Darrell Whitley and Michael D. Vose, editors, *Foundations of Genetic Algorithms 3*, pages 23–49. Morgan Kaufmann, San Mateo, CA, 1995.
- [32] X. Qi and F. Palmieri. Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space. part i: Basic properties of selection and mutation. *IEEE Transactions on Neural Networks*, 5(1):102–119, January 1994.
- [33] M. D. Vose and G. E. Liepins. Punctuated equilibria in genetic search. *Complex Systems*, 5(1):31–44, February 1991.
- [34] G. Rudolph. Convergence analysis of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 5(1):96–101, January 1994.
- [35] S. W. Mahfoud. Finite Markov chain models of an alternative selection strategy for the genetic algorithm. *Complex Systems*, 7(2):155–170, April 1993.
- [36] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, April 1997.
- [37] D. Hagedorn and F. Rodriguez-Mateos. Modelling of the quenching process in complex superconducting magnet systems. *IEEE Trans. Magn.*, 1992.
- [38] J. Lucas. Private communication. CERN/LHC-ICP.
- [39] S. Ramberger and S. Russenschuck. Genetic algorithms with niching for conceptual design studies. *IEEE Transaction on Magnetics*, 34(5):2944–2947, September 1998.
- [40] S. Ramberger and S. Russenschuck. Genetic algorithms for the optimal design of superconducting magnets. In S. Myers, L. Liljeby, Ch. Petit-Jean-Genaz, J. Poole, and K.-G. Rensfelt, editors, *EPAC 98 – Sixth European Particle Accelerator Conference*, pages 2014–2016. Institute of Physics, November 1998.
- [41] S. Russenschuck. Comparative study of different coils for the LHC main dipoles. LHC-Project-Report 159, CERN, October 1997.
- [42] R. Ratnajeevan, H. Hoole, K. Weeber, and S. Subramaniam. Fictitious minima of object functions, finite element meshes, and edge elements in electromagnetic device synthesis. *IEEE Trans. Magn.*, 27(6):5214–5216, November 1991.
- [43] O. A. Mohammed, F. G. Üler, S. Russenschuck, and M. Kasper. Design optimization of a superferric octupole using various evolutionary and deterministic techniques. *IEEE Trans. Magn.*, 33(2):1816–1821, March 1997.
- [44] G. F. Üler, O. A. Mohammed, and S. K. Chang. Utilizing genetic algorithms for the optimal design of electromagnetic devices. *IEEE Trans. Magn.*, 30(6):4296–4298, November 1994.

- [45] D. N. Dyck and D. A. Lowther. Automated design of magnetic devices by optimizing material distribution. *IEEE Trans. Magn.*, 32(3):1188–1193, May 1996.
- [46] J. Ahlbäck, J. Ikäheimo, J. Järvi, D. Leroy, L. Oberli, R. Perin, D. Perini, S. Russenschuck, J. Salminen, M. Savelainen, J. Soini, and G. Spigo. Electromagnetic and mechanical design of a 56 mm aperture model dipole for the LHC. Technical Report AT/93-38, CERN, October 1993.
- [47] T. Shintomi, A. Yamamoto, Y. Doi, T. Haruyama, N. Higashi, H. Kawamata, S. W. Kim, N. Kimura, V. Kovachev, T. Nakamoto, N. Ohuchi, T. Ogitsu, K. Tanaka, A. Terachima, K. Tsuchiya, R. Perin, D. Leroy, S. Kato, K. Makishima, T. Orikasa, and A. Tanaka. Development of a 56 mm aperture superconducting dipole model magnet for LHC. *IEEE Trans. Appl. Supercond.*, 7(2):558–561, June 1997.
- [48] Christine Völlinger. Design of yoke cross-sections for LHC main dipoles using numerical field calculation and mathematical optimization techniques. Master’s thesis, Technische Universität Berlin, September 1998.