Novel Beam Optics of Achromatic Superconducting Gantries for Proton Therapy

INAUGURAL-DISSERTATION

ZUR ERLANGUNG DES DOKTORGRADES
DER MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
DER HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF

VORGELEGT VON

ANTHONY MICHAEL HUGGINS
AUS COLUMBIA, SOUTH CAROLINA

DÜSSELDORF, AUGUST 2020

AUS DEM INSTITUT FÜR LASER- UND PLASMAPHYSIK
DER HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. PROF. DR. ANTONY J. LOMAX
   Zentrum für Protonentherapie, Paul-Scherrer-Institut, Villigen
   Department of Physics, Medical Physics, ETH Zürich

2. PROF. DR. GEORG PRETZLER
   Institut für Laser- und Plasmaphysik, HHU Düsseldorf

Tag der mündlichen Prüfung:

Novel Beam Optics of Achromatic Superconducting Gantries for Proton Therapy

by
Anthony Huggins

To Jimmy and Jessica

PREFACE

The research for this thesis was conducted during my employment with Varian Medical Systems Particle Therapy GmbH, and mostly during my assignment to the joint collaboration project[1] between Varian, the Paul Scherrer Institute and Lawrence Berkeley National Laboratory (LBNL). While the subject of the project was the development of a new superconducting and achromatic gantry magnet, the underlying question was really: What can a next-generation proton therapy gantry look like? And if its future is superconducting, what are the best possible superconducting magnet and optics designs?

As such, more aspects than the optics alone had to be considered or at least put in perspective. Albeit this text focuses on the scientific and novel part of the optics of the superconducting achromats, it also mentions today's proton therapy technology, market situation and clinical best practices as well as derived requirements. If new proton therapy gantry concepts shall become reality, they must bring advantages beyond beautiful optics and exciting technology but must provide benefits for the treatment of cancer and the economics thereof.

Due to this background, and the inter- / multidisciplinary nature of the subject, the thesis also comprises a few more explanatory chapters. Since some back-and-fourth jumping between the various aspects cannot be avoided, a glossary is added where many terms are briefly defined.

Furthermore, some relevant parts of the magnet technology are described, particularly in chapter 4. The magnet design regarding both, the *AGCCT* and the *fixed-field magnet*, is the work of Lucas Brouwer and the Supercon group at LBNL, not mine. Some other pieces, especially in context of the conception of the *fixed-field magnet* are the results of group efforts, mostly with Weishi Wan and Lucas, as well as of inputs from all other project group members.

Finally, the educated reader may skip chapters 2 (beam optics), 3 (proton therapy) and/or 4.1 to 4.5 (superconducting magnets). Actually, jumping right to the results, discussion and conclusion may spare the pain of going through all of the (somewhat lengthy) text – a very proton thing to do!

[1] Project team and involved people: LBNL: Soren Prestemon, Adrian Hodgkinson, Ray Hafalia, David Robin, Shlomo Caspi, Lucas Brouwer, Scott Myers, Weishi Wan. Varian: Michael Schillo, Roberto Totato, Arno Godeke, Vladimir Anferov, Jürgen Heese, Rob Nast. PSI: Mike Seidel, Marco Schippers, Stéphane Sanfilippo, Ciro Calzolaio, Alexander Gerbershagen, David Meer and Tony Lomax.

ACKNOWLEDGEMENTS

This thesis would not have been feasible without the support from Varian Medical Systems Particle Therapy GmbH. Varian has many helpful and knowledgeable people from whom I benefited during my studies. Therefore, I would like to explicitly thank Wolfgang Kaissl, Jürgen Heese, Michael Schillo, Arno Godeke, Martin Wegner, Jürgen Schultheiss, Manuel Benna, Vladimir Anferov and Rob Nast as well as all of my other former colleagues at Varian. Martin Wegner directly contributed to this thesis by helping with coding of the post-processing tool and Jürgen Schultheiss provided valuable inputs for the mechanics of the gantry as well as the always useful CAD drawings. I also appreciate the help by Rob Nast regarding the evaluation of the cooling needs of the *fixed-field magnet*.

Essential for this project was also the supportive and very much appreciated supervision of the thesis by Prof. Tony Lomax from PSI and ETH Zürich and by Prof. Georg Pretzler from the Heinrich-Heine University Düsseldorf. Thank you for accepting this project and the trouble that comes with it as well as your helpful guidance and advice.

Further, I would like to highlight the contributions and help of two special people and friends: From Prof. Weishi Wan, from ShanghaiTech University, who was until recently at LBNL, I learned most of what I know regarding beam optics and the tool *COSY INFINITY*. Lucas Brouwer, magnet scientist at LBNL, has been an extremely helpful and creative co-worker in our joint project. He and Weishi Wan are co-inventors of the *fixed-field magnet* discussed in this thesis. Thank you, Weishi and Lucas, for the countless discussions we had, all your explanations and teaching me so much about superconducting magnets and accelerator physics. Thank you also for your hospitality during my many visits and your friendship. It has always been fun with you in- and outside the office.

Also from LBNL, I would like to thank Soren Prestemon, for providing access to and support at the Lab during my visits as well as for his advice and helpful suggestions. Soren Prestemon is the principal investigator of the *AGCCT* project.

I am grateful for the many fruitful discussions with the colleagues from PSI and their important contributions to the project. Thank you, Prof. Marco Schippers, Mike Seidel, Ciro Calzolaio and Alex Gerbershagen (now at CERN).

Much appreciated were also the discussions with my former Varian colleagues Jörg Wulff and Alexander Winnebeck.

Lastly and most importantly, I am truly thankful for the love of my family, first and foremost my wife Sally, daughter Jessica and little son James, but not forgetting the unconditional support from my parents, brothers and sisters and grandparents: John, Katrin, Emily, Sophia (Horlacher), Benedikt, Jonas, Prof. Robert and Patricia Huggins.

ABSTRACT

Proton beam therapy can be advantageous for cancer treatment due to the depth dose curve of protons described by the Bragg peak. Because of the relatively large mass of protons, the equipment for proton therapy is, however, much larger and more expensive compared to conventional radiotherapy devices based on linear electron accelerators. Superconducting (SC) magnets have been identified to help reduce the size of proton therapy gantries for their high magnetic fields and small bend radii. Due to low thermal margins, one of the most challenging issues of SC magnets for their application in ion therapy beamlines is their inability to quickly ramp the magnetic field – usually an important requirement for therapeutic beam delivery. Two different designs of superconducting combined function magnets, the *alternating gradient canted cosine theta* magnet *(AGCCT)*, and the so-called *fixed-field magnet* are studied with respect to their optical properties as local achromats, facilitating the transport of multiple beam energies without changes in field.

Based on the local achromat feature of the magnets a simplified and optimized gantry beamline for proton therapy is developed and discussed. Both magnets were optimized for their employment in the proposed gantry concept and use for therapeutic proton beam application addressing energy and emittance acceptance, focusing properties as well as weight, size and cryogen-free cooling considerations. The gantry optics of dipole and quadrupole fields were studied with a 1st order lattice code in *Transport*, while the higher order multipole fields of the achromats were developed and analyzed with the code *COSY INFINITY* based on transfer maps and differential algebra.

The *AGCCT* is a curved version of the *CCT* design with additional winding layers to produce alternating quadrupole and sextupole fields. The results of 5th order tracking studies show that the *AGCCT* achieves a momentum acceptance of 16.5 % *dp/p* of proton beams with clinical properties regarding emittance and energy spread. The momentum acceptance is mainly limited by the radius of the warm bore of the magnet. Deviations from clinically acceptable beam properties were found with respect to focusing for large momentum deviations, leading to elliptical profiles, as well as positional shifts throughout the energy range. The feasibility of corrections of all unwanted effects by means of trim quadrupole coils is demonstrated. However, a comparison with clinical treatment energy range statistics revealed that the momentum acceptance of the *AGCCT* is insufficient to significantly alleviate ramping requirements, making the magnet less favorable for an application in a proton therapy gantry.

The *fixed-field magnet* is a double-bend achromat of two straight racetrack coil designs with a c-shaped iron yoke and additional resistive quadrupoles. The two-dimensional field of the magnet facilitates a momentum acceptance over the entire clinical energy range from 70 to 225 MeV which is demonstrated with 7th order tracking. It is further shown that the beam properties at isocenter meet the clinical requirements throughout all energies. The feasibility of static field operation has significant economical and practical implications for conductor choice, cryostat design and operation of the magnet.

Key words: Proton therapy, beam optics, compact gantry, high momentum acceptance, superconducting achromat.

TABLE OF CONTENTS

GLOSSARY

Fraction	The portion of a total treatment dose which is applied during one treatment session on one day. A typical radiotherapy treatment comprises 20 and more fractions. Typically, 5 to 6 fractions per week are applied.
Global achromat	A beam optics concept where all elements of the beamline combined function as an achromat, i.e. at the endpoint the dispersion is zero. In-between bends of the beamline the dispersion function is non-zero.
Hypofractionation	A therapeutic concept which allows larger than typical doses per fraction, e.g. more than 2 Gy, and thereby reduces the number of fractions and the overall duration of the treatment.
Isocenter	A fixed reference point in space to which certain values are calibrated. The treatment isocenter is the reference point for any treatment relevant properties for the proton beam or mechanical motion of gantry and treatment table.
Isochronous cyclotron	A cyclotron which works with a non-varying acceleration radiofrequency. In order to compensate the increase of mass of the accelerated relativistic particles, isochronous cyclotrons increase the magnetic field with the radius (independent of time).
Layer (CCT magnet)	A layer of a CCT magnet is one toroidal coil of the CCT. It consists of a metal mandrel (tube shaped) with milled groves to hold the canted conductor windings. Two layers with different diameters, slid over one another, complement each other.
Layer (treatment field)	A layer of a *treatment field* is one transversal plane of the treatment volume to be irradiated with the scanning method and a given proton beam range. The 3D target is irradiated with multiple stacked layers (see the description of “pencil beam scanning” in section 3.1.6). During the irradiation of one layer the beam energy is held constant.
Local achromat	An optics concept for a beamline bend that corrects the dispersion. In-between multiple local achromats of a beamline the dispersion function is zero.
Magnet set point	Values of a magnet (e.g. the power supply currents) that refer to a *reference energy*. E.g. the current I that refers to the dipole field strength which relates to protons of energy E via the magnetic rigidity relation, see equation (2.11).
Momentum / energy spread	The distribution of momenta (or energies) within the proton beam. The relative values given in % refer typically to 1 or 2 RMS of an assumed Gaussian distribution of these momenta (energies).

Momentum acceptance	The relative range of proton momenta that are correctly transported by an achromat. E.g. +/- 2 % *dp/p* means that protons up to +2 and down to -2 % of deviating momenta are still transported correctly. For the AGCCT the deviating momenta refer to the mean momentum of a proton beam.
Momentum deviation / offset	The relative deviation or shift of a mean proton beam momentum to the momentum which corresponds to the *magnet set point* or *reference momentum*.
Quench	An event which leads to the loss of the superconducting state in a conductor.
Ramping	The process of changing the current in an electromagnet's coil to change the field strength.
Reference trajectory	The ideal particle trajectory of a beam optics that corresponds to the *reference energy / momentum*. In most cases the reference trajectory is the same as the optical axis or orbit.
RMS	The standard deviation of a (Gaussian) distribution, often also called sigma (σ).
Section (CCT magnet)	A part of a CCT magnet that has been divided longitudinally into *sections* that differ in their magnetic properties. The five sections of the *AGCCT* differ with regards to the multipole *B*-fields.
Set / reference energy or momentum	The energy or momentum which corresponds to the *set point* of a magnet.
Synchro-cyclotron	A cyclotron with time-varying acceleration radiofrequency in order to compensate the increase of mass of the relativistic particles. Synchro-cyclotrons can only accelerate one bunch of particles at a time leading to a pulsed beam structure.
Treatment couch / table	A moveable robotic table to hold and position the patient (usually in supine or prone position) during the treatment.
Treatment field	The part of a treatment fraction that is applied under a given gantry angle. Beam application of multiple fields under different gantry (incidence) angles combined deliver the dose of one treatment fraction.

1 INTRODUCTION

1.1 Treating cancer with protons

1.1.1 The significance of cancer

The relevance of cancer is on the rise. The increase in life expectancy and the medical successes in the cure of many other diseases, moved cancer to the second leading cause of death, being responsible for 1 of 6 deaths worldwide [1]. Annual cancer incidence is expected to raise to about 23.6 million globally by 2030, almost 70 % more than 2012 (14.1 million) [2]. Hence the cancer burden weighs heavy, and it is growing.

Cancer occurrence is even more pronounced in highly developed countries where the population is older and higher standards in medical care have decimated many diseases from commutable, nutritional, maternal or perinatal conditions [3, 4]. In Germany, the probability for a 45-year-old woman or man to get cancer at some point in her or his life is 42 or 51 %, respectively. Their respective risk to die from the disease is about half as large with 20 and 26 % [5]. Incidence rates are strongly depending on age [3, 4, 5], which means that, luckily, only about 1 out of 100 of the exemplary 45-year-olds will die of cancer within the next 10 years.

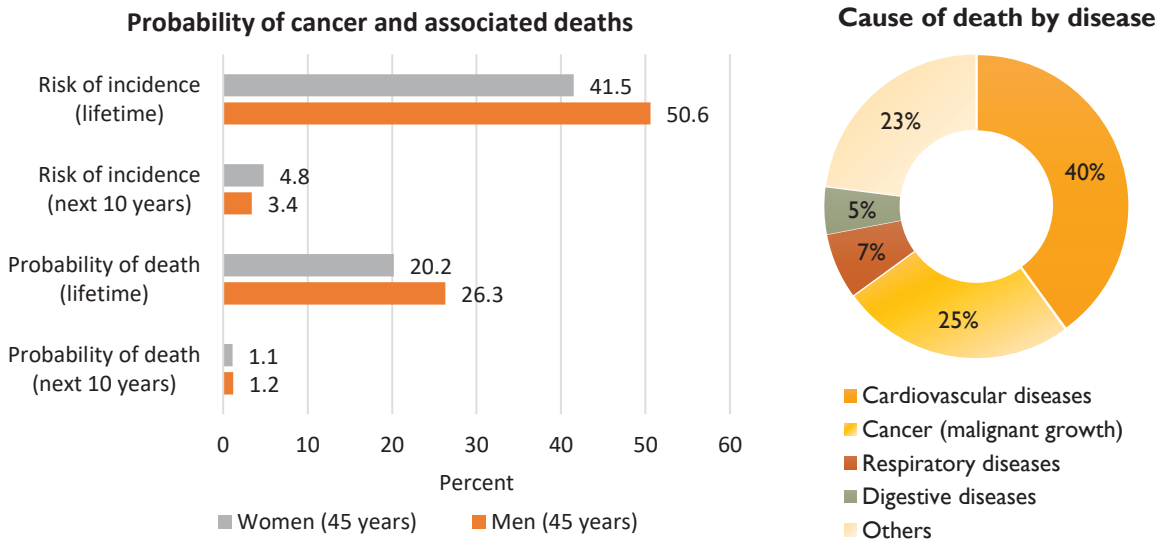

Figure 1.1: *Statistics from the cancer report of the Robert-Koch-Institut [5]. The data refers to Germany in the year 2013.*

Cancer is a random occurrence of cell mutation, so-called malignant neoplasms, which can spread throughout the body. Damaged DNA of any body cell, if not correctly repaired, can lead to new, typically dysfunctional, types of cells that may increase by amitosis without control and form a tumor [6]. The space and energy consumption of the proliferating cells weakens the body and becomes life threatening, either directly by affecting vitally important organs or indirectly, e.g. when patients lack the strength to overcome infections [6].

1.1.2 The fight against cancer

The fight against cancer has many aspects in prevention, diagnostics and therapy and many advancements have been achieved in all these fields [3] by

- identifying preventable causes (e.g. carcinogenic materials/food, hepatitis and other infections or environmental and behavioral causes like smoking, obesity or sun bathing),
- earlier diagnoses (e.g. improved medical imaging and biopsy methods, screening programs or knowledge of inheritable genetic defects) and
- improved treatments (e.g. new technologies and pharmaceuticals as well as a better understanding of curative effects).

The three established treatment options are:

1. Surgery: The physical removal of the tumor.
2. Chemotherapy: The body-wide treatment with pharmaceutical agents. Often the only option for metastasized tumors.
3. Radiotherapy: The destruction of tumor cells with ionizing radiation.

Radiotherapy, often in combination with surgery and/or chemotherapy, is involved in approximately 50 % of treatments [7]. Its underlying mechanism is the ionization of molecules which leads to chemically aggressive radicals. These radicals can kill tumor cells by damaging their DNA. However, the major problem of ionizing radiation – and why it is dangerous – is that, in principle, healthy body cells are just as vulnerable. Moreover, DNA damage not only leads to cell death, it can also trigger cell mutation via incorrect DNA repairs. Hence treatments with ionizing radiation potentially cause secondary cancers [8]. This risk and other radiation side effects motivate the most important goal of radiotherapy: tumor control while minimizing the irradiation of healthy tissue. Basically all developments in the field address this essential problem.

1.1.3 Proton therapy

The radiation is most commonly applied externally (teletherapy) by linear electron accelerators which produce hard x-rays (photons) via bremsstrahlung. The treatment with hard x-rays is also referred to as conventional radiotherapy. Proton therapy, the context of the following studies, is a special form of radiotherapy. Instead of photons, accelerated protons, directly applied to the tumor, serve as the ionizing radiation. The amount of cell damage corresponds to the amount of ionization, which again is proportional to the energy deposited in a given mass. The deposited energy is measured in Grays (J/kg) and called dose. The advantage of protons is their depth dose curve which exhibits the *Bragg peak* as demonstrated in **Figure 1.2**. In this way, healthy tissue proximal and distal to the tumor can be spared. Hence the risk of side effects, especially long-term effects (secondary tumors), is reduced [9]. The depth of the *Bragg peak* depends on the kinetic energy of the proton beam. About 70 to 220 MeV, corresponding to a range of about 4 to 30 cm in water, are necessary to reach the tumor in most patients. Shallower depths are treated with a pre-absorber in the beam, i.e. a plastic sheet to reduce the residual range. The disadvantage of protons is their rest mass, which is three orders of magnitude larger than that of an electron. Consequently, accelerating and steering proton beams require much larger, heavier and more expensive equipment.

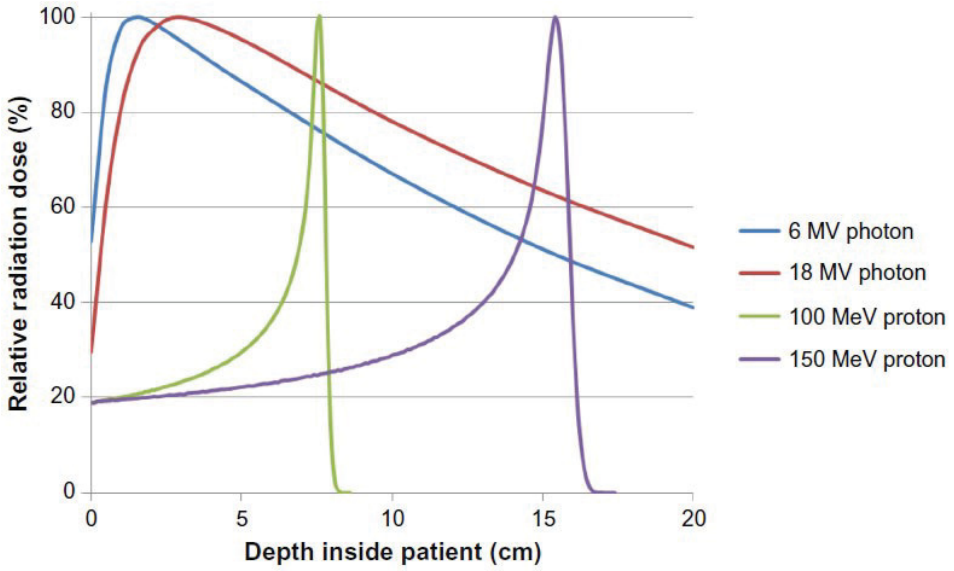

Figure 1.2: *Relative depth dose curves of proton and photon beams for selected energies [10]. Most of the dose (energy) of a proton beam is delivered at an energy-specific depth, which can be adjusted to the depth of the tumor. In contrast to photon beams, protons stop and deliver no dose beyond their range. The depth dose curve for photons exhibits a maximum near the entrance to the patient, thus deposit higher doses to healthy tissue in the proximal region.*

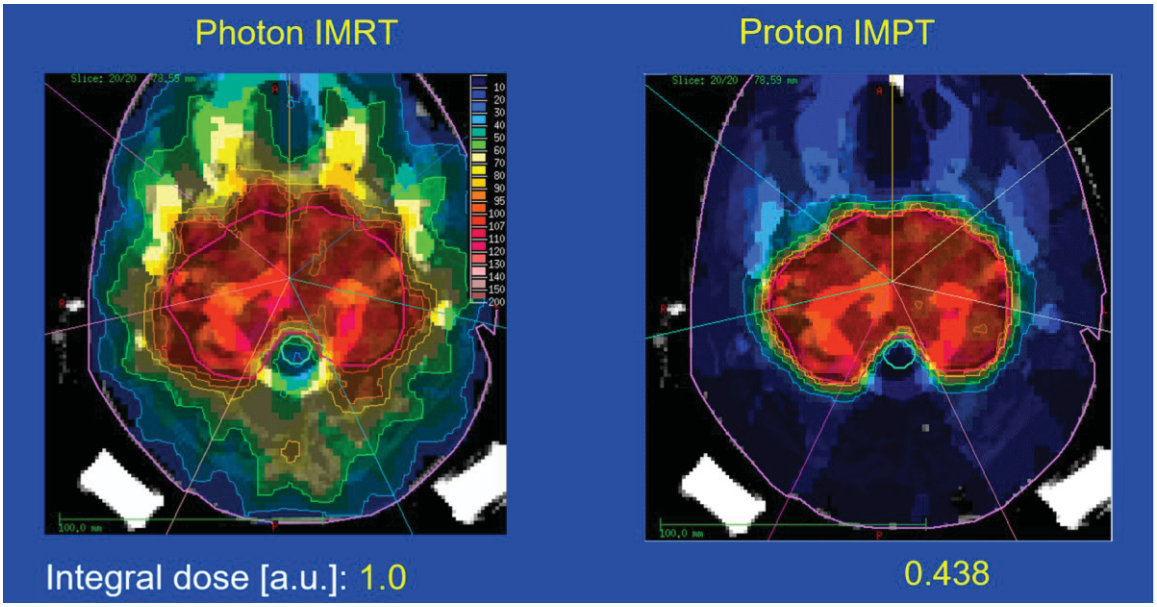

Figure 1.3*: Comparison of photon and proton intensity modulated treatment plans (planning CT, axial view) at the example of a clivus chordoma [11]. The pink contour marks the targeted area and the colors indicate relative iso-dose levels. Ideally the target area receives at least 100 % relative dose and the surrounding tissue zero. The proton plan not only shows significant better conformity but the integral dose is more than a factor of 2 smaller.*

In the broader sense one speaks of particle therapy as other ion species have also been used for therapy. Carbon beam therapy, particularly popular in Japan, benefits from an even sharper *Bragg peak* and a higher biological effectiveness [12]. But due to the larger mass to charge ratio, its equipment is even more costly. Outside Japan only four sites in Europe and two in China treat with ions other than protons [13].

1.1.4 The history of proton therapy

Robert R. Wilson is considered the pioneer of proton therapy as he first suggested their medical use for cancer therapy in 1946 [14]. First treatments started roughly ten years later [15]. However, for a long time cyclotrons with sufficient energy were only available in few parts of the world to treat patients. Notable sites were LBL[2], the Gustav-Werner Institute[3] and HCL[4] [15]. The Loma Linda University Medical Center in California opened the first dedicated proton therapy facility in a hospital environment in 1990. They also built the first proton therapy gantry, a proton beamline that is rotatable about the patient to apply beam from multiple angles (see chapter 3.1) [16]. Since then particle therapy expanded mainly in the US, Japan and Europe. Other important milestones were the first demonstration of the modern pencil beam scanning, the "painting" of the tumor with discrete narrow proton beams (details in section 3.1.6), at PSI[5] in 1996 [17] and the first clinical superconducting cyclotron development by ACCEL[6] based on the design proposal by Henry Blosser et al [18, 19]. It was first installed at PSI in 2005 [20] and meant the introduction of superconductivity in proton therapy.

1.2 Motivation

1.2.1 Towards compact systems (1st problem statement)

The commercialization of proton therapy and recent shift to single-room solutions

Despite its history, proton therapy is often viewed as a very young field because its more recent commercialization led to a strong increase in treatment centers in the past years. Between 2011 and 2017 the number of proton therapy facilities more than doubled as 39 new centers went operational. Yet, another 55 facilities are currently under construction. By the end of 2017 a total of almost 150,000 patients have been treated with proton radiation [13].

This fast growing market has attracted numerous vendors, namely IBA[7], Varian, Hitachi, Mevion, Pronova, Sumitomo, Protom, Mitsubishi, Toshiba and Advanced Oncotherapy. A very interesting list when compared to the only three suppliers for the much larger market of conventional radiotherapy (Varian, Elekta and Accuray). The most important markets are the United States, Europe, China and Japan. Whereas China is relatively new to the field and immensly investing in the typical large particle therapy centers with multiple treatment rooms (8 facilities with a total of 32 treatment rooms are under construction), the traditional big markets in the United States and Japan show a shift. There, a trend towards small single-room facilities with lower capacity and capital investment can be observed. All five current projects in Japan and nine of the fifteen planned facilities in the United States are single-room systems. This trend is somewhat less pronounced in Europe, where seven of the thirteen facilities under construction are of the multi-room type, but there is no doubt that the market share of single-room systems is growing rapidly, as demonstrated in **Figure 1.4**.

---

[2] Lawrence Berkeley Laboratory in California (now: Lawrence Berkeley National Laboratory, LBNL)
[3] In Uppsala, Sweden
[4] Harvard Cyclotron Laboratory in Massachusetts
[5] Paul Scherrer Institute in Villigen, Switzerland, previously „Swiss Institute for Nuclear Research"
[6] ACCEL Instruments, now Varian Medical Systems Particle Therapy, Germany
[7] Ion Beam Applications

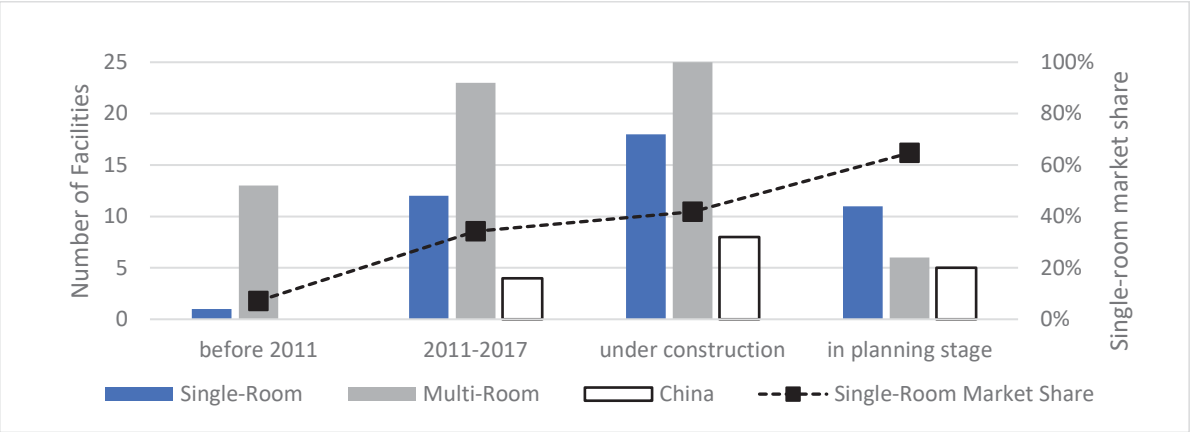

Figure 1.4: *Market share of single-room gantry systems without China. The black squares (dashed line to guide the eye) indicate the percentage of single-room systems of all facilities that went operational before and after 2011 as well as those currently under construction or being planned. Systems without gantries are excluded in the statistics. The number of facilities in China, all multi-room, are displayed by the empty columns [13].*

The main reason for the ermergence of the single-room systems – which are also called *compact systems* – is the lower "entry price-tag" to proton therapy and their reduced space consumption. However, also compact systems are still much bigger and significantly more expensive than conventional x-ray devices. The rotating gantry is the largest piece of equipment in proton radiation systems with typical dimensions of about 10 by 10 meters in length and diameter. Due to the need of heavy magnets and high stiffness requirements, the weight of a gantry is in the order of 100-200 tons. Its size particularly stands out in single-room systems and, together with its cost, has long been identified as a key obstacle for a faster spread of proton therapy [21, 22].

This leads to the first problem statement which motivates this thesis:

> The facilities are still big and expensive. Proton therapy could be made available to many more cancer patients if cost and size of the system were reduced.

Figure 1.5*: The compact proton therapy systems of Varian, called ProBeam® Compact. The two main components are the cyclotron on the right and the gantry on the left which rotates the beamline about the patient (not visible here, indicated with blue arrow). Both components are situated in a concrete bunker to shield the radiation. Source: Courtesy of Varian.*

A number of product developments in the industry, as well as design proposals from the science community, have been introduced in recent years to improve the size and cost of gantries, especially for single-room systems. Chapter 3 provides more background regarding these developments, compact systems, and proton therapy facilities in general.

1.2.2 The application of superconducting magnet technology (2nd problem statement)

Superconducting magnets can achieve much higher field strengths than resistive magnets thus present an opportunity for sharper bends of the beamline and thereby to reduce the size of the gantry (compare chapter 3.4). On top of that they do not necessarily rely on magnetized iron for field generation, which significantly reduces the weight of the magnets. These properties have already been applied to reduce the cost and size of cyclotrons, demonstrated by the developments of superconducting clinical cyclotrons by Varian, Mevion and IBA [20, 23, 24, 25].

However, there are also two major drawbacks to superconducting magnets:

1. They can only operate below a critical temperature T_C [115].
2. Secondly, it is rather difficult to rapidly change the field of a superconducting magnet, since that always implies heating of the conductor and structure, making it difficult to maintain a temperature below T_C (see chapter 4). As shall be seen, this second problem, also called *ramping*, is pivotal for any superconducting beamline magnet.

Varian, LBNL and PSI initiated a collaborative project in 2014 to explore the application of superconducting magnet technology for proton therapy gantries. The project focuses on the development of the *Alternating Gradient Canted Cosine Theta* magnet concept (short: *AGCCT*), an invention by LBNL [27, 28] made in the context of earlier beam optics concepts by PSI [29], and is partially funded by the Accelerator Stewardship program of the US Office of High Energy Physics, governed by the Department of Energy [30, 31]. The *AGCCT* design combines a bending dipole with alternating focusing quadrupoles to form a locally achromatic bend. Local achromatic bends achieve a bending of protons of different energies by the same angle. In other words, the bent beam is dispersion-free – in contrast to pure dipoles that act like energy filters, see **Figure 1.6**. While the *AGCCT* magnet design work is mainly driven at LBNL, the concept is simultaneously being studied at PSI where the development of a Nb_3Sn racetrack-based magnet system is pursued [62].

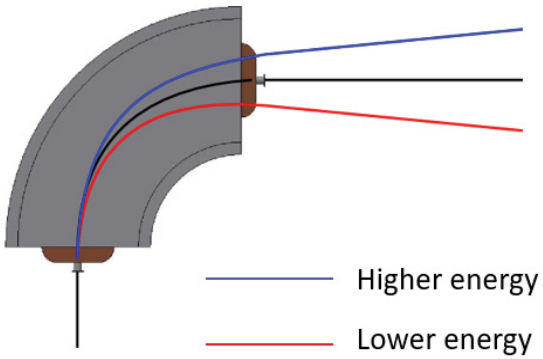

Figure 1.6: *Dispersion effect: A dipole acts like a proton energy filter.*

Such an achromatic feature is limited to a maximal energy acceptance, which describes to what extent the achromatic property is preserved. Within the energy acceptance window local achromats allow the application of multiple beam energies without changing the magnetic field, addressing problem number two above. They also make novel beamline designs feasible. The advantages of a proposed novel design, that aims at a simplification and size reduction of compact proton therapy gantries, are described in chapters 3.5 and 6.1. Details of the

corresponding magnet designs can be found in chapters 4.6 and 4.7 as well as in [27, 32, 33]. The differences of local and global achromats are described in section 3.5.1.

Consequently, the second problem statement that motivates the following studies can be formulated as follows:

> Can beam optics, utilizing the possibilities of superconducting magnets, achieve achromaticity with an *energy acceptance* large enough to significantly alleviate the ramping requirements for superconducting magnets in the context of clinical use?

1.2.3 The speed of beam delivery

An additional motivation for new gantry beamlines is the speed of beam delivery. Precise treatment is very sensitive to tumor motion, complicating the irradiation of targets in the vicinity of the lung or liver. There are different motion mitigation strategies:

- In the most established method, called *gating*, which was adopted from conventional radiotherapy, the motion is monitored and the beam is only applied when the target is in an acceptable window [34].
- In the concept of *volumetric repainting* the irradiation is repeated multiple times, each with a fraction of the dose, and thereby the applied protons are smeared out over the moving target which increases robustness [35, 36].
- A third and most promising approach to deal with respiratory motion are so-called *breath-hold treatments*. The simple idea is that if the beam delivery is fast enough, say applied within 10 seconds, the treatment can be completed while the patient holds his or her breath [37]. Today's proton therapy machines cannot provide this speed of treatment yet.
- For the sake of completeness, the method of *tumor tracking* shall be mentioned. The concept foresees that by tracking the tumor location, the beam can be actively steered to follow its motion. However, this approach brings many complications such as a change of the penetration depth due to the shifted proton beam path. Since this would require a real-time recalculation of the treatment plan, the concept is, to-date, not feasible.

1.2.4 The promise of local achromats

As shown later, local achromats allow beamline designs that are much more efficient, i.e. have a much higher transmission than state-of-the-art gantries. Hence the dose rate at the patient can be higher. If the energy acceptance of the achromat is large, changes in the beam energy can occur very quickly, avoid the need for magnet ramping and speed up the in-depth variation of the *Bragg peak.* Consequently, the beam application of a treatment can be accelerated, facilitating breath-hold treatments as well as volumetric repainting. The higher dose rate would possibly also be useful for hypofractionation – a trend to apply larger portions of doses per treatment fraction in proton therapy (see chapter 3.2).

1.2.5 Current developments in the field

Both the application of superconducting magnets as well as the concept of (local) achromats is not new: The ProNova SC 360 gantry also uses superconducting magnets in an achromatic double bend. There, superferric dipoles and quadrupoles are arranged sequently (instead of combined). The resultant energy acceptance, however, is not sufficient to alleviate the ramping

problem [38, 39]. To this date, no data regarding the performance or cost of the gantry design have been published. In fact, ProNova offers not a single- but a 2-room system. The company received FDA 510(k) approval for clinical use early 2018 [40] and successfully completed the first patient treatment using the SC360 in November 2018 [41].

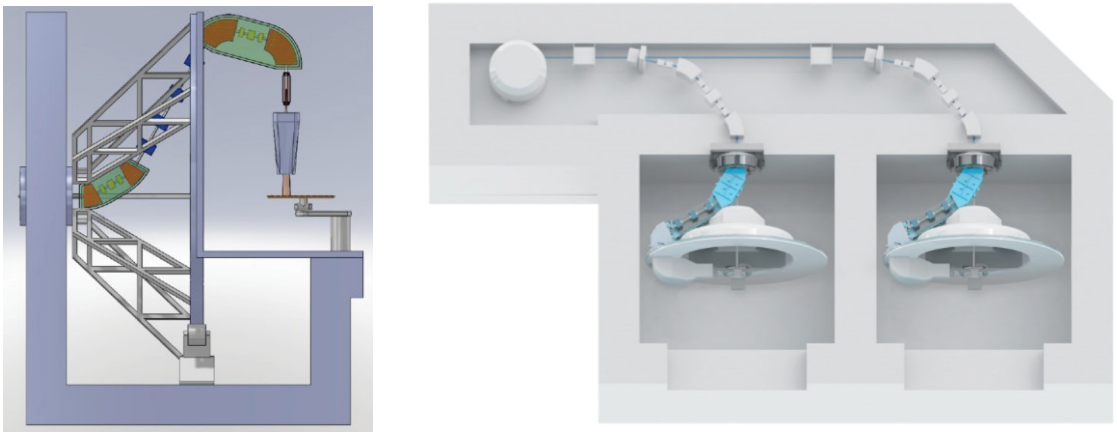

***Figure 1.7**: The ProNova SC360 superconducting gantry (left) and 2-room system (right). The cryostats housing the superconducting magnets, which each make an achromatic bend, are colored in green (left picture). Source: [39]*

Another proposal (that has already been made in 2005) by Trjobevic, Sessler et al is the fixed field alternating gradient (FFAG) gantry [42]. It is based on a continued sequence of alternating strong quadrupoles and derived from the FFAG principle known from accelerators [43]. It originally aimed to achieve an energy acceptance that covers the entire clinical energy range from 70 to 220 MeV and thereby facilitating the use of permanent magnets that minimize operational cost. However, more recent results imply this is not achievable [44]. The FFAG gantry is very large due to the large bend radius of the beamline, even exceeding the existing designs. Fenning et al addressed the size but found problems with beam stability [45] and their new design trades diameter for an enlarged length of the gantry. The FFAG gantry was mentioned in context of the PAMELA project (carbon and protons) [46], but since the study ended in 2011 no plans to actually build a FFAG gantry are known.

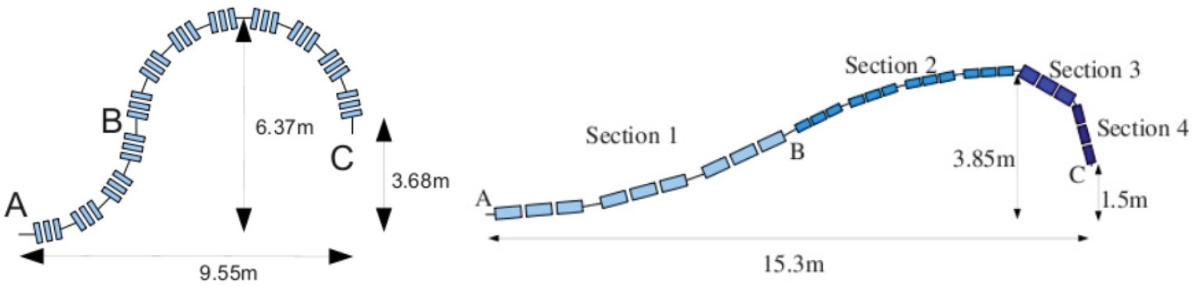

***Figure 1.8:** Schematic views of the FFAG gantry in the original design from Trbojevic et al (left) and a new design from Fenning et al (right). The improvement in diameter increases the length of the gantry to over 15 meters. Source: Fig. 1, 2 of [45].*

For carbon beam therapy, where the only existing gantry with resistive magnets at HIT[8] weighs 450 tons (rotating parts) and is impressive in length (25 m) and diameter (13 m) [47], a superconducting version has been built at NIRS[9] in Japan [48]. It consists of ten superconducting magnets (developed by Toshiba) and went clinical in May 2017 [49]. The

[8] Heidelberg Ionentherapie Zentrum (center for ion beam therapy in Heidelberg, Germany)
[9] National Institute of Radiological Sciences in Chiba, Japan

design, that indeed achieves a size reduction of approximately 50 %, is not transferable to proton therapy as the number of magnets (and their cooling) are not cost effective nor necessary due to the much smaller magnetic rigidity of protons. Carbon beam therapy is not commercialized yet and it will be interesting to see if the NIRS design or a derivate thereof will be implemented elsewhere: Toshiba announced the intention to develop a further downsized version [50]. Another superconducting gantry proposal aims at 330-350 MeV to facilitate proton radiography/tomography [53, 54, 55]. Since this is not related to any cost or space issues the idea is not discussed here.

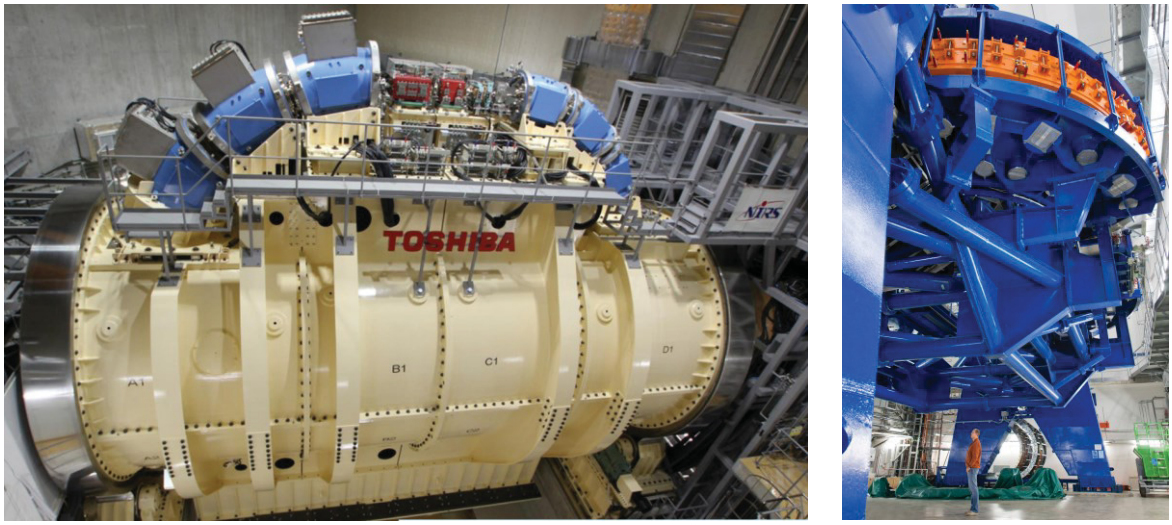

Figure 1.9: *The superconducting (left) and resistive (right) carbon gantries at NIRS [51] and HIT [52]. With a length of 13 and a diameter of 11 meters, the superconducting gantry is 12 meters shorter and 2 meters slimmer than the HIT gantry.*

1.3 Scope of the thesis

The goal of the joint LBNL-Varian-PSI project is to utilize the stronger fields of superconducting magnets in general, as well as the unique advantage of the *AGCCT* concept, that allows for a combined function magnet with alternating quadrupoles. The achromatic property, facilitated by the latter, shall be exploited to alleviate the requirement for fast field changes. In a sense, this unique feature tackles a self-made problem, one that only resulted from choosing superconducting magnets in the first place.

The scope of this thesis comprises three studies:

1. Exploration and optimization of the beam optics of the *AGCCT:* Results of the optimization defined the magnet design, which in turn was then analyzed regarding its optics properties as an actual result. These results were put in context of clinical operation.

During the course of the *AGCCT* project, as the design was refined and prototyping began, it became apparent that most of the cost and complications originate from the remaining need for magnet ramping, which could not be eliminated despite the achromatic feature. However, as a converse argument, an achromatic magnet with an energy acceptance as large as the entire clinical range – if achievable – would have such significant impact on cost and performance of the system, it can be viewed as a “holy grail”. Whereas the energy

acceptance of the *AGCCT* concept is just not large enough, some explorative thinking was spent in that direction and indeed, in a group effort, an idea for a beam optics concept, that would fulfill this requirement, was conceived[10]. The resulting new design is called the *fixed-field magnet* and is described in chapter 4.7.

2. The second part of the thesis analyzes the beam optics of a preliminary design of the *fixed-field magnet.* The properties of the system over the clinical energy range are evaluated and compared with typical clinical requirements.

Local achromats in general – including both the *AGCCT* and the *fixed-field magnet* – can be used to simplify the beamline of a gantry and to improve its performance with regards to the transmission, the ratio of transported versus lost protons in a beamline.

3. The design of a simplified gantry beamline maximizing the benefits of the two magnets is studied. It is further discussed in context of other compact gantry designs and performance requirements of the future.

1.3.1 Overview of the following chapters

The following text covers the most important results of the beforementioned three studies, starting with some theoretical basics regarding beam optics in chapter 2, particularly the concept of transfer maps for beam transport. Some (mainly technical) aspects of proton therapy facilities, the context of this study, are described in chapter 3 and chapter 4 goes into some relevant details of superconductivity and magnets. Therein, also the design of the *AGCCT* is explained as well as methods that led to the *fixed-field magnet* concept. The magnet baseline design for the corresponding beam optics analysis is shown 4.7. Chapter 5 summarizes the methods employed for the treatment of proton passage through magnetic fields. The results of the studies for the *AGCCT,* the *fixed-field magnet* and the proposed gantry design are presented in chapter 6 and discussed in their respective context. Finally, chapter 7 draws a conclusion and gives an outlook on recommendable next steps.

[10] Patent application submitted. Inventors are: L. Brouwer, W. Wan and A. Huggins

2 THEORY AND BASIC PHYSICS

2.1 Proton interactions with matter

2.1.1 Stopping Power and Range

As shown in the introduction protons of a given kinetic energy stop at a corresponding range, exhibiting the Bragg peak. Apart from the kinetic energy, the properties of the material penetrated by the proton influence its range. The stopping power S is material and particle dependent, both particle species and velocity, and describes the rate of loss of kinetic energy E along the path x of the particle. In other words, the stopping power of a material is the kinetic energy a charged particle loses per unit of path length as it penetrates the material. It is usually measured in MeV/cm and defined as

$$S \stackrel{\text{def}}{=} -\frac{dE}{dx} \tag{2.1}$$

The interaction of protons with matter were described by Bethe and Bloch in 1933 [56]. The mass stopping power S/ρ given in the Bethe formula is the stopping power normalized for material density:

$$\frac{S}{\rho} = 0.3072\ \frac{Z}{A}\ \frac{1}{\beta^2}\left(\ln\frac{W_m}{I} - \beta^2\right)\ \frac{\text{MeV}}{\text{g/cm}^2} \tag{2.2}$$

where Z and A are atomic and mass number, $\beta = v/c$, I the mean excitation energy and the maximum energy loss during a single collision with a free electron is

$$W_m = \frac{2\, m_e c^2 \beta^2}{1-\beta^2}. \tag{2.3}$$

However, the mean excitation energy I is an experimental value. Hence it is most convenient to obtain the (mass) stopping power or the range experimentally. The database from NIST[11] for example provides values for many materials [57].

The use of correct values for the stopping power is a delicate part of treatment planning algorithms. From CT images Hounsfield units of tissue (x-ray attenuation) must be translated to stopping power to accurately calculate the range of the proton beam in the human body. Since there is no analytical relation, calibration methods determine the stopping powers. Precision and errors in these calibrations are frequently discussed topics in the proton therapy community [58].

The physical process behind the deceleration of protons described with the stopping power is the electromagnetic interaction of the charged particle with the electron shell or nucleus of atoms (for protons the latter is irrelevant for the energy range of interest, $E_k > 1$ MeV). Since

[11] National Institute of Standards and Technology

the cross-section depends on the kinetic energy of the proton, the amount of energy that is transferred from the proton to the medium increases as the proton incrementally (via numerous discrete interactions) slows down. This explains the rise of the Bragg curve (see **Figure 2.11**) until the protons travel with such reduced velocity that they lose large amounts of energy in these events and quickly come to an ultimate stop. The range of protons is commonly defined as the *d80* point of the Bragg curve: the range at which the energy deposition drops to 80 % of the Bragg peak on its distal side.

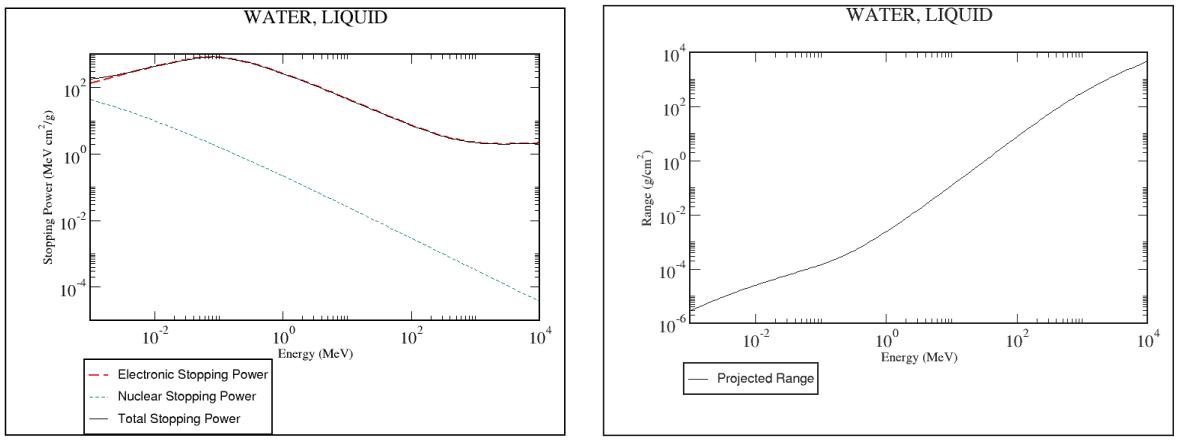

Figure 2.10: *The graph on the left shows that the stopping power is dominated by the electronic stopping power and is energy dependent. The projected range (normalized for density) is displayed on the right. For the energy region of interest (10-300 MeV) the graph is approximately linear. Hence the range-energy relation can be approximated with a power function. Source: PSTAR from NIST [57].*

2.1.2 Range Straggling

Since the interactions are a statistical phenomenon, not all particles experience the exact same number of interactions or lose the exact same amount of energy. Thus, the ranges of all protons in a proton beam (even if initially monoenergetic) differ slightly. They are normally distributed with increasing RMS for longer stopping paths and cause a broadening of the Bragg peak. This is why a pristine 70 MeV proton beam has a much sharper Bragg peak than a 250 MeV proton beam. The Gaussian broadening of ranges $s(x)$ is given by

$$s(z) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-\frac{(z-R)^2}{2\sigma_z^2}} \tag{2.4}$$

with

$$\sigma_z = 0.012\, R^{0.951} \tag{2.5}$$

and where z is the direction of travel and R the mean range of the corresponding energy [59].

Instead of a deviation in range, the straggling effect can also be expressed as a deviation in kinetic energy after a fixed penetration depth. A monoenergetic proton beam, e.g. 250 MeV, penetrating a certain thickness of a material, e.g. 10 cm of graphite, will exit the material with an induced energy spread. This energy spread can also be approximated with a Gaussian distribution (an approximation which is not always accurate enough, as discussed in chapter 6.3). The energy spread caused by straggling is quadratically added to any pre-existing energy spread of the beam.

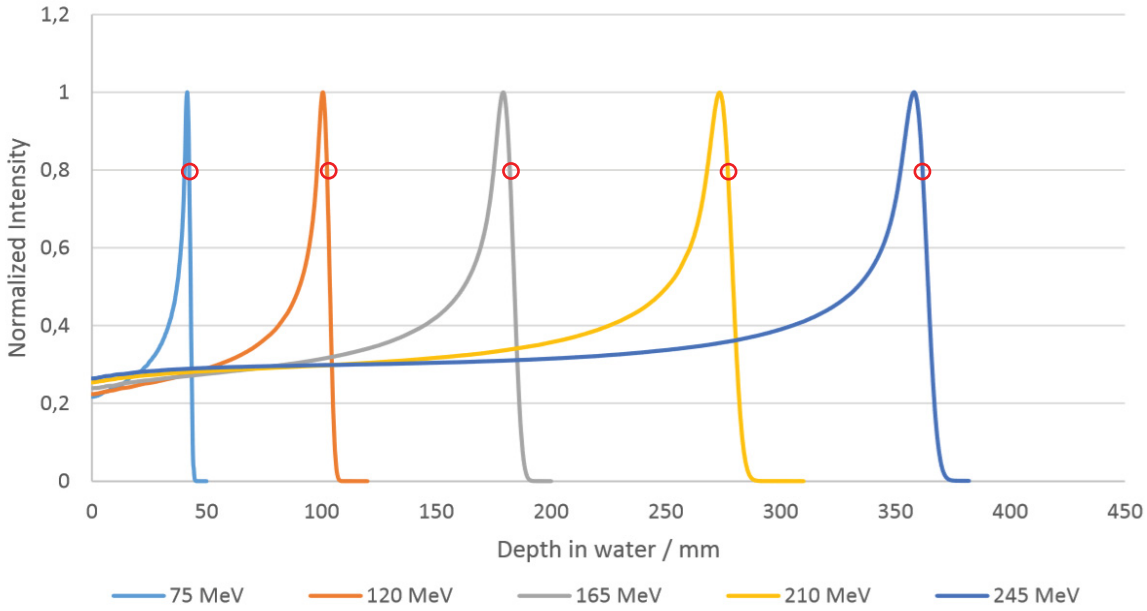

Figure 2.11 Bragg peaks for various energies as measured at RPTC, Munich: *The graphs display the range straggling effect as the Bragg peaks broadens with increasing beam energy (penetration depth). The red circle marks the d80 point, the reference for range.*

2.1.3 Nuclear collisions

While interactions with the atom shell are dominant, protons can also directly collide with a nucleus of an atom (roughly 20 % of protons of a 160 MeV beam do [60]). Those collisions can be elastic or inelastic. The latter meaning that a part of the kinetic energy is transformed into a change of state, e.g. excitation or nuclear reaction. The initial proton will lose a significant amount of energy and scatters under a large angle. It is then, along with possible other produced particles such as fragments or emitted neutrons, protons or alphas, called secondary. Nuclear collisions cause a loss in intensity as the proton beam penetrates matter (as indicated above a 160 MeV proton beam has about 80 % intensity right before the Bragg peak). It also causes a somewhat increased dose in the proximal plateau region of the Bragg curve as energy is mostly deposited in the vicinity of the nuclear events.

2.1.4 Multiple Coulomb Scattering

A third important interaction of fast protons with matter is the Multiple Coulomb Scattering (MCS). MCS describes very small angular deflections caused by electrostatic interactions. Along the travel path of a particle the small angles of the deflections add up and result in a Gaussian (angular) distribution of particles in the beam. It is important to note that through MCS the proton beam broadens and also becomes more symmetric when traversing material as can be seen in the approximation of Chu et al [59], simplified for protons:

$$\sigma'(R) = 0.0294\, R^{0.896}, \tag{2.6}$$

where R is the depth in water and σ' the additive transverse size of the proton beam in both directions x and y. The size increase σ' from MCS is quadratically added to the initial sizes of the proton beam in the transverse plane x-y, σ_x and σ_y [61].

$$\sigma_{x/y,total}(R) = \sqrt{{\sigma_{x/y}}^2 + \sigma'(R)^2} \tag{2.7}$$

2.2 Beam optics

As a direct consequence of the Maxwell equations electric fields accelerate charged particles while magnetic fields allow to steer them [64]. An ensemble of many particles with similar velocity and direction of propagation which are in close vicinity to a reference particle is called a particle beam. Similar to a lens system for glass optics, magnets are used to guide and focus particle beams. The (theoretical) reference particle with the average position, angle and energy in combination with the statistical distribution of the other particles' values are used to describe the beam and calculate its propagation through conservative systems – such as vacuum and magnetic fields.

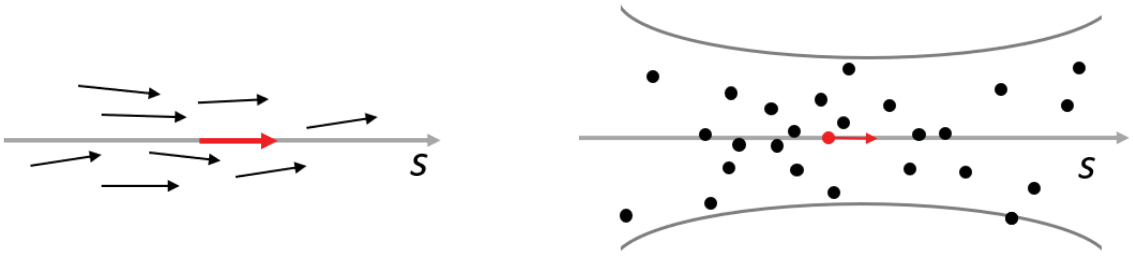

Figure 2.12: *Left: The direction of velocity of an ensemble of particles is similar to a reference particle (all depicted as vectors) on a reference trajectory s. Right: A particle beam is often visualized by its envelope which contains most particles and is of varying width along the reference trajectory s.*

2.2.1 Motion of charged particles in magnetic fields

A charged particle such as a proton interacts with electromagnetic fields via the Lorentz force,

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right). \tag{2.8}$$

When no electrical field $\vec{E}$ is present, the equation becomes

$$\vec{F} = \frac{d\vec{p}}{dt} = q\,\vec{v} \times \vec{B}. \tag{2.9}$$

Thus, magnets can be used to induce a momentum perpendicular to the magnetic flux $\vec{B}$ and the direction of flight of a proton.

2.2.2 Magnetic rigidity

Uniform magnetic fields change the direction of a proton with momentum $\vec{p}$, which is also called bending. A proton travelling on a circular bent path with curvature radius ρ experiences the centripetal force depending on momentum p:

$$F = \frac{mv^2}{\rho} = \frac{pv}{\rho} \tag{2.10}$$

The magnetic rigidity $B\rho$ relates the scalar Lorentz force of equation (2.9) with the centripetal force of protons of equation (2.10) with given relativistic momentum p and charge $q = e$:

$$B\rho = \frac{p}{e} \tag{2.11}$$

2.2.3 *Momentum and kinetic energy*

In proton therapy, instead of the momentum, typically the kinetic energy E is used (in MeV). Common accelerator energies go up to about 250 MeV which corresponds to a proton velocity of approximately 0.6 c. Hence the motion of protons becomes a relativistic problem. Momentum and energy can be converted using Paul Dirac's energy-momentum relation:

$$E_{tot}{}^2 = (E + E_0)^2 = (pc)^2 + (m_0c^2)^2. \tag{2.12}$$

With rest energy $E_0 = m_0 c^2$, the momentum of a proton can be expressed with its kinetic energy:

$$p = \sqrt{\frac{1}{c^2} E^2 + 2\,E\,m_0}, \tag{2.13}$$

where the proton rest mass $m_0 = 938.27$ MeV/c^2. In the following the energy of protons always refers to the kinetic energy and is denoted E for simplicity.

2.2.4 *Beam path and coordinate systems*

In proton therapy systems the so-called beamline transports the protons from the accelerator to the patient. Its most essential elements are magnets that guide and focus the beam along a designed beam path. According to equations (2.9) and (2.11) a homogenous magnetic field perpendicular to the particle motion causes the proton beam to travel on a circular path. Dipole magnets provide such magnetic fields and build bends of the beamline. The designed beam path defines the reference trajectory for the proton beam: It is an assembly of straight sections and curved bends (see **Figure 2.13**).

Now, with the reference trajectory of an ideal reference particle defined, it is clear how the proton beam is supposed to move from the accelerator to the target. The vast majority of protons in the beam, however, does not exactly travel on the ideal trajectory. The task for the beam optics is hence to transport all or at least most of the other particles as well.

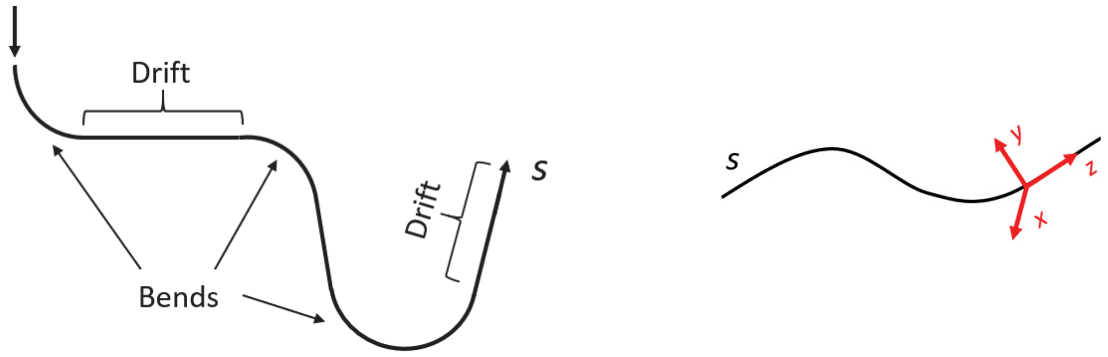

Figure 2.13 Beam path and coordinate system: *Left: The beamline is an arrangement of curved bends and straight sections, which are sometimes called drifts. Right: The cartesian local coordinate system follows the reference trajectory s so that z is always aligned to s and the other two axes define the orthogonal transverse plane.*

Because their coordinates are very similar to the ideal reference particle, they are described in relation to the reference trajectory s. The cartesian coordinate system (x,y,z), also called relative or local coordinate system, moves along s with z always aligned to s while x and y define an orthogonal transversal plane, as depicted in **Figure 2.13**. Typically, the layouts of beamlines,

including the studied cases in this thesis, are two-dimensional and all bends stay in the horizontal plane. Some applications, however, require a mixture of bending in the horizontal and vertical axes, leading to sometimes not insignificant complexity (one example is the corkscrew gantry described in chapter 3.4). Note that the rotation of the gantry does not pose a problem to the plane separation as long as the coupling point, where the rotatable and fixed beamline couple, is a point of symmetry in the optics.

The transversal coordinates of the particles in the beam are given in relation to the reference particle which follows the designed beam path s. If the reference particle is described by a vector $\vec{Z}_0$ then the other particles are defined by $\Delta\vec{Z} = \vec{Z} - \vec{Z}_0$ [63]. In contrast to the local / relative coordinate system is the lab coordinate system, which is employed to define the position of beamline elements (see also **Figure 5.67**).

2.2.5 Focusing with quadrupoles

Since the beam diverges as it propagates, "lenses" are necessary to provide focusing in order to keep the transversal spread of the beam confined. This is an important part of the beam transport as the better confined a beam is, the smaller the apertures of magnets and the vacuum beam pipe can be chosen which has significant practical advantages. Quadrupole magnets serve as electromagnetic lenses for the proton beam and are usually placed in the straight sections of the beamline. A quadrupole always focuses in one transversal plane while it defocuses in the other plane as demonstrated in **Figure 2.14**.

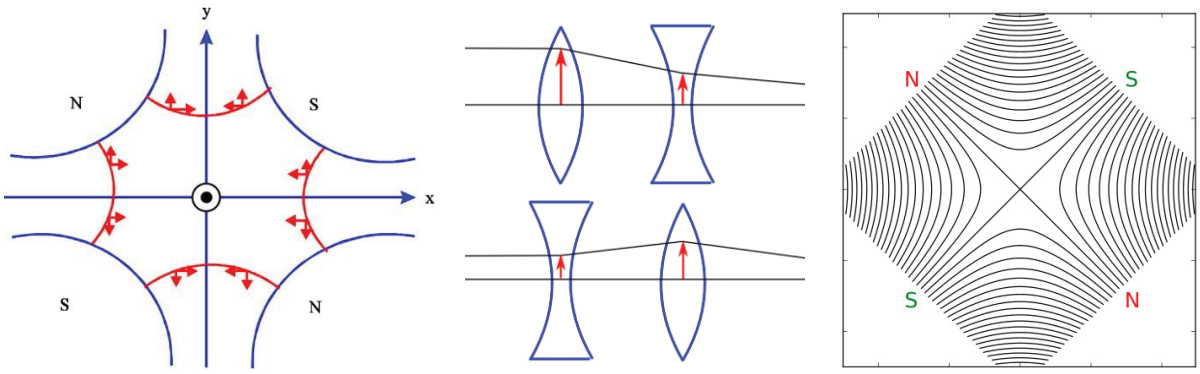

Figure 2.14 A,B,C: *(A) On the left the forces acting on protons passing through a quadrupole are indicated in red. It shows that protons deviating from the center of the magnet are pushed back in the x-plane but pushed farther away in the y-plane. (B) The sketch in the middle indicates the strength of the acting forces which depend linearly on the distance to the magnet's center which is a consequence of the density of magnetic flux lines as shown on the right (C). Credit (C): A. Holzner/Wikipedia.org.*

A quadrupole doublet, two sequential quadrupoles in a beamline, can, however, still provide focusing in both planes. This works because the *magnetic flux* in a quadrupole increases with the square of the radius (in the center it is zero). As depicted in **Figure 2.14** the focusing effect of one quadrupole is always stronger than the defocusing of the other quadrupole due to the fact that the focusing/defocusing power increases with radial distance to the central beam path. The combination of two quadrupoles of inverse polarity is called a FODO cell. A less descriptive but mathematical sound derivation of its combined focusing in x and y can be found in chapter 9.1 of [63].

2.2.6 Dispersion

According to equation (2.11) the bend radius is proportional to the particle momentum for a fixed *B*-field:

$$\rho \propto p, \qquad B = const \tag{2.14}$$

Consequently, particles with a higher momentum are carried outwards and are bent by less degrees than the reference particle. Particles with lower momenta experience the opposite effect. In other words, constant *B*-fields, i.e. dipole fields, act as energy filters, an effect called dispersion. The dispersion function is a function of the beam path *s* that indicates by how far a particle with a larger momentum would deviate from the reference trajectory. Quadrupoles, with their linear field gradient, can change the dispersion function (cause a kick to outwards travelling particles) but cannot eliminate dispersion. A second dipole, however, can under the right circumstances undo what the first dipole produced and cancel the dispersion function in the beam.

The dispersion function is used as an additional term in the equation of motion describing by how much a particle with a deviating momentum is displaced with respect to the reference trajectory along *s*. The displacement is proportional to the relative deviation in momentum $\Delta p/p$.

$$x(s) = x_0(s) + x_D(s) = x_0(s) + D(s)\frac{\Delta p}{p} \tag{2.15}$$

The dispersion function itself is another inhomogeneous differential equation of the general form:

$$D''(s) + \frac{1}{\rho^2} D(s) = \frac{1}{\rho} \tag{2.16}$$

as explained and solved e.g. in Wille's textbook (pp. 74 ff. [64]) and only applies to the transverse direction of beam bending.

2.2.7 Phase space

With the particle beam described as an ensemble that is in relative vicinity of a reference particle, it is important to note that in this context the density of particles is still small so that the space charge remains negligible and interactions between particles within the beam have no measurable effect.

To generally describe the motion of the ideal particle on the reference beam path, the state vector $\vec{Z}$ with six relevant coordinates can be used. In a cartesian coordinates system the state vector defines the position (*x, y, z*) and the momentum (p_x, p_y, p_z) of the particle [63]:

$$\vec{Z} = (x, p_x, y, p_y, z, p_z) \tag{2.17}$$

The dynamics of all other particles in a particle beam are then expressed in relative coordinates to such a reference particle since velocity and direction of flight of all particles are very similar. In single pass-through proton therapy beamlines only the particle distribution in the transversal plane is of interest – at any given point along the design path *s*. If and by how much the particles are distributed along their direction of propagation has no effect since the time constants for all

observable effects are orders of magnitudes larger than those caused by a longitudinal displacement. Accordingly, only the distribution in position and momentum in the transversal directions are needed, called the *transversal dynamics*, which leaves four relevant coordinates.

There are two additional parameters to describe the *longitudinal dynamics* of the proton beam: the bunch length l and the energy spread of the beam δ. The bunch length has no relevance in this context (again for the given application there is no time dependence on this scale). But the energy spread is an important property of the beam.

These parameters, which define the so-called phase space, can be applied to individual particles – providing their space coordinates, transversal momenta and energy deviations, all with respect to the reference particle. But the phase space parameters can also describe the entire beam. In this case they refer to the standard deviation of the respective distribution of values (usually these distributions resemble a Gaussian). So, the complete phase space vector $\vec{Z}$ consists of five relevant components:

$$\vec{Z} = (x, \frac{p_x}{p_0}, y, \frac{p_y}{p_0}, \delta) \tag{2.18}$$

At any point along a trajectory s the proton beam is described using these phase space coordinates which give information regarding the beam size and divergence defined as follows, while the energy spread δ typically remains unaltered.

Table 2.1: *The five relevant phase space coordinates.*

Phase space component	**Name**	**Explanation**
x	Horizontal size	RMS of the particle distribution on the horizontal axis.
$a = p_x/p_0$	Horizontal divergence	RMS of the angle of divergence of the particle distribution in the horizontal direction. Sometimes x', the derivative of x, is used.
y	Vertical size	RMS of the particle distribution on the vertical axis.
$b = p_y/p_0$	Vertical divergence	RMS of the angle of divergence of the particle distribution in the vertical direction. Sometimes y', the derivate of y, is used.
$\delta = dE/E_0$	Energy spread	RMS of the energy distribution with respect to the energy of reference particle E_0. Usually given in percent.

Since the horizontal and vertical axes are orthogonal to each other, they are independent (although quadrupoles always affect both). If one makes sure that those axes are not mixed, the phase space for linear beam optics can be treated separately for x and y. Furthermore, when all bends in the beamline are placed in the horizontal plane, not only the horizontal and vertical planes can be studied separately but the vertical plane is also decoupled from dispersion effects. In this case the appropriate vector for linear motion in the horizontal plane is (x, a, δ) and for the motion in the vertical plane is (y, b).

2.2.8 Phase ellipse, emittance and Twiss parameters

Since particles in a beam typically resemble a Gaussian distribution in real space as well as in phase space, one can draw an ellipse in the phase space diagram which includes a certain fraction of all particles. This ellipse can, for example, depict the one- or two-RMS line of the

corresponding Gaussian (an isodensity line), thus represents approximately 68 or 95 % of the particle beam, respectively. As the horizontal and vertical phase spaces are independent from each other, the phase space diagrams of *x* and *y* can be studied separately.

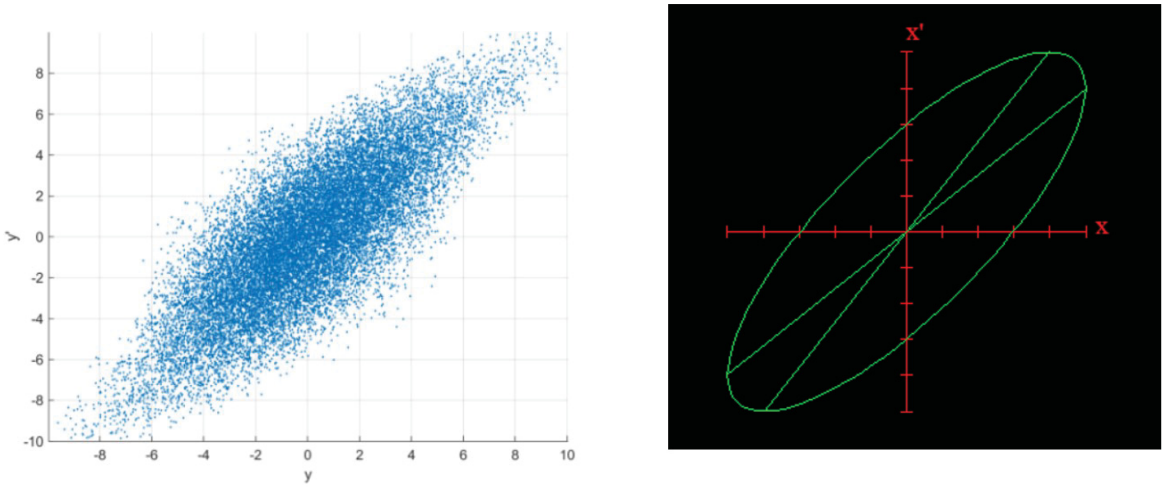

Figure 2.15: *Left: Simulated distribution of particles in phase space. The contour of an ellipse is clearly visible. Right: A phase ellipse as it is used in the beam optics code Transport (see chapter 5).*

According to Liouville's theorem, the volume in phase space that is occupied by the beam is conserved as long as the beam travels through conservative systems. Hence, the area $A_{ellipse}$ of the phase ellipse remains constant. There are a few important parameters, the emittance and the so-called Twiss parameters, that are very useful to characterize the beam. They are elements of the Courant-Snyder invariant and the respective formulas are listed here directly for the sake of convenience. The derivation of the Courant-Snyder formula is shown in the appendix.

The *Courant-Snyder invariant* is the area of the phase ellipse with half-axes *r* and *r'*:

$$\gamma x^2 + 2\alpha x x' + \beta x'^2 = rr' = const \tag{2.19}$$

with $x = r \cos\varphi$, $x' = r' \sin\varphi$ and $\varphi = \{0{:}2\pi\}$.

The *emittance* ε is *the* important constant value in beam optics, defined as

$$\varepsilon = rr'/\pi = A_{ellipse}/\pi \tag{2.20}$$

and given in the units mm·mrad, as it can be calculated as the product of the size and divergence of the beam (for each dimension, horizontal and vertical, separately). The emittance can be given for a single particle or a particle beam. In the latter case it is important to note what the emittance exactly refers to. One common notation is the 2-RMS notation where the emittance is calculated as the product of the 2-RMS values of initial size and divergence.

$$A_{ellipse,x} = \pi\,\varepsilon_x, \qquad A_{ellipse,y} = \pi\,\varepsilon_y \tag{2.21}$$

The *Twiss parameters* are the α, β and γ of equation (2.19) and have besides their relation to the emittance a direct meaning for the size, the divergence as well as maxima and minima of the beam envelope. They can be expressed and calculated as functions of the beam path *s* and are dependent from each other with

$$\gamma = \frac{1+\alpha^2}{\beta}, \tag{2.22}$$

and

$$\alpha(s) = -\frac{1}{2}\beta'(s). \tag{2.23}$$

2.2.9 Envelope and beta function; divergence and gamma function

The envelope *E* of a particle beam is often depicted along the beamline to check that the beam size stays sufficiently confined. A large envelope can cause beam losses at apertures of magnets or in the beam pipe. There is a direct relation between the envelope of the beam, so the size of the beam, and the beta function via the constant emittance. The maximal size at location *s* can be calculated with:

$$x_{max}(s) = E(s) = \sqrt{\varepsilon_x}\sqrt{\beta_x(s)} \tag{2.24}$$

Hence the beta function is proportional to the square of the beam size.

Analogously the gamma function directly indicates the maximal divergence *A* in the beam:

$$x'_{max}(s) = A(s) = \sqrt{\varepsilon_x}\sqrt{\gamma_x(s)}. \tag{2.25}$$

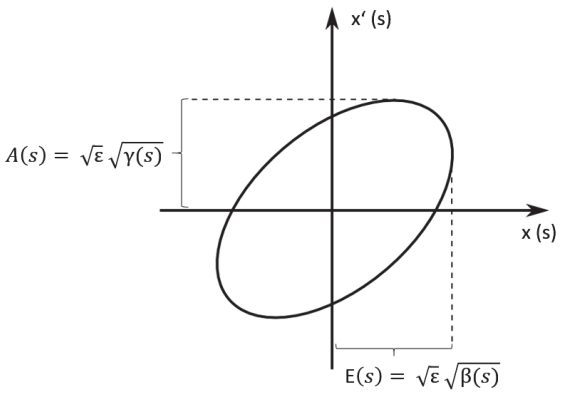

Figure 2.16: *The phase ellipse indicates the envelope E and divergence A as a function of s which can be calculated with the functions of the Twiss parameters β and γ, respectively. The orientation of the phase ellipse also indicates whether the beam is converging or diverging.*

2.2.10 Beam waists and alpha function

However, the maximal divergence itself does not provide too much useful information for the beam optics. A beam focus, which is a (local) minimum in size in real space, is called a beam waist. If a beam is focused in both transversal dimensions, *x* and *y*, at the same location *s*, it is called a double waist. Often a double waist is desired, e.g. to minimize the beam size cross section for maximal treatment precision. As can be seen in **Figure 2.17** the envelope of the beam exhibits a minimum, thus the following condition must be fulfilled:

$$\beta'_{s=waist} = 0 \tag{2.26}$$

And with equation (2.23) follows

$$\alpha_{s=waist} = 0. \tag{2.27}$$

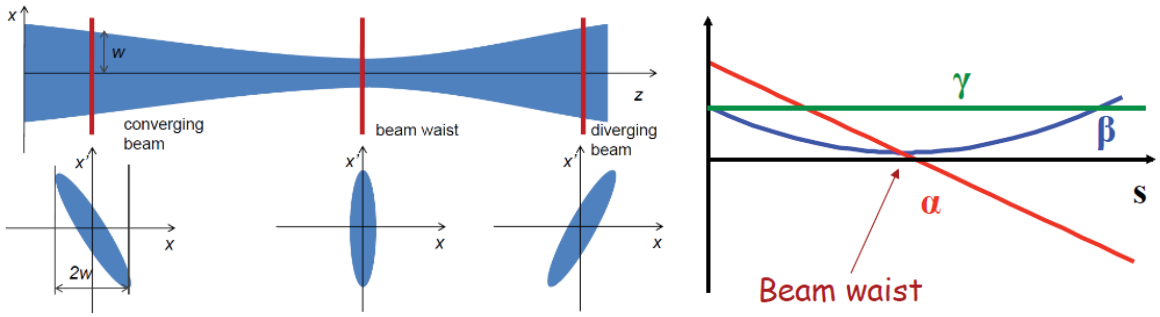

Figure 2.17: *Left: A beam envelope exhibits a converging and diverging beam with a beam waist. From the orientation of the phase ellipses below the divergence can be deduced [26]. Right: The course of the Twiss functions for a similar situation is depicted. As α crosses zero the beam exhibits a waist.*

2.3 Magnetic fields

2.3.1 Biot-Savart law

When studying beam optics, there are three different ways to determine the magnetic fields encountered by the proton beam when passing through electromagnets. In the simplest case, it is sufficient to use idealized formulas for the beamline elements like drifts and magnets, e.g. dipoles, quadrupoles, sextupoles, and others. Whenever there are deviations from the standard description a closer look at the individual field is necessary. There are multiple tools available that allow to numerically calculate the *B*-field of any magnet geometry. They produce field maps that give any voxel at location $\vec{r}$ a calculated $\vec{B}_{\vec{r}}$. The precision depends on the granularity of the map which in turn is decisive for the computing time to solve the magnet model. Even with high performing clusters the computing time can become very significant (order of hours or days), thus iteratively working with this approach is rather slow and tedious. A third way to obtain the *B*-field at any given point $\vec{r}$ is the application of the Biot-Savart law. While this is very quick and elegant, it can only be employed if the *B*-field is solely produced by electric currents and not with the help of (ferro-)magnetic materials, such as iron. Resistive magnets with the discussed field strengths always employ iron, but with superconductors so-called air-coils become feasible. Then with Biot-Savart one can analytically obtain at any point of interest:

$$\mathrm{d}\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi} I \, \mathrm{d}\vec{l} \times \frac{\vec{r} - \vec{r}'}{|\vec{r} - \vec{r}'|^3} \tag{2.28}$$

where *I* is the electric current running through a piece of conductor with length and orientation $\vec{l}$ and location $\vec{r}'$. Note that this formula only holds true when the conductor width can be neglected, which is the case for the studied applications. Otherwise integration of the electric current density of the conductor volume becomes necessary.

2.3.2 Taylor expansion of the B-field

In the transversal coordinate system the equation of magnetic rigidity can be written for a particle moving in the horizontal plane (*x*-*z*):

$$\frac{1}{\rho(x,y)} = \frac{e}{p}\, B_y(x,y) \tag{2.29}$$

The *B*-field can be expanded in a Taylor series:

$$\frac{e}{p}B_y(x) = \frac{e}{p}B_{y0} + \frac{e}{p}\frac{dB_y}{dx}x + \frac{1}{2}\frac{e}{p}\frac{d^2B_y}{dx^2}x^2 + \frac{1}{6}\frac{e}{p}\frac{d^3B_y}{dx^3}x^3 + \cdots \tag{2.30}$$

$$= \frac{1}{\rho_0} + kx + \frac{1}{2}mx^2 + \frac{1}{6}ox^3 + \cdots \tag{2.31}$$

The first term is a constant *B*-field, produced by a dipole, the second term has a linear dependence on *x*, produced by a quadrupole of strength *k*, the third term with the quadratic gradient is produced by a sextupole with strength *m* and the fourth term by an octupole with strength *o*. Hence the magnetic field can be expressed as a sum of multipoles [64].

The first two terms, the fields of dipoles and quadrupoles, are relevant for the **linear beam optics** which are the basis for beam guidance and focusing. Most beamlines for proton therapy systems only use dipoles and quadrupoles. Sextupoles or other higher multipole fields are employed to correct nonlinear aberrations or unwanted field components which then have to be dealt with. Sextupoles are typically employed in storage rings to compensate chromatic aberration of focusing magnets, the energy-dependent focal length of quadrupoles.

2.3.3 Equations of motion

The equations of motion for the particle optical coordinates (see **Table 2.1**) can be directly derived from the Lorentz force:

$$F_L = \frac{dp_x}{dt} = -q\, v_z\, B_y(x,y), \quad if\ B_z = 0\ \ and\ z \| s \tag{2.32}$$

$$\frac{dp_x}{ds} = -q\, B_y(x,y) \quad | \ with\ \frac{ds}{dt} = v_z \tag{2.33}$$

$$\frac{da}{ds} = -\frac{q}{p_0}\, B_y(x,y) \quad | \ with\ \frac{p_x}{p_0} = a \tag{2.34}$$

$$\frac{dx}{ds} = \frac{p_x}{p_s} = a\,\frac{p_0}{p_s} = \frac{a}{\sqrt{1-a^2-b^2}} \tag{2.35}$$

where p_0 is the total momentum previously denoted *p*. Analogously, the equations for the coordinates *y* and *b* of the *y*-*s* plane are:

$$\frac{db}{ds} = -\frac{q}{p_0}B_x(x,y) \tag{2.36}$$

$$\frac{dy}{ds} = \frac{b}{\sqrt{1 - a^2 - b^2}} \tag{2.37}$$

The fifth coordinate δ describing the energy spread remains unaltered in the beamline:

$$\frac{d\delta}{ds} = 0 \tag{2.38}$$

There are also other forms of the equations of motion. In the literature, particularly the form with ODEs of second order are often found. However, for the methods employed in this study the first-order differential equations are much easier to deal with.

Equation of motions in the form of 2nd-order ODEs:

$$x''(s) + \left(\frac{1}{\rho^2(s)} - k(s)\right) x(s) = \frac{1}{\rho(s)} \frac{\Delta p}{p} \tag{2.39}$$

$$y''(s) + k(s)\, y(s) = 0 \tag{2.40}$$

2.4 Transfer maps

2.4.1 Mapping functions to deterministically calculate beam transport

The transport of charged particles through a magnetic field is a completely deterministic system just by virtue of the equations above. This means, if the magnetic field is known, it acts like a function: calculating the final particle coordinates from the initial coordinates.

$$\vec{Z}_{final} = f(\vec{Z}_{initial}) \tag{2.41}$$

In the case of linear beam optics, this function can be written as a matrix and by straightforward vector matrix multiplication the final *Z*-vector can be calculated which is shown in more detail in chapter 5. But more generally this function is nothing but a correlation that maps the particle coordinates before and after the magnetic field, hence this function is called a transfer map *M* [65].

With the beam transport through a known magnetic field described in terms of a transfer map, the output coordinates of the beam, as a function of the initial coordinates, can be obtained by plugging in the starting coordinates of the beam. *M* itself is completely deterministic and returns the optical coordinates at the end of the studied beamline. Of course, obtaining *M* in the first place can be an involved exercise.

In idealized cases the magnetic fields stay constant for a longer stretch along the beam path *s*, for example when the field is zero for a certain length of beam drift or the field profile within a dipole or other magnet stays constant along its physical length. However, in many real cases this is not true. In contrary, the *B*-field seen by the beam can change continuously along the trajectory due to fringe fields or other inhomogeneities. While in an idealized beamline lattice one has a rather small number of elements and each of these beamline elements can be described with one transfer matrix (linear optics) or transfer map (nonlinear optics), a detailed study of an inhomogeneous system requires much smaller steps and many more transfer maps.

In any case, a concatenation of the many maps leads to a single map describing the transport through the complete magnetic system. The advantage of this method is the possibility to compute large numbers of particles in a very short time. The disadvantage is that the detailed information of the beam coordinates along its way through the system is lost as the only output information is the set of coordinates at the final position. A good compromise can be to intentionally break up the system to check the beam coordinates at certain locations in between the individual transfer maps. In the study of the *AGCCT* this was done e.g. at the midpoint or the entrance of the magnet to investigate possible beam scraping at the magnet aperture. The details of the transfer maps used for this study are described in chapter 5.4.

2.4.2 Non-linear optics in transfer maps

As described above any *B*-field encountered by the particle beam can be described in terms of a Taylor expansion, see equation (2.30). When higher order terms become relevant one speaks of non-linear optics. For example, the effect of a sextupole on the motion of the particle is proportional to the square of the particle position. That effect is nothing else than a "kick" due to the Lorentz force acting on the charged particle. It is described with *da/ds* in the formula below for the *x*-*s* plane. By integrating over *s* the resultant transversal dynamics of the particle are obtained. From equation (2.34) follows:

$$\frac{da}{ds} = \frac{e}{p_0}B_{y0} + \frac{e}{p_0}\frac{dB_y}{dx}x + \frac{1}{2}\frac{e}{p_0}\frac{d^2B_y}{dx^2}x^2 + \frac{1}{6}\frac{e}{p_0}\frac{d^3B_y}{dx^3}x^3 + \cdots \tag{2.42}$$

With the expanded form of the *B*-field in a power series, one can directly see how the various orders of the *B*-field terms lead to respective orders for the particle motion. As the sextupole term has quadratic *x* dependence, the leading contribution to the resultant particle position (after integration) will have quadratic dependence on the initial *x* position.

2.4.3 Non-linear effects from quadrupoles

Apart from higher multipoles that directly generate non-linear fields, also for the linear beamline elements of quadrupoles non-linear beam optics play a role. This can be seen from equation (2.34):

$$\frac{da}{ds} = -\frac{q}{p_0}B_y \tag{2.43}$$

Because with

$$p_0 = \sqrt{{p_x}^2 + {p_y}^2 + {p_s}^2} \tag{2.44}$$

and

$$p_s = p_0\sqrt{1 - \frac{p_x^2}{p_0^2} - \frac{p_y^2}{p_0^2}} = \sqrt{1 - a^2 - b^2} \tag{2.45}$$

follows that the "kick" experienced in the quadrupole has third order dependence as well. These third order aberrations of quadrupoles are a common issue with strong quadrupoles.

2.5 Differential algebra

A *Differential algebra* (DA) is applied to perform the algebra and differentiation of polynomial functions. More details and a deeper explanation can be found in [63] and [65] but a brief introduction is outlined here.

A polynomial function, i.e. a power series, is characterized by its order n and number of variables v. For the handling of such a power series a vector $\boldsymbol{a}$ is defined where each component represents the coefficient of a monomial a_i of the series and the number of monomials N, i.e. vector components, is:

$$N(n, v) = \frac{(n+v)!}{n!\, v!} \tag{2.46}$$

The rules of the DA define how addition, scalar and vector multiplication are performed. When the ordering of the monomials and corresponding coefficients is defined in a consistent way, two power series represented by two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ can be manipulated obeying these rules:

$$(a_1, a_2, \dots, a_N) + (b_1, b_2, \dots, b_N) = (a_1 + b_1, a_2 + b_2, \dots, a_N + b_N) \tag{2.47}$$

$$t \cdot (a_1, a_2, \dots, a_N) = (t \cdot a_1, t \cdot a_2, \dots, t \cdot a_N) \tag{2.48}$$

$$(a_1, a_2, \dots, a_N)\,(b_1, b_2, \dots, b_N) = (c_1, c_2, \dots, c_N) \tag{2.49}$$

Where the coefficients c_i are calculated with

$$c_i = \sum_{\substack{1 \le \nu,\mu \le N \\ M_\nu \cdot M_\mu = M_i}} a_\nu\, b_\mu \tag{2.50}$$

and M_i is the monomial of the i^{th} vector component according to the defined ordering. Note that Berz in the quoted source uses a slightly different notation where the factorials of the monomials are treated separately. However, in this general form the here simplified *DA* rules are correct.

By the multiplication rule the terms in each vector component stay the same and thereby the order of the represented power series stays the same, too. In other words, this is nothing but a straightforward multiplication of two polynomials where all terms with higher orders than the original function are discarded. The rule $M_\nu \cdot M_\mu = M_i$ guarantees that only those monomials are multiplied that lead to the same terms as the original monomial of the i^{th} vector component.

This is an important feature, since otherwise the multiplication by going through many steps (individual transfer maps) would lead to a fast increase of the order of polynomials.

As a consequence, it is sufficient to only deal with the coefficients of the monomials as the variables and their exponents remain unaltered for each vector component. Hence the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ only carry the coefficients a_i and b_i and the term of the component i is defined by the ordering convention.

The differentiation of the polynomials is done by the operation

$$\partial_\nu(a_1, a_2, \dots, a_N) = (c_1, c_2, \dots, c_N) \tag{2.51}$$

with

$$c_i = O_v(M_i\ x_v)\ a_{I_{(M_i\ x_v)}} \tag{2.52}$$

which moves the coefficients in the vector to the component of the lower order regarding the variable of differentiation. Where I_M is the position i of the monomial M and M_i is the i^{th} monomial according to the ordering convention.

3 PROTON THERAPY EQUIPMENT

3.1 Proton therapy facilities

In proton therapy facilities the beam is guided from an accelerator to one or more treatment rooms where it is applied to a patient. From acceleration to stopping in the tumor, the protons pass through the following essential subsystems in sequential order:

- ***Accelerator**:* The accelerator produces the proton beam. Cyclotrons output a constant beam energy. Synchrotrons can adjust the beam energy.
- ***Energy degrader:*** The degrader reduces the energy of the proton beam as required. This device is only mandatory for cyclotron powered facilities. It is usually a part of the so-called energy selection system (ESS).
- ***Beam transfer line*:** An arrangement of magnets that guides the beam from the accelerator to the patient, also called *beamline*.
- ***Rotational gantry*** (optional): the last part of the beam transfer line is mounted onto a rotatable mechanical structure which allows a range of angles for beam application.
- ***Scanning nozzle*:** Two magnets deflect the beam in the transversal plane to "scan" a lateral field. Note that the in-depth beam position is defined by its energy (and adjusted with the energy degrader). There are various forms of scanning systems as well as the older scattering technique where the beam is spread out and collimated instead. Since modern systems (and all recent developments in the field) employ scanning, the scattering approach is not further described here.

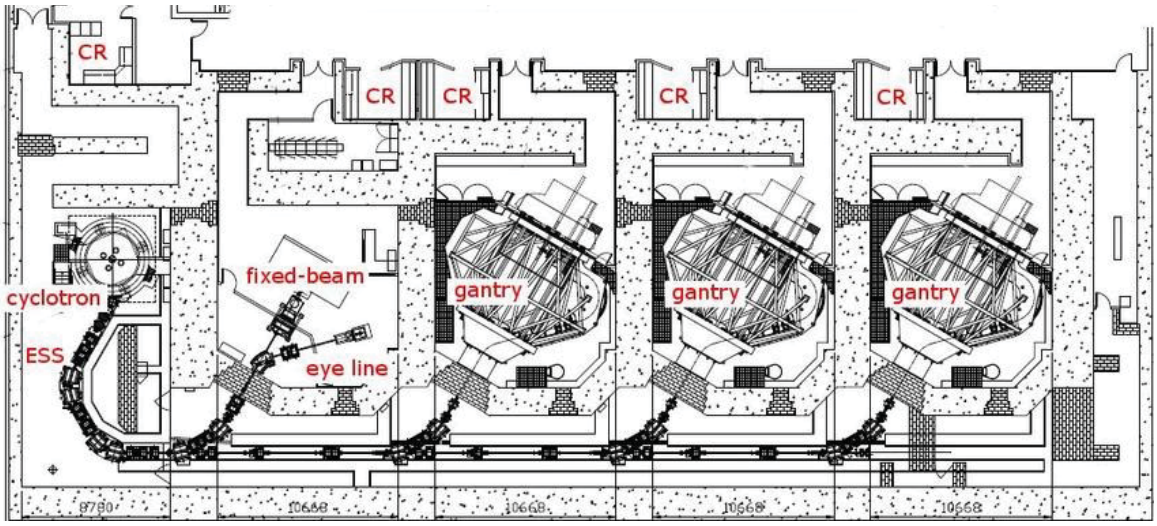

Figure 3.18: *Layout of the proton therapy facility WPE in Essen, equipped by IBA [66]. The proton beam is produced in the cyclotron and the energy is adjusted within the "ESS". The beam is then distributed to one of the treatment rooms via the beamline which either feature a rotating gantry or a beam port with fixed horizontal angle. Eye treatments need special equipment, thus require a dedicated beam exit. "CR" denotes the control rooms for the technical and clinical staff.*

3.1.1 Nomenclature

Upstream and downstream

Upstream of a given beamline element refers to the part of the beamline that protons pass before they reach said element. Therefore "upstream of the degrader" means the part of the beamline between cyclotron and the degrader. *Downstream* is the counterpart, in the example the beamline following the degrader.

Dipoles and bend magnets

Dipole magnetic fields bend the proton beam on a circular trajectory. But not only "pure dipoles" are used as magnet bends, also a combined function magnet – including a dipole field component – can bend the beam. Hence magnets that are used for beam bending are simply called *bends* in the following and only further distinguished when necessary. Outside bends the beam in principle always travels on a straight path.

3.1.2 Accelerators

Synchrotrons and cyclotrons are the technologies of choice to accelerate protons to the energies needed for proton therapy. They have specific advantages and disadvantages but overall the cyclotron dominates in the market as it has the lowest cost [67]. A third option is a linear accelerator which is even more costly because, in contrary to cyclic accelerators, all the RF-structures are only seen once by the beam. Advanced Oncotherapy, a spin-off from CERN, is in the process of commercializing such a linear accelerator for proton therapy but are still in the development phase [67, 68, 69].

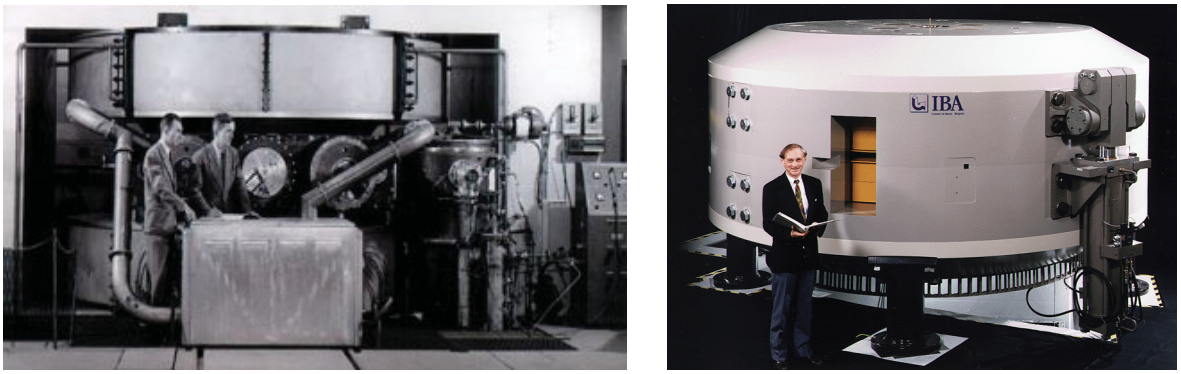

Figure 3.19 Examples of normal conducting cyclotrons: *Left: The synchro-cyclotron at the Harvard Cyclotron Laboratory (HCL), Boston, was operated an impressive 54 years from 1949 to 2003. Picture taken from [70]. Right: The isochronous normal conducting cyclotron C230 developed by IBA weighs over 200 tons. Photo from [71].*

More novel ideas are dielectric wall and laser-plasma acceleration. But both concepts are still in the research state and it is unclear when (or if at all) it will become technically feasible to effectively produce proton beams for clinical purposes with these methods [67, 72].

Hitachi, Mitsubishi and Protom offer synchrotrons. Varian, IBA, Mevion and Sumitomo developed dedicated cyclotrons (ProNova employs the Sumitomo cyclotron). Varian (then ACCEL Instruments) was first to develop a superconducting cyclotron for proton therapy [20] which is isochronous producing a quasi-continuous beam. IBA and Mevion have more recently developed superconducting synchro-cyclotrons with pulsed beam operation [23, 67].

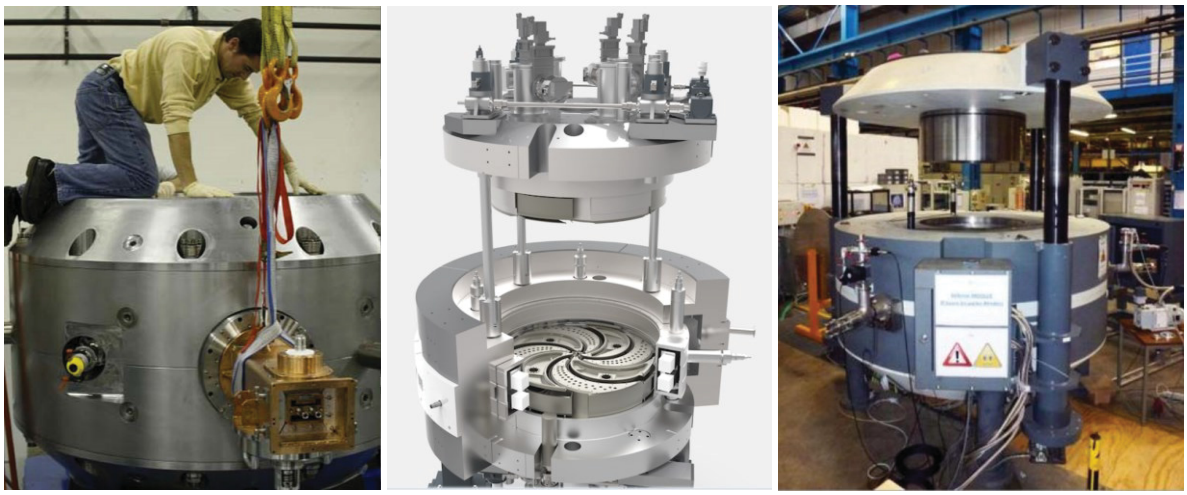

Figure 3.20 The three medical superconducting cyclotrons: *Left: The Mevion superconducting synchro-cyclotron operates at a maximal field of 9 T and has a diameter of only 1.8 m. Source: [73]. Middle: The ACCEL/Varian isochronous cyclotron was the first superconducting cyclotron for clinical beam energies. It is about 3.5 m in diameter and was first installed at PSI. (Picture: Courtesy of Varian). Right: The S2C2 superconducting synchro-cyclotron produces a maximal field of about 5.7 T and is about 2.5 meters in diameter [74, 75]. Picture taken from [76].*

Here is an overview of the most important accelerator types and their pros and cons:

Table 3.2*: Overview of accelerator types for clinical beam production.*

Accelerator	**Advantages**	**Disadvantages**
Synchrotron	- Variable energy: - No beam degradation necessary: less neutrons produced. - Option for multiple ion species	- Large - Expensive (ring + injector) - Macro-pulses - Asymmetric beam profiles
Linear accelerator	- Variable energy - Very small emittance	- Expensive (many RF-structures) - Pulsed beam - Long (unfavorable form factor)
Isochronous cyclotron	- Continuous beam - High intensity - Compact (diameter ~ 3 m)	- Limited in maximal field - Fixed energy - Protons only
Synchro-cyclotron	- Most compact (diameter ~ 2m) - High fields feasible - Low energy consumption - Most affordable	- Pulsed beam - Low intensity - Fixed energy - Protons only

In the following compact proton therapy systems powered by cyclotrons are studied further as they achieve lowest cost and footprint. The beam optics considerations do not have any time dependence. Hence it does not matter whether the beam is pulsed or continuous and the results are applicable to both, isochronous and synchro-cyclotrons.

3.1.3 Energy degrader

The biggest disadvantage of cyclotrons is their inability to vary the output energy. As pointed out in chapter 2 (see **Figure 2.11**) the range of the proton beam depends on its energy. Therefore,

in order to reach different depths in the patient body, the energy of the beam needs to be adjusted with the help of an energy degrader.

The method is rather simple and crude: By placing a piece of material in the beam, the protons lose energy via the stopping power. And by controlling the thickness of the material one controls the residual beam energy as needed.

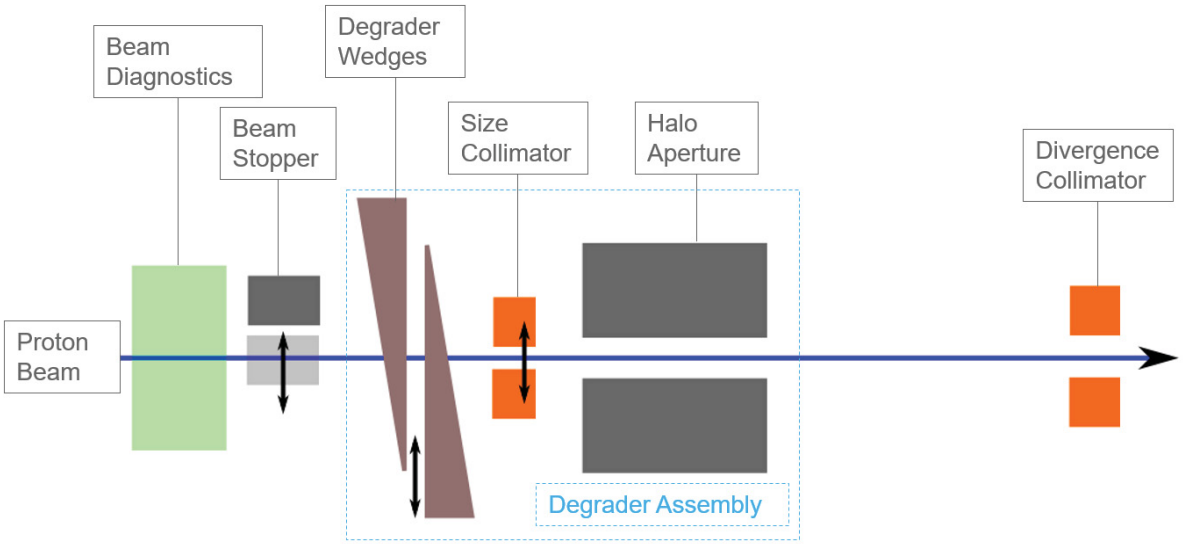

Figure 3.21: *Schematics of the degrader and associated components. The residual energy/range of the protons is controlled via the thickness of the inserted wedges (typically carbon). Scattering in the wedges spreads the beam (not depicted here). The size collimator defines an initial beam size and the divergence collimator stops protons that passed the size collimator under larger angles. Protons with very large angles (that would miss the divergence collimator) are stopped in the halo aperture. Beam currents and position are monitored with a diagnostic box of transmission ionization chambers. The beam stopper is a safety feature.*

While the simplicity of the method is robust and cheap, it comes with some negative side effects:

- The proton beam experiences range straggling (see section 2.1.2) and an energy spread is introduced.
- Through scattering the proton beam blows up. Due to the limited emittance acceptance of the subsequent beamline only a small portion of the beam actually gets through to the patient. The major portion of protons (~ 99 %) have too large angles and have to be stopped in collimators.
- The stopped protons produce neutrons causing a radiation hazard that requires adequate radiation shielding. Hence the use of an energy degrader is quite dirty.

3.1.4 Beam transfer line

The beamline guides the beam from one location to another, i.e. from a source to a target. This is achieved with magnets that steer the beam (dipole fields) and keep it confined by focusing (quadrupole fields). Typically, the beamline comprises bending dipoles, small dipole steerers for corrections, quadrupoles as well as beam monitors for diagnostics. In this study the beamline is separated into two segments. The first, also called *front-end*, starts at the accelerator and ends at the energy degrader. The second starts at the degrader (with a new defined beam) and ends at the isocenter – the reference point for treatment.

3.1.5 Gantries

Rotational gantries are needed to apply the beam from multiple angles. The gantry rotates the beamline about the isocenter (the fixed reference point in the treatment room) on a vertical circular path. With a robotic patient couch, rotatable in the horizontal plane, any beam angle of incidence with respect to the patient can be achieved. In fact, in combination with the couch rotation only 180 degrees of gantry rotation are sufficient. Such solutions were built by IBA, the Proteus®One gantry and by PSI, the Gantry 2. The advantages are simplified gantry mechanics and reduced space requirements.

The beamline of a gantry can be described as follows. The beam enters the gantry horizontally is then bent upwards (along the so-called *gantry spine*) and then finally bent into a vertical plane orthogonal to the gantry rotational axis. Other solutions have a total of three bends (e.g. the Pavlovic design [77]) but the general scheme remains the same (see **Figure 3.22**).

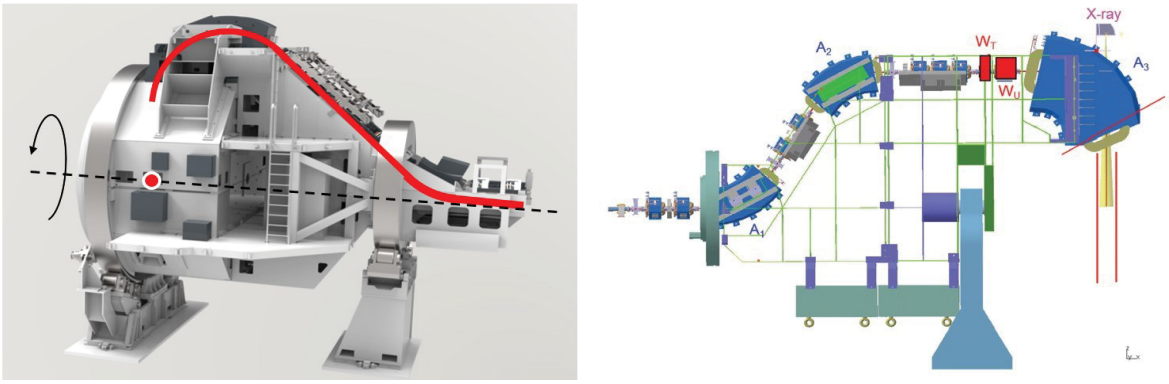

Figure 3.22: Examples of normal conducting proton gantries. *Left: The Varian Gantry with 2 bends (45° and 135°) is rotatable by 360°. The red dot indicates the isocenter, the red line the beam path and the dashed lined the rotational axis. Source: Courtesy of Varian. Right: Schematic view of the PSI Gantry 2 as an example for the Pavlovic design. Since there is no large bearing at front, the rotational range is limited (about 220°). The scanning magnets, WT and WU (red), are placed upstream of the last bend. To allow the scanned beam to pass through, this dipole has a large bore, leading to a significant size increase of the magnet. The benefit is the reduced gantry radius: 3.2 m instead of the 5 m for the Varian example [78, 79].*

3.1.6 Pencil beam scanning and Spread-out-Bragg-peaks (SOBP)

Pencil beam spot scanning is a technique to apply the proton beam according to a prescribed dose distribution as specified in a treatment plan. The relatively narrow proton beam (FWHM ≈ 1 cm) is moved across a lateral raster, the scanning field, by two orthogonal deflection magnets. At each position on this raster the beam is held until the prescribed number of protons are accumulated (beam current times duration) – which is also called a *beam spot*. Once all spots of a given layer in depth is irradiated, the beam energy is changed by the energy degrader and the next layer is scanned. The time interval needed between the energy layers is

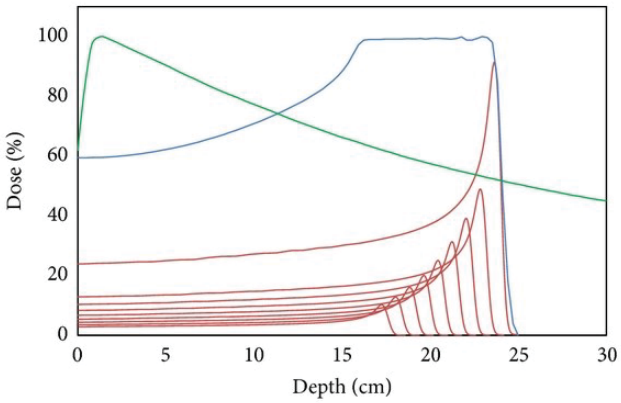

Figure 3.23 Relative depth doses: *The SOBP (blue) is the summation of multiple pristine Bragg peaks (red). The green line depicts the relative depth dose curve of a photon beam for comparison. Note that the dose level for photons must be raised significantly to generate a comparable integral dose over the target region [81].*

called the *layer-switch time*, during which the beam is turned off. With step sizes of typically 5 mm (in water) between the layers, the accumulated "dead-time" can become significant – especially in the context of rescanning or breath-hold techniques (see section 1.2.3).

The energy variation is needed to "stack" multiple depth dose curves in order to achieve dose conformity in all three dimensions. **Figure 3.23** shows an example to achieve a flat homogenous dose distribution in depth, called a spread-out-Bragg-peak, *SOBP* [80].

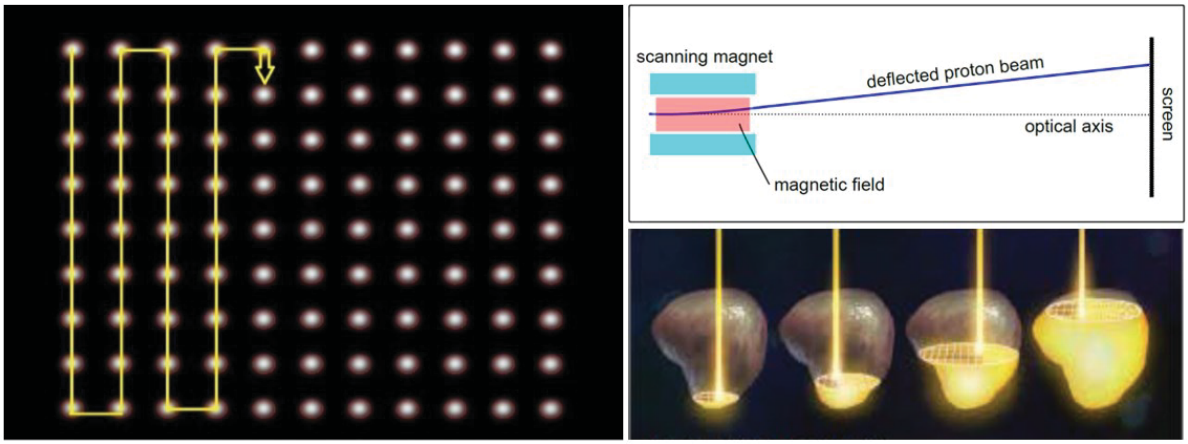

Figure 3.24: *Left: View of a 2D layer with a regular grid of beam spots that are irradiated in the typical meander pattern. Right top: Schematic sketch of scanning a pencil beam by deflecting it with a dipole field. Right bottom: Non-physical visualization of the scanning process in 3D. Pencil beam scanning over the tumor volume is also called "painting" [82].*

3.2 Proton therapy treatment on a gantry and derived beam requirements

Treatment of a localized tumor in a patient starts with treatment planning. With diagnostic methods, foremost CT imaging, the tumor as well as the surrounding organs are identified in a 3D map of the patient in a process called *contouring*. Next a physical dose is prescribed to the target volume, which includes the tumor and safety margins. A maximum dose limit is allocated to vulnerable organs in the vicinity – called organs at risk (OAR). The task of treatment planning is to calculate irradiation plans that achieve the prescribed dose in the target but minimize the dose in healthy tissue, especially obeying the maximum dose values for the OARs.

Dose is defined as the energy per mass that is absorbed in the irradiated volume, measured in Gray (Gy) with:

$$1\,Gy = 1\,J/kg \tag{3.1}$$

Note that the definition is per mass, hence a 60 Gy treatment plan for 1 liter target volume requires, to first order, twice as many protons as a 60 Gy treatment plan for a 500 milliliter target volume. The total treatment dose is not applied at once but in daily fractions of typically 2-3 Gy per fraction (with some exceptions, e.g. eye treatments). Consequently, the completion of a radiation treatment takes a couple of weeks and typically 20 - 30 fractions. However, a current trend in proton therapy is to apply larger doses per fraction in so-called *hypofractionated treatments* [83, 84].

In order to achieve the prescribed dose distribution different treatment angles are chosen under which the proton beam enters the body. In many cases it is advantageous to use two, three or even more different treatment angles to dilute the undesired dose in the healthy tissue over a

larger volume. Hence, in one fraction multiple irradiations from different angles take place. For each of these angles an individual irradiation plan is made, which is called a treatment or irradiation *field*.

The most important and state-of-the-art technique to apply pencil beam scanning is called intensity modulated proton therapy (IMPT). With IMPT different treatment angles are taken into consideration for the planning optimization, and only the summation of all applied beams from these multiple angles will yield a uniform dose distribution in the target volume [85]. It is considered the counterpart to IMRT [86] in conventional radiotherapy where multiple treatment angles and 2D field shapes are optimized. However, the 3D shape of Bragg peaks adds another dimension, i.e. free parameter, to the optimization.

IMPT also means that different spots of the treatment field can have very different dose weights. This in combination with the multiple fields and the fast motion mitigation techniques (rescanning or breath-hold), leads to some important requirements for the beam at isocenter:

Table 3.3: *Important technical beam requirements deduced from clinical needs.*

Beam size	Smaller beams enhance precision. However, the effective size of beams with in-vacuo size smaller than about 3 mm (1 sigma) is dominated by MCS. Hence there is hardly clinical benefit from minimizing the beam size beyond this point [87].
Beam current	High beam currents are needed for fast beam application (e.g. hypofractionation and breath-hold techniques).
Layer / energy switches	Minimizing dead-times, prominently layer-switch time, is essential for a short treatment duration (especially for the breath-hold technique).
Beam current variability	A high and fast dynamic of the beam current is desired to rapidly irradiate spots with different dose weights.
Angles	360 degrees of treatment angles are required. Due to the interplay of gantry and patient table rotation, the gantry does not necessarily need a 360° rotatability. For mechanical simplicity some designs restrict the gantry rotation to less than 220°.

3.3 Compact proton therapy systems

Compared to conventional photon radiotherapy, the equipment for proton therapy is much larger, more complex and very expensive. The original idea to maximize cost efficiency of proton therapy was to connect multiple treatment rooms to one accelerator and thereby dilute the significant accelerator cost.

While it is achievable to schedule treatments in four or even five rooms, so that beam is readily available when needed, such facilities are quite massive in size, total cost and patient capacity. They really are treatment centers that require a large number of patients to be economically efficient. Many hospitals, however, do not have these numbers of patients (>1000 per year) and may find it difficult to finance such a massive project as the initial investment can easily surpass 100 or even 200 million USD.

As a consequence, a market for compact proton therapy systems featuring a single treatment room emerged which has grown rapidly in the last couple of years (see **Figure 1.4**). While efforts were made to reduce the size of the systems, they are in fact even more designed for lowest cost. At the same time new and cheaper generations of accelerators, such as the superconducting synchro-cyclotrons S250™ from Mevion and S2C2 from IBA or the compact synchrotron Radiance 330® from PROTOM, were developed. Now the need to distribute the accelerator cost to multiple treatment rooms is weakened but more importantly the total investment sum for a single-room proton therapy facility drops into a region (approx. 30 million USD) that is by far easier to finance for health care providers. Thus, compact proton therapy systems do not require extraordinary high patient numbers to be operated efficiently and become affordable for a larger number of hospitals.

In the figures below some commercial compact single-room systems are shown.

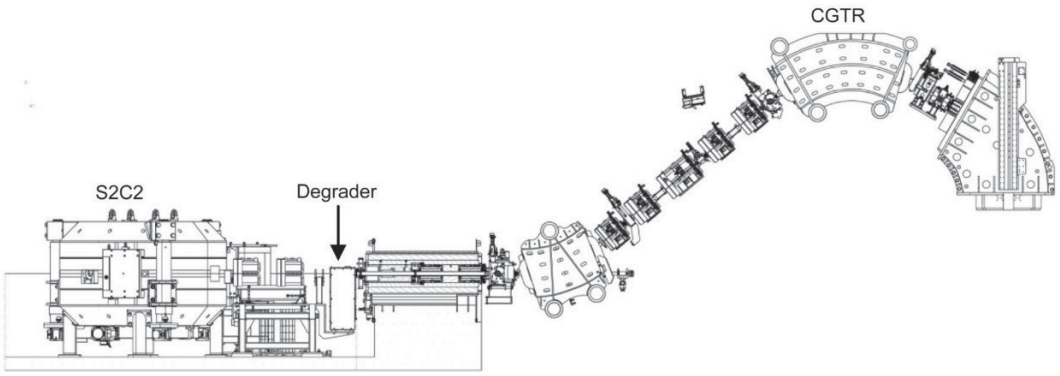

Figure 3.25: *Side-view of the Proteus®One system. One novelty is the short beamline between cyclotron and gantry with only four quadrupoles, two inside a drum that is put in the radiation shielding wall (not depicted), and the degrader. Another novelty is upstream scanning, where the deflection magnets are situated before the final 60° dipole [88].*

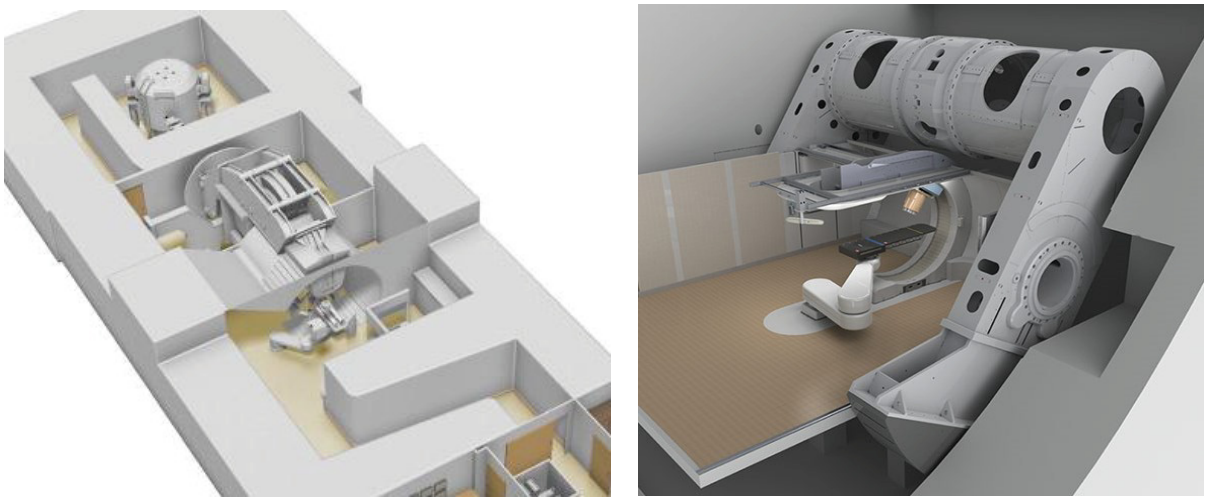

Figure 3.26: *Left: The IBA Proteus®One System with superconducting cyclotron S2C2 and the half-rotating gantry directly attached, makes a compact footprint of approx. 14 x 27 m² [89]. Right: The Mevion S250™ system has the superconducting accelerator mounted on the rotatable mechanical structure. It occupies a volume of approx. 14 x 14 x 14 m³ [90, 91].*

The IBA Proteus®One, the Mevion S250™ series and the Varian ProBeam®Compact (see **Figure 1.5**) products are examples of compact systems that have found success in the market [13]. The specialty of the IBA system is a new design type with upstream scanning through a 60 degrees big bore dipole (compare section 3.4.1). It is probably the most compact gantry design with

resistive magnets. Varian recently announced a size optimized version of the downstream scanning gantry with 360° of gantry angle rotation (in contrast to mechanical solutions by IBA and Mevion that limit the rotation to about 190-220° [90, 96]), but it is not clinical yet [92]. The Mevion system follows a very different concept by rotating the cyclotron, mounted to a gantry, about the patient – a concept first proposed by Blosser in 1982 [93]. Other compact gantries include two designs from PSI. The newer Gantry 2 at PSI, see **Figure 3.22**, was the first proton gantry to feature full upstream scanning (both directions) and the 'old' PSI Gantry 1 from the mid-90s is in fact still the most compact gantry in the world – achieved by an eccentric mechanical concept (see **Figure 3.31**).

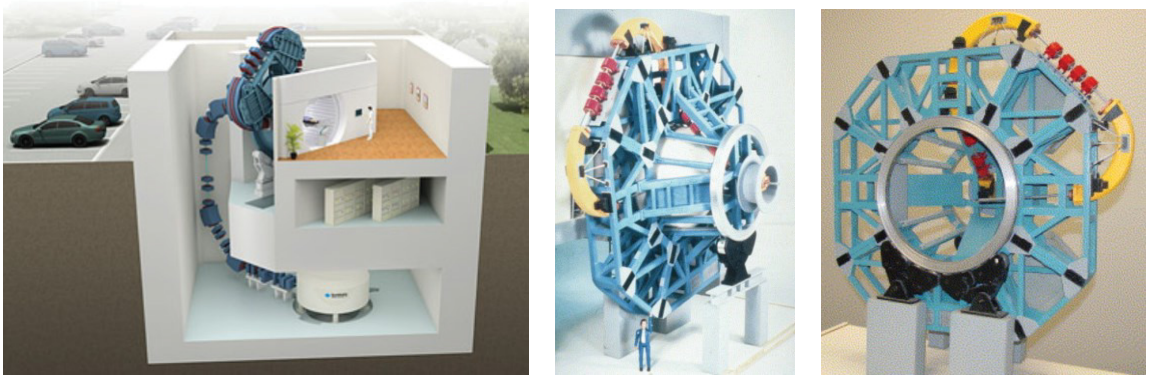

Figure 3.27: *Left picture: The corkscrew gantry of Sumitomo for the Aizawa Hospital [94]. Middle and right: A model of the Loma Linda corkscrew gantry [95].*

Besides the already mentioned ProNova SC360 (see **Figure 1.7**) another very short gantry design is the so-called corkscrew gantry. It was first built at Loma Linda, actually the first gantry overall, and has recently been turned into a commercial product by Sumitomo Heavy Industries with modern pencil beam scanning. Sumitomo combines the gantry with a vertical system layout for compact one- or two-room facilities. The first system went clinical in 2014 at the Aizawa hospital in Matsumoto, Japan.

Table 3.4: *List of vendors for compact proton therapy equipment.*

Manufacturer	**Product name**	**More information**
IBA	Proteus®One	www.iba-protontherapy.com
Mevion Medical Systems	S250™ series	www.mevion.com
Sumitomo Heavy Industries	SpaceSavingCompact PTS	www.shi.co.jp/english/products/industry/proton/index.html
ProTom International	Radiance 330®	www.protominternational.com
Provision Healthcare	ProNova SC360	www.provisionhealthcare.com
Varian Medical Systems	ProBeam®Compact ProBeam®360°	www.varian.com

3.4 Gantry diameter

The most obvious difference between conventional and proton therapy gantries is the size difference in diameter. Conventional machines easily fit into a bunker with regular ceiling height of approximately 3 meters while a proton therapy gantry has a typical diameter of about 10 meters.

Figure 3.28: *Varian's proton therapy and conventional radiotherapy products: ProBeam® (left) and TrueBeam® (right). Note that the black couch-tops are of similar size. Source: Courtesy of Varian.*

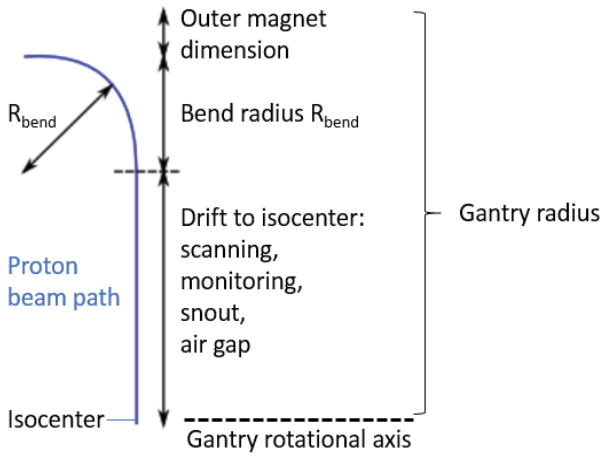

Figure 3.29: *Sketch of radial size contributors for proton therapy gantries. The blue line represents the simplified beam path of the proton beam.*

There are two main contributors to the radial size of the proton gantry: the bend radius of the dipole sitting atop of the gantry and whatever distance is needed down-stream this *final bend* before the isocenter. Additionally, the outer dimension of the bend magnet itself adds to the gantry radius: the space taken by the aperture and the magnet hull etc. that adds to the designed proton path (blue line in **Figure 3.29**).

For a more compact gantry one can try to minimize the drift after the bend to the isocenter and/or reduce the bend radius. However, the diameter of a proton therapy gantry is a tricky thing to tackle – at least as long as focused pencil beam scanning shall be possible. All concepts described in the table blow that reduce the gantry radius have severe disadvantages and require a thorough win/loss analysis. Simply put, they do not seem to change the game and may not be worth the trouble.

Table 3.5: *An overview of three well-discussed options to reduce the radius of proton therapy gantries.*

Concept	Radius reduction	Disadvantages
Upstream scanning	~ 1 meter	Significantly complicating the beam optics, not feasible with achromatic bends, requires large bore bend magnets which are expensive and heavy. Superconducting dipoles are not feasible because of their slow ramp speed.
Eccentric gantry	~ 0.5 – 1 meter	The complication for the patient positioner, both from a technical and user point of view, should be well understood and weighed against the modest radius improvement.
Stronger (SC) magnets	~ 0.5 – 1 meter	Simply exchanging a regular dipole with a superconducting magnet can make a meter difference in gantry radius, which is not insignificant. However, one should be aware of the complications that come with superconductivity: High cost, limited ramping speed and technical complexity. The often-cited additional advantage of low weight, is mostly overestimated as weight savings of even 10 tons, do not dramatically influence the cost of a gantry.

3.4.1 Upstream and downstream scanning

In the context of the location of the scanning magnets, “upstream” and “downstream” refer to the final bend. Upstream scanning means that the scanning magnets are located upstream the final bend, whereas a downstream scanning solution puts them in the vertical path between the final bend and the isocenter. By placing the scanning magnets upstream, the drift length downstream the final bend can be reduced, yielding a more compact gantry design. Another often-cited advantage is that the distance between the deflecting scanning magnets and the isocenter, called the source-to-axis-distance (SAD), is enlarged. Actually, one can design the beam optics to achieve an infinite (or at least close to infinite) SAD with upstream scanning which is also called parallel scanning [97]. The two proton gantry designs that employ upstream scanning are the PSI Gantry 2 and the IBA Proteus®One gantry.

3.4.2 The SAD and short-SAD downstream scanning

The SAD is said to have some clinical relevance. Many citations lead to the list of specifications for proton therapy, defined by Chu et al from LBNL in 1993 [98], who demand an SAD of at least 3 m. Chu also published a renowned standard reference for proton therapy instrumentation with Ludewigt and Renner in the same year [59]. For a short SAD the angles for the proton beams are larger to achieve the same field size at isocenter. It implies that the magnification factor between the intermediate field at the skin of the patient and at target depth is rather large, hence the proximal dose at the skin is increased. As the skin is a radiosensitive organ, this has caused concerns in the past [99], e.g. in a survey result that asked for a minimum of 4 m SAD [100]. However, if this is viewed with respect to x-ray treatments the argument is questionable. The radiation source in these machines are much closer to the patient and the entry dose level at the skin higher.

Indeed, recent discussions regarding the SAD for proton therapy, move towards accepting smaller SADs, especially in context of IMPT. A study by J. Wulff et al [101] summarized in [102] suggests that an SAD of 1-1.5 meters is sufficient. Proximal dose enhancements for IMPT plans were found to be negligible and the SAD should not be the primary rationale for upstream scanning.

The short-SAD downstream scanning approach aims at minimizing the drift between final bend and isocenter. It is technically limited by the maximal deflection angle achievable with reasonable magnets and by space requirements of other components. Furthermore, the resolution of the beam monitoring ionization chamber (see figure below) is a bottle neck geometrically. A realistic assumption is about 1.8 meters for the total drift length downstream the final bend magnet [101].

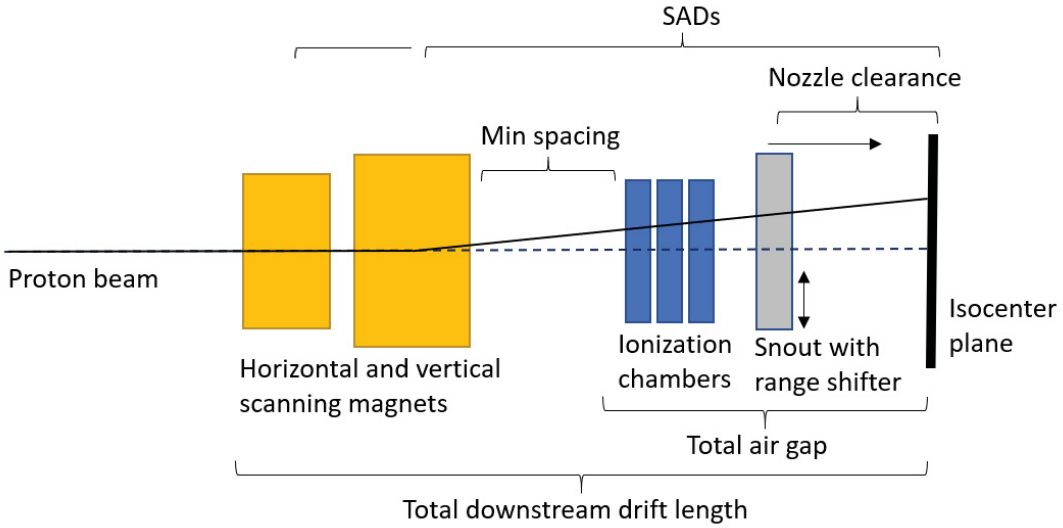

Figure 3.30: *Sketch of a downstream scanning nozzle. The beam enters from the left on the optical axis (dashed line) and is deflected in the two orthogonal directions by scanning dipole magnets. The beam is passing ionization chambers with spatial resolution to monitor the beam position (and dose). The ionization chambers can neither be moved to close to the isocenter (clearance for patient) nor to close to the scanning magnets (magnification factor for resolution). The optional range shifter reduces the beam energy. It can be moved close to the patient (isocenter plane) to minimize beam growth through scattering. The vacuum pipe (not depicted) extends up to the ionization chambers that are in air environment. Scattering in the chamber foils and in air contributes to the resultant beam size as well.*

3.4.3 Isocentric and eccentric gantries

Isocentric means that the reference point for the treatment coincides with the isocenter of the gantry rotation. Alternatively, one can move the treatment reference point beyond the gantry rotational axis. So-called eccentric gantries have been proposed many times and PSI did build one: the PSI Gantry 1 [103]. In this case the patient table position rotates with the gantry. However, clinicians have some concerns with this practice. Partly because of positioning uncertainty that comes with patient motion but foremost because of accessibility limitations to the patient in a high lifted table position (for irradiations from below). In fact, these concerns caused PSI to choose to not irradiate from certain angles at Gantry 1 [104]. And even though the problem seems manageable, no practical solution for the issue has been shown yet – and, notably, no commercial vendor has chosen an eccentric gantry concept. Thereby the PSI Gantry 1, designed in the 1990s, remains, until today, the most compact proton gantry with respect to diameter.

***Figure 3.31 Eccentric gantries**: The PSI Gantry 1 (left) during its installation in 1994 [105]. The patient table is mounted to the rotational gantry. The patient is positioned off the rotational axis thus moves on a circle during gantry rotation, see sketch on the right.*

Modern proton therapy gantries also offer CBCT[12] imaging at treatment position. A method that was adopted from conventional radiotherapy. This technique requires an isocentric gantry design.

3.4.4 Superconducting and normal conducting bend magnets

Another way to reduce the gantry diameter is to increase the magnetic field of the bend magnets. Since resistive magnets are limited to fields of about 1.5 T (see chapter 4), the employment of superconducting magnets has the potential of reducing the bend radius significantly. A 4.5 T dipole field, which would correspond to a bend radius of about 50 cm (see graph below), is easily reached and much stronger fields are certainly feasible (e.g. for the next LHC[13] upgrade 16 T dipoles are discussed and being developed [106]). However, as indicated in the introduction the technology brings a few complications. Chapter 4 deals with the potential and challenges of superconducting magnets for compact proton therapy gantries.

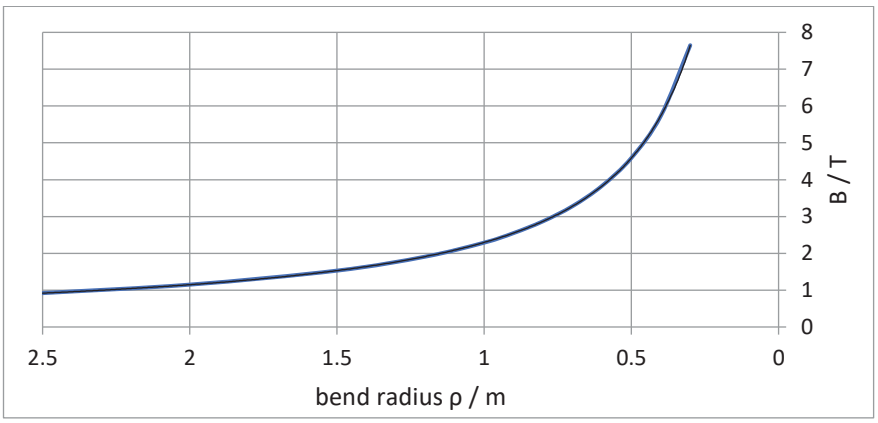

***Figure 3.32:** Bend radii and dipole fields for a 230 MeV proton beam (magnetic rigidity = 2.3 Tm). A field of 1.5 T (typical for resistive dipoles) means a bend radius of about 1.5 meters.*

[12] Cone Beam Computer Tomography, see [107].
[13] Large Hadron Collider, Geneva

3.5 Achromatic bends

3.5.1 Dispersion function and momentum acceptance

The problem with dispersion is that it is generated in one dipole and a second dipole is required to compensate the effect. Only *local achromats* can eliminate dispersion in an individual bend of the beamline. These local achromats include serially alternating focusing of the beam that manipulates the dispersion function in a way so that the same dipole field of the bend compensates it. The discussed *AGCCT* magnet is a local achromat. Double-bend achromats, such as the ProNova type, also fall into this category. The contrary is the concept of *global achromats*. Here, all bends of the gantry beamline are combined to correct dispersion. Apart from ProNova all other existing gantries employ the concept of global achromaticity. An example is shown in **Figure 3.35**. Due to the distance between the bends the dispersion function often becomes relatively large, limiting the *momentum acceptance* of the optics. The momentum acceptance is the extent of relative momentum deviation of protons with respect to the reference momentum that are still transported by the beam optics. It is decisive for the transmission efficiency of the beamline which not only impacts performance but alleviates neutron radiation shielding needs as well. The momentum acceptance of a typical (globally achromatic) gantry is +/- 1% *dp/p*. The achromatic double bend of the ProNova SC360 design achieves a moderately improved +/- 3% d*p/p* [39] and follows a similar concept as depicted in **Figure 3.33**.

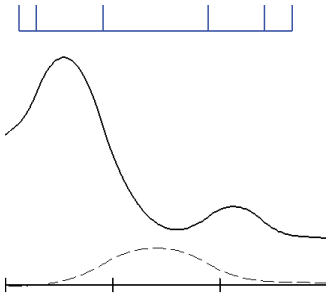

Figure 3.34: *Transport graphics of an earlier version of the AGCCT (design work by A. Gerbershagen). The continuous lines display the beam envelope. The dashed line the dispersion function. The blue boxes are sections of combined function magnets (dipole and quadrupole fields).*

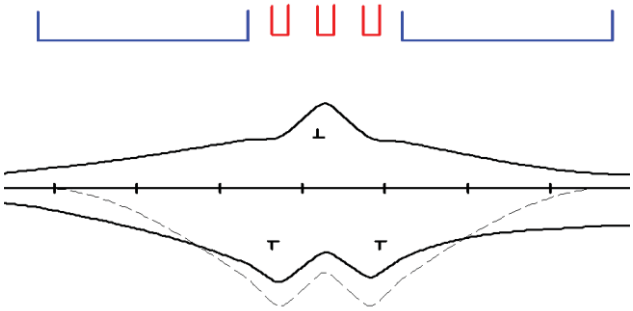

Figure 3.33: *Achromatic beam optics of a conventional DQQQD layout (blue boxes = dipoles (D), red boxes = quadrupoles (Q)). The dashed line shows dispersion, continuous line the envelope.*

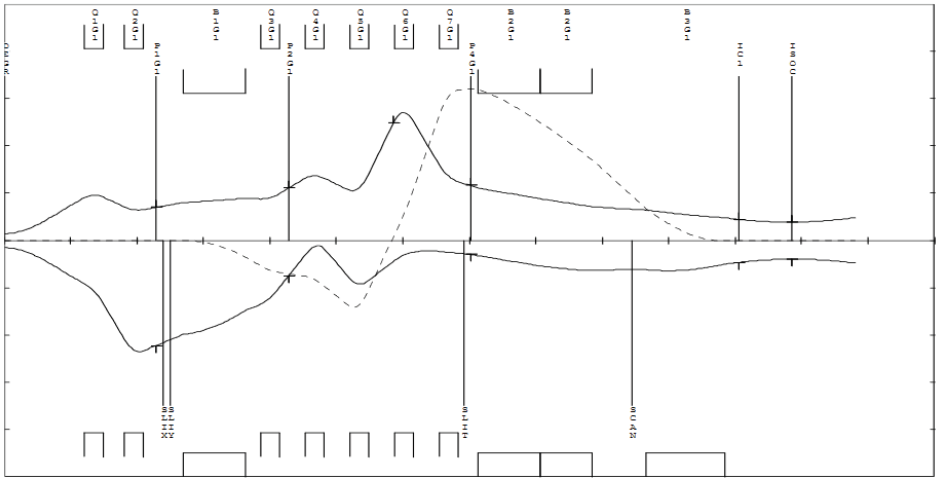

Figure 3.35 Example global achromat: *Beam optics of the ProteusOne gantry [96]. The beam enters the gantry from the left. The first dipole introduces dispersion (dashed line) which is cancelled with the other dipoles, which are indicated by the rectangular boxes touching the bottom frame of the picture. The thinner boxes indicate quadrupoles. The continuous lines show the envelope of the beam.*

3.5.2 Gantry size reductions with local achromats

However, apart from an improved momentum acceptance a local achromat can help in the grand scheme of reducing size and cost of compact proton therapy.

Compact proton therapy systems powered by a cyclotron, the accelerator of choice for cost effectiveness, require an energy degrader to adjust the beam energy. The straggling effect in the energy degrader will, however, induce an energy spread in the beam. The most logical arrangement of these components, which is also found in Varian's ProBeam® and IBA's Proteus®One systems, is:

Figure 3.36: *Conventional sequence of major elements of a compact gantry-based proton therapy system. For radiation safety, the cyclotron-room and the treatment-room with the gantry are usually in two different bunker vaults, separated by a radiation shielding wall (with a conduit for the beam).*

Even though the material thickness of the degrader is only in the order of 20 cm it consumes much more space than that. To achieve optimal and stable transmission, the beam must be centered and focused on the degrader, which calls for additional steering and focusing magnets. Unfortunately, the beam exiting the cyclotron is neither well focused nor necessarily stable regarding its position or angle. It is neither symmetric [108]. Therefore, a set of quadrupole and steering dipoles are required that in connection with an appropriate drift distance can prepare suitable beam properties. Those elements and their spacing take about 2 to 3 meters in length. As described above, the output beam of the degrader requires collimation. With the divergence collimator located about 1 meter downstream of the degrader, a total length of 3 to 4 meters is required between the cyclotron and the rotating gantry. This section of the beamline is commonly called the *front-end*. The accelerator (and front-end) are separated from the treatment-room with the gantry by a radiation shielding wall with a typical thickness of about 1.5 meters.

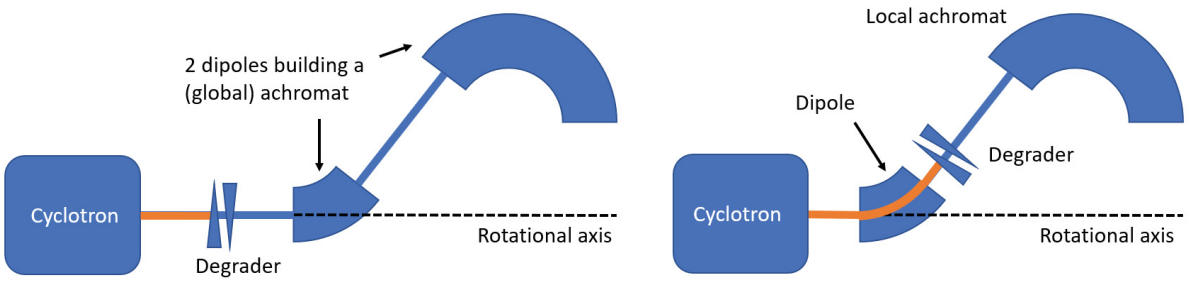

Figure 3.37: *Schematics of compact gantry systems with conventional global achromatic optics (left) and with a locally achromatic final bend (right), where the degrader has been placed on the gantry spine. The orange section of the beamline indicates the part where the energy spread in the beam is minimal (before beam degradation). The illustrations do not show quadrupoles and other relevant details of the beamline.*

The bend magnets in the gantry build a global achromatic system, where a first dipole produces dispersion and the following bend(s) compensate it. If, however, the final bend of the gantry is a local achromat, it allows to place the degrader, *the* source of the energy spread in the beam, on the gantry spine and thereby merge the beamline of the front-end with the gantry optics [109].

This has quite a few advantages:

- The focusing and beam steering between cyclotron and degrader can now be done “through” the radiation shielding wall and the first bend. Instead of arranging the front-end, the shielding wall and the gantry beamline successively, they can be merged into another. As a result, the system is shorter.
- The shielding wall thickness can be reduced since the degrader, the major neutron source during irradiation, is not behind the wall and, more importantly, not pointing towards the patient any longer.
- Since the beam up to the degrader is quasi-monoenergetic and of fixed energy, the first bend of the gantry becomes trivial. It is later shown that the small energy spread in the beam does not require achromatic optics. Theoretically, the first bend can then even be a permanent-magnet dipole.
- The emittance from the cyclotron is much smaller than after beam degradation. Hence aperture of the magnets and the beam pipe of the front-end can be significantly reduced (by more than 50 %).
- A local achromat with sufficient momentum acceptance increases the transmission and thereby facilitates larger dose rates (beam currents) at the isocenter. Its effect on beam delivery time is quite crucial for many applications as discussed above in sections 1.2.3 and 3.2.

4 SUPERCONDUCTING MAGNETS FOR PROTON THERAPY BEAMLINES

4.1 Superconductivity

General superconducting behavior, the resistance-free transport of electric current below a critical temperature T_c, was first discovered in 1911 by Heike Kamerling Onnes [110] and in the fifties and sixties the first useful superconducting Niobium-alloys for magnet coils were found [111, 112]. Due to their need of very low temperatures (below about 30 K) they are called low-temperature superconductors (LTS) and magnets built with Niobium-conductors typically operate at liquid helium temperatures. Another milestone was reached in 1986 when ceramic superconductors, known as high-temperature superconductors (HTS), were found to have much higher critical temperatures and can even exceed the important 77 K mark, where cooling with inexpensive liquid nitrogen becomes feasible [113]. Eight Nobel prizes [114] have been awarded to achievements in this field of physics and material science. However, besides the critical temperature also the magnetic field and the current density in the conductor define the so-called critical surface, i.e. the threshold of preservation of the superconducting property. Thus, all three parameters are important.

One differentiates between type I and type II superconductors. The first type are superconductors that completely rely on the Meissner effect, the shielding of an applied magnetic field in the conductor up to a critical state defined by temperature or field [115]. If the field or temperature becomes too high, the conductor instantly turns resistive. Unfortunately, the critical fields for type I conductors are so low that they are not suitable for magnets. In contrast, type II superconductors show a mixed state, above the Meissner state, where the conductor is partially penetrated by the field but is still superconducting up to a second critical field strength H_{c2} [115]. These conductors allow material dependent maximal fields that by far surpass the fields generated by conventional magnets. All LTS and HTS conductors discussed for beamline magnets are type II superconductors.

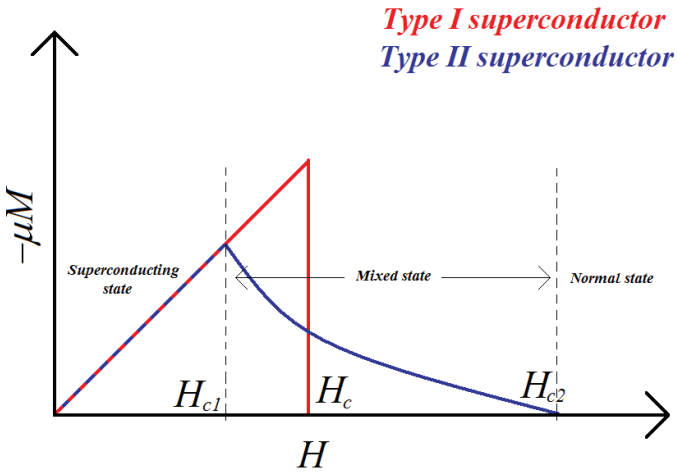

Figure 4.38: *Magnetization against H-field of type I and type II superconductors. Source: [116].*

4.2 Field limitations of iron yoke magnets

Due to the saturation of magnetization in iron [117] the maximum field in room temperature magnets with normal conducting coils is limited. For higher magnetic flux densities (*B*-field) in the saturation region, the magnetic field strength (*H*-field) strongly increases as the magnetic permeability of the material μ_r drops.

$$\vec{B} = \mu_0 \mu_r \vec{H} \tag{4.1}$$

The non-linearity starts at about 1.3 T and the *B*-*H* curve quickly flattens above 1.5 T. Room temperature magnets with greater *B*-fields are disproportionally heavy due to the large amount of iron that is needed to provide the required *H*-field. The weight difference of a 1.5 T and a 1.8 T 135° dipole as used in conventional proton gantries is about 10 tons or a factor of two. Hence it is not practical (nor economical) to use room temperature magnets with fields surpassing 1.5 T by much. Being limited to this field, the bend radius of such magnets for 230 MeV protons is about 1.5 meters.

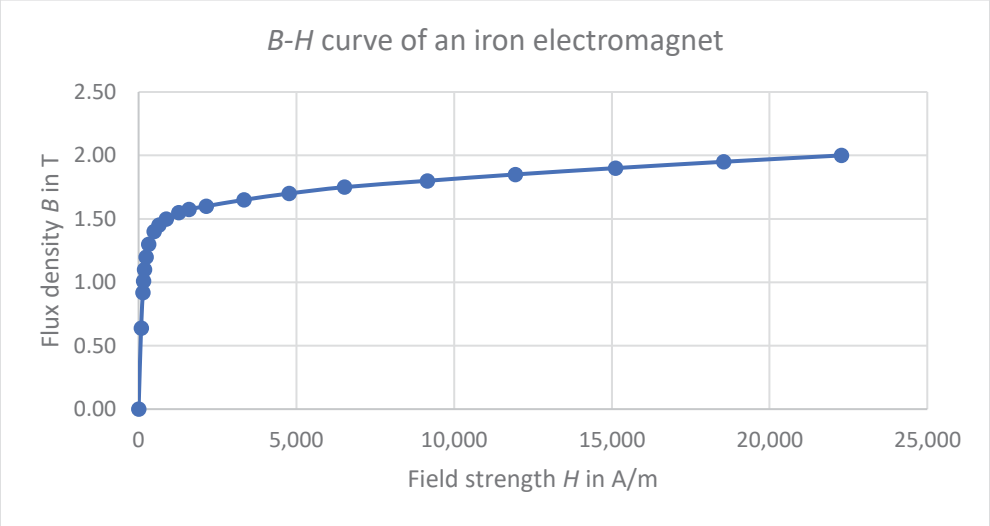

Figure 4.39: B-H curve of an exemplary iron magnet, designed for proton therapy [118].

Of course, stronger fields are desirable to reduce the bend radius and achieve more compact beamlines (compare **Figure 3.32**). Superconducting materials can carry such high currents that the ferromagnetic iron is no longer needed as the coils directly generate enormous magnetic fields without the boosting factor μ_r of ferromagnetic materials. For a solenoid example *B* is:

$$B = \cancel{\mu_r}\, \mu_0 I, \tag{4.2}$$

where $n = N/l$ is the number of windings divided by the length of the coil. The term $I \cdot n$ is called ampere turns or abbreviated *ampturns*.

Iron may still be used in some cases for field contribution (savings of expensive conductor), for field shaping (directing of flux lines) or for stray field shielding. However, in high magnetic fields the iron is fully saturated and no longer an equipotential. As a consequence the shape of the field cannot be simply correlated to the iron pole shape. It limits the possibility to utilize iron to maintain field homogeneity, especially since the iron-contribution becomes non-linear. Superconducting magnets that employ large amounts of iron (as poles or yokes) are called

superferric. For example, the superconducting bends in the ProNova gantry are superferric and produce up to 4.1 Tesla [119].

4.3 The promise of superconducting magnets

Additional to the already discussed benefit of higher fields and smaller bend radii, superconducting magnets allow more degrees of freedom to design combined function magnets. These magnets not only produce a dipole field but superimpose another or multiple multipole fields, which can for example add strong focusing power or a correcting sextupole to a bend magnet. The *AGCCT* design, which is detailed below, is a good example. With their high values of ampturns, superconducting magnets are often designed as air coils which significantly reduces the weight of the magnets. Another advantage of air coil magnets is the absence of any hysteresis effects in the ramping cycle. Also, the power consumption of the non-resistive coils is very low – although the overall power balance is very case specific and highly depending on the power efficiency of the cooling system, see section 4.4.2.

The potential benefits:

- High fields leading to compact beamlines
- Combined function magnets
- Low weight
- Low power consumption
- No hysteresis effects of iron

4.4 Problems of superconducting magnets

A relatively broad overview over the application of superconducting magnet technology in particle therapy equipment, its advantages and challenges, can also be found in the PSI paper [120]. Some statements therein with regards to size benefits and required momentum acceptance are picked up again in the results discussion of chapter 6. Following the main issues when employing superconducting technology for proton therapy gantries are discussed.

4.4.1 Material costs

Unfortunately, the benefits of superconductivity come at very literal costs. The conductor itself can be very expensive with a huge dynamic depending on the conductor type. The cheapest NbTi wires start at about 1 $/m while HTS tapes typically exceed 20 or 30 $/m.

4.4.2 Cooling

The second main cost factor is the cooling that is required to operate at superconducting conditions. For a rotating gantry system using a liquid helium bath is not a practical option which leaves one with cryocoolers for cryogen-free conduction cooling. These cryocoolers are not only expensive themselves (the suitable Sumitomo RDK-415 which cools with 1.5 W at 4.2 K costs approximately 40 k$) but also cause recurring costs due to their 24/7 power consumption (7 kW for the RDK-415) and the periodical servicing needs. More expensive conductors can have the advantage of a larger temperature margin or a higher operating temperature. Depending on many boundary conditions the cost of more advanced conductors can be offset – to some extent – by savings on the cooling side.

4.4.3 Ramping

The most challenging problem for the application of superconducting magnets in proton therapy beamlines is ramping. Ramping means changing the magnitude of the *B*-field by "ramping" the power supplies, i.e. the electric current in the coils. The variation of magnetic field causes AC-losses in the conductor (energy that is converted to heat) and eddy currents in other resistive materials in the cold mass, e.g. metal mandrels that physically support the coil. The losses stem from the voltage inductance by the *B*-field variation and always lead to a local temperature rise reducing the temperature margin of the superconducting state. The AC losses in the conductor can be divided in two effects [121]:

i) Type II superconductors operating above the Meissner state show hysteretic *B(H)* behavior generated via the magnetization due to self-shielding currents in the conductor. For a typical rectangular multi-filamentary conductor the power of the *hysteresis losses* per unit volume can be estimated with

$$P_{hystersis} = \frac{2}{3\pi} J_c(B)\, d_{eff} \frac{dB_t}{dt} \left(1 + \frac{J^2}{Jc^2}\right), \tag{4.3}$$

where J_c is the critical current at field B, d_{eff} is the filament diameter and dB_t/dt is the rate of field variation transverse to the conductor, i.e. the ramp rate. The correction factor $(1 + J^2/J_c^2)$ takes the transport current into account. Consequently, the hysteresis losses are proportional to the ramp rate and the filament size.

ii) *Inter-filament coupling losses* occur when current flows in between the filament through the resistive copper matrix. They are different in the filamentry and the outer copper region and can be calculated per unit volume

$$P_{filamentry} = \frac{2}{\mu_0} \left(\frac{dB_t}{dt}\right)^2 \tau, \tag{4.4}$$

with where τ is a time constant for the decay time and is proportional to the square of the twist pitch T_P, and

$$P_{outer\ region} = \left(\frac{dB_t}{dt}\right)^2 \frac{{T_P}^2}{48\,\rho_{cu}} \frac{u}{c}. \tag{4.5}$$

The power loss in the outer copper region is also proportional to the square of the ramp rate and the square of the twist pitch. The factor u/c takes the ratio of outer and inner region of the conductor into account but is of no further interest here.

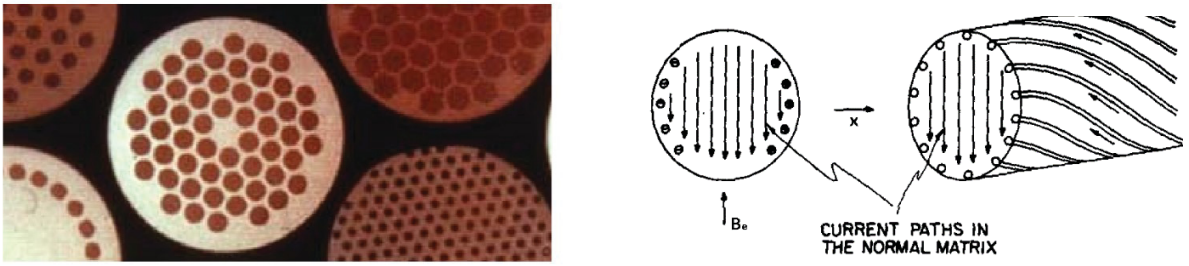

Figure 4.40: *Left: Picture of various NbTi wires, differing in filament number and sizes. The SC material is embedded in copper. Source: [122]. Right bottom: Schematic illustration of the helical twist [123].*

The conclusion is that to minimize the heat loads during ramping, one wants to keep the filament size and the twist pitch as small as possible, which drives costs up (especially the reduction in filament size). And most importantly, the ramp rate is a very decisive factor as it contributes quadratically to the coupling losses.

Eddy currents that appear in the resistive metals of the support structure of the cold mass, can effectively be dealt with by means of lamination. Laminating the structure, however, not only has possible influence on the rigidity of the supporting structure but can prevent good heat conductance from one end of the magnet to the other. Electrically insulating the individual laminations and maximizing the heat conductance across laminations is a contradicting task. The problem is even more pronounced due to the fact that materials with good heat conductance are also good electric conductors which in turn enhance eddy currents.

4.4.4 Cryostat and cold mass

The typical operating temperature for LTS coils is 4.2 K which is the boiling point of Helium at normal pressure. A cryostat is needed that shields the heat of the room-temperature environment from the cold mass. In the studied cases the cold mass is cooled with conduction cooling. The cryostat assembly basically consists of two stages: The first stage is a shield cooled to an intermediate temperature in the range of 30 to 70 K. Inside the shield sits the cold mass (the coil) at the second stage of about 4 K. The outer cryostat shell is at room temperature. Insulation is achieved with evacuation between the stages and low-heat conducting connections (e.g. G-10[14]) as well as superinsulation foils to prevent heating form thermal radiation.

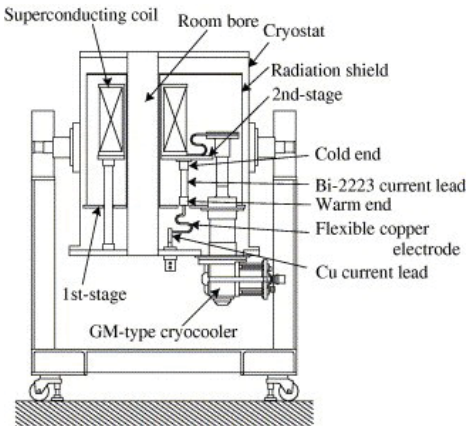

Figure 4.41: *Example of a cryostat for a conduction cooled SC magnet with two cooling stages [124].*

The cryocoolers connect to either the first or both stages. Heat generated in the cold mass, e.g. by eddy currents during ramping, must be extracted by heat conduction only. Therefore, the use of materials with adequate heat conductance is mandatory.

The ability of a superconductor to carry large currents has the complication that the current leads, the electric cable connection from room temperature to the cold mass, require a respective cross section. These thick conductor leads constitute a major heat leak to the magnet. The higher the design current, the higher is the heat load from the leads. The leads are divided into a first set from room temperature to the first stage (copper cables) and a second HTS lead set carrying the current from the first stage at e.g. 50 K to the coil at ~ 4 K.

4.4.5 Manufacturability

The various conductor options also greatly vary regarding their ease of coil manufacturing. NbTi wires or cables are mechanically quite strong thus easy to wind. Some other conductors, particularly Nb_3Sn, require heat treatment and become brittle and sensitive afterwards. HTS

[14] G-10 is a widely used glass-fibre epoxy composite and has less than factor 20 heat conductance compared to steel [125].

conductors which are ceramic compositions are very difficult to produce. The few HTS tapes commercially available are limited in maximal length as well as minimal bend radii, tensile stress tolerance etc. Hence, due to the individual mechanical properties, the initial coil design should consider the conductor choice. Manufacturing is a third major cost contributor which is strongly depending on coil geometry and conductor type. Particularly difficult are concave winding shapes for brittle conductors (e.g. Nb_3Sn).

4.4.6 Stray fields

The strong magnetic fields of superconducting magnets have the drawback that one has to deal with the corresponding strong stray fields. Stray fields, if high, can be problematic for the beam optics, other magnets in the vicinity (crosstalk) and field sensitive equipment nearby such as x-ray tubes for example. Apart from these technical considerations, a stray field can cause complications from the regulatory side as well. If the area accessible to the patient is clear of magnetic fields of 5 mT or higher, no precautions, e.g. with regards to pacemakers, must be taken. While other solutions are possible (see MRI applications) it is still desirable to keep the stray field below 5 mT in the patient area.

The stray field not only depends on the strength of the primary field but is also strongly influenced by the coil design. Especially large apertures in magnets are problematic. Iron yoke magnets, in contrast to air coil designs, have the advantage that the return flux is guided through the iron, greatly reducing the stray field. Stray fields can be reduced with passive or active shielding, which is either heavy (iron) or expensive (additional magnet coils that create a counter field).

4.4.7 Special issues of HTS

HTS conductors have a few special physical issues which make their use for beamline magnets more difficult [126]. Here, they shall be listed briefly. However, the dominant concerns regarding HTS conductors for magnets remain their costs and unfavorable mechanical material properties.

- *Magnetization*: When the current through an HTS tape is changed, the generated magnetic field does not follow instantly but asymptotic with a time delay. This is due to magnetization effects in the material. The time constants in this effect are orders of magnitudes larger than required by a typical application in a proton beamline which is under 1 second for a field change of 1 % B_{max}.

- *Quench detection:* The slow propagation of a quench due to the high T_C margin delays the point of detection. Preventing coil damage in HTS magnets becomes difficult because, before a measurable voltage is reached, the quench location may have reached a temperature high enough to damage the conductor. On the other hand, HTS magnets are typically very stable so that conservative design can be used to minimize risk of quench events.

- *Flux flow resistance:* The current transported through a superconductor applies a Lorentz force on any flux vortex in the conductor causing it to drift sideways. The motion of the flux vortex (also called flux flow) on the other hand induces an electric field. This means with flux flow present the conductor is no longer non-resistive. Pinning forces hold the

vortices in place but can only match the Lorentz force (also called Magnus force) up to a maximum current. Hence the maximal current in the conductor is limited which reduces the critical current density of HTS significantly, especially for BSCCO.

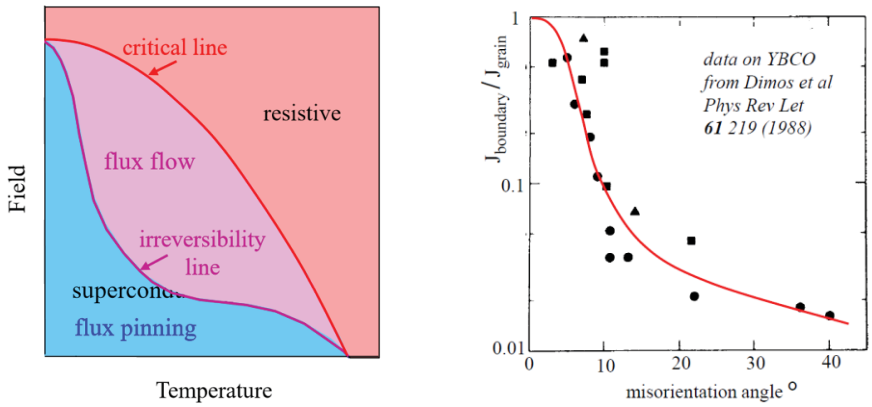

Figure 4.42: *Left: Flux flow reduces the critical field for HTS conductors, only the blue region is superconducting. Right: Measurements of the effect of grain misorientation by Dimos et al. Note the logarithmic scale. Source: [127].*

- *Grain boundary mismatch:* The grains of the conductor have various orientations of the crystal planes. Within the grains the critical current is high, but at the intersection of grains it depends on the angle of the different plane orientations (and quickly drops with few degrees of misalignment). BSCCO conductor tapes usually do not have this problem but the alignment in YBCO is more difficult.

4.5 Overview of suitable superconductors

To build a beamline magnet a number of parameters of superconductors must be considered. Due to the conduction cooling the temperature margin in the conductor is very important. It depends on the magnetic field, the current density and the operating temperature. The biggest issue in this regard are AC-losses and eddy currents that occur during magnet ramping (dynamic case) and reduce the margin as present in the static case.

To maintain superconducting conditions (and prevent quenching), the conductor must stay below the critical surface in the *I-B-T* plot (see **Figure 4.43**). Often these plots are shown in 2D graphs with one of the variables held constant, suitable to determine the boundaries for the static case. In these graphs superconductors can be compared. Conductors with high critical temperatures, such as MgB_2 with $T_c = 39$ K, are not necessarily advantageous for magnets. E.g. the low critical current density of MgB_2 makes it difficult to generate high fields at higher temperatures (or even at the common 4.2 K).

It is worth noting that to date almost all superconducting accelerator magnets are still made of NbTi (55 years after its discovery) while only a few special high field magnets use Nb_3Sn.

Since their discovery in 1986 high-temperature superconductors have been the promise to facilitate breakthrough superconducting magnet designs. Advancements in field strength or much higher operating temperatures seem feasible but no beamline magnets have been built yet beyond a prototype stage. Operating at much higher temperatures, possibly at 77 K, the boiling point of nitrogen, would indeed simplify the cooling of superconductors tremendously. But even an operating temperature at about 30 K could lead to less expensive single stage cryostats and much more efficient cooling. HTS, however, are difficult to produce, very expensive and

have some delicate drawbacks as explained above. High-temperature conductors that have shown the biggest promise for the manufacturing of beamline magnets are YBCO and BSCCO (2212, 2223) tapes. They are commercially available in various versions which differ regarding their magnetic and mechanical properties. Interesting as well, considering the similarity to LTS properties, is multi-filamentary Bi-2212, which however requires very high temperature heat treatment [165].

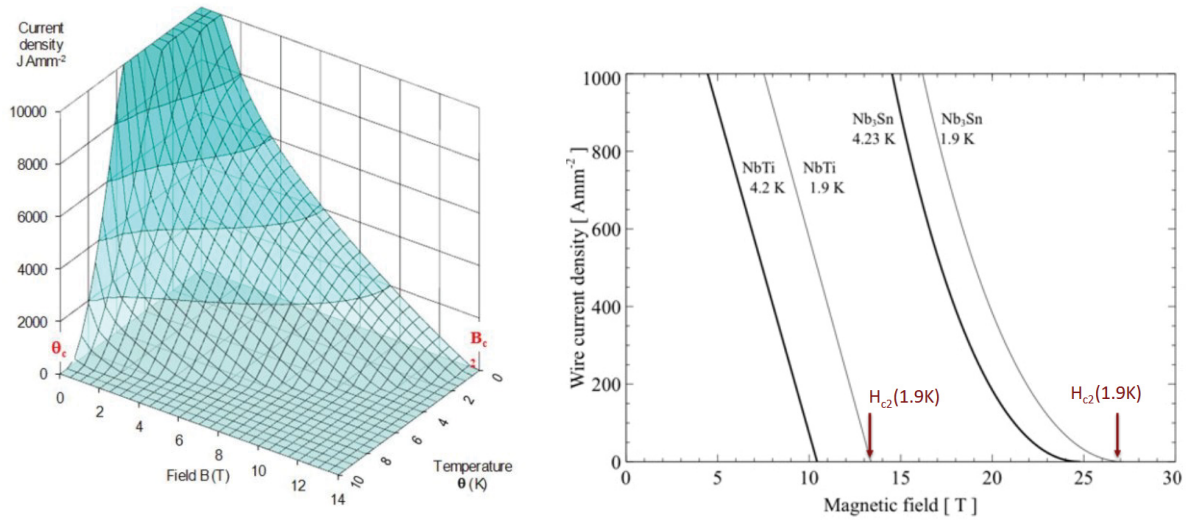

Figure 4.43: *Left: I-B-T surface plot of NbTi [122]. Right: Comparison of NbTi and Nb_3Sn for two temperatures showing the maximal field dependence on current density [128].*

For LTS conductors other properties than the material compositions have decisive influence on the performance of the conductor as well. LTS wires are built of NbTi or Nb_3Sn filaments in a copper matrix, see **Figure 4.40**. As explained above, the ratio of copper to conductor, the filament diameter and the twist pitch are most important for the dynamic performance of the wire. The copper matrix is needed for multiple reasons. Copper has good mechanical properties and can protect the SC filaments in case of instabilities or quenches with its heat and electric conductance.

4.6 Design of the *AGCCT*

The *AGCCT* design is described in detail in [32]. Here a short overview of its principles and design is given for a clearer picture of the context of the studied beam optics. It is based on the *CCT* (canted-cosine-theta) winding scheme where the combination of two layers produce an orthogonal field with respect to the symmetry axis of the torus they are wound upon [129, 167, 168]. The canted, i.e. inclined, winding turns also produce a parallel field component (solenoid field), but the inverted field of the complementing layer cancels this component, see **Figure 4.44**.

The conductor (wires or Rutherford cable) sits in channels that are machined in each metal torus that makes a layer. The so-called ribs between the channels intercept the Lorentz forces and provide exceptional stability, warranting minimal conductor displacement during magnet excitation. This feature makes the *CCT* also promising candidate for high-field accelerator magnets [130, 131]. Below is an example of a straight *CCT* which produces a dipole field.

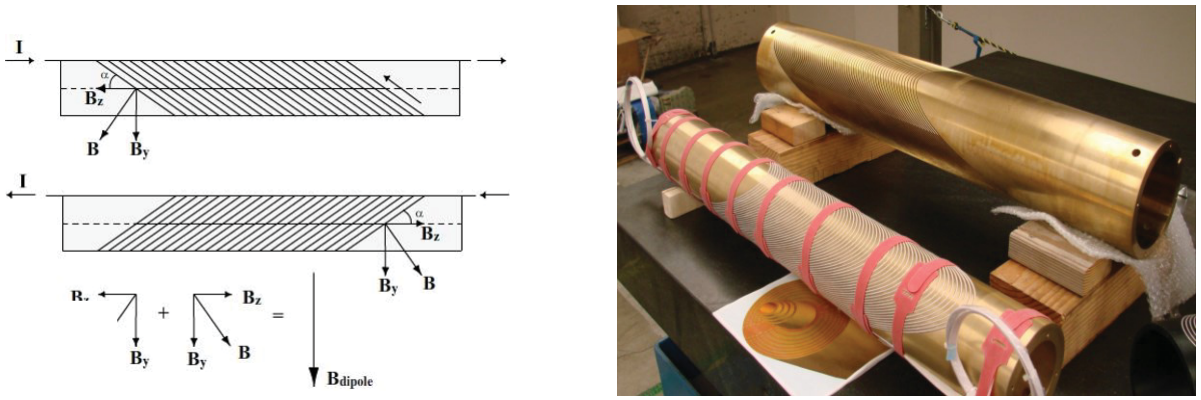

Figure 4.44: *Left: Principle of the canted-cosine-theta field generation. The two B_z components cancel each other and the two layers generate a B_y dipole field. Right: Photo of two straight CCT layers taken at LBNL (different magnet). Source: Courtesy of LBNL.*

The *AGCCT* is a curved version of the *CCT*[15]. The particular design studied here has a total of four layers. The outer two generate a dipole and the inner two a quadrupole field. A specialty in the quadrupole layers is a partition into five different sections along the length of the magnet (see **Figure 4.45**). From one section to the next the quadrupole alternates between focusing (F) and defocusing (D) which is facilitated by a reversal of the winding direction, i.e. a reversal of the direction of the electrical current flow. In contrast to other alternating-gradient concepts the spacing between the quadrupole sections is minimal (see plotted field in **Figure 5.64** in chapter 5).

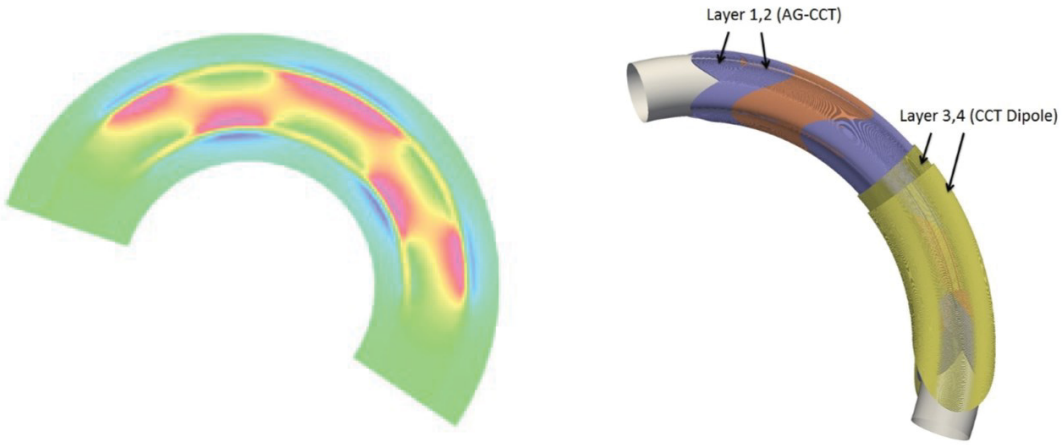

Figure 4.45: *Left: The two inner layers of the AGCCT produce alternating quadrupole fields as depicted by a field map (generated with Opera [32]), which shows the vertical field in the bending plane. Right: The two dipole layers are wrapped around the quadrupole layers. The image shows a cut model, the dipole actually extends over the entire length of the magnet. Source: Courtesy of LBNL.*

This feature not only makes a combined function magnet but enables achromatic beam optics. Hence the *AGCCT* is designed as a local achromatic bending magnet. Additionally, the second and fourth quadrupole sections contain a sextuple field component (for the correction of chromatic aberrations). It is generated by a slight modification of the winding path in these sections, displayed in **Figure 4.47**. **Table 4.6** gives an overview over the fields and gradients produced by the *AGCCT*.

[15] Before the concept of the *AGCCT* was developed between PSI, LBNL and later Varian, D.S. Robin et al proposed in 2011 to use a curved *CCT* for an ion beam therapy gantry [163].

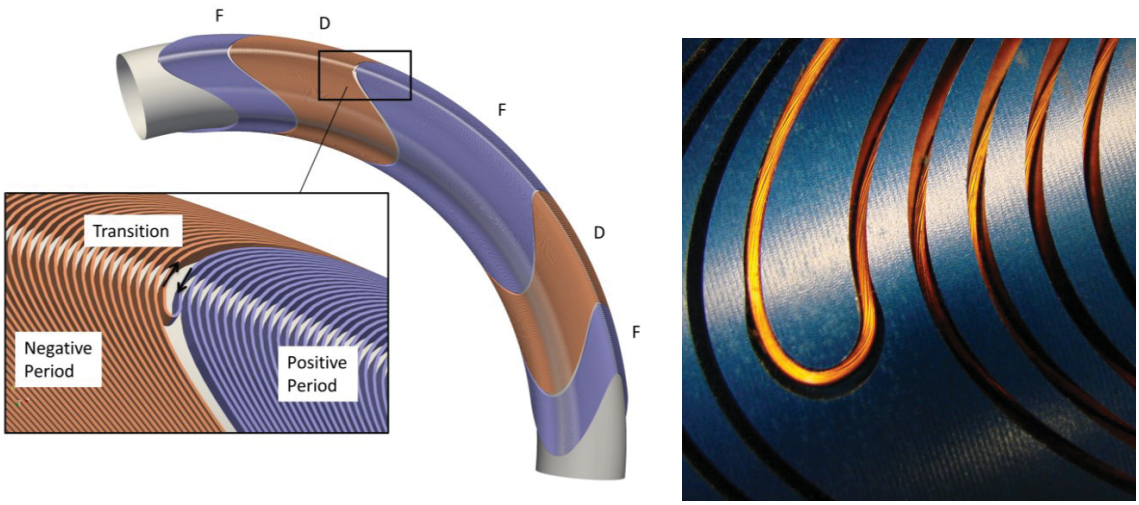

Figure 4.46: *A quadrupole layer with its five sections (left) depicting an earlier version of 90 degrees bend. A photo of a wire being placed into the transition area is shown on the right. Source: Courtesy of LBNL.*

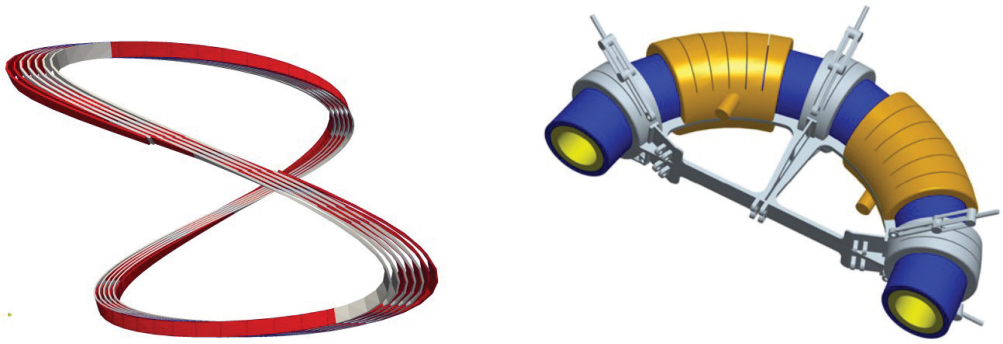

Figure 4.47: *Left: The simulated winding path of the quadrupole with added sextupole component (red) and without (white). Right: A CAD model of the AGCCT. The dipole (blue) and quadrupole (yellow) layers build a bent torus of nearly 180°. The mechanical structure in grey is designed to take stresses from straightening forces when the magnet is powered. The copper clamps shall provide heat conductance and connect to the cryocoolers (not depicted). Source: Courtesy of LBNL.*

The magnet has a curvature radius of 0.9 m and bends the beam by approx. 135 degrees. Since the canted windings extend further than the magnetic length, the physical length of the magnet is actually longer adding almost 20 degrees to both ends of the magnet. Hence the torus is bent by close to 175 degrees. The inner bore of the magnet is 210 mm. With 20 mm for the cryostat radially, the clear bore at room temperature for the beam is then 170 mm in diameter.

Table 4.6: *The five angular sections of the AGCCT and their designed fields.*

Type	Angle (deg)	R (m)	B_0 (T)	B_1 (T/m)	B_2 (T/m^2)	B_3 (T/m^3)
Bend	22.71	0.900	2.289	21.28	0	0
Bend	26.58	0.900	2.289	-21.28	-17.60	0
Bend	36.42	0.900	2.289	21.28	0	0
Bend	26.58	0.900	2.289	-21.28	-17.60	0
Bend	22.71	0.900	2.289	21.28	0	0

Each layer of the metal mandrel is divided into pieces of approx. 5 degrees of bending, which are called laminations, since they interrupt any electric currents along the magnet length. Hence these laminations were anodized, giving them the blue color visible in the pictures.

Table 4.7: *AGGCT design specification [32]*

AGCCT design specifications	
Bend radius	900 mm
Designed magnetic bend angle	135°
Physical bend angle	~170° (mandrel)
Number of layers	4
Inner bore diameter	210 mm
Outer diameter	~ 350 mm
Conductor	NbTi, rectangular wire 1.7 mm x 1.7 mm, formvar insulated
Peak field at conductor	5.2 T (at quadrupole layer)
Cooling	4 double-stage cryocoolers
Operating current (max.)	~ 800 A
Weight	~ 1,000 kg (cold mass)

A prototype of the *AGCCT* is currently being built at LBNL. It is downsized to 90 degrees of bending but with the full-size bore of 210 mm. It will be tested in a liquid helium Dewar vessel in Berkeley. A few pictures below show some of the hardware. This is the first *AGCCT* ever being built and expected to be tested in liquid helium in early 2019.

Figure 4.48 Assembly of the prototype magnet at LBNL: *(A): The individual laminations of the layers are assembled by shrinking one piece with liquid nitrogen cooling, so that a "dovetail" tip can slide into the other lamination [32]. When both pieces assume the same temperature again, the laminations are locked together tightly. (B) and (C): The first layer is ready with the conductor wound into the machined channels. (D): After winding of the first layer is completed, the second layer is slid over. (E): The combined two layers are then mounted on the winding table to finish the second layer of the magnet. Photos: Courtesy of LBNL.*

4.7 Conception of the *fixed-field magnet*

This section shall briefly reflect the conception of what turned into the *fixed-field magnet* invention which is currently in the process of being patented. It is important to emphasize that the development of this new concept was a true group effort between Weishi Wan (ShanghaiTech University), Lucas Brouwer (LBNL) and myself. The majority of beam tuning with the theoretical model to find new optics solutions was driven by Weishi, the design and generation of magnet models by Lucas and the beam tracking and analysis of the optical properties by me. Further detailing of the magnet design (possibly leading to additional patents on the hardware side) also included contributions from Soren Prestemon (LBNL). The magnet details are, however, only discussed to the necessary extent for the present text. Furthermore, it is important to note that the found solution looks in principle similar to the Mainz Microtron (MAMI) which was developed and built at the Johannes Guttenberg University in Mainz, Germany [132], to accelerate electrons up to 1.6 GeV [133], see discussion in section 6.7.

4.7.1 Goal

The goal of the new concept is to tackle some of the most bothersome problems of superconducting magnets with respect to an application to proton therapy gantries which became more and more apparent during the development of the *AGCCT* magnet. The issues are high cost, limited ability to quickly change the field and the involved cooling. The biggest driver for all three problems is the non-trivial task to apply fast ramp rates to a superconducting magnet. High magnet ramping requirements, mandatory for the timely application of multiple proton beam energies during patient treatment, cause high cost and complexity of the magnet, especially in the conductor and the cooling system. As discussed in the previous chapter, all means to increase the ramping speed raise the costs, some very significantly:

- Advanced conductor wires with small filament size and short twist pitch or even more advanced (expensive) conductors with higher critical temperatures
- Lamination of structures and other means to limit Eddy currents
- Increased cooling power / number of cryocoolers
- Design for higher margins

A hypothetical magnetic system, however, able to transport all clinical energies without the need to change the magnetic field, i.e. without any ramping of the superconducting magnet, would break free of all the compromises and troubles listed above. This goal was the main motivation during the conception of the *fixed-field magnet*.

4.7.2 From the idea to a proof of concept

The original idea is somewhat derived from the concept of the alpha magnet which bends a charged particle beam through a loop of approximately 270 degrees via a linear field gradient along the bend radius and indeed achieves achromatic energy acceptance over a very large range [134]. A simple scaling to protons of the requested energy, even though considered previously, seems not practical due to problems with the size, necessary field strengths, as well as fringe fields and beam focusing.

Studying the trajectories of various beam energies through an alpha magnet (depicted in different colors in **Figure 4.49**) one observes that at a symmetry plane in the loop all beams are parallel. Hence a hypothetical half-alpha-magnet would bend all energies by the same angle, although with different bend radii. Beams of various energies exiting such a magnet would be parallel to each other but offset in the bending plane as shown in the figure.

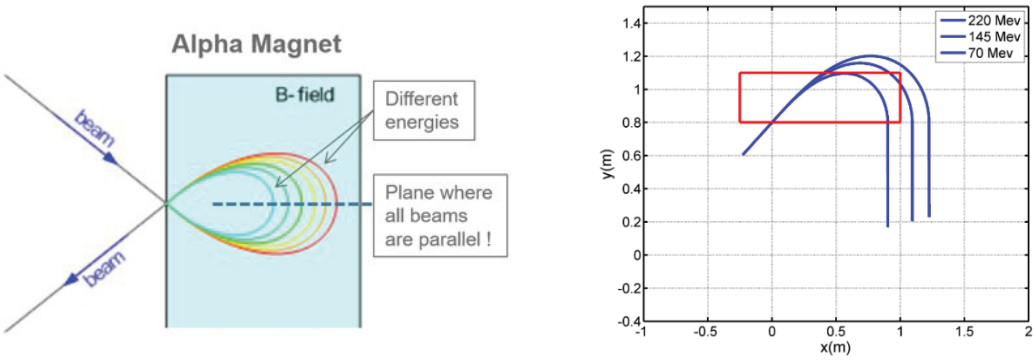

Figure 4.49: *Left: Schematic of rays through the field of an alpha magnet. Right: A hypothetical magnetic field resulting in a "half-alpha" behaviour where the beams exit on parallel trajectories.*

When combining two such magnets a system that indeed is able to transport a very large range of energies is obtained, as the second identical magnet reverses the "splitting" of beam energies by the first magnet.

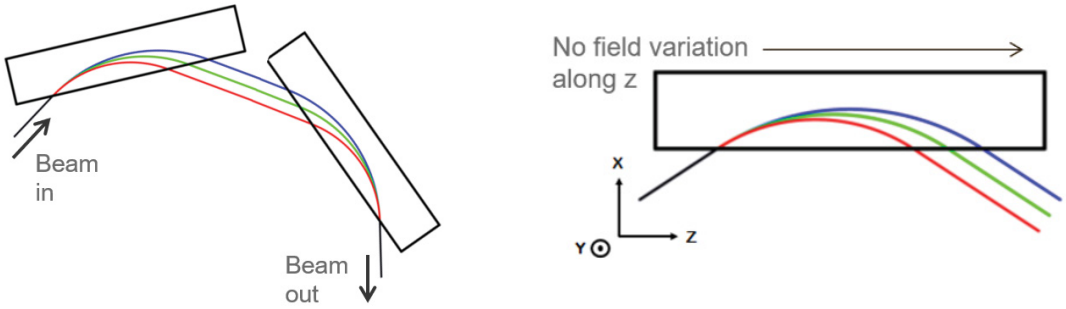

Figure 4.50: *Left: Rays of different energies passing through the proposed achromatic system. After being split up the second magnets unites all beams to an identical trajectory. Right: The important property of the 2D magnet is to not have any field variation along its length z.*

It can theoretically be proven that a rectangular magnet in a *x*-*z* plane with a B_y-field variation only along the in-depth axis *x*, always returns particles under the same angle as the angle of incidence [135].

From the Lorentz force follows:

$$\vec{F} = \frac{d\vec{p}}{dt} = q\vec{v} \times \vec{B} \tag{4.6}$$

$$\vec{p} = \gamma m_0 \vec{v} \tag{4.7}$$

Because there is no acceleration or deceleration, the velocity $|\vec{v}| = v$ and the relativistic factor γ are constants. m_0 is the proton's rest mass. The equations of motion become:

$$\gamma m \frac{d\vec{v}}{dt} = q\vec{v} \times \vec{B} \tag{4.8}$$

The vector $\vec{v}$ of a particle moving in the mid-plane (*x*-*z*) of the magnet, can be written as

$$\vec{v} = v_x\hat{x} + v_z\hat{z}\,, \tag{4.9}$$

while the B-field is perpendicular to this plane and only dependent on *x* (independent of *z*):

$$\vec{B} = B_y(x)\,\hat{y} \tag{4.10}$$

The equations of motion in components are then:

$$\gamma m_0 \frac{dv_x}{dt} = -qv_zB_y(x) \tag{4.11}$$

$$\gamma m_0 \frac{dv_z}{dt} = qv_xB_y(x) \tag{4.12}$$

Integrating the second equation yields with $v_x = dx/dt$:

$$\gamma m_0\, dv_z = q\, dx\, B_y(x) \tag{4.13}$$

$$v_z(x) = v_z(-\infty) + \frac{q}{\gamma m_0}\int_{-\infty}^{x} B_y(x')\,dx' \tag{4.14}$$

As a result, v_z is a function of *x* only and independent of *z*. Hence $v_x = \sqrt{v^2 - v_z^2}$ is also a function of *x* only. This means that $v_{z,1} = v_{z,2}$ and $v_{x,1} = -v_{x,2}$ where *1* and *2* denote the entrance and exit to the magnet, respectively. This is true because all particles entering the rectangular magnet have the same *x*-coordinate as when they exit the magnet. Hence, if particles of different energies enter such a magnet under the same angle of incidence, they will be parallel when exiting the magnet, too. From this, two more conclusions can be deduced. First, the focal length in the horizontal plane (*x*-*z*) is infinite and secondly the slope of the dispersion at the magnet exit is zero.

What also follows is the very important prerequisite of having a rectangular magnet with a one-dimensional field profile in the mid-plane.

From the theoretical result it seems that the specific shape of the $B_y(x)$ field profile can be arbitrary. While this is true for the mere condition of having equal entrance and exit angles, beam focusing must be considered as well. Especially the relatively flat entrance angle to the magnet results in strong edge-focusing effects.

4.7.3 Approach to find beam optics and magnet solution

In order to find a magnet design that can produce a field fulfilling the beam optics requirements multiple approaches were undertaken, which are shortly summarized here.

First, it is important to note that the one-dimensional field requirement is actually very beneficial for the magnet design. It allows straight magnets (in contrary to the curved *AGCCT*) and hereby simple straight racetrack coils can be used. The biggest question at this stage is how simple (and cost-effective) a magnet can be and still work for the optics.

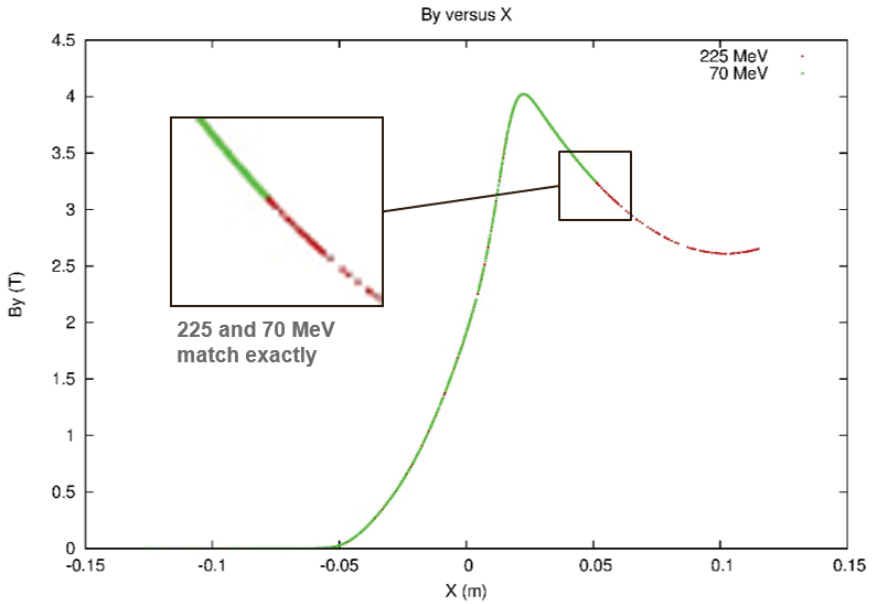

Figure 4.51: Theoretical field profiles fulfilling the focusing conditions. The field profiles are a result of the beam optics calculation and do not reflect the field of an actual magnet. Calculations show, however, that an identical field can serve as a solution for the lowest (70 MeV) and highest beam energy (225 MeV).

Starting with the profile from **Figure 4.51**, various coil combinations were tried: single, and double racetracks as well as tilted coils. However, the requested profile has such "unnatural" shape that no good agreement could be obtained. Instead, the process was then reversed. In the previous step it was understood that the combination of two racetrack coils with a boosting iron yoke gets relatively close to the desired fields. The iron is used for returning the majority of the flux on the side opposite to beam entry, greatly reducing the problematic reverse field at the entry location. On the other hand, the iron poses a difficulty for calculating the beam optics as an analytical approach with Biot-Savart is no longer feasible. For iron magnets with inhomogeneous fields commonly magnet field maps are used and optics checked with slow FEA tracking methods – an involved approach for an iterative search for a beam optics solution. Here an intermediate step was made. An analytical function was fitted to the field profile produced by the magnet model. The function, a set of three Gaussians, was then used in a code in *COSY INFINITY*[16] to automatically search for beam optical solutions. Since the analytical function can be varied with its multiple parameters, it can be easily tuned until the optics solution is found. The clue here is that the altered function – after being successfully adapted to the optics requirements – is still similar in shape to the original profile. As a consequence, it is much easier to generate a matching field profile from a magnet design perspective. Indeed, as demonstrated in **Figure 4.52**, this approach led to a technically feasible solution.

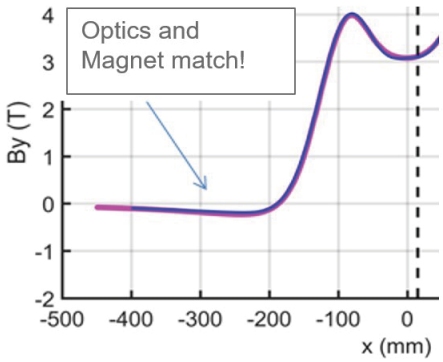

Figure 4.52 First success: *Magnetic profiles as requested from the beam optics (blue) and delivered by the magnet model (pink) match for a first intermediate result.*

[16] More details about the software *COSY INFINITY* are given in chapter [5].

4.7.4 Design of the fixed-field magnet

The system consists of the two identical magnets and two normal conducting quadrupole-doublets or -triplets, one on each side. These quadrupoles are necessary because of the very short focal length of the double bend achromat, which is mostly due to the shallow angle of incidence of the proton beam which causes strong edge focusing. The first working design solution that was studied in depth (see chapter 6) with good beam shape results, uses a quadrupole-triplet on each side. More recently, another design, which potentially works with a quadrupole-doublet on each side, was found and reported [136]. The general feasibility of this doublet-design is demonstrated by the imaging of the field and axial rays, see **Figure 4.54**. In the following chapters, however, all results refer to the more mature quadrupole-triplet solution.

A steeper entrance angle, and thereby an increased bending angle, reduces the edge focusing. Hence a total bend angle of 155 degrees is chosen, i.e. 77.5 degrees per bend. The quadrupoles on the downstream side focus the beam to the isocenter approximately 1.8 m away.

The magnetic field in the superconducting bend is highly inhomogeneous containing various higher order multipole fields. It is necessary to achieve as much symmetry in the system as possible in order to cancel out non-linear aberrations. Therefore, an upstream drift of equal length to its downstream counterpart is chosen and the quadrupoles assure point-to-parallel (axial ray), and parallel-to-point (field ray) imaging for anti-symmetric optics from the start to the mid-point, i.e. the symmetry point of the system. For the whole system this means point-to-point imaging in the conventional sense.

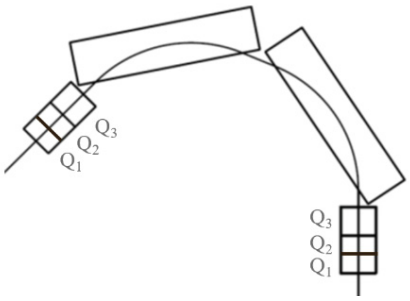

Figure 4.53: *Design of the achromatic bend: a total of six resistive quadrupoles arranged in two identical triplets and the two rectangular superconducting magnets.*

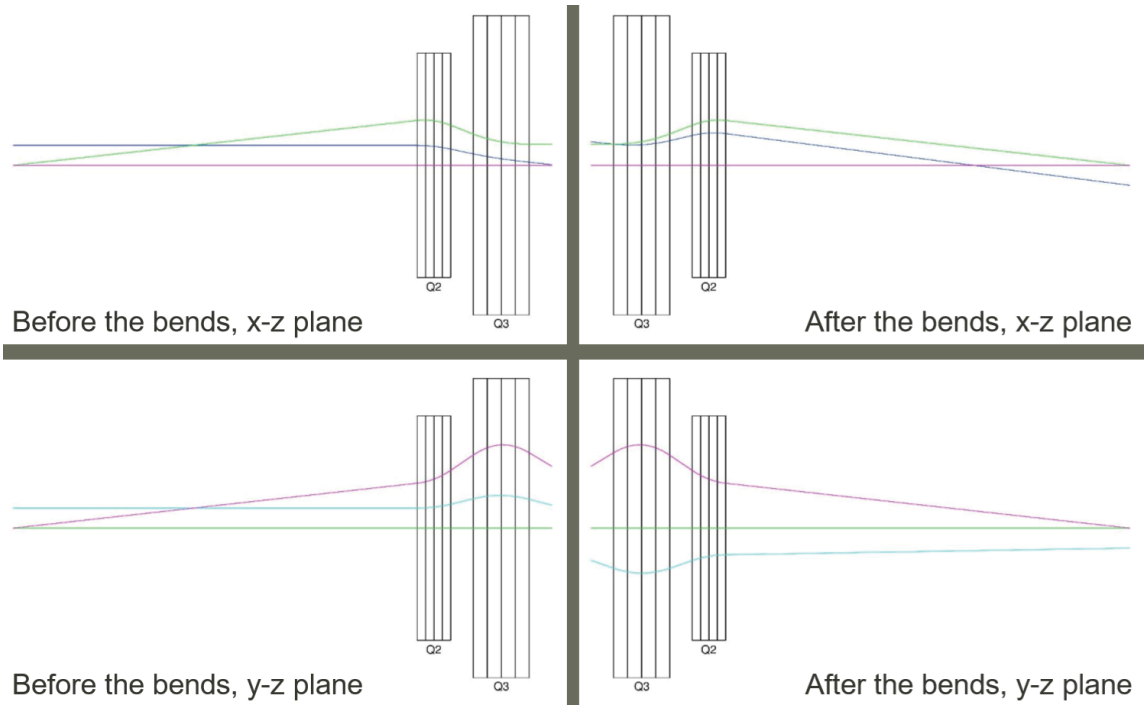

Figure 4.54: *The field and axial rays of both bending planes before and after the two bending magnets are shown for the quadrupole-doublet solution. In total the achromatic system achieves 1:1 imaging and nearly perfect symmetry.*

The normal conducting quadrupoles are further optimized for minimal space consumption as they add directly to the gantry radius (compare **Figure 4.58**). Their properties are listed in the table below[17].

Table 4.8: *Dimensions and fields of the quadrupole-triplet.*

Triplet	Length	Full Aperture	Gradient	Pole tip field
NC Quad 1	160 mm	38 mm	43 T/m	0.82 T
NC Quad 2	190 mm	38 mm	43 T/m	0.82 T
NC Quad 3	110 mm	38 mm	43 T/m	0.82 T

As mentioned, the identical superconducting bends each consist of two pairs of racetrack coils embedded in a "C-shaped" iron yoke. The "C" opening is necessary to let the beam enter and exit the magnet sideways.

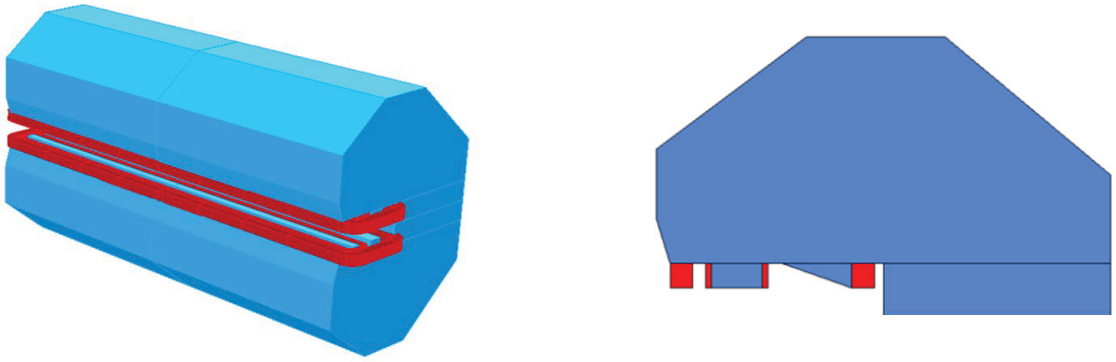

Figure 4.55: *Left: the two coil pairs are shown in the Opera model where only one quart of the C-shaped iron yoke is visualized. Right: Cross section of the upper half of the magnet. The blue area depicts the iron and the red rectangles are the cross sections of the winding packs of the two racetrack coils.*

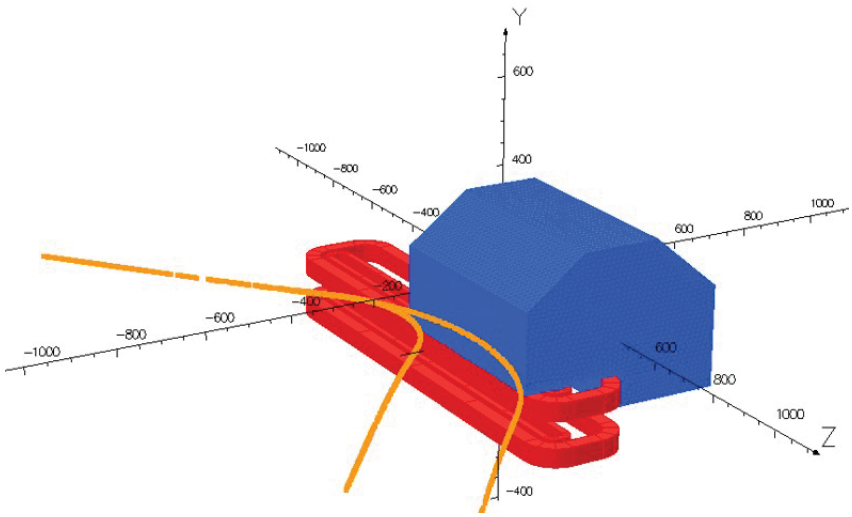

Figure 4.56: *The trajectories of two protons with 70 and 225 MeV through the fixed-field magnet are displayed overlapping with the magnet model. They are obtained through straightforward FEA tracking directly in Opera.*

[17] The numbers specified here are indicative, representing an intermediate first solution for which the desired magnetic and optical properties could successfully be achieved. The design is still being optimized and adapted before a first prototype will be manufactured.

The outer racetrack coil generates a first field with a plateau region (blue line in **Figure 4.57**) while the inner racetrack coil adds another bump to it (yellow line). The complete field is then significantly boosted by the iron, a gain of approximately 2 T. Another very important effect of the iron is the reduction of return flux outside the magnet (negative *x* region in the graph).

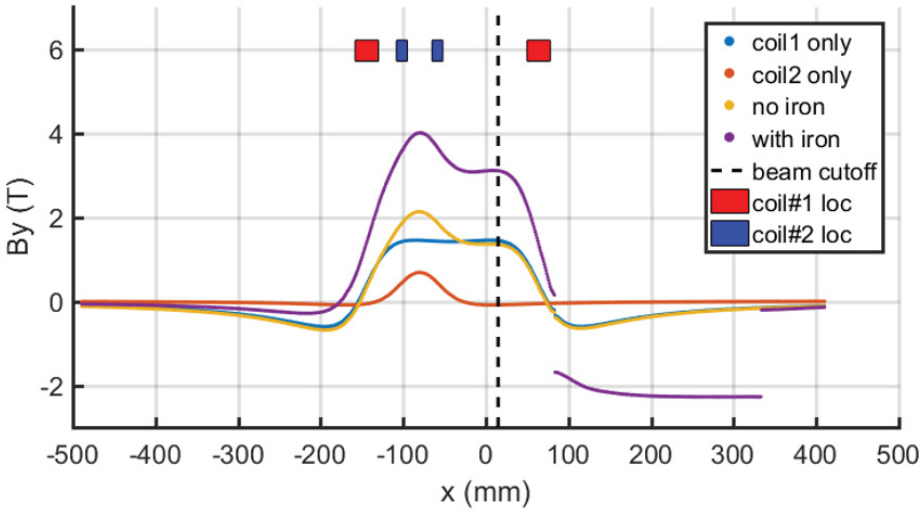

Figure 4.57: *The one-dimensional field profiles of the outer coil (#1) producing the blue line, the inner coil (#2) producing the red line, as well as the sum of both fields shown in yellow. The purple profile is the final field when the iron is taken into account. Note that not only the field is raised but the negative field on the left side (around x = -200) is significantly reduced. The dashed vertical line indicates the farthest x-location the protons reach (coming from negative x values).*

As the beam enters the magnet from negative *x* (from the left in the graph) it first sees the negative field of the return flux which causes a bending in the opposite direction. As a result, the effective bend radius of the complete system becomes larger which also makes the gantry radius larger. A second, and even more problematic, effect is the same field extension at the exit, so in between the two magnets. For maximal compactness the gap between the magnets is minimized but field changes due to potential crosstalk would violate the condition of field independence of coordinate *z*. Furthermore, even an unchanged fringe field like of the purple case of **Figure 4.57** extending about 15 cm is problematic since the gap between the magnets is less than 30 cm for the highest energies. Consequently, protons of different energies would see different integrated fields along their trajectories, leading to different beam angles and ultimately to non-parallel beams at the isocenter. Hence field clamps at entrance and exit of the magnets are necessary to suppress the field extension to less than half of the gap distance for the highest energy which is 200 – 250 mm.

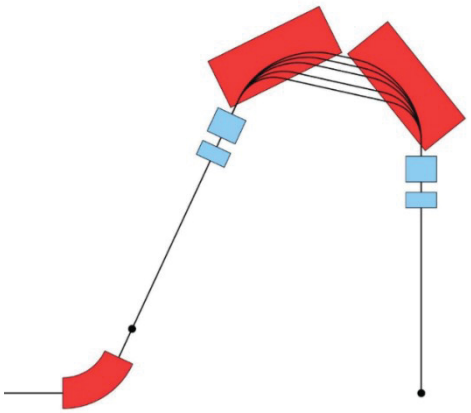

Figure 4.58: *The layout of the proposed gantry (here with the quadrupole-doublet). A detailed solution is shown in chapter 6.*

An additional advantage of returning a large portion of the field through the iron is that the stray field of the magnets is generally much reduced. This plays an important role for the reasons listed in section 4.4.6. Here a plot of the stray field of the combined magnets up to the isocenter shows that the magnetic field in the patient area can easily be reduced below the 5 mT limit.

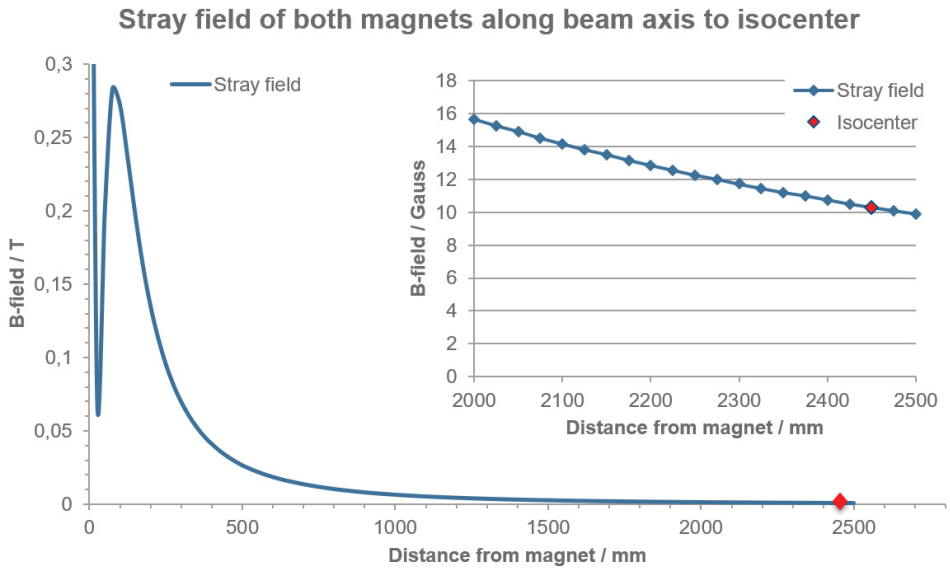

Figure 4.59: *Stray field of both superconducting magnets plotted along the direct path to the isocenter which is 2.45 m downstream of the second magnet and marked with a red diamond. One can also see that the B-field rapidly drops to values below 0.3 T thus should not be an issue for the quadrupole or scanning magnets.*

Following some detailed specifications of the studied magnet are listed. They reflect the design work from L. Brouwer[18].

Table 4.9: *Specifications of the fixed field magnet.*

Design specifications of the fixed field magnet	
Length	1.1 m
Width	0.45 m
Height	0.5 m
Weight (coils + iron)	ca. 1.9 t
Cold aperture	60 mm
Warm aperture	25 mm
Peak field at conductor	4.7 T
Peak field in aperture	4.0 T

The efficiency of the design can be seen in the small ratio of the peak fields at conductor and aperture, enhancing the margin for the magnet. While the peak field is 4 T, the average field seen by the beam during bending is about 3 T leading to an effective bend radius of about 0.75 m for the highest energy.

Compared to superconducting magnets requiring ramping, this design allows the use of a very simple and cheap conductor. With an estimated wire length of about 4.2 km per magnet, the total costs for the conductor are less than $ 10k, making a typical major cost driver almost insignificant. The chosen operating current below 500 A allows the use of less complicated current leads and relaxes the cooling requirements (compare section 4.4.2). The specifications

[18] The numbers specified here are indicative, representing an intermediate first solution for which the desired magnetic and optical properties could successfully be achieved. The design is still being optimized and adapted before a first prototype will be manufactured.

also show a disadvantage of the iron: its weight. Both magnets together with a cryostat are estimated to weigh somewhere between 5 and 6 tons, which is more than a factor 2 heavier than the *AGCCT* but also a factor 2 lighter than a normal conducting dipole of 155 degrees.

Table 4.10: *Proposed conductor and operating parameters.*

Conductor and operating parameters	
Conductor	NbTi wire
Cu:SC ratio	4:1
Wire dimensions	1.65 mm × 1.0 mm
Operating temperature	4.2 K
Operating current	400-500 A
Static case margin	~ 30 %
Coil packing fraction	80 %
Wire cost	1-2 $/m

5 TOOLS AND METHODS

5.1 Tools

For the calculation of proton transport through the discussed magnet systems, the software COSY INFINITY version 9.1 (further referred to as "*COSY*"), developed by Martin Berz and Kyoko Makino from Michigan State University, was employed [137]. The post processing (Gauss-fits, histograms, 2D and 3D plots) of the tracking results of larger numbers of protons was done with MatLab®. The linear beam optics calculations in the conventional beamline, mainly the front-end of the gantry, were performed with the *Transport* code, developed by Urs Rohrer from the Paul Scherrer Institute.

5.1.1 Linear beam optics with Transport

The *PSI Graphic Transport Framework* developed by Rohrer [138] is based on a CERN-SLAC-FERMILAB version by K.L. Brown et al [139]. The PSI code is now widely used for simple beam optics calculations of first order (linear optics). It is based on the transport matrices of the phase space of the beam. For proton therapy beamlines with conventional quadrupoles and dipole magnets the accuracy of this linear optics code is sufficient. Its advantage is the high speed computation of the fitting algorithm. A disadvantage is that it does not take any scattering or scraping effects into account, e.g. beam losses at apertures or collimators. The related code *Turtle* [140] does offer this feature. However, since beam losses in a single-pass system are not critical, the calculations with *Transport* provide a convenient and precise enough method of calculation.

Naturally, at some locations, e.g. at the collimators, beam losses are more important for predictions of the transmission and for estimates of the amount of produced neutron radiation. For these questions Monte Carlo methods are typically employed, wherein the code *TOPAS* [141], an extension to *Geant4*, offers the option to simulate a beamline by importing magnetic fields or using standard quadrupole and dipole elements. It has been shown that – if setup properly – a complete beamline can be simulated with *TOPAS* with very exact results on spot shapes and transmission [142]. This method, however, is not useful for the design or optimization of beamlines, as the imported magnetic fields cannot be easily adjusted.

For the beam optics design and optimization with the more complicated magnetic fields of these studies, that deviate from standard dipole or quadrupole fields i.e. have non-linear components of significance, other tools must be used. For both studied magnet systems, the AG-CCT as well as the Fixed-Field Achromat the *COSY* code was employed as described in the following chapters.

Excepted is the portion of the beamline upstream of the superconducting bend. Here *Transport* calculations through the standard quadrupole and dipole elements are appropriate. This

beamline, the *front-end,* transports the beam from the cyclotron to the degrader which is mounted on the rotating gantry as described in section 3.5.2.

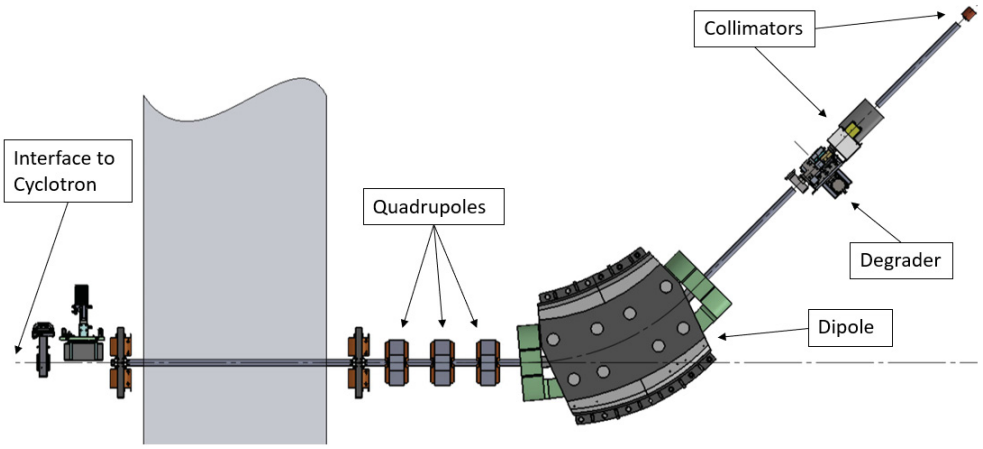

Figure 5.60: *Front-end of the beamline. The beam is fed in from the left as it exits the cyclotron (not depicted). The three quadrupoles downstream the shielding wall focus the beam through the dipole onto the degrader. They are also the first elements that rotate with the gantry (horizontal dashed line is the rotational axis). The quadrupole triplet provides enough flexibility to change the beam envelopes as required by the gantry rotation angle and the dispersion effect. The collimators downstream of the degrader define a "new" beam for the subsequent optics.*

The front-end is a very simple piece of beamline, basically consisting of three quadrupoles and one dipole. Its only unique characteristic is that is does not fulfill achromatic conditions. Hence effects of the dispersion must be taken into account. Furthermore, it is desirable to minimize the apertures of the magnets for cost reasons.

5.1.2 Mathematical description of the linear beam transport

A beamline lattice is a sequence of beamline elements. These elements can be either magnets (dipoles and quadrupoles), drifts or other beam defining elements, e.g. initial phase ellipse definition. In this lattice also other points of interest can be introduced and labeled, such as collimator positions or the isocenter, but in *Transport* they do not affect the beam optics (apart from representing possible fit conditions).

Each of the beamline elements are represented by a transport matrix (*R*-matrix) which by simple multiplication gives a resulting output phase space vector depending on an input phase space vector. The input vector can be multiplied with all transport matrices of all beamline elements in sequential order. The result will be the phase space vector describing the proton beam at the final position. In the example of the studied front-end the final position is at the degrader.

Following is an exemplary evolution of a two-dimensional vector through three beamline elements:

$$\begin{pmatrix} x \\ a \end{pmatrix}_{final} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}_3 \cdot \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}_2 \cdot \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}_1 \cdot \begin{pmatrix} x \\ a \end{pmatrix}_{input} \tag{5.1}$$

Each R-matrix describes the relative change of x and a as well as the coupling of x and a by the matrix elements R_{12} and R_{21}. The matrix indices *1,2,3* denote the various elements (magnets or

drifts) indicating the order of sequence as seen by the beam. One can also summarize all *n* *R*-matrices of the complete beamline in one transport matrix R_T:

$$R_T = R_n \cdot \ldots \cdot R_3 \cdot R_2 \cdot R_1 \tag{5.2}$$

And the transport of a particle beam can be calculated and written as:

$$\vec{x}_{final} = R_T \cdot \vec{x}_{input} \tag{5.3}$$

The *Transport* code takes user defined beam input parameters and calculates the evolution of the corresponding phase space vector through the lattice. Data for the input emittance (size and divergence) were taken from empirical values as measured at commissioned cyclotrons in use [143]. The momentum spread of the beam when exiting the cyclotron is not exactly known but a value of 0.03 % *dp*/*p* is used as an educated assumption according to information from PSI (who use the same Varian cyclotron [144]) and which complies with the original cyclotron design parameters [145]. However, this value may deviate in practice from one cyclotron to another (e.g. due to multi-turn extraction). Hence, ideally, the beam optics are flexible enough to work for larger momentum spreads as well.

As mentioned above, the front-end beamline does not fulfill achromatic conditions so the optics have to take the non-vanishing dispersion function into account. This is achieved by preparing an asymmetric beam before the dipole so that the dispersive broadening in the bending plane matches the larger beam size of the vertical plane, resulting in a symmetric beam at a defined location downstream of the bend[19].

5.1.3 Focusing and imaging

An essential task of the optics is to keep the beam confined and focus it to the isocenter. A double beam waist, as outlined in chapter 2, defines a focus point. When a given beamline (or part of a beamline) transports the beam from one focus point to another focus point, one speaks of imaging optics. In theory, the resultant spot size is then only dependent on the initial spot size but independent of the initial divergence angle. The magnification factor, which in the case of 1:1 imaging is one, determines the final-size to initial-size dependence. In general, it is advantageous to keep the magnification as small as possible or, if possible, even achieve demagnification. Typically, the final beam size is defined by clinical requirements, thus, the smaller the magnification factor, the larger may be the initial beam size, which usually improves transmission. In the cases studied, an imaging of 1:1, i.e. without magnification, is desired because of the inherent symmetry that comes with it.

5.1.4 COSY INFINITY

The code *COSY INFINITY* using the customized language COSYScript serves to combine methods of numerical integration with the calculation of transfer maps based on Taylor expansions that provide insight into the action of the field multipoles on the particle phase space coordinates. This is achieved at high computational speed by the use of differential algebra techniques (as described in section 2.5). Thus, *COSY* is not limited to a specific order of calculation, i.e. order of magnetic multipole fields. Compared with ray tracing, which is purely

[19] A patent application for this concept is submitted. Inventor: A. Huggins.

based on numerical integration, this method is much faster and the computation of transfer maps allows for very efficient tracking of large number of particles. Additionally, the calculation in terms of phase space coordinates is very useful (almost mandatory) for the goal-oriented optimization of optical systems. Details of the algorithms and language can be found in [146, 147] but some important elements are highlighted in the following chapters describing the SCOFF and the "realistic" models.

5.1.5 Transfer maps in COSY

The transfer map is a set of functions, as many as there are phase space particle coordinates: in this case five. As described in section 2.4 each particle coordinate is written as a function of all input coordinates:

$$M = \begin{pmatrix} f_x(x,a,y,b,\delta) \\ f_a(x,a,y,b,\delta) \\ f_y(x,a,y,b,\delta) \\ f_b(x,a,y,b,\delta) \\ f_\delta(x,a,y,b,\delta) \end{pmatrix} \tag{5.4}$$

The functions are – depending on the order n of calculation – long polynomials of the five independent variables. The map in *COSY* is a list of the coefficients for the monomials of all contributing combinations of the variables. An example is shown in **Figure 5.61**.

I	COEFFICIENT	ORDER	EXPONENTS
1	0.3536967910906734	1	1 0 0 0 0 0
2	0.9958255941745762	1	0 1 0 0 0 0
3	0.3489516745114685E-01	1	0 0 1 0 0 0
4	0.7198927903848527E-01	1	0 0 0 1 0 0
5	0.1766884780123016E-01	1	0 0 0 0 0 1
6	-8.528461827426103	2	2 0 0 0 0 0
7	6.043626703650448	2	1 1 0 0 0 0
8	2.351715402523489	2	0 2 0 0 0 0
9	-2.071714138766994	2	1 0 1 0 0 0
10	0.5333424113180074E-01	2	0 1 1 0 0 0
11	1.960263347880448	2	0 0 2 0 0 0
12	-5.233606223266324	2	1 0 0 1 0 0
13	0.8688583928330124E-01	2	0 1 0 1 0 0
14	5.037750967461124	2	0 0 1 1 0 0
15	1.719105358806541	2	1 0 0 0 0 1
16	-1.553301622744706	2	0 1 0 0 0 1
17	0.5330434471112030	2	0 0 1 0 0 1
18	1.206642550854440	2	0 0 0 2 0 0
19	1.137416595332998	2	0 0 0 1 0 1
20	0.8381142830840617E-01	2	0 0 0 0 0 2
21	42.79245513767219	3	3 0 0 0 0 0
22	7.369570962002967	3	2 1 0 0 0 0
23	-116.8336876222541	3	1 2 0 0 0 0
24	53.22380744305503	3	0 3 0 0 0 0
25	-4.919228383509308	3	2 0 1 0 0 0

Figure 5.61: *First 25 coefficients of an exemplary transfer map. The exponents refer to the variables x, a, y, b and at the sixth position δ, which are the relevant phase space coordinates. The longitudinal displacement l is not of interest; thus the fifth variable is always 1. Each line represents one monomial of the polynomial function.*

The polynomial function yielding the respective final phase space coordinate is the sum of all its N monomials. So, for the example of the final x-coordinate f_x:

$$f_x = \sum_{k=1}^{N} c_k \; x_i^{r(k)} a_i^{s(k)} y_i^{t(k)} b_i^{u(k)} \delta_i^{v(k)} \tag{5.5}$$

where *k* is the line number (or number of the monomial), *c* is the coefficient and *r(k), s(k), t(k), u(k), v(k)* are the exponents of the five initial variables (denoted with the index *i*) of the respective monomial, whereby the sum of all exponents gives the order of the monomial:

$$r(k) + s(k) + t(k) + u(k) + v(k) = n(k). \tag{5.6}$$

The other four final coordinates are calculated in analogous manner.

In *COSY* the ordering of the monomials is maintained following a general rule which facilitates faster computation by applying the rules of differential algebra [148].

Producing the transfer map

At any point along the trajectory of a particle the *B*-field is expanded in a Taylor power series where the polynomial terms represent the various multipole field orders. The Taylor series in general reads:

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}\, x^n, \tag{5.7}$$

where *x* is the relative coordinate (like the particle coordinate). So, x is the distance to the position x_0 of the reference particle, which is also the location of expansion. With increasing *n,* not only the order of the variable is increased but also the order of the n^{th} derivate of f is needed. This is why it is advantageous, for more precise calculations, to use functions that are infinitely derivable, as will be shown below, in the calculation of the fixed field magnet. If the field is written in form of the Taylor expansion, each term represents a multipole field component and gives the effect on the particle trajectory according to the equations of motion.

$$\frac{da}{ds} = -\frac{q}{p_0}\left(B_{y0} + \frac{dB_y}{dx}x + \frac{1}{2}\frac{d^2B_y}{dx^2}x^2 + \frac{1}{6}\frac{d^3B_y}{dx^3}x^3 + \cdots\right) \tag{5.8}$$

Thus, from the field a kick, *da/ds*, results which in its differentiated form requires integration to calculate the change in *a* as well as in *x* according to equations (2.32 – 2.35):

$$\frac{dx}{ds} = \frac{a}{\sqrt{1 - a^2 - b^2}} \tag{5.9}$$

COSY uses an eighth order Runge-Kutta integrator with variable step size $\bar{s}$. The step size of the integration is iteratively reduced until the solution converges. If the solution does not converge, a warning message is displayed as soon as a configurable minimal step size is reached. As a result, a transfer map over the step size $\bar{s}$ is obtained. By choosing a maximal order of the Taylor expansion, the order of this transfer map is directly defined. It is important to note, that the high order derivatives of the field are not from numerical differentiation but from smooth analytical models with known Taylor expansions (stored in a library). That way the high order derivatives are all exact, making the coefficients in the transfer map accurate up to machine precision in principle.

In order to describe the beam through a complete magnetic field, not just for the short distance of $\bar{s}$, its transportation is calculated in sequential steps, whereby a subsequent step is calculated from the endpoint of the previous and the end-coordinates of the reference particle are used as

new input for the subsequent step. At the same time the transfer maps for the many small steps are concatenated by plugging in the polynomial function for each coordinate, e.g. for x, of one step into the polynomials of the next step.

For the expansion of the plugged-in polynomials, the methods of the differential algebra are employed. As pointed out before any produced terms with higher orders than the prior chosen order of calculation are omitted. Furthermore, the ordering of the monomials in the vector is kept unchanged for computational simplicity.

According to this procedure the reference particle is tracked through the magnetic field stepwise, wherein the step size depends on the gradients of the field. At the same time phase space coordinates are used, which apply to a complete particle beam. Consequently, after concatenation of all maps, the final transfer map provides a direct relation of input phase space coordinates and resultant final phase space coordinates describing the properties of the particle beam when exiting the magnetic field as a function of its input properties.

Additionally, more parameters can be included in the equation. If the magnetic field is calculated from a parametrized model, the final beam properties also depend on these parameters, e.g. the strength, position or orientation of individual quadrupole components. This is particularly useful for sensitivity analyses, as it provides insight into how certain parameters affect the beam properties. Thus, the transfer map M relates the final coordinates not only to the input coordinates but to a parameter set ω influencing the magnetic field:

$$\vec{Z}_f = M\left(\vec{Z}_i, \omega\right). \tag{5.10}$$

5.2 Methods for the AGCCT

5.2.1 Methodology of the AGCCT design

The process of designing the *AGCCT* followed these main steps: The layout of the optics is drafted as a concept, i.e. the local achromat with five alternating quadrupoles[20]. From previous experience or by estimates of maximal fields etc. the bend angle as well as the aperture radius is chosen. These constitute the *initial assumptions*. From there a simplified model called *SCOFF* is developed which allows the fitting of the magnetic parameters to a first beam optics solution. It basically defines the dipole and quadrupole strengths and can be used for a first optimization based on idealized field profiles. This information is then used in the next step to actually design the magnet (a magnet engineering task). In case of the *AGCCT* this means the geometric definition of the conductor winding path and the corresponding electric currents to generate the requested fields. In more abstract words: a current density in space is defined. (Of course, many other aspects drive this engineering exercise as technical and economic feasibility of the magnet must be considered). With the information on the detailed magnet geometry, a *realistic model* in *COSY* is created which reflects the actual conductor winding. It can be used to calculate the optics including fringe fields and other field inhomogeneities that were previously idealized. Here, another optimization is performed leading to a number of iterations

[20] The original optics concept was developed by J.M. Schippers and A. Gerbershagen at PSI, and later adapted in the collaboration project with LBNL and Varian.

between the magnet design and the optics. At last, when the optimization is completed, tracking studies can be conducted to analyze details of the beam transport that characterize the beam optics performance of the magnet.

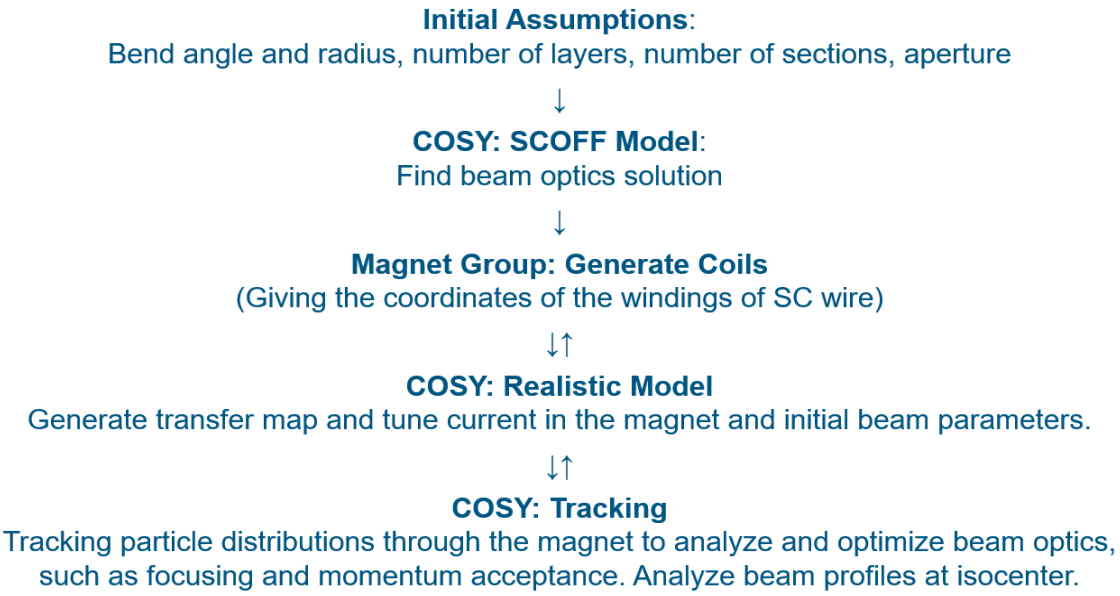

Figure 62: *Methodology of designing the AGCCT.*

5.2.2 SCOFF Method

This simplified model helps the search for optical solutions in a quick manner. In a *sharp cut-off fringe field* (SCOFF) approximation model the various magnet orders are considered uniform over a defined length and can be combined without restrictions. Such solutions do not consider technical limitations from the magnet design, nor the physical details of fringe fields and non-uniformities within the magnets.

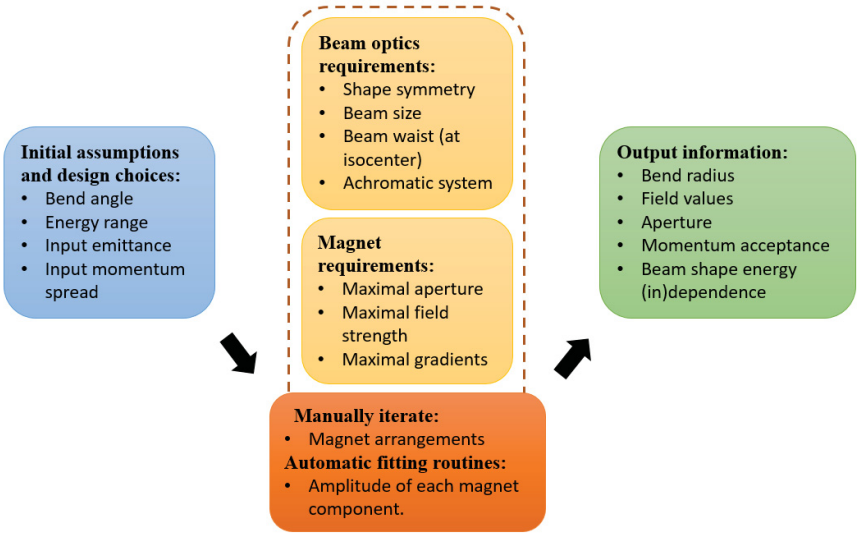

Figure 5.63: *Workflow of the SCOFF model step.*

In the SCOFF model in *COSY* one can define magnet arrangements (lattices) in a similar way as in *Transport*. The difference is that higher orders of fields, e.g. sextupoles or octupoles, can be included. For the design of the *AGCCT* the *COSY* procedure "MS" is used that allows the

combination of a bending magnet and multipoles up to 5th order [147]. The beam optics methods remain unaltered using the differential algebra methods described. As such, higher order effects even of linear elements such as quadrupoles are taken into account – an important distinction versus *Transport* calculations. However, all magnets in the system are idealized, thus some deviations to a real system are to be expected.

5.2.3 Realistic model in COSY

Based on the SCOFF model a magnet can be designed such that its field profiles match the SCOFF "prescription". Suitable programs are for example Opera FEA Simulation Software [149] and ANSYS Maxwell [150]. This step was carried out by Lucas Brouwer who also uses a self-developed code [33]. While the magnet is designed to achieve the fields to mimic the SCOFF model as close as possible, some deviations due to fringe fields etc. remain unavoidable.

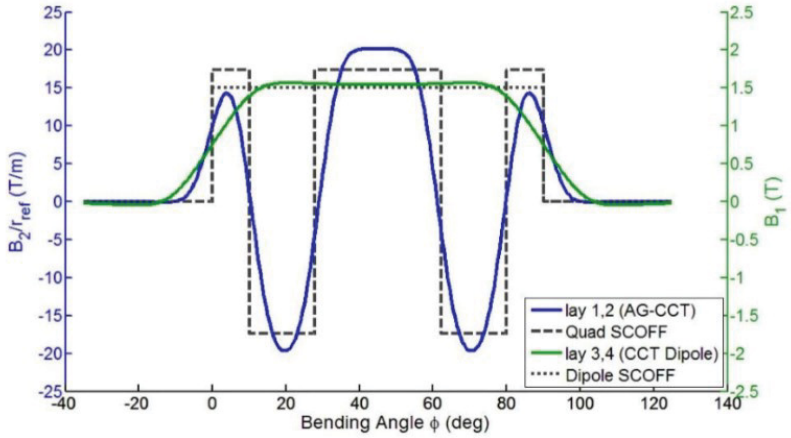

Figure 5.64: *Exemplary comparison of SCOFF (dashed) and real fields (solid lines). The dipole is shown in green, the quadrupoles in blue.*

The *AGCCT* magnet is a so-called "air-coil", meaning that the magnetic field is solely produced by the electric currents running through the conductor and no ferromagnetic materials are employed. Hence the *B*-field is defined by the geometric path of the conductor windings. As described in chapter 4.6 the *AGCCT* is built in multiple layers where every two layers combine to produce a dipole, quadrupole or sextupole. A list of coordinates representing the geometric information of the winding path then defines the actual current distribution in space.

With the realistic model of the *AGCCT* in *COSY* the information of the winding path coordinates allows the calculation of the *B*-field via the Biot-Savart law (only facilitated by the absence of any iron). In the code the coordinates of all points along the conductor winding path are read in for each layer. These points, 60-120 per winding turn, are connected as incremental straight pieces of the conductor (the $d\vec{l}$ in the Biot-Savart equation). Then, depending on a tunable current in the conductor, the resulting *B*-field at any point in space $\vec{r}$ can be calculated directly by integration and summation.

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi} \int_C \frac{I\, d\vec{l} \times \vec{r}'}{|\vec{r}'|^3} \tag{5.11}$$

A variation of the current *I* allows to adapt the magnetic field strengths for the individual layers of the magnet.

As explained in chapter 4.7 the *AGCCT* consists of one continuous dipole, five quadrupoles in five sequential sections (see also **Figure 5.64**) and two sextupoles which overlap with sections 2 and 4 of the quadrupole layer.

While the magnet design foresees that the (relatively weak) sextupoles will be incorporated in the two quadruple sections by slight manipulation of their winding paths (compare **Figure 4.47**), this was in fact not respected in the *COSY* model. There, the sextupoles are treated as separate layers to facilitate independent fine tuning of the currents of the quadrupole and sextupole. However, the length of the two sextupole sections are fitted to the quadrupole sections 2 and 4 and the winding radius is matched to the quadrupole as well. This is noteworthy because of the r^2 dependence of the field.

Apart from adjusting the current of the magnet, the partition of the alternating quadrupoles can also be changed in the model. The length of each quadrupole is defined by the number of winding turns. This makes another deviation from the real magnet design: The inversion of the quadrupole polarity between neighboring sections of the (physical) magnet is achieved by reversing the winding direction (see **Figure 4.46**). In the *COSY* model, however, the inversion is done by (unphysically) inverting the sign of the electric current (but keeping an uninterrupted continuous winding pattern). The position of the current inversion can theoretically be chosen arbitrarily, as it does not affect the geometry of the winding. This potentially provides for more degrees of freedom for the optimization of the realistic model. But there are two constraints regarding the point of inversion: For one, the physical winding reversal in the actual magnet requires some space due to the minimal bend radius of the wire. And since the distance between the wire grooves is largest at the poles, it makes them the natural choice for the reversal location (see **Figure 5.65**).

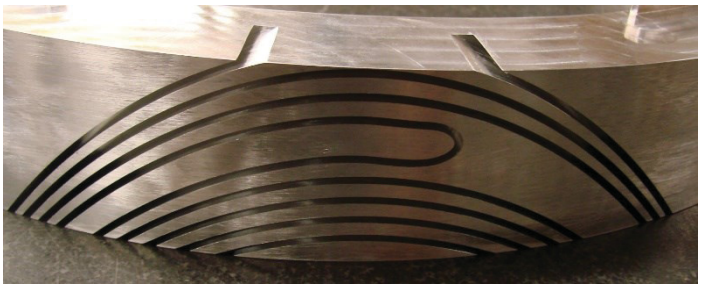

Figure 5.65: *Photo of a quadrupole lamination of a "practice magnet". It shows the machined conductor grooves and the point of reversal between two quadrupole sections.*

And secondly, breaking up a winding turn is not reasonable. A quadrupole turn that by its winding pattern creates four poles - but some poles have inverted polarity – would not create a quadrupole field by virtue of physics. Hence it is not allowed to divide the quadrupole sections into fractions less than an integer number of winding turns. Note that with a total of 340 turns per quadrupole layer this constraint actually impacts the minimal "resolution" by which a length of a section can be tuned.

The strength of a quadrupole is typically given in Tesla as the product of the quadrupole gradient B_1 and the length of the quadrupole l_Q:

$$k_Q = B_1 \, l_Q \tag{5.12}$$

For the AGCCT model a similar value is introduced to measure the strength of a quadrupole section i.

$$k_i \propto n_i \, B_1(I_i) \,, \quad k_i \propto Q_i = n_i \, I_i \tag{5.13}$$

Here the quadrupole strength k is proportional to the product of the number of turns n of the quadrupole section i and the B-field which is a function of the current I. The value Q is a useful quantity to compare the strength of the quadrupole sections and is proportional to k_i for a fixed winding radius. It is important to note that the current I in this context is the total current running through all wires in the winding channel, thus the quantity Q must not be confused with *ampturns*. In the design studied, the dipole windings are made of a stack of 6 wires (connected in series) so that the total current I is the sum of the actual operating current ($I_{operating} = I/6$). Analogously, this is true for the quadrupole layers, each comprising a stack of 8 wires.

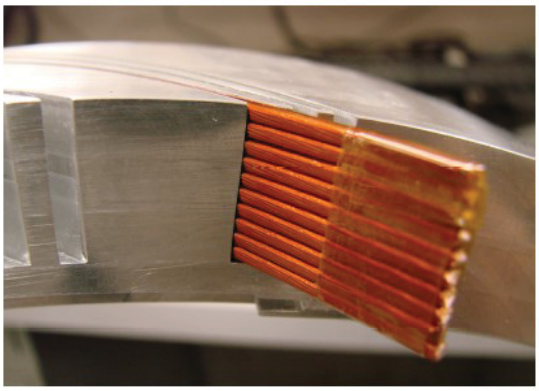

Figure 5.66: *A stack of 10 wires is placed in a channel of an experimental mandrel lamination piece of the AGCCT. Later the design of mandrel and conductor was changed, hence the different number of wires per stack.*

To track the particles through the magnet, the B-field can be directly calculated with the Biot-Savart law along any point of their trajectories. With the currents, i.e. magnet strengths, and partition of the quadrupoles as parameters, automatic fitting routines are used to optimize the system in order to meet the beam optics requirements, whereby the quadrupole layer partition is optimized in an indirect way which is explained below in section 5.3.2.

As a single particle is tracked through the magnet, the methods described above are employed to calculate the transfer map of the system. The order of the calculation is chosen based on the problem to be studied, since this strongly impacts computation time.

5.3 Detailed design and optimization

The following sub-chapters describe the steps of further detailing and optimizing the magnet design. After fixing the bend radius and aperture, i.e. the winding radius, further iterations between the *COSY* realistic model and the magnet model are not necessary. In general, all fitting routines employed are iterative fitting loops, wherein a number of parameters are varied in search of a parameter set that fulfills the specified “endfit” conditions. For a sufficient definition of the system, the number of free parameters must not be greater than the number of “endfit” conditions.

5.3.1 Orbit and start coordinates

The first thing to check with the realistic model is whether the integrated dipole field indeed matches the SCOFF model. From that integrated field, the true bend angle of the magnet can be determined. The coordinate list of the winding geometry is first transformed to a new coordinate system which places a particle on the reference trajectory (optical axis) halfway through the magnet at position (R, 0, 0), wherein *x-z* is the bending plane in the lab coordinate system and R is the bend radius.

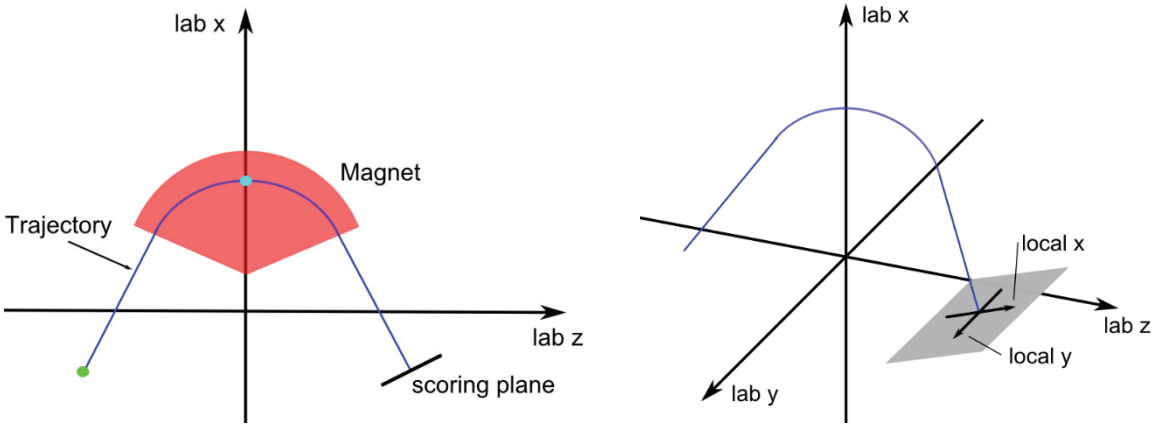

Figure 5.67 The two coordinate systems: *Left: The arrangement of the magnet in the lab coordinate system, which was chosen for symmetry reasons. The green dot indicates the starting point of the optics (and tracking) and the light blue dot marks the "midpoint" of the magnet, which is a symmetry point for the optics. Its coordinates are (0.9, 0, 0) with 0.9 m being the bend radius. The scoring plane, perpendicular to the beam path, and the starting point are equidistant to the magnet (symmetric drift lengths of 1.8 m each). Right: The local coordinate system (2D) is drawn in the light grey plane which is always orthogonal to the beam trajectory.*

For an ideal reference trajectory, the following conditions should be met, where *i, m* and *f* denote the input, midpoint and final values:

$$a_i = -a_f \tag{5.14}$$

$$a_m = 0 \tag{5.15}$$

$$x_i = x_f \tag{5.16}$$

$$x_m = R \tag{5.17}$$

$$y_i = y_m = y_f \tag{5.18}$$

5.3.2 First order fitting

True bend angle and initial coordinates

With the transformed coordinates the beam is supposed to be horizontal at the midpoint ($a_m = 0$). To determine the true bend angle of the dipole, the reference particle is set to (R,0,0,0,0) at $z_i = 0$ and is tracked to z_f which is calculated as the assumed endpoint from simple geometric considerations (bend angle ≈ 67.5° and drift length = 1.8 m). In this first step only the dipole layers of the magnet are turned "on" (all other currents *I* are set to zero). The result gives the coordinates *x, a, y* and *b* at z_f. The dipole current defines the dipole strength, i.e. the bending power. It must be tuned such that the curvature of beam bending matches the physical curvature of the magnet in order to keep the reference trajectory exactly centered. This orbit-

optimization step is very important because a deviating beam would experience kicks by asymmetrically passing through the quadruple fields (once they are switched "on"). Hence it is not possible to adjust the bend angle of the *AGCCT* by changing the dipole current. Instead the dipole bend radius and angle are defined by the physical geometry of the magnet.

Table 11: *First step for the realistic model optics fitting.*

Step 1				
Input	**Endpoint**	**Tune conditions**	**Free parameters**	**Results**
$(R, 0, 0, 0, 0)_{z=0}$	$z = z_f$	Minimize quadrupole effect	Dipole current	Bend angle: a_f

To check that the beam is indeed centered, the quadrupole layers are turned on and the resultant coordinates at z_f are compared. The dipole correction factor (applied to the dipole current) is then iteratively adjusted to minimize the effect of the quadrupoles on the reference orbit. The resultant coordinate a_f gives the actual bend angle of the magnet-half. These coordinates can be used as the set of initial coordinates for the input beam with the relationships given above.

In a next step the results are confirmed by tracking from the new initial coordinates to the midpoint through the first half of the magnet. Due to small beam shifts (or twists) from the residual solenoid in the field the initial coordinates have to be adjusted. This is done with an automatic fitting routine that maximizes the symmetry of the system by demanding x and y to be on the ideal trajectory at the midpoint (x_m, $y_m = 0$). As a result, the values of x_i and y_i are adapted.

Table 5.12: *Second step of the fitting procedure.*

Step 2				
Input	**Endpoint**	**Tune conditions**	**Free parameters**	**Results**
$(x_i, a_i, y_i, 0, 0)_{z=z(i)}$	$z = 0$	$a = 0$ $x = R$ $y = 0$	x_i y_i	Set of initial coordinates

After tracking through the complete magnet confirms that the reference trajectory exactly coincides with the magnetic center of the quadrupoles, the true bend angle and the optimized initial coordinates are found.

Next, the focusing of the magnet is optimized. The quadrupole currents are tuned individually providing five degrees of freedom. A fitting routine automatically searches for the optimal focusing strengths in the various quadrupole sections whereby the length of each segment remains constant.

Table 5.13: *Third Step: Fitting the individual quadrupoles.*

Step 3				
Input	**Endpoint**	**Tune conditions**	**Free parameters**	**Results**
$(x_i, a_i, y_i, b_i, 0)_{z=z(i)}$	$z = z_f$	me(1,6) = 0 me(2,6) = 0 $\alpha_x = 0$ $\alpha_y = 0$ $\beta_x - \beta_y = 0$	Quadrupole currents of sections: 1, 2, 3, 4, 5	Quadrupole strengths

The fit conditions for the focusing at z_f (which corresponds to the isocenter), reflect the beam optics requirement: The symmetry of the beam is warranted by equal beta functions in *x* and *y*, a double beam waist requires the alpha functions in *x* and *y* to be zero and the achromaticity of the system demands that the dispersion function vanishes. The map elements (1,6) and (2,6) of the linear map are the linear *x* and *a* dependence on the energy. Since there is (almost) no beam bending in *y*, me (3,6) and me (4,6) are naturally zero.

The fitting routine described above can consume a significant amount of computing time which usually means an overnight job for a DELL Latitude 7480 with an Intel® Core™ i7-7600U processor. Apart from the variation of five parameters also the integration over all conductor pieces of all layers contributes to the computing time. The calculation results are five different currents I_i for the five quadrupole sections i. In contrast to the *COSY* model the five quadruple sections of the real magnet are connected in series and thus a single identical current runs through them. Therefore, instead of the currents the lengths of the sections are adapted so that the quadrupole strength *k,* approximated by field gradient times its length, remains constant. This is proportional to the value Q_i, the current of a section I_i times the number or turns of the section n_i, see equation (5.13).

Table 5.14: *Redistribution of windings per section.*

i:	**1**	**2**	**3**	**4**	**5**
Fit result:	$Q_1 = I_1 n_1$	$Q_2 = I_2 n_2$	$Q_3 = I_3 n_3$	$Q_4 = I_4 n_4$	$Q_5 = I_5 n_5$
Intermediate step:	$Q_{tot} = \sum_{i=1}^{5} I_i n_i$; $I_{tot} = Q_{tot}/ n_{tot}$ which is the common current for all sections.				
Equal current result:	$n_1^* = Q_1/I_{tot}$	$n_2^* = Q_2/I_{tot}$	$n_3^* = Q_3/I_{tot}$	$n_4^* = Q_4/I_{tot}$	$n_5^* = Q_5/I_{tot}$

Apparently, this method can lead to non-integer number of turns for a given section which would not be practical for the reasons listed above. How this problem can be handled is discussed in section 6.5.5 in context of the overall tracking results.

Effects of higher orders (check feed down effects, alpha, beta etc.)

In the next step the linear solution is tested in a higher order calculation to examine any feed down effects to the linear measures of the fit condition, namely the linear dispersion and the alpha and beta functions. The sextupole layers are turned on and by choosing a higher order of calculation, e.g. fifth or seventh, respective field components due to inhomogeneities or those of fringe fields can be taken into account. The optimization of the sextupoles, which compensate for the chromatic aberrations of the quadrupoles, is done manually in an iterative process starting again with values from the SCOFF result.

5.3.3 Higher order transfer maps

As the last step with this *COSY* code, individual transfer maps for tracking are produced in various orders. Given the large momentum acceptance and the consequently deviating trajectories for the different beam energies, multiple transfer maps need be calculated, one for each case of relative momentum offset, spread out over the considered range from -10.5 to +14 % *dp/p*.

5.4 Tracking with transfer maps.

In a separate code, also in *COSY*, the tracking of large number of protons is performed with the help of the transfer maps following the methods described above. Since this only means plugging in starting coordinates into an equation, the method is fast. Using a regular laptop, a million protons can be tracked through a fifth order map in only a few minutes.

The initial proton distribution is set up according to a definable emittance and follows a 2D Gaussian distribution. In the case of the proposed compact proton therapy gantry the *AGCCT* follows the degrader and its collimation system (compare with **Figure 3.21**). Hence the actual proton distribution feeding into the magnet can be expected to be a small cutout of a wider Gaussian distribution. In principle, this situation could also be emulated by the code using a configurable cutoff value. However, in previous calculations of the beam optics from the degrader to isocenter for existing proton therapy beamlines (at Varian), a so-called effective beam emittance, taken as an initial Gaussian emittance without cutoffs, was found to have very good agreement with measurements [151]. One should also keep in mind that all calculations refer to vacuum, omitting any scattering effects due to air or the many foils of the ionization chambers and the beamline exit window.

The number of protons in the calculations, typically between 100,000 and 1,000,000, is large enough to ensure good statistics in the analysis of the final proton distribution. In fact, this range is close to the order of magnitude of delivered protons in an actual clinical pencil beam spot (approximately $10^7 – 10^8$).

The tracking results are a list of particle phase space coordinates, all relative to the reference particle. This makes an analysis in real space as well as in phase space possible. Only at this step do higher order effects (depending on the order of the transfer map and the chosen order of tracking) become visible. Using MatLab, graphs and histograms of the resulting particle distributions were produced, including the application of 1D and 2D “Gauss-fits”. The fitted Gauss curves are used to determine the size (RMS values) in *x* and *y* of the beam spot. These

are the same analyses and procedures that are also employed for experimentally measured proton distributions.

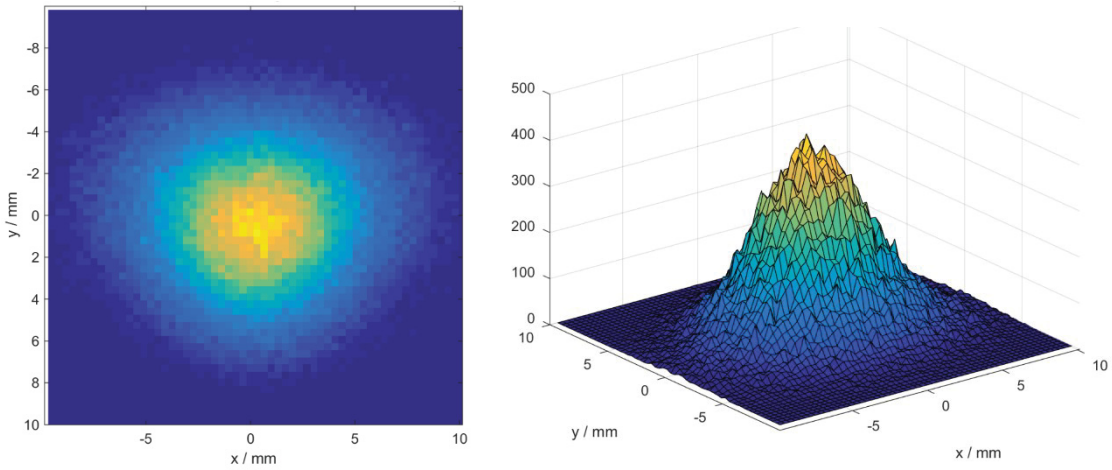

Figure 5.68: *Left: 2D plot of the beam spot as a tracking result at isocenter in real space. Right: The same spot is shown in a 3D view. Note that on the z-axis the number of protons per pixel (0.5 x 0.5 mm²) are counted. All protons have the same z-coordinate in the scoring plane.*

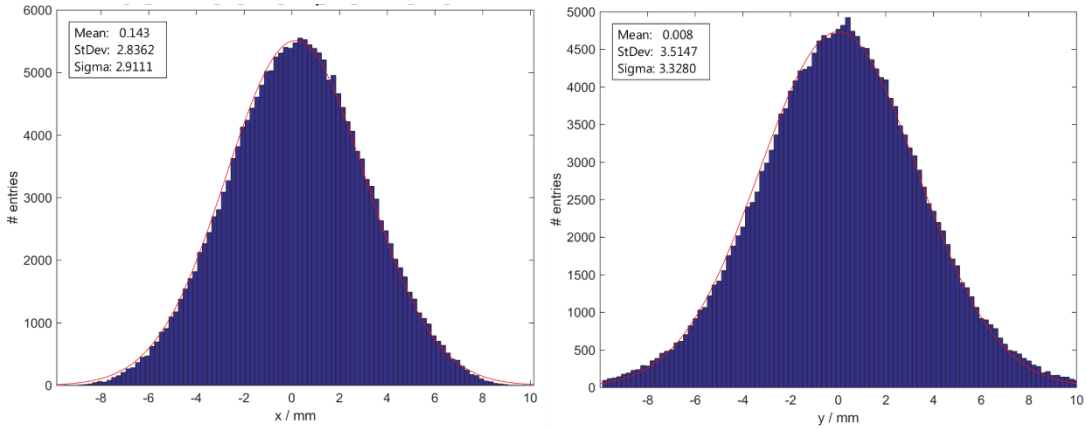

Figure 5.69: *Two histograms of the same spot of figure 5.68 along the x and y axes. The red line shows the Gaussian fit for which some characteristic values are printed in the upper left corner.*

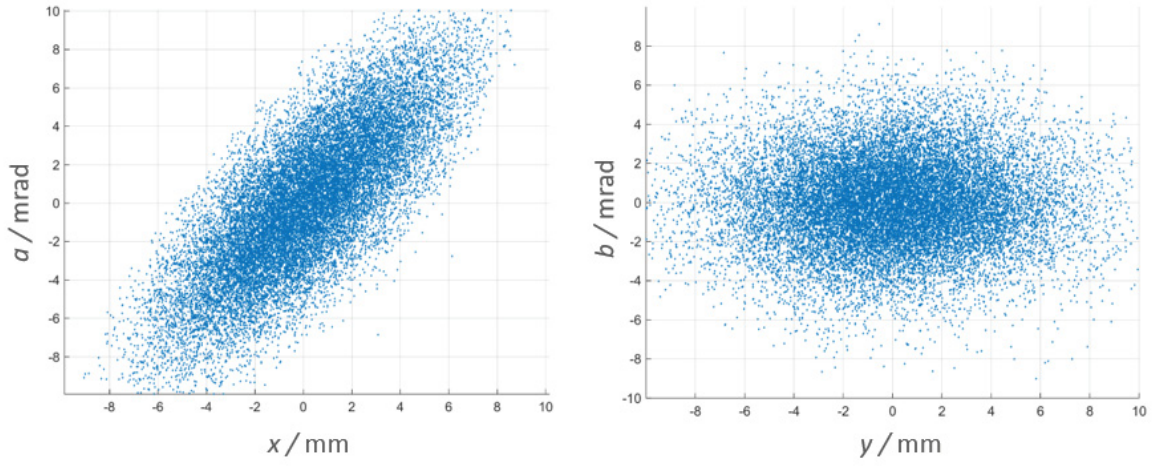

Figure 5.70*: The two plots show the phase space diagram in x (left) and in y (right). From the shape one can tell that the beam in this example does not have a waist in x at isocenter. Instead it is diverging.*

5.5 Beam scraping

Besides the isocenter, other locations along the beamline are also of interest. It is important to examine the beam size throughout the beamline and assess possible beam losses due to “beam scraping”. Since the protons in the beam are normally distributed, it is often unavoidable that some protons will be lost at tight locations or where the beta or dispersion functions assume large values. As mentioned before, minimizing apertures has advantages from a technical and cost perspective, hence, in conventional single-pass-through systems, such as proton therapy beamlines, relatively large beam losses are tolerated. Values of 50 % and more are quite common but considering the immense intensity losses at the energy degrader, these additional losses can be considered insignificant. This is in complete contrast to periodic beamlines, such as storage rings, where the beam passes through the same optics many times, making beam loses prohibitive.

Of course, one could reduce the beam losses by simply reducing the emittance of the beam. But in context of transmission maximization, it can be better to choose a larger emittance and accept extensive beam scraping by filling the apertures as much as possible. However, without going into too much detail, this approach is limited since it can also affect beam shapes at isocenter as well as lead to activation and radiation hazards. In this particular example significant beam loss in the superconducting magnet should be avoided. The main reason for caution is the sensitivity of the epoxy, whereas only larger proton numbers can cause meaningful heating effects above 0.1 W. The heating effect can be exemplarily demonstrated by an estimate for the maximal heat load P_{max} if the complete kinetic energy of the protons is transformed to heat:

$$P_{max} = I_p\, E \overset{exempl.}{\Longrightarrow} 1\ \text{nA} \cdot 220\ \text{MeV} = 0.22\ \text{W} \tag{5.19}$$

There are different methods to assess possible beam scraping. Instead of following the trajectories of individual particles – a rather tedious task due to the large numbers that are required to ensure meaningful statistics – tracking with transfer maps to the location in question is more practical. Since the transfer maps do not contain any information about the beam in-between start and finish, the locations where beam scraping could occur, or is most pronounced, have to be identified first.

5.6 Methods for the fixed field magnet

5.6.1 Workflow

The workflow and methods are described in detail in chapter 4.7. With the indirect approach employing an analytical fit function the beam optics of the *fixed field magnet* were calculated. In the respective code in *COSY* the function $B_y(x)$ directly returns the B-field for any particle coordinate according to its x value. The quadrupoles in the triplet were handled as standard SCOFF models, which is accurate enough given that they represent regular quadrupoles with small apertures. The triplet allows three parameters of freedom for optics tuning while the second triplet (downstream of the bend) is always set to identical values to preserve symmetry.

It must be noted that Weishi Wan did the majority of work to tune and find the beam optics solution, which is why the results and descriptions detailed here and in chapter 6 focus on the beam tracking and the results from the beam profile analysis.

5.6.2 The analytical fit function

A good fit to the shape of the field of two nested racetrack coils with a c-shaped iron yoke, was found by use of three Gaussian functions plus an additional offset B_0.

$$B_y(x) = a_1 e^{-\left(\frac{x-b_1}{c_1}\right)^2} + a_2 e^{-\left(\frac{x-b_2}{c_2}\right)^2} + a_3 e^{-\left(\frac{x-b_3}{c_3}\right)^2} + B_0 \tag{5.20}$$

This function allows for variation of ten parameters to adjust the field profile for the optics. Furthermore, the function is infinitely derivable which makes a Taylor expansion to arbitrary orders feasible, enhancing the precision of the optics calculation.

5.6.3 Tracking through the fixed field magnet

The origin of beam path that was used for the tracking is 1.8 meters upstream of the entrance to the quadrupole-triplet. The intention is to keep it symmetric with the 1.8 meter drift downstream of the system. This particular value was chosen to respect the assumed space requirement for a compact short-SAD scanning system, compare section 3.4.2.

Since the upstream drift is relatively long, this part, including the first quadrupole-triplet, was additionally investigated with *Transport* to check for issues with beam scraping. Due to the absence of non-linear fields up to the entrance point of the bend magnet, the method is accurate.

The tracking with transfer maps is in principle identical to the treatment of the *AGCCT*, see chapter 5.4. A 7th order map includes all relevant effects, which extend to a slightly higher order than for the *AGCCT* (5th order) to respect the stronger non-linearity of the field.

6 RESULTS AND DISCUSSION

In this chapter the results of the three studies, the novel proton therapy gantry layout, the analyses of the beam optics of the *AGCCT* and of the beam optics of the *fixed field magnet*, are detailed. Consequences and meaning of the individual results are discussed and interpreted within the respective subchapters.

6.1 Proposed proton therapy gantry

Both discussed achromatic solutions, the *AGCCT* as well as the *fixed field magnet*, can be utilized in a novel beamline concept for a compact proton therapy gantry system. The proposed design aims to combine compactness, high treatment performance and simplicity. It is based on the achromatic bend and the incorporation of the degrader onto the gantry spine as described in chapter 3.5, an idea jointly developed by PSI and Varian [109, 152]. This setup maximizes the potential of achromatic bending magnets not only because of the length reduction due to the new degrader position but because of significantly increasing beam transmission while apertures in all other beamline parts are downsized. The layout of this gantry is shown in **Figure 6.71** and **Figure 6.72** for both examples, the *AGCCT* and the *fixed field magnet* versions.

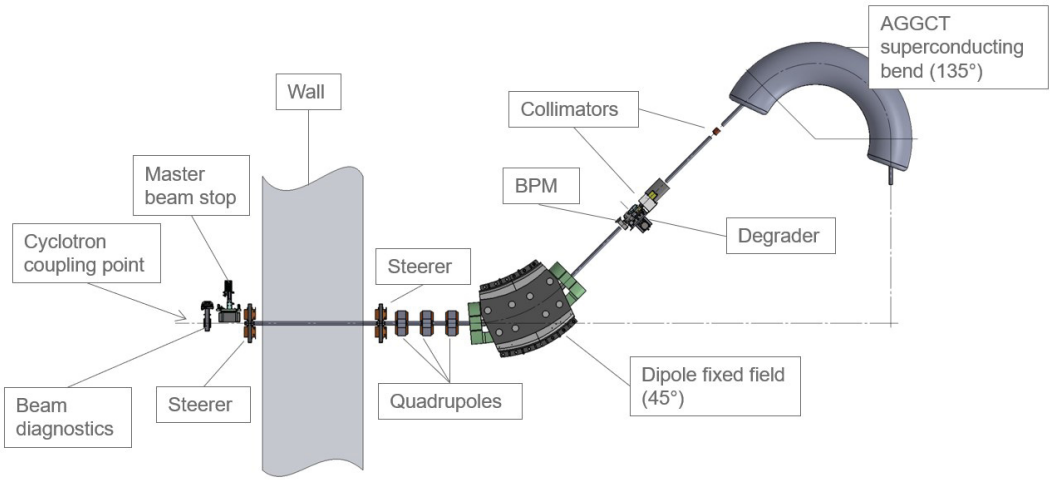

Figure 6.71: *2D view of the optimized gantry layout with AGCCT magnet. The starting point for the beam optics is the coupling point to the cyclotron. The quadrupole triplet focuses the beam to the degrader position. The set of steerers (both combined steerers for both axes) correct any beam misalignments with respect to the gantry. The collimators downstream of the degrader define a new beam emittance for the beam transport to the isocenter which is at the intersection of the two dashed lines.*

The depicted wall is separating the cyclotron room (on the left) from the gantry and treatment room. Here, deviating from other existing systems, only very few components are outside the treatment room. Hence the size of the cyclotron room is minimized.

Along the beam path starting at the cyclotron exit the following beamline elements are essential: A first diagnostic box is needed to measure the beam current and position before the required master beam stopper. Next, two combined steerer magnets that are placed on both sides of the

radiation shielding wall adjust the beam angle in the horizontal and vertical planes. Together, they can correct the beam position and angle to exactly match the rotational axis of the gantry. Note that a drift length between the steerers is always necessary to effectively correct beam shifts, which coincides with the drift through the shielding wall in the proposed layout.

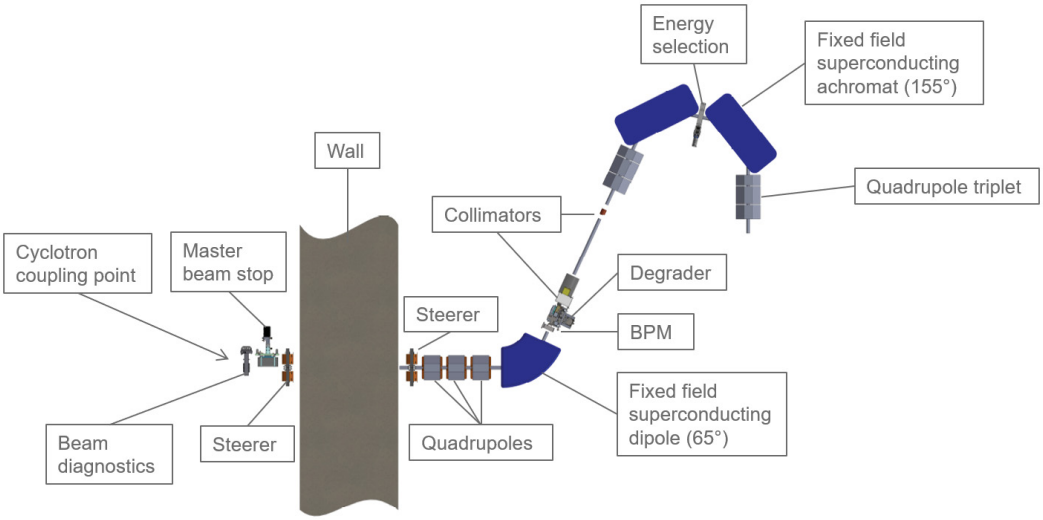

Figure 6.72: *2D view of the CAD drawing of the fixed field magnet gantry displaying the most important elements of the beamline. The beam is fed in from the left at the cyclotron coupling point.*

The three quadrupoles at the gantry entrance are the first elements that rotate with the gantry. They are used to focus the beam from the cyclotron coupling point onto the degrader through the first dipole by point-to-point imaging. Since the cyclotron produces an asymmetric beam profile the quadrupoles manipulate the magnification of the imaging in order to obtain a round beam at the degrader. The beam profile monitor ("BPM") at the degrader location is necessary to measure the beam parameters at this important location.

An important distinction of the proposed optics is the neglection of the dispersion function. Instead of correcting the dispersion, the one-dimensional broadening of the beam (in the horizontal plane) is compensated by an asymmetric input beam profile. In other words, starting with a small horizontal size and a medium vertical size yields a symmetric beam when dispersion is included. The resultant size can be as small as 1 mm (1 RMS) – see **Figure 6.73** where the markers at the turquoise line correspond to 1 mm. The effect of this method is that neither a second dipole for global achromaticity nor a local achromatic bend is required, saving costs and reducing the size of the beamline.

The studied front-end ends at the degrader position where the beam properties are almost completely destroyed (and a new emittance must be defined). Hence small deviations from symmetry in the beam are not very critical.

Since the beam energy is constant throughout the front-end, the dipole can be run in static mode (no ramping). It is proposed to employ a simple and cost-effective SC racetrack dipole (NbTi) to provide a *B*-field of approximately 4 Tesla to maximize the compactness of the gantry. Of course, a resistive electromagnet or even a permanent magnet would also work.

The degrader is pointing upwards and away from the patient area. This is an advantageous feature of the gantry design since the degrader is a major neutron source – but emitting most neutrons in the forward direction. Thereby, a relatively slim local shielding of the degrader

becomes sufficient. In the degrader the beam properties are not preserved and a new beam emittance must be defined as a consequence. This is achieved with the standard means of the size and divergence collimators as described in section 3.1.3. It is known empirically that the 2/3 length of the degrader block makes a good assumption for the new starting point of the following beam optics [151].

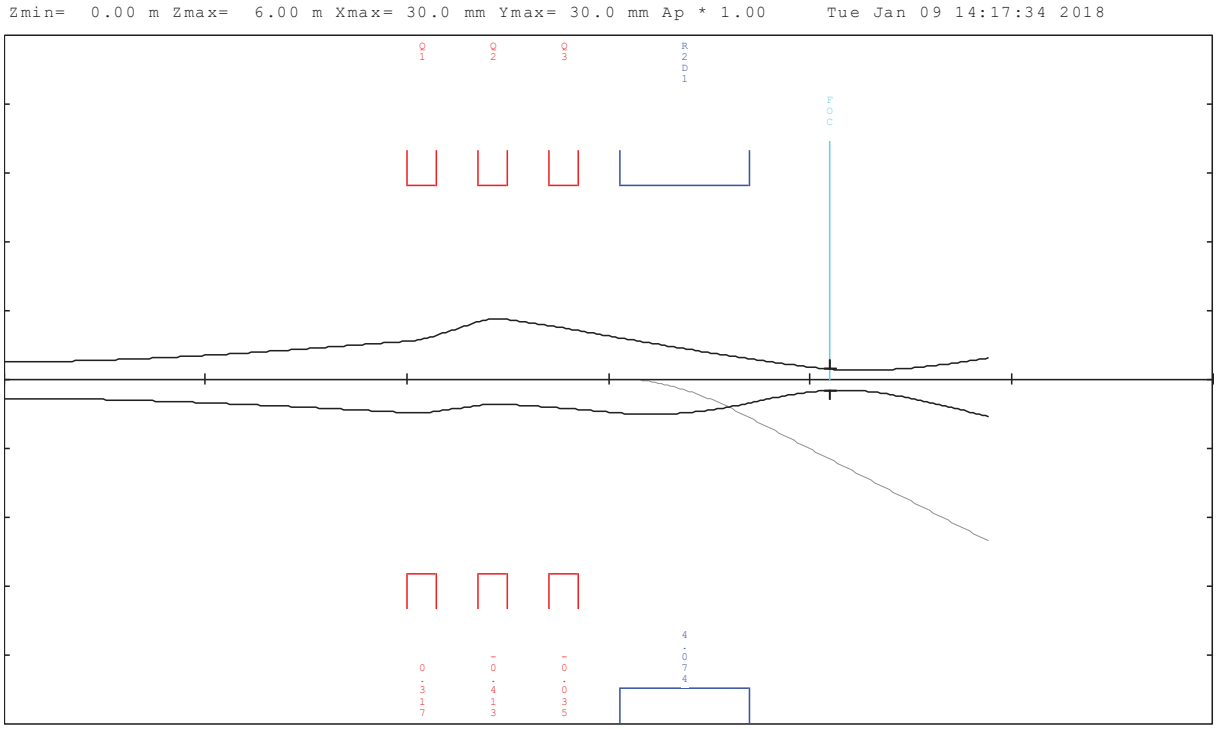

Figure 6.73: *Screenshot of the Transport graphic result for the front-end beam focusing. The beam is tightly focused to the degrader while the dispersion function (dashed line) remains uncorrected. The red rectangles indicate the quadrupoles and the blue one the dipole.*

In summary, the main advantages of the proposed gantry are its short length, the reduced number of beamline elements and the elegant simplicity of the optics, which enhance the performance, i.e. the transmission, and potentially the reliability of the beam transport.

6.1.1 Simplicity of gantry optics

The simplicity of the optics lies in the point-to-point-to-point concept with approximately 1:1:1 imaging, from the cyclotron to the degrader and from the degrader to the isocenter. Note, that these are actually two independent optics. The quadrupole triplet of the front-end complies with simple textbook optics while the optics of the *AGCCT* and *FF magnet* are, admittedly, more complex. However, when considering these magnets as black box beamline elements, the beam path from degrader to patient is simplified to a minimum. Another consequence of the setup is that the beam properties at isocenter are independent of the gantry angle, an attribute not necessarily granted for existing proton gantries.

6.1.2 *Size considerations*

One important goal of the developments for future proton therapy systems is compactness. It is not exactly clear how the mere size of the gantry influences buyers' decisions but it is also a driver for cost. The cost of the bunker of a compact system is approximately 25 % of the overall investment, hence savings there can be significant. **Figure 6.74** shows some dimensions of the *fixed-field magnet* based gantry and system. The gantry radius of 3.5 meters is 30 % smaller than the conventional 5 meters but about the same as the compact Proteus®One gantry (3.6 meters). Furthermore, it is mostly achieved by a reduced SAD of the downstream scanning system – not by the high fields in the bends. This is mostly due to the height contributions of the straight geometry of the magnets as well as the space consumption of the quadrupoles. On the other hand, the length of the gantry is only 4 meters, a reduction of more than 50 % compared to the current Varian ProBeam® gantry. With a system length of 6.1 meters (excluding the accelerator) it is in total about 50 % shorter than the IBA Proteus®One design.

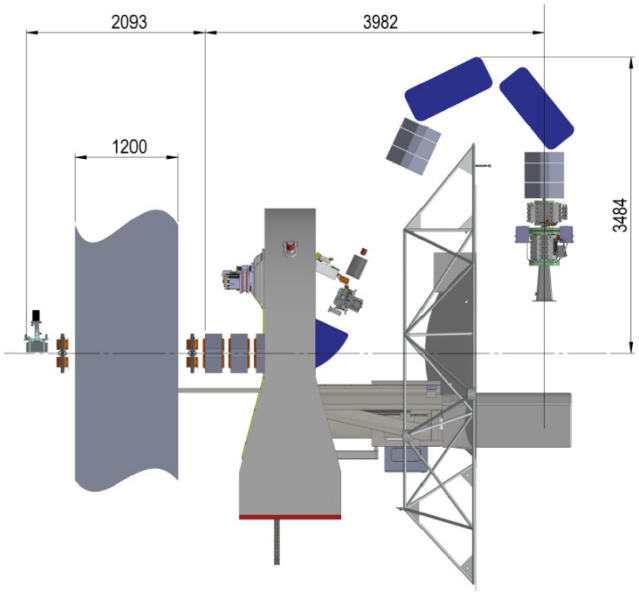

Figure 6.74: *The proposed system with the fixed-field magnet. Dimensions are given in mm.*

Table 6.15: Size comparison of the two studied system with the most compact current gantries [119, 120]

Proposed and existing systems	System Length (w/o cyclotron)	Gantry Length	Diameter
AGCCT based system	7.8 m	5.7 m	6.6 m
Fixed-field magnet based system	6.1 m	4.0 m	7.0 m
IBA Proteus®One	11.7 m	7.9 m	7.2 m
ProNova SC360	n/a	6.4 m	8.2 m

6.2 Beam scraping

6.2.1 *AGCCT*

In the AGCCT system there are three bottlenecks for the beam. These are the entrance and exit of the curved toroidal magnet and the symmetry point at the middle of the magnet, see **Figure 6.75**. Due to the symmetry (minor deviations can be neglected) it is sufficient to study the entrance losses as they also cover the exit problem: a trivial task since there are no magnetic elements between the degrader and the entrance of the magnet. Hence the resulting beam size at the entrance edge only depends on the drift (1.8 meters) and the initial emittance. Here it is worthwhile to remember that the input proton distribution is in fact a Gaussian with cut tails.

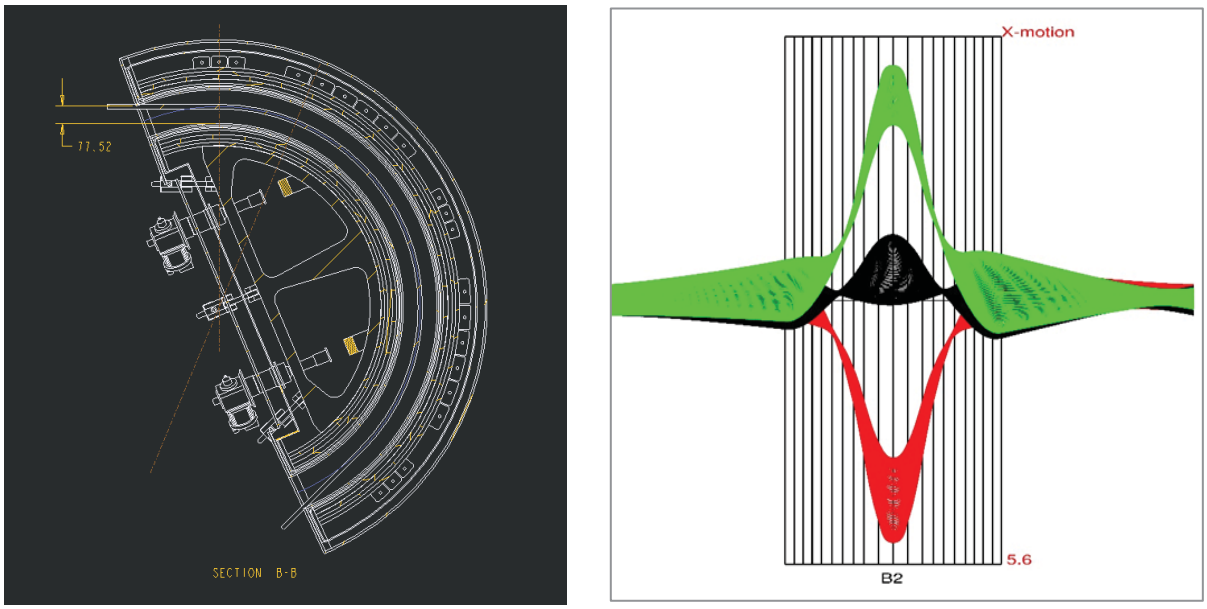

Figure 6.75: *Bottlenecks with regards to beam clearance. Left: The drawing shows the effect of the extending ends of the AGCCT. The straight beam path gets close to the edge of the curved structure at entrance and exit. Right: Proton trajectories of three different momenta: green = positive deviation, black = reference momentum and red = negative deviation. Halfway through the AGCCT (at the symmetry point) the deviations are largest from the optical axis, marking a third bottleneck for the beam.*

Actually, the main issue for beam scraping is that, when passing the entrance edge, the beam is offset by about 77.5 mm as shown in the figure above. With a clear half-aperture of 85 mm this leaves only 7.5 mm for the radial beam size. Due to the divergence during the drift the beam grows from 3 mm to 5.2 mm (1 RMS). Particles outside 1.44 RMS of the distribution are therefore lost. How large of a percentage really is affected is not easy to judge due to the cut tails of the Gaussian at 1 RMS of the distribution at the location of the size collimator). A thorough Monte Carlo study could answer this question but, in any case, these losses are tolerable for a single-pass beamline.

The situation halfway through the magnet is different. There the deviation of the beam path from the optical axis depends on the relative momentum. In fact it is the aperture of the magnet that limits the maximum momentum acceptance of the system. If beam losses inside the magnet shall be avoided, i.e. three RMS of the beam must stay clear of the aperture, the maximal permissible momentum deviations are +10.5 % and - 6 % *dp/p* as listed in **Table 6.16**. Hence, depending on whether the definition refers to the set energies of beams or the lowest and highest momenta of protons that are correctly transported, the *AGCCT* has a momentum acceptance of 16.5 or 23 % *dp/p*. Since for therapeutic treatments the mean energy and the corresponding range are of primary interest, the definition with an acceptance of 16.5 % is adopted here.

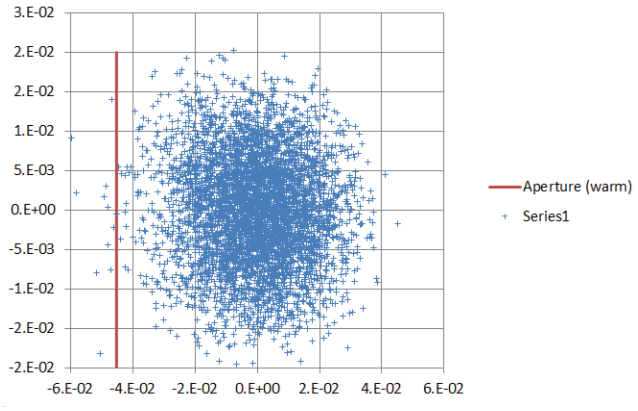

Figure 6.76: *5,000 protons at + 10.5 % dp/p including the intrinsic momentum spread tracked halfway through the AGCCT (to its middle). The red line indicates the relative location of the aperture of the magnet. The units for both axes are meters.*

Table 6.16: *Resultant momentum acceptance of the AGGCT.*

Definition	Momentum acceptance	Lowest relative momentum	Highest relative momentum
Beamline set energies: The average momentum of a given proton beam.	16.5 % *dp/p*	- 6 % *dp/p*	+ 10.5 % *dp/p*
Proton momenta that are transported: Taking the momentum spread of a beam into account.	23 % *dp/p*	- 10 % *dp/p*	+ 13 % *dp/p*

6.2.2 Fixed-field magnet

In the optics of the *fixed-field magnet* the beam size is largest in the quadrupole triplet. A linear optics calculation in *Transport* was conducted to determine the beam size and required apertures at this location. The maximal size (1 RMS) is close to the end of the 2nd quadrupole, see **Figure 6.77**, at 8.2 mm. The aperture of the quadrupole is limited by the maximal acceptable pole tip field from a magnet perspective in combination with the required quadrupole gradient. With a 45 T/m gradient and a maximal pole tip field of ~ 0.8 T, the half aperture is limited to about 18 mm. Considering 2 mm for the vacuum pipe and tolerance this leaves room for almost 2 RMS of the maximal beam size (8.2 mm). Since the proton beamline is a single-pass through system this clearance is acceptable. Proton beam losses in the room-temperature quadrupole is less critical from a radiation safety point of view as they are located far away from the patient and the primary direction any produced radiation is point away from the treatment area (compare **Figure 6.72**). Inside the superconducting bend only the vertical size is critical but stays below 12.5 mm and thus clear of the bore (value obtained with transfer maps).

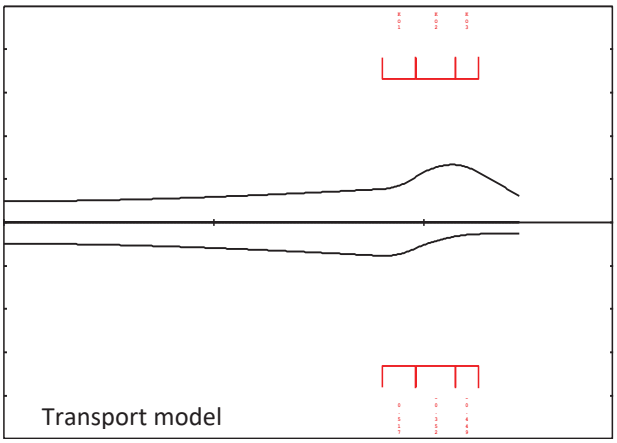

Figure 6.77: *Exemplary Transport graphic output displaying the beam envelope through the quadrupole triplet.*

6.3 Momentum spread

Range straggling in the energy degrader induces a momentum spread in the beam. To reflect this momentum spread for the various beam energies, an approximation was used and conservative estimates on the high side were used for the tracking.

In [153] the momentum spread for various degraded beams, starting with monoenergetic 250 MeV, is calculated with various theoretical models [154, 155] and experimentally measured as well, see **Figure 6.78**.

For the present study an accelerator with a max energy of 230 MeV is assumed. Hence the values from above are scaled to this lowered input energy, in a simple first order approximation. The largest energy spread at 70 MeV is assumed to be similar to the 90 MeV momentum spread

from the reference data and thus 2 % *dp/p* (1 RMS) is used as the largest momentum spread in the beam. For all tracked beam energies four different cases with regards to the momentum spread have been studied (see **Table 6.18**).

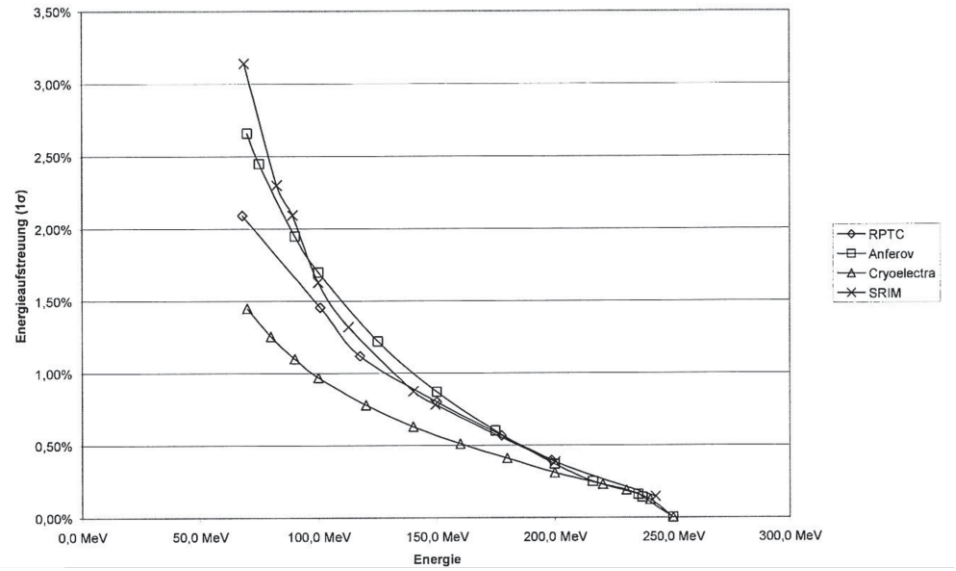

Figure 6.78: *Simulated, calculated and measured momentum spreads after a proton beam passes through a carbon degrader [153]. Note that the German description of the ordinate is wrong. The values correspond to the 1 RMS (σ) value of the momentum spread ("Impulsaufstreuung") not energy spread ("Energieaufstreuung").*

As pointed out in [162] the actual energy distribution after degradation is not a pure Gaussian but overlapped with a linearly dropping tail that extends to very low energies. According to *FLUKA* [164] simulations at lower beam energies (more "degradation") the tail of the energy distribution can contain up to a few percent of the overall beam intensity, thus a non-negligible fraction. As discussed in [162] for achromatic bends this poses a problem: Due to the large momentum acceptance even particles with strongly deviating momenta could potentially get through the system and affect dose delivery. They could also get lost within the system and if those proton numbers are large, potential damage or heating of the magnet must be studied in detail. For the *AGCCT* in combination with the proposed gantry this can become an issue because of the lack of an energy selection system. Hence, once a final design of magnet and gantry has been selected, the subject should be studied in more detail.

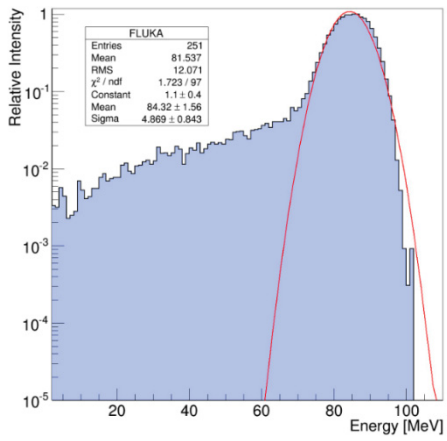

Figure 6.79: *Energy distribution after degradation from 250 to approx. 84 MeV, simulated with FLUKA [162]. Note the logarithmic scale for intensity.*

6.4 Optimized fitting results of the *AGCCT*

According to the methods and steps described in chapter 5 the quadrupoles and sextupoles of the *AGCCT* are optimized. Below a table shows the final partition of the five sections and the respective fields and gradients.

Table 6.17: *Layers and sections of the AGCCT after optimization.*

Type	Layers	Section 1	Section 2	Section 3	Section 4	Section 5
Length in turns	1,2	54	69	94	69	54
Length in degrees	1,2	22.71	26.58	36.42	26.58	22.71
Dipole in T	3,4	2.29	2.29	2.29	2.29	2.29
Quadrupole in T/m	1,2	21.28	-21.28	21.28	-21.28	21.28
Sextupole in T/m^2	1,2	0	-17.6	0	-17.6	0

The sextupole showed two strong sections and three very weak sections which were rounded to zero. Since the exact length of the sections is of minor importance it is adjusted such that the sextupole partition length is equal to the quadrupole. Thereby, the sextupole can be integrated into the quadrupole layers (as described in chapter 4.6). This is a very important step since a six-layer magnet would have been technically much more challenging and costlier as well. The other three sextupole sections (1,3,5) are rounded to zero to simplify the design.

From this setup the magnet model was created and, working with the *realistic model* in *COSY*, the magnet was optimized. The orbit optimization (compare section 5.3.1) resulted in a true bending angle of 133.4°. The deviation from the designed 135° can be explained by fringe field effects (the non-equivalence of the dipole integral along the beam trajectory). This can be taken into account in the gantry design by adapting the other bend angle of the first dipole. However, in contrast to other dipole magnets, changing the dipole field of the *AGCCT* to tune the bend angle is not allowed because the proton beam bend radius must not differ from the mechanical bend radius of the magnet.

The individual quadrupole strengths are tuned to achieve the demanded conditions for the alpha, beta and the dispersion functions at isocenter for the on-momentum case. Note that while these are parameters of linear optics, the results here include feed-down effects from higher orders as they were calculated in 5^{th} order. The same exercise (with fixed values) is then repeated for the expanded momentum bandwidth.

The graph (**Figure 6.80**) shows that the *AGCCT* is not a perfect achromat over a larger momentum range. While the dispersion function (at isocenter) stays relatively flat for positive momentum deviations, it quickly rises for negative deviations. Hence, increasing aberration effects, visible in the horizontal spot size, can be expected.

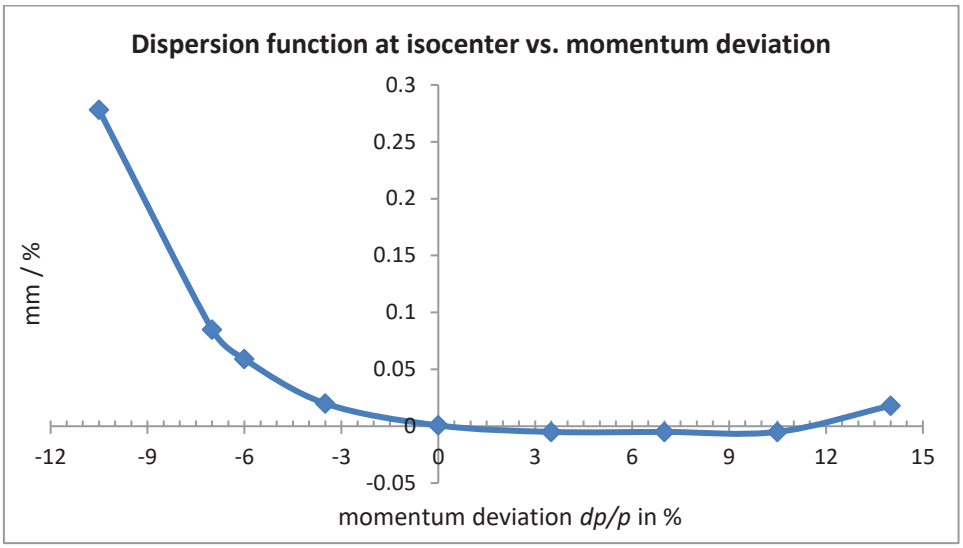

Figure 6.80: *First order dispersion at isocenter, matrix element (1,6), against momentum deviation (line to guide the eye). For negative momentum deviations the dispersion does not vanish. Effects on spot shapes and sizes must be expected.*

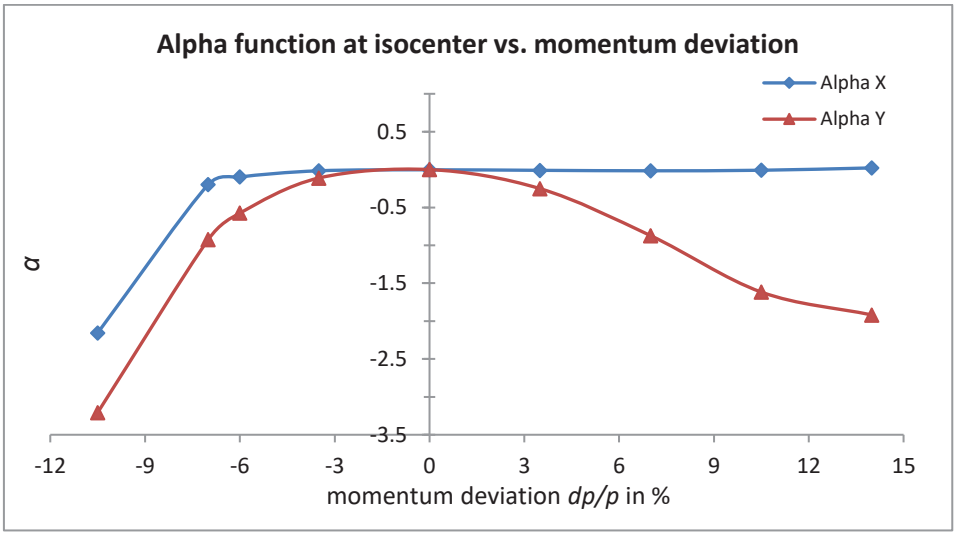

Figure 6.81: *Alpha function at isocenter against momentum deviation (lines to guide the eye). In the horizontal plane (x) the alpha function stays flat at zero over a wide momentum range. In the vertical plane (y) this is different and to both sides the graph falls off.*

The alpha function differs between the vertical (y) and the horizontal (x) plane. The horizontal focusing stays constant over a wide momentum range. In the vertical plane this is not the case. Instead, for both, positive and negative momentum deviations, the alpha function assumes negative values. That means the beam is diverging and the waist is no longer at isocenter but shifted upstream. This is most likely due to the fact that with increasing momentum deviation the curvature of beam trajectories become more pronounced (compare with **Figure 6.95**). With larger angles the proton path is less aligned to the reference trajectory, thus field gradients seen by the beam become slightly distorted. However, the shift of the beam waist can be corrected with the help of quadrupoles which is discussed in more detail in section 6.5.5.

The plot of the beta functions at isocenter (**Figure 6.82**) is consistent with the observation of the shifted beam waist.

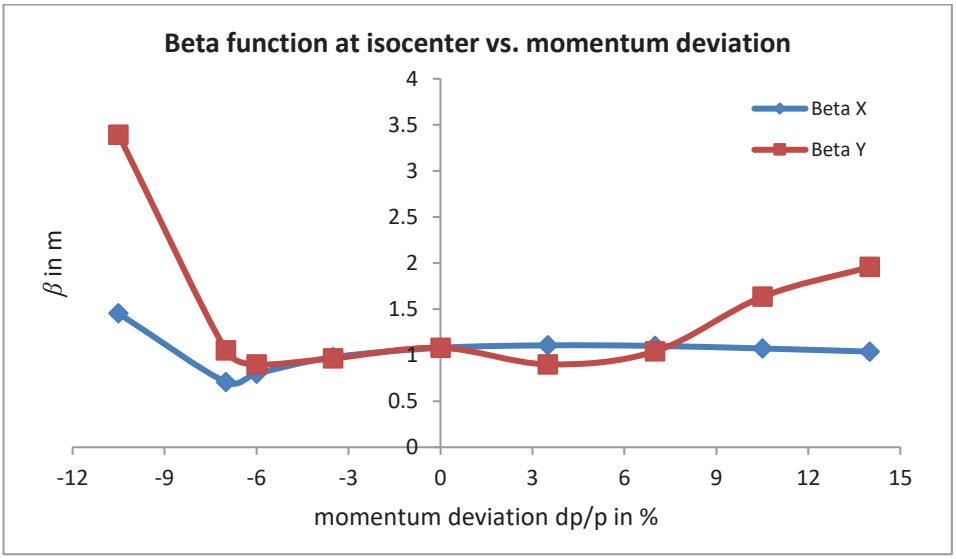

Figure 6.82: *Beta function at isocenter against momentum deviation (line to guide the eye). Between -7 and +7 % dp/p the beta function stays relatively close to 1 m. Consistent with the alpha function the horizontal beam size (x) stays confined even for momentum deviations above 10 %.*

From equation (2.24) the beta function gives a direct indication of the beam envelope by its square root ($x_{max} = \sqrt{\varepsilon \beta_x}$). As the initial value of β is 1.08 m (at the starting point of the optics), the optics achieve close to 1:1 imaging over a relatively large range from -7 to +7 % *dp/p*. However, these results only resemble the linear part of the optics. Higher order effects can still influence the spot sizes at isocenter and are only unveiled by higher order tracking results, as investigated in the following section.

6.5 Results and discussion of the AGCCT

6.5.1 Overview of tracking studies

This section summarizes the tracking studies that have been performed for the *AGCCT* magnet. All results are from 5th order tracking, thus take all relevant higher order effects into account. If not stated otherwise the tracking was performed with 500,000 particles. A realistic emittance (30 π mm mrad) with an initial Gaussian shape of the beam was assumed. Multiple cases of momentum spreads in the beam were tracked (between 0 and 2 % *dp/p* as the 1 RMS value of the corresponding Gaussian).

In summary a total of 108 tracking studies were performed for the described system.

Table 6.18: *List of performed tracking studies. Note that there is no logical relation in between the rows. All momentum deviations have been calculated in 1st, 3rd and 5th order and for all four cases.*

Momentum deviations in % *dp/p*	-6	-5	-3.5	0	3.5	5	7	9	10.5
Orders of transfer maps	**1st**			**3rd**			**5th**		
Momentum spreads, 1 RMS values	**Case 1** 0 % dp/p			**Case 2** 0.5 % dp/p		**Case 3** 1 % dp/p		**Case 4** 2 % dp/p	

The results from 3rd and 5th order tracking did not show any significant differences. Hence it is concluded that a 5th order calculation / tracking is sufficient to cover all relevant higher order effects. This is visually demonstrated for an artificially "bad case" in the figure below.

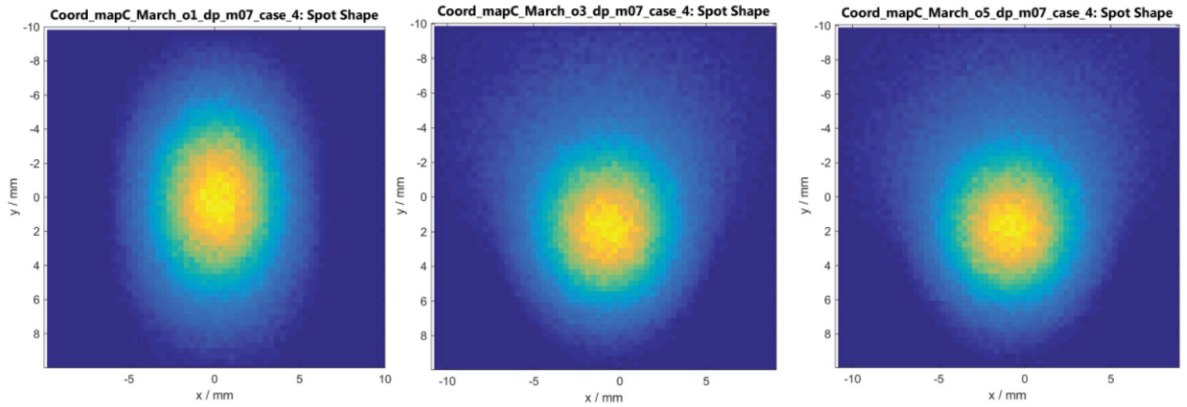

Figure 6.83: *Comparison of 1st, 3rd and 5th order spot results. Exemplary beam profiles for a momentum deviation of -7 % dp/p is shown. The only difference between the three plots is the order of tracking: Left 1st, center 3rd and right 5th order. The inclusion of 2nd and 3rd order effects makes a visible difference. Expanding the calculation to 5th order, however, does not affect the spot profile.*

6.5.2 Results from tracking

The tracking results were analyzed at a scoring plane at isocenter, compare **Figure 5.67**. Below, the 2D plots of various beams across the momentum acceptance window are displayed. Each relative momentum is tracked with four different momentum spreads in the beam, classified in cases 1 to 4 as defined in **Table 6.18**. Each result resembles the real space distribution of the 500,000 protons, scored in pixel sizes of 0.5 mm times 0.5 mm. The coordinates *x* and *y* relate to the mean coordinates of the beam (hence positional shifts are not visible here). However, inconsistent with the nomenclature in other parts of this text, *x* refers to the vertical and *y* to the horizontal direction (bending plane) in the 2D plots.

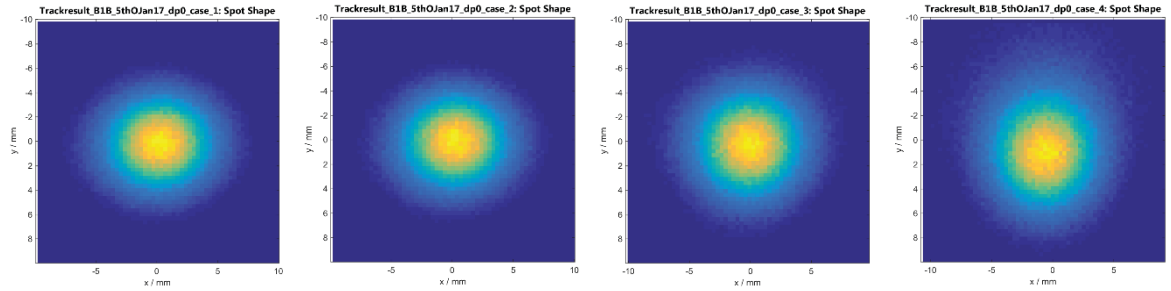

Figure 6.84 (-6 % dp/p): *Spot profiles of beams with a mean relative momentum of -6 %. From left to right the 2D plots of cases 1, 2, 3 and 4 are shown. The colors scale with the number of scored protons per scoring pixel (0.5 x 0.5 mm²). Dispersion effects for higher momentum spreads (case 4) are apparent.*

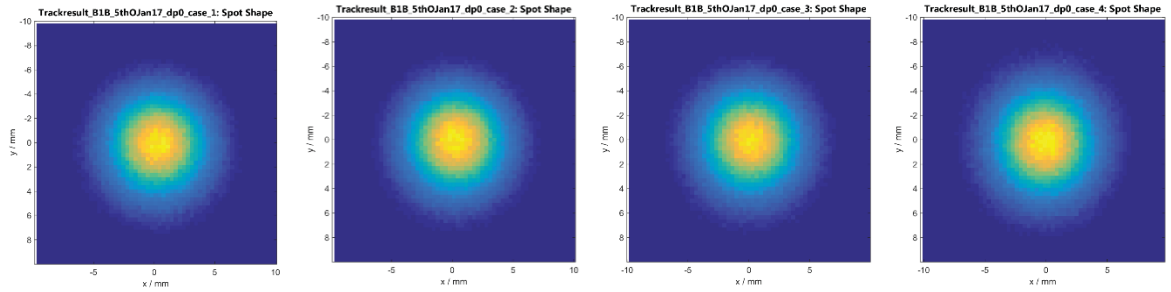

Figure 6.85 (-3.5 % dp/p): *Spot profiles of beams with a mean relative momentum of -3.5 %. From left to right the 2D plots of cases 1, 2, 3 and 4 are shown. The colors scale with the number of scored protons per scoring pixel (0.5 x 0.5 mm²).*

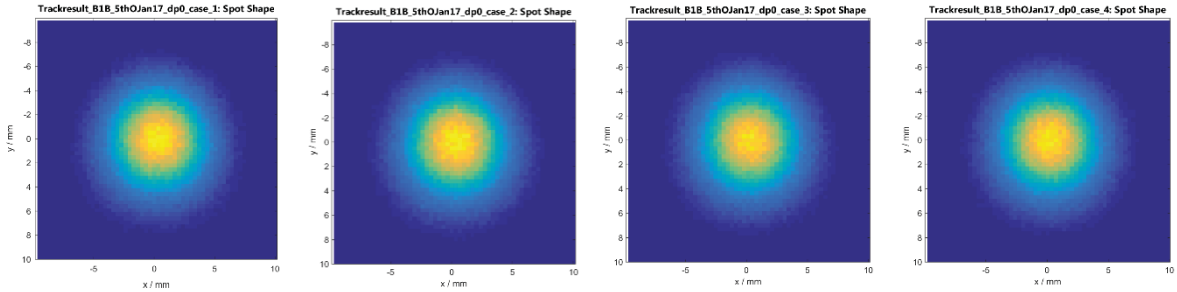

Figure 6.86 (0 % dp/p): *Spot profiles of beams with mean reference momentum (no deviation). From left to right the 2D plots of cases 1, 2, 3 and 4 are shown. The colors scale with the number of scored protons per scoring pixel (0.5 x 0.5 mm²).*

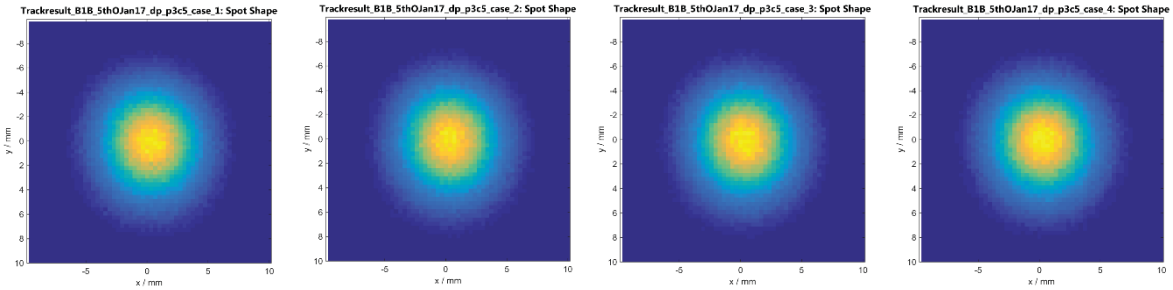

Figure 6.87 (+3.5 % dp/p): *Spot profiles of beams with a mean relative momentum of +3.5 %. From left to right the 2D plots of cases 1, 2, 3 and 4 are shown. The colors scale with the number of scored protons per scoring pixel (0.5 x 0.5 mm²).*

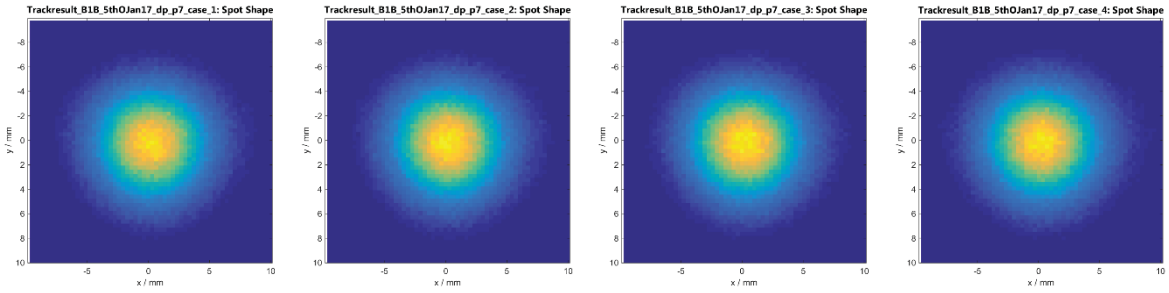

Figure 6.88 (+7 % dp/p): *Spot profiles of beams with a mean relative momentum of +7 %. From left to right the 2D plots of cases 1, 2, 3 and 4 are shown. The colors scale with the number of scored protons per scoring pixel (0.5 x 0.5 mm²).*

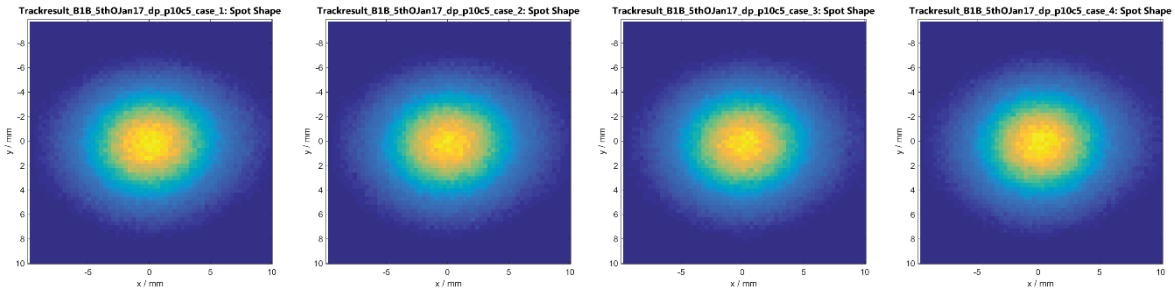

Figure 6.89 (+ 10.5% dp/p): *Spot profiles of beams with a mean relative momentum of +10.5 %. From left to right the 2D plots of cases 1, 2, 3 and 4 are shown. The colors scale with the number of scored protons per scoring pixel (0.5 x 0.5 mm²). Here the effects of the defocusing in y are clearly visible.*

In the acceptance window from -6 to +10.5 % *dp/p* the spots apear round and of similar size. Hence the analysis with Gauss-fits is appropriate. An elliptical shape for high momentum deviation is visible as well as some blurring for larger momentum spreads at low momentum deviations. These observations were expected due to the results of the alpha, beta and dispersion functions (compare **Figure 6.80**, **Figure 6.81** and **Figure 6.82**).

6.5.3 Analysis of beam profiles and positions

Here, the tracking results are summarized in which only the most relevant results are shown in the summary graphs for better clarity. For that only the 5th order results in combination with the applicable momentum spread, classified in cases 1-4 as described in **Table 6.18**, are chosen.

Beam position at isocenter

The beam position at isocenter is not constant for different momenta in the horizontal coordinate *x*. It is a direct consequence of the dispersion function, compare graph **Figure 6.80**. Essential for treatment precision, the position must be corrected even for smallest deviations. While this is possible with help of the deflection magnets of the scanning system, quite involved calibration measurements can be the consequence. This calibration dilemma, which also applies to any other corrections, is detailed below in section 6.6.2.

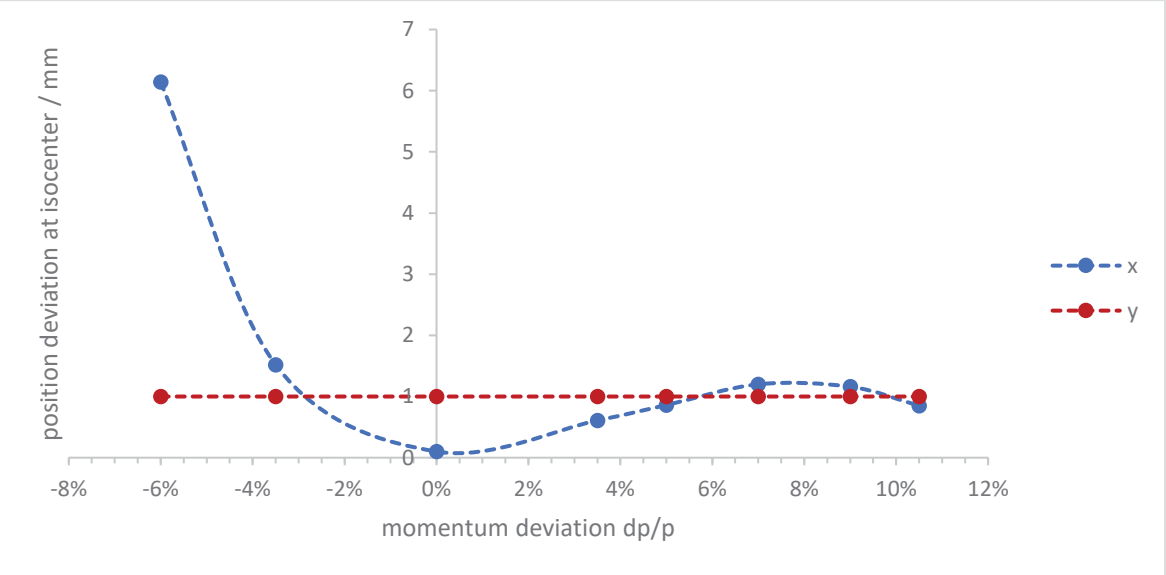

Figure 6.90: *Beam position in the horizontal (x) and vertical direction (y) at isocenter against the momentum deviation (lines to guide the eye).*

The beam position in *y* remains invariant over the complete momentum range at an offset of 1 mm. This offset is an intrinsic effect of a residual solenoid field and not an issue since it can be easily calibrated.

Beam sizes and shapes at isocenter

There are two obvious differences between the horizontal and vertical beam sizes at isocenter. In the horizontal direction the dispersion for negative relative momenta causes a size increase with increasing momentum spread. For all other momenta the size stays close to the initial input value of 2.8 mm, thus 1:1 imaging is achieved.

In the vertical direction the shifted beam waist for high relative momenta causes an increase in the beam size. An effect also visible for -6 % *dp/p*. This is in good agreement with the prediction from the alpha function (see **Figure 6.81**). The fact that the graphs of all four cases are not separating proves that no energy dependence exists – which is expected in the non-bending plane. In general, the beam size is slightly reduced by about 0.2 mm with respect to the input value, resulting in a modest demagnification.

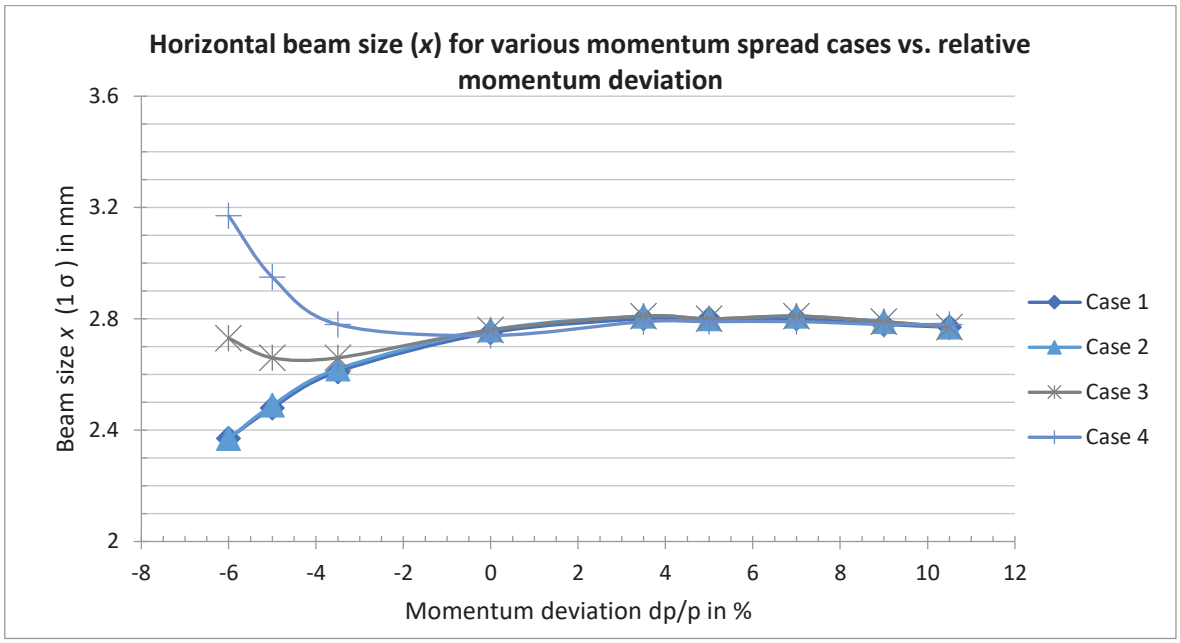

Figure 6.91: *Spot size results (1 RMS of corresponding Gauss-fit) in the horizontal direction at isocenter of the transfer map tracking of various beams of 500,000 protons with different relative momenta and momentum spreads (line to guide the eye).*

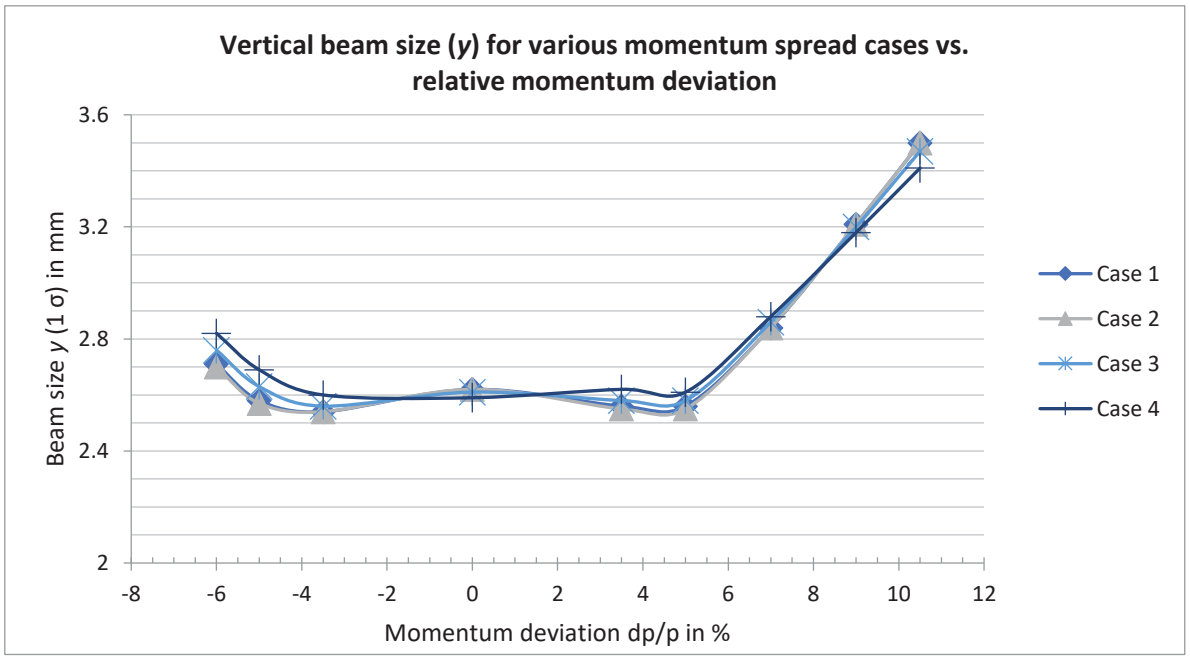

Figure 6.92: *Spot size results (1 RMS of corresponding Gauss-fit) in the horizontal direction at isocenter of the transfer map tracking of various beams of 500,000 protons with different relative momenta and momentum spreads (line to guide the eye).*

The combined spot size was calculated as the arithmetic mean. Generally, some spot size changes must be expected in any proton therapy machine and can have numerous causes. However, the spot size as measured in air is not the size at the target which increases with penetration depth via MCS, diminishing a relative change to the small in-air size of about 4 mm. A typical threshold to which spot size deviations can be tolerated without any corrections is +/- 15 %, which can be deduced e.g. from [156]. Note that all sizes in the following refer to the theoretical *in vacuo* value.

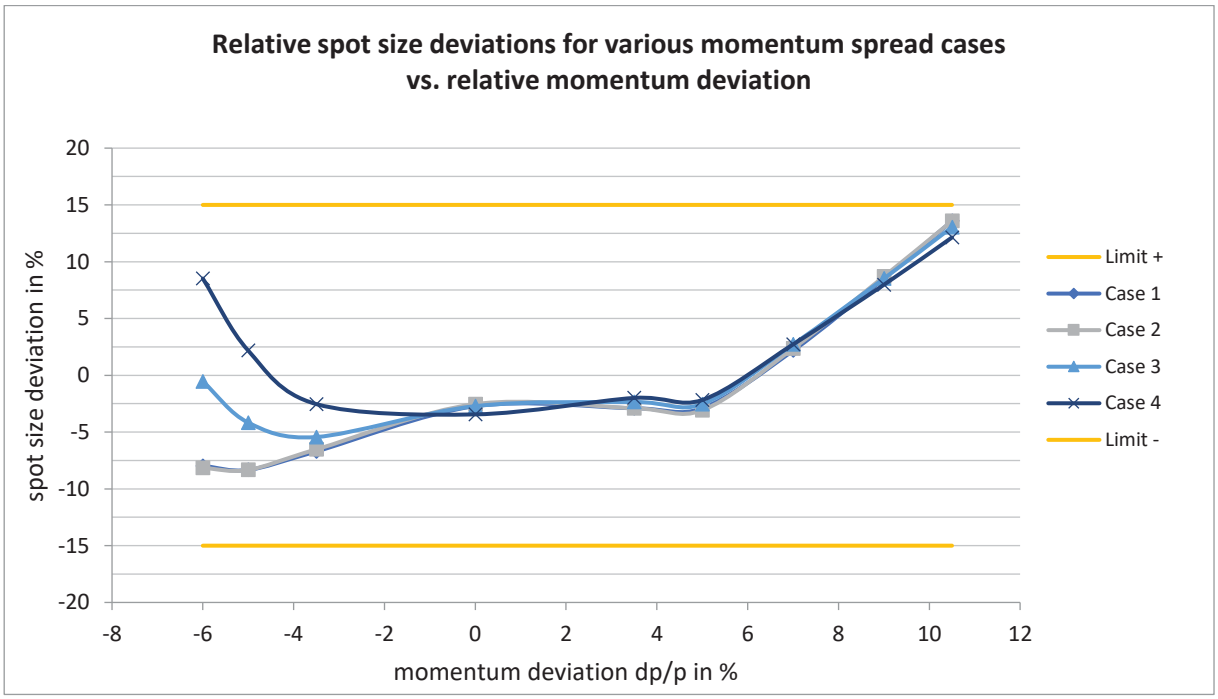

Figure 6.93: *Relative spot sizes at isocenter with respect to the input spot size of 2.8 mm = 1 RMS (line to guide the eye).*

The differences in the horizontal and vertical sizes also lead to elliptical spot shapes as visible in **Figure 6.89** where the defocusing in y for the +10.5 % dp/p momentum clearly influences the vertical size. The asymmetry of elliptical spot shapes is measured as ellipticity e, calculated with the half-axes a and b which reflect the 1 RMS values of the horizontal and vertical size, respectively:

$$e = \frac{2(a-b)}{a+b} \tag{6.1}$$

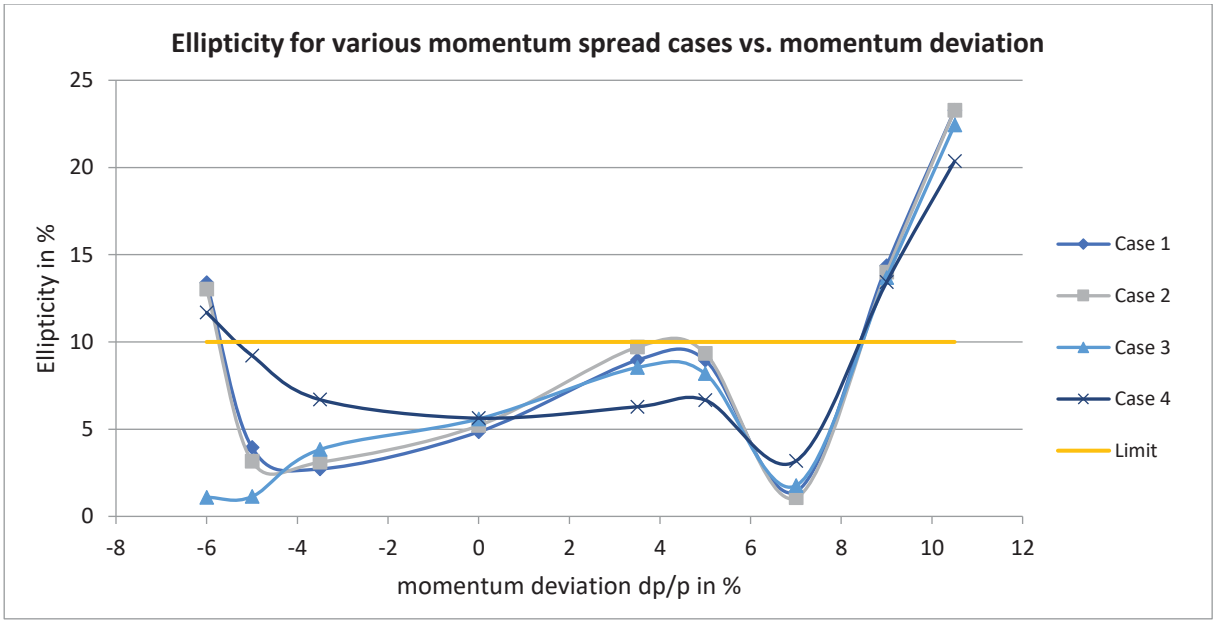

Figure 6.94: *Ellipticity of spot profiles at isocenter for various momentum deviations and momentum spread cases (line to guide the eye). The limit (orange line) is clearly exceed at multiple points. Due to the unequal size even at 0 %, see demagnification in y, ellipticity is non-zero throughout the intermediate momenta.*

Effects of elliptical spots have not thoroughly been investigated yet in context of proton therapy. Chu et al. specified a value of 30% in [98] for energies of 150 MeV and up. In the industry a limit of 10 % measured in air is commonly used in order to still assume round spots for the treatment planning system [157]. However, the topic has not been analyzed in depth yet and some results, e.g. [158], indeed suggest that larger values could be acceptable.

Even larger deviations from the nominal round spot shape could still be overcome if they were reflected in the machine parametrization of the treatment planning system – an option that is less preferred because it is adding another variable of complication (and potential error source).

6.5.4 Discussion of beam sizes and shapes

Results from **Figure 6.94** clearly show that the uncorrected spot shapes would not be compliant with an ellipticity requirement of 10 %. It is again the shifted beam waist in *y* that causes the largest abnormality. Dispersion at low momenta and the unequal size due to the demagnification in *y* at intermediate momenta also cause significant ellipticities. However, one should keep in mind that these results do not take any scattering effects (foils, air) into account – thus some improvement in experimental reality can be expected.

In summary, the spot sizes and shapes are significantly affected by the shifted beam waist in the horizontal plane. The presence of dispersion for negative momentum deviations, on the other hand, is not as pivotal. Consequently, flattening the alpha function over the momentum range is the most crucial correction in need – apart from the mandatory correction of beam position.

6.5.5 Trim coils

Whether the beam properties described above are viewed as satisfying or not, the lack of tunability of the *AGCCT* means an unpractical risk. The five strongly focusing quadrupoles connect in series to the same power supply. After a magnet is built it becomes impossible to change the quadrupole field ratios with respect to each other. Since some deviations in a real magnet will most certainly appear, some degree of tunability of the quadrupoles must be allowed. It is, however, technically challenging to modify the currents in the different quadrupole sections, especially those in the second layer (sandwiched between layers 1 and 3) where access with a differential power supply would be very difficult.

A feasible and less involved way to adjust individual quadrupoles could be the implementation of trim coils inside the bore of the magnet. Since the corrections only require relatively weak fields, these trim coils could be designed as normal conducting air coils (the strong magnetic fields inside the magnet forbid the use of iron-based magnets anyhow). In fact, the warm bore of the magnet is only filled with beam in the center region (section 3) where dispersion becomes largest and the different momenta cause the beam trajectories to travel farthest from the magnet center.

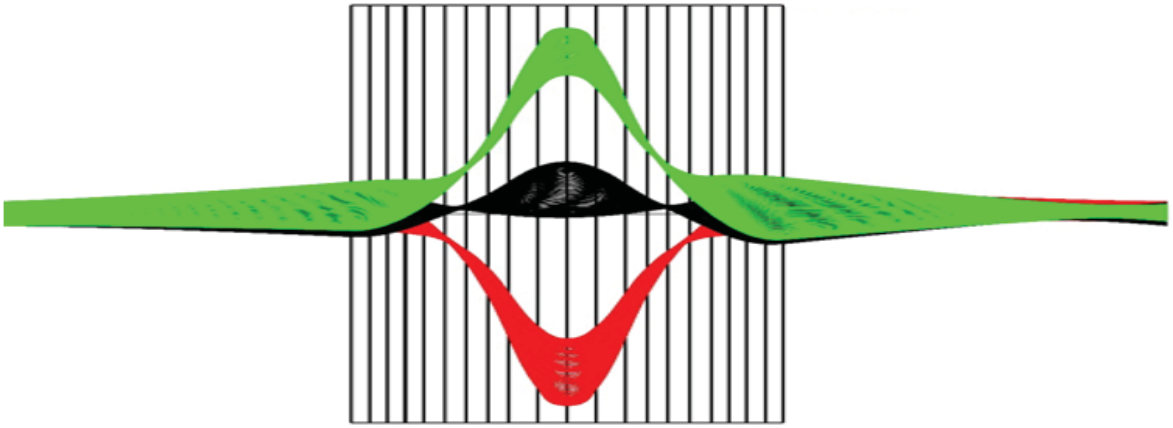

Figure 6.95: *Rays through the AGCCT magnet (straightened). The green, black and red rays show the path of beams with maximal, reference and minimal momentum of the acceptance window. With the transversal motion magnified, the picture shows qualitatively the ample space at entrance and exit in the aperture.*

This leaves room to add quadrupole air coil windings to the inside of the warm bore, so outside the cryostat and at room temperature. The addition of two trim coils allow the adjustment of the quadrupoles in sections 1 and 5, providing another degree of freedom. This is sufficient to achieve beam symmetry at isocenter.

In **Figure 6.96** the results with trim coils for quadrupoles in sections 1 and 5 are displayed for the worst case of dp/p = +10.5 %. This option, addressing the ability to change the focusing properties of the magnet, is needed eminently. The quadrupole currents are optimized for equality of the beta values in *x* and *y*, which leads to current changes of 0.92 % for section 1 and 2.23 % in section 5 which equal to additional field gradients of 0.21 and 0.51 T/m, respectively. This translates to approximately 100 *ampturns* for an air-coil design.

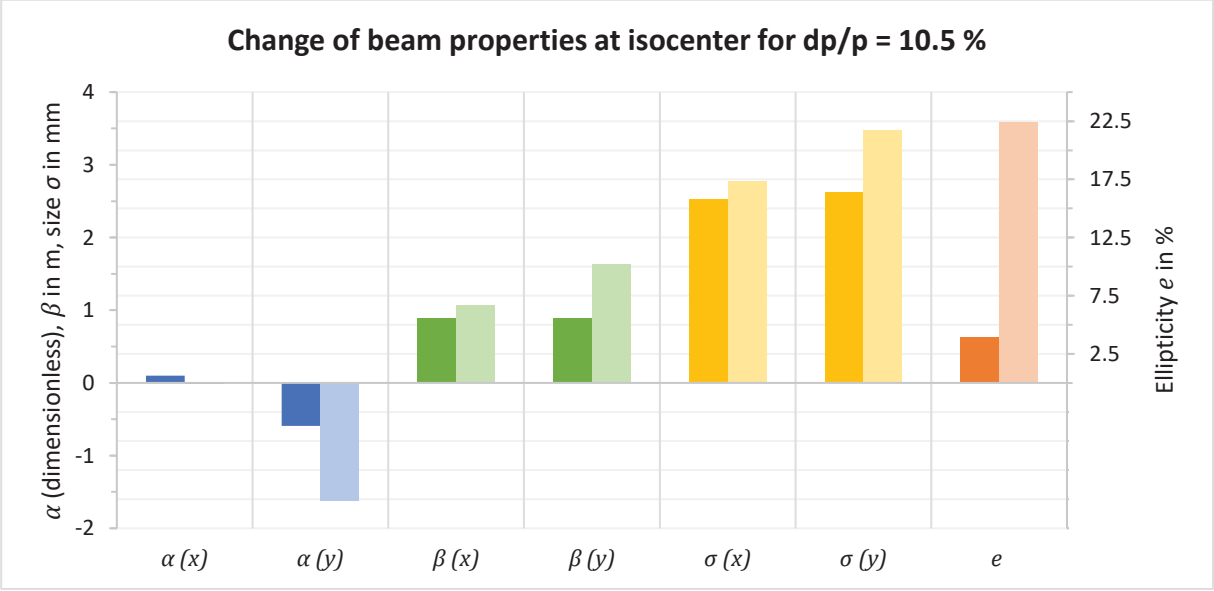

Figure 6.96: *Comparison of the beam properties at isocenter before and after an individual adjustment of the quadrupoles in sections 1 and 5 for the example of a momentum deviation of +10.5 % dp/p. The light and dark colors represent the values before and after the correction, respectively. Ellipticity e is plotted on the secondary axis. All values reflect particle tracking with a 5th order transfer map for case 3 (regards to the momentum spread).*

The results show that with the help of two trim coils the symmetry in the beam can indeed be restored in the linear optics, while some higher order effects are still visible (compare β and σ in **Figure 6.96**). However, the resultant ellipticity is well within the requirement.

If another set of trim coils, complementing sections 2 and 4 was implemented, all five quadrupoles could be tuned individually. In this case, also the correction of dispersion for all momenta becomes possible. **Figure 6.97** shows the relative adjustments to the quadrupoles when tuned to exact $\beta_x = \beta_y$ and first order dispersion to exact zero over the complete momentum range.

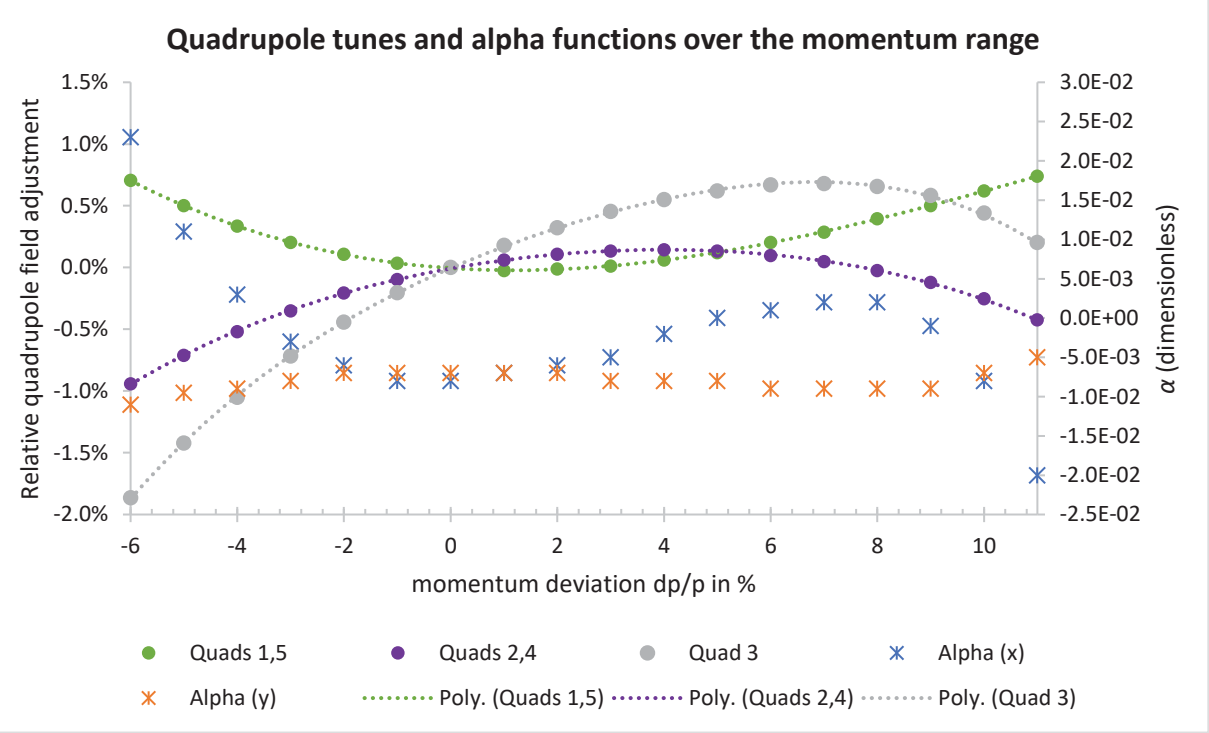

Figure 6.97: *Trim coil settings in relative quadrupole strengths depicted as round dots. The dashed curves show fourth order fit polynomials. The alpha functions (at isocenter) are not tune conditions but indirect results and are approximately zero throughout the momentum range. These settings achieve exact 1:1 imaging (beta) and zero first order dispersion.*

These tunes, which are less than 2 % of the quadrupole strengths are well fitted by a polynomial function, enabling simple parametrization and calibration of a trim coil system. Note that these results correspond to the first order fitting routines described in section 5.3.2 and not higher order tracking. Yet the trim coils achieve a clear improvement compared to an uncorrected *AGCCT* with respect to focusing and dispersion, which are the source of the significant effects shown above.

6.6 Consequences and other aspects of the *AGCCT* results

6.6.1 Realistic quadruple transitions

For the sake of flexibility when working with the *COSY* model the quadrupole transitions have been simplified. In order to verify that this simplification does not affect the beam optics results meaningfully, realistic quadrupole transitions have been implemented in the model and studied. Instead of reversing the sign of the current flow, the actual winding path as designed between the sections is reproduced in the *COSY* model. **Figure 6.98** displays the very small effects from the differences between the simplified and realistic transitions.

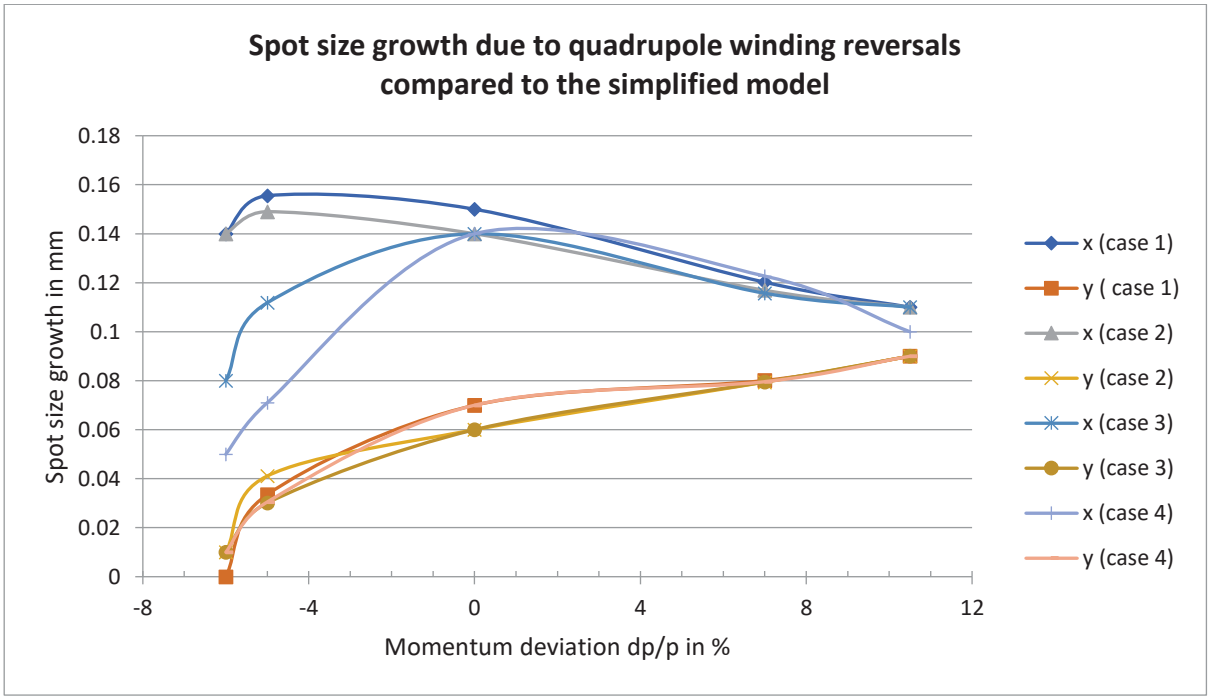

Figure 6.98: *The winding reversal between the quadrupole sections cause a spot size growth at isocenter compared to the values of the simplified model. The blue/grey colors and the orange/brown colors show the size in x and y, respectively, for the various momentum spreads over the acceptance window (lines to guide the eye).*

The size growth in either plane is insignificant, thus the effects on the shapes will be insignificant as well. Hence the conclusion is safe that the simplification used in the *COSY realistic model* is accurate.

6.6.2 Calibrations

Some effects unveiled in the tracking studies of the previous sections require active corrections, namely the beam position in *x* and the focal length in *y*. To both problems, solutions which are technically not too difficult are mentioned (steering and focusing corrections), but the calibration of the *AGCCT* system could become very bothersome as a consequence.

Recalling the first problem statement of the motivation of this study, an important goal in the development of compact proton therapy systems is their affordability with the noble purpose of enhancing the access to this form of treatment worldwide and, of course, the more economic purpose of being more successful in the market. But, with higher volumes also the commissioning effort moves more into focus. Hence, simplicity, automation and ease of handling become essential design goals. Activities in the commissioning of facilities,

particularly activities that require beam time – such as the here discussed calibrations – are costly in the financial sense as well as in time. Due to the many prerequisites, readily commissioned accelerator and degrader for example, they occur at the end of the I&C[21] phase and cannot be performed in parallel to other tasks. Also, although depending on the tools and level of automation, they call for expert staff to perform measurements and calibrate the system. These efforts must not be underestimated thus attention early in the design phase is mandatory.

Calibration of the beam position

The number of required measurements is by far larger than a few points in the curve plotted above in **Figure 6.90**. Even though the positional shifts remain constant for all reference momenta, the correction by the deflection magnet still is energy dependent. Furthermore, there is the complication that one treatment energy can be equivalent to multiple points on the relative scale of the momentum bandwidth of the magnet. This means that one energy, say 160 MeV, can correspond to any relative momentum deviation on the spectrum – depending on the magnet set point or reference energy, respectively, as shown in **Figure 6.99**.

The number of possibilities for the relative momentum deviation of a given energy depends on the number of set points for the magnet that are allowed. For sure one can choose to reduce this number to reduce the calibration effort but that in turn can affect the ramp speed. Minimizing complexity and maximizing the speed of treatment are contradicting in this case. The ramping problem is detailed in the following text.

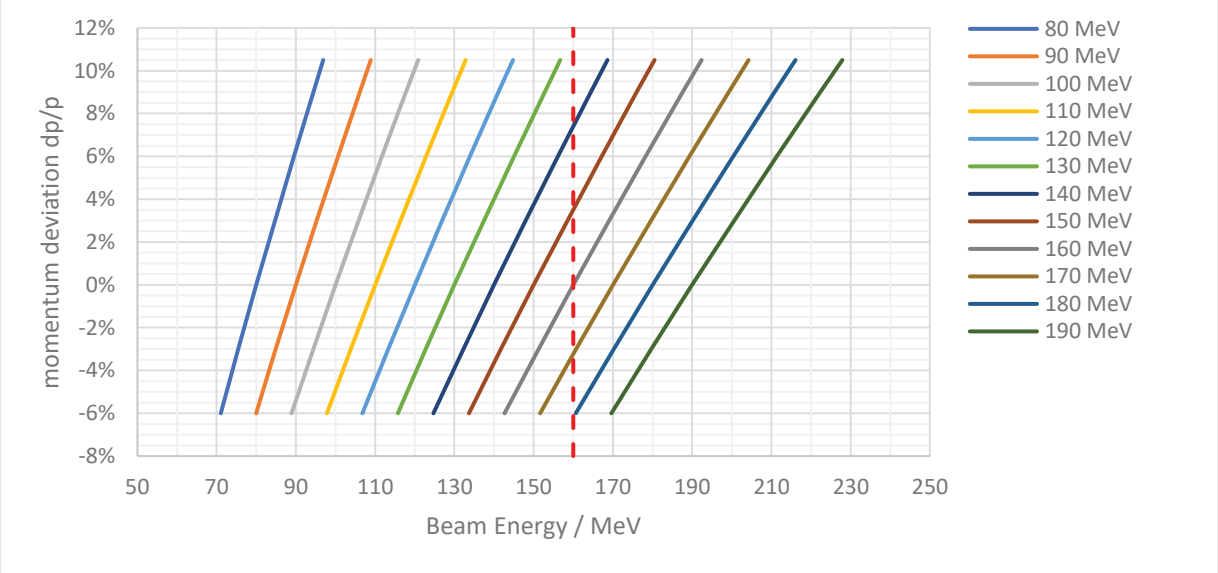

Figure 6.99: *Ranges of the momentum acceptance of the AGCCT transformed into beam energy. Displayed are a number of magnet settings, differing and denoted by their set energy. The red dashed vertical line indicates an exemplary beam energy of 160 MeV. A 160 MeV beam can be applied by numerous different magnet settings – in the graph by five different settings.*

6.6.3 Consequences of the momentum acceptance and ramping strategies

As pointed out many times in this text ramping is the big problem of superconducting magnets. In fact, the large momentum acceptance of the *AGCCT* is designed with the goal to reduce the required ramping and at the same time facilitate advanced treatment methods. The results above

[21] Installation and Commissioning

show that – at least for the chosen design with 170 mm warm bore – this momentum acceptance is limited to about 16.5 % *dp/p*. The 16.5 % are still quite extraordinary for a proton therapy beamline, unreachable with *global achromats* and surpassing the ~ 3% *dp/p* of the double bend achromat of the ProNova SC360 gantry [39]. Translating the momentum acceptance into ranges for SOBPs that can be applied, the in-depth coverage is still significant. As detailed in the analysis in **Figure 6.100** numerous cases can work without ramping, while the majority still require the magnet to ramp at least once during the beam application of a treatment field.

Momentum acceptance in terms of ranges in water

The momentum acceptance sets a window for energies that can be transported with a corresponding magnet setting. This window again translates to ranges in water that can be reached with the respective beam energies. The range coverage in water, however, decreases significantly for lower energies: For one, the momentum acceptance is a relative value, depending on the reference momentum. Thus, with lower reference momenta the absolute momentum acceptance becomes smaller (visible for example in **Figure 6.99**). Secondly, the energy-range-relation is non-linear, so that the range of protons falls disproportionally with reduced energy [166], as can be seen in **Figure 2.10** where the range in water is plotted against energy.

For the *AGCCT* this means: the deeper the target, the larger the SOBP that can be applied within the momentum acceptance. The difference for shallow and deep targets is quite large. On the low side the range coverage is only 3 cm while at highest energies it is extended to over 12 cm.

***Table 6.19**: Energy dependent range coverage.*

Magnet set point	Lowest and highest energies of acceptance window	Shortest and longest range in water	Range coverage in water: maximal applicable SOBP
78.9 MeV	70 – 95 MeV	4.1 – 7.1 cm	3 cm
106.9 MeV	95 – 129 MeV	7.1 – 12.1 cm	5 cm
144.9 MeV	129 – 174 MeV	12.1 – 20.4 cm	8.3 cm
185.0 MeV	165 – 221 MeV	18.6 – 30.8 cm	12.2 cm

A detailed study of the range requirements for such systems has been carried out using range data of nearly 28 thousand proton treatment fields as clinically applied at CPTC and MPTC[22]. The number of the corresponding SOBP ranges that would require magnet ramping if treated with an *AGCCT* based system, is analyzed and summarized in **Figure 6.100**.

[22] CPTC: California Proton Therapy Center (previously Scripps) in San Diego, MPTC: Maryland Proton Therapy Center in Baltimore. Both treatment centers are equipped with the Varian ProBeam® system.

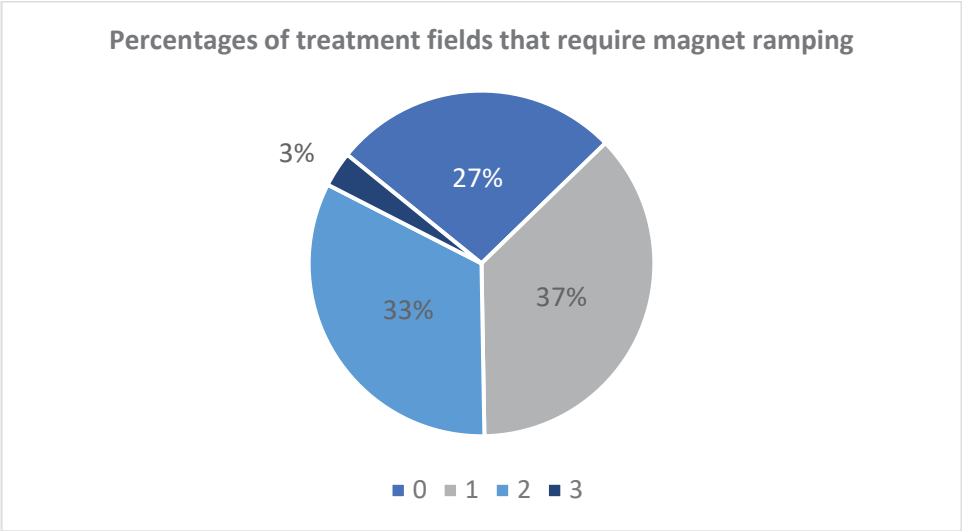

Figure 6.100: *Chart displaying the percentages of clinical treatment fields sorted by the number of magnet ramps they require. 27 % of all fields can be applied with a constant magnet setting. The majority of 73 %, however, need one, two or three magnet ramps (1: 37 %, 2: 33 %, 3: 3 %). The total number of analyzed clinical treatment fields is 27,907.*

The result of the SOBP analysis is clear: the momentum acceptance of the *AGCCT* is not large enough to reduce the ramp speed as doing so would increase treatment durations. Consequently, the magnet ramping calls for a more detailed study.

Ramping strategies

There are four possibilities to perform the ramping of the magnet in order to extend the range coverage.

1) Fixed segments:
One can organize the clinical energy spectrum (70 to 220 MeV) in a minimal number of segments that can be covered with the momentum acceptance of the *AGCCT*, for the discussed design these are four different settings, shown in **Figure 6.101.**

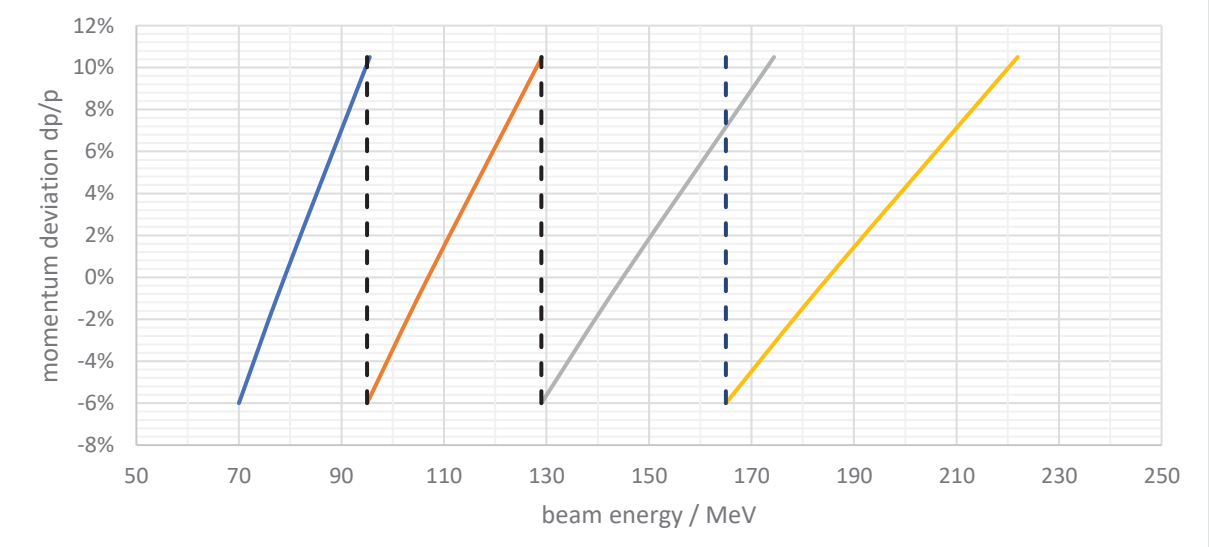

Figure 6.101: *Energy coverage of four different set points for the AGCCT. With four different settings for the magnet the whole clinical energy range from 70 to 220 MeV can be covered. Each line indicates which energies correlate to which momentum deviation of the momentum acceptance (-6 to +10.5 % dp/p) window. The blue, orange, grey and yellow line refer to magnet set points of 78.9, 106.9, 144.9 and 185 MeV, respectively. The dashed vertical lines indicate the conjunction of the setting where ramping has to take place. The clinical energy range is not exactly fitting to the four acceptance windows thus some overlap remains, which is here between the grey and yellow line but could be moved anywhere else as well.*

2) Field adjusted set points:
   Instead of fixed segments, the setpoint of the magnet can be set such that the highest energy (the first treatment layer) of the treatment field corresponds to the highest energy of the energy acceptance. That way the number of ramps for treatments is minimized. This ramp strategy was applied within the analysis of the clinical data above. The question here is, how many setpoints should be allowed. Even a relatively coarse resolution of 5 MeV steps leads to 22 different setpoints, which would require a corresponding commissioning effort.

3) Dead-time ramping:
   Even within the acceptance window a layer switch takes some duration greater than zero. The degrader, and possibly other elements have to be adjusted. A conservative estimate would be that the beam will be turned off for 100 ms during the layer switch. To make effective use of the dead time, one could ramp the magnet in short ramps during the layer switch. Thereby the effective momentum acceptance window can be extended.

4) On-the fly ramping:
   To maximize the complexity of the system but avoid treatment delays in between fields, one could – at least theoretically – continuously ramp the magnet during beam application. However, with layer switch times as short as 100 ms and a high performing, i.e. high dose rate, beamline, as many as twenty layers could be irradiated under 10 maybe even under 6 seconds. At least the order of 10 to 15 seconds should be met, a reasonable time for a breath-hold treatment method. But, at this speed, the required ramp rate for the "on-the-fly-approach" becomes very high. Here is an estimate for a 20 layer irradiation in 10 seconds:

$$\frac{dB}{dt} \approx \frac{\Delta B(20\ layers)}{10\ \mathrm{s}} \approx 2.2\ \% \frac{B_{max}}{\mathrm{s}}, \tag{6.2}$$

   where B_{max} is the maximal field of the magnet according to the highest setpoint and ΔB the difference in field that is required to cover the respective in-depth range of 20 irradiated layers. It is not very likely the described method will prove to be feasible in a real system. Apart from the high ramp rate, on-the-fly corrections for the beam optics are technically very challenging and could easily compromise safety as well as the economic goals for a compact proton therapy machine.

The following graph shows the schematics of the four discussed ramping strategies for an 18-layer example. It shows that strategy no. 3 with dead-time ramping yields an extension of about 2 layers in depth. Overall, none of the four concepts can overcome the fact that the momentum acceptance of the *AGCCT* is not large enough to effectively alleviate ramping needs in a clinical use case.

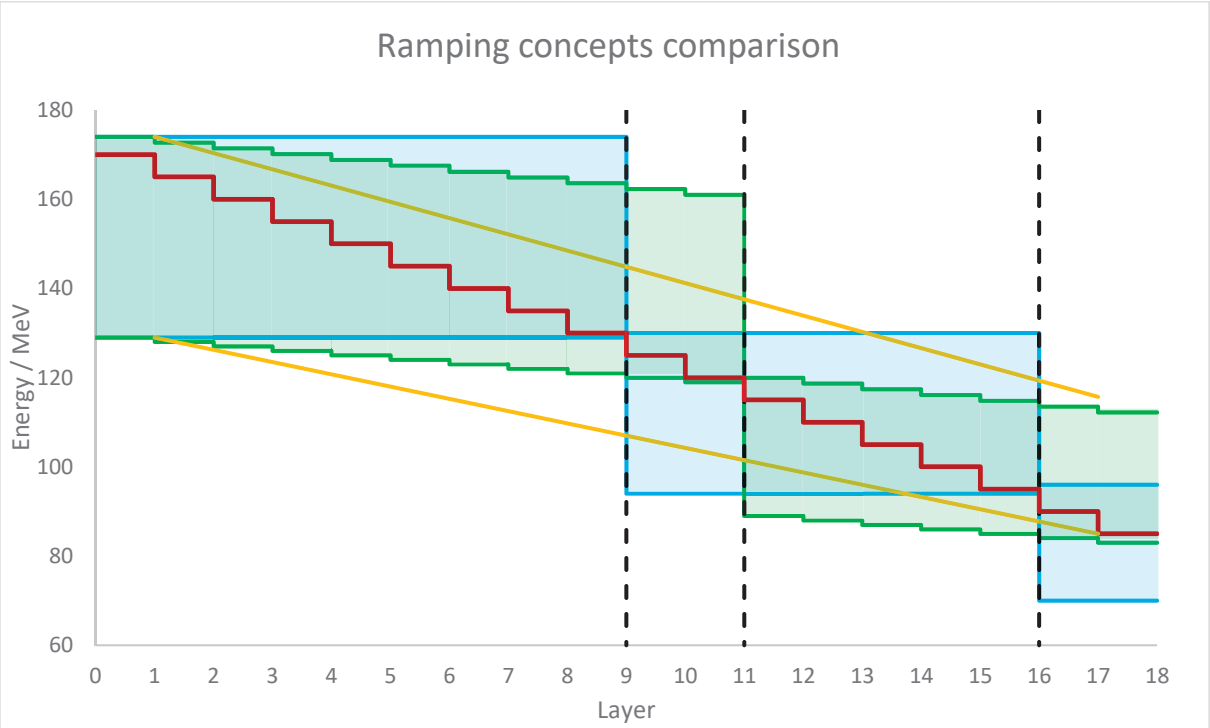

Figure 6.102: *Schematic comparison of ramping schemes: the "field adjusted setpoints" (option 2), the "dead-time ramping" (option 3) and "on-the-fly ramping" (option 4). The blue area displays the second option where the magnetic field is held constant until ramping becomes necessary when a next layer would be outside the acceptance window. The green area shows the acceptance window of the third option where the magnetic field is adjusted in small steps during each layer switch, hereby making use of approximate 100 ms dead-time between layers. The yellow lines indicate the upper and lower limit of the acceptance window if it was continuously ramped as discussed in option 4 ("on-the-fly slow ramping"). The red line indicates the beam energy for each layer, the horizontal steps shall show the time duration of beam application. Here the red step function is only exemplary. Both, layer duration and energy steps between layers are typically varying. The black dashed vertical lines indicate where magnet ramping (in a larger step) would be required which would lead to a respective time delay in the treatment.*

6.6.4 Other ramping, heat capacity and cool-off times

Besides the ramping during treatment fields, the magnet needs to ramp in other events as well. When a treatment plan is loaded, which is minutes before actual beam application, the magnet has ample time to adjust the field to prepare for the first treatment field. Here a rather slow ramp rate can be allowed. In between treatment fields (for the same patient), when the gantry is rotated, the magnet may need to ramp again. If no x-ray imaging is performed at this time, a time interval of approximately 30 seconds can be assumed. This leads to higher ramp rates up to a worst-case of

$$\frac{dB}{dt} = \frac{B_{max} - B_{min}}{t_{field\ change}} \approx 1.2\ \% \frac{B_{max}}{\mathrm{s}} \tag{6.3}$$

where B_{max} and B_{min} are the maximal and minimal set points of the magnet, respectively.

One advantage of the air-coil design of the *AGCCT* is the absence of any hysteresis effect. This allows ramping upwards or downwards without completing full ramp cycles. As such, in contrary to the conventional methods, not all fields have to start with the highest energy. One could just as well start at the bottom and move layer-by-layer upwards. Unfortunately, this does not change the worst-case calculation above, but it adds some flexibility and can thereby reduce the likelihood of larger ramps in between treatment fields.

During magnet ramping heat loads occur in the cold mass as outlined in section 4.4.3. However, they do not necessarily have to be matched by the cooling power immediately. Whenever the magnet is not ramping cool-off times apply. The heat capacity of the cold mass works as a buffer and only on an average over a larger time scale the cooling power should be greater than the heat loads (stemming from both, ramping as well as static heat loads).

6.6.5 Conclusions on ramping

Apparently, optimizing the ramping of the *AGCCT* is a complicated topic. Since the cryostat of the *AGGCT* is not fully designed at this stage and results from the prototype tests could cause changes to the coldmass structure which again influences heat capacity and possible heat flux barriers, it is futile to go too deep into the ramping optimization at this point. One conclusion that will remain valid is that the beam optics results prove ramping will still be required for the *AGCCT* and that ramp rates of at least 1 % B_{max}/s must be expected. Note that according to equations (4.4) and (4.5) the thermal margins design and the cooling needs of the magnet are very sensitive to the ramp rate as it contributes quadratically to the losses in the conductor.

6.7 Results and discussion of the *fixed-field magnet*

6.7.1 General results and optics feasibility

The following results of high order tracking studies demonstrate that the *fixed-field magnet* indeed transports proton beams over the clinical energy range without the need to adjust the magnetic field produced by the superconducting magnet. How far the protons travel inside the magnet along the transversal coordinate *x* (see graph below) depends on their energy. The integrated gradients of the fields, which differ for each energy, determine the focusing / defocusing property of the 2D magnet. Together with the quadrupole triplet clinically acceptable beam properties at isocenter are achieved, proving the feasibility of the concept from the optics perspective. The similarity in field profile between the found solution and the MAMI magnet is shown in **Figure 6.103**. Note that a decisive difference is that the MAMI magnet generates and shapes the field with an iron-dominated magnet limited to approximately 1.5 Tesla whereas the influence of the iron on magnetic field shape in the superconducting *fixed-field magnet* is much reduced, yet essential.

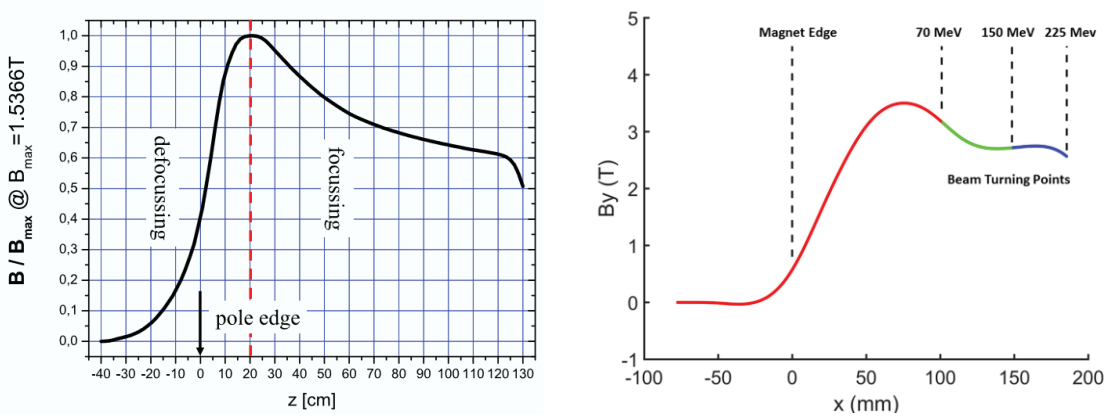

Figure 6.103: *A comparison of the field profiles of the MAMI 90° bend (left) [159] and the studied fixed-field 77.5° bend (right). Note that the MAMI bend has a peak field of approx. 1.5 Tesla. There, no return flux is in the beam entrance area (negative z). In the plot for the fixed-field magnet, return points (along x) for various proton energies are displayed.*

6.7.2 Results and analysis of tracking studies

The *fixed-field magnet* has been characterized and analyzed with 7th order tracking studies over the entire energy range from 70 to 225 MeV. As an input beam a Gaussian distribution with an emittance of 30 π mm mrad was used, comparable to the *AGCCT* study. The initial beam size was chosen as 3 mm for 1 RMS.

Table 6.20: *Momentum spreads of the beam energies used for the tracking.*

Beam energy	70 – 110 MeV	120 – 170 MeV	180 – 225 MeV
Case number	4	3	2
Momentum spread (1 RMS)	2 % *dp/p*	1 % *dp/p*	0.5 % *dp/p*

Analogously to the *AGCCT*, tracking was performed with different momentum spreads in the beam, grouped into three different cases. However, since the *fixed-field magnet* covers the complete range of energies, the tracked beam momenta / energies are not described with relative but with absolute values. Hence to each tracked beam energy a single case of momentum spread

can be applied, which is chosen, conservatively, based on the estimates for the respective degraded beam energy according to the data discussed in chapter 6.3. The table above shows the grouping and the chosen case of momentum spread for the various energies.

While the magnetic setting of the superconducting bends remains static, the fields in the normal conducting quadrupoles are adjusted for each energy.

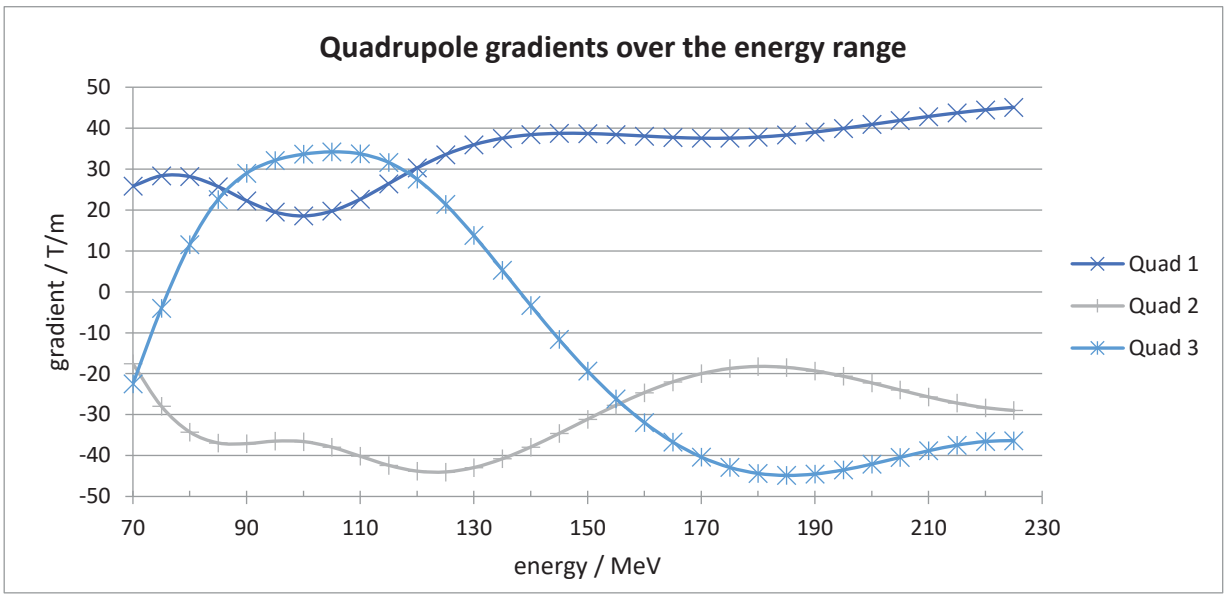

Figure 6.104: *Field gradients of the three quadrupoles adjusted to the beam energy (lines to guide the eye). The same fields are used for both quadrupole triplets.*

The ramp curves for the quadrupoles are easy to parametrize despite the fact that they are quite unconventional – they are not only not *Broh* scaled but also not monotone with energy. The field gradients in the quadrupoles of up to 45 T/m are actually relatively high – a result of the length and aperture optimization. As shown above in section 6.2.2 a half-aperture of 18 mm is sufficient. This leads to pole tip fields of ~ 0.8 T which is technically conservative for normal conducting quadrupoles.

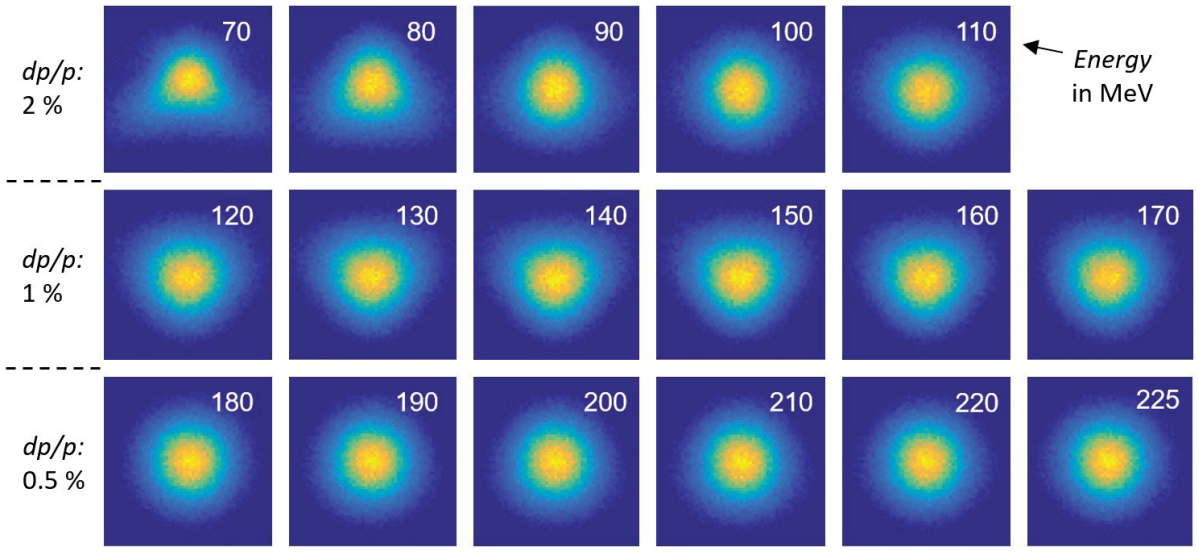

Figure 6.105: *Beam shapes at isocenter in the transversal x-y-plane grouped according to their momentum spread. The number in the upper right corner of each image is the beam energy in MeV. Some triangular shape distortions are visible for low and intermediate energies. The size of the spots remains constant over the energy range.*

The tracking in 7th order was performed in 10 MeV steps with 200,000 particles per tracked beam and initial beam parameters as listed above. For selected cases tracking in 9th order was done to verify the absence of aberrations of these high orders. Resultant spot shapes at isocenter are displayed in **Figure 6.105**. In this figure some spot shape deviations are apparent. However, the 2D images do not reveal the number of protons in the light blue halos.

Here the worst cases for 70 and 150 MeV are shown in 3D plots.

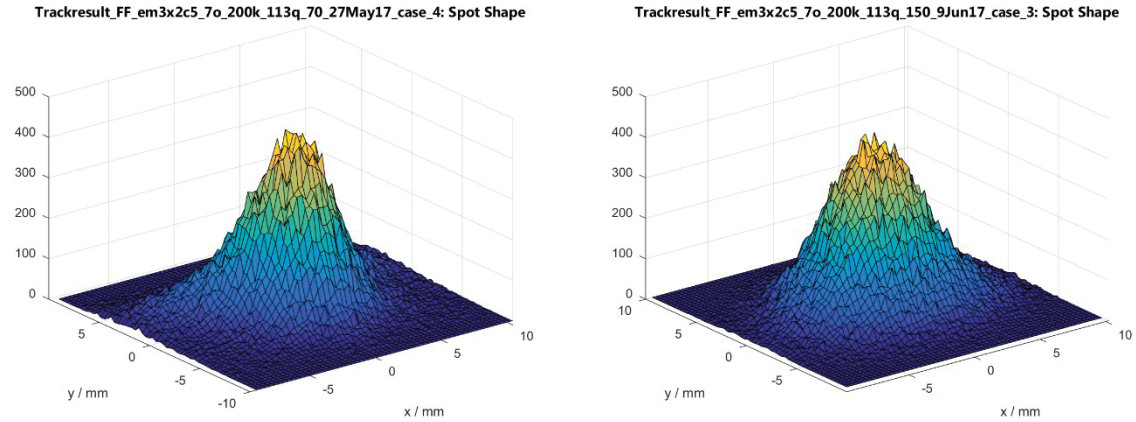

Figure 6.106: *3D plots of 200,000 tracked protons at isocenter with mean energy of 70 MeV (left) and 150 MeV (right). The numbers of protons per pixel (0.5 x 0.5 mm²) are counted on the vertical z-axis.*

At least the 70 MeV case shows triangular tails of relevance. The Gauss conformity along the *x* and *y* axes is also checked by a comparison of the one-dimensional beam profiles to the Gauss-fit. These profiles are histograms in which all particles are projected by the respective coordinate in bins (instead of only using a single binned line through the beam center).

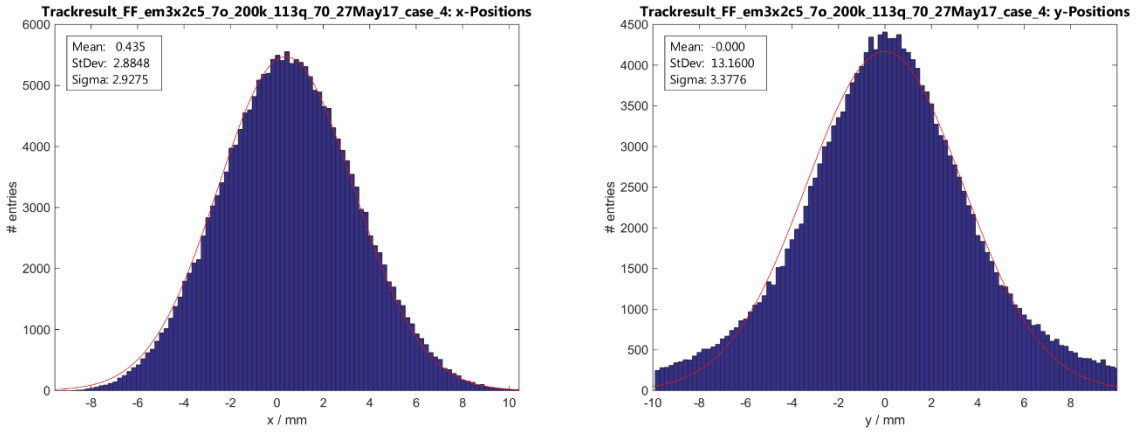

Figure 6.107: *Histograms of the one-dimensional profiles in x and y for a 70 MeV beam (200,000 tracked protons) at isocenter with corresponding momentum spread case 4. All protons are projected on the respective axis and counted in bins of 0.2 mm.*

Indeed, the histogram for coordinate *y* shows some tails of the distribution in the 5-10 % level which leads to a wider Gauss function. Since the shape of the profile is still close to a Gaussian, the change in the RMS value of the fit function does reflect the broadening of the beam by the tails. For the purpose of this study this method is adequate. If, however, the system is brought into clinical use, the deviations of shapes with respect to the Gaussian function should be evaluated more closely, e.g. as suggested in [160].

The tracked beams are analyzed at isocenter analogously to the *AGCCT* results.

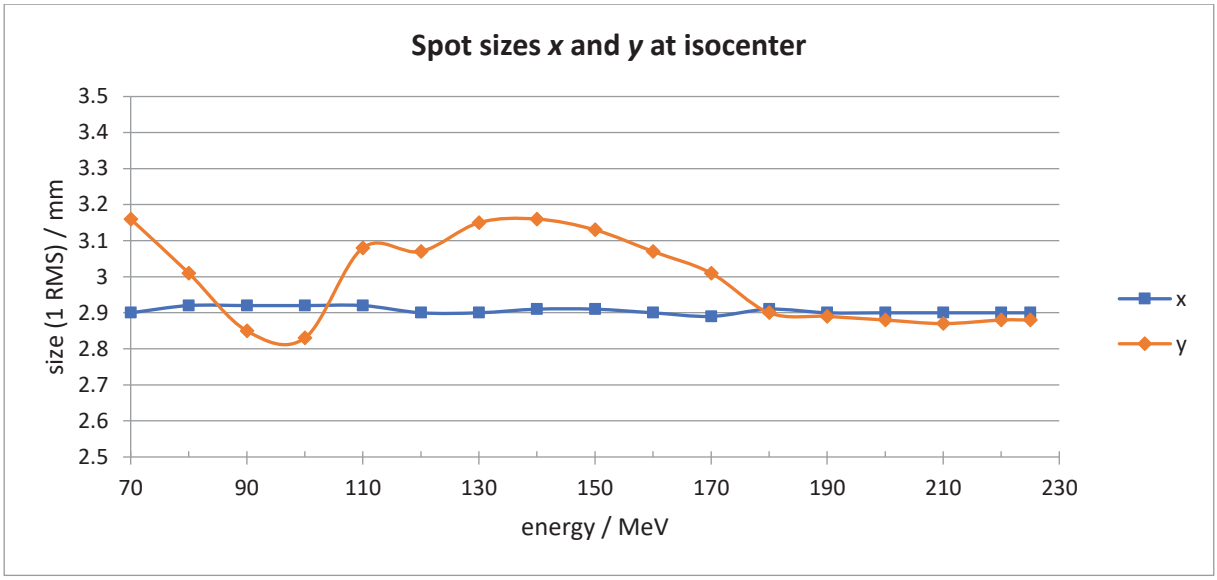

Figure 6.108: *1 RMS values of the one-dimensional beam profiles in x and y (lines to guide the eye). Some deviations in y for low and intermediate energies occur. For each energy the result from 7th order tracking with the momentum spread case according to the grouping from table 6.20 is depicted.*

The graph (**Figure 6.108**) confirms the qualitative impression from the 2D spot images. Low and intermediate energies exhibit some asymmetry. However, a difference of less than 0.3 mm is not too significant, which can be seen in the ellipticity plot (see **Figure 6.110** below).

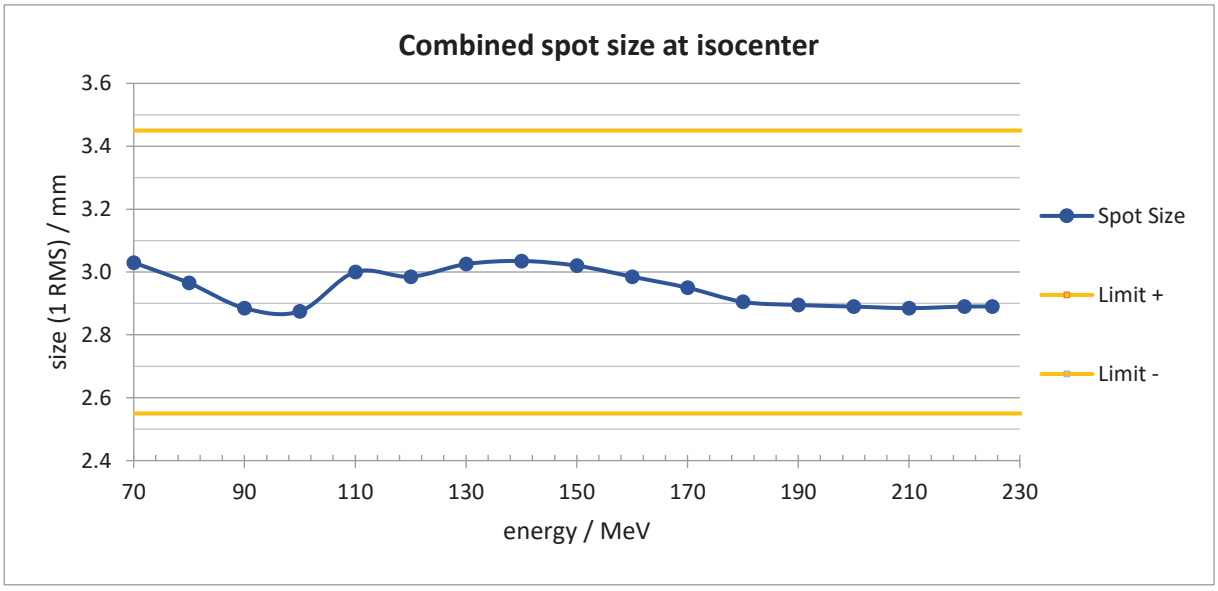

Figure 6.109: *Combined spot sizes at isocenter calculated as the arithmetic mean of the data from figure 6.106 (line to guide the eye). The yellow lines depict the +/- 15 % limit with respect to 3 mm.*

In vacuum the spot sizes stay close to the initial 3 mm across the entire energy range and far clear from the +/- 15 % requirement. Of course, when measured in air, an energy dependence of the spot size will be induced.

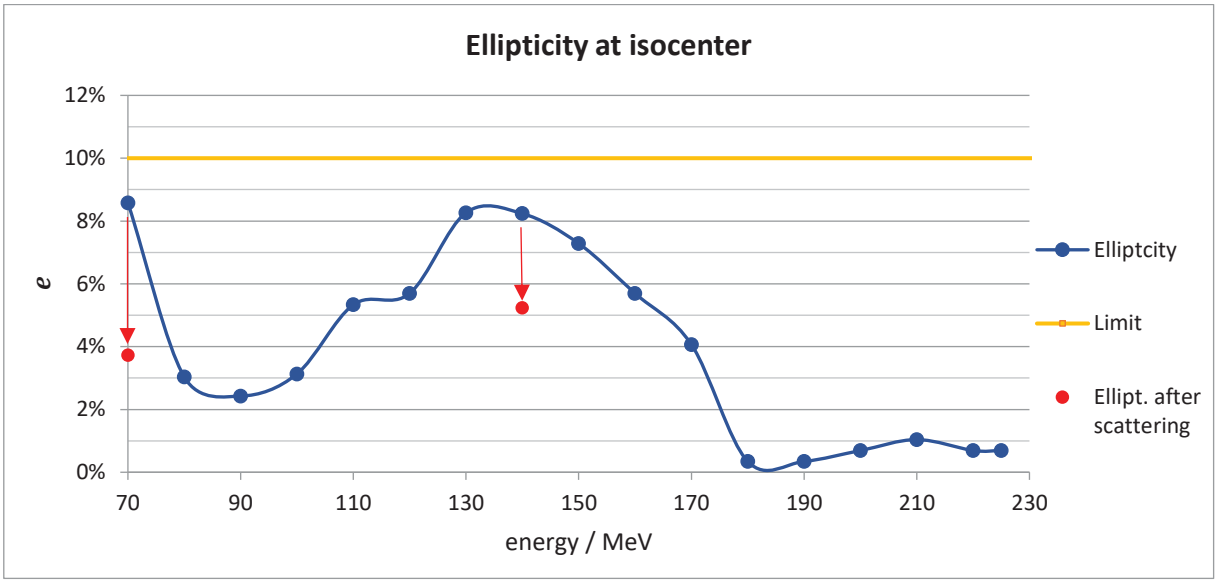

Figure 6.110: *Ellipticity of the beam at isocenter against beam energy (line to guide the eye) calculated from histograms (example in figure 6.105). The red arrows indicate the reduced ellipticity due to scattering effects in air and foils. The yellow line marks the 10 % requirement.*

Spot shape asymmetry is not an issue for the *fixed-field magnet* as demonstrated in the graph above. When scattering effects are taken into account (like they would occur in measurements with a real system), the ellipticity values stay well below the 10 % limit.

6.7.3 Beam path separation and energy selection

The separation of the parallel beams between the two superconducting bends is about 30 cm, so relatively large, as also shown together with the magnet model in **Figure 4.56** displaying the trajectories of two protons of 70 and 225 MeV. The drawback is that a wide vacuum chamber bridging the two cryostats is needed while the advantage is that for individual proton beams containing an energy spread, all particles will be nicely sorted by energy. Hence this makes a suitable location for an energy slit, filtering out unwanted energies. Hence, for the *fixed-field magnet* the problem of the low energy tail in the beam after degradation can easily be solved (compare chapter 6.3). Additionally, an energy measuring device could enhance certainty in the applied beam energy, addressing the respective requirement from the IEC norm 60601-2-64 [161].

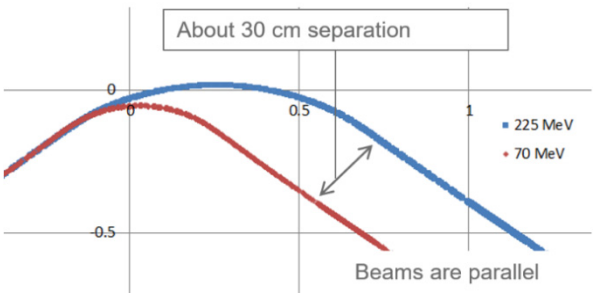

Figure 6.111: *The trajectories of two protons of 70 MeV (red) and 225 MeV (blue) as tracked through the superconducting bend are displayed. The separation of their flight path is approximately 30 cm.*

7 CONCLUSIONS AND OUTLOOK

In this thesis the beam optics of two current developments of achromatic superconducting magnets, the *AGCCT* and the *fixed-field magnet*, for compact proton therapy systems were investigated. Their properties with regards to requirements from clinical operation and industrial production, highlighting the challenges and specialties of the superconducting technology, are critically discussed. Furthermore, an optimized gantry layout is shown which is designed to exploit the strengths of locally achromatic bending magnets.

The initial problem statements that motivated these studies referred to the size and cost of single-room proton therapy systems and the question whether achromatic superconducting bending magnets can help to achieve a reduction of the same. The gantry study demonstrated that beamline optics built around the concept of a local achromat can indeed achieve significant size reductions of single-room systems – especially in length. As an essential enabler of such a concept, superconducting technology allows for the design of combined function magnets with strong fields and gradients as well as large bores. Increasing the momentum acceptance of the achromats alleviates the pain of ramping. Although the study also revealed that only a range coverage of the full treatment SOBP – not reached by the *AGCCT* – or ideally even the coverage of the entire clinical energy range – indeed accomplished by the *fixed-field magnet* – significantly reduces the complexity and cost of the magnet and its cooling.

7.1 Gantry

In summary, the largest benefit of superconducting magnets is not their mere increase in fields. A proton gantry design that simply replaces conventional magnets with superconducting versions is in the end disadvantageous – too little weigh the benefits of reduced bend radii and weight if the cost and complexity consequences are considered. However, with the approach of local achromats new gantry concepts become feasible. Moving the energy degrader onto the gantry spine reduces the length of a compact system by multiple meters (in case of the steep *fixed-field magnet* version even by about 50 %). In addition, the upstream gantry optics, described in chapter 6.1 become trivial and with small apertures, minimized number of elements, and a static-field-dipole cost-effective, too.

7.2 AGCCT

The *AGCCT* in its discussed design, which is currently in the prototyping stage, achieves a momentum acceptance of 16.5 % *dp/p*, limited by its warm aperture of 170 mm. This acceptance window, however, does not cover the majority of treatment SOBPs (only 27 %) as shown with the respective statistics in section 6.6.3.

Further, it is demonstrated that the shift of the vertical beam waist for higher momenta, which causes non-tolerable beam properties at isocenter, can be corrected with a set of trim coils.

These trim coils inside the magnet aperture, ideally four to also correct linear dispersion, additionally provide tunability of the magnet – most likely a very valuable feature for beam commissioning. Other than the vertical focus and some dispersion effects for lower momenta, the spot profiles are universally within typical limits. But the beam position at isocenter is depending on the relative momentum deviation with respect to the magnet set point, which would require corrections with the scanning deflection magnets.

From those pivotal results it can be concluded that the best use of the *AGCCT* concept may not necessarily be for proton therapy. The acceptance window is not quite large enough and the resultant ramping requirements of at least 1 % B_{max}/s become problematic due to the AC losses (see section 4.4.3). A larger bore size would help but the cost implications are not insignificant. Furthermore, its complexity, especially concerning the calibration needs, make it less suited for a standardized low-cost product. For other applications, however, the *AGCCT* offers the advantage of relatively high momentum acceptance for large emittances. It becomes particularly attractive if more time for the ramping or liquid helium cooling can be allowed. The possible addition of the four trim coils also promises excellent beam properties across the acceptance window in a tunable version of the *AGCCT*.

7.3 Fixed-field magnet

The *fixed-field magnet* on the other hand has the potential to mark a breakthrough for the application of superconducting magnets in proton therapy. The results of chapter 6.7 demonstrate the transport of beams from 70 to 225 MeV considering respective momentum spreads with spot shape and size requirements met at isocenter. This makes a unique feature with an immense impact on the conductor and cooling requirements, typically the two dominating cost drivers for superconducting magnets.

The relatively large weight (approx. 5 – 6 tons) and the not too compact size of the double bend assembly can be seen as its disadvantages. Indeed, the quadrupole triplet contributes to the gantry radius, which assumes about 3.5 m, resulting in a similar gantry diameter as the normal conducting Proteus®One for example. Nevertheless, the local achromat with 155° bend angle enables the short and simple gantry system described above, making a very compact system overall. Apart from its compactness, the sketched gantry with the *fixed-field magnet* has a number of appealing advantages. It promises fast treatments with high transmission (thus high dose rates) and approximately 100 ms layer switching, facilitating advanced treatment methods such as volumetric repainting and breath-hold treatments as well as hypofractionation.

Maybe the most important aspect of the *fixed-field magnet* design is the simplicity and lack of novelty in the needed hardware. The magnetic fields, required for the achromatic beam optics, are generated by simple racetrack coils with c-shaped iron yokes and standard normal conducting quadrupoles. The moderate field of 4.7 T at the conductor and the absence of ramping allow for the use of low-cost NbTi conductor. Also, the cryostat and cooling capacity is much relaxed for static-field operation. These properties seem to be well suited for commercial products where cost, robustness, reliability and ease of commissioning and use are high priorities.

7.4 Outlook

The completion and testing of the prototype of the *AGCCT* will add new insight into the magnet and best practices concerning its manufacturing. As mentioned, the beam optics study of its properties does not necessarily recommend its application to proton therapy in its current form. It would, however, be very interesting to investigate the optimization of the bore size depending on the application. For instance, the *AGCCT* may be well suited for an application in carbon beam therapy (where it would have more significant size advantages) or beamlines for laser-plasma accelerated particles where high momentum acceptance optics are crucial.

For the *fixed-field magnet,* which is currently a pure paper study, it is recommended to move forward with building and testing a prototype. Since the underlying magnet technology is well established, a prototype (including cryostat) for tests with beam can be targeted directly.

A final design of the *fixed-field magnet* in combination with a corresponding final adaption of the discussed gantry can potentially make a very attractive proton therapy product leading to improvements in cost and size as well as in speed of beam delivery and energy precision. Furthermore, the proposed system of the simplified gantry and a superconducting magnet with constant field, mean a very significant reduction of complexity and technical fragility.

I personally hope that, one day, the here discussed concepts will contribute to a better access to proton therapy treatments and better cancer care for patients.

REFERENCES

[1] World Health Organization (WHO). URL: www.who.int/cancer/en/ (Jan 2018)

[2] Cancer Research UK. URL: www.cancerresearchuk.org/sites/default/files/cs_report_world.pdf (Jan 2018)

[3] International Agency for Research on Cancer, WHO (2014). *World Cancer Report 2014.* B.W. Stewart and C.P. Wild (editors). Lyon.

[4] M. Plummer, C. de Martel et al. (2016). *Global burden of cancer attributable to infections in 2012: a synthetic analysis.* The Lancet Global Health **4** (9), pp e609-e616.

[5] Robert Koch Institut (2015). *Krebs in Deutschland.* 10th Edition.

[6] J.S. Tobias and D. Hochhauser (2014). *Cancer and its management.* 7th edition. Wiley-Blackwell. West Sussex.

[7] Deutsche Krebsgesellschaft. Onko Internetportal. URL: https://www.krebsgesellschaft.de/onko-internetportal/basis-informationen-krebs/therapieformen/strahlentherapie-bei-krebs.html (Jan 2018)

[8] S. Kumar (2012). *Second Malignant Neoplasms Following Radiotherapy.* Int. J. Environ. Res. Public Health **9** (12), pp 4744-4759.

[9] W.D. Newhauser et al. (2009). *The risk of developing a second cancer after receiving craniospinal proton irradiation.* Phys. Med. Biol. **54** (8), pp 2277-2291.

[10] D. Wang (2015). *A critical appraisal of the clinical utility of proton therapy in oncology.* Figure 1. Dovepress **8**, pp 439-446.

[11] J.J. Wilkens (2007). *Optimierung der Protonentherapie mit KonRad.* Presentation. DKFZ.

[12] T. Kamada (2012). *Clinical evidence of particle beam therapy (carbon).* Int. J. Clin. Oncol. **17** (2), pp 85-88.

[13] Statistics of particle therapy facilities in operation, under construction and planned. URL: https://www.ptcog.ch (Jan 2018)

[14] R.R. Wilson (1946). *Radiological use of fast protons.* Radiology **47**, pp 487-491.

[15] H. Paganetti (2012). *Proton Therapy: History and Rationale.* Proton Therapy Physics, pp 1-6. CRC Press. Boca Raton.

[16] J.M. Slater et al. (1992). *The proton treatment center at Loma Linda University Medical Center: Rationale and description of its development.* Int. J. Radiat. Oncol. Biol. Phys., **22** (2), pp 383-389.

[17] E. Pedroni et al. (1995). *The 200-MeV proton therapy project at the Paul Scherrer Institute: conceptual design and practical realization.* Med. Phys. **22** (1), pp 37-53.

[18] H. Blosser et al (1989). *Medical accelerator projects at Michigan State University,* Proceedings of 13th Particle Accelerator Conference (PAC89), Chicago, pp. 742-746.

[19] J. Kim and H. Blosser (2001). *Optimized Magnet for a 250 MeV Proton Radiotherapy Cyclotron.* Proceedings of the 16th Int. Conf. on Cyclotrons and Their Applications, East Lansing, pp 345-347.

[20] A. Geisler et al. (2004). *Status Report of the ACCEL 250 MeV Medical Cyclotron.* Proceedings of the 17th Int. Conf. on Cyclotrons and Their Applications, Tokyo, pp 178-182.

[21] J.M. Schippers (2016). *Miniaturizing Proton Therapy: A Technical Challenge with Unclear Clinical Impact.* Int. J. Radiat. Oncol. Biol. Phys. **95** (1), pp 149-153.

[22] J. Flanz and T. Bortfeld (2013). *Evolution of Technology to Optimize the Delivery of Proton Therapy: The Third Generation.* Semin. Radiat. Oncol. **23** (2), pp 142-148.

[23] W. Kleeven et al. (2013). *The IBA Superconducting Synchrocyclotron Project S2C2,* Proceedings of 20th Int. Conf. on Cyclotrons and Their Applications, Vancouver.

[24] E. Pearson et al. (2016). *Developments of Cyclotrons for Proton and Particle Therapy.* Particle Radiotherapy, pp 21-35. A. Rath, N. Sahoo (editors). Springer. New Delhi.

[25] URL: www.mevion.com/newsroom/press-releases/mevion-medical-systems-delivers-worlds-first-superconducting (Jan 2018)

[26] U. Raich (2009). *Emittance Measurements.* Presentation at Accelerator Beam Diagnostics Conference USPAS09 in Albuquerque, NM.

[27] W. Wan et al. (2015) *Alternating-gradient canted cosine theta superconducting magnets for future compact proton gantries.* Phys. Rev. Accel. Beams **18** (10), 103501.

[28] S. Caspi et al. (2013), *A superconducting magnet mandrel with minimum symmetry laminations for proton therapy.* Nucl. Instrum. Meth. Phys. Res., Sect. A **719**, pp. 44-49.

[29] A. Gerbershagen et al. (2016). *Novel beam optics concept in a particle therapy gantry utilizing the advantages of superconducting magnets.* Z. Med. Phys. **26**, pp. 224-237.

[30] Accelerator Stewardship Awards. URL: https://science.energy.gov/hep/research/accelerator-stewardship/awards/ (Jan 2018)

[31] K. Greene (2015). *Grants Give Particle Accelerator Technologies a Boost.* Berkeley Lab News Center, URL: http://newscenter.lbl.gov/2015/02/12/grants-give-particle-accelerator-technologies-boost/ (Jan 2018)

[32] L. Brouwer et al. (2017). *Design of an Achromatic Superconducting Magnet for a Proton Therapy Gantry.* IEEE Trans. Appl. Supercond. **27** (4), pp 1-6.

[33] L. Brouwer (2015). *Canted-cosine-theta superconducting accelerator magnets for high energy physics and ion beam cancer therapy.* Dissertation. University of California, Berkeley.

[34] D. De Ruysscher et al. (2015). *Tumour Movement in Proton Therapy: Solutions and Remaining Questions: A Review*. Cancers (Basel) **7** (3), pp 1143-1153.

[35] Y. Zhang et al. (2016). *An evaluation of rescanning technique for liver tumour treatments using a commercial PBS proton therapy system.* Radiotherapy and Oncology **121** (2), pp 281-287.

[36] S.M. Zenklusen, E. Pedroni and D. Meer (2010). *A study on repainting strategies for treating moderately moving targets with proton pencil beam scanning at the new Gantry 2 at PSI.* Phys. Med. Biol. **55**, pp 5103-5121.

[37] J. Dueck et al. (2016). *Robustness of the Voluntary Breath-Hold Approach for the Treatment of Peripheral Lung Tumors Using Hypofractionated Pencil Beam Scanning Proton Therapy.* Radiat. Oncol. **95** (1), pp 534-541.

[38] J.M. Cameron, V. Anferova and T.A. Antaya (2013). *Compact isocentric gantry.* US-Patent 8575563B2.

[39] V. Derenchuk (2014). *The Pronova SC360 Gantry.* Presentation at Modern Hadron Therapy Gantry Developments, Cockcraft Institute, Daresbury.

[40] ProNova SC360 FDA 510(k) clearance.
URL: https://www.businesswire.com/news/home/20161208005859/en/Provision-Healthcare-Receives-FDA-Clearence-ProNova-SC360 (Jan 2018)

[41] ProNova SC360, First patient treatment completed.
URL: https://provisionhealthcare.com/2018/11/15/provision-nashvilles-first-patient-celebrates-treatment-completion-with-victory-bell-ceremony/ (Dec 2018)

[42] D. Trbojevic et al. (2007). *Carbon/proton therapy: A novel gantry design.* Phys. Rev. Accel. Beams **10** (5), 053503.

[43] C. Johnstone et al. (1999). *Fixed Field Circular Accelerator Designs*. Proceedings of PAC99, New York.

[44] D. Trbojevic et al. (2007). *Superconducting non-scaling FFAG gantry for carbon/proton cancer therapy.* Proceedings of PAC07, Albuquerque.

[45] R.J.L. Fenning et al. (2010). *A non-scaling FFAG gantry design for the PAMELA project,* Proceedings of the 1st Int. Particle Accel. Conf. (IPAC2010), Kyoto, pp 4593-4595.

[46] K. Peach et al. (2010). *PAMELA Overview and Status,* Proceedings of IPAC2010, Kyoto, pp 112-114.

[47] M. Galonska et al. (2011). *Commissioning of the carbon beam gantry at the Heidelberg Ion Therapy (HIT) Accelerator.* Proceedings of IPAC2011, San Sebastián, pp 2874ff.

[48] Y. Iwata et al. (2012). *Design of a superconducting rotating gantry for heavy-ion therapy.* Phys. Rev. Accel. Beams **15** (4), 044701.

[49] Y. Iwata et al. (2019) *Development of Carbon-Ion Radiotherapy Facilities at NIRS,* IEEE Trans. Appl. Supercond. **28** (3), (published online Dec 2017).

[50] Toshiba press release (2017). URL: www.toshiba.co.jp/about/press/2017_05/pr1001.htm (Jan 2018).

[51] Y. Iwata (2016). *SC Gantry at NIRS and new developments in SC magnets for gantries & synchrotrons.* Presentation at Academia-Industry Event on Superconductivity for Accelerators for Medical Applications 2016, Madrid.

[52] Photo with courtesy from Universitätsklinikum Heidelberg. URL: https://www.klinikum.uni-heidelberg.de/ShowSingleNews.176.0.html?&no_cache=1&L=0&tx_ttnews%5Btt_news%5D=6428&cHash=6cd0b72e579869713a8c3fa2bba668fd

[53] M. Schaer and J.M. Schippers (2012). *A Design of a Gantry for 350 MeV Protons.* PSI report (unpublished).

[54] H. Owen (2015). *Gantries and Beam Delivery for 350 MeV p+.* Presentation at EUCARD2 Workshop on SC Gantries, Manchester.

[55] E. Oponowicz and H. Owen (2017). *Superconducting gantry design for proton tomography.* Proceedings of IPAC 2017, Copenhagen.

[56] H. Bethe and J. Ashkin (1953). Passage of radiation through matter, in *Experimental Nuclear Physics* (Ed. E. Segrè). Vol.1 part 2, p 253. Wiley. New York.

[57] URL: http://physics.nist.gov/PhysRefData/Star/text/PSTAR.html (Feb 2018)

[58] J. Heese (2013). Private Communication.

[59] W.T. Chu, B.A. Ludewigt and T.R. Renner (1993). *Instrumentation for Treatment of Cancer Using Proton and Light-Ion Beams.* Rev. Scientific Instrumentation **64**, pp 2055-2122.

[60] B. Gottschalk B (2012). Physics of Proton Interactions in Matter, in *Proton Therapy Physics* (Ed. H. Paganetti), pp.19-59, CRC Press, Tayler & Francis, Boca Raton.

[61] L. Hong et al. (1996). *A pencil beam algorithm for proton dose calculations.* Phys. Med. Biol. **41** (8), pp 1305-1320.

[62] C. Calzolaio, S. Sanfilippo et al. (2016), *Preliminary Magnetic Design of a Superconducting Dipole for Future Compact Scanning Gantries for Proton Therapy.* IEEE Trans. Appl. Supercond. **26** (3), 4401005.

[63] M. Berz, K. Makino and W. Wan (2015). *An introduction to beam physics.* CRC Press, Taylor & Francis, Boca Raton.

[64] K. Wille (2000). *The physics of particle accelerators,* Oxford University Press, Oxford.

[65] M. Berz (1999). *Modern Map Methods in Particle Beam Physics.* Academic Press, London.

[66] V. de Smet (2016). *Neutron measurements in a proton facility and comparison with Monte Carlo shielding simulations.* Dissertation. Université Catholique de Louvain.

[67] H. Owen, A. Lomax and S. Jolly (2016). *Current and future accelerator technologies for charged particle therapy*. Nucl. Instr. Meth. A **809**, pp 96-104.

[68] Advanced OncoTherapy (AVO) update of LIGHT development on medicalphysicsweb.org (2018). URL: http://medicalphysicsweb.org/cws/article/newsfeed/70751 (Jan 2018).

[69] J.M. Schippers and A.J. Lomax (2011). *Emerging technologies in proton therapy.* Acta Oncologia **50** (6), pp 838-850.

[70] S. Meyroneinc (2014). *Particle therapy and accelerators: part 1: on reliability of accelerators.* Presentation at JUAS 2014, Archamps.

[71] K. Peach (2011). *Applications of particle accelerators outside particle phyiscs.* Presentation at APPEAL 2, Oxford.

[72] U. Linz and J. Alonso (2016). *Laser-driven ion accelerators for tumor therapy revisited.* Phys. Rev. Accel. Beams **19**, 124802.

[73] V. Derenchuk (2013). *Particle Beam Technology and Delivery – Cyclotrons.* Presentation at the 55th AAPM Meeting, Proton Symposium, Indianapolis.

[74] E. Pearson et al (2013). *The new IBA superconducting synchrocyclotron (S2C2): From modelling to reality.* Conference paper, 11th International Tpoical Meeting on Nuclear Applications of Accelerators, Bruges.

[75] W. Kleeven (2014). *Accelerators for medical and industrial applications.* Presentation at JUAS 2014, Archamps

[76] M. Seidel (2016). *Cyclotrons.* Presentation at CERN Accelerator School 2016, Budapest.

[77] M. Pavlovic (2005). *Beam-transport study of an isocentric rotating ion gantry with minimum number of quadrupoles.* Nucl. Instr. Meth. A **545**, pp 412-426.

[78] S. Safai et al. (2012). *Improving the precision and performance of proton pencil beam scanning.* Translational Cancer Research **1** (3).

[79] U. Weinrich (2006). *Gantry design for proton and carbon handrontherapy facilities.* Proceedings of EPAC 2006, Edinburgh.

[80] A.M. Koehler, R.J. Schneider and J.M. Sisterson (1975). *Range modulators for protons and heavy ions.* Nucl. Instr. Meth. **131** (3), pp 437-440.

[81] J.D. Grant and J.Y. Chang (2014). *Proton-Based Stereotactic Ablative Radiotherapy in Early-Stage Non-Small-Cell Lung Cancer*. Bio. Med. Research Int. **2014**, 389048. Figure 1.

[82] http://www.protonpals.org/why-md-anderson-proton-therapy-center/ (Jan 2019)

[83] C.E. Vargas et al. (2018). *Hypofractionated versus standard fractionated proton-beam therapyfor low-risk prostate cancer: interim results of a randomized trial PCG GU 002.* Am. J. Clin. Oncol. **41** (2), pp 115-120.

[84] A.M. Laine et al. (2015). *The role of hypofractionated radiation therapy with photons, protons and heavy ions for treating extracranial lesions.* Front. Oncol. **5**, 302.

[85] A. Lomax (1999). *Intensity modulation methods for proton radiotherapy*. Phys. Med. Biol. **44**, p 188. Section 2.1.4.

[86] S. Webb (2001). *Intensity-modulated radiation therapy.* CRC Press, Taylor & Francis, Boca Raton.

[87] J. Flanz (2012). Particle Beam Scanning, in *Proton Therapy Physics* (Ed. H. Paganetti), pp 157-189. CRC Press, Taylor & Francis, Boca Raton.

[88] F. Stichelbaut, M. Closset and Y. Jongen (2014). *Secondary neutron sources in a compact proton therapy system.* Radiation Protection Dosimetry **16**, 1-4, pp 368-372.

[89] URL: www.iba-protontherapy.com (Jan 2019)

[90] URL: www.mevion.com (Jan 2019)

[91] K.P. Gall (2012). The single-room ion beam therapy facility, in *Ion Beam Therapy–Fundamentals, Technology, Clinical Application.* (Ed. U. Linz), pp 661-672. Springer, Berlin.

[92] Varian Press Release (2018). URL: https://www.varian.com/news/varian-introduces-new-probeam-360°-proton-therapy-system (Dec 2018)

[93] H. Blosser et al. (1985). *Rotatable superconducting cyclotron adapted for medical use*. US-Patent 4507616A.

[94] URL: http://www.shi.co.jp/english/products/industry/proton/img/index-ph01.jpg (Jan 2018)

[95] M.F. Moyers and W. Lesyna (2004). *Isocenter characteristics of an external ring proton gantry*. Int. J. Rad. Oncol. Biol. Phys. **60** (5), pp 1622-1630. Figures 2 and 3. Presented at the PTCoG Meeting 19 in Boston, 1993.

[96] W. Kleeven (undated). *IBA proton therapy systems, accelerators, beamlines and gantry technology.* Presentation slides published on the website of the Belgian Society for Radiation Protection, URL: http://www.bvsabr.be/js/tinymce/plugins/moxiemanager/data/files/20170626/IBA_WKleeven.pdf (Jan 2019)

[97] E. Pedroni, D. Meer et al. (2011). *Pencil beam characteristics of the next-generation proton scanning gantry of PSI: design issues and initial commissioning results*. Eur. Phys. J. Plus **126** (66).

[98] W.T. Chu et al. (1993). *Performance specification for proton medical facility.* LBL-33749, Lawrence Berkeley Laboratory, University of California.

[99] H. Owen et al. (2014). *Technologies for Delivery of Proton and Ion Beams for Radiotherapy.* Int. J. Mod. Phys. A **29**, 1441002.

[100] D. Robin (2011). *An Overview of Next Generation Gantries.* Presentation. 2nd Workshop on Hadron Beam Therapy of Cancer, Erice.

[101] J. Wulff, A. Winnebeck et al. (2014). *Short SAD Scanning Nozzle.* Varian Medical Systems, internal report (unpublished).

[102] J. Heese et al (2015), *Clinical aspects of compact gantry designs.* Presentation at EUCARD2 workshop: Superconductivity and other new Developments in Gantry Design for Particle Therapy, Bad Zurzach.

[103] URL: https://www.psi.ch/protontherapy/gantry-1 (Jan 2019)

[104] A.J. Lomax (2017). Private communication.

[105] URL: https://www.psi.ch/media/MM20161125JubilaeumProtonentherapieDE/igp_97c45ee47faca3a0643450bc7df546f6_bi1994m03_f052_n30__an_v.jpg (Jan 2018)

[106] D. Tommasini et al. (2017). *Status of the 16 T dipole development program for a future hadron collider.* IEEE Trans. Appl. Supercond. **28** (3).

[107] M. Guckenberger (2011). *Image-guided Radiotherapy Based on Kilovoltage Cone-beam Computed Tomography – A Review of Technology and Clinical Outcome*. European Oncology and Haematology **7** (2), pp 121-124.

[108] M. Grewe (2016). Private communication.

[109] A. Huggins and M. Benna (2016). *Proton beam delivery gantry with energy modulation functionality.* Varian Medical Systems, internal report (unpublished).

[110] S. Blundell (2009). *Superconductivity: A Very Short Introduction*. Oxford University Press, 1st edition, p 20. Oxford.

[111] T. Mathias et al. (1954). *Superconductivity of Nb_3Sn*. Phys. Rev. **95** (6), p 1435.

[112] T.G. Berlincourt and R.R. Hake (1962*). Pulsed-Magnetic-Field Studies of Superconducting Transition Metal Alloys at High and Low Current Densities*. Bull. Am. Phys. Soc. II **7** (408).

[113] J.G. Bednorz and K.A. Müller (1986). *Possible high T_c superconductivity in the Ba-La-Cu-O system*. Zeitschrift für Physik B **64** (2), pp 189-193.

[114] URL: http://ieeecsc.org/pages/nobel-laureates-superconductivity (Feb 2018)

[115] M.N. Wilson (1987). *Superconducting Magnets*. Clarendon Press, Oxford.

[116] URL: https://en.wikipedia.org/wiki/File:Magnetisation_and_superconductors.png (Jan 2019)

[117] G. Pretzler (2010). *Elektrizität & Magnetismus, Kapitel 3-4.* Lecture notes, HHU Düsseldorf.

[118] A. Winnebeck (2017). Varian Medical Systems, internal report (unpublished).

[119] V. Derenchuk (2016). *Application of Superconducting Technology for Proton Therapy*. Presentation at NAPAC 2016, Chicago.

[120] A. Gerbershagen et al. (2016). *The advantages and challenges of superconducting magnets in particle therapy*. Supercond. Sci. Technol. **29**, 083001.

[121] S. Prestemon, P. Ferracin and E. Todesco (2015). *AC Losses in Superconductors.* Presentation at USPAS 2015, New Jersey.

[122] M.N. Wilson (2013). *Superconducting Accelerator Magnets*. Lecture 1, held at Joint Universities Accelerator School, Feb. 2013, Archamps.

[123] G. Morgan (1970). *Theoretical Behavior of Twisted Multicore Superconducting Wire in a Time-Varying Uniform Magnetic Field*. Journal of Applied Physics **41**, p 3673.

[124] M. Ishizuka and J. Sakaruba (2006). *Critical current density (J_C) and mechanical characteristics of a Bi-2223 bulk current lead for cryocooler-cooled superconducting magnets.* Physica C: Supercond. **433** (3-4), pp 173-181, Fig. 4.

[125] V. Parma (2015). *Cryostat design.* CERN Yellow Report CERN-2014-005, pp 353-399.

[126] A. Godeke (2017). Private communication.

[127] M.N. Wilson (2013). *Superconducting Accelerator Magnets*. Lecture 1, held at Cockcroft Institute, Jan. 2013. Daresbury.

[128] A. Godeke et al (2007). *Limits of NbTi and Nb3Sn, and Development of W&R Bi-2212 High Field Accelerator Magnets*. IEEE Trans. Appl. Supercond. **17** (2), pp 1149-1152.

[129] D. Meyer and R. Flasck (1970). *A new configuration for a dipole magnet for use in high energy physics applications.* Nucl. Instr. Meth. **80**, pp 339-341.

[130] B. Auchmann, L. Brouwer, S. Caspi et al. (2018). *Electromechanical Design of a 16-T CCT Twin-Aperture Dipole for FCC.* IEEE Trans. Appl. Supercond. **28** (3), 4000705.

[131] S. Caspi et al. (2017). *Design of a Canted-Cosine-Thea Superconducting Dipole Magnet for Future Colliders.* IEEE Trans. Appl. Supercond. **27** (4), 4001505.

[132] V. Veksler (1945). *A new method to accelerate relativistic particles*. J. Phys. **9**, pp 153-158.

[133] M. Dehn et al. (2011). *The MAMI C accelerator.* Eur. Phys. J. Spec. Top. pp 198-219.

[134] H.A. Enge (1963). *Achromatic Magnetic Mirror for Ion Beams.* Rev. Sci. Instr. **34** (4), pp 385-389.

[135] W. Wan (2016). Informal note (unpublished).

[136] A. Huggins, L. Brouwer and W. Wan (2018). *Design and Simulation of High Momentum Acceptance Gantries for Ion Beam Therapy.* Presentation at ICAP 18, Key West.

[137] K. Makino and M. Berz (2006). *COSY INFINITY Version 9.* Nucl. Instr. Meth. Phys. Res. A **558** (1), p 346-350.

[138] *PSI Graphic Transport Framework* by U. Rohrer based on a CERN-SLAC-FERMILAB version by K.L. Brown et al. URL: http://aea.web.psi.ch/Urs_Rohrer/MyWeb/trans.htm

[139] K.L. Brown, D.C. Carey, Ch. Iselin and F. Rothacker. *Transport, a Computer Program for Designing Charged Particle Beam Transport Systems.* See yellow reports CERN 73-16 (1973) & CERN 80-04 (1980).

[140] *PSI Graphic Turtle Framework* by U. Rohrer based on a CERN-SLAC-FERMILAB version by K.L. Brown et al. URL: http://aea.web.psi.ch/Urs_Rohrer/MyWeb/turtle.htm

[141] *TOPAS Tool for Particle Simulation* by TOPAS MC Inc. URL: http://www.topasmc.org/

[142] J. Wulff and E.T. Abel (2016). *Calculation of Proton Pencil Beam Properties with Full Beamline Model in TOPAS.* Medical Physics **43** (6), p 3728.

[143] M. Benna and A. Rusanov (2017). Private Communication.

[144] J.M. Schippers (2017). Private Communication.

[145] H. Roecken (2011). *The VARIAN 250 MeV Superconducting Compact Proton Therapy Cyclotron.* Presentation. 2nd Workshop on Hadron Beam Therapy of Cancer, Erice.

[146] M. Berz and K. Makino (2015). *COSY INFINITY 9.2 Beam Physics Manual.* MSU Report MSUHEP 151103. Michigan State University.

[147] M. Berz and K. Makino (2015). *COSY INFINITY 9.2 Programmer's Manual.* MSU Report MSUHEP 151102. Michigan State University.

[148] M. Berz (1989). *Differential Algebraic Description of Beam Dynamics up to Very High Orders.* Particle Accelerators **24**, pp 109-124.

[149] Opera FEA Simulation Software. URL: https://operafea.com/

[150] ANSYS Maxwell. URL: https://www.ansys.com/products/electronics

[151] M. Benna (2015). Private Communication.

[152] J.M. Schippers, D. Meer, A. Gerbershagen (2017). EU patent application EP3167933A1.

[153] F. Kubo (2007). *Determination of the energy slit's width*. ACCEL internal report 1291-BP-6387-00 (unpublished).

[154] W.C. Hsi et al. (2009). *Energy spectrum control for modulated proton beams*. Med. Phys. **36** (6), pp 2297-2308.

[155] V. Anferov et al. (2015). *Estimating proton beam energy spread using Bragg peak measurement*. Med. Phys. **42** (6), 3365. Presented at PTCoG 54 (2017), San Diego.

[156] M.A. Chanrion et al. (2013). *Dosimetric consequences of pencil beam width variations in scanned beam particle therapy*. Phys. Med. Biol. **58**, pp 3979-3993.

[157] M. Klodowska et al (2016). *Criteria for spot asymmetry in proton radiotherapy pencil beam scanning – Monte Carlo study.* Poster at ICRE-PHE, Geneva.

[158] A. Huggins (2013). *Investigation and characterization of quality factors in proton radiation fields*. Master thesis, HHU Düsseldorf.

[159] A. Jankiowak (2006). *The Mainz Microtron MAMI – Past and future.* Eur. Phys. J. A **28**, s01, pp 149-160.

[160] J. Wulff and A. Huggins (2016). *Deriving Spot Shape Criteria for Proton Pencil Beam Scanning.* Medical Physics **43** (6), 3652.

[161] IEC 60601-2-64 (2014). *Particular requirements for the basic safety and essential performance of light ion beam medical electrical equipment.*
Published here: URL: https://webstore.iec.ch/publication/2680

[162] A. Gerbershagen et al. (2017). *Simulations and measurements of proton beam energy spectrum after energy degradation.* J. Phys.: Conf. Ser. **874**, 012108.

[163] D.S. Robin et al. (2011). *Superconducting toroidal combined-function magnet for a compact ion beam cancer therapy gantry.* Nucl. Instr. Meth. A **659**, pp 484-493.

[164] The *FLUKA* code for particle transport and interaction with matter, copyright owned by CERN and INFN. URL: http://www.fluka.org/fluka.php?id=about&mm2=1

[165] Godeke et al. (2015). *Bi-2212 Canted-Cosine-Theta coils for high-field accelerator magnets.* IEEE Trans. Appl. Supercond. **25** (3), 4002404.

[166] J. Kempe and A. Brahme (2008). *Energy-range relation and mean energy variation in therapeutic particle beams.* Medical Physics **35** (1), pp 159-170.

[167] C. Goodzeit, M. Ball, and R. Meinke (2003). *The double-helix dipole - a novel approach to accelerator magnet design.* IEEE Trans. Appl. Supercond. **13** (2), pp 1365–1378.

[168] C. Goodzeit, R. Meinke, and M. Ball (2007). *Combined function magnets using double-helix coils.* Proceedings of IPAC2007, pp 560–562.

APPENDIX

Derivation of the Courant-Snyder invariant from the phase ellipse

In the case of an upright ellipse with half-axes a and b the area can easily be calculated with:

$$A = \pi\, a\, b \tag{A.1}$$

The equation for the ellipse is with $\varphi = \{0 : 2\pi\}$:

$$x = a\cos\varphi \tag{A.2}$$

$$x' = b\sin\varphi \tag{A.3}$$

The equation of a tilted ellipse in Cartesian coordinates takes the angle of rotation ψ into account (see also appendix C of [64]):

$$x = a\cos\varphi\cos\psi + b\sin\varphi\sin\psi \tag{A.4}$$

$$x' = -a\cos\varphi\sin\psi + b\sin\varphi\cos\psi \tag{A.5}$$

Combining the two equations, and using $sin^2\,\varphi + cos^2\,\varphi = 1$, yields:

$$a^2b^2 = b^2(x\cos\psi - x'\sin\psi)^2 + a^2(x\sin\psi + x'\cos\psi)^2 \tag{A.6}$$

$$ab = \frac{b}{a}(x\cos\psi + x'\sin\psi)^2 + \frac{a}{b}(x\sin\psi + x'\cos\psi)^2 \tag{A.7}$$

$$\begin{aligned} ab = {} & x^2\left(\frac{b}{a}\,cos^2\psi + \frac{a}{b}\,sin^2\psi\right) + x'^2\left(\frac{b}{a}\,sin^2\psi + \frac{a}{b}\,cos^2\psi\right) \\ & +2xx'\cos\psi\sin\psi\left(\frac{b}{a} - \frac{a}{b}\right) \end{aligned} \tag{A.8}$$

Now by substituting

$$\cos\psi\sin\psi\left(\tfrac{b}{a} - \tfrac{a}{b}\right) = \alpha, \tag{A.9}$$

$$\frac{b}{a}\,sin^2\psi + \frac{a}{b}\,cos^2\psi = \beta, \tag{A.10}$$

$$\frac{b}{a}\,cos^2\psi + \frac{a}{b}\,sin^2\psi = \gamma, \tag{A.11}$$

the parametrized ellipse equation for a (tilted) ellipse reads:

$$\gamma x^2 + 2\alpha xx' + \beta x'^2 = ab \tag{A.12}$$

This is the Courant-Snyder invariant which is directly giving the constant area of the phase ellipse:

$$\frac{A}{\pi} = \gamma x^2 + 2\alpha xx' + \beta x'^2 \tag{A.13}$$

The parameters α, β and γ are the Twiss parameters and emittance ε is defined as A/π. From there by substituting the equations of the Twiss parameters, their inter-relation can be obtained:

$$\gamma = (1 + \alpha^2)/\beta \tag{A.14}$$