Investigations of the automatic design and optimisation of beam transport lines in particle accelerators with dedicated imaging properties

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Physik
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Stephan Reimann
aus Dresden

GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

Frankfurt am Main 2021
(D 30)

vom Fachbereich Physik der

Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan:	Prof. Dr. Harald Appelshäuser
erster Gutachter:	Prof. Dr. Holger Podlech
zweiter Gutachter:	Prof. Dr. Oliver Kester

Datum der Disputation: 20.12.2021

Zusammenfassung

Teil 1: Autotune

Nach einer Zusammenfassung der notwendigen Hintergründe zur Strahldynamik und zur Beschleunigerphysik in Kapitel 2, werden in dieser Arbeit zunächst verschiedene Optimierungstechniken besprochen und deren Anwendbarkeit auf die Transmissionsoptimierung und das Acceptance-Matching in Strahltransportstrecken analysiert. Dafür wurden Teilchen-Tracking-Simulationen mit verschiedenen künstlichen Beispielszenarien durchgeführt. Es zeigte sich, dass vor allem naturanaloge Optimierungsverfahren und unter diesen, speziell die genetischen Algorithmen, geeignet sind, optimale Lösungen dieser synthetischen Problemstellungen zu finden.

Anschließend wird ein Genereller Autotune Algorithmus (GAA) zur automatischen Optimierung von Strahltransportlinien entwickelt. Das Problem wird dabei auf die Optimierung einer Black-Box-Funktion $f : \mathbb{R}^n \to \mathbb{R}$ reduziert, welche einen n-dimensionalen Parameterraum auf die Transmission T abbildet. In diesem Fall wird dafür eine Ziel- bzw. Fitnessfunktion

$$F(\boldsymbol{x}) = 1 - T(\boldsymbol{x}) + \epsilon(\boldsymbol{x}) \tag{0.1}$$

verwendet, deren Minimum $F_{\min}$ der größten Transmission $T_{\max}$ entspricht. Der Term $\epsilon(\boldsymbol{x})$ ist entweder gleich Null, oder er repräsentiert die Parametrisierung weiterer Optimierungsziele, wie beispielsweise eine gleichzeitige, möglichst geringe Erregung der Quadrupole. Für die Optimierung wurde ein genetischer Algorithmus verwendet.

Die Leistungsfähigkeit dieser Methode wurde zunächst durch Teilchen-Tracking-Simulationen am Modell untersucht und schließlich online an einer 70 m langen Strahltransportstrecke im GSI-Helmholtzzentrum getestet (Abbildung 0.1). Um das Minimum von F zu
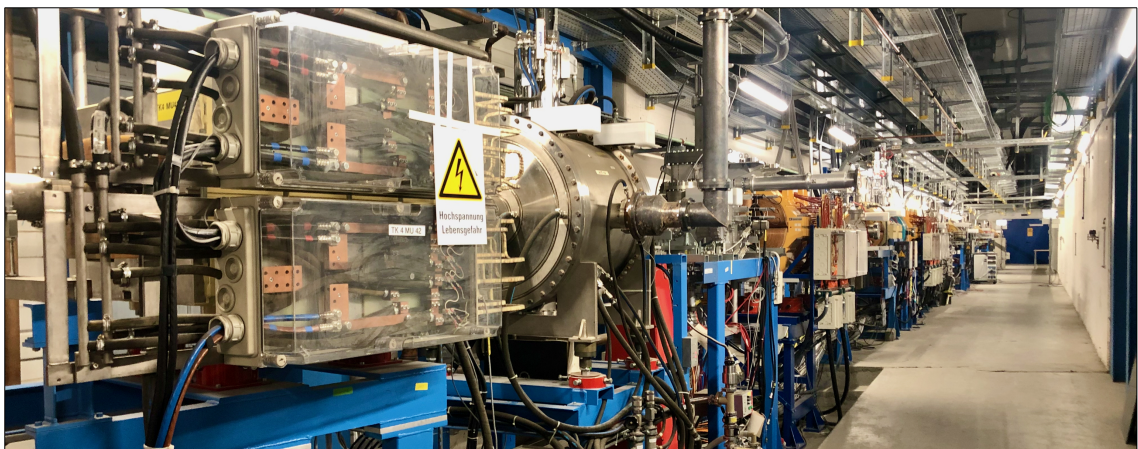

Abbildung 0.1.: Teilstück des Transferkanals vom universellen Linearbeschleuniger UNI-LAC zum Schwerionensynchrotron SIS18 am GSI Helmholtzzentrum für Schwerionenforschung in Darmstadt.

finden, mussten verschiedene Parametersätze, die sich aus dem Algorithmus ergeben, in die Maschine geladen und die jeweils resultierende Transmission gemessen werden.

In der getesteten Strecke standen 6 Quadrupolmagnete und 6 Steerer für das automatische Tuning zur Verfügung. Der Parameterraum hatte in diesem Fall also 12 Dimensionen. Ein einfacher Parameterscan ist in so einer Situation offensichtlich unpraktikabel, da jeder Punkt im Parameterraum als Maschineneinstellung mit Strahlpuls hätte getestet werden müssen. Je nach Rastergröße und Wiederholrate, würde ein solches Vorgehen mehrere Tage oder sogar Jahre in Anspruch nehmen. Ein genetischer Algorithmus benötigt dafür wesentlich weniger Zyklen.

In Abbilung 0.2 ist die Entwicklung der Transmission von 4 aufeinanderfolgenden Strahltransformatoren (GTK4DT3, GTK5DT1, GTK6DT2, GTK7DT3), bezogen auf eine Referenzmessung an GTK3DT4, dargestellt. Man erkennt, dass sich die Gesamttransmission (gemessen an GTK7DT3) nach 19 Generationen von 0% auf ca. 70% deutlich verbessert hat. Die konstant eingeschränkte Transmission zwischen dem Referenztransformator GTK3DT4 und dem nachfolgenden GTK4DT3 weist darauf hin, dass die Einstellung der Strahlführung vor dem Referenztransformator nicht optimal war und die daraus resultierende Fehlanpassung mit dem Zwischenquadrupol-Doublet GTK4QD2 nicht kompensiert werden konnte. Für die Bewertung der Performance soll dieser Abschnitt daher nicht herangezogen werden. Im darauf folgenden, 54 m langen Abschnitt von GTK4DT3 bis GTK7DT3 betrug die finale Transmission $T = 86\%$ und befindet sich damit im Bereich dessen, was in vergleichbarer Zeit auch mit manuellem Tuning erreicht werden kann.
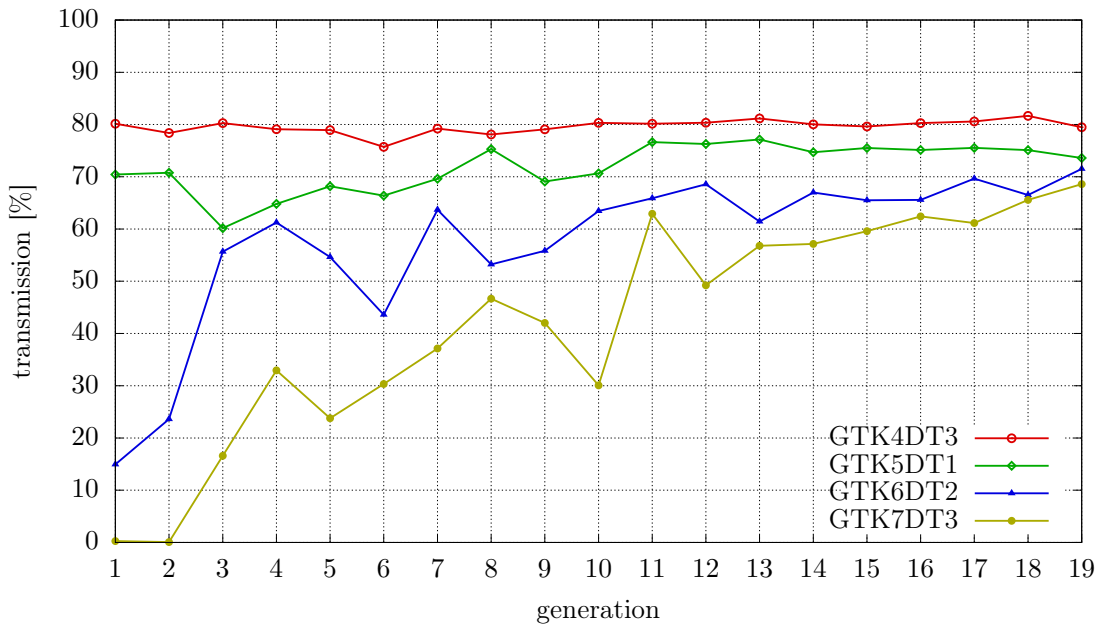

Abbildung 0.2.: Entwicklung der Transmission bezüglich der Referenzmessung am Transformator GTK3DT4 während der Onlineoptimierung mit GAA.

Für dieses Ergebnis mussten 2400 Beschleunigereinstellungen getestet werden und bei der zur Verfügung stehenden Wiederholrate von $f_r = 1\,\mathrm{Hz}$ dauerte die Optimierung insgesamt 40 Minuten. Diese Zeit ließe sich mit verschiedenen Verbesserungen allerdings weiter reduzieren. Eine Möglichkeit ist zum Beispiel, solche Settings vom Online-Test auszuschließen, die von einer parallel laufenden Trackingroutine als sichere Totalverluste identifiziert werden. Diese Maßnahme würde auch im Bezug auf den Strahlenschutz eine Verbesserung

bedeuten, denn aufgrund des statistischen Ansatzes des Verfahrens, werden im Rahmen der Optimierung regelmäßig auch solche Einstellungen getestet, die größere Transmissionsverluste produzieren. Bei den Experimenten mit echtem Ionenstrahl, war es daher erforderlich, die Aspekte des Strahlen- und Maschinenschutzes in besonderer Weise zu berücksichtigen. So wurde für den Test ausschließlich ein Pilotstrahl verwendet. Das heißt, der maximale Pulsstrom wurde auf $I < 100\,\mu\text{A}$ begrenzt, um Strahlverluste durch exotische Settings zu begrenzen, was außerdem den Vorteil hatte, dass sämtliche Strahltransformatoren stabil im kleinsten Messbereich betrieben werden konnten. Mit dem Versuch konnte gezeigt werden, dass der Algorithmus ein hervorragendes Werkzeug für die assistierte Strahloptimierung ist. Darüber hinaus gibt es Szenarien in denen GAA als vollständig autonomer Beschleuniger-Controller eingesetzt werden kann.

Teil 2: Lattice-Konstruktion

Im zweiten Teil der Arbeit liegt der Fokus auf dem Design von Strahltransportstrecken mit dedizierten Abbildungseigenschaften im sechsdimensionalen Phasenraum. Mit einer Erweiterung des Ansatzes aus Teil 1, lässt sich die Auslegung einer beliebigen Strahltransportstrecke ebenfalls als Optimierungsproblem formulieren, das sich theoretisch algorithmisch lösen lässt. Ziel war es dabei, eine optimale Anpassung einer gegebenen Teilchenverteilung an eine beliebige Akzeptanz, mit minimalem Energie- und Materialaufwand (dh. mit möglichst geringen Feldstärken und möglichst wenigen Elementen) zu erreichen. Die Formulierung der individuellen Fragestellung wird dabei wieder auf das Problem der Minimierung einer einwertigen Funktion über einem multidimensionalen Kubus reduziert. Anhand von Testbeispielen mit bekannter Parameterraumtopologie zeigte sich, dass die Anzahl der Dimensionen und die Topologie erwartungsgemäß einen starken Einfluss auf die Performance der Optimierung haben. Beide Aspekte lassen sich aber in gewissem Rahmen durch eine spezielle Formulierung der Randbedingungen reduzieren. Um das Problem auf ähnliche Weise über eine Fitnessfunktion lösen zu können, wurde zunächst eine eindeutige Parametrisierung der Transportstreckengeometrie und der verwendeten Felder (Phenotyp) über eine Liste von Zahlen (Genotyp) benötigt.

Ausgangspunkt war der allgemeinste Fall einer Transportstrecke der Länge L, in der mindestens ein Abschnitt der Länge L_{opt} mit neuen Elementen bestückt werden soll. Wenn $L = L_{\text{opt}}$, dann gibt es keine Beschränkungen bezüglich der Position von neuen Elementen. Weiterhin gibt es eine Menge von μ frei positionierbaren, strahlmanipulierenden Komponenten $\{C_1, \cdots, C_\mu\}$. Diese können auch Meta-Elemente sein, die aus einer festen Kombination von Komponenten bestehen. Anschließend müssen diejenigen ν Komponenten hinzugefügt werden, deren Positionen unter Optimierungsgesichtspunkten fest sind, die sich also außerhalb von L_{opt} befinden. In der Summe ergibt das $N = \mu + \nu$ Komponenten. Jede dieser Komponenten $C_n \in \{C_1, \cdots, C_{\mu+\nu}\}$ hat eine bestimmte Anzahl von λ_n freien Parametern $x_{n,1} \cdots x_{n,\lambda_n}$. Diese Parameter können z. B. Quadrupolstärken, Gapspannungen, aber auch die Längen der jeweiligen Elemente sein. Die Gesamtzahl der freien Parameter ist damit

$$\eta_p = \sum_{n=1}^{N} \lambda_n \,. \tag{0.2}$$

Zusätzlich muss nun für jede bewegliche Komponente aus $\{C_1, \cdots, C_\mu\}$ eine Position definiert werden. Die Möglichkeit beliebiger Permutationen der Komponenten soll dabei im-

plizit enthalten sein. Zu diesem Zweck wird für jede Komponente eine relative Position $s_n \in [0,1]$ definiert. Die tatsächliche Position l_n ergibt sich dann aus

$$l_n = s_n \cdot \left(L_{\mathrm{opt}} - \sum_{n=1}^{\mu} L_n \right), \tag{0.3}$$

wobei L_n die Länge der Komponente C_n ist. Die Anzahl der freien Parameter erhöht sich also durch die freie Positionierung um μ, womit $\eta = \eta_p + \mu$.

Jede mögliche Realisierung einer Transportstrecke, welche den gegebenen Randbedinungen entspricht, kann damit einem Punkt in dem η-dimensionalen Vektorraum zugeordnet werden. Genauer gesagt, handelt es sich um einen Unterraum von $\mathbb{R}^\eta$, da die Parameter unter Beachtung der Gerätegrenzen auf den Bereich $[-1,1]$ normiert werden. Dies entspricht also einem Punkt

$$\boldsymbol{x} = \begin{pmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_{\eta_p} \\ \tilde{s}_1 \\ \vdots \\ \tilde{s}_\mu \end{pmatrix} \tag{0.4}$$

im η-Hyperwürfel. Dabei sind $\tilde{x}_n$ und $\tilde{s}_n$ jeweils die auf das Intervall $[-1,1]$ normierten Werte von x_n und s_n. Ziel ist es nun, einen Punkt in dieser Menge zu finden, dem die Transportstrecke mit der höchsten Transmission zugeordnet werden kann. Auf diese Weise werden sowohl Geometrie als auch Felder gleichzeitig optimiert.

Um einen Eindruck von der Topologie des Parameterraums bzw. der Funktion F über dem Parameterraum zu bekommen, wurden zweidimensionale Projektionen der Funktion F für verschiedene Beispiele gescannt, wobei alle anderen Koordinaten festgehalten wurden. Abbildung 0.3 zeigt die Resultate zweier solcher Scans für ein künstliches Problem, in
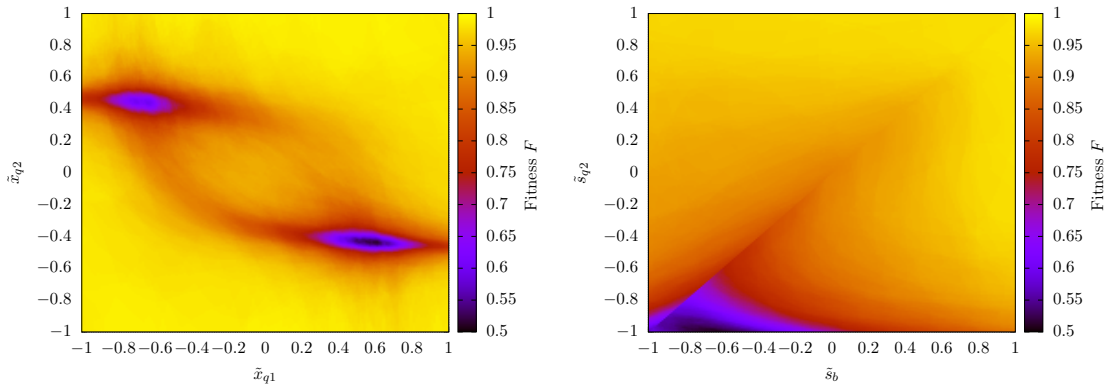

Abbildung 0.3.: Das Resultat von Scans zweier exemplarischer Unterräume des Parameterraums dieses Problems. (links: Fitnessfunktion in Abhängigkeit von der Fokussierstärke beider Quadrupole, rechts: Fitnessfunktion in Abhängigkeit von der Position des Bunchers und der des zweiten Quadrupols)

welchem zwei Quadrupole und ein Buncher optimiert werden sollen, um eine gegebene Akzeptanz zu treffen, während nur einer der Quadrupole eine feste Position hat und die Positionen des zweiten Quadrupols, sowie die des Bunchers, ebenfalls optimiert werden sollten. Das Problem hat $\eta = 5$ Freiheitsgrade und es zeigt sich, dass F in diesem Fall

mehrere lokale Minima besitzt, die sich vor allem aus der Quadrupolsymmetrie ergeben. Jede Lösung entspricht hier einem Punkt

$$\boldsymbol{x} = \begin{pmatrix} \tilde{x}_{\mathrm{q1}} \\ \tilde{x}_{\mathrm{q2}} \\ \tilde{x}_{\mathrm{U}} \\ \tilde{s}_{\mathrm{q2}} \\ \tilde{s}_{\mathrm{buncher}} \end{pmatrix} \tag{0.5}$$

im dazugehörigen Penterakt. Der Algorithmus konnte für dieses Problem eine plausible Lösung finden (Abbildung 0.4). Da sich die Parameterräume höherer Dimension nicht mehr in einer sinnvollen Granularität komplett scannen lassen, ist es dann nicht mehr
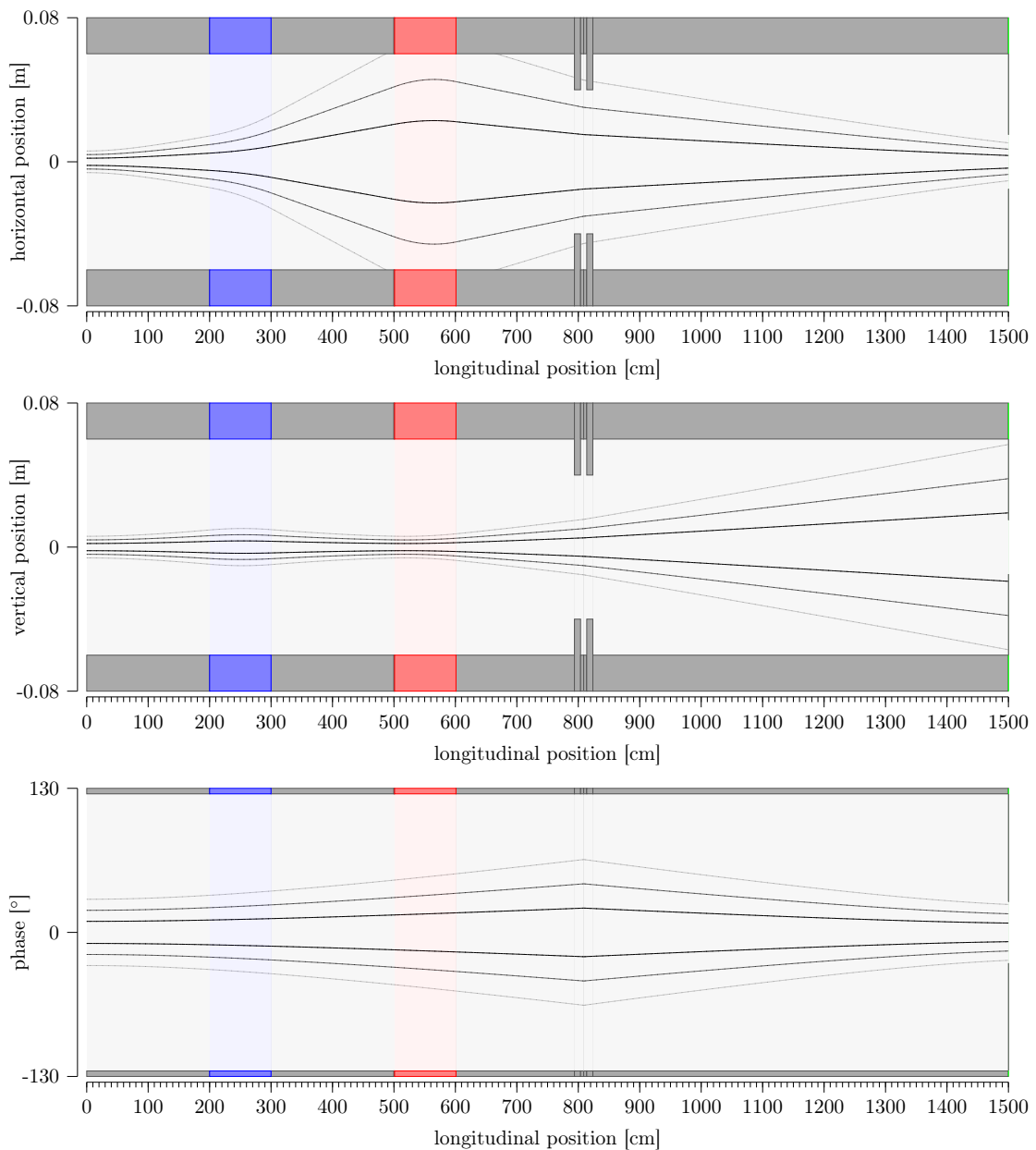

Abbildung 0.4.: Illustration einer Lösung mit maximaler Transmission ($T = 48.1\%$). Der vertikal fokussierende Quadrupol ist blau dargestellt, der horizontale rot.

möglich, jede Lösung des Algorithmus quantitativ zu beurteilen. Es lassen sich nur noch qualitative Aussagen machen. Ob diese Lösung das Optimum ist, lässt sich also nicht direkt verifizieren, aber sie erscheint zumindest sehr plausibel. Wäre die Phasenakzeptanz exakt identisch mit der Anfangsphasenbreite des Strahls, dann sollte der Spalt genau in der Mitte liegen. Die Phasenakzeptanz ist jedoch etwas größer, so dass eine Position hinter der Mitte ideal ist. Der zweite Quadrupol muss dann so weit vorne wie möglich positioniert werden, damit der Strahl horizontal gut durch die Driftröhren des Bunchers fokussiert werden kann. Wurden im Rahmen der Untersuchungen Positionierungsfehler bereits während der Optimierung induziert, dann ergaben sich Hinweise darauf, dass bei der genetischen Latticekonstruktion solche Lösungen bevorzugt werden, die sich als robust gegenüber Positionierungsfehlern erweisen. Tabelle 0.1 zeigt den Vergleich von Fehlerstudien nach einer Optimierung mit induzierten Fehlpositionierungen gegenüber einer ungestörten Konstruktion. Auf diese Weise ließe sich eine Fehlertoleranz bereits zu Beginn der Designphase implementieren und die Notwendigkeit von Fehlerstudien und daraus folgenden Iterationen des Designs könnten entfallen.

Tabelle 0.1.: Vergleich der Optimierung mit und ohne Positionsfehler. Die Tabelle enthält die jeweiligen Mittelwerte der Transmission, die Standardabweichung vom ungestörten Optimum und die Standardabweichung vom Mittelwert.

Optimierung	$\overline{T}$	$\sqrt{EW((T - T_{\mathrm{opt}})^2)}$	$\sqrt{EW((T - \overline{T})^2)}$
ohne Fehler	46.71%	1.7%	1.00%
mit Fehlern	46.71%	1.6%	0.95%

Die Anwendung dieser Methode auf zwei reale Beschleunigerprojekte ist Gegenstand von Kapitel 5.3. Für den geplanten Betrieb der CERN North-Area mit Protonenstrahlen bei Strahlimpulsen von $p < 10\,\mathrm{GeV}/c$, konnte die optimale Position verschiedener Kombinationen von Gaborlinsen ermittelt werden. Außerdem wurde die Konstruktion einer RFQ-Matching-Sektion (Abbildung 0.5) im Niedrigenergiebereich des zukünftigen Hoch-Brillanz-
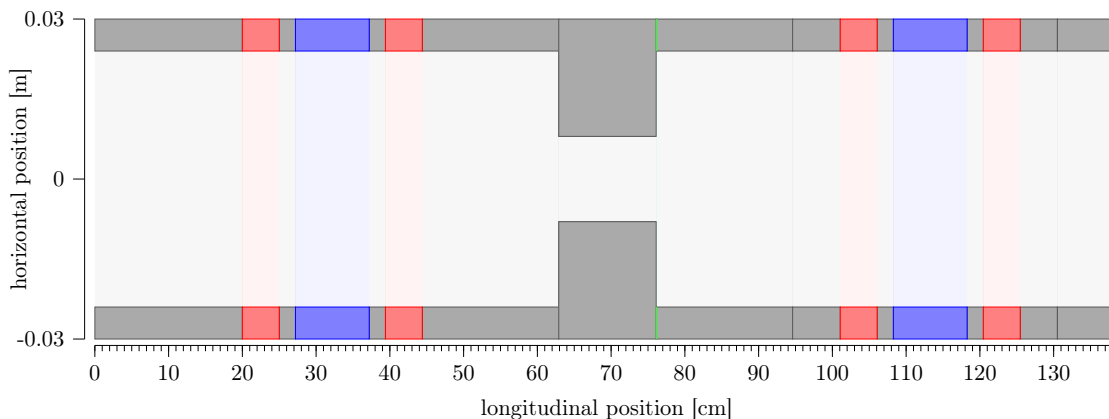

Abbildung 0.5.: Optimiertes Lattice für die RFQ-Matching-Sektion.

Protonenbeschleunigers für HBS am Forschungszentrum Jülich untersucht. Unter Beachtung verschiedener Randbedingungen (Größe der verwendeten Komponenten und reali-

sierbare Abstände), wurde die Matching-Sektion für die gegebene Ausgangsverteilung des RFQ 1 und die Zielakzeptanz des RFQ 2 automatisch konstruiert.

Dabei konnte eine Transmission von 93% erreicht werden. Aus der Studie ergab sich direkt die Anforderung, eine der gegebenen Randbedingungen zu ändern. Es wurde deutlich, dass die Phasenbreite am Buncher zu groß ist, um den Strahl vollständig fokussieren zu können (Abbildung 0.6) und dieser bei der gegebenen Eingangsverteilung daher weiter vorne positioniert werden sollte. Die GLC-Methode erwies sich in beiden Beispielen als hervorragendes Werkzeug und lieferte überzeugende Ergebnisse.
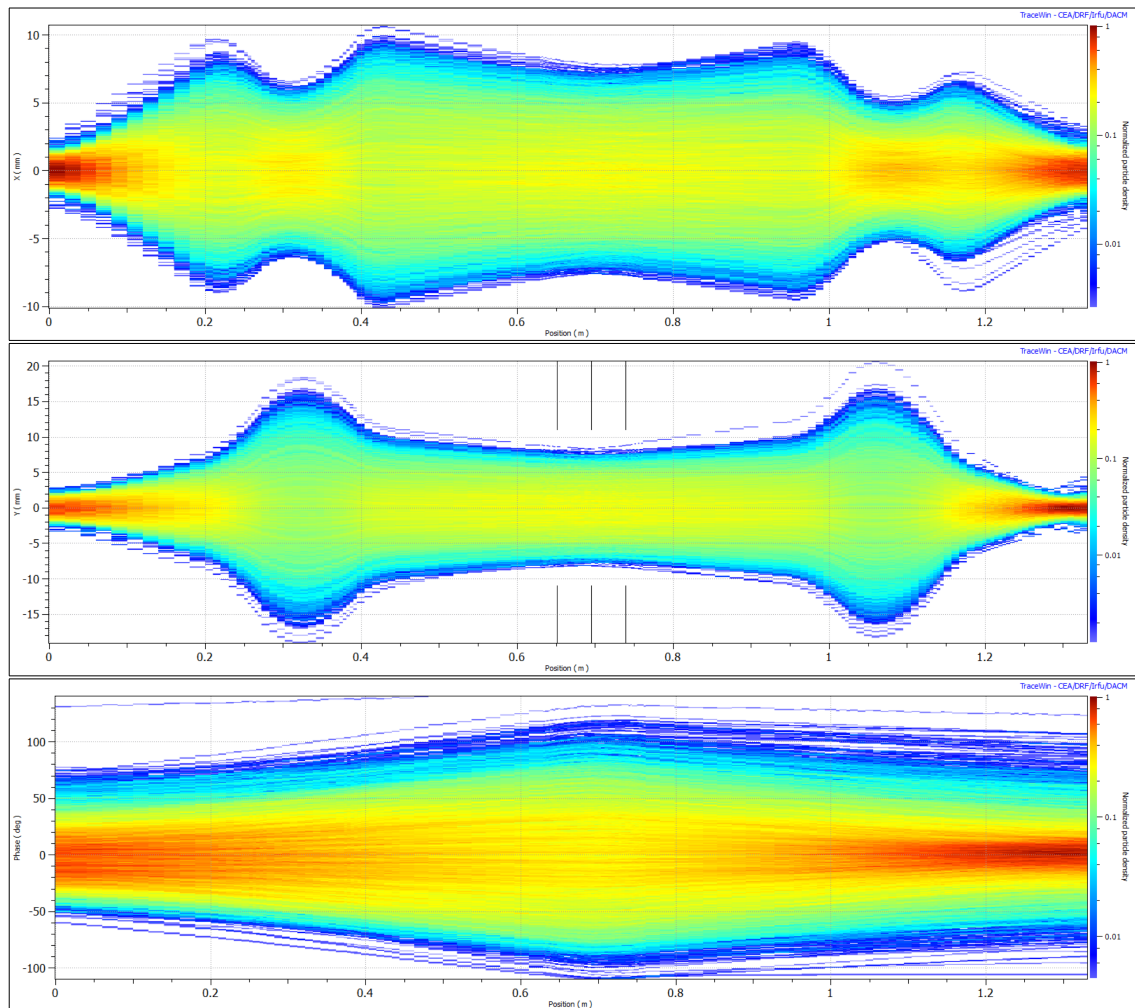

Abbildung 0.6.: Ergebnis der Optimierung (Tracewin-Darstellung) (oben: horizontale Projektion, Mitte: vertikale Projektion, unten: Längsbewegung (Phase)).

Die Referenzimplementierungen der verschiedenen, in dieser Arbeit entwickelten Methoden wurden in eine neue, interaktive Softwareanwendung Accelerator Construction Set (ACS) integriert, die in Kapitel 6 kurz vorgestellt wird. Dieses Werkzeug ermöglicht die Zusammenstellung von Strahlführungen aus einem Satz von vordefinierten Elementen und bietet einen einfachen Zugang zu den vorgestellten Algorithmen. Auf diese Weise kann die Anwendung sowohl zur Optimierung von Optikparametern als auch als automatischer Lattice-Designer verwendet werden.

Abstract

The first part of this work addresses the automatic online tuning of transfer lines in particle accelerator facilities. In the second part the focus lies on the automatic construction and optimisation of such transport lines. It can be shown that genetic algorithms can be used very well for optimisation in both cases. Automatic online tuning can be performed very efficiently at accelerators under certain boundary conditions and is particularly well suited for initial beam commissioning with low intensity pilot beams. The construction of transfer lines can also be formulated and solved as an minimisation problem with an adopted parameterisation. Thereby, both the imaging properties of the beam transport and the robustness against error studies can be optimised at the same time.

List of acronyms

GSI GSI Helmholtz Centre in Darmstadt, Germany

FAIR Facility for Antiproton and Ion Research

UNILAC Universal Linear Accelerator at GSI

SIS18 18 T m Heavy Ion Synchrotron at GSI

HSI High Current Injector

HLI High Charge Injector

FRS Fragment Separator

ESR Experimental Storage Ring at GSI

CRYRING@ESR Crying Storage Ring downstream the ESR

CERN European Organisation for Nuclear Research in Geneva, Switzerland

SLAC SLAC National Accelerator Laboratory in Stanford, USA

HBS High Brilliance Neutron Source at Research Centre Jülich, Germany

TK Transfer Channel between UNILAC and SIS18

AGS Alternating Gradient Synchrotron at the Brookhaven National Laboratory, USA

Contents

1. Introduction

1.1. Motivation

Despite great progress in automation over the past decades, a lot of manual tuning still occurs at many accelerator facilities. Especially for transfer lines in flexible user facilities, a lot of time is spent on manual setup. One example is the accelerator complex at the GSI Helmholtz Centre in Darmstadt, Germany.

For the users of the GSI accelerators, almost all elements of the periodic table, from Hydrogen to Uranium, and various isotopes can be provided. In addition, rare radioactive isotopes can be produced for storage ring or fixed-target experiments via the fragment separator FRS. At the beginning of the accelerator chain is the UNILAC, which is supplied with beam via two injectors (HSI and HLI) from three ion source terminals. The UNILAC can be pulsed with 50 Hz and for each pulse in principle a different source and thus a different ion species can be requested and accelerated to an energy up to 11.4 MeV/u. The system is specially designed for complex and highly flexible parallel operation. This allows the UNILAC to supply up to five different beam targets with beam quasi simultaneously. The principle of virtual accelerators is used for this purpose. For each running experiment, a complete set of accelerator settings is kept, which is optimised to the specific requirements of the experiment (isotope, charge state, beam energy, intensity, target focusing, pulse length). This data set can then be used on demand. A pulse controller unit organises the timing of the virtual accelerators in the 50 Hz pattern according to the specified sharing ratio.

Further acceleration to energies corresponding to a beam rigidity of 18 T m, is performed via the SIS18, which requests up to one beam pulse per second from the UNILAC, depending on the requested extraction time. In each cycle, a different beam target can be supplied with the desired beam (e.g. the ESR storage ring or the fragment separator FRS). This system is highly efficient, but also makes high demands on hardware and stability. That is why field-regulated magnets and power supplies with high precision and short response times are used. It is planned to extend this principle further to the beam production chains of the Facility for Antiproton and Ion Research FAIR, which is presently under construction at GSI Helmholtz Centre.

Due to the various usage scenarios and the associated frequently changing requirements for the beam parameters, it is necessary to often readjust or at least optimise the machine settings of the virtual accelerators. The theory settings available for each beam path are generally scalable with magnetic rigidity $B\rho$, but it turns out that the transport properties are not particularly reproducible. Without further optimisation, this often leads to settings with poor overall transmission and not seldom to complete beam loss. The reasons are manifold. For example, different source types or instances are used alternately, so that the beam properties already vary from source to source to a certain extent. For a good theory setting, a precise knowledge of the input beam parameters is crucial. However, this is not always known or is subject to changes. Especially in the range of low beam rigidities,

where power supplies are operated close to their lower limits, these variations add up and have a large influence on the overall transmission. The effect is especially noticeable in highly parallel operation modes with many users, or when timing settings are changed. As a consequence, the machine parameters have to be manually re-adjusted, which not only results in a loss of efficiency, but also requires increased personnel effort. There is therefore a great interest in automating such processes. When performing automatic settings, it is particularly important to consider not only the desired beam parameters and optimum transmission, but also aspects of machine protection and radiation safety.

A related optimisation problem can be formulated for the construction of beam transport lines. Even though this case is usually not primarily time-critical, an automatic design is expected to optimise the design process in other ways. For many beam transport problems, more than one solution can be considered, and it is ultimately up to the experience of the beam dynamists which one is preferred. Whether this is the best possible solution cannot ultimately be determined. In addition to the realisation of the required imaging properties, it is particularly important to develop a solution that also proves to be robust against field and alignment errors in error studies.

1.2. Contents

After summarising the necessary background on beam dynamics and accelerator physics in chapter 2, this thesis first reviews various optimisation techniques and analyses their applicability to transmission optimisation and acceptance matching (chapter 3).

Subsequently, a **G**eneral **A**utotune **A**lgorithm for fully automatic optimisation of beam transport lines is developed. The performance of this algorithm was first investigated by particle tracking simulations and eventually tested online at the GSI transfer channel TK. It is shown that the algorithm is a great tool for assisted tuning and that there are even scenarios where GAA can be used as a fully autonomous controller. The results are presented in chapter 4. From the investigations, principles can be derived according to which accelerator sections or transport channels must be designed and equipped with appropriate diagnostics so that the algorithm can achieve optimal results.

Using an extension of this approach, a complete accelerator section can be automatically designed in such a way, that an optimal adaptation of a given particle distribution to an arbitrary acceptance can be achieved with minimal energy and material requirements. For this purpose, the **G**enetic **L**attice **C**onstruction method was developed and presented for the first time in frame of this work [1]. In this study, both geometry and fields are optimised simultaneously. The method and its application to new projects, as a specific transport line at CERN and a matching section of the future high brilliance proton accelerator for HBS, are subject of chapter 5.

The reference implementations of the various methods developed in this thesis have been integrated into an interactive software application (**A**ccelerator **C**onstruction **S**et), which is briefly presented in chapter 6. This tool allows the composition of beam lines from a set of pre-defined elements and provides an easy access to the presented methods. In this way it can be used both for automatic tuning and as an automatic lattice designer.

2. Beam dynamics

A theoretical description of the influence of different accelerator elements on the dynamics of charged particle beams is required for detailed studies on the optimisation of beam transport lines. The theoretical background, which is the basis for the present work is summarised in this chapter.

Besides important beam diagnostic devices (beam transformers, luminous screens, profile grids, phase probes, rest gas monitors, etc.), beam transfer lines contain several beam manipulating elements such as deflection or correction magnets and optical elements which can change the focusing of the ion beam or correct aberrations caused by momentum deviations of the particles. Figure 2.1 shows examples of standard elements with magnetic mid-plane symmetry, which are commonly used in transfer lines.
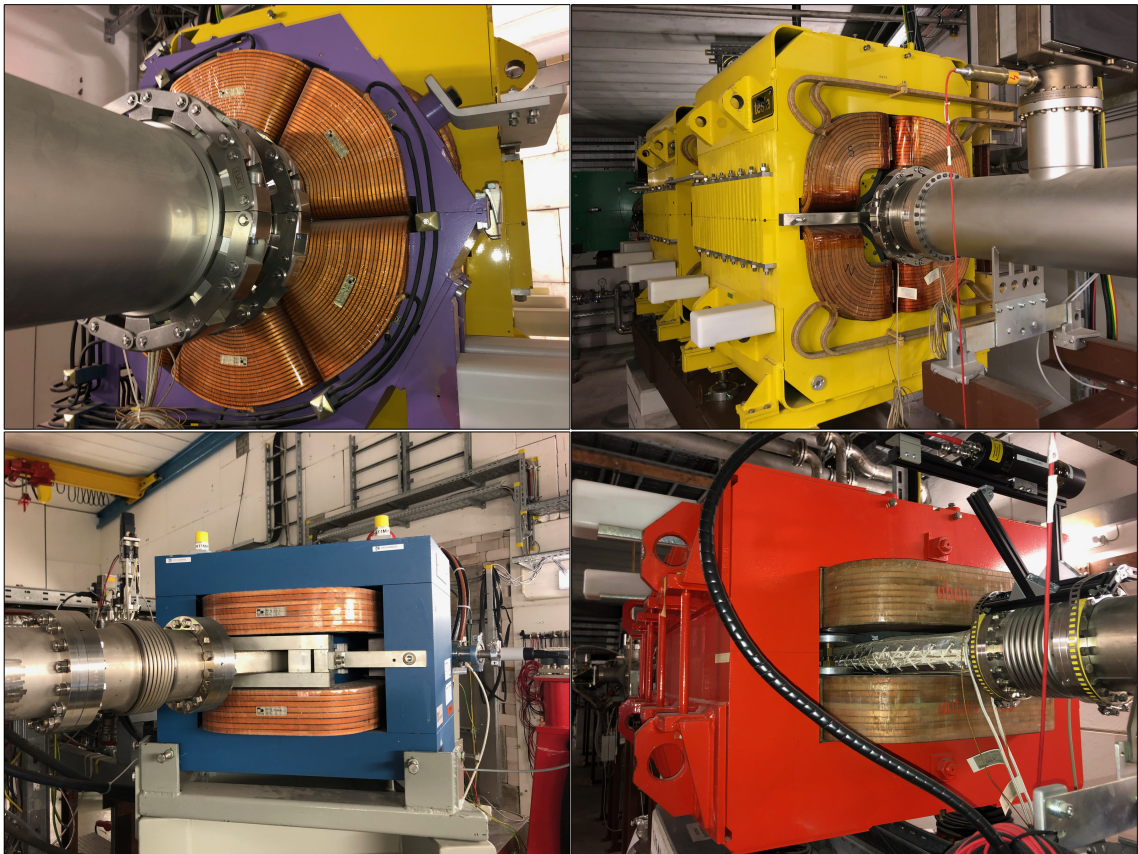

Figure 2.1.: Examples of beam optical elements of the magnetic mid-plane from the high-energy beam lines at GSI Helmholtzzentrum, designed for ion beams with a magnetic rigidity of up to $18\,\mathrm{T\,m}$. (top left: sextupole magnet, top right: quadrupole magnet, bottom left and bottom right: bending dipole magnet)

The Lorentz force (2.1) provides two ways to influence the trajectory of moving charged particles (with charge q and velocity $\boldsymbol{v}$), namely by electric and magnetic fields.

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \tag{2.1}$$

Although electrostatic elements are also sometimes used for the above purposes (e.g. at the local Cryring injector at GSI), they play a role only at very low particle momenta or in injection and extraction systems of synchrotrons, because of the technical limitation in generating high electrostatic field strengths ($E_{max} \approx 10\,\mathrm{MV/m}$). Therefore, only the magnetic elements are considered in the first place. But the following derivations can be carried out analogously also for electrostatic elements.

The dynamics of charged particle beams in accelerators have been described in various textbooks and other publications [2–5]. The fundamental derivations for the particle dynamics in the mid-plane (Figure 2.2) are based on the description of the matrix theory from K. L. Brown [2] and its adoptions in the lectures of F. Hinterberger [3].
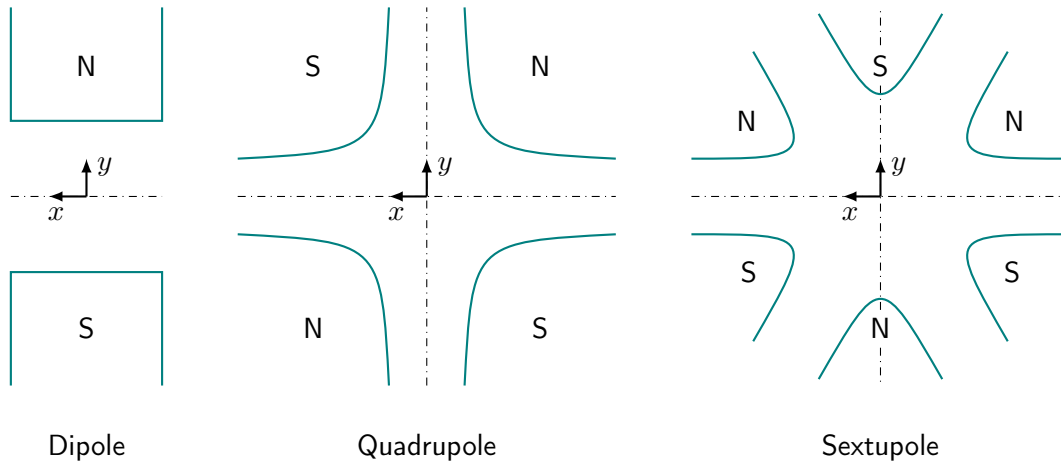
Figure 2.2.: Schematic view of the magnetic mid-plane for dipole, quadrupole and sextupole elements [2, page 26]. The magnetic fields of these elements are antisymmetric with respect to the y-coordinate.

The standard coordinate system usually used in accelerator physics (Figure 2.3) is a right-handed cartesian coordinate system (x, y) perpendicular to the central orbit s. The origin of the coordinate system is shifted along the nominal orbit s, following the central particles path. The functions $x(s)$ and $y(s)$ therefore describe the path of a particle relative to the nominal orbit. In practice, the magnetic centre plane is usually identical with the horizontal plane (x, s). Therefore, it is first assumed that there is only a horizontal curvature of the nominal path with a bending radius ρ around the y-axis. So the curvilinear coordinate system (x, y, s) locally corresponds to a cylinder coordinate system with the radial coordinate $r = \rho + x = \rho(1 + hx)$, the axial coordinate y and the angular coordinate $\alpha = s/\rho$, where the curvature $h = \rho^{-1}$. The equations of transformation are

$$\tilde{x} = r \sin \alpha = \rho(1 + hx) \sin\left(\frac{s}{\rho}\right),$$

$$\tilde{y} = y, \tag{2.2}$$

$$\tilde{z} = r \cos \alpha = \rho(1 + hx) \cos\left(\frac{s}{\rho}\right).$$
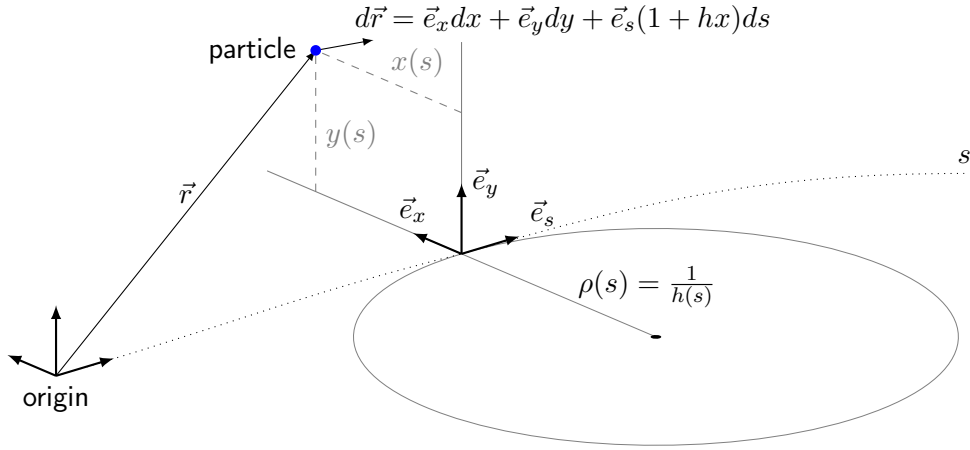
Figure 2.3.: The curvilinear standard coordinate system of particle physics [2, page 12]. The metrics for the line element is defined such that in case of a disappearing curvature h, the coordinate system changes to an ordinary cartesian coordinate system.

With the total differentials

$$d\tilde{x} = \frac{\partial \tilde{x}}{\partial x}dx + \frac{\partial \tilde{x}}{\partial y}dy + \frac{\partial \tilde{x}}{\partial s}ds = \sin\left(\frac{s}{\rho}\right)dx + (1+hx)\cos\left(\frac{s}{\rho}\right)ds\,,$$

$$d\tilde{y} = \frac{\partial \tilde{y}}{\partial x}dx + \frac{\partial \tilde{y}}{\partial y}dy + \frac{\partial \tilde{y}}{\partial s}ds = dy\,, \tag{2.3}$$

$$d\tilde{z} = \frac{\partial \tilde{z}}{\partial x}dx + \frac{\partial \tilde{z}}{\partial y}dy + \frac{\partial \tilde{z}}{\partial s}ds = \cos\left(\frac{s}{\rho}\right)dx - (1+hx)\sin\left(\frac{s}{\rho}\right)ds\,,$$

the line element is

$$dr^2 = dx^2 + dy^2 + (1+hx)^2 ds^2\,. \tag{2.4}$$

Therefore the metric tensors are

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (1+hx)^2 \end{pmatrix} \quad \text{and} \quad g^{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (1+hx)^{-2} \end{pmatrix} \tag{2.5}$$

and the gradient and Laplace operator can be derived to

$$\nabla = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{1}{1+hx}\frac{\partial}{\partial s}\,, \tag{2.6}$$

$$\nabla^2 = \frac{1}{1+hx}\frac{\partial}{\partial x}\left((1+hx)\frac{\partial}{\partial x}\right) + \frac{\partial^2}{\partial y^2} + \frac{1}{1+hx}\frac{\partial}{\partial s}\left(\frac{1}{1+hx}\frac{\partial}{\partial s}\right)\,. \tag{2.7}$$

Since the particle moves mainly in the longitudinal direction, the transversal velocity components can be neglected. So the particle velocity is approximately $\boldsymbol{v} = (0,0,v_s)$ and the curvature of the particles path depends only on the axial magnetic field component $B_0 = B_0(s) = B_y(x=0, y=0, s)$. For a particle with charge q to follow the curvature of

the orbit path, Lorentz force and centrifugal force must cancel each other out

$$qv_s B_0 = \frac{\gamma m v_s^2}{\rho}\,.$$

(2.8)

Using the relativistic momentum $p_0 = \gamma m v_s$, one obtains the important relation

$$\boxed{B_0\rho = \frac{p_0}{q}\,,}$$

(2.9)

where $B_0\rho$ is called the magnetic rigidity.

2.1. Taylor-series expansion of the magnetic field

For the further steps a special representation of the Taylor-series expansion of the magnetic field $\boldsymbol{B} = (B_x, B_y, B_s)$ is used, as it was originally developed by L. C. Teng in 1962 [4]. Using Maxwell's equations for the vacuum and the fact that there is no time-varying electric field, Ampère's circuital law is reduced to

$$\nabla \times \boldsymbol{B} = 0\,.$$

(2.10)

The rotation of a scalar gradient disappears, therefore in this case the magnetic field can be written as a gradient of a scalar potential $\boldsymbol{B} = \nabla\phi_m$. For the potential $\phi_m(x,y,s)$ a Taylor expansion can be performed at the position $\boldsymbol{0}$. Due to the mid plan symmetry (Figure 2.2), ϕ_m is an odd function regarding y

$$\phi_m(x,y,s) = -\phi_m(x,-y,s)\,.$$

(2.11)

This eliminates all terms that do not contain y with odd exponents. The detailed derivation can be found in appendix A.1. The resulting expression of the magnetic field up to the second order can be written in the general form

$$\boxed{\begin{aligned}
B_x(x,y,s) &= \frac{p_0}{q}\left(-nh^2 y + 2\beta h^3 xy\right),\\
B_y(x,y,s) &= \frac{p_0}{q}\left(h - nh^2 x + \beta h^3 x^2 - \frac{1}{2}(h'' + 2\beta h^3 - nh^3)y^2\right),\\
B_s(x,y,s) &= \frac{p_0}{q}\left(h'y - (n'h^2 + 2nhh' + hh')xy\right),
\end{aligned}}$$

(2.12)

where

$$h = \frac{q}{p_0}B_0\,, \qquad\qquad\qquad h' = \frac{\partial h}{\partial s}\,,$$

$$n = -\frac{1}{hB_0}\frac{\partial B_y}{\partial x}\bigg|_{(x=0,y=0)}\,, \qquad\qquad n' = \frac{\partial n}{\partial s}\,,$$

(2.13)

$$\beta = \frac{1}{2!h^2 B_0}\frac{\partial^2 B_y}{\partial x^2}\bigg|_{(x=0,y=0)}\,.$$

2.2. The equations of motion

The starting point for the equation of motion in linear approximation is the Lorentz force which acts on a particle with the charge q and mass m. Converted to acceleration follows

$$\ddot{\boldsymbol{r}} = \frac{q}{\gamma m}\boldsymbol{v} \times \boldsymbol{B}\,. \tag{2.14}$$

In the area of a deflecting magnet, the orbit describes a circular path with the curvature radius ρ and the angular velocity ω. This results in the radial acceleration $a_r = -\omega^2(\rho + x)$. Thus the total radial acceleration is $\ddot{x} - \omega^2(\rho + x)$ and in combination with (2.14) follows

$$\ddot{x} - \omega^2(\rho + x) = \frac{q}{\gamma m}(v_y B_z - v_z B_y)\,,$$
$$\ddot{y} = \frac{q}{\gamma m}(v_z B_x - v_x B_z)\,. \tag{2.15}$$

Since the transverse components of speed v_x and v_y are very small compared to v_z and B_x and B_z are very small compared to B_y, the terms $v_y B_z$ and $v_x B_z$ can be neglected. The longitudinal component holds the largest part of the velocity, $v_z \approx v$ and thus $p = \gamma m v \approx \gamma m v_z$. With $v_z = \omega(\rho + x)$ the equation converts to

$$\ddot{x} - \omega^2(\rho + x) = -\frac{q}{p}\omega^2(\rho + x)^2 B_y\,,$$
$$\ddot{y} = +\frac{q}{p}\omega^2(\rho + x)^2 B_x\,. \tag{2.16}$$

The time derivatives are substituted by derivatives after s

$$\frac{d}{dt} = \frac{ds}{dt}\frac{d}{ds} = \rho\omega\frac{d}{ds} \quad \text{or rather} \quad \frac{d^2}{dt^2} = (\rho\omega)^2\frac{d^2}{ds^2}\,. \tag{2.17}$$

Dividing by $(\rho\omega)^2$ and with $h = \rho^{-1}$ follows

$$x'' - h(1 + hx) = -\frac{q}{p}B_y(1 + hx)^2\,, \tag{2.18}$$
$$y'' = +\frac{q}{p}B_x(1 + hx)^2\,. \tag{2.19}$$

Using the Taylor expension of the magnetic field (2.12) up to first order,

$$x'' - h(1 + hx) = -(h - nh^2 x)(1 + 2hx + h^2 x^2)\,, \tag{2.20}$$
$$y'' = +(0 - nh^2 y)(1 + 2hx + h^2 x^2) \tag{2.21}$$

and after substitution $k_x = (1 - n)h^2$ and $k_y = nh^2$ and neglecting second order terms, one obtains the homogeneous equations of motion for mono-energetic particles:

$$\boxed{\begin{array}{l} x'' + k_x x = 0 \\ y'' + k_y y = 0 \end{array}} \tag{2.22}$$

These equations are the equations of the harmonic oscillator, whose general solutions are the linear combination of the two linear independent characteristic solutions

$$x(s) = x_0 c_x(s) + x_0' s_x(s) \quad \text{and} \quad y(s) = y_0 c_y(s) + y_0' s_y(s)\,. \tag{2.23}$$

2.2.1. Transverse characteristic solution

For $k > 0$ the characteristic solutions of the equations (2.22) are:

$$c_x(s) = \cos\left(\sqrt{k_x}s\right); \qquad s_x(s) = \frac{\sin\left(\sqrt{k_x}s\right)}{\sqrt{k_x}},$$

$$c_y(s) = \cos\left(\sqrt{k_y}s\right); \qquad s_y(s) = \frac{\sin\left(\sqrt{k_y}s\right)}{\sqrt{k_y}}. \tag{2.24}$$

For $k < 0$ this is:

$$c_x(s) = \cosh\left(\sqrt{|k_x|}s\right); \qquad s_x(s) = \frac{\sinh\left(\sqrt{|k_x|}s\right)}{\sqrt{k_x}},$$

$$c_y(s) = \cosh\left(\sqrt{|k_y|}s\right); \qquad s_y(s) = \frac{\sinh\left(\sqrt{|k_y|}s\right)}{\sqrt{k_y}}. \tag{2.25}$$

For $k_x = k_y = 0$ the result simplifies to:

$$c_x(s) = 1; \qquad s_x(s) = s,$$

$$c_y(s) = 1; \qquad s_y(s) = s. \tag{2.26}$$

2.2.2. The particular solution of dispersion

In reality particles are not mono-energetic but have a momentum difference $\Delta p = p - p_0$ to the central particle. Inserting the relative momentum deviation $\delta = \Delta p/p_0$ up to the linear approximation into equation (2.18), the differential equation for the radial position deviation becomes inhomogeneous:

$$x'' + k_x x = h(s)\delta. \tag{2.27}$$

The solution is the linear combination of a particular solution of the inhomogeneous equation and the characteristic solution of the homogeneous equation

$$x(s) = x_0 c_x(s) + x_0' s_x(s) + \delta d_x(s). \tag{2.28}$$

The function $d_x(s)$ can be determined using Green's function G:

$$d_x(s) = \int_0^s h(\bar{s})G_x(s, \bar{s})d\bar{s} \tag{2.29}$$

$$= \int_0^s h(\bar{s})\left(s_x(s)c_x(\bar{s}) - c_x(s)s_x(\bar{s})\right)d\bar{s}. \tag{2.30}$$

The solution is the dispersion function, which is only non-zero when curvature $h \neq 0$ (e.g. in bending magnets). Usually curvature in dipoles is constant, so $h(s) = h$

$$d_x(s) = \frac{h}{k_x}\left(1 - \cos\left(\sqrt{k_x}s\right)\right) \tag{2.31}$$

and for negative k

$$d_x(s) = \frac{h}{|k_x|}\left(\cosh\left(\sqrt{|k_x|}s\right) - 1\right) \tag{2.32}$$

or for $k = 0$

$$d_x(s) = 0\,. \tag{2.33}$$

The dispersion is a measure for the momentum dependency of the transversal focusing and is only caused by dipole fields.

2.2.3. Longitudinal characteristic solution

Considering now the dynamics of particles with an initial longitudinal position deviation l_{initial} and a velocity v. The change of this positional deviation after the design particle has traveled the distance s, results from the difference in flight time of the particle to the design particle

$$\Delta l = l_{\text{final}} - l_{\text{initial}} = v_0(t_0 - t) = v_0\left(\frac{s}{v_0} - \frac{S}{v}\right) = -(S - s) + \frac{\Delta v}{v}S \tag{2.34}$$

where $\Delta v = v - v_0$. The path length difference $S - s$ can be calculated using the line element

$$S - s = \int_0^s \left(\sqrt{x^2 + y^2 + (1 + hx)^2} - 1\right)d\bar{s}\,. \tag{2.35}$$

In linear approximation only contributions from the field of deflecting magnets are considered.

$$S - s = \int_0^s h(\bar{s})x(\bar{s})d\bar{s} \tag{2.36}$$

$$= x_0\int_0^s h(\bar{s})c_x(\bar{s})d\bar{s} + x_0'\int_0^s h(\bar{s})s_x(\bar{s})d\bar{s} + \delta\int_0^s h(\bar{s})d_x(\bar{s})d\bar{s} \tag{2.37}$$

and with $\frac{\Delta v}{v}S \approx \frac{\Delta v}{v_0}s = \frac{1}{\gamma^2}\frac{\Delta p}{p_0}s = \frac{s}{\gamma^2}\delta$ follows

$$l(s) = -x_0\int_0^s h(\bar{s})c_x(\bar{s})d\bar{s} - x_0'\int_0^s h(\bar{s})s_x(\bar{s})d\bar{s}$$
$$+ l(0) - \delta\left(\int_0^s h(\bar{s})d_x(\bar{s})d\bar{s} - \frac{s}{\gamma^2}\right)\,. \tag{2.38}$$

The effect on $l(0)$ is only existing in the presence of a curvature $h \neq 0$ (e.g. in bending magnets). Assuming a constant curvature $h(s) = h$, the solution for $k_x > 0$ is

$$l(s) = -x_0\frac{h\sin(\sqrt{k_x}s)}{\sqrt{k_x}} - x_0'\frac{h(1 - \cos(\sqrt{k_x}s))}{k_x}$$
$$+ l(0) + \delta\left(\frac{s}{\gamma^2} - h^2\left(\frac{s}{k_x} - \frac{\sin(\sqrt{k_x}s)}{k_x^{3/2}}\right)\right) \tag{2.39}$$

and for negative k_x

$$l(s) = -x_0\frac{h\sinh(\sqrt{k_x}s)}{\sqrt{k_x}} - x_0'\frac{h(1 - \cosh(\sqrt{k_x}s))}{k_x}$$
$$+ l(0) + \delta\left(\frac{s}{\gamma^2} - h^2\frac{\sinh\left(\sqrt{|k_x|}s\right) - \sqrt{|k_x|}s}{|k_x|^{3/2}}\right)\,. \tag{2.40}$$

2.2.4. Matrix formalism

The solutions (2.24), (2.25), (2.26), (2.31), (2.32), (2.33) and (2.39) of the first order equations of motion can be formulated in matrix notation for different optical elements. In this way an initial position of a particle in the 6-dimensional phase space $\boldsymbol{\pi}_0 = (x_0, x'_0, y_0, y'_0, l_0, \delta_0)$ can be assigned the point in the phase space which the particle occupies after passing through the element.

$$
\begin{pmatrix} x \\ x' \\ y \\ y' \\ l \\ \delta \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} & R_{13} & R_{14} & R_{15} & R_{16} \\ R_{21} & R_{22} & R_{23} & R_{24} & R_{25} & R_{26} \\ R_{31} & R_{32} & R_{33} & R_{34} & R_{35} & R_{36} \\ R_{41} & R_{42} & R_{43} & R_{44} & R_{45} & R_{46} \\ R_{51} & R_{52} & R_{53} & R_{54} & R_{55} & R_{56} \\ R_{61} & R_{62} & R_{63} & R_{64} & R_{65} & R_{66} \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ x'_0 \\ y_0 \\ y'_0 \\ l_0 \\ \delta_0 \end{pmatrix} \tag{2.41}
$$

The matrix R_{ij} is called transfer matrix and must be developed for each optical element.

Considering a particle with the position in the phase space $\boldsymbol{\pi}_0$, which first crosses the element a and then b with the transfer matrices $\boldsymbol{R}_a$ and $\boldsymbol{R}_b$, it finally takes the position

$$
\boldsymbol{\pi} = \boldsymbol{R}_b \cdot (\boldsymbol{R}_a \cdot \boldsymbol{\pi}_0) = (\boldsymbol{R}_b \cdot \boldsymbol{R}_a) \cdot \boldsymbol{\pi}_0 \, . \tag{2.42}
$$

In this way, also a complex optical system, consisting of n elements can be described by the product of the transfer matrices of its components

$$
\boldsymbol{\pi} = (\prod_n \boldsymbol{R}_n) \boldsymbol{\pi}_0 \, . \tag{2.43}
$$

If the determinant of the transfer matrices $\det(\boldsymbol{R}) \neq 0$, the matrix is invertible and initial phase space vectors can be calculated from final states backwards. If the matrix also satisfies the equation (2.44), it is called symplectic, which correspond to canonical transformations.

$$
\boldsymbol{R}^T \boldsymbol{M} \boldsymbol{R} = \boldsymbol{M} \quad \text{where} \quad \boldsymbol{M}^T = \boldsymbol{M}^{-1} = -\boldsymbol{M} \quad , \text{e.g.} \quad \boldsymbol{M} = \begin{pmatrix} \mathbf{0} & \mathbf{1} \\ -\mathbf{1} & \mathbf{0} \end{pmatrix} \, . \tag{2.44}
$$

The first order matrices for the magnetic mid-plane, presented in section 2.3, are always symplectic, therefore Liouville's theorem applies in all cases. Thus, the phase space volume occupied by a particle ensemble remains constant below these transformation $\boldsymbol{R}$.

2.2.5. The second order equations of motion

The approximation up to the first order is in many cases sufficient for the description of the phenomena, e.g. for simple transfer lines. However, in ring accelerators it is necessary also to correct for chromaticity, which arises from different values of the betatron phase advance produced by particles with an energy deviation. Also in linear accelerators, sextupole fields can help to compensate for over or under focusing of particles with deviating energies. A development up to the first order is not sufficient to be able to describe these phenomena. In order to determine the second order equations of motion, the development of the magnetic field up to the second order (2.12) must be substituted into the Lorentz equation (2.15).

The result is a system of coupled second order differential equations:

$$x'' + h^2(1 - n) = h\delta + (2n - 1 - \beta)h^3 x^2 + h'xx' + \frac{1}{2}hx'^2$$

$$+ (2 - n)h^2 x\delta + \frac{1}{2}(h'' - nh^3 + 2\beta h^3)y^2$$

$$+ h'yy' - \frac{1}{2}hy'^2 - h\delta^2, \tag{2.45}$$

$$y'' + nh^2 y = 2(\beta - n)h^3 xy + h'xy'$$

$$- h'x'y + hx'y' + nh^2 y\delta. \tag{2.46}$$

The solution of this system of differential equations can be found in a same way as the first order solutions, which is by multiplying with Green's function and subsequent integration. The result can be expressed using the corresponding first order matrix R_{ij} and a second order matrix T_{ijk}:

$$x_i(s) = \sum_{i=1}^{6} R_{ij}(s)x_j(0) + \sum_{k=1}^{6}\sum_{j=1}^{k} T_{ijk}(s)x_j(0)x_k(0). \tag{2.47}$$

The complete list of all components R_{ij} and T_{ijk} for the mid plane can be found in [6].

2.3. Accelerator components and transfer matrices

Subsequently, the first order transfer matrices for those elements required for the present work are derived. These matrices are also the basis of the common ion optics programs as (Transport [7], Mirko [8], MADX [9], TraceWin [10] or WinAgile [11])

2.3.1. Drift space

In a field free drift space $h = \rho^{-1} = 0$ and therefore $k_x = k_y = 0$. The dispersion is zero and using initial conditions $x_0 = x(0)$ and $x_0' = x'(0)$ the transverse solution (equation (2.26)) becomes

$$x(s) = x_0 + x_0's \qquad\qquad x'(s) = 0 + x_0'$$

$$y(s) = y_0 + y_0's \qquad\qquad y'(s) = 0 + y_0' \tag{2.48}$$

and the longitudinal part (2.39) becomes

$$l(s) = l_0 + \frac{s}{\gamma^2}\delta_0. \tag{2.49}$$

Since the momentum is not changed $\delta(s) = \delta_0$, the transfer matrix for a drift space of length L is

$$\boldsymbol{R}_0 = \begin{pmatrix} 1 & L & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & L & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \frac{L}{\gamma^2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.50}$$

2.3.2. Quadrupole magnets

In a quadrupole magnet (Figure 2.4) the nominal path is a straight line and so the radial acceleration disappears in equation (2.15). The solution for the radial component is thus equivalent to the axial one with a correspondingly changed definition of $k_x = -k_y = -nh^2$. For a quadrupole the general focusing strength $k = k_x = -k_y$ is introduced, given in $[1/m^2]$. The quadrupole gradient g is given in $[T/m]$ and describes the gradient of the transverse

Figure 2.4.: Examples of quadrupole magnets. (left: quadrupole doublet from the transfer channel UNILAC to SIS18, centre: quadrupole doublet from high energy beam line, right: quadrupole doublet from Cryring local injector line)

magnetic field which corresponds to the focusing strength of a quadrupole [12]:

$$g = \frac{\partial B_y}{\partial x} = \frac{\partial B_x}{\partial y} = B' \,. \tag{2.51}$$

The focusing strength is the rigidity normalised gradient

$$k = -nh^2 = \frac{1}{(B\rho)}g \,. \tag{2.52}$$

For a quadrupole of length L the integral field gradient kL corresponds to the inverse focus length, analogous to an optical lens:

$$kL = \frac{1}{f_{\text{foc}}} \,. \tag{2.53}$$

Depending on the sign of the focusing strength there are two possible transfer matrices for quadrupole magnets. Strictly speaking, there is a third one, namely for $k = 0$, which ultimately corresponds to a drift space of length L.

Horizontal focusing quadrupole

For $k > 0$ (2.24), the quadrupole is horizontal focusing and vertical defocusing. The longitudinal solutions for $l(s)$ and $\delta(s)$ are similar to a drift space. So, using $\chi = \sqrt{k}L$,

the transfer matrix for a horizontal focusing quadrupole of length L can be written as

$$
\boldsymbol{R}_{\mathrm{F}} = \begin{pmatrix}
\cos\chi & \frac{1}{\sqrt{k}}\sin\chi & 0 & 0 & 0 & 0 \\
-\sqrt{k}\sin\chi & \cos\chi & 0 & 0 & 0 & 0 \\
0 & 0 & \cosh\chi & \frac{1}{\sqrt{k}}\sinh\chi & 0 & 0 \\
0 & 0 & \sqrt{k}\sinh\chi & \cosh\chi & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & \frac{L}{\gamma^2} \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}. \tag{2.54}
$$

Vertical focusing quadrupole

For $k < 0$ (2.25), the quadrupole is vertical focusing and horizontal defocusing. So for the transfer matrix, the planes are just exchanged and with $\chi = \sqrt{|k|}L$

$$
\boldsymbol{R}_{\mathrm{D}} = \begin{pmatrix}
\cosh\chi & \frac{1}{\sqrt{|k|}}\sinh\chi & 0 & 0 & 0 & 0 \\
\sqrt{k}\sinh\chi & \cosh\chi & 0 & 0 & 0 & 0 \\
0 & 0 & \cos\chi & \frac{1}{\sqrt{|k|}}\sin\chi & 0 & 0 \\
0 & 0 & -\sqrt{|k|}\sin\chi & \cos\chi & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & \frac{L}{\gamma^2} \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}. \tag{2.55}
$$

2.3.3. Dipole magnets

A horizontal sector dipole is a pie-shaped bending magnet (Figure 2.5) of length L with a fixed curvature $h = \frac{1}{\rho}$ and a constant bending angle $\alpha = \frac{L}{\rho} = hL$.

Figure 2.5.: Examples of sector magnets. (left: 7.5° bending dipole of the GSI fragment separator FRS, right: 3.3° super conducting dipole of the SIS100 synchrotron)

Weak focusing dipole magnets

The constants of the equation of motion (2.22) are $k_x = (1 - n)h^2$ and $k_y = nh^2$. For positive k_x and k_y, that means $0 \le n \le 1$ and the dipole is called a weak focusing dipole. The longitudinal solution is given with equation (2.39) and using $\xi = \sqrt{1 - n}$ and $\psi = \sqrt{n}$

the transfer matrix can be written as

$$
\boldsymbol{R}_{\mathrm{B}} =
\begin{pmatrix}
\cos(\xi\alpha) & \frac{\rho\sin(\xi\alpha)}{\xi} & 0 & 0 & 0 & \frac{\rho(1-\cos(\xi\alpha))}{\xi^2} \\
-\frac{\xi\sin(\xi\alpha)}{\rho} & \cos(\xi\alpha) & 0 & 0 & 0 & \frac{\sin(\xi\alpha)}{\xi} \\
0 & 0 & \cos(\psi\alpha) & \frac{\rho\sin(\psi\alpha)}{\psi} & 0 & 0 \\
0 & 0 & -\frac{\psi\sin(\psi\alpha)}{\rho} & \cos(\psi\alpha) & 0 & 0 \\
-\frac{\sin(\xi\alpha)}{\xi} & -\frac{\rho(1-\cos(\xi\alpha))}{\xi^2} & 0 & 0 & 1 & \frac{L}{\gamma^2} - \rho\left(\frac{\alpha}{\xi^2} - \frac{\sin(\xi\alpha)}{\xi^{3/2}}\right) \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} .
$$

Homogeneous dipole magnets

A homogeneous sector dipole has no field gradient $\frac{\partial B_y}{\partial x} = 0 \Rightarrow n = 0$. In this case the constants of the equation of motion become $k_x = \frac{1}{\rho^2}$ and $k_y = 0$ and by using $\lim_{\psi\to 0}(\sin(\psi\alpha)/\psi) = \alpha$ the matrix $\boldsymbol{R}_{\mathrm{B}}$ simplifies with $\xi = 1$, $\psi = 0$ to

$$
\boldsymbol{R}_{\mathrm{B}_0} =
\begin{pmatrix}
\cos(\alpha) & \rho\sin(\alpha) & 0 & 0 & 0 & \rho(1-\cos(\alpha)) \\
-\frac{sin(\alpha)}{\rho} & \cos(\alpha) & 0 & 0 & 0 & \sin(\alpha) \\
0 & 0 & 1 & L & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
-\sin(\alpha) & -\rho(1-\cos(\alpha)) & 0 & 0 & 1 & \frac{L}{\gamma^2} - \rho(\alpha - \sin(\alpha)) \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} . \qquad (2.56)
$$

Strong focusing dipole magnets

With a field index $|n| > 1$ one speaks of a strong focusing system. In synchrotrons also dipoles with an alternating and high field gradient can be used, to achieve a focusing in both transverse planes without the need for additional quadrupole magnets [13]. This concept has been realised in the Alternating Gradient Synchrotron at the Brookhaven National Laboratory, USA (AGS) [14] (Figure 2.6) and with the Proton Synchrotron at CERN. In case of $(n > 1) \iff (k_x < 0) \wedge (k_y > 0)$ and with $\xi = \sqrt{|1-n|}$ and $\psi = \sqrt{|n|}$
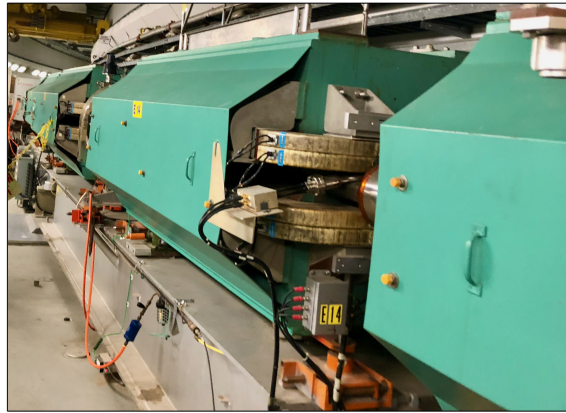

Figure 2.6.: Strong focusing dipole of the Alternating Gradient Synchrotron at the Brookhaven National Laboratory, USA

the transfer matrix for strong focusing dipole magnets is

$$
\boldsymbol{R}_{\mathrm{B}^+} = \begin{pmatrix}
\cosh(\xi\alpha) & \frac{\rho\sinh(\xi\alpha)}{\xi} & 0 & 0 & 0 & \frac{\rho(1-\cosh(\xi\alpha))}{\xi^2} \\
-\frac{\xi\sinh(\xi\alpha)}{\rho} & \cosh(\xi\alpha) & 0 & 0 & 0 & \frac{\sinh(\xi\alpha)}{\xi} \\
0 & 0 & \cos(\psi\alpha) & \frac{\rho\sin(\psi\alpha)}{\sqrt{n}} & 0 & 0 \\
0 & 0 & -\frac{\psi\sin(\psi\alpha)}{\rho} & \cos(\psi\alpha) & 0 & 0 \\
-\frac{\sinh(\xi\alpha)}{\xi} & -\frac{\rho(1-\cosh(\xi\alpha))}{\xi^2} & 0 & 0 & 1 & \frac{L}{\gamma^2}-\rho\left(\frac{\sinh(\xi\alpha)}{\xi^3}-\frac{\alpha}{\xi^2}\right) \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$

and for $(n<0) \iff (k_x>0)\wedge(k_y<0)$

$$
\boldsymbol{R}_{\mathrm{B}^-} = \begin{pmatrix}
\cos(\xi\alpha) & \frac{\rho\sin(\xi\alpha)}{\xi} & 0 & 0 & 0 & \frac{\rho(1-\cos(\xi\alpha))}{\xi^2} \\
-\frac{\xi\sin(\xi\alpha)}{\rho} & \cos(\xi\alpha) & 0 & 0 & 0 & \frac{\sin(\xi\alpha)}{\xi} \\
0 & 0 & \cosh(\psi\alpha) & \frac{\rho\sinh(\psi\alpha)}{\psi} & 0 & 0 \\
0 & 0 & -\frac{\psi\sinh(\psi\alpha)}{\rho} & \cosh(\psi\alpha) & 0 & 0 \\
-\frac{\sin(\xi\alpha)}{\xi} & -\frac{\rho(1-\cos(\xi\alpha))}{\xi^2} & 0 & 0 & 1 & \frac{L}{\gamma^2}-\rho\left(\frac{\alpha}{\xi^2}-\frac{\sin(\xi\alpha)}{\xi^3}\right) \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} .
$$

Edge focusing of brick shaped dipole magnets

In case of a brick shaped dipole magnet (Figure 2.7), the entrance and exit of the dipole are rotated by the angle $\beta = \frac{\alpha}{2}$ with respect to the transverse plane. For a particle with radial

Figure 2.7.: Two examples of brick shaped dipole magnets. (left: GSI high energy transfer line, right: transfer channel TK to SIS18)

position deviation x_0 the path within the dipole field is therefore shorter by $x_0(\tan\beta)$ at each edge. This corresponds to a positive angular kick

$$
\Delta x' = \frac{\tan(\beta)}{\rho}x_0 . \tag{2.57}
$$

Using the approximation of an infinitely extended field edge and integrating over a straight path from inside the magnet to outside on the rotated z-axis in a coordinate system rotated

by angle β, a radial displacement leads to a negative angular kick

$$\Delta y' = \frac{1}{B\rho} \int B_x dz = -y_0 \frac{\tan(\beta)}{\rho} \tag{2.58}$$

However, due to the finite nature of the boundary field in reality, the effect of the edge angle β must be corrected. For this purpose an effective edge angle is defined:

$$\beta_{\text{eff}} = \beta - \frac{g}{\rho} \frac{1 + \sin^2(\beta)}{\cos(\beta)} K \, , \tag{2.59}$$

where g is the pole shoe distance and K depends on the shape of the iron edge. So this resulting kick can be represented by the matrix

$$\boldsymbol{R}_\beta = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{\tan(\beta)}{\rho} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{\tan(\beta_{\text{eff}})}{\rho} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} . \tag{2.60}$$

The matrix form of a brick shaped dipole magnet can be constructed by combining the usual dipol matrix with the edge matrix at the entrance and the exit

$$\boldsymbol{R}_{\text{B}_\beta} = \boldsymbol{R}_\beta \cdot \boldsymbol{R}_{\text{B}} \cdot \boldsymbol{R}_\beta \, . \tag{2.61}$$

2.3.4. Solenoids
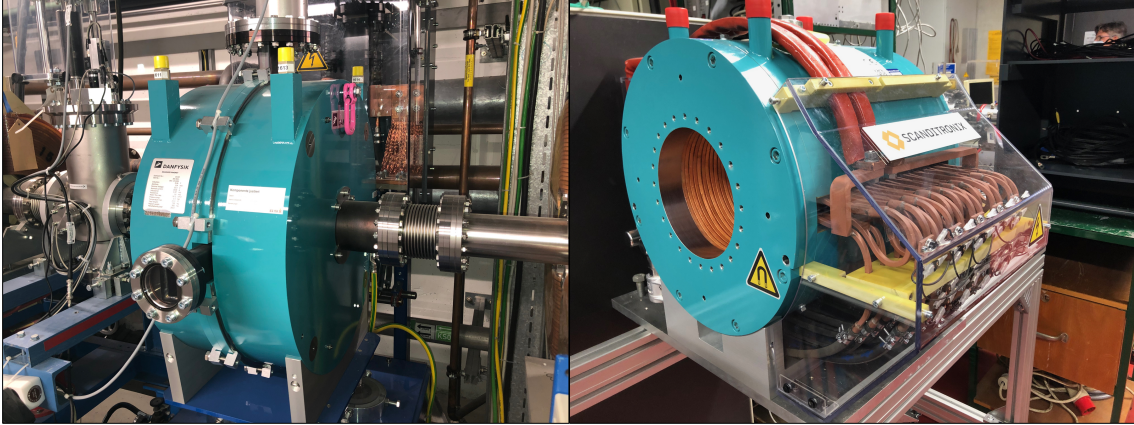
Figure 2.8.: Examples of solenoids (left: a $0.96\,$T solenoid from TK, right: a solenoid for the electron gun of a future SIS18 electron lens [15])

Solenoids are optical elements with a longitudinal magnetic field (Figure 2.8). Since this means that the transverse velocity components v_t of the particles are perpendicular to the magnetic field B_s, particles with such components are forced onto a circular path with the radius r around the magnetic field lines:

$$\gamma m \frac{v_t^2}{r} = q v_t B_s \, . \tag{2.62}$$

After passing a distance L in the solenoid, the angle α is

$$\alpha = -\frac{v_t t}{r} = -\frac{qB_s}{\gamma m}\frac{L}{v_z} \,. \tag{2.63}$$

In linear approximation $p_s \approx p_z = \gamma m v_z$ and with equation (2.9) follows

$$\alpha = -\frac{B_s L}{(B\rho)} \,. \tag{2.64}$$

This transverse rotation is achieved with the matrix

$$\boldsymbol{R}_{\mathrm{rot}} = \begin{pmatrix} 1 & \frac{L\sin\alpha}{\alpha} & 0 & -\frac{L(1-\cos\alpha)}{\alpha} & 0 & 0 \\ 0 & \cos\alpha & 0 & -\sin\alpha & 0 & 0 \\ 0 & \frac{L(1-\cos\alpha)}{\alpha} & 1 & \frac{L\sin\alpha}{\alpha} & 0 & 0 \\ 0 & \sin\alpha & 0 & \cos\alpha & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \,. \tag{2.65}$$

However, an additional transverse effect is added to the rotation due to the Lorentz force in the marginal field of the solenoid. This effect results from the longitudinal velocity v_z in the radial field component B_r. Because of the rotational symmetry $B_x = \frac{x}{r} B_r$ and $B_y = \frac{y}{r} B_r$. Gaussian law applies to the magnetic field

$$0 = \nabla \cdot \boldsymbol{B} = \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 2\frac{B_r}{r} + \frac{\partial B_z}{\partial z} \tag{2.66}$$

and so

$$B_r = -\frac{r}{2}\frac{\partial B_z}{\partial z} \,. \tag{2.67}$$

Using the approximation that the boundary field only affects the change of direction, but not the positions,

$$\Delta x' = -\frac{1}{B\rho}\int B_y dz = \pm y\frac{B_s}{2(B\rho)} = \pm yK \,, \tag{2.68}$$

$$\Delta y' = +\frac{1}{B\rho}\int B_x dz = \mp x\frac{B_s}{2(B\rho)} = \mp xK \,. \tag{2.69}$$

For the boundary field at input $(+)$ or output $(-)$

$$\boldsymbol{R}_{\pm} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \pm K & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \mp K & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \,. \tag{2.70}$$

With $K = \frac{B_s}{2(B\rho)} = \frac{-\alpha}{2L}$, $C = \cos(KL)$ and $S = \sin(KL)$ the combined transfer matrix for

a solenoid can be written

$$
\boldsymbol{R}_{\mathrm{S}} = \boldsymbol{R}_{+} \cdot \boldsymbol{R}_{\mathrm{rot}} \cdot \boldsymbol{R}_{-} =
\begin{pmatrix}
C^2 & SCK^{-1} & SC & S^2K^{-1} & 0 & 0 \\
-KSC & C^2 & -KS^2 & SC & 0 & 0 \\
-SC & -S^2K^{-1} & C^2 & SCK^{-1} & 0 & 0 \\
KS^2 & -SC & -KSC & C^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & \frac{s}{\gamma^2} \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} . \tag{2.71}
$$

2.3.5. Gabor lens

Space charge lenses, introduced 1947 by D. Gabor [16], are electron traps that can also be used for radial focusing of ion beams (Figure 2.9).
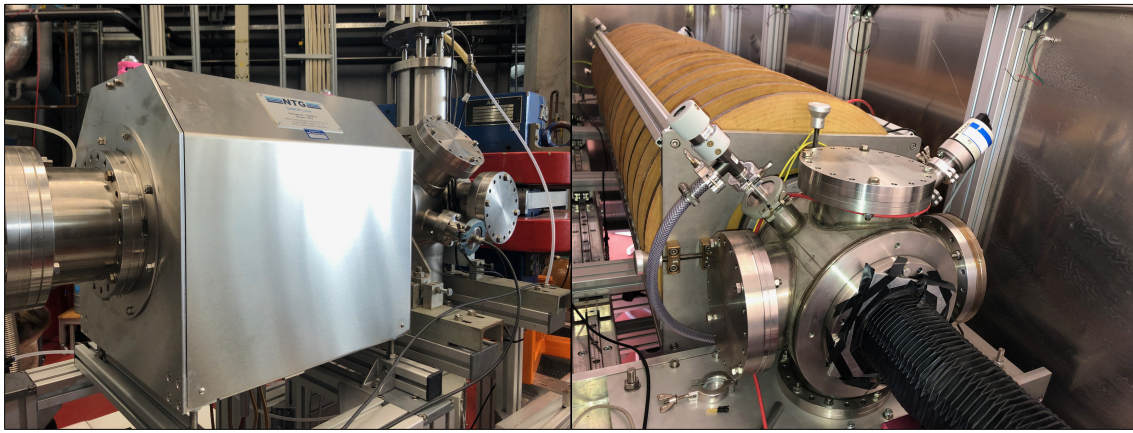
Figure 2.9.: Two gabor lenses (left: the 40 cm long GL9000, right: the 2 m lens GL2000).

Here a solenoid field is superimposed on an electrostatic potential to confine a homogeneously distributed electron cloud (Figure 2.10).
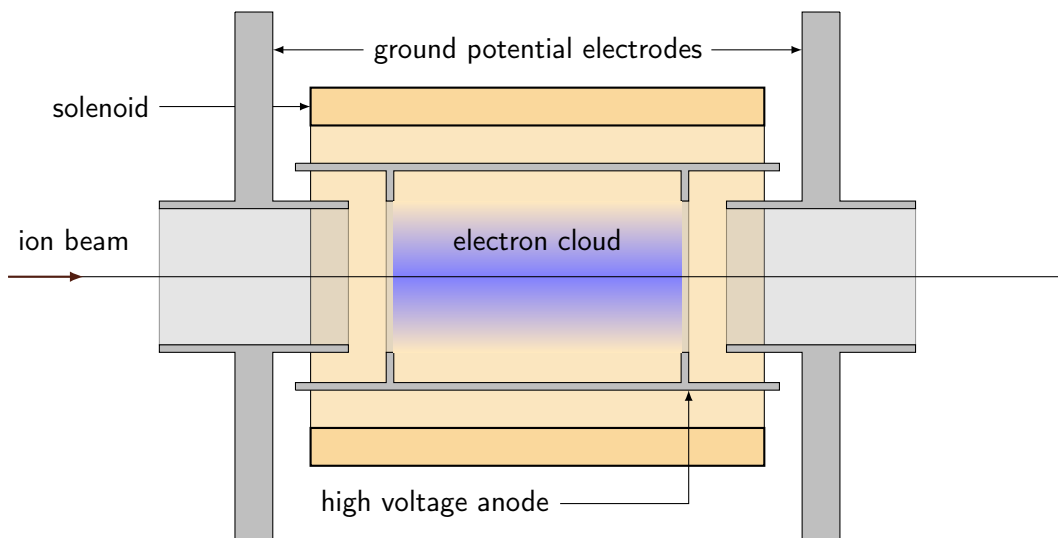
Figure 2.10.: Schematic representation of a Gabor lens [17, page 5]. For high field Gabor lenses, instead of a solenoid, a pair of Helmholtz coils at the entrance and exit of the device, is used to generate the longitudinal magnetic field.

In the case of positive ion beams, the space charge of the confined electron cloud causes an overcompensation of the ion beam's space charge force and the resulting cylinder symmetrical potential has a focusing effect on the beam in both transverse planes [17]. So ion beams can be transported and focused with high quality at relatively low field strengths of $30 - 200\,\mathrm{mT}$ [18].

For the functionality of such a lens it is very important that the uniform electron density and hence the linearity in radial force can be achieved in reality. The loss and confinement conditions for a volume element in the Gabor lens are a function of the magnetic and electric fields [18]. If the minimum magnetic field is not reached, then no electrons are confined. The confinement is then ultimately limited by the Brillouin flow [19]. Here it is important to set the conditions so that the electrons have constant radial drift velocities.

For the derivation of the transfer matrix, the effects of the solenoid field and the external electric field are not taken into account, since they are approximately negligible due to the low field strength, which is even more valid for relativistic beams. Thus only the field of the confined electrons is considered (Figure 2.11). In a first approximation, the electron
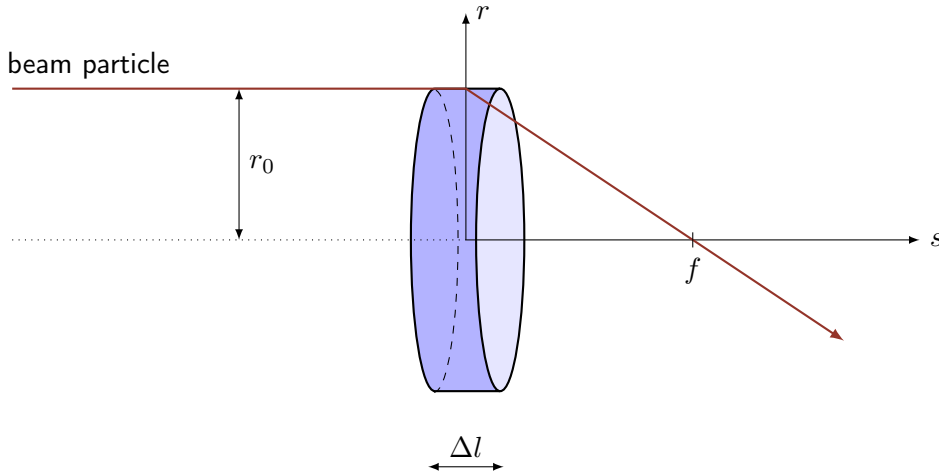
Figure 2.11.: The effect of an electron cloud on a charged particle. In first approximation the electrons are in equilibrium and the electron density distribution can be considered as homogeneous ($\eta = \mathrm{const}$) and cylindrically symmetric around the beam axis. A small longitudinal section (Δl) of this cylinder has then the effect of a thin lens.

cloud is now to be approximated as an infinitely long cylinder with a homogeneous charge distribution. The integral charge enclosed by a small cylinder segment of length dl with radius r_0 is $Q = \eta e \pi r_0^2 \Delta l$, where η is the electron density and e is the elementary charge. The electric field has only a radial component $\vec{E} = E_r \vec{e}_r$. So, following Gauss's law

$$\oint \vec{E} d\vec{A} = \frac{Q}{\epsilon_0}\,, \tag{2.72}$$

the electric flux through the cylinder surface is

$$E_r \cdot 2\pi r_0 dl = \frac{\eta e \pi r_0^2 \Delta l}{\epsilon_0}\,. \tag{2.73}$$

The radial electric field is therefore

$$E_r = \frac{\eta e r_0}{2\epsilon_0} \,. \tag{2.74}$$

Considering a particle with positive charge q, which has a longitudinal velocity component v_s and a radial momentum p_r, then a radial electric field causes a Lorentz force

$$qE_r = F_r = \frac{dp_r}{dt} = \frac{v_s dp_r}{\Delta l} \tag{2.75}$$

and so the change of radial momentum of the charged particle is

$$dp_r = \frac{qE_r \Delta l}{v_s} = \frac{q\eta r_0 \Delta l}{2\epsilon_0 \beta c} \,. \tag{2.76}$$

From geometry follows

$$\frac{r_0}{f} = \frac{dp_r}{p_s} \tag{2.77}$$

and so the focusing strength is

$$\frac{1}{f} = \frac{\eta e \Delta l}{2\epsilon_0 \beta c} \frac{q}{p_s} = \frac{\eta e \Delta l}{2\epsilon_0} \frac{q}{\gamma \beta^2 m_0 c^2} \,. \tag{2.78}$$

With (2.9) the result can be expressed as a function of the magnetic rigidity of the beam:

$$\frac{1}{f} = \frac{\eta e \Delta l}{2\epsilon_0 \beta c} \frac{1}{(B\rho)} \,. \tag{2.79}$$

Since $1/f$ is always negative for positively charged beams and

$$-\frac{1}{f} = \frac{\Delta x'}{x} = \frac{\Delta y'}{y} \,, \tag{2.80}$$

a thin Gabor lens has the effect of a thin lens with focal length f in both transversal planes

$$\boldsymbol{R}_{\mathrm{G}^*}(\Delta l) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{f} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \,. \tag{2.81}$$

The effect of a long Gabor lens with length L can be approximated by N thin lens kicks of length $\Delta l = \frac{L}{N}$ alternating with field free drifts

$$\boldsymbol{R}_{\mathrm{G}} = \prod_{n=1}^{N} \boldsymbol{R}_0\left(\frac{1}{2}\Delta l\right) \cdot \boldsymbol{R}_{\mathrm{G}^*}(\Delta l) \cdot \boldsymbol{R}_0\left(\frac{1}{2}\Delta l\right) \,. \tag{2.82}$$

The Gabor lens GL9000 (Figure 2.9) has a length of $L = 40\,\mathrm{cm}$ and shall achieve an electron density of at least $\eta = 1 \times 10^{15}/\mathrm{m}^3$. For a proton beam with an energy of $10\,\mathrm{MeV/u}$, this results in a focal length of $f = 5.54\,\mathrm{m}$ according to (2.78) with $\Delta l = L$. To get a sense of how the accuracy evolves with number of drift-kick-drift elements, Table 2.1 lists the

corresponding values of three matrix elements for an increasing number of slices.

Table 2.1.: Value of the resulting matrix elements R_{11}, R_{21}, R_{22} as a function of the number of thin lenses. Already at a division of 10, the error of R_{21} is only in the order of 10^{-5} and the uncertainty of the focal length goes down to the millimetre range.

number of slices N	R_{11}	R_{22}	$R_{21} = -\frac{1}{f}$	f
1	0.96387935232	0.39277587046	-0.18060323842	5.54 m
2	0.96404243996	0.39459821161	-0.17897236193	5.59 m
10	0.96409413094	0.39517719865	-0.17845793465	5.60 m
15	0.96409532467	0.39519057747	-0.17844606874	5.60 m
100	0.96409625808	0.39520103897	-0.17843679093	5.60 m
1000	0.96409627934	0.39520127736	-0.17843657952	5.60 m

2.3.6. Accelerating gap

Interaction with an alternating radio frequency (RF) field in an accelerating gap is typically used to accelerate charged particles in modern particle accelerators. Since the magnetic component of the Lorentz force (2.1) is perpendicular to the velocity vector of the particles, only[1] electric fields can be used for longitudinal acceleration. The particles are accelerated when the varying electric field in the gap is in the right phase as they move through the gap. During the rest of the cycle, they are shielded from the then decelerating field in field-free regions inside of drift tubes. To allow efficient acceleration of the entire beam,
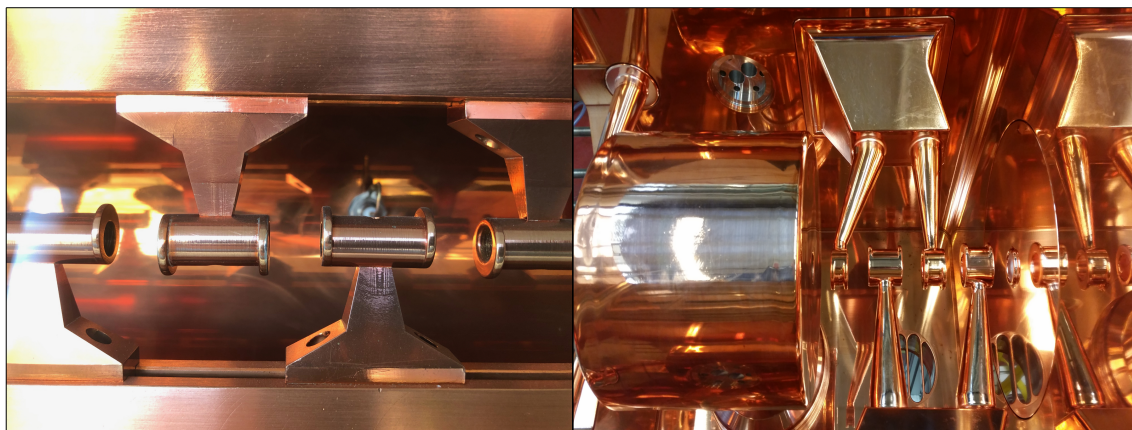
Figure 2.12.: (left) Accelerating gaps and drift tubes of an IH-type linear accelerator. (right) IH tank with internal quadrupole lens [22].

the beam must be divided into particle bunches. These bunches must have a longitudinal separation equal to an integer multiple of $\beta\lambda$ (or resp. $\beta\lambda/2$ for H-Type structures). This principle is used in RF accelerators and buncher resonators of the crossbar H-mode (CH) or interdigital H-mode (IH) type (Figure 2.12), in conventional Alvarez-type accelerators, as well as in non-resonant variable-frequency cavities commonly used in synchrotrons.

---

[1] The law of induction provides another possibility for acceleration via varying magnetic fields. Changes of the magnetic flux density lead to an electric curl field perpendicular to the variable magnetic field, which can also be used for acceleration. This principle is the basis for induction LINACS [20, 21].

It can be shown that the particle trajectory in a gap can be approximately described by a discontinuous parameter change in the gap centre. The fundamental derivations for the particle dynamics are based on the description of linear accelerators from P. Lapostolle [23] and its adoptions in the lectures of T. P. Wangler [24].

Energy gain within a gap

Energy can be transferred to a particle from an electromagnetic standing wave in an RF cavity. The energy gain of a particle on axis ($r = 0$) with charge q in a gap of length g and with constant voltage $U_0 = U(r = 0)$ is

$$\Delta W = qU_0 \,. \tag{2.83}$$

Since the electric field also penetrates into the field free drift tube (Figure 2.13), the path integral between the field free areas must be calculated.
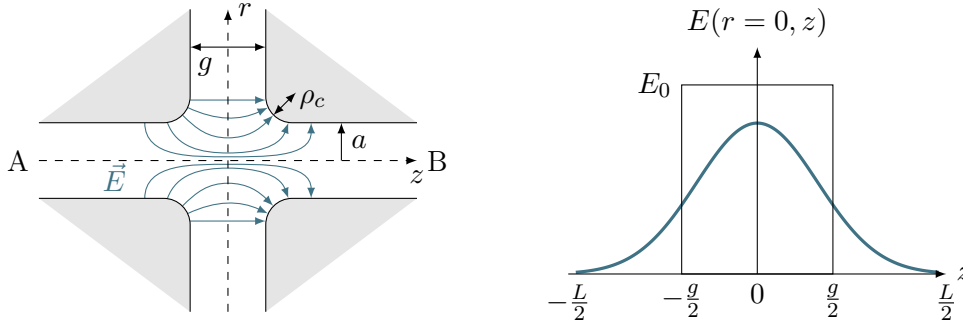
Figure 2.13.: (left) Schematic representation of an accelerating gap of length g with a drift tube radius a. (right) Longitudinal electric field in the gap [23, page 16].

$$U_0 = \int_{-\frac{L}{2}}^{\frac{L}{2}} E_z(r = 0, z)dz \,. \tag{2.84}$$

In RF-accelerators time varying electric fields are used. The time dependent electric field along the axis is given by

$$E_z(r = 0, z, t) = E_z(r = 0, z) \cos\left(\omega t(z) + \phi\right) \,. \tag{2.85}$$

The time $t = 0$ is chosen such that a particle experiences the maximum acceleration in the middle of the gap at this time. If $\phi = 0$, the particle arrives in the middle of the gap at time $t = 0$, if ϕ is negative, it arrives earlier. If $\phi \neq \left(\frac{\pi}{2} + n\pi\right) \quad \forall\, n \in \mathbb{Z}$, the energy gain within the gap is

$$\begin{aligned}
\Delta W &= q \int_{-\frac{L}{2}}^{\frac{L}{2}} E_z(r = 0, z) \cos\left(\omega t(z) + \phi\right)dz \\
&= q \cos\left(\phi\right) \int_{-\frac{L}{2}}^{\frac{L}{2}} E_z(r = 0, z)(\cos\left(\omega t(z)\right) - \sin\left(\omega t(z)\right) \tan\left(\phi\right))dz \,.
\end{aligned} \tag{2.86}$$

Using 2.84 this can be written as $\Delta W = qU_0T_0 \cos\left(\phi\right)$. T_0 is called the transit time factor and expresses the decrease of the energy gain during the particle transit caused by the

time variation of the electric field,

$$T_0 = \frac{\int_{-\frac{L}{2}}^{\frac{L}{2}} E_z(r = 0, z) \cos\left(\omega t(z)\right) dz}{U_0} - \tan\phi \frac{\int_{-\frac{L}{2}}^{\frac{L}{2}} E_z(r = 0, z) \sin\left(\omega t(z)\right) dz}{U_0} . \tag{2.87}$$

Due to the time dependence $T_0 \cos\left(\phi\right) < 1$, the energy gain ΔW is always smaller than the energy gain in a constant dc field seen by the particle at the gap centre qU_0, which is called the transit time effect. Usually one defines an average axial electric-field amplitude $E_0 = \frac{U_0}{L}$ and the energy gain can be expressed in form of Panofsky's equation [25]

$$\Delta W = qE_0 T_0 L \cos\left(\phi\right). \tag{2.88}$$

The quantity $E_0 T$ is called the accelerating gradient and for a multi cell cavity, the natural choice for L is the geometric cell length. Using the Lorentz transformation $W = \gamma mc^2$ with $\Delta W = W_\mathrm{f} - W_\mathrm{i} = \gamma_\mathrm{f} mc^2 - \gamma_\mathrm{i} mc^2$ one obtains for the Lorentz-factor of the synchronous particle ($\phi = \phi_s$) after acceleration in the gap

$$\gamma_\mathrm{f} = \gamma_\mathrm{i} + \frac{qE_0 T_0 L \cos(\phi_s)}{mc^2} . \tag{2.89}$$

The transit time factor

Approximately $E(z)$ is an even function about the geometric centre of the gap and so

$$\int_{-\frac{L}{2}}^{\frac{L}{2}} E_z(r = 0, z) \sin(\omega t(z)) dz = 0 \tag{2.90}$$

and the transit factor simplifies to

$$T_0 = \frac{\int_{-\frac{L}{2}}^{\frac{L}{2}} E_z(r = 0, z) \cos\left(\omega t(z)\right) dz}{U_0} . \tag{2.91}$$

With $\omega t(z) \approx \frac{\omega z}{v_z} = \frac{2\pi z}{\beta\lambda}$, where $\beta\lambda = \frac{v_z}{c}\frac{1}{f}$ is the distance, a particle travels within one RF period at frequency f.

$$T_0 = \frac{\int_{-\frac{L}{2}}^{\frac{L}{2}} E_z(r = 0, z) \cos\left(\frac{2\pi z}{\beta\lambda}\right) dz}{U_0} \tag{2.92}$$

Assuming the field has a square profile and $E_z(r = 0, z) = E_0$ within the gap of length g and zero outside (see Figure 2.13) the transit factor simplifies to

$$T_0 = \frac{\sin\left(\frac{\pi g}{\beta\lambda}\right)}{\frac{\pi g}{\beta\lambda}} . \tag{2.93}$$

To achieve maximum energy gain, T_0 must be equal to 1, which corresponds to minimising the gap distance because of $\lim_{x\to 0}\frac{\sin(x)}{x} = 1$. In reality, a certain distance must be maintained to avoid arcing. For the gap factor, an additional empirical correction can be made to account for the finite drift tube edge radius ρ_c [24], introducing an effective gap

$$g_c = g + 0.85\rho . \tag{2.94}$$

Solution of Maxwell's equations in the gap

The electric field on the axis can be expressed as a Fourier integral. The parameter L can be chosen freely as long as the electric field disappears at this location. This is especially the case for $L \to \infty$ and so

$$U_0 T(k_z) = \int_{-\infty}^{\infty} E_z(z,0) \cos(k_z z) dz \tag{2.95}$$

with the Fourier transformation

$$E_z(z, r = 0) = \frac{U_0}{2\pi} \int_{-\infty}^{\infty} T(k_z) \cos(k_z z) dk_z . \tag{2.96}$$

These equations can be extended with the transit-time factor and the energy gain of off-axis particles. In circular symmetry, electromagnetic fields can then be expressed in form of modes and for synchronous waves, only TM or E modes interact with the particle. So $T(k_z)$ can be considered as the amplitude function with wave number k_z in the spectrum of a stationary wave. Assuming azimuthal symmetry and in a paraxial approximation, a complete solution of Maxwell's equations for the electromagnetic field in the gap can be derived [24]. The components of the electric field are

$$E_z(r, z, t) = \frac{U_0}{2\pi} \int_{-\infty}^{\infty} T(k_z) I_0(k_r r) \cos(k_z z) \cos(\omega t + \phi) dk_z , \tag{2.97}$$

$$E_r(r, z, t) = \frac{U_0}{2\pi} \int_{-\infty}^{\infty} T(k_z) \frac{k_z}{k_r} I_1(k_r r) \sin(k_z z) \cos(\omega t + \phi) dk_z , \tag{2.98}$$

$$E_\phi(r, z, t) = 0 \tag{2.99}$$

and the components of the magnetic field are

$$B_z(r, z, t) = 0 , \tag{2.100}$$

$$B_r(r, z, t) = 0 , \tag{2.101}$$

$$cB_\phi(r, z, t) = \frac{U_0}{2\pi} \int_{-\infty}^{\infty} T(k_z) \frac{\omega}{ck_r} I_1(k_r r) \cos(k_z z) \sin(\omega t + \phi) dk_z , \tag{2.102}$$

where $I_i(k_r r)$ are the Bessel functions resulting from the cylindrical symmetry.

Longitudinal focusing

For the energy and phase change in the gap, one finally obtains

$$\Delta W = + qVT(k) I_0(k_r r) \cos(\phi)$$
$$+ qV \frac{\partial}{\partial k} \Big(T(k) k_r I_1(k_r r) \Big) r' \sin(\phi) , \tag{2.103}$$

$$\Delta \phi = + \frac{qV}{2W} k \frac{\partial}{\partial k} \Big(T(k) I_0(k_r r) \Big) \sin(\phi)$$
$$- \frac{qV}{2W} k \frac{\partial^2}{\partial k^2} \Big(T(k) k_r I_1(k_r r) \Big) r' \cos(\phi) \tag{2.104}$$

where

$$k = \frac{\omega}{\beta c} \tag{2.105}$$

and

$$k_r^2 = \left(\frac{\omega}{\beta_s c}\right)^2 - \left(\frac{\omega}{c}\right)^2 . \tag{2.106}$$

Transverse focusing

The equations for the transversal plane are

$$\Delta r' = -\frac{qV}{2W}\frac{k_r}{k}T(k)I_1(k_r r)\sin(\phi)$$
$$+ \frac{qV}{2W}\left(\frac{\partial}{\partial k}\left(T(k)\frac{k_r^2}{k}I_1'(k_r r)\right) - T(kI_0(k_r r))\right)r'\cos(\phi), \tag{2.107}$$

$$\Delta r = -\frac{qV}{2W}\frac{\partial}{\partial k}\left(T(k)\frac{k_r}{k}I_1(k_r r)\right)\cos(\phi)$$
$$- \frac{qV}{2W}\left(\frac{\partial^2}{\partial k^2}\left(T(k)\frac{k_r^2}{k}I_1'(k_r r)\right) - \frac{\partial}{\partial k}\left(T(k)I_0(k_r r)\right)\right)r'\sin(\phi). \tag{2.108}$$

An exemplary derivation of the specific effect of a gap on the radial momentum coordinate can be found in appendix A.2.

The transfer matrixes

Equation (A.17), the solutions for transverse position and those for the longitudinal plane can also be represented in matrix notation. Here the formulation is shown as it is also used in the TraceWin [10] documentation:

$$\boldsymbol{R}_{\text{gap}} = \begin{pmatrix} k_1 C & 0 & 0 & 0 & 0 & 0 \\ \frac{k_{xy}}{\beta_f\gamma_f} & k_2 C & 0 & 0 & 0 & 0 \\ 0 & 0 & k_1 C & 0 & 0 & 0 \\ 0 & 0 & \frac{k_{xy}}{\beta_f\gamma_f} & k_2 C & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \frac{k_z}{\beta_f\gamma_f} & \frac{\beta_i\gamma_i}{\beta_f\gamma_f} \end{pmatrix} \tag{2.109}$$

where

$$k_1 = 1 - \frac{qE_0TL\cos(\phi_s)}{2mc^2\beta_s^2\gamma_s^3}(\gamma_s^2 + kT'/T),$$
$$k_2 = 1 - \frac{qE_0TL\cos(\phi_s)}{2mc^2\beta_s^2\gamma_s^3}(\gamma_s^2 - kT'/T),$$
$$k_{xy} = -\frac{q\pi E_0TL\sin(\phi_s)}{mc^2\beta_s^2\gamma_s^2\lambda},$$
$$k_z = \frac{2q\pi E_0TL\sin(\phi_s)}{mc^2\beta_s^2\lambda} \tag{2.110}$$

and $C = \sqrt{\frac{\beta_i\gamma_i}{\beta_f\gamma_f}\frac{1}{k_1 k_2}}$. The combined matrix for a complete cell of length L is

$$\boldsymbol{R}_{\text{cell}} = \boldsymbol{R}_0(L/2)\boldsymbol{R}_{\text{gap}}\boldsymbol{R}_0(L/2) \tag{2.111}$$

and for a tracking simulation it is necessary that besides the phase space coordinates of the particles also the momentum of the bunch at the position of the gap centre is recorded.

2.4. Emittance and phase space ellipsoid

Since the particles of a bunch in phase space are all located in the neighbourhood of the central particle, it is always possible to define an ellipsoid in phase space, whose centre is the position of the central particle and which encloses all particles. The general equation for a 6 dimensional ellipsoid can be written as

$$\frac{x^2}{x_{\text{max}}^2} + \frac{x'^2}{x'^2_{\text{max}}} + \frac{y^2}{y_{\text{max}}^2} + \frac{y'^2}{y'^2_{\text{max}}} + \frac{l^2}{l_{\text{max}}^2} + \frac{\delta^2}{\delta_{\text{max}}^2} = 1 \,. \tag{2.112}$$

Since the ellipse can also be tilted, the more general form for conic sections must be used. A conic section arises as the intersection of a plane ($\alpha x + \beta y + \gamma z + \delta = 0$) with a double cone ($x^2 + y^2 = \epsilon z^2$) . This results in an equation that has the following form:

$$ax^2 + 2bxy + cyx^2 + dx + ey = 1 \,. \tag{2.113}$$

Applied to horizontal phase space coordinates (x, x'), this is

$$ax^2 + 2bxx' + cx'^2 + dx + ex' = 1 \,. \tag{2.114}$$

In case of $ac - b^2 > 0$, the equation describes an ellipse. Since the centre of the ellipse is at the coordinate origin and because of rotational symmetry, the equation for $(-x, -x')$ must also be satisfied, if it is satisfied for (x, x'). This is generally only the case when $d = 0$ and $e = 0$. So eventually remains

$$ax^2 + 2bxx' + cx'^2 = 1 \,. \tag{2.115}$$

This equation can also be expressed in matrix notation:

$$\begin{pmatrix} x & x' \end{pmatrix} \cdot \begin{pmatrix} a & b \\ b & c \end{pmatrix} \cdot \begin{pmatrix} x \\ x' \end{pmatrix} = 1 \,. \tag{2.116}$$

For ellipses, the determinant must be positive, so the matrix is invertible and can be written as the inverse of a matrix that is usually called the sigma matrix $\boldsymbol{\sigma}_x$

$$\begin{pmatrix} x & x' \end{pmatrix} \cdot \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}^{-1} \cdot \begin{pmatrix} x \\ x' \end{pmatrix} = 1 \,. \tag{2.117}$$

Executing the inversion leads to

$$\begin{pmatrix} x & x' \end{pmatrix} \cdot \frac{1}{\det \boldsymbol{\sigma}_x} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{pmatrix} \cdot \begin{pmatrix} x \\ x' \end{pmatrix} = 1 \,, \tag{2.118}$$

which can be expanded to

$$\sigma_{22}x^2 - 2\sigma_{12}xx' + \sigma_{11}x'^2 = \det \boldsymbol{\sigma}_x = \epsilon_x^2 \,. \tag{2.119}$$

The area A_x surrounded by the phase ellipse corresponds to the emittance and is equal to $A_x = \pi \epsilon_x = \pi \sqrt{\det \boldsymbol{\sigma}_x}$. The Emittance is given in units of mm · mrad and is conserved under canonical transformations.

Figure 2.14 shows the corresponding maximum beam width and the maximum angular divergence. Under the assumption of a Gaussian distribution of the particles around the
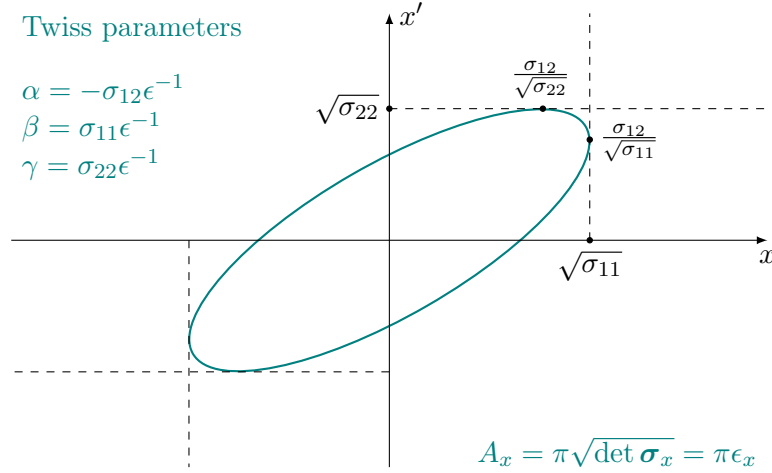

Figure 2.14.: The phase space ellipse in horizontal plane. In case of a 2 dimensional gaussian density distribution, the edge of the ellipse can be understood as the level curve of the 1-σ emittance, if 39.3% of the particles are located within the ellipsoid. The emittance-normalised Twiss parameters are commonly used to characterise optical systems.

origin in the phase space, an ellipse $\boldsymbol{\sigma}_x$ can be defined in such a way that its edge encloses 39.3% of all particles. The resulting emittance then corresponds to the 1-σ emittance $\epsilon_x^{1\sigma}$. The 2-σ-emittance, which includes 86.5% of the particles, and the 3-σ-emittance, which includes 98.9% of the particles, can be derived accordingly.

$$\epsilon_x^{2\sigma} = 4\epsilon_x^{1\sigma} \tag{2.120}$$

$$\epsilon_x^{3\sigma} = 9\epsilon_x^{1\sigma} \tag{2.121}$$

In this case, the values $x_{\max}$ correspond to the standard deviation of the projection on the x-axis. With a horizontal profile grid this can be measured directly. Analogously, a phase space ellipsoid can be generated for all 6 dimensions,

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} & \sigma_{45} & \sigma_{46} \\ \sigma_{15} & \sigma_{25} & \sigma_{35} & \sigma_{45} & \sigma_{55} & \sigma_{56} \\ \sigma_{16} & \sigma_{26} & \sigma_{36} & \sigma_{46} & \sigma_{56} & \sigma_{66} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}_{xx} & \boldsymbol{\sigma}_{xy} & \boldsymbol{\sigma}_{xz} \\ \boldsymbol{\sigma}_{yx} & \boldsymbol{\sigma}_{yy} & \boldsymbol{\sigma}_{yz} \\ \boldsymbol{\sigma}_{zx} & \boldsymbol{\sigma}_{zy} & \boldsymbol{\sigma}_{zz} \end{pmatrix}. \tag{2.122}$$

With magnetic mid-plane symmetry, the density distributions in the horizontal subspace are approximately decoupled from the density distributions in the vertical and longitudinal subspace. The sigma matrix thus has the form

$$\boldsymbol{\sigma} \approx \begin{pmatrix} \boldsymbol{\sigma}_{xx} & 0 & \boldsymbol{\sigma}_{xz} \\ 0 & \boldsymbol{\sigma}_{yy} & 0 \\ \boldsymbol{\sigma}_{zx} & 0 & \boldsymbol{\sigma}_{zz} \end{pmatrix}. \tag{2.123}$$

2.4.1. Transformation of the phase space ellipsoid

An initial phase space vector $\boldsymbol{\pi}_0 = (x_0, x_0', y_0, y_0', l_0, \delta_0)$ is transformed from an element, represented by the matrix $\boldsymbol{R}$, according to $\boldsymbol{\pi} = \boldsymbol{R}\,\boldsymbol{\pi}_0$. For a particle distribution, which is initially described by the phase space ellipse $\boldsymbol{\sigma}_0$, the equation (2.117) can also be written in the form

$$\boldsymbol{\pi}_0^T \cdot \hat{\sigma}_0^{-1} \cdot \boldsymbol{\pi}_0 = 1\,. \tag{2.124}$$

Using conversion rules for matrices, the following transformations can be made

$$\boldsymbol{\pi}_0^T \cdot [\boldsymbol{R}^T (\boldsymbol{R}^T)^{-1}] \cdot \boldsymbol{\sigma}_0^{-1} \cdot [\boldsymbol{R}^{-1}\boldsymbol{R}] \cdot \boldsymbol{\pi}_0 = 1\,, \tag{2.125}$$

$$[\boldsymbol{R} \cdot \boldsymbol{\pi}_0]^T \cdot [(\boldsymbol{R}^T)^{-1}\boldsymbol{\sigma}_0^{-1}\boldsymbol{R}^{-1}] \cdot [\boldsymbol{R} \cdot \boldsymbol{\pi}_0] = 1\,, \tag{2.126}$$

$$[\boldsymbol{R} \cdot \boldsymbol{\pi}_0]^T \cdot [\boldsymbol{R}\,\boldsymbol{\sigma}_0\boldsymbol{R}^T]^{-1} \cdot [\boldsymbol{R} \cdot \boldsymbol{\pi}_0] = 1\,, \tag{2.127}$$

$$\boldsymbol{\pi}^T \cdot [\boldsymbol{R}\,\boldsymbol{\sigma}_0\boldsymbol{R}^T]^{-1} \cdot \boldsymbol{\pi} = 1\,, \tag{2.128}$$

$$\boldsymbol{\pi}^T \cdot \boldsymbol{\sigma}^{-1} \cdot \boldsymbol{\pi} = 1\,. \tag{2.129}$$

The phase space ellipsoid can be transformed using the same transfer matrices as used for the phase space vectors

$$\boldsymbol{\sigma} = \boldsymbol{R}\,\boldsymbol{\sigma}_0\boldsymbol{R}^T\,. \tag{2.130}$$

2.5. Space charge

In the previous sections, space charge effects were not considered, but especially at low particle energies or high intensities, the defocusing effect has a significant impact on particle dynamics.

To determine the electromagnetic field resulting from the positions and relative velocities of all n particles of a bunch and the resulting Lorentz force on each particle, an n-body problem would have to be solved, which is not analytically possible. Also, the summation approximation by smaller time steps is practically not feasible in reasonable time for a larger number of particles as usually encountered in modern high-intensity particle accelerators.

2.5.1. Particle-in-cell

One way to address the space charge problem is to use the particle-in-cell method as usually applied in plasma physics [26]. Here, a grid is placed over the phase space. For each time step, the charge and momentum of each particle are then distributed to macro particles, which are placed at the intersection points of the grid lines. The distribution is made according to the relative position of the particle to the nearest macro particles of the surrounding cell. From then on, only the fields of the macro particles are considered. The resulting electromagnetic field at each intersection point is calculated and the corresponding effect of the Lorentz force on the macro particle momentum is then redistributed again to the particles, according to their relative position to the macro particles of the surrounding cell.

To apply this method to beam dynamics, the usual matrix calculation is first performed for a small step Δs. Then a coordinate transformation into a Cartesian or a cylindrical coordinate system is performed, in which the Particle-In-Cell calculations are then carried out. Afterwards the back transformation takes place.

The number of macro particles can then be balanced with the computation time by increasing the number of grid points and decreasing the step size Δs as needed. Very accurate results can be obtained in this way, but for the investigations to be made in this work, a very large number of particle tracking simulations must be performed in a short time. Therefore, at first a fast method that can be easily integrated into the known matrix formalism is desirable.

2.5.2. The Kapchinskij-Vladimirskij (KV) envelop equation

For a first estimation, the beam can be described as a cylinder with homogeneous charge distribution [27]. A cylinder with particle density η generates a defocusing radial electric field

$$E_r = q\frac{\eta}{2\epsilon_0}r \,. \tag{2.131}$$

The particle current density generates an azimuthal magnetic field

$$B_\phi = q\frac{\eta}{2\epsilon_0}\frac{v}{c^2}r \,. \tag{2.132}$$

The resulting radial Lorentz force is

$$F_r = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) = \frac{q^2\eta}{2\epsilon_0\gamma^2}r \,. \tag{2.133}$$

Ionisation of the residual gas results in partial or complete compensation of the space charge component due to the generated electrons accumulating in the beam potential. For this purpose, the neutralisation factor $0 \leq f_e \leq 1$ is introduced. In case of complete compensation, this value is 1.

$$F_r = \frac{q^2\eta}{2\epsilon_0\gamma^2}(1 - \gamma^2 f_e)r \tag{2.134}$$

Using the space charge density as an expression of beam cross section πa and current I,

$$q\eta = \frac{I}{\beta c\pi a^2} \,, \tag{2.135}$$

the space charge force can be written

$$F_r = \frac{qI}{2\epsilon_0\beta c\gamma^2\pi a^2}(1 - \gamma^2 f_e)r \,. \tag{2.136}$$

In a field-free drift section

$$F_r = m\gamma\frac{d^2r}{dt^2} = m\gamma(\beta c)^2 r'' \,. \tag{2.137}$$

The equation of motion is therefore

$$r'' - \frac{K}{a^2}r = 0 \,, \tag{2.138}$$

where K is called the generalised perveance

$$K = \frac{I}{I_0}\frac{2}{\beta^3\gamma^3}(1 - \gamma^2 f_e) \tag{2.139}$$

and I_0 the characteristic current

$$I_0 = \frac{4\pi\epsilon_0 mc^3}{q}\,.\tag{2.140}$$

In general, the beam is not cylindrical and the beam envelopes are a function of s. By I. Kapchinskij and V. Vladimirskij the problem was solved for a special phase space distribution [28], where the particles are homogeneously distributed on the surface of a four-dimensional phase space ellipsoids with semi-major and semi-minor axes $r_x(s)$ and $r_y(s)$. In this case, as well, an unbunched beam is assumed and the charge density is

$$q\eta = \frac{I}{\beta c\pi r_x(s) r_y(s)}\,.\tag{2.141}$$

Analogous to the previous considerations, the KV equation of motion can be derived. Combining the corresponding terms with the general solution of the mid-plane (2.22), one obtains the equations of motion in linear approximation

$$x'' + k_x - \frac{2K}{r_x(s)[r_x(s) + r_y(s)]}x = 0\,,\tag{2.142}$$

$$y'' + k_y - \frac{2K}{r_y(s)[r_x(s) + r_y(s)]}y = 0\,.\tag{2.143}$$

Sacherer [29] and Lapostolle [30] were able to show that the KV envelope equation can also be applied to other charge distributions with elliptic symmetry (for example, Gaussian, parabolic, and hollow) if one expresses the semi-axes by the corresponding 2σ-RMS values

$$r_x(s) = 2\sigma_x(s)\,,\tag{2.144}$$

$$r_y(s) = 2\sigma_y(s)\,.\tag{2.145}$$

In practice, the general equation of the mid-plane is solved for an element and then the space charge matrices are applied equivalently to the drift-kick method by dividing the element into smaller sections of length L. The corresponding 2D-KV matrix depends on the modified envelopes on the position s and can be written as

$$\boldsymbol{R}_{\mathrm{KV2D}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{2KL}{2\sigma_x(2\sigma_x+2\sigma_y)} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{2KL}{2\sigma_y(2\sigma_x+2\sigma_y)} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}\,,\tag{2.146}$$

which is equivalent to the effect of a de-focusing thin lens.

2.5.3. The 3 dimensional envelope equation

For bunched beams, the longitudinal component cannot be neglected and the cylindrical approximation is not suitable anymore. Therefore the effect of the bunch length must also be considered. The prerequisite in this case is also a particle distribution with ellipsoidal

symmetry. Using the 3D-space charge parameter and the approximation for $f(s)$,

$$K_3 = \frac{3qI\lambda}{20\sqrt{5}\pi\epsilon_0 mc^3\beta^2\gamma^3} \quad \text{and} \quad f(s) \approx \frac{\sqrt{\sigma_x\sigma_y}}{3\gamma\sigma_z}\,, \tag{2.147}$$

the RMS-envelope equations for 3 dimensions can be derived [3] to

$$(\sigma_x)'' + \sigma_x k_x - \frac{K_3(1-f(s))}{(\sigma_x+\sigma_y)\sigma_z} - \frac{\epsilon_x^2}{\sigma_x^3} = 0\,, \tag{2.148}$$

$$(\sigma_y)'' + \sigma_y k_y - \frac{K_3(1-f(s))}{(\sigma_x+\sigma_y)\sigma_z} - \frac{\epsilon_y^2}{\sigma_y^3} = 0\,, \tag{2.149}$$

$$(\sigma_z)'' + \sigma_z k_z - \frac{K_3 f(s)}{\sigma_x\sigma_y} - \frac{\epsilon_z^2}{\sigma_z^3} = 0\,. \tag{2.150}$$

The corresponding space charge matrix for a drift space of length L is

$$\boldsymbol{R}_{\text{KV3D}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{(1-f(s))K_3 L}{(\sigma_x+\sigma_y)\sigma_x\sigma_z} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{(1-f(s))K_3 L}{(\sigma_x+\sigma_y)\sigma_y\sigma_z} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \frac{f(s)K_3 L}{\sigma_x\sigma_x\sigma_z} & 1 \end{pmatrix}\,, \tag{2.151}$$

which is equivalent to a de-focusing thin lens in 3 dimensions. Thus, the defocusing effect of the space charge increases linearly with the current. At high energies, however, the effect plays less and less of a role, since it decreases with $\beta^{-2}\gamma^{-3}$. It can also be seen that the space charge force is particularly strong in regions of small beam cross sections (e.g. in focal points).

3. Optimisation methods

The optimisation of a particle accelerator usually has many degrees of freedom, since several parameters are available for the adjustment, depending on the function of the corresponding accelerator section. In the case of transfer lines, this includes the field strengths of the installed optical elements or the phases and amplitudes of buncher cavities. On the other hand, the quality of the settings can be assessed by means of the respective beam diagnostics. These are, for example, beam positions, beam widths, beam currents or the measured data of loss monitors or phase probes.

For further considerations an accelerator section is now abstracted as a black box, representing a mapping of n setting parameters $x_1 \ldots x_n$ to m observables $y_1 \ldots y_m$. The problem of accelerator optimisation is thus equivalent to the search for extrema in nonlinear functions $f : \mathbb{R}^n \to \mathbb{R}^m$, with the difficulty that the function f is noisy and generally not analytically known. In addition, the extremum generally cannot be reached in all m dimensions at the same time, so that from a certain degree of optimisation, a further optimisation of one observable can only be bought with a deterioration in another. A so-called Pareto front is formed (Figure 3.1).
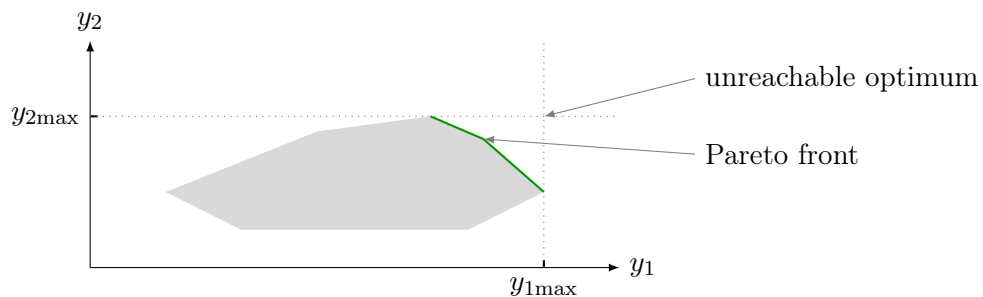
Figure 3.1.: Schematic representation of a Pareto front (green) for a 2-dimensional set of values. For both observables the maximum can only be reached at the expense of the other. The optimum is not part of the set.

One possibility is the prioritisation of observables by means of weights and thus to select a position on this front. For the transmission optimisation the problem of the Pareto front can be left aside, because with the transmission $T(\boldsymbol{x}) = I(\boldsymbol{x})/I_0$ only one single observable is needed. Thus f is reduced to a function $f : \mathbb{R}^n \to \mathbb{R}$, which in the following will be referred to as fitness function F and must be minimised

$$F(\boldsymbol{x}) = 1 - T(\boldsymbol{x}) + \epsilon(\boldsymbol{x}) . \tag{3.1}$$

The term $\epsilon(\boldsymbol{x})$ is either zero or represents the parametrisation of further optimisation goals.

3.1. Exhaustive search

For beam lines with few optical elements, a parameter scan over the entire parameter space is generally possible. However, its dimension increases with each additional element. In this context the question of a reasonable step size arises. If the step size is chosen too coarse, the optimum may be missed. If, on the other hand, the smallest possible step size is chosen, which corresponds to the precision of the power supply, then, for example, the scan of a $1\,\mathrm{kA}$-power supply in $1\,\mathrm{A}$-steps will directly lead to 1000 values and thus, with n magnets, to 1000^n different settings to be tested. Even a rough 1% scan (step size of $(x_{\max} - x_{\min})/100$) leads to 100^n settings, which must be set in each case at a real accelerator and whose effect must then be measured with at least one beam pulse. For example, with only three quadrupoles at UNILAC (magnets can be pulsed with $50\,\mathrm{Hz}$), it takes $(100^3/50)s = 5.55\,\mathrm{h}$. With four values this takes 100 times as long, i.e. 23 days. So for a general optimisation a scan is not practicable, especially for longer and more complex transfer lines.

3.2. Gradient descent

If the fitness function $F(\boldsymbol{x})$ is differentiable, an extremum can be approximated by the gradient descent algorithm (3.2)

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \gamma \nabla F(\boldsymbol{x}_i)\,. \tag{3.2}$$

With each step $i \to i+1$, the vector $\boldsymbol{x}_i$ moves in the direction of the strongest descent. The step size γ is reduced each time the function value has not decreased after one iteration step. In accelerator control, a gradient could be approximated with a scan over small deviations of all parameters.

If one scans the parameter range of a single steerer or a single quadrupole and measures the change of the transmission behind an elliptical aperture diaphragm placed at a fixed distance from the optical element (Figure 3.2), one may well find a definite maximum. However, a quadrupole doublet already has two local transmission maxima (Figure 3.3).
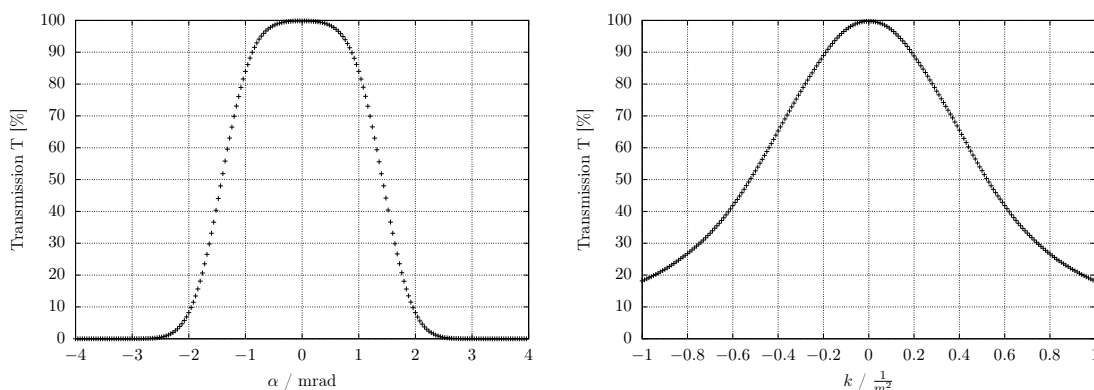

Figure 3.2.: Results of single parameter scans with a matrix tracking code (on the left: dependence of transmission on steerer angle α_s, on the right: quadrupole scan). In both cases the element is followed by a drift line and an elliptical aperture limitation. Both scans show a global transmission maximum.
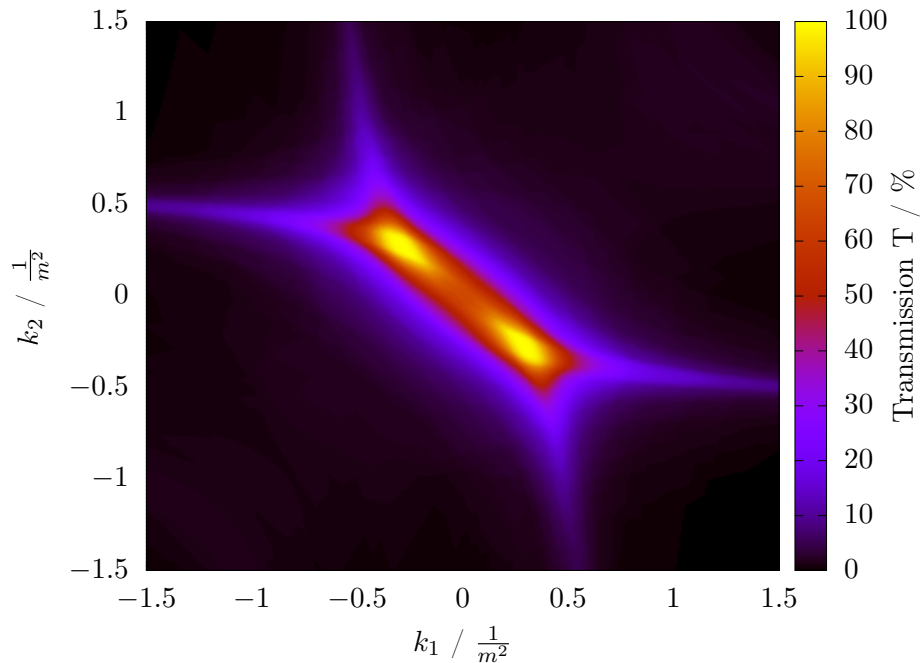
Figure 3.3.: The result of a doublet scan (simulation). There are two extrema, symmetrical with respect to the quadrupole strengths k_1 and k_2, where the transmission becomes maximum. If the particle distribution of the beam is slightly asymmetric, one of the extrema becomes the global maximum.

The number of extrema potentially increases even more when additional optical elements are used. Also certain inhomogeneous beam distributions can have a corresponding effect. However, gradient methods are generally not suitable for finding global minima. Since this property is necessary for the described optimisation, in the following especially such methods are considered, which are able to escape from local extrema. Furthermore, gradient-free methods are used for discrete, discontinuous, or noisy functions, where the gradient is difficult or impossible to determine.

3.3. Evolutionary algorithms

Evolutionary algorithms are a class of optimisation methods that follow a stochastic, meta-heuristic approach and whose mode of operation is inspired from the natural evolution of living organisms. Due to the stochastic approach, convergence to the global optimum is not guaranteed, however, very good results are shown in practice, especially when the algorithm is specifically adapted to the dedicated problem. For the problem of minimising a single value function on a subspace of $\mathbb{R}^n$, a genetic algorithm with direct real-valued problem representation can be used.

The genetic algorithm

Genetic algorithms (Figure 3.4, Algorithm 1) with binary representation were first described in 1975 by John H. Holland [31]. It was shown that such an algorithm significantly outperforms a random search. David E. Goldberg showed in 1989 [32] that this is also true for real-valued representations if suitable arithmetic operators are used.

Initially, K genomes G_k with N random real-valued genes $g_{k,n}$ are generated. Additionally, a genotype-phenotype mapping $\boldsymbol{x}_k = X(G_k)$ must be defined, with which the genes are mapped into the parameter space (in the case of direct, real-valued problem representation, this is the identity $x_{k,n} = g_{k,n}$). For all genomes, the phenotype is then tested for fitness (in the simplest case of accelerator optimisation, this means loading the corresponding settings into the machine and then measuring the transmission T_k). This results in the value of the fitness function $F_k = 1 - T_k$. The smaller this value is for an individual, the more successful is its phenotype with respect to transmission. The s most unsuccessful genomes are then removed from the population (selection), while the most successful ones are used to construct the next generation. Along with K, the selection criterion s is a parameter of the algorithm.
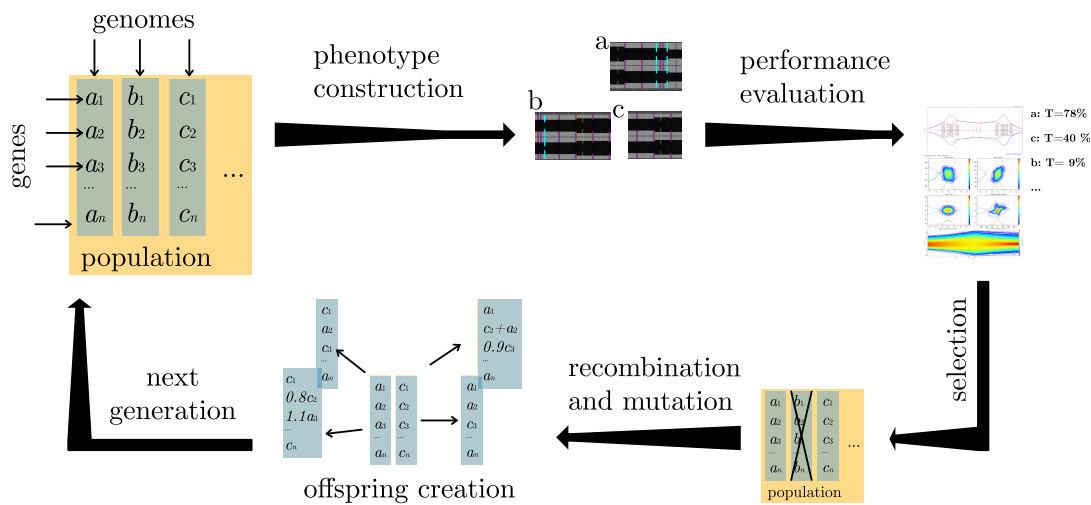

Figure 3.4.: Schematic view of the standard cycle of a genetic algorithm. The process will repeat until a termination criterion is reached.

Various mutation (Figure 3.5) and recombination operators (Figure 3.6) are available for the exploration of the search space. In addition, it is possible to transfer the best genomes directly into the next generation.
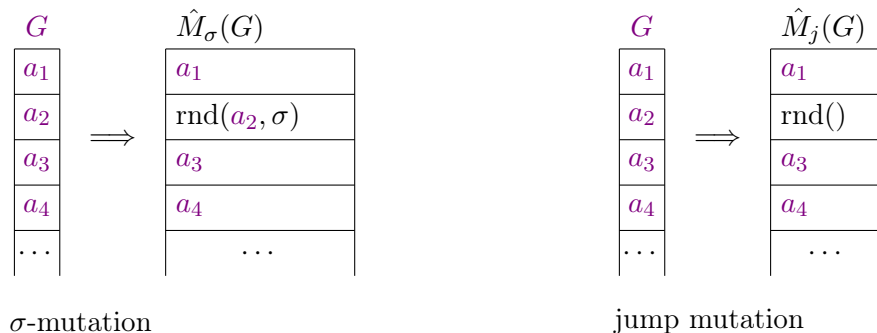

Figure 3.5.: The two mutation operators $\hat{M}$ using the example of gene a_2 in genome G. The function $\mathrm{rnd}(\mu, \sigma)$ returns a random number with the expected value μ and the variance σ^2, the function $\mathrm{rnd}()$ returns a random value equally distributed over the whole definition space. The value σ could itself also be subject to mutation.

The specific performance ultimately depends on the exact choice of parameters and operators. An appropriate genotype-phenotype mapping also plays an important role in this context.
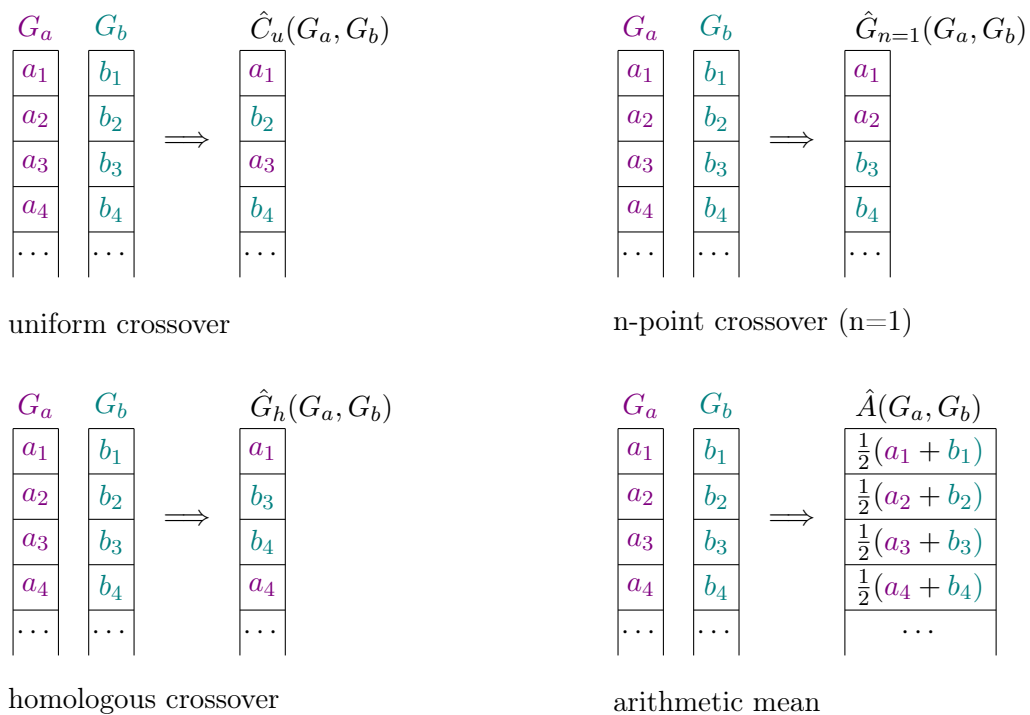

Figure 3.6.: The four basic recombination operators for genetic algorithms with real-valued representation. Additional operators are possible.

3.4. Particle swarm optimisation

The particle swarm optimisation [33] also belongs to the nature-analogue optimisation methods and is inspired by the swarm behaviour of animals (Figure 3.7, Algorithm 2). First, K individuals (resp. swarm particles) are randomly distributed in the n-dimensional search space and each individual is additionally assigned to an initial velocity and thus to a point $(\boldsymbol{x}_k, \boldsymbol{v}_k)$ in the corresponding $2n$-dimensional phase space. In each optimisation step $i \to i+1$, the value of the objective (or fitness) function is evaluated for all individuals, analogous to the genetic algorithm. The velocities are adjusted according to equation (3.3). Subsequently, the swarm particles are shifted by the vector $\boldsymbol{v}_k \Delta t$. The time step Δt is set to 1 for simplicity.

$$\boldsymbol{v}_{k,i+1} = m\boldsymbol{v}_{k,i} + w_c r_1 (\boldsymbol{x}_{k,\min} - \boldsymbol{x}_{k,i}) + w_s r_2 (\boldsymbol{x}_{\min} - \boldsymbol{x}_{k,i}) \tag{3.3}$$

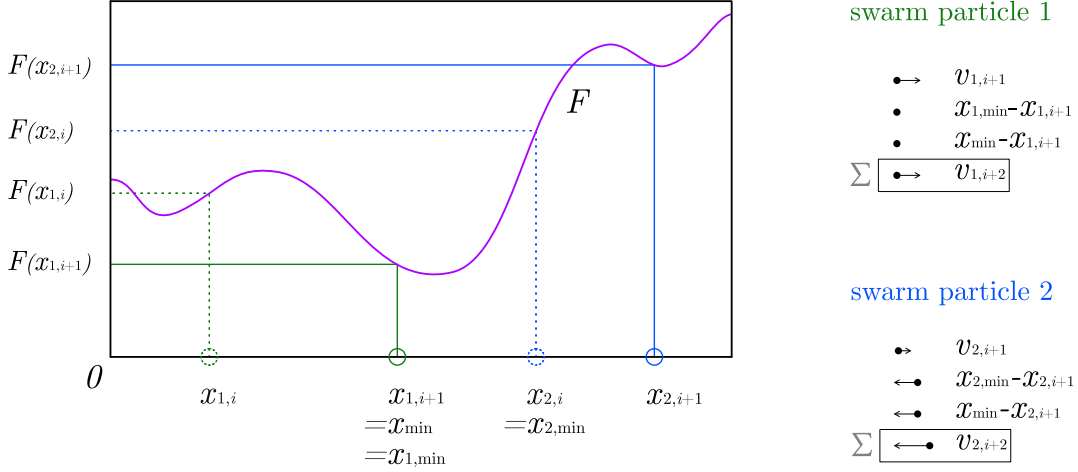|		
m	...	inertia parameters for the motion of the swarm particles
w_c	...	cognitive weighting factor
w_s	...	social weighting factor
r_1, r_2	...	randomly generated weighting factors
$\boldsymbol{x}_{k,\min}$	...	position of best value, seen by k^{th} particle
$\boldsymbol{x}_{\min}$	...	global optimum position, seen so far

Figure 3.7.: Illustration of an optimisation cycle for a swarm with only two particles. The purple line represents the fitness function F over a one-dimensional search space and its minimum is to be found. The various weighting factors have been omitted from this figure for the sake of clarity.

After each step, all positions corresponding to the best fitness value of each particle so far ($F_{k,\min} = F(\boldsymbol{x}_{k,\min})$), as well as the position corresponding to the best fitness value of all particles so far ($F_{\min} = F(\boldsymbol{x}_{\min})$) are stored. In this way, the particles move analogously to a swarm through the parameter space with two varying attractors each, namely their own and the global best value. In this way, the algorithm is also able to escape from local optima and the particles accumulate with a certain probability after some time in the global optimum. Terminal conditions can be defined analogous to the genetic algorithm.

3.5. Simulated annealing

Based on the Metropolis algorithm [34], M. Pincus developed in 1970 an algorithm for minimisation using Monte-Carlo methods [35], which could be improved again decisively by S. Kirkpatrick [36].

The basic principle of simulated annealing (Figure 3.8, Algorithm 3) is based on the behaviour of physical systems which, when cooled, occupy a micro-state with the lowest possible energy.

The probability that a state with energy E is occupied in thermal equilibrium at temperature T is, according to Boltzmann statistics, proportional to

$$p \propto e^{-\frac{E}{k_B T}} . \tag{3.4}$$

The ratio of the probabilities for occupation of two different states E_1 and E_2 is therefore

$$\frac{p_2}{p_1} = e^{\frac{E_2 - E_1}{k_B T}} . \tag{3.5}$$

Analogously, the value of the fitness function for a certain state is interpreted as its energy, which has to be minimised. The Boltzmann constant is set to 1. For each optimisation step a second state $\boldsymbol{x}'$ in the neighbourhood $|\boldsymbol{x}' - \boldsymbol{x}| < r$ is randomly selected. Subsequently,

$F(x'_{\text{max}})$
$F(x)$
$F(x'_{\text{min}})$

The probability of keeping a new position in this range is smaller the larger F'=F(x') is and decreases with falling temperature.

The probability of keeping a new position x' within this area is 1

r r

F

0 x'_{min} x x'_{max}

x is the last particle position
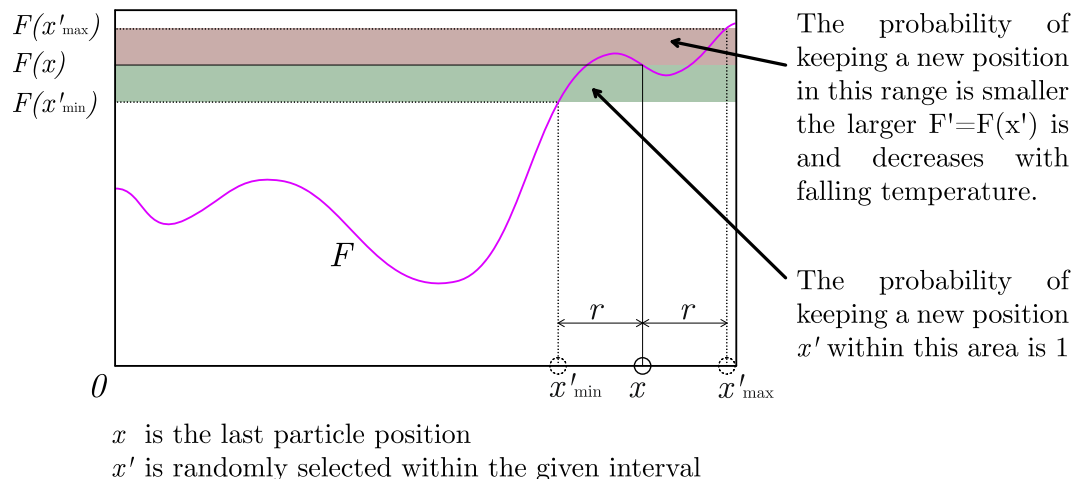x' is randomly selected within the given interval

Figure 3.8.: Illustration of an optimisation cycle of the simulated annealing algorithm for a one-dimensional problem. In practice, multiple particles are used to increase the probability that one of them is already located in a promising region.

the values of the fitness function of both states are evaluated. If $F(\boldsymbol{x'}) < F(\boldsymbol{x})$, then the next optimisation step is performed for the new value $\boldsymbol{x'}$. Otherwise, the next optimisation step for the new value $\boldsymbol{x'}$ is performed only with probability

$$p = e^{-\frac{F(\boldsymbol{x'})-F(\boldsymbol{x})}{T}} \ .$$
(3.6)

S. Kirkpatrick [36] then extended the algorithm in such a way that for the temperature a monotonically falling sequence T_t with $t \in \mathbb{N}$ striving towards 0 is used. So, at the beginning the probability is still quite high that also unfavourable neighbouring energy states are taken. With decreasing temperature this becomes increasingly improbable. In 1994, the method was further developed towards the quantum annihilation algorithm [37], including also tunnelling effects.

3.6. Bayesian optimisation with Gaussian processes

Besides artificial neuronal networks that have gained great reputation in solving a lot of pattern recognition problems in recent years, the Bayesian optimisation [38] is one of the standard machine learning techniques, which is frequently used especially for tuning hyper-parameters or finding structure parameters of artificial neural networks.

The basis for this optimisation is Bayes' theorem (3.7). It relates the a-posteriori probability (the probability that a hypothesis H holds, given data D) with the probability that data D are measured, if the hypothesis H is true (Likelihood)

$$P(h|D) = \frac{P(D|h) \cdot P(h)}{P(D)} \ .$$
(3.7)

Starting with an a-priori probability for the hypothesis $P(h)$, the accuracy of the predictions can be iteratively improved with integrating new data. Usually, but not necessarily, a multidimensional Gaussian distribution is used as the prior for Bayesian optimisation, what can be well justified for physical phenomena.

3.6.1. Gaussian processes

The probability density function of a multi-variant Gaussian distribution in N dimensions (mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$) has a probability density function

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}} \cdot \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right). \tag{3.8}$$

According to the Central Limit Theorem [39], the distribution of independent vector-valued random variables converges to a multivariate Gaussian distribution. When one performs an optimisation of parameters, to minimise empirical measurement of an observable like transmission $T = 1 - F(\boldsymbol{x})$, the value of $F(\boldsymbol{x}_i)$, evaluated at various points $\boldsymbol{x}_i$, can be represented as the average of a set of so called component loss functions

$$\begin{pmatrix} F(\boldsymbol{x}_1) \\ \vdots \\ F(\boldsymbol{x}_N) \end{pmatrix} = \frac{1}{N} \sum_{j=1}^{n} \begin{pmatrix} f_j(\boldsymbol{x}_1) \\ \vdots \\ f_j(\boldsymbol{x}_N) \end{pmatrix}. \tag{3.9}$$

It can be assumed that such a vector for transmission values is approximately Gaussian distributed (Figure 3.2), and a multidimensional Gaussian distribution should preferably be used as a prior for Bayesian optimisation in this case. Since there are usually infinitely many possible values for the parameters, Gaussian processes are used instead, which are the infinite-dimensional analog of a multidimensional Gaussian.

Usually, the null-vector is used for the mean $\boldsymbol{\mu} = \boldsymbol{0}$, and the covariance matrix $\boldsymbol{\Sigma}$ is replaced with a Kernel function K, so that the following applies to the function to be modelled,

$$E(F(\boldsymbol{x}_i)) = 0\,, \tag{3.10}$$

$$E(F(\boldsymbol{x}_i) \cdot F(\boldsymbol{x}_j)) = K(\boldsymbol{x}_i, \boldsymbol{x}_j)\,. \tag{3.11}$$

Here, E is explicitly not the expected value of some statistical random distribution, but the expression of a prior assumption that has to be made about the problem. The distribution of a finite number of variables of a Gaussian process is again a multivariate Gaussian distribution, and for each point $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N$ of the parameter space, the vector

$$\begin{pmatrix} F(\boldsymbol{x}_1) \\ \vdots \\ F(\boldsymbol{x}_N) \end{pmatrix} \tag{3.12}$$

is a multivariate Gaussian distribution, where mean $\boldsymbol{\mu} = \boldsymbol{0}$ and covariance

$$\boldsymbol{\Sigma} = \begin{pmatrix} K(\boldsymbol{x}_1, \boldsymbol{x}_1) & K(\boldsymbol{x}_1, \boldsymbol{x}_2) & \cdots & K(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ K(\boldsymbol{x}_N, \boldsymbol{x}_1) & K(\boldsymbol{x}_N, \boldsymbol{x}_2) & \cdots & K(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{pmatrix}. \tag{3.13}$$

The Gaussian process can then be used as a prior for the unknown values of the loss function f. To find now a new distribution $F(\boldsymbol{x}')$ at a new test value $\boldsymbol{x}'$ in parameter

space, a vector $\boldsymbol{k}'$ is defined

$$\boldsymbol{k}' = \left(\begin{array}{cccc} K(\boldsymbol{x}_1, \boldsymbol{x}') & K(\boldsymbol{x}_2, \boldsymbol{x}') & \cdots & K(\boldsymbol{x}_N, \boldsymbol{x}') \end{array}\right)^T . \tag{3.14}$$

Using all observations made so far $\boldsymbol{y} = \left(\begin{array}{cccc} y_1 & y_2 & \cdots & y_N \end{array}\right)^T$, with $y_i = F(x_i)$, the marginal distribution of $F(\boldsymbol{x}')$ can be written as

$$F(\boldsymbol{x}')|(F(\boldsymbol{x}) = \boldsymbol{y}) \sim N(\boldsymbol{\mu}', \sigma'^2) , \tag{3.15}$$

where N is a normal distribution and

$$\boldsymbol{\mu}' = \boldsymbol{k}'^T \boldsymbol{\Sigma}^{-1} \boldsymbol{y} , \tag{3.16}$$

$$\sigma'^2 = K(\boldsymbol{x}', \boldsymbol{x}') - \boldsymbol{k}'^T \boldsymbol{\Sigma}^{-1} \boldsymbol{k}' . \tag{3.17}$$

3.6.2. Bayesian optimisation

In Bayesian optimisation (Figure 3.9, Algorithm 4), a certain number of positions in the parameter space are first selected according to a given criterion (usually random selection). The objective function is then evaluated at these points. Furthermore, an acquisition function and a kernel are defined. The kernel should be chosen to match the similarity of two parameters assignments. Usually this is done on the basis of a deeper insight into the problem gained from previous work or simply intuition.

The acquisition function determines the manner in which new points are selected for acquisition. For a Gaussian process prior, the acquisition functions are generally a function of $F(\boldsymbol{x}')$, its standard deviation and the best value seen so far during optimisation y_{best}.
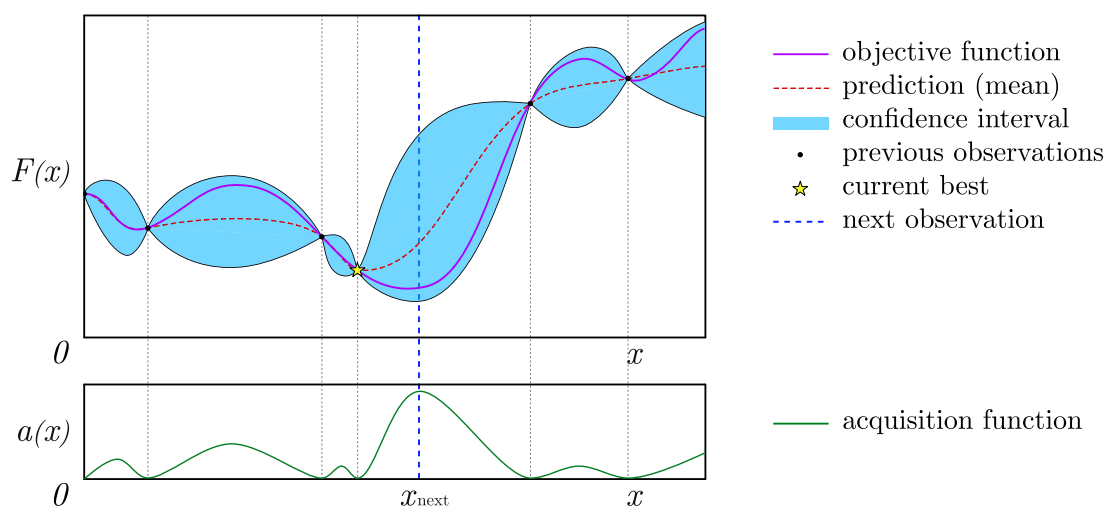

Figure 3.9.: Schematic illustration of the principle of Bayesian optimisation [40, page 11]. With each additional point evaluated, the confidence interval shrinks and the prediction approaches the objective function. In this way, new information about the objective function is constantly incorporated and increasingly better predictions for the minimum are obtained.

3.7. Bound optimisation by quadratic approximation

In 2009, M.J.F. Powell presented a derivative-free algorithm using a quadratic approximation model [41]. BOBQA (Bound optimisation BY Quadratic Approximation) is a package of subroutines based on NEWUOA [42], that can be used to find minima of 'black-box' functions on closed N-dimensional hypercubes.

Since a quadratic function of N variables has $\frac{1}{2}(N+1)(N+2)$ degrees of freedom, the number of interpolation conditions m imposed on a quadratic approximation must be defined in the interval $\left[N+2, \frac{1}{2}(N+1)(N+2)\right]$. In each iteration, m points $\boldsymbol{x}_i$ in the parameter space are automatically selected and the corresponding function values $F(\boldsymbol{x}_i)$ of the objective function are evaluated. Using confidence intervals, a quadratic approximation $Q(x)$ is then computed, which has the property $Q(\boldsymbol{x}_i) = F(\boldsymbol{x}_i)$. The quadratic models are updated periodically, allowing new information about the objective function to be integrated continuously. For this purpose, around the parameter $\boldsymbol{x}_{\min}$ that led to the smallest function value $F_{\min} = F(\boldsymbol{x}_{\min})$ so far, another value $F(\boldsymbol{x}_{\min} + \Delta\boldsymbol{x})$ is determined for a $\Delta\boldsymbol{x}$ that lies in a defined confidence region around $\boldsymbol{x}_{\min}$. If $F(\boldsymbol{x}_i) > F(\boldsymbol{x}_{\min} + \Delta\boldsymbol{x})$, one of the points $\boldsymbol{x}_i$ is then replaced by $\boldsymbol{x}_{\min} + \Delta\boldsymbol{x}$. At the new location, $Q(\boldsymbol{x})$ is now determined again. The trust region is then successively reduced up to a given threshold.

It has been shown, that this method can outperform Bayesian optimisation in accelerator online optimisation [43].

3.8. Other methods

An inevitably incomplete list of the most promising algorithms considered in this work is shown in Table 3.1. In general, there are many different further optimisation methods.

Table 3.1.: An incomplete overview of the main classes of optimisation methods for non-linear objective function in multidimensional space and their general properties with respect to the need for a gradient evaluation and their ability to explore areas remote from local optima.

class	examples	properties
gradient based	gradient descend, Gauss-Newton, BHHH, BFGS, DFP, SR1	good convergence properties, difficulties with local minima, gradient evaluation needed
heuristic	evolution strategies, genetic algorithm, particle swarm, simulated annealing	statistical methods with good exploration properties, can escape local minima, are robust in case of noisy data, can be adapted to the problem, need frequent evaluation of the objective function, have many hyper parameters
model functions	Bayesian Opt., BOBQA	gradient free, evaluation of the objective function is only necessary at a few points, have few hyper parameters, have difficulties with fractured parameter spaces

Some of them are specialisations for certain problems, such as the ant colony algorithm for the traveling salesman problem [44], or variations of the algorithms presented in the previous sections. For example, the hill-climbing algorithm is equivalent to the genetic algorithm without operators and the evolution strategies [45] are similar to the genetic algorithm without crossing over. It is to be expected that there will be further developments in this direction, since there is a great need for fast optimisation methods in science and technology.

Artificial neural networks have also proven to be very powerful recently and it can be shown that such networks are generally able to approximate any given function [46]. However, the backtracking algorithm, which is used for the training, is a gradient method itself [47] and therefore also not able to escape local extrema. It is true that there is also the possibility to perform the learning process using a genetic algorithm [48]. But even if an artificial neural network perfectly approximates the objective function over the search space, it's extremum still cannot be found without additional testing. This would only shift the optimisation problem. Therefore, for the problem of general accelerator optimisation, the use of artificial neural networks is not indicated.

Motivated by the use of the specific advantages of different optimisation methods, hybrid optimisation methods often represent a combination of different methods. Usually, stochastic and deterministic methods are combined. However, purely stochastic or purely deterministic combinations can also be used. In addition, statistical experimental designs, meta models and heuristics are applied. Corresponding optimisation algorithms are often very specialised for the respective application [49]. However, it can generally be shown that there is no optimal algorithm that can be applied equally successfully to all classes of problems, which is exactly the key message of the no-free-lunch theorem [50].

Table 3.2.: Terminology of genetics, mapped to the field of accelerator optimisation.

term	general meaning	meaning in accelerator optimisation
generation	one optimisation cycle	checking multiple settings for resulting transmission
gene	encoding of a specific property	specific value for a particular parameter e.g. a quadrupole strength
genome	the set of all encoded properties of an individual	the set of all values of a given accelerator setting
population	the set of all genomes of a generation	the set of all settings of an optimisation cycle
genotype	the representation of an individual as the set of its encoded properties	the representation of a transfer line as a set of all device parameters
phenotype	real representation of an individual in the form of its measurable properties	a transfer line, with the settings corresponding to the genotype and the resulting transmission
individual	a particular combination of genotype and phenotype	a particular setting with one value for each parameter

Ultimately, it is always necessary to analyse the given problem and to find, adopt or design the appropriate algorithm. Therefore, in the following chapter, the presented algorithms and their parameterisations are examined with respect to their performance in

connection with the particular problem of accelerator optimisation. For this purpose, the terms objective function and fitness function are used synonymously. Other terms from the evolution theory, that are used frequently in the following sections, are listed in Table 3.2. For simulated annealing and particle swarm optimisation, the table applies analogously if one substitutes the terms 'gene' with a single 'particle coordinate' and 'genome' with 'particle position' in parameter space.

4. Automatic beam line tuning

Depending on the specifics of the particular laboratory and the dedicated beam quality requirements, manual setup of beam transport paths in particle accelerators can be a time-consuming process. This is especially the case during initial commissioning, when determination of standard parameters is the major task, while machine and equipment specifics are not yet very well understood and the software modelling is still at the very beginning. A good accelerator model and a carefully executed alignment of all optical elements is particularly important and the understanding must improve as the operation progresses.

However, sometimes this is not easy to accomplish because the boundary conditions or the initial beam parameters are not constant. Especially in multipurpose accelerators with massive parallel operation, there are always situations that make readjustment necessary or result in known settings not being reproducible. Reasons include: changing ion sources or isotopes, changes in timing, new charge states, degradation of stripper foils, or adding new parallel cycles or including new devices. At GSI Helmholtzzentrum, the initial setup of a beam for a new experiment is usually estimated to take about 8 hours. In normal operation, this situation occurs several times per week. Automation therefore has the potential to save a lot of time and personnel resources and eventually increase availability.

4.1. Comparison of optimisation algorithms

For the following investigations, the four most promising algorithms from Table 3.1 (Genetic Algorithm, Particle Swarm, Simulated Annealing and BOBQA) were compared. To get a first overview, a simple artificial beam line with four quadrupoles and a central accelerating gap (Figure 4.1) was tested, for which it is easily possible to find a solution by manual tuning. However, due to the four quadrupoles, several equivalent solutions are possible in this case. If a minimum excitation of the quadrupoles is introduced as a further optimisation criterion, for example to find a solution which minimises the electrical power consumption, then the setup is also suitable to test the abilities of the algorithms to leave local minima.

For the given example, the search space is a hypercube with five dimensions (four quadrupole strengths k_i and the gap voltage U_g). Using a direct real-valued representation, the fitness function F to be minimised is then

$$F(k_1, k_2, k_3, k_4, U_g) = 1 - T(k_1, k_2, k_3, k_4, U_g) + w \cdot \sum_{i=1}^{4} \frac{1}{4} \frac{|k_i|}{k_{\max}} \qquad (4.1)$$

where T is the transmission and all k_i as well as U_g are normalised to the range $[-1, 1]$ with respect to the corresponding individual limits. To minimise the power consumption, the sum of the quadrupole strengths normalised to the maximum value and weighted with factor w is added to the fitness function. The factor w must be chosen smaller compared
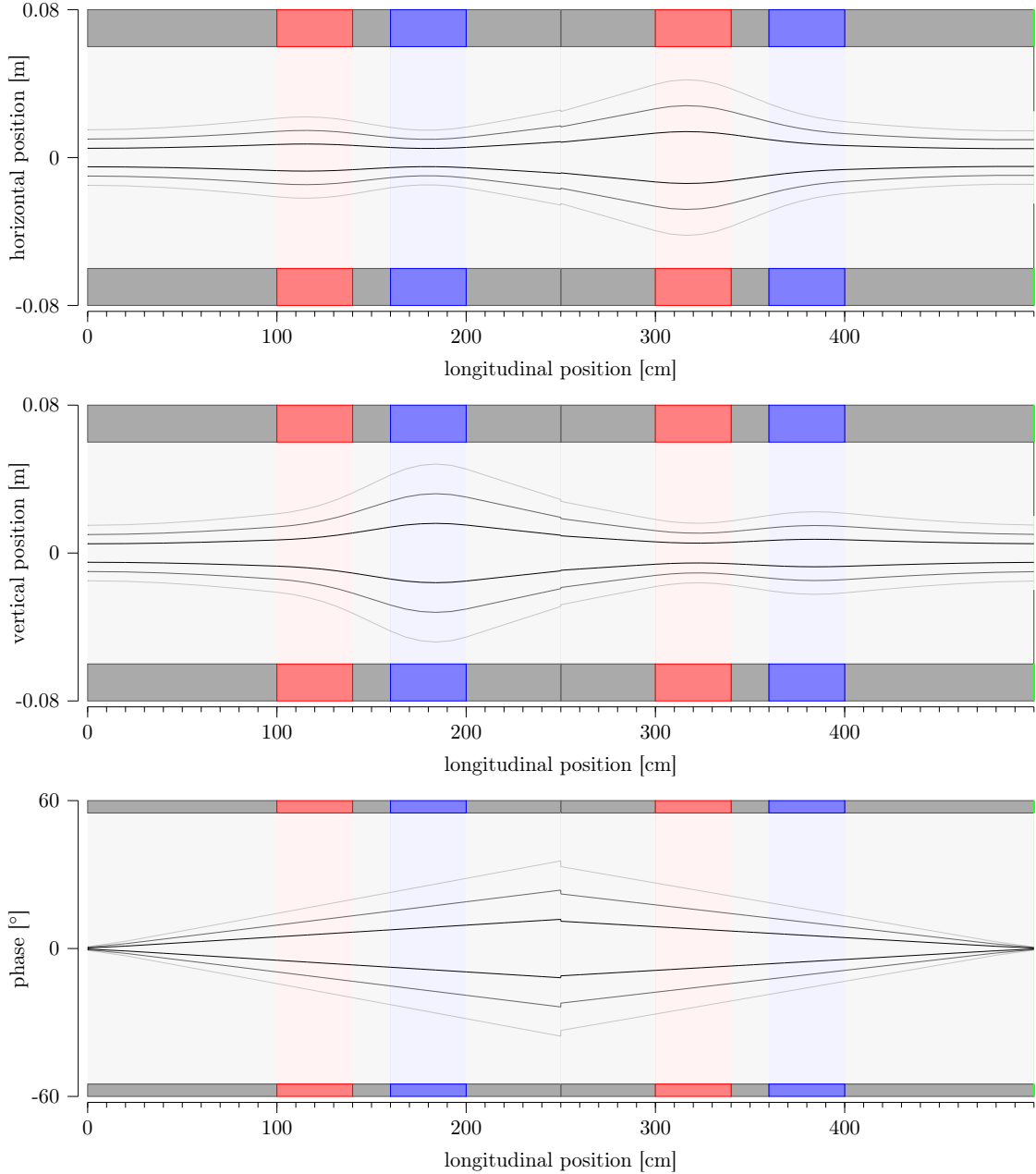
Figure 4.1.: The test beam line is 5 m long, contains four quadrupoles (blue and red elements) of length 40 cm each and an central acceleration gap. At the end, there is an elliptical aperture boundary with semi axes 25 cm horizontally and 20 cm vertically. The phase acceptance is $\pm 5°$. A particle distribution was chosen for which it is easy to find a setting with which the acceptance can be met. One solution with 100% transmission is shown here. The black lines represent 1σ of the distribution, the dark grey 2σ and the light gray 3σ.

to the minimal transmission change, because the maximum transmission is the actual goal of the optimisation and even the smallest improvement should be weighted higher in any case. To determine the value of T for a certain point in the search space, a matrix tracking simulation with 1000 single particles is performed. The smallest possible change in the transmission is then $\Delta T_{\min} = 0.001$, which means that $w = 0.001$ can be used.

All algorithms used, provide different hyper-parameters to adapt them to problems and to fine-tune their performance. However, the best values for hyper-parameters are not known a priori. Thus, the problem arises that the optimisation algorithms must also be optimised in some sense and if their parameters are not robust against changes, the problem of optimisation is just shifted to the hyper-parameters of the algorithms. In order to find reasonable initial values for the hyper-parameters and to be able to assess their robustness, each of these hyper-parameters was first scanned individually in a reasonable range. Since the nature-analog algorithms are based on statistical processes, in these cases each simulation was done 100 times, to collect some statistics.

4.1.1. Hyper-parameter-scans

The results of the hyper-parameter scans are listed below. Not scanned were those hyper-parameters which, with increasing value, definitely lead to an improvement of the outcome of the respective algorithm, but also have a decisive influence on the runtime. In the case of the genetic algorithm this is the number of individuals and in the case of simulated annealing or particle swarm it is the number of particles. For the scan, this values were all arbitrarily set to 100. Additionally the optimisation was stopped after the 100th cycle. So for each run 10^4 accelerator settings had to be simulated with a matrix tracking code.

Genetic algorithm

The mutation rate is a component of the genome and is thus itself subject to the evolutionary process and does not need to be scanned. It turns out that the genetic algorithm, in this particular example, is very robust to parameter changes. Figure 4.2 shows for all hyper-parameters a large stable range in which a solution with maximum transmission can be reliably found. Only σ_s apparently must not be chosen too high. If individuals with too poor fitness are included in the reproduction process, this has a negative effect on the performance of the optimisation. It is also noticeable that p_r has very small deviations from the optimum at values between 0.02 and 0.16. The crossing over probability p_x seem to have no effect at all on the performance of the algorithm, at least in this example. The values that were ultimately used for the comparison can be found in Table 4.1.

Table 4.1.: Hyper-parameters of the genetic algorithm considered as good after scan.

| Parameter	Description	Min.	Max.	Best
N_{keep}	number of best individuals, kept for next generation unchanged	0	20	3
σ_s	standard deviation of the fitness of individuals selected randomly for reproduction to the fitness of best individual of their generation	0.01	1	0.15
p_m	mutation probability for a single gene	0	1	0.50
p_r	probability for a gene to be replaced by a new random number	10^{-5}	1	0.10
p_x	probability for the use of the crossing over operator	0	1	0.50
p_a	probability for the use of the arithmetic mean operator	0	1	0.10
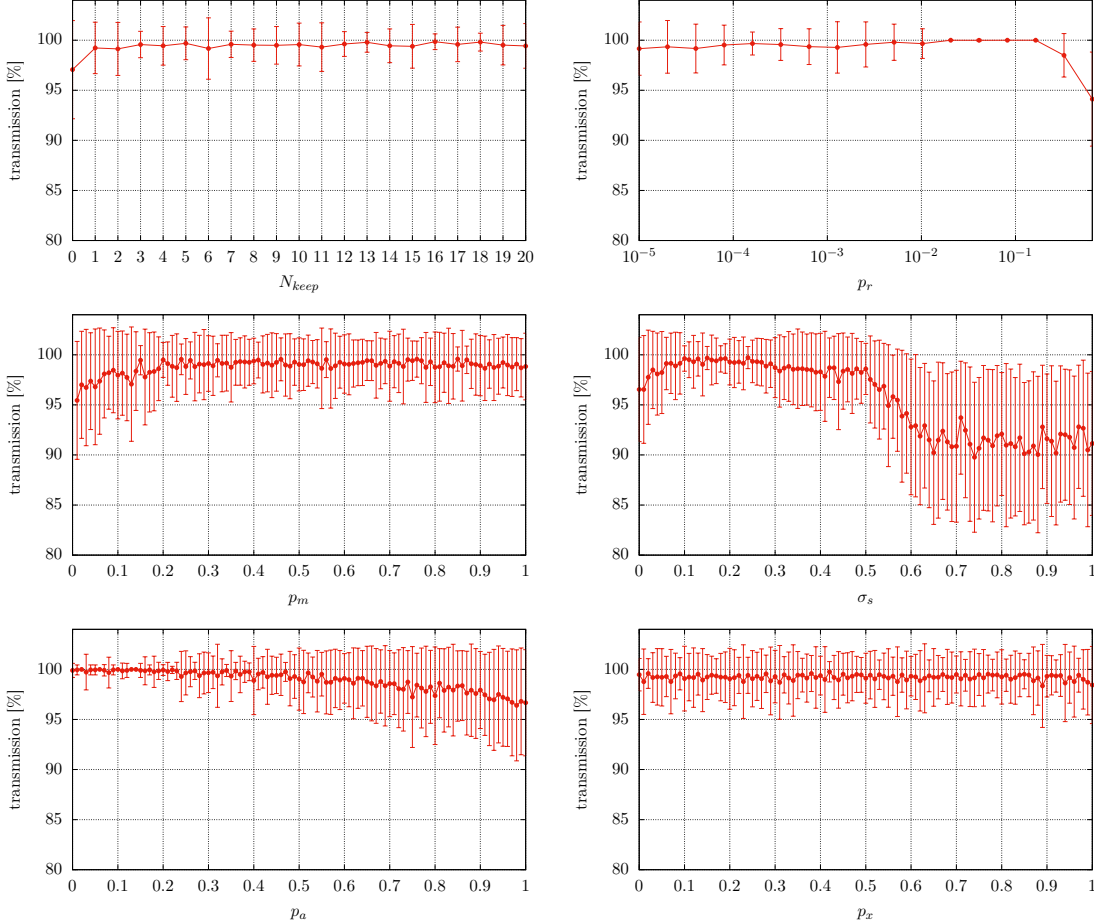
Figure 4.2.: Hyper-parameter scans for a transmission optimisation of the test beam line with a genetic algorithm. For each parameter value the optimisation was performed multiple times. The corresponding standard deviation is also shown in the plots.

Simulated annealing

The geometric progression $T_t = t^{-2}$ was chosen as the base temperature function. The rate at which the series converges to 0 is then determined by the parameter Δt. Figure 4.3 shows that it is disadvantageous if the temperature parameter falls too slowly. However, from about $\Delta t > 1$ on, no particular influence on the performance can be observed. The algorithm is also insensitive to the choice of r. In general, it can be stated that simulated annealing is extremely robust with respect to hyper-parameter changes, but apparently also has a poor performance in general.

Table 4.2.: Hyper-parameters of the simulated annealing algorithm considered as good after scan.

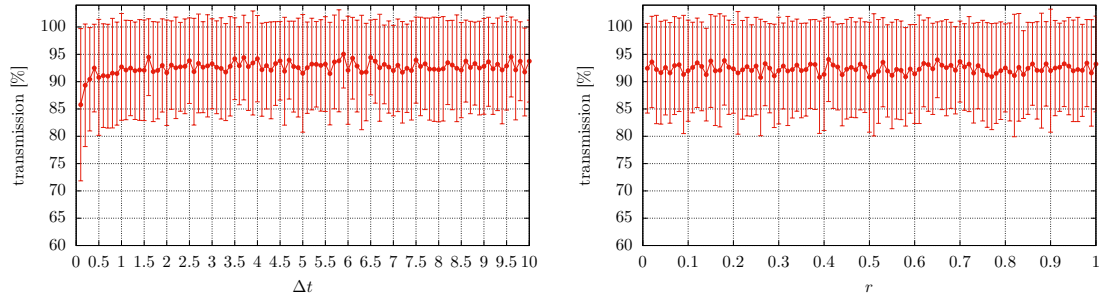| Parameter	Description	min	max	best
Δt	time step	0	4	3.5
r	radius of random movement of the particles in parameter space	0	1	0.1

Figure 4.3.: Hyper-parameter scans for simulated annealing optimisation of the test beam line. For each parameter value the optimisation was performed multiple times. The corresponding standard deviation is also shown in the plots.

Particle swarm

Besides the number of swarm particles, the algorithm has only three hyper-parameters, the inertia parameter m and the two weights (w_c and w_s). For the scans (Figure 4.4), the two fixed parameters were set to 1 in each case. It can be noticed that the inertia parameter is of special importance. It must not be set too low, because it has a special meaning

Table 4.3.: Hyper-parameters of the particle swarm algorithm.

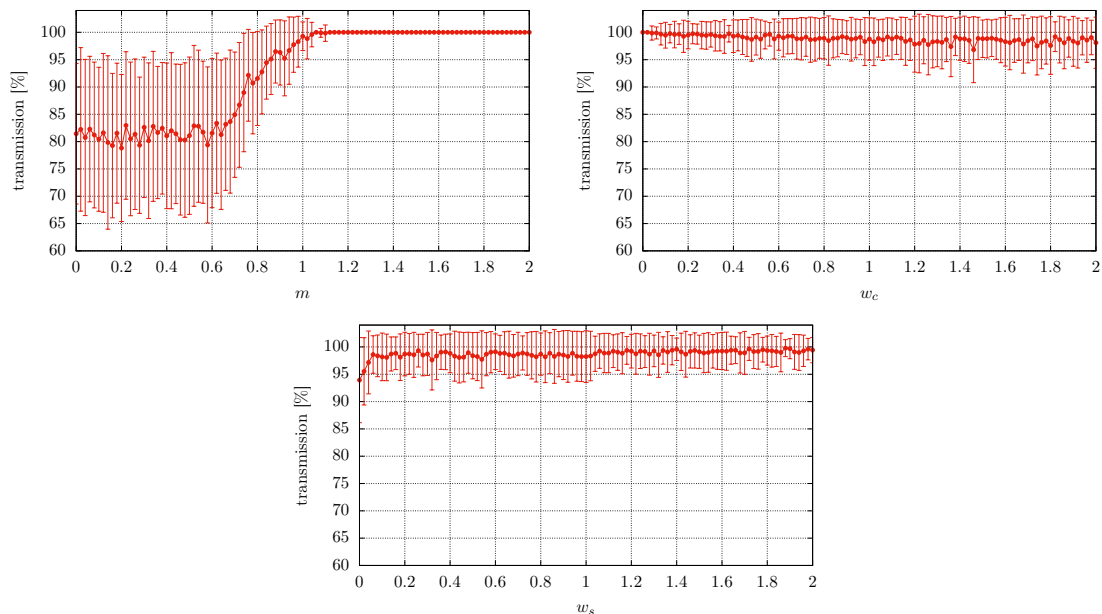| Parameter	Description	min	max	best
m	inertia parameter for the particles	0	2	2
w_c	cognitive weight factor	0	2	0.1
w_s	social weight factor	0	2	2

Figure 4.4.: Hyper-parameter scans for particle swarm optimisation of the test beam line. For each parameter value the optimisation was performed multiple times. The corresponding standard deviation is also shown in the plots.

for the ability of the swarm to leave local minima. The algorithm is quite robust against changes of the weighting factors, however, the cognitive weighting factor should be chosen rather smaller and the social weighting factor larger than 1.

Bound optimisation by quadratic approximation

The BOBYQA-Algorithm has only one hyper-parameter, which is the number of interpolation conditions m. Its value must be in the interval $[N + 2, (N + 1)(N + 2)/2]$, where N is the number of dimensions of the parameter space. So, for the test example with 5 parameters, the range from 7 to 21 was scanned (Figure 4.5). Since the algorithm is deterministic for identical start values, each simulation was done only once. For a value of $m = 14$, the optimisation achieved a transmission of 100%. For the values 11,15,16 and 17 it was still 99.8%.
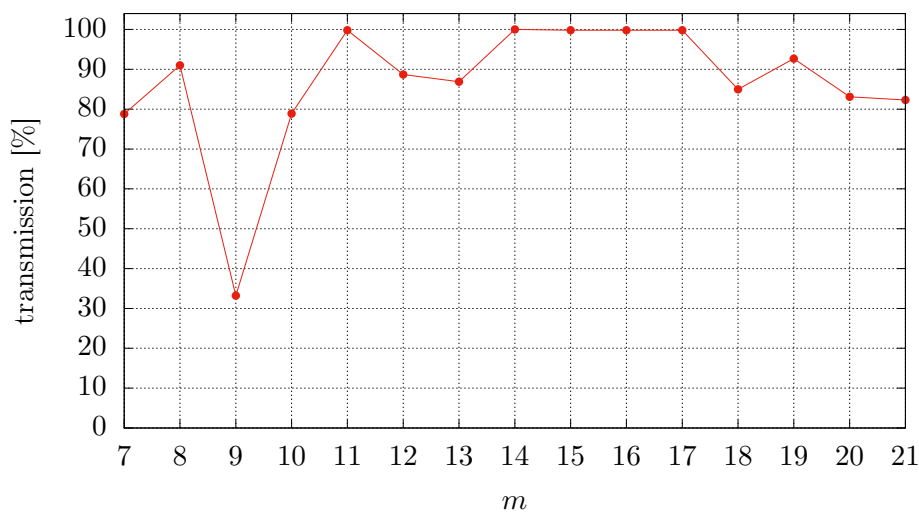

Figure 4.5.: Hyper-parameter scan for BOBYQA optimisation of the test beam line.

4.1.2. Comparison

For the given example, both the genetic and the particle swarm algorithm can be used to find parameters which, despite the statistical character of these optimisations, lead very reliably to optimal transmission after 100 cycles. Simulated annealing, on the other hand, achieved a maximum of 95% on average. The BOBYQA algorithm also reliably finds a 100% solution with the correct parameters. However, the maximum achievable transmission and the reliability with which it is achieved are not the only decisive factors. The main factor for online optimisation of an accelerator is the number of parameter sets that have to be tested on the accelerator in order to achieve a certain transmission. Depending on the maximum repetition rate of the accelerator, the duration of one cycle is the bottleneck for any optimisation. All algorithms were therefore set with their optimal parameters and compared with each other. This time, no fixed number of cycles was specified, but for the statistical methods a termination criterion for the transmission was defined. Once the optimisation was stopped at $T > 95\%$ and once at $T > 99.5\%$. The results are shown in Table 4.4. Of the nature-analog optimisation methods, the genetic algorithm converges the fastest. After already 5.4 ± 3 generations 95% transmission is

reached. Thus, a corresponding optimisation would require 540 accelerator cycles and would be finished after 9 min at a repetition rate of 1 Hz. An optimisation to 99.5% would need approx. 15 min.

Table 4.4.: Comparison of the investigated optimisation methods. Each nature-analog algorithm was tested twice, each with a different termination criterion. Once the algorithm was terminated when a transmission $T > 95\%$ was reached and once at $T > 99.5\%$. Each optimisation was again performed 100 times. The mean number of accelerator cycles $\overline{N}_c$, the mean integral quadrupole strength $\overline{K}$ and the mean gap voltage $\overline{U}$ found for the tested cases are shown in the table. The BOBYQA algorithm was run only once with the optimal parameter. Therefore, there are no standard deviations and no cut-off value.

Algorithm	T_{cut}	$\overline{N}_c$	$\sigma^2_{N_c}$	$\overline{K}$	σ^2_K	$\overline{U}$	σ^2_U
Genetic Algorithm	95%	540	300	17.40/m²	1.45/m²	0.21 MV	0.02 MV
Genetic Algorithm	99.5%	920	3.3	18.04/m²	0.79/m²	0.22 MV	0.01 MV
Particle Swarm	95%	580	3.1	17.86/m²	1.56/m²	0.22 MV	0.02 MV
Particle Swarm	99.5%	1500	13	18.19/m²	0.85/m²	0.22 MV	0.01 MV
Sim. Annealing	95%	9200	61	13.69/m²	1.61/m²	0.20 MV	0.02 MV
Sim. Annealing	99.5%	18600	95	14.73/m²	0.88/m²	0.21 MV	0.01 MV
BOBQA	100%	213	0	15.54/m²	0/m²	0.21 MV	0 MV

While the genetic algorithm and the particle swarm optimisation are still very close to each other, at least for the 95% criterion, simulated annealing falls off very sharply although the corresponding transmissions can be achieved with much less quadrupole excitation. The mean voltage at the gap is also reliably lower. It seems that with Simulated Annealing the exploration to exploitation ratio is shifted more towards exploitation, while the minimisation of the excitation has not yet really started with the other two methods. The genetic algorithm is therefore to be preferred among the statistical methods for an online optimisation.

In the end, it must be said that the BOBYQA algorithm in this particular example uses only a quarter of the accelerator cycles. Moreover, since it is a deterministic method, it reliably leads to an optimal solution in terms of total transmission while maintaining very low quadrupole strengths. BOBYQA is thus initially particularly suitable for accelerator optimisation. This is why it was successfully tested off-line for finding a new optic for the SIS18-HADES transfer line at GSI Helmholtz Centre [43]. To what extent the deterministic character is a disadvantage in the case of noisy data, as it arises in an online optimisation, is investigated later in the specific example.

4.2. The autotune algorithm

The genetic algorithm was selected for the first experiments on online optimisation of a real accelerator section. As part of the Swedish in-kind contribution to the FAIR project, the CRYRING@ESR was commissioned at GSI Helmholtz Centre in 2015 and has been ready for initial investigations and machine studies since 2016 [51]. Since the CRYRING@ESR has its own local injector (Figure 4.6), it can be operated stand-alone independently of the GSI accelerator chain. During the two years long major shutdown of the GSI accelerator
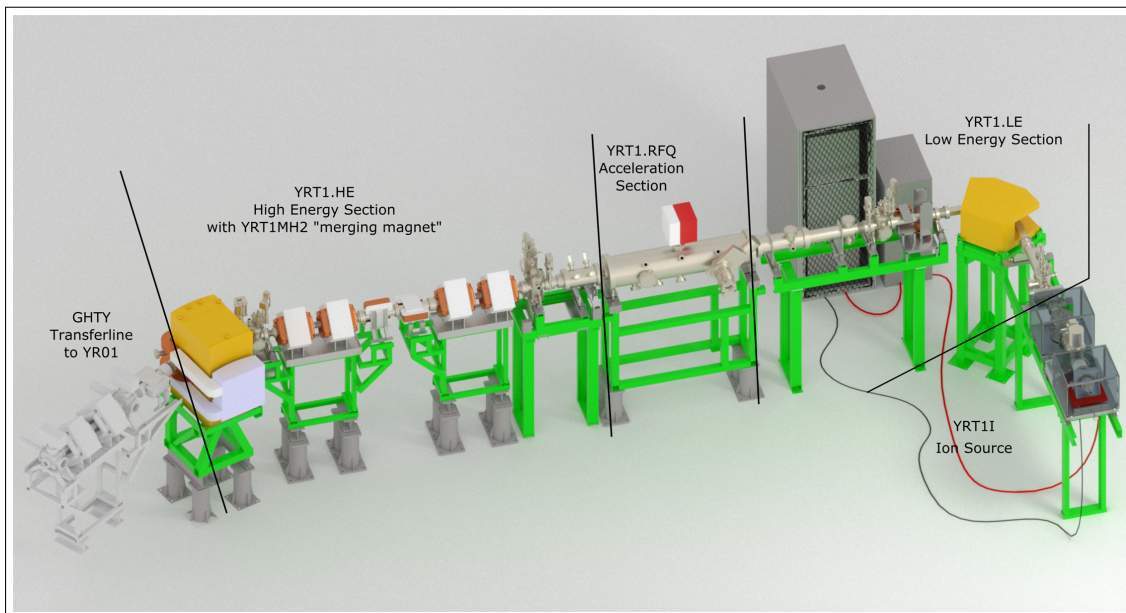
Figure 4.6.: The local injector of CRYRING@ESR at GSI Helmholtz Centre [52].

facilities between August 2016 and April 2018 [53], the local injector provided the only possibility to perform such a study at GSI, but also presented ideal conditions, since the FAIR control system was already available at the CRYRING. The investigations were then carried out in 2017 together with S. Appel et al. [54].

Three electrostatic quadrupoles and one electrostatic steerer were tuned in the low energy section between ion source and RFQ. The transmission measurement was performed between the source transformer and the Faraday Cup in front of the RFQ. The results were promising but not completely convincing. Due to the long settling time of the installed electrostatic optical elements, 20 s had to be waited before each transmission measurement, to make sure, the actual voltage has reached the desired value. In a 30 min run, it was only possible to run an optimisation for five generations and the parameter space was limited to a narrow band around the parameters found manually during commissioning. The maximum achieved transmission in the last generation was nevertheless only slightly better than the randomly chosen init values. However, a significant improvement in the mean transmission of individuals was observed over all five generations. A follow-up study with improved tools, showed slightly improved results [55].

It was shown that online optimisation of a short transport line (approx. 6 m) in a narrow parameter band is possible to perform, using a genetic algorithms. But it became obvious, that the CRYRING injector is not particularly suitable for fast tuning due to the slow response of the electrostatic elements.

The UNILAC, on the other hand, is designed to run pulsed at 50 Hz. Thus, it would be theoretically possible to test fifty individuals per second, which would lead to an enormous acceleration of the optimisation. However, the transfer lines are longer by an order of magnitude and contain accordingly many more parameters to be optimised, but also several diagnostic elements to measure the transmission. Nevertheless, it was very attractive to make this investigation, because just the manual tuning of the UNILAC needs a large part of the setup time of beams for experiments at GSI, which is usually around 10% of the beam time available [56].

4.3. Online optimisation of the GSI transfer channel TK

The study was conducted in 2019 during a beam study campaign at GSI Helmholtz Centre. The object of the investigation was a 116 m long section of the transfer line from UNILAC to SIS18 (Figure. 4.7). Starting point was the beam transformer GTK3DT4, where the

Figure 4.7.: overview of transfer channel from UNILAC accelerator to SIS18 synchrotron [GSI overview drawings, status 2010].

reference current I_0 for the optimisation was measured. The manual setup of the transfer line (36 parameters) usually takes 1-2 hours. Since the SIS18 allows cycle times of $< 1\,\mathrm{s}$ (in FAIR-booster mode 2.7 Hz operation is foreseen [57]) , the transfer channel is designed for a corresponding repetition rate and all magnets can be pulsed with 10 Hz. The high repetition rate is an ideal condition for automatic optimisation procedures.

4.3.1. Simulation

A simulation was carried out to check the general feasibility and to find the optimal operating parameters for the genetic algorithm. For the particle tracking the TK Lattice information verified by Y.El Hayek [58] (Figure 4.8) was used. The standard deviation of
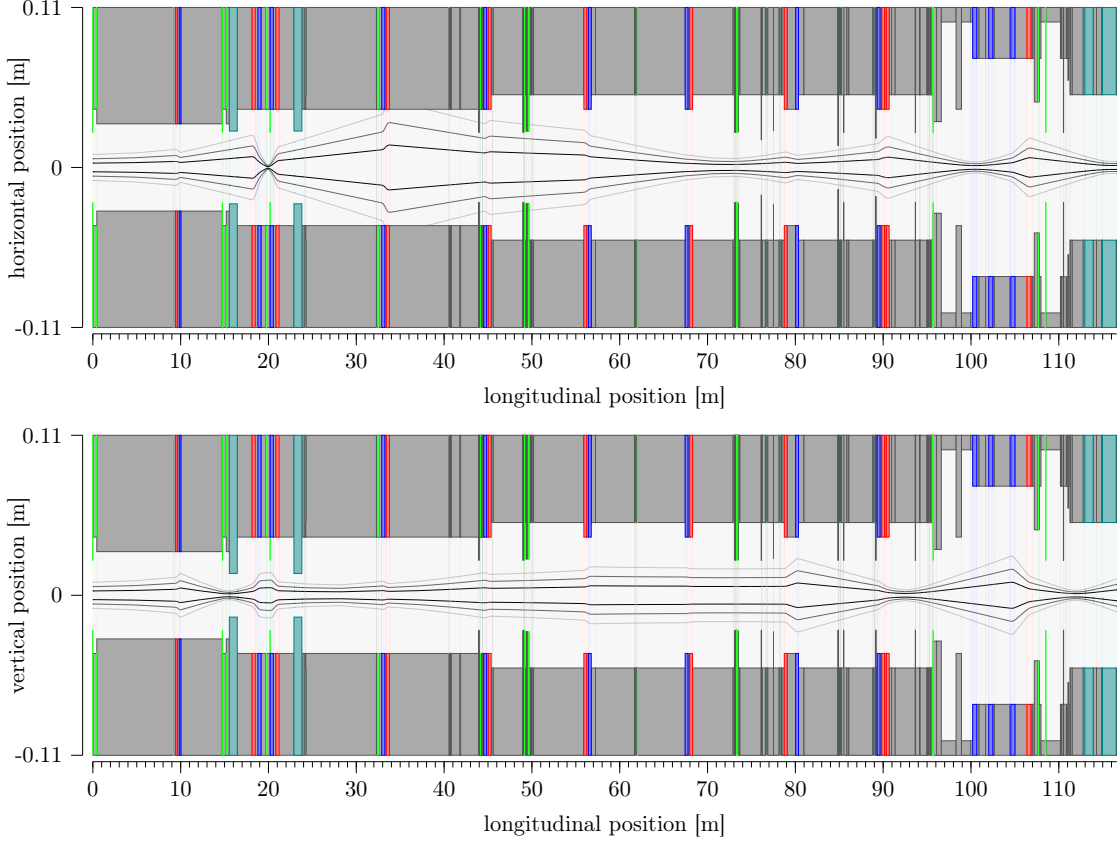
Figure 4.8.: lattice of the SIS18 transfer channel (top = horizontal, bottom = vertical): blue elements are vertical focusing and red elements are horizontal focusing quadrupoles, cyan elements are bending dipoles. All beam instrumentation devices are displayed in green.

the actual current of the power supplies were measured to 0.1% for quadrupoles and 0.04% for steerer magnets. These values were included in the simulation. The measuring accuracy of the transformers (0.3%) was also considered in order to represent the real conditions as correctly as possible. Again a parameter scan was carried out to find the parameter set for the fastest possible convergence of the genetic algorithm. The optimal parameters for the given transfer line are listed in Table 4.5.

The fitness function to be minimised slightly differs from equation (3.1)

$$F = \sum_{i=0}^{4} c^i (1 - \tau_i). \tag{4.2}$$

It represents the sum of the beam losses, measured via four consecutive beam transformers with the transmission $\tau_i = \frac{I_i}{I_0}$ weighted with a factor c^i. Compared to the fitness function which directly evaluates the total transmission, the slightly modified version converges more reliably. The ideal value of c could be determined to $c = 2.7$.

Table 4.5.: GA hyper-parameters for TK optimisation.

name	parameter description	value
I	number of individuals	120
N_k	number of most performing individuals kept for next generation	1
σ_s	variance of individuals chosen for reproduction (0...best, 1..worst)	0.13
p_m	probability of each gene to mutate	0.38
p_r	probability of a gene to be replaced	0.0001
p_c	probability for crossing over at each gene	0.52

With optimised parameters a convergence in 250 generations could be achieved. A transmission of 90% was achieved on average after 180 generations, which corresponds to 21600 accelerator cycles using 120 individuals and so, at a repetition rate of 10 Hz, the optimisation would take 36 min.

For the BOBYQA algorithm, a parameter scan was also performed with the noisy quadrupole strengths and currents measured from noisy beam transformers. Since BOBYQA results are also no longer deterministic in this case, optimisation was performed multiple times for each hyper-parameter value (Figure 4.9). The best average transmission
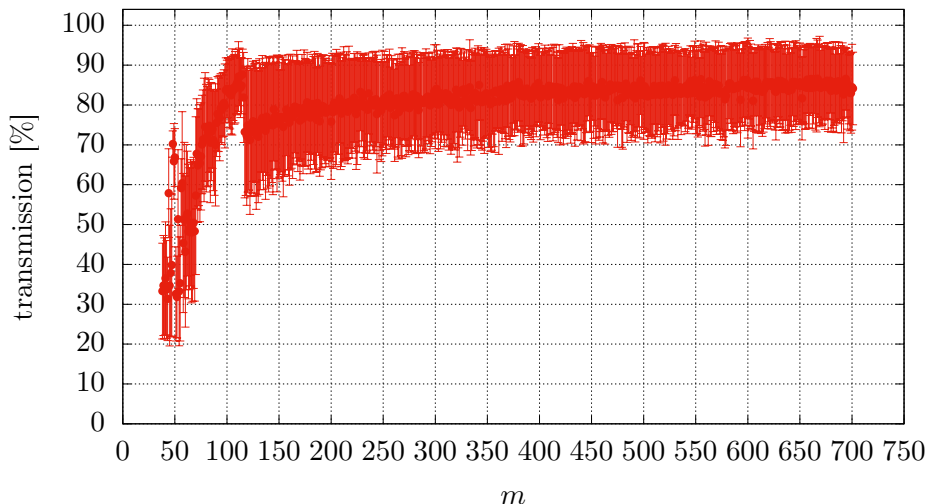
Figure 4.9.: Parameter scan of m with respect to total transmission.

is $T = 88\%$ and could be achieved at $m = 115$. In contrast to the genetic algorithm, a transmission of $T = 100\%$ could not be achieved with BOBYQA in any case. However, m can be adjusted such that values of $T > 75\%$ can be reliably achieved. Another advantage over the genetic algorithm is the extremely low number of accelerator cycles required. For the best value, only 1066 ± 71 cycles were needed, which means that the optimisation requires only about 5% of the time needed by the genetic algorithm. Although BOBYQA proved to be quite suitable for the usage in an online optimisation, the corresponding comparative study on this particular algorithm had not yet been carried out at the time of the experiment. Therefore, the BOBYQA algorithm was not considered for the TK study.

4.3.2. Experiment setup

For the experiment a completely stripped argon beam $^{40}\text{Ar}^{18+}$ with an energy of $8.5\,\text{MeV/u}$ was used. In order to prevent unnecessary activation, a pilot beam was used for the optimisation. The beam current was reduced to less than $100\,\mu\text{A}$ on the reference transformer GTK3DT4 to additionally prevent the measuring range limits of the beam transformer from being crossed and thus to avoid switching the measuring range during the optimisation process.

End point was the beam transformer GTK7DT3. The original intention was to optimise the entire beam transport up to the injection point of the SIS18, but the last two transformers GTK8DT7 and GTK9DT8 could not be used due to a technical defect. The total length of the beam line decreased to $75\,\text{meter}$ and the number of tuneable parameters reduced to 20. Another 8 magnets are locked for manual re-tuning (and thus also for autotune), leaving 6 quadrupoles, 3 horizontal steerers, 3 vertical steerers for optimisation and 4 beam transformers that could be used for transmission measurements. Such a strong reduction of the number of free parameters generally led to the expectation of better performance during optimisation process.

For the beam time, the high current source was in operation, which limited the repetition rate to $f_r = 1\,\text{Hz}$. Unfortunately a promising run could not be completed due to a technical defect. Since the remaining experiment time was only $t = 40\,\text{min}$ and because $t \cdot f_r = G \cdot I$, the final run had to be limited to $G = 20$ generations.

The genetic algorithm was used with the parameters from Table 4.5. During the optimisation, 2400 different settings were tested with one beam pulse each.

4.3.3. Results

After the optimisation time of 40 minutes, a total transmission of $T = 70\%$ could be achieved. Figure 4.10 shows the development of the transmission over the optimisation process at 4 consecutive beam transformers. The constantly limited transmission between reference transformer GTK3DT4 and GTK4DT3 suggests that the setup of the beam line before the reference transformer was not optimal and the resulting mismatch could not be compensated with the intermediate quadrupole doublet GTK4QD2. For the $54\,\text{m}$ long section from GTK4DT3 to GTK7DT3 the final transmission was $T = 86\%$. For comparison: with manual tuning, values of up to 83% have been set up for this section within a similar optimisation time during the FAIR phase 0 [59] physics run in 2020.

Fitness value developed as in Figure 4.11. It is not yet visible that it is approaching a lower plateau, so it is to be expected that a longer optimisation time could result in a further improvement. It can also be seen that the fitness function does not fall monotonically continuously, as would actually be expected. This is due to minimal field fluctuations of the magnets and deviations in the transmission measurement, but can also be caused by slight changes of conditions of the incoming beam. The large rebound from Generation 11 to Generation 12, is specifically due to a short quadrupole failure. The magnet could be reactivated after a few seconds with a reset, but all genomes evaluated during this time, were evaluated without beam current. The measured transmissions are therefore bogus, but were still used to evaluate fitness. Therefore, potentially good individuals have also been screened out by this failure, while others have been evaluated too optimistically. During the experiment, the w factor was left zero and so low magnet excitations were not particularly preferred by the algorithm, thats why steerer magnets were not optimised to
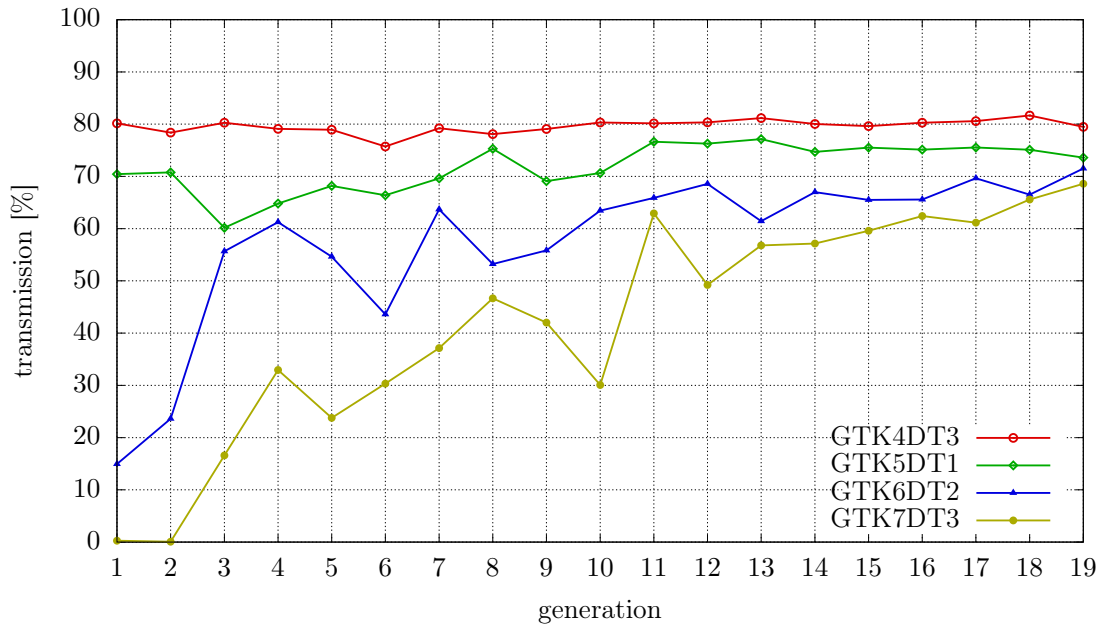
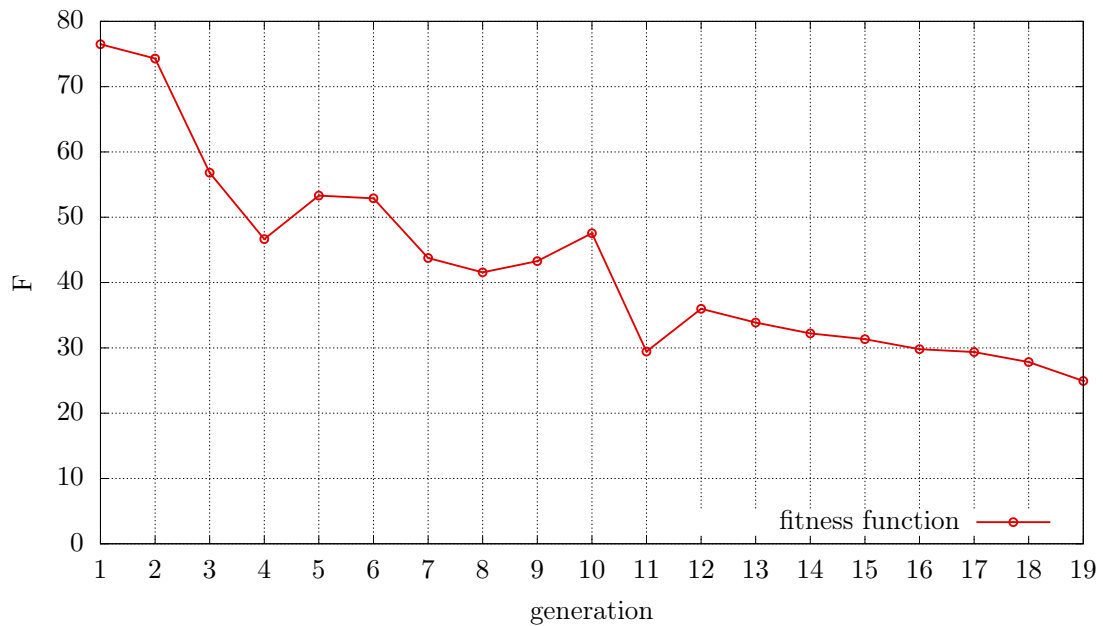Figure 4.10.: Development of the transmission with regard to the reference transformer GTK3DT4

Figure 4.11.: Development of fitness value during the online optimisation process.

a low angle if this was not explicitly necessary to avoid losses. Therefore all profile grids show significantly more offset than one would expect from a manual setup (Figure 4.12).

4.3.4. Conclusion & outlook

The transfer channel turns out to be an ideal application for the autotune algorithm and should be further investigated. Both, setup time and transmission results are at least equivalent to manual tuning and it can be expected, that any further improvement shifts the assessment in favour of the algorithm.
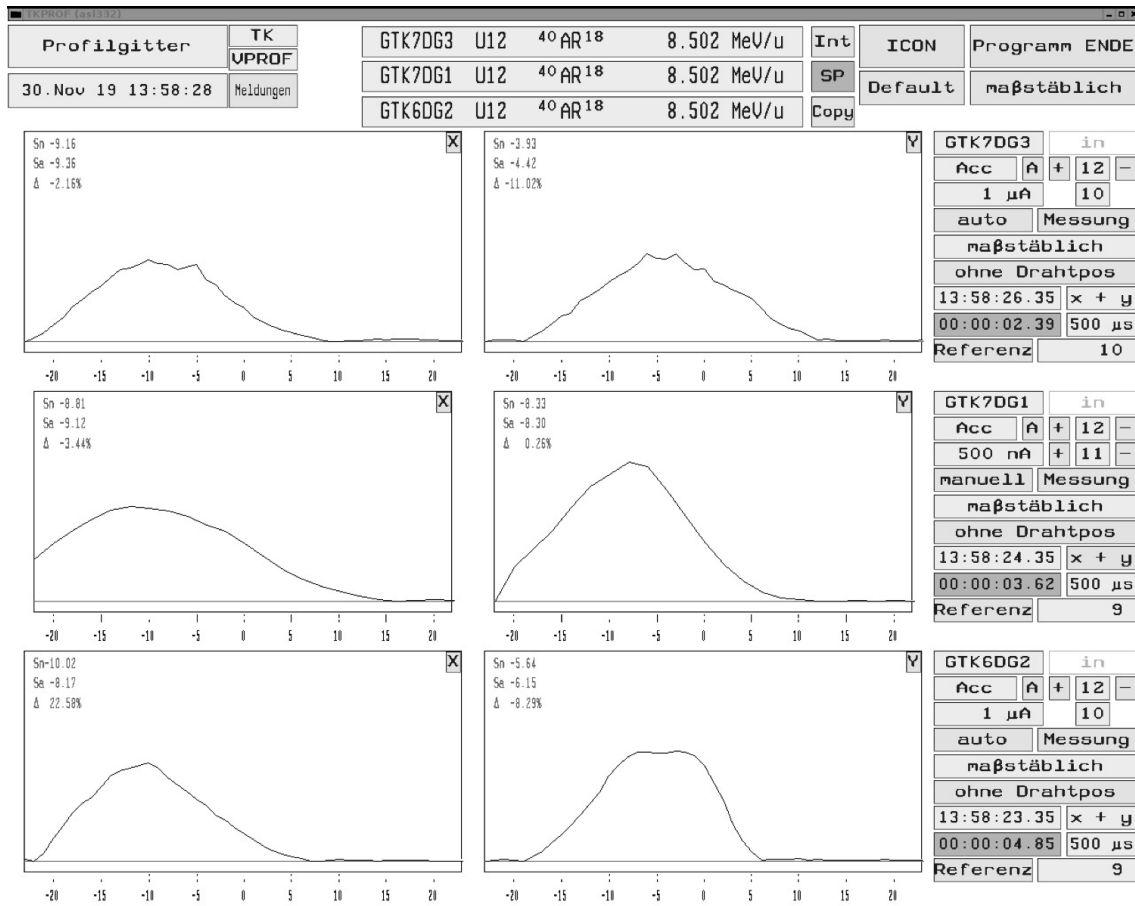
Figure 4.12.: Profile grid measurements of the best setting, left - the horizontal pane left and right - the vertical pane

The entire channel should be optimised again in full length under normal operating conditions. For this purpose, an interruption protection must be implemented so that after interruptions it is not necessary to start the optimisation again from the beginning.

If the SIS18 fast transformer can be integrated into the optimisation process, the injection efficiency could be optimised directly. In this case the difficulty is to coordinate the synchronous access to parameters of the new FAIR settings management system [60], which is already in use from SIS18 downstream, and the legacy GSI control system, used at UNILAC. If successful, further top level parameters could be used for optimisation and thus the complete injection process could be optimised, as already suggested in [61].

A parallel running particle tracking simulation could identify settings which definitely lead to total beam losses. The corresponding individuals should be evaluated by the simulation instead by testing their settings on the real accelerator. This pre-check can potentially reduce or even eliminate total losses, which is also desirable in terms of machine protection and radiation safety. Such an approach would at least reduce the number of accelerator cycles required. It has been suggested that the BOBYQA algorithm instead of a genetic algorithm leads faster to convergence. There are also recent investigations to improve the BOBYQA algorithm with respect to safety [62], with promising results. Preference is given to test only settings that do not exceed a certain fitness value, which intrinsically minimises beam losses during the optimisation process. Corresponding investigations on the transfer channel should therefore be carried out in the future.

5. Automated lattice construction

Inspired by the successes in automatic tuning of accelerator settings and the improvement of transfer line optics, the question arose whether the design of transport lines could possibly also be partially or completely automated. In frame of this work, a corresponding study was first presented at the international particle conference 2019 [1] and will now be summarised in the following section 5.1. In Section 5.2, a more generalised parameterisation is developed from this approach, which can be applied to a large class of design problems.

5.1. The genetic lattice construction method (GLC)

The goal was to find an algorithm to optimise the geometry and field strengths of the elements of a transport line for a specific purpose. In the approach described in this section, there are three optimisation criteria. The transmission should be maximised and, in parallel, this goal should be achieved with as few elements as possible and with as low field strengths as possible.

5.1.1. Genotype parameterisation

Initially, the problem was artificially limited to quadrupoles and to first-order ion optics. The goal was to find the minimum number of quadrupoles, their optimal position within a transport line of fixed length L and the associated field gradients. A further simplification was to restrict the element selection first to two quadrupole types Q_a and Q_b with the corresponding lengths L_a and L_b, with $L_a < L_b$ and each quadrupole type was used for a different area of values for the normalised integrated field gradient $kl = B'l/(B\rho)$. These areas are distinct and do not overlap.

$$\text{area 1: } (kl)_0 < |kl| \leq (kl)_a$$
$$\text{area 2: } (kl)_a < |kl| \leq (kl)_b$$

The transfer line was divided into n sections so that there would be room for one quadrupole type in each section and still some space left for relative positioning. Since Q_b is the longer element, the maximum number of quadrupole magnets is

$$n = \frac{L}{L_b} - 1 \,. \tag{5.1}$$

A matching genotype that can encode all solutions to this problem requires $2n$ genes per genome. Even genes ν_i represent the $(kl)_i$ value of quadrupole i and odd genes μ_i represent the relative distance between the quadrupoles. The genome is therefore

$$G = \{\mu_1, \nu_1, \mu_2, \nu_2, ..., \mu_n, \nu_n\} \,. \tag{5.2}$$

5.1.2. Phenotype construction

The phenotype of the genome G was constructed in the following way. For all genes ν_i, the element E_i is either one of the standard quadrupoles or a drift D, depending on the value of ν_i:

$$0 < |\nu_i| \le (kl)_0 \implies E_i = D$$
$$(kl)_0 < |\nu_i| \le (kl)_a \implies E_i = Q_a \qquad (5.3)$$
$$(kl)_a < |\nu_i| \le (kl)_b \implies E_i = Q_b$$

For quadrupoles, ν_i is an expression of the kl value. In cases where the kl was lower than the threshold $(kl)_0$, a drift D was used and the length of this D was set to

$$l_i = \frac{|\nu_i|}{(kl)_0} L_a \,. \qquad (5.4)$$

In this way, continuity could be achieved in the transition from the drift line D to the small quadrupole Q_a (Figure 5.1) and a quadrupole with more focusing power is used only,
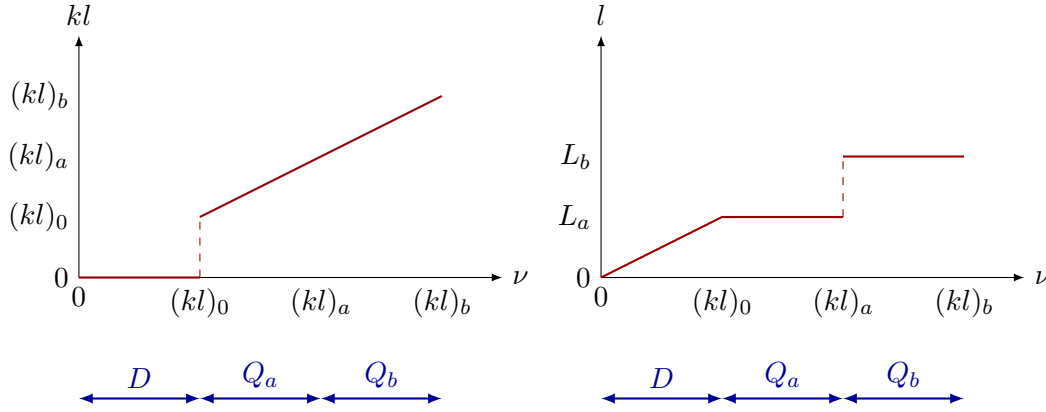
Figure 5.1.: Dependence of kl (left) and l (right) on ν value.

if more focusing power is really needed. The length (l_i) of the constructed quadrupole is defined by its type ($E_i = Q_a \implies l_i = L_a$, $E_i = Q_b \implies l_i = L_b$) and its strength can be calculated from

$$k_i = \frac{\nu_i}{l_i} \,. \qquad (5.5)$$

Table 5.1 contains the properties of the standard element types as they can be extracted from the genome. Once the standard quadrupoles are constructed, the distances between

Table 5.1.: Standard elements E_i (quadrupoles and drift).

element $\mathbf{E}_i$	condition	length l_i	strength k_i
D	$0 < |\nu_i| \le (kl)_0$	$\frac{\nu_i}{(kl)_0} L_a$	0
Q_a	$(kl)_0 < |\nu_i| \le (kl)_a$	L_a	$\frac{\nu_i}{L_a}$
Q_b	$(kl)_a < |\nu_i| \le (kl)_b$	L_b	$\frac{\nu_i}{L_b}$

the elements (and therefore their positions) are calculated. The remaining space is

$$L_r = L - \sum_{i=1}^{n} l_i \tag{5.6}$$

and the distances between the elements are proportional to the share of μ_i,

$$d_i = \frac{\mu_i}{\sum_{i=1}^{n} \mu_i} L_r \,. \tag{5.7}$$

5.1.3. Fitness evaluation

A given genotype had to be tested with regard to its fitness. Therefore the phenotype or rather the lattice of the transfer line was constructed for each genome. If the solution should match a given acceptance, an additional aperture was appended. Subsequently m particles with a given distribution were tracked through each solution and the following indicators were collected:

$$
\begin{array}{lll}
L_p & \dots & \text{the flight distance for each particle} \\
T & \dots & \text{the overall transmission } (N_{end}/N_{init}) \\
N_q & \dots & \text{the resulting number of quadrupoles} \\
\Sigma_{|kl|} & \dots & \text{the sum of absolute value of strengths}
\end{array}
$$

The evaluation function for each genome $F(G)$ was chosen to be the weighted sum of all indicators normalised to 1:

$$
\begin{aligned}
F(G) = {} & \omega_1 \cdot (1 - (1/m) \cdot \Sigma_p(L_p/L)) \\
& + \omega_2 \cdot (1 - T) \\
& + \omega_3 \cdot N_q/n \\
& + \omega_4 \cdot \Sigma_{|kl|}/(n \cdot (kl)_b) \,.
\end{aligned}
\tag{5.8}
$$

The weights ω_i can be chosen according to the priorities of the different indicators. A genetic algorithm was used to minimise F.

5.1.4. Results

The developed algorithm has been tested against theoretical cases (trivial, doublet and target focusing) and for the transfer line between SIS18 and HADES experiment [63] at GSI. A detailed discussion of the results can be found in [1] but should be briefly summarised here again.

Theoretical test cases

In the trivial case, the solution was searched for a transfer line, which for a given particle distribution is short enough that a complete transmission is possible even without additional focusing elements. The algorithm was able to achieve this result reliably in each test run.

In the second test case, the length L was longer, so that additional focusing was needed in both planes. The algorithm was able to find two solutions with quadrupole doublets. In 70 % of the runs, solutions resulted where the doublet starts with a vertically focusing quadrupole and in 30 % with a horizontally focusing quadrupole. The reason for this

asymmetry was that the required quadrupole strength for the first solution was slightly lower due to a corresponding asymmetry of the particle distribution.

In the third test case, a 100 m transfer line was used with a circular aperture limitation of 5 mm (diameter) at the end of it. Three solutions have been found with 100 % transmission, where the prevalence of solutions could be suppressed by varying the weights ω_i.

Transfer line SIS18 to HADES experiment

The last case of study concerned the 160 m long transfer line that transports the beam from the SIS18 synchrotron to the HADES experiment [63]. This transfer line have been under review [64], as well as all the existing GSI transfer lines , due to need for improving the transmission when running at high intensity [65]. The algorithm assumed a 5 cm long target with 2.2 mm diameter [66] at the end of the transfer line. The result obtained with 100 % transmission and smallest F is shown in Table 5.2 against the present transfer line.

Table 5.2.: Both HADES lattices (length ca. 160 m), the existing version (a) and the solution from the algorithm (b), blue are vertical focusing quadrupoles, red elements are horizontal focusing quadrupoles and cyan elements are bending dipoles.

The result for HADES beam line varies significantly from the existing version. The integral quadrupole field and the element number (14 vs. 20) is much lower and the strong quadrupole (Q_b) is only used twice. Still, one wouldn't use this lattice in practice. One reason is, the beam line is also used for other purposes. The dipoles distribute the beam to other target stations and many optics must be supported, including for example one for the pion production. The line is also designed for fast extracted beams, where the horizontal emittance is higher by a factor of 6. One must also consider that for heavy ion operation, the found kl-values result in much higher field strengths, which may then require the use of stronger or possibly additional quadrupoles.

So, even if especially the HADES example is not of much practical use, the investigation showed that it is in principle possible to design transfer lines automatically. However, it is important to specify the boundary conditions and the purpose of the transfer line exactly and calculate with realistic particle distributions. In the following, the method is further developed for more realistic applications.

5.2. Generalisation of the genetic lattice construction method

The further development of the automatic lattice construction method shall now be more adopted to questions that are asked in reality. Usually, there are 3 basic types of requests on beam dynamics with respect to transfer lines.

1 The objective is to identify a new optics or a new general setting for an existing facility to either improve performance or prepare for an application for which the existing facility was not originally designed.

2 How an existing transfer line can be upgraded in the most cost effective way to either improve performance or prepare for a new application.

3 The design of a completely new transfer line with specific imaging properties for given particle distributions at the entrance.

As has been shown, the first question can usually be solved excellently using an autotune algorithm. The other two problems are usually solved in iterative processes that require a certain amount of experience, while the problem itself is often split in logical sections. In both cases, one or more initial particle distributions are specified, which the transfer line should match to a given acceptance. In addition, there may be other boundary conditions, such as the space available for additional elements or specified element types and their properties. In the second case, a large part of the transfer line may be fixed. In the end, a beam line has to be designed which satisfies all the boundary conditions and ensures optimal beam transmission. It would be ideal if the solutions found, proved to be particularly stable against displacement of elements or field errors in the error studies, which are usually carried out in the final design phase.

5.2.1. Adaptations to the parametrisation

The procedure is analogous to that of the autotune algorithm or the first version of Genetic Lattice Construction. First, a suitable parameterisation and a fitness function must be found whose minimum represents the ideal solution of the given problems of the second and third type.

Given a transfer line of length L, in which at least one section of length L_{opt} is free for optimisation. If $L = L_{\mathrm{opt}}$, there are no restrictions on the position of new elements. Furthermore, there is a set of μ freely positionable beam manipulating components $\{C_1, \cdots, C_\mu\}$. These can also be meta-devices consisting of a fixed combination of components. Then those ν components must be added whose positions are fixed from the optimisation point of view. In total, these are $N = \mu + \nu$ components. Each component $C_n \in \{C_1, \cdots, C_{\mu+\nu}\}$ has a certain number of λ_n free parameters $x_{n,1} \cdots x_{n,\lambda_n}$. These parameters can be for example quadrupole strengths, gap voltages but also the length of elements. The total number of free parameters is

$$\eta_p = \sum_{n=1}^{N} \lambda_n \,. \tag{5.9}$$

In addition, a position must now be defined for each movable component from $\{C_1, \cdots, C_\mu\}$. The possibility of arbitrary permutations of the components is to be contained implicitly. For this purpose, a relative position $s_n \in [0, 1]$ is defined for each component.

The actual position l_n results then from

$$l_n = s_n \cdot \left(L_{\mathrm{opt}} - \sum_{n=1}^{\mu} L_n \right) \qquad (5.10)$$

where L_n is the length of the component C_n. Thus, the number of free parameters increases by μ due to the free positioning,

$$\eta = \eta_p + \mu \, . \qquad (5.11)$$

Each solution of the problem corresponds to a point in the η-dimensional vector space. More precisely, it is a subspace of $\mathbb{R}^\eta$, since the parameters are usually normalised to the domain of $[-1, 1]$, respecting the device limits. It actually corresponds to a point

$$\boldsymbol{x} = \begin{pmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_{\eta_p} \\ \tilde{s}_1 \\ \vdots \\ \tilde{s}_\mu \end{pmatrix} \qquad (5.12)$$

in the η-hypercube. Here $\tilde{x}_n$ and $\tilde{s}_n$ are respectively the values of x_n and s_n normalised to the interval $[-1, 1]$. This parametrisation maps to a large class of lattice construction problems. For example, if one wishes to position a quadrupole doublet as a fixed meta-component in a free section, then its two quadrupole strengths would be free parameters, hence $\lambda_1 = 2$. Moreover, $\mu = 1$, $\nu = 0$ and $N = 1$. So there would be a total of $\eta = 3$ free parameters describing this problem completely, namely the two quadrupole strengths x_{q1} and x_{q2} and the relative position of the doublet s. Each possibility corresponds to a position

$$\boldsymbol{x} = \begin{pmatrix} \tilde{x}_{q1} \\ \tilde{x}_{q2} \\ \tilde{s} \end{pmatrix} \qquad (5.13)$$

in the 3-dimensional parameter space. Another example would be a section where a triplet is fixed ($\nu = 1$ and $\lambda_1 = 3$) and a buncher cavity ($\lambda_2 = 1$) and another triplet ($\lambda_3 = 3$) should be added ($\mu = 2$). The number of components is then simply $N = 3$ and the number of free parameters is $\eta = 9$. Each solution of the problem corresponds to a position

$$\boldsymbol{x} = \begin{pmatrix} \tilde{x}_{t11} \\ \tilde{x}_{t12} \\ \tilde{x}_{t13} \\ \tilde{x}_U \\ \tilde{x}_{t21} \\ \tilde{x}_{t22} \\ \tilde{x}_{t23} \\ \tilde{s}_{\mathrm{gap}} \\ \tilde{s}_{t2} \end{pmatrix} \qquad (5.14)$$

in the 9-dimensional parameter space. Thus, for each problem of this class there is the corresponding parameter space and for each point in this space a corresponding fitness

value $F(\boldsymbol{x})$ can be determined by constructing the intrinsically described lattice (phenotype construction) and performing a tracking simulation with the given particle distribution. At the end, all particles are examined to verify that they survived the transport distance and matched the desired target acceptance. From this information a corresponding value for the transmission T is determined. The simplest fitness function is then the important equation

$$F(\boldsymbol{x}) = 1 - T(\boldsymbol{x}) + \epsilon(\boldsymbol{x}) \tag{5.15}$$

for which only a minimisation with the described algorithms has to be performed afterwards. The quantity ϵ is either 0 or an expression for the excitation, if minimisation of the energy consumption is to be emphasised. This could be a norm in the sub vector space $\mathbb{R}^{\eta_p}$, for example

$$\epsilon(\boldsymbol{x}) = ||(\tilde{x}_1, \cdots, \tilde{x}_{\eta_p})|| = \sqrt{\tilde{x}_1^2 + \cdots + \tilde{x}_{\eta_p}^2} \,. \tag{5.16}$$

This method can also be used to minimise the number of components, if necessary. If there are solutions, where the value of a parameter of a component is negligibly small, then it is recommended to accomplish a further optimisation without this component. One could also define a cutoff value and leave the element out of the lattice construction, if its value is fallen below. In general, the simulation speed is not critical for the offline design, but it is nevertheless important to find the most efficient hyper parameters, otherwise the number of simulated macro particles and thus the accuracy may be limited too much.

Another interesting possibility is to include positioning errors or field errors in the phenotype construction of the lattice, to test whether an error-tolerant solution could be found in this way.

In the limit case of a simulation with an infinite number of particles, the fitness function is continuous everywhere, except for the few positions where a swapping of elements takes place. This is a great advantage for optimisation, because small changes in the parameter space result in small changes of the fitness function, which is for example a justification for the use of Gaussian processes as in the BOBYQA algorithm. If the relative position of the free components is fixed and a swapping is not necessary, then F is even continuous everywhere. So if the permutation is not necessary, it should not be enabled. The definition of s_n must then be adapted accordingly in such a way that s_n describes the distances between the components, instead of their relative position.

However, when describing the particular problem, one should always try to minimise the number of dimensions. For example, if the focal length before and after a quadrupole triplet should be the same and thus the first and the last quadrupole should have the same gradient. In this case, it makes no sense to tune three parameters in a triplet where two should be sufficient. The same is true for re-buncher cavities with multiple gaps whose voltages have a fixed ratio between them. So here it is sufficient to optimise one parameter, the total voltage.

5.2.2. Test cases

The principle is to be demonstrated in the following with two simple artificial test cases. The results of different algorithms are compared once again.

A single free quadrupole

Given is a drift section of length $L = 20\,\text{m}$ at the end of which an aperture limitation is attached. This has a diameter of 4 cm horizontally and 10 cm vertically. An asymmetric, divergent particle distribution is chosen such that most of the beam is lost on the aperture limitation (Figure 5.2). For a quadrupole of length $L_q = 1\,\text{m}$, the goal is to find a position
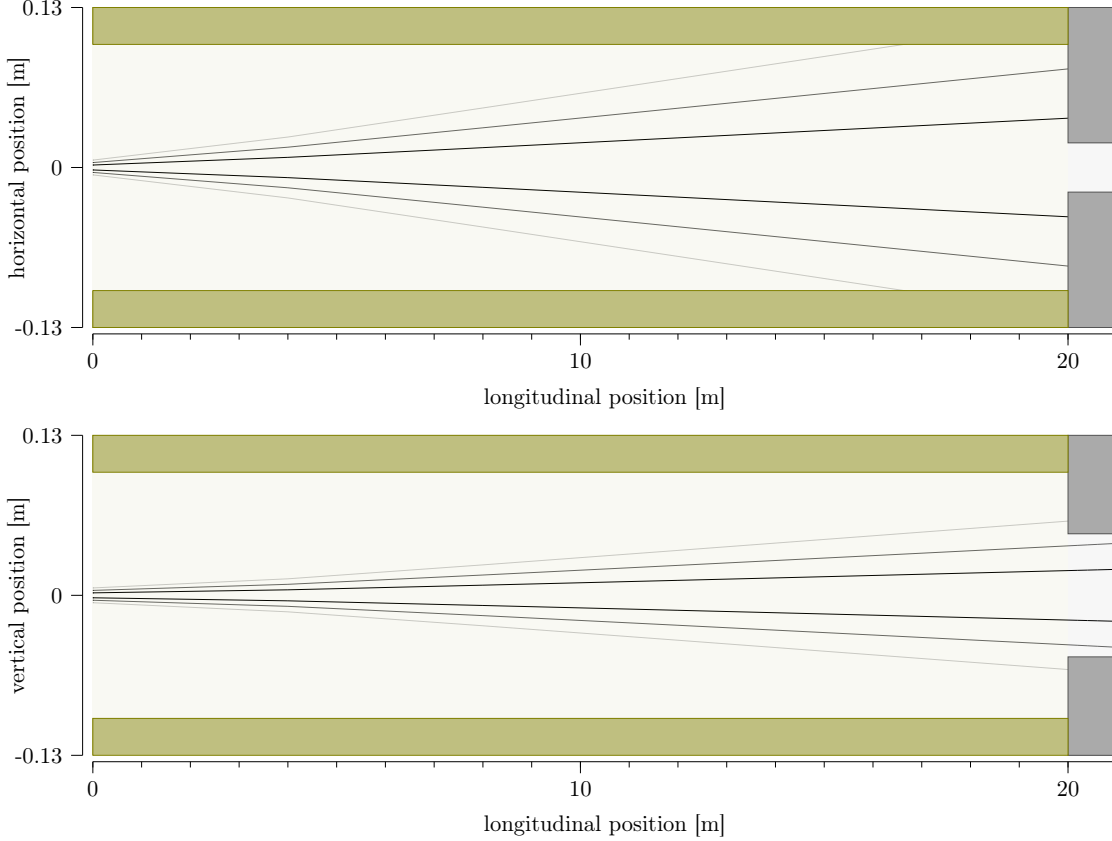

Figure 5.2.: Representation of the initial situation. The beam line is $20\,\text{m}$ long and has a circular cross section with $20\,\text{cm}$ diameter. The beam distribution used is divergent and does not match the final aperture limitation. The transmission is only 31.8%. Within the yellow area, the ideal position and strength for a quadrupole should be found

.

$l = s \cdot (L - L_q)$ within this distance and a gradient (kL_q) at which the transmission becomes maximum. According to equation (5.10) and (5.12) the parameters

$$x_1 = (kL_q)\,, \tag{5.17}$$

$$s_1 = \frac{l}{(L - L_q)} \tag{5.18}$$

are used again for the general optimisation. Thus, it is a 2-dimensional problem $F = F(\tilde{x}_1, \tilde{s}_1) = 1 - T(\tilde{x}_1, \tilde{s}_1)$. The problem is chosen this way because it can still be illustrated very well via a parameter scan (Figure 5.3). Both parameters were scanned within their definition range in steps of $\Delta = 0.005$. For each pair of parameters, a particle tracking simulation was done with 1000 particles and the resulting transmission was determined.
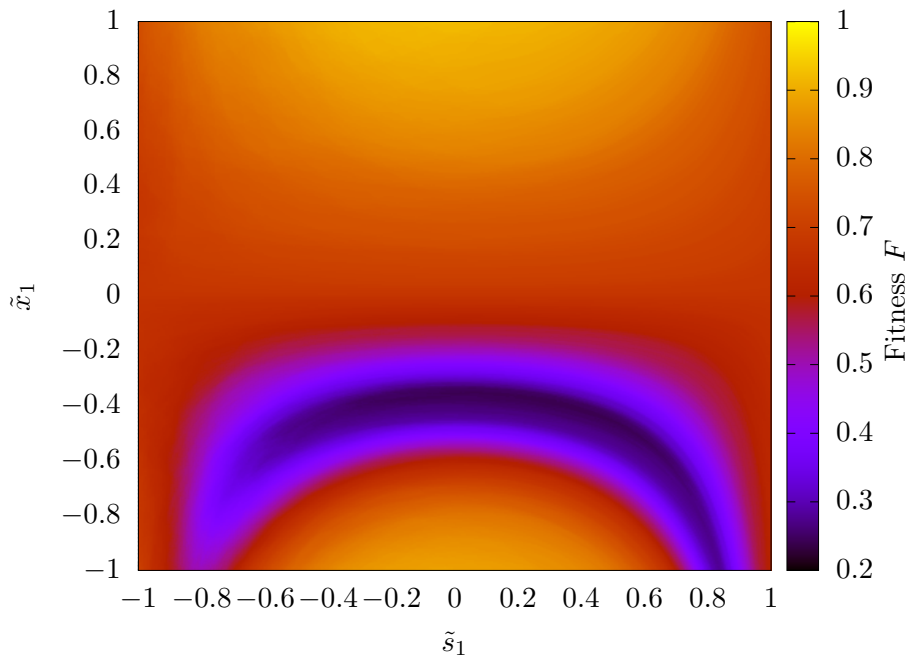
Figure 5.3.: The result of the parameter scan. Shown is the value of the fitness function depending on the normalised parameters $\tilde{s}_1$ and $\tilde{x}_1$. The minimum is located at $\tilde{x}_1 = -0.35$, $\tilde{s}_1 = 0.135$

For the scan, 400 settings were tested on 400 different lattices. Thus 160000 tracking simulations had to be performed.

As can be seen in Figure 5.4, there is a large range with respect to position and field strength in which similarly good transmissions can be achieved. Position and field strength are strongly correlated in this range. In the end, however, there is only one optimum. The best possible transmission in this grid is $T = 74.5\%$ and is located at the position

$$\tilde{x}_1 = -0.35\,,$$
$$\tilde{s}_1 = 0.135\,.$$

The corresponding position of the quadrupole is $l = 10.78\,\mathrm{m}$ and the normalised quadrupole strength $(kL_q) = -0.175/\mathrm{m}^2$. Since in this case no secondary minima exist, the minimum of the fitness function is very reliable and relatively easy to determine. Even a gradient method can be considered.

The optimisation was performed with different methods. The results are listed in Table 5.3. There is a large variance in the number of necessary optimisation cycles and thus the number of necessary tracking simulations for all natural analog meta-heuristics, which is due to the fact that there is a large chance that the randomly selected initial values are already relatively close to the optimum. In these cases all algorithms are converging very quickly. Even though the results are only comparable to a limited extent due to the arbitrarily selected hyper-parameters, the picture is similar to all previous investigations. It can be said that the genetic algorithm is very reliable and converges fastest among the statistical methods. The BOBYQA optimisation is nevertheless significantly faster and requires 2 orders of magnitude less tracking simulations. However, depending on the choice of starting points, BOBYQA is not very reliable in finding the correct minimum.
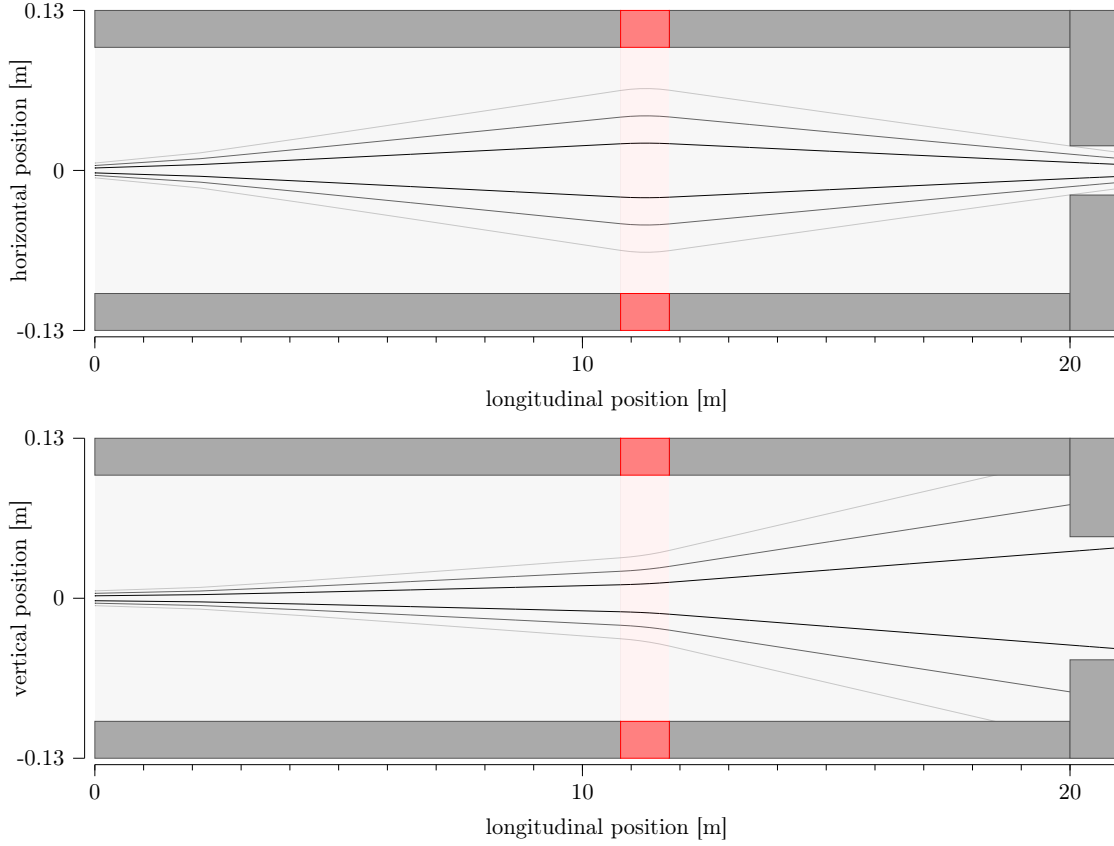
Figure 5.4.: Representation of the solution with the maximum achievable transmission of $T = 74.5\%$ in this test case, as it is also found by all nature analogous algorithms. The black lines represent 1σ of the distribution, the dark grey 2σ and the light gray 3σ.

Table 5.3.: Comparison of the optimisation methods with respect to their performance. All heuristic algorithms are reliably able to find the optimum and differ only in the number of necessary optimisation cycles and therefore in the number of necessary tracking simulations. For the comparison, no special optimisation of the algorithms' hyper-parameters was performed. Experience-based, reasonable default values were chosen. Nevertheless, it is reasonable to assume that in some cases better performance would be possible after tuning the hyper-parameters. The table contains the mean transmission value $\overline{T}$ and the best transmission $T_{\max}$ found by the algorithms within 10 optimisations.

optimisation method	$\overline{T}$	$T_{\max}$	# tracking simulations
gradient descent	74.13%	74.30%	63 ± 12
scan $\Delta = 0.005$	74.50%	74.50%	160000 ± 0
genetic algorithm	74.43%	74.50%	1790 ± 1048
particle swarm	74.50%	74.50%	7500 ± 6100
simulated annealing	74.48%	74.50%	8600 ± 5800
BOBYQA	72.78%	74.30%	26 ± 6

The same is true for the gradient algorithm. The cause is the same in both cases and it is related to the finite number of particles used in the tracking simulations. When determining the gradient, it inevitably comes to the point that the selected step size is either too small to cause a difference in transmission, causing the gradient to disappear, or it is too large and the minimum is repeatedly skipped, causing the gradient to be determined too low. When using model functions, a basic assumption is that the functions are continuous and there are no jumps. Of course, this is only true in the limit case of infinitely many particles.

A buncher and two quadrupoles

In the second example, two quadrupoles and a re-buncher are to be tuned. While one of the quadrupoles gets a fixed position (Figure 5.5), the position of the other one, as well as that of the re-buncher, can be chosen freely. For simplicity the buncher has only one accelerating gap. The particle distribution is again chosen such, that it cannot be transported to 100% without additional elements. A $2\,\mathrm{MeV/u}$ bunched proton beam is simulated with $1\,\mathrm{mA}$ pulse current considering linear space charge forces. The base frequency of the buncher is set to $100\,\mathrm{MHz}$.

The number of free parameters in this example is $\eta = 5$ and therefore the solution set is a 5-hypercube. Each solution can be described by a point

$$\boldsymbol{x} = \begin{pmatrix} \tilde{x}_{\mathrm{q1}} \\ \tilde{x}_{\mathrm{q2}} \\ \tilde{x}_{\mathrm{U}} \\ \tilde{s}_{\mathrm{q2}} \\ \tilde{s}_{\mathrm{buncher}} \end{pmatrix} \tag{5.19}$$

in the penteract. Due to the high dimension, the problem can no longer be solved by a scan. A scan, equivalent to the first example, would require 10^{13} particle tracking simulations. Assuming that such a simulation would take $1\,\mathrm{ms}$, the scan would take 317 years to complete. Illustrating a 5-dimensional problem is also difficult. However, if three parameters are kept fixed, at least projections on different 2-dimensional subspaces can be generated and also visualised (Figure 5.7).

The study was now conducted in such a way that a hyper-parameter scan was first performed in rough steps for each algorithm. The associated tracking simulations were performed with only 100 particles and no space charge. With the best hyper-parameters found, the test case was then optimised. The best solution is shown in Figure 5.6.

Although it is no longer possible to verify that this solution is the optimum, without performing a reference scan, it seems very plausible. If the phase acceptance were exactly identical with the initial phase width of the beam, then the gap would have to be located exactly in the centre. However, the phase acceptance is slightly larger, so a position slightly in front of, or behind the centre is reasonable. The second quadrupole must be positioned as far forward as possible, so that the beam can be well focused transversely through the drift tube of the re-buncher. Table 5.4 once again compares the performance of the different algorithms for this particular example.

Given 5 dimensions, the combinatorics yields $\binom{5}{2} = 10$ possible 2-dimensional subspaces. To get a better understanding of the problem and its parameter space, scans were performed for 4 different projections on 2-dimensional subspaces while keeping the remaining dimensions fixed (Figure 5.7). The projection on the quadrupole strengths $F(\tilde{x}_{q2}, \tilde{x}_{q1})$
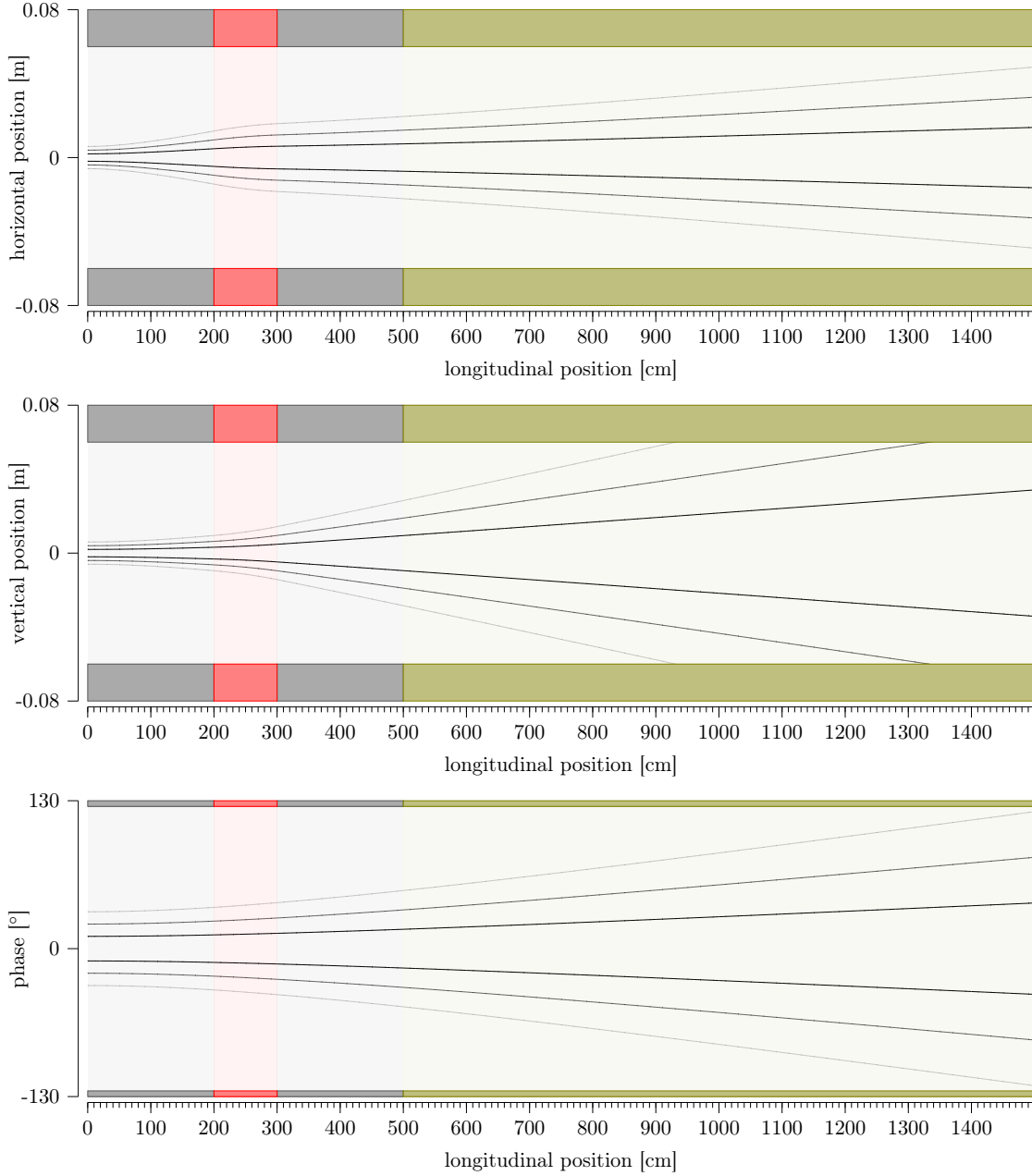
Figure 5.5.: Representation of the initial situation. The beam line is 150 cm long and has a circular cross section with 12 cm diameter. The beam distribution used is divergent and does not match the final aperture limitation. The transmission is only 3.3%. Within the yellow area, the ideal position for a second quadrupole and a re-buncher should be found, while also tuning the quadrupole strengths and the gap voltage.

shows directly that F has more than one minimum. Moreover, both local minima in this projection are far from the limits with respect to the quadrupole strengths. Thus, one could use much weaker quadrupoles in this example. However, further insight can be gained from the other projections. For example, the minima of $F(\tilde{s}_{q2}, \tilde{s}_b)$ and $F(\tilde{x}_{q2}, \tilde{s}_{q2})$ are both at the edge, where $\tilde{s}_{q2} = -1$. So, to improve the transmission, one should try to install the second quadrupole even further forward, even if it deviates from the requirements spe-
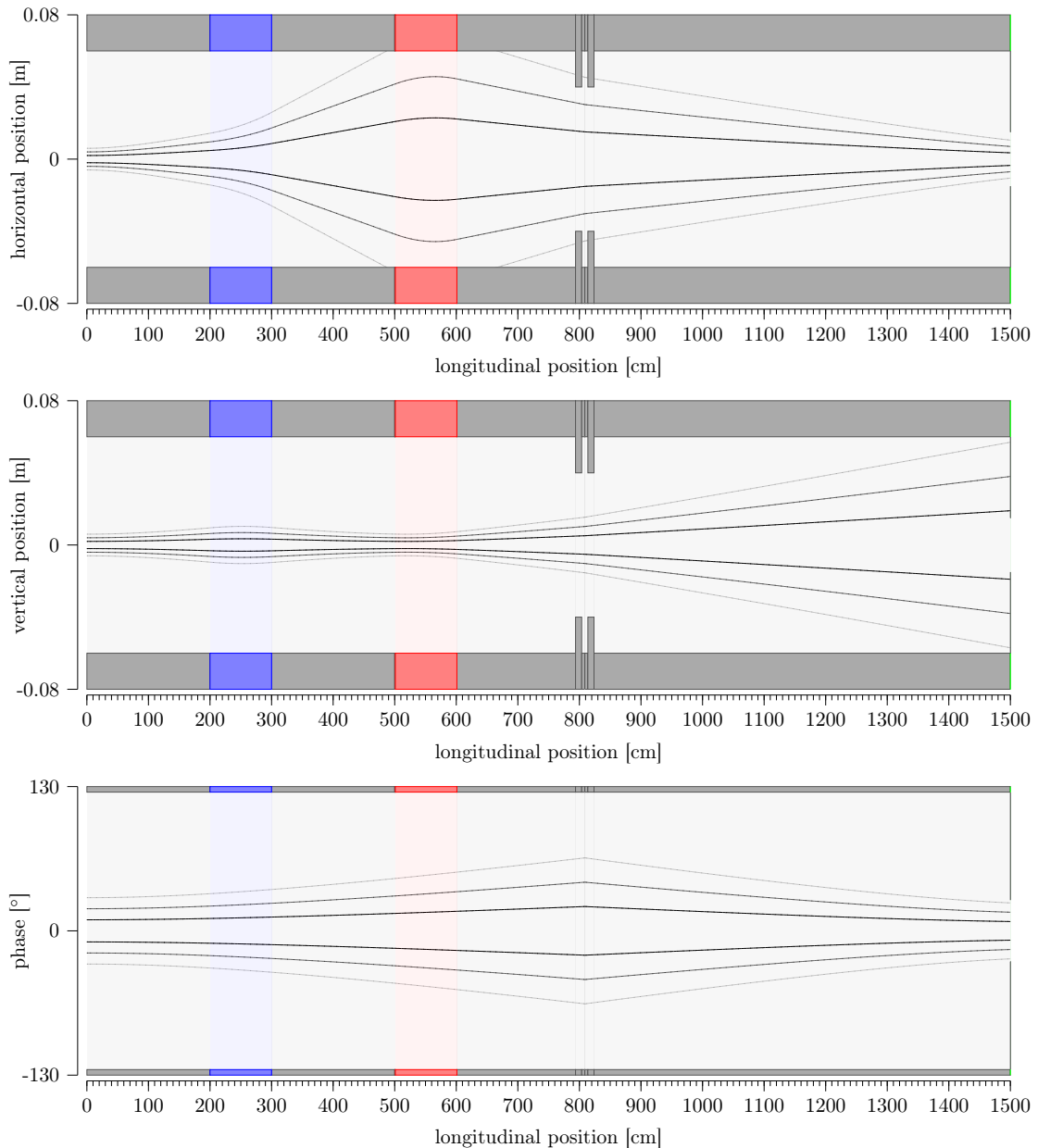
Figure 5.6.: Representation of the best solution that was found. The second quadrupole is positioned directly at the first possible position in this solution. The gap is located approximately in the middle. The transmission of this solution is $T = 48.1\%$.

cified initially. In the projection $F(\tilde{s}_b, \tilde{x}_b)$ one can see that both, position and voltage of the re-buncher cover a large range with maximum fitness. So the example provides some flexibility in the design.

So, for a found solution, it can be useful to run some scans of subspaces to study the effects of boundary conditions and find optimisation opportunities that lie outside the original problem specification.

Table 5.4.: Comparison of the optimisation methods with respect to their performance. In this case, too, the genetic algorithm is best suited to find the optimum, but the BOBYQA optimisation requires much fewer tracking simulations, while its solutions never reach the quality of those found by the genetic optimisation.

optimisation method	$\overline{T}$	$T_{\max}$	# tracking simulations
genetic algorithm	47.0%	48.1%	37440 ± 0
particle swarm	43.3%	47.0%	16900 ± 6568
simulated annealing	44.5%	47.0%	51290 ± 23261
BOBYQA	12.3%	44.3%	123 ± 13
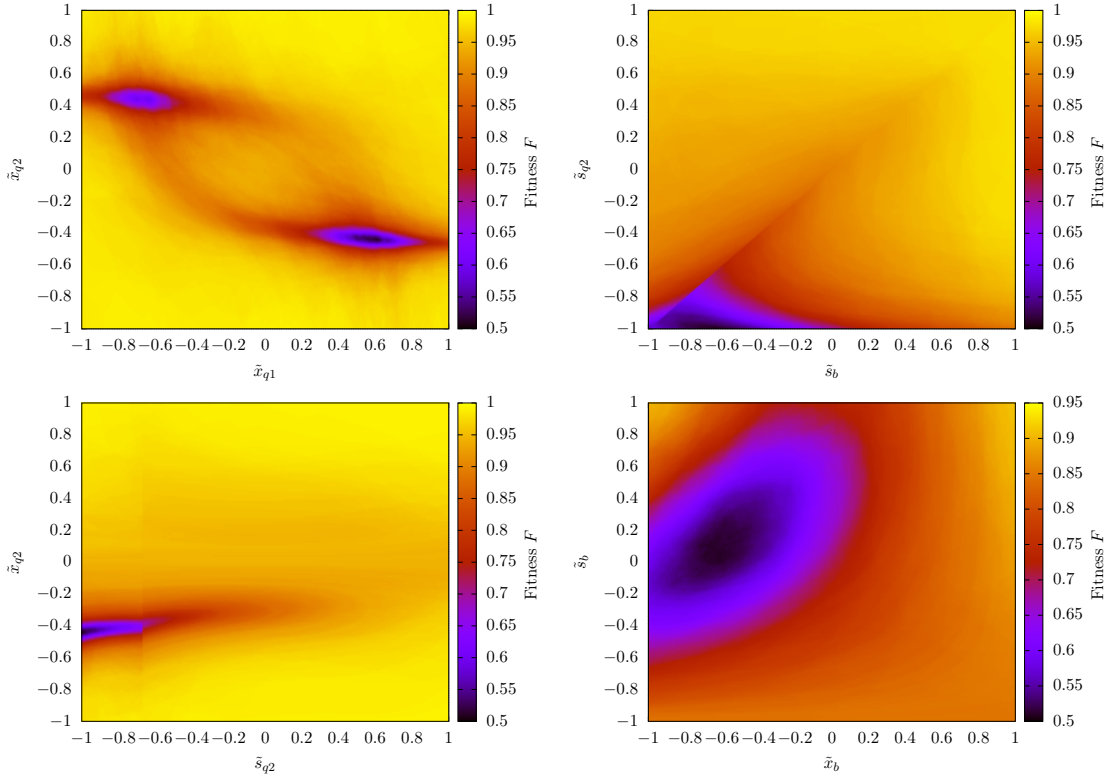
Figure 5.7.: The result of 4 parameter scans of 2 dimensional sub-spaces of the parameter space with respect to the fitness function. (top left: scan of the quadrupole strengths, top right: position of the second quadrupole vs. buncher position, bottom left: position vs. strength of the second quadrupole, bottom right: position vs. voltage of the re-buncher)

Integration of error studies

For error studies, a lot of simulations have to be performed in later design phases to understand the influence of field or positioning errors and to minimise them. For example, in the design of the prototype for the replacement of the post-stripper Alvarez accelerator at GSI-UNILAC, 1000 lattices with different positioning errors were generated and in this way it was tested how stable the design is against alignment errors in terms of emittance growth [67]. Therefore, it seems obvious to include the error studies right away in an automatic design.

The idea is now to introduce random positioning errors for all optical elements each time the phenotype construction of the lattice is done during the optimisation process. The hypothesis is, during the process only those solutions will prevail in the long term that are less susceptible to positioning errors and therefore still produce good transmission in the subsequent cycle.

This is now to be investigated with the preceding example. First, an error study was performed for the best solution found so far (Figure 5.8). A total of 50000 transfer lines
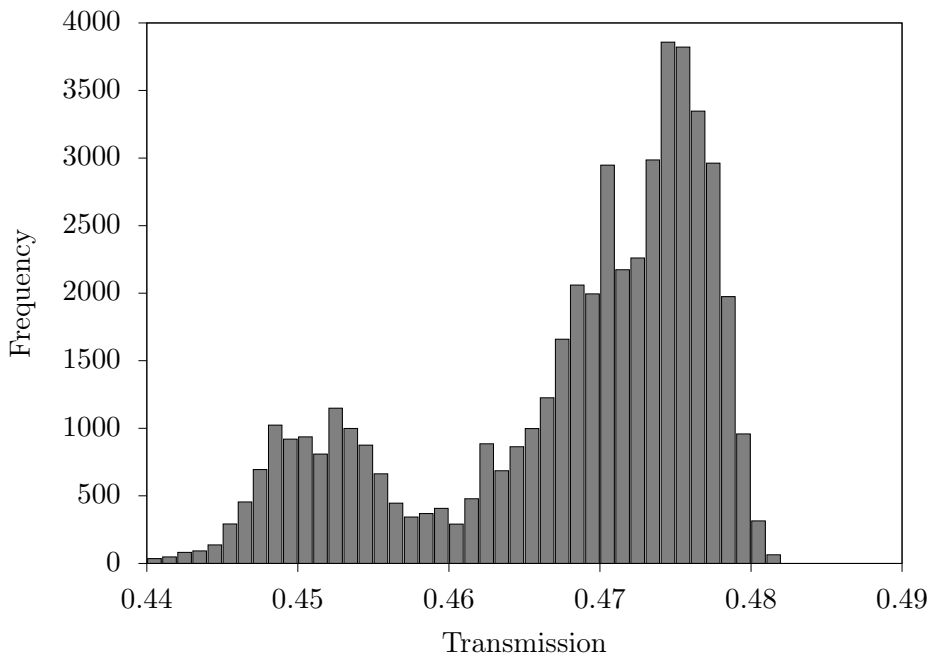
Figure 5.8.: The histogram of the error study using the result from example 2 of the previous section. A total of 50000 accelerators with random alignment errors were generated and the transmission was measured. The mean value of the transmission is $\overline{T} = 46.71\%$.

were generated and all optical elements were provided with positioning errors normally distributed around the ideal position. The mean deviation from the ideal position in all directions corresponds to $\sigma = 0.1\,\text{mm}$, concerning the angles $\sigma = 0.1\,\text{mrad}$. All spatial directions and all rotational axes were varied. Each tracking simulation was performed analogous to those in the optimisation with 1000 particles using the same distribution. As expected, it can be seen that the ideal transmission with positioning errors is less likely to be achieved. The transmission is between 44% and 48.1% the mean value is $\overline{T} = 0.467$.

Subsequently, the complete optimisation of the transport line was performed again, with equivalent positioning errors being integrated directly during the evaluation. The best result of this optimisation was then subjected to another error study. The comparative distribution of the transmission can be seen in Figure 5.9. The mean value is $\overline{T} = 0.467$ and therefore identical to that of the unperturbed optimisation. An interesting result can be found at the upper end of the distribution. The perturbed optimisation apparently produces a result which, in the case of misalignment, also makes higher transmissions possible than are at all possible with an unperturbed machine. But this result is a feature of the chosen example and cannot necessarily be generalised. The optimal solution is in the parameter space in some dimensions at the edge of the penteract. A de-adjustment
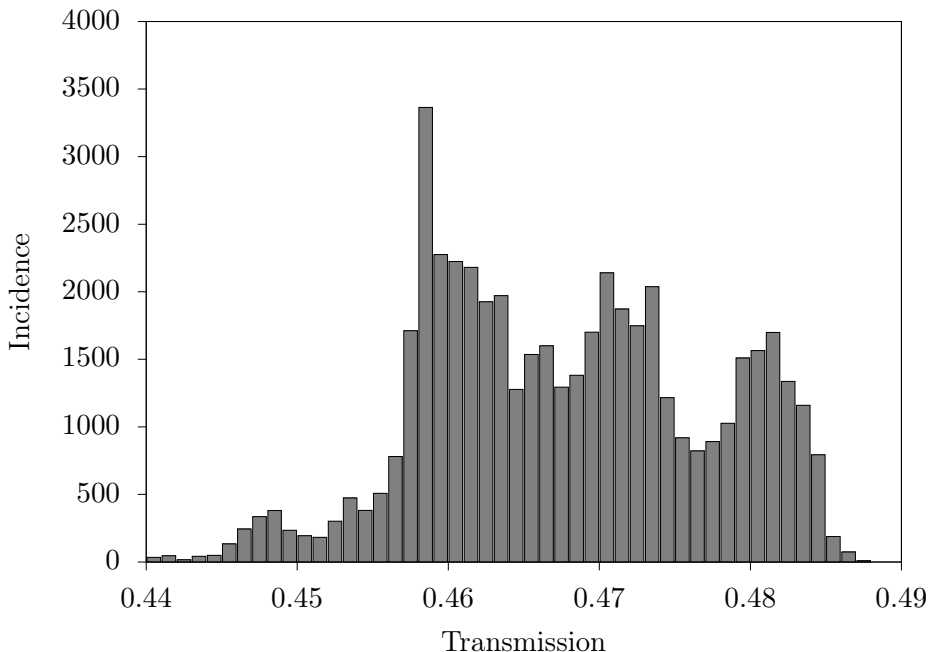
Figure 5.9.: The histogram of the error study using the result of the perturbed optimisation. A total of 50000 accelerators with random alignment errors were again generated and the transmission was measured. The mean value of the transmission is the same as for the unperturbed optimisation $\overline{T} = 46.71\%$.

in this case can therefore lead to the realisation of solutions outside the penteract, which in the end allow a higher transmission. As a measure of the robustness of a solution, the standard deviation from the unperturbed maximum transmission can be used. This can be found next to the standard deviation from the mean transmission in Table 5.5. An

Table 5.5.: Comparison of the lattice optimisation with and without position errors. The table contains the respective mean values of the transmission, the standard deviation from the undisturbed optimum and the standard deviation from the mean value.

optimisation	$\overline{T}$	$\sqrt{EW((T - T_{\mathrm{opt}})^2)}$	$\sqrt{EW((T - \overline{T})^2)}$
without errors	46.71%	1.7%	1.00%
with errors	46.71%	1.6%	0.95%

optimisation with position errors leads in this case to a slightly more robust transfer line with respect to positioning errors. To what extent the result can be generalised to other configurations remains to be investigated. A positive effect can be assumed if there are 2 or more similarly good solutions, and the optimal solution in comparison can only be realised on a small area of the parameter space or even a point in the parameter space, while alternative solutions have a larger area of the parameter space available. So it makes sense in any case to perform the optimisation once with and once without disturbances and to compare the results in error studies.

5.3. Applications

In the following, the presented method is applied to two problems, each of which comes from a real accelerator project.

5.3.1. Optimisation of the CERN H2-beam line using Gabor lenses

NA61/SHINE (SPS Heavy Ion and Neutrino Experiment) [68] is a multi-purpose experimental facility to study hadron production in hadron-proton, hadron-nucleus and nucleus-nucleus collisions at the North area of CERN Super Proton Synchrotron (Figure 5.10).
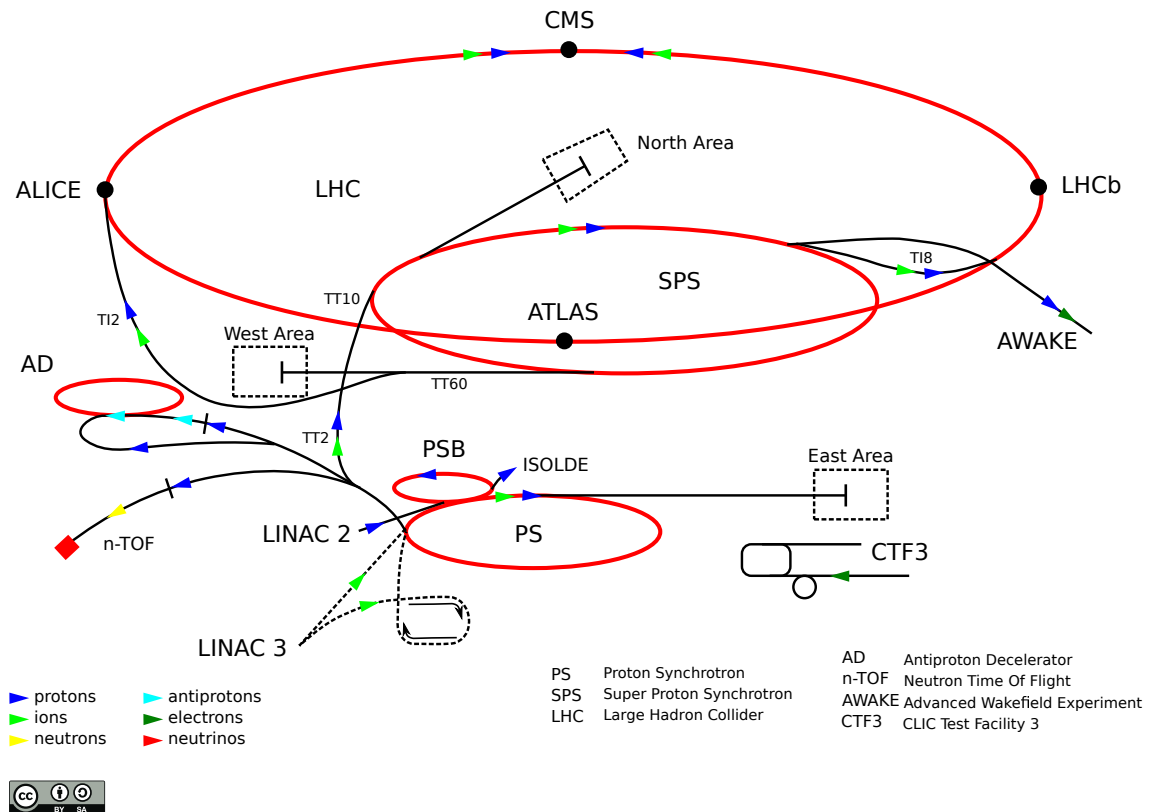
Figure 5.10.: The CERN accelerator complex by Forthommel.

After extraction, the beam is transported over about 1 km. The last part of transfer line from SPS towards NA61 area (H2-line) is a precise magnetic spectrometer. The target is located 8 meter below the surface and is followed downstream by a fragment separator with an acceptance of $\frac{dp}{p} = 2\%$. The transfer line was optimised for several purposes and optimised optics for different physics configurations have been prepared which also led to very fast optimisation times. The H2 beam line is prepared to transport charged particles in a wide range of momenta from ca. 30 GeV/c up to the top SPS energy of 400 GeV.

Presently the NA61/SHINE collaboration is exploring the potential physics opportunities for NA61 in the 1-20 GeV/c region. But for lower momenta (≤ 30 GeV/c), the line was considered not to be optimally designed. Those energies are produced with very large transverse angles. So the spot size and divergence at NA61 is affected. Additionally, old power supplies with a deviation of 0.2 A could lead to $\frac{dp}{p} = 0.6\%$ and corresponding 50% acceptance at 10 GeV/c and even worse with lower momenta. The Idea is, to create the low energy particles closer to the experiment using a secondary target, momentum-select
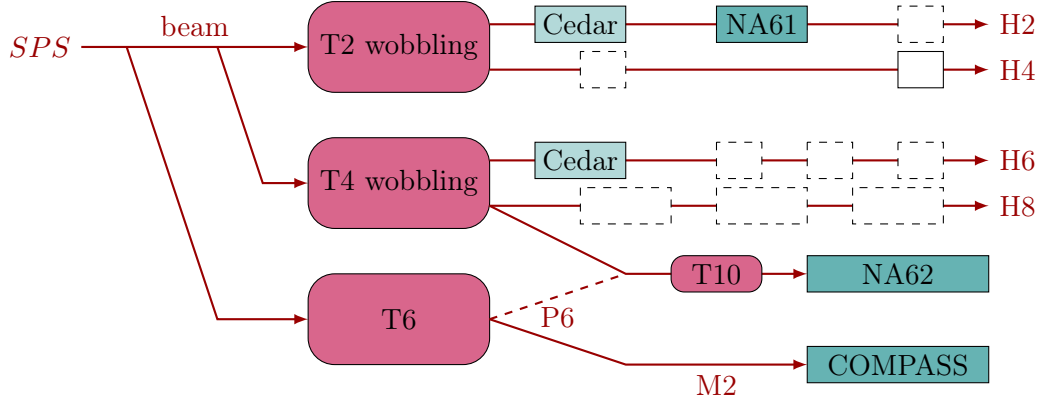
Figure 5.11.: Schematic view of the CERN north area [69]. The transfer lines H2-H8 are located in the Experimental Hall North 1 (EHN1), a 300 m long industrial type building. H2 contains a precise magnetic spectrometer, designed for high momenta >300 GeV/c. H4 is considered to be the test beam line for a possible upgrade with Gabor lenses.

particles between 0.5-12 GeV/c and transport them to the final target. In any case, such a production target will produce particles with high momentum spread and also widen the transverse size of the beam.

First particles lost are usually part of the halo. So one option is, to equip the H2-line with Gabor lenses to open a transport channel also for halo particles and therefore improve the transmission significantly. The focal length of such a lens can be estimated using equation (2.78). For four Gabor lenses of 2 m length each, with an electron density of $\eta = 1 \times 10^{15}/\text{m}^3$ the effect on a proton beam of different energies is listed in Table 5.6. In all high energy cases, the focal lengths are relatively large. For momenta between

Table 5.6.: The estimated focal lengths for a proton beam with different beam momenta in a Gabor lens of 8 m and an electron density of $\eta = 1 \times 10^{15}/\text{m}^3$.

Momentum	Energy	β	γ	focal length
1 GeV/c	432.9 MeV	0.729256	1.46147	10 m
5 GeV/c	4.149 GeV	0.982845	5.42196	68 m
10 GeV/c	9.106 GeV	0.995627	10.7047	138 m
20 GeV/c	19.08 GeV	0.998901	21.3392	276 m
80 GeV/c	79.07 GeV	0.999931	85.2690	1105 m
140 GeV/c	139.1 GeV	0.999978	149.214	1934 m

80 GeV/c and 140 GeV/c this is 1-2 km. So for high momenta the expected focusing effect is rather low, but for the aimed low energy beams, the installation of Gabor lenses could avoid the need for the installation of additional quadrupoles, especially since the charge density of a Gabor lens is a tuneable parameter. Gabor lenses also have the advantage that they focus the beam in both planes simultaneously. Additionally the focal lengths depends linearly on the charge density. A factor of two shortens the focal length at 80 GeV/c for example by half a kilometre.

It can be assumed that the best position of the lenses will be momentum-dependent and one has to place them one focus length before the final target, if possible. Since the

distance between collimator 9 (which is the starting point of the H4 lattice) and the final target is only about 200 m, an optimal focusing effect can be expected for all momenta below 15 GeV/c. Which means that for higher momenta, they should be placed close to collimator 9. First simulations were done under this assumption, using the H4 beam line, which is considered to be the test bench for a possible upgrade with Gabor lenses. The lattice data are imported from H4 Beatch-file [70] and the particle distribution was derived from Twiss-parameters of the official H4 filter mode optics [71] starting with a horizontal focus at collimator 9.

Positioning of a single Gabor element

To keep the dimension of the parameter space small, a single Gabor lens of 8 m length shall be positioned first. Since a Gabor lens changes the optics of the transfer line, at least the last two quadrupoles at the beginning of the drift line have to be tuned as well to allow a reasonable comparison of the results. Thus, it is only a 3-dimensional problem and

$$\boldsymbol{x} = \begin{pmatrix} \tilde{x}_{\mathrm{q1}} \\ \tilde{x}_{\mathrm{q2}} \\ \tilde{s}_{\mathrm{gabor}} \end{pmatrix} . \tag{5.20}$$

The goal of the optimisation is to achieve a beam as parallel as possible on the target. To translate this into a transmission, an artificial aperture limitation must be inserted before and after the target. An aperture limitation with a diameter of 1 cm and a total length of 10 m limits the maximum angular divergence at the target to 1 mrad. In this way a rather parallel beam with a focus on the target is automatically achieved, by just optimising transmission. The degree of parallelism can then be adjusted by varying the diameter of the aperture limitation. Ultimately, a tradeoff between parallelism and transmission has to be found or respectively, the maximum angular divergence of the beam at the target must be specified beforehand.

For an estimated particle distribution, the optimisation was performed for beams with momenta of 1 GeV/c, 5 GeV/c, and 10 GeV/c, respectively. The result for 5 GeV/c is exemplarily shown in Figure 5.12 and the corresponding values for all three momenta are listed in Table 5.7. The respective distances of the Gabor lenses from the target

Table 5.7.: The results of the optimisation. Listed are the ideal positions of the Gabor lens for different beam momenta, the strengths of the two quadrupoles and the maximum transmissions achievable with this lattice. For comparison, the maximum achievable transmission without an additional Gabor lens is 15.71 %.

Momentum	s_{gabor}	k_{q1}	k_{q2}	$T_{\max}$
1 GeV/c	170.95 m	$-0.0059/\mathrm{m}^2$	$-0.0106/\mathrm{m}^2$	68.93 %
5 GeV/c	125.64 m	$-0.0191/\mathrm{m}^2$	$0.0031/\mathrm{m}^2$	51.50 %
10 GeV/c	101.45 m	$-0.0178/\mathrm{m}^2$	$0.0015/\mathrm{m}^2$	27.93 %

correspond approximately to the estimates from Table 5.6. The effect of the Gabor lens on the transmission is very clearly visible in all three cases. As expected, the effect is much stronger at low energies.
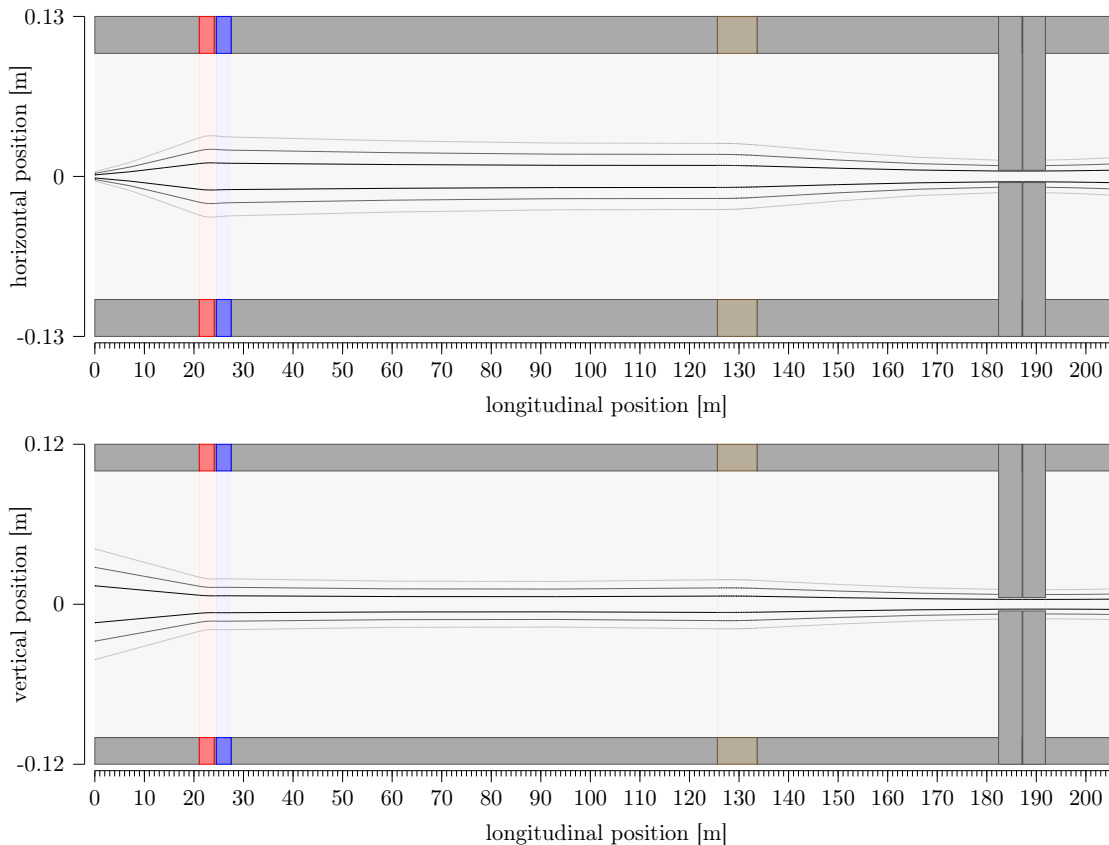
Figure 5.12.: Optimised optics of the last section of the H4 beam line for a $5\,\mathrm{GeV}/c$ proton beam. An $8\,\mathrm{m}$ long Gabor lens (brown) was used, which was positioned $125\,\mathrm{m}$ behind the starting point (collimator 9). The transmission could be increased from 16% to 52%. The target MWPC0910 is located at position $187.27\,\mathrm{m}$.

Free positioning of all 4 Gabor elements

The four actually available Gabor lenses have a length of $2\,\mathrm{m}$ each and can be positioned individually. The minimum distance for the mount is $0.5\,\mathrm{m}$. In this second optimisation, the positions of all four lenses plus the two quadrupole strengths have now been optimised. The dimension of the parameter space thus increases to six.

The optimisation was performed for different beam momenta of between $1\,\mathrm{GeV}/c$ and $15\,\mathrm{GeV}/c$. In all cases, solutions in which the lenses were positioned immediately behind each other prevailed (Figure 5.13). From $8\,\mathrm{GeV}/c$ the position of the array is about $100\,\mathrm{m}$, which is about half the distance between collimator 9 and the aperture limiting element. Many equivalent solutions with multiple intermediate foci can be found for $1\,\mathrm{GeV}/c$, however, it would make more sense to adjust the electron density than the position in these cases. At higher energies, the transmission becomes increasingly insensitive to the exact position of the lenses, resulting in sections of several meters where the lenses can be freely positioned. The simulation was made with 2000 protons. As long as no additional particle survives the aperture limit by shifting a lens, the positions are equivalent. To determine the ideal position of the lenses more precisely at high energies, correspondingly more particles must be tracked. The maximum transmission for different beam momenta can be seen in Figure 5.14.
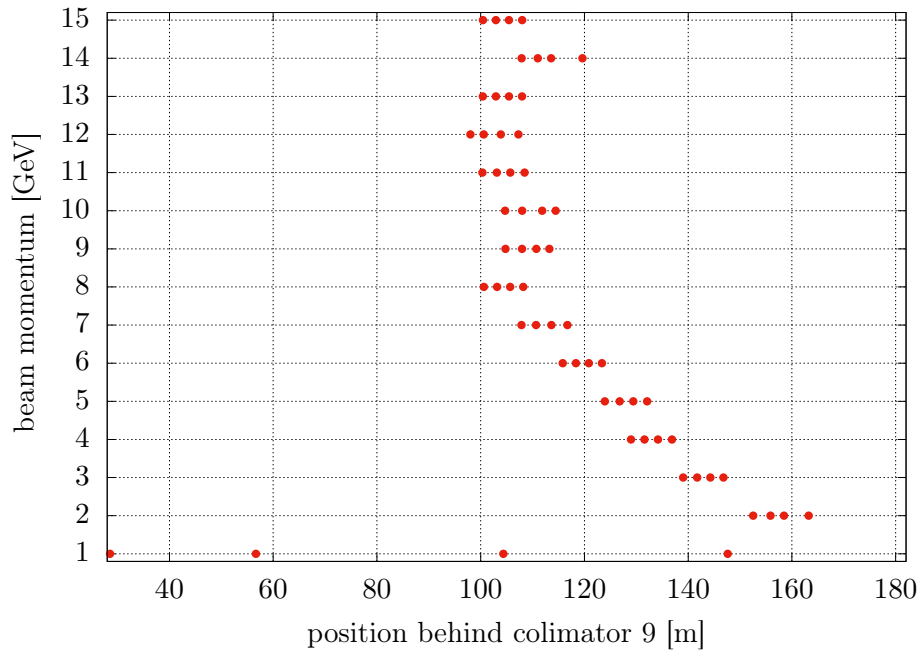
Figure 5.13.: Optimised position of the 4 Gabor lenses in the H4 beam line at different beam momenta. From about $8\,\mathrm{GeV}/c$ the position of the array is stable at about $100\,\mathrm{m}$. The unusual distribution of lenses at $1\,\mathrm{GeV}/c$, results from overfocusing at very low beam momenta ($p < 2\,\mathrm{GeV}/c$).
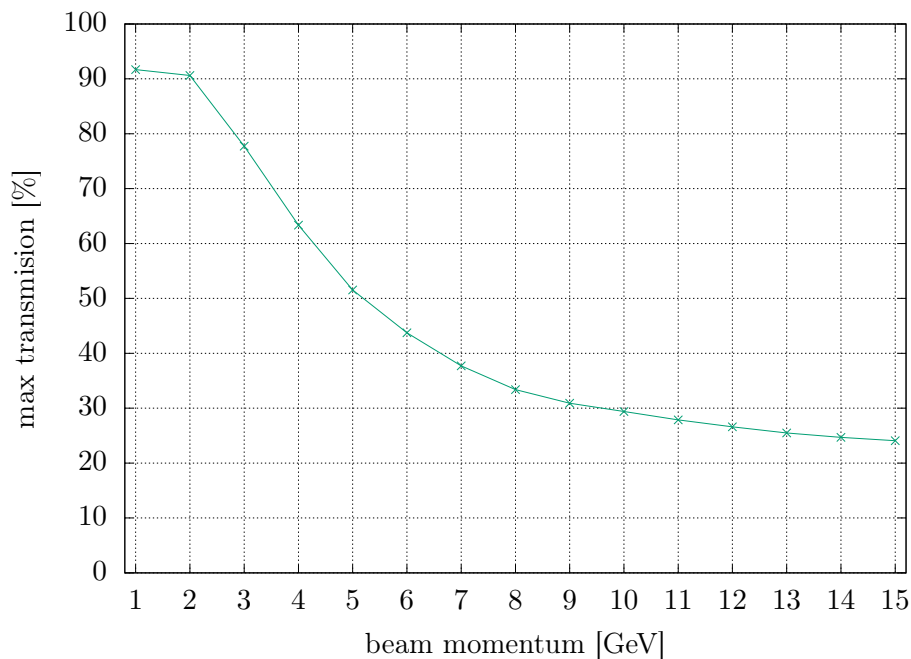

Figure 5.14.: The maximum transmission, achievable with four Gabor lenses at different beam momenta. The reference transmission without Gabor lenses is 16%. Since, according to equation (2.78), the focusing power of a Gabor lens is $\frac{1}{f} \propto \frac{1}{\gamma}$, the effect of the Gabor lens approaches 0 with $\frac{1}{\gamma}$ as the momentum increases. To compensate for the drop in transmission at higher energies, additional lenses would have to be installed.

The transmission drops very quickly with increasing beam momentum and then slowly approaches the value of $T = 16\%$, which corresponds to a beam path without additional lenses. If good transmissions are also to be achieved at higher energies, then additional lenses would have to be installed. Figure 5.15 shows the results of optimisations with different numbers of Gabor lenses at a beam momentum of $5\,\text{GeV}/c$. Here, between one
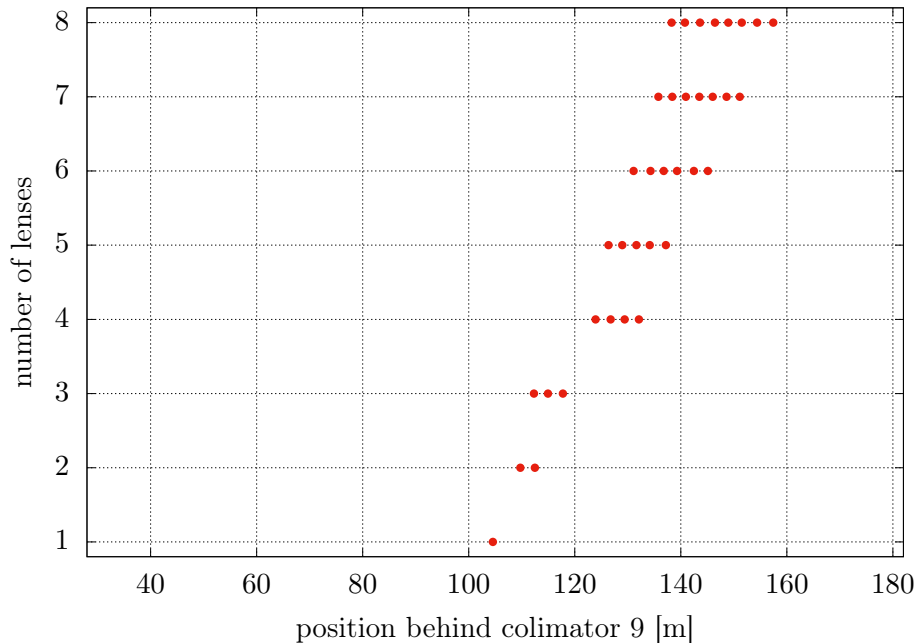
Figure 5.15.: Optimised position of Gabor lenses at $5\,\text{GeV}/c$. As the number of lenses increases, the array must be positioned closer and closer to the target. In the limit case of sufficient focusing power, positioning directly in front of the target is recommended.

and eight lenses were positioned. It can be seen that even with a higher number of lenses, it is favourable to position them directly behind each other. However, the optimal position of the lens array shifts further back as the number increases. Adding lenses increases the integral focusing power and thus decreases the focal length.

Conclusion

With the use of Gabor lenses, an improvement in transmission and beam properties can theoretically be achieved in the H4 beam line. However, it must be said that the same result can also be achieved with quadrupoles. Although these are much more expensive, including the power supply unit, their positioning is not dependent on the beam momentum. But this fact does not matter in practice, because the electron density in the Gabor lens can also be varied. So a position has to be chosen which is optimal for the highest energy used. For low energies, then, only the electron density in the lens needs to be reduced.

Quadrupole doublets also have certain disadvantage compared to Gabor lenses, especially for transversely particularly wide beams. Since the first quadrupole can only focus in one plane and defocuses in the other plane, additional particles are lost in the defocusing plane in contrast to Gabor lenses where there is no defocusing plane.

The parameter space of this problem has only three dimensions and there are at most two local optima due to the quadrupole doublet. In any case, there is one solution that

stands out. For each configuration (energy and number of lenses), 20 optimisations were made and each time the same solution could be found. Depending on the specification of the beam properties at the target and on the input distribution, the ideal position of the Gabor lenses can be determined with the GLC method.

Based on the present results, it is recommended to position the existing 4 lenses centrally between the last quadrupole and the target. This is the ideal position for momenta above $8\,\mathrm{GeV}/c$. If the transmission is to be optimised for smaller beam momenta, a position further back can be selected, or alternatively the electron density of the lenses must be reduced. This also applies to further expansion stages, using more lenses. The optics in the previous beam line can remain unchanged, only the last doublet has to be adopted.

5.3.2. Construction of a matching section for HBS linac

The accelerator-driven High Brilliance Neutron Source (HBS) [72] is to be built at Forschungszentrum Jülich by the end of the decade. The facility will produce particularly small and high-intensity beams with high brilliance and will be used in particular to study nanostructures and biological materials. The driver is a 70 MeV linear accelerator for protons with an intensity of 100 mA with a duty cycle of 6%. For effective operation, the beam is distributed to 3 individual target stations via a proton multiplexer, operated at 3 different macro pulse frequencies (384 Hz, 96 Hz and 24 Hz), to serve the specific experimental requirements. The average beam power is 420 kW and the peak beam power can reach up to 7 MW.

The linear accelerator (Figure 5.16) is fed by an ECR ion source and consists of the LEBT, the RFQ-section, the MEBT and a room temperature CH-drift tube linac with high-beta cavities in the final section. The RFQ-section is divided into two parts. The subject of the investigation is the matching section between the two RFQs. The low energy area is the region with the highest proton density and therefore with the highest space charge forces acting on the beam. That is why a transport lines should be as short as possible to prevent excessive emittance growth.
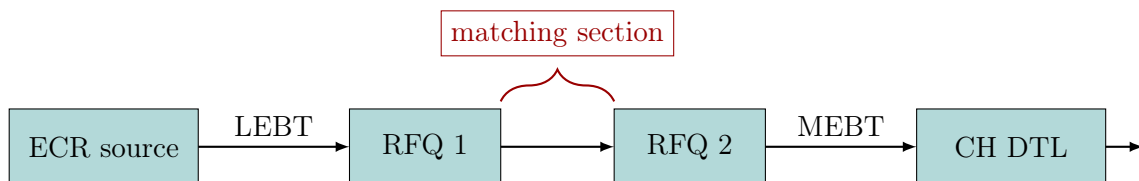

Figure 5.16.: Schematic view of the HBS proton linear accelerator. The matching section to be constructed is located between the two RFQs and is used to manipulate the beam so that a perfect matching with 100% transmission can be achieved.

The difficulty here is that all the necessary elements must be placed within the shortest possible distance to meet the acceptance of the second RFQ. Within less than 2 m, quadrupoles for transverse focusing, a re-buncher cavity for phase focusing, a steerer pair for position correction and diagnostic elements have to be accommodated.

Variable triplet position

The input distribution of the particles (Figure 5.17) is a direct result from the design of the first RFQ and was provided for the simulation. The acceptance of the second RFQ is also
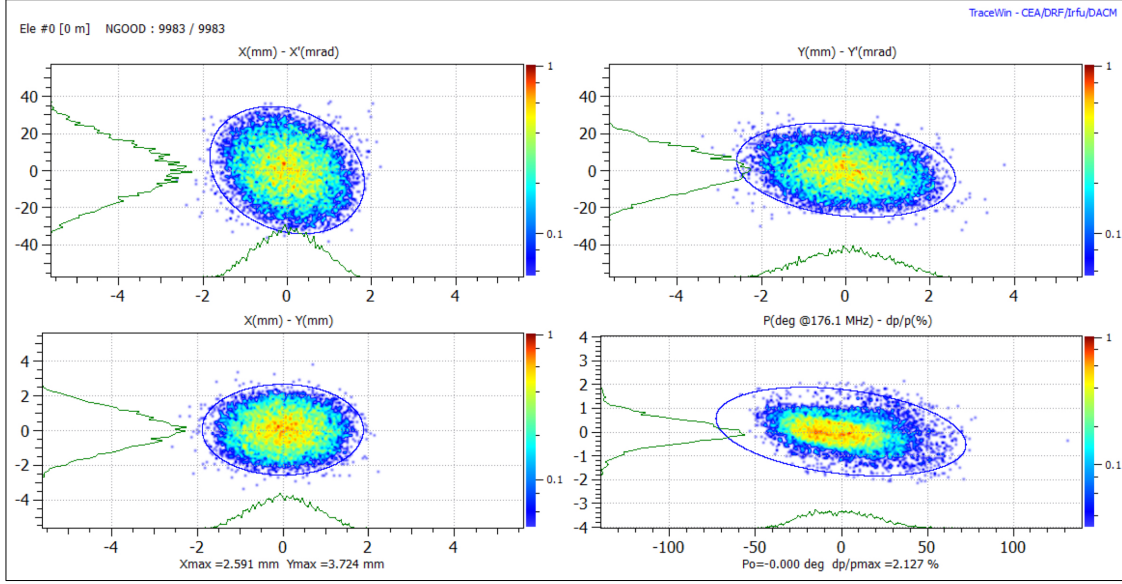

Figure 5.17.: Input particle distribution, Tracewin visualisation.

given by its design and was first estimated using the output parameters of the first RFQ (Table 5.8). For all particle tracking simulations, a bunched beam at a base frequency of 176.1 MHz and 3D linear space charge force was used. The beam current is 105 mA.

Table 5.8.: Acceptance of the second RFQ assumed and used for optimisation. Listed are the limits of a non-rotated six-dimensional phase space ellipsoid.

direction	acceptance	direction	acceptance	direction	acceptance
$|x|_{max}$	2 mm	$|y|_{max}$	3 mm	$|z|_{max}$	18 mm
$|x'|_{max}$	40 mrad	$|y'|_{max}$	30 mrad	$|z'|_{max}$	50 mrad

The total length of the matching section is first set to 1 m. In order to perform phase focusing, the section must contain a re-buncher cavity. Since its drift tubes are generally aperture limiting, a transverse focusing element must be installed before the re-buncher. A quadrupole triplet shall be used for this purpose. A second triplet is then required for transverse focusing into the second RFQ. The aperture of the section is circular and has a diameter of 48 mm. The diameter of the drift tubes is planned to be 22 mm. For optimisation, it was set to 16 mm to avoid particle trajectories too close to the tube wall and to achieve better focusing into the buncher. The initial situation for the first optimisation can be seen in Figure 5.18.

The re-buncher contains three accelerating gaps whose gap voltages have a fixed ratio of (0.261 : 0.478 : 0.261) to each other. Therefore, it is not necessary to optimise the voltage of each gap individually. It is sufficient, to use the total voltage, which has the advantage that one can reduce the parameter space by two dimensions. Now the optimal
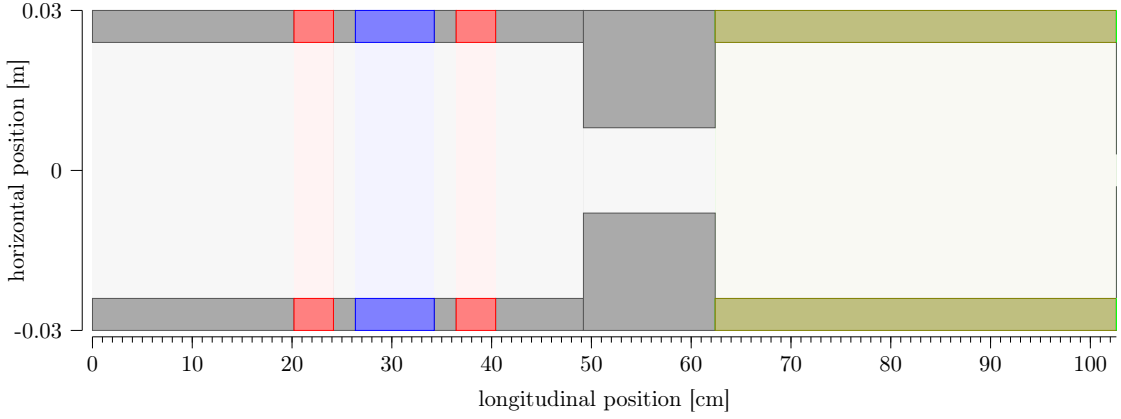
Figure 5.18.: Initial situation, top view. The first triplet is located at position $s = 20\,\text{cm}$ and followed by the re-buncher cavity. A second identical triplet is to be positioned within the yellow area.

quadrupole strengths, the buncher voltage and the position of the second triplet are to be found. In total, this problem has eight degrees of freedom and so the parameter space has eight dimensions. These are the normalised gradients of the quadrupoles of the first triplet $k_{t11}, k_{t12}, k_{t13}$, the normalised gradients of the second triplet $k_{t21}, k_{t22}, k_{t23}$ and its distance s_d to the centre of the last gap within the buncher, as well as the buncher voltage U_b. Each solution corresponds to one point $\boldsymbol{x}$ in the normalised 8-cube.

In the first optimisation, the bounds on the field gradients were set large enough, so that the optimisation was not a priori limited. The best result found with GLC was a transmission of 93%. The corresponding quadrupole strengths and triplet position are given in Table 5.9.

Table 5.9.: Parameter values of the solution found by GLC with 93% transmission.

parameter	value	parameter	value	parameter	value
k_{t11}	$-311.2/\text{m}^2$	k_{t21}	$297.1/\text{m}^2$	U_b	$118\,\text{kV}$
k_{t12}	$200.6/\text{m}^2$	k_{t22}	$-262.7/\text{m}^2$	Δs_d	$15.9\,\text{cm}$
k_{t13}	$-222.2/\text{m}^2$	k_{t23}	$407.5/\text{m}^2$		

However, it turns out that the strength values are too high. For a realistic application, the magnetic field should be smaller than $B_{\text{max}} = 1\,\text{T}$ at the pole iron. The radius of the quadrupole chamber is $r = 0.024\,\text{m}$. So the maximum field gradient is

$$g_{\text{max}} = \frac{B_{\text{max}}}{r} = 41.67\,\frac{\text{T}}{\text{m}}\,. \tag{5.21}$$

The magnetic rigidity of a proton beam with $1.27\,\text{MeV}/c$ is $(B\rho) = 0.163\,\text{T\,m}$ and so the upper limit for the value of the normalised integrated field gradient is

$$k_{\text{max}} = \frac{g_{\text{max}}}{(B\rho)} = 255\,\frac{1}{\text{m}^2}\,. \tag{5.22}$$

Variable triplet position and limited field gradient

It is possible to achieve the same integral field with a limited field gradient by increasing the quadrupole length. However, a second optimisation with identical geometry, but correspondingly reduced limits was performed first. The resulting parameter values are given in Table 5.10. The optics were changed and the second triplet was positioned 5.7 cm closer

Table 5.10.: Parameter values of the solution found by GLC with 88% transmission, using limited normalised quadrupole gradients ($k \leq 255\,\mathrm{m}^{-2}$).

parameter	value	parameter	value	parameter	value
$k_{\mathrm{t}11}$	$247.8\,\mathrm{m}^{-2}$	$k_{\mathrm{t}21}$	$253.6\,\mathrm{m}^{-2}$	U_{b}	$106\,\mathrm{kV}$
$k_{\mathrm{t}12}$	$-185.1\,\mathrm{m}^{-2}$	$k_{\mathrm{t}22}$	$-198.9\,\mathrm{m}^{-2}$	Δs_{d}	$10.2\,\mathrm{cm}$
$k_{\mathrm{t}13}$	$172.5\,\mathrm{m}^{-2}$	$k_{\mathrm{t}23}$	$252.3\,\mathrm{m}^{-2}$		

to the buncher cavity. Even though the lower field strength could be partially compensated by a different, now symmetrical optics, the maximum achievable transmission of 88% is still significantly lower. Furthermore, three of the quadrupoles have to be operated close to their upper limit for this setting, which is not very desirable for stability and long term operational safety.

Realistic element dimensions and distances

But if one looks at the quadrupole matrix (2.54), it can be seen that the position of the particle after passing through the quadrupole depends on the cosine of $\chi = \sqrt{k}L$. Thus, the influence of the length on the effect of the quadrupole is even stronger than that of the gradient. The value of

$$\chi = \sqrt{400\,\mathrm{m}^{-2}} \cdot 4\,\mathrm{cm} = 0.8\,\mathrm{m}^{-1}\,, \tag{5.23}$$

reached in the initial example can therefore also be achieved by a quadrupole of length

$$L = \frac{\chi}{\sqrt{k_{\mathrm{max}}}} = 5\,\mathrm{cm}\,. \tag{5.24}$$

Ultimately an optimisation has been carried out in which the quadrupole lengths have been adjusted accordingly. In addition to this, reasonably technically feasible distances are also to be used. The minimum distance between gap centre and quadrupole was defined as 20 cm, inspired from the experience of other projects, e.g. the proton accelerator for MYRRHA [73]. The length of the final drift path is introduced as an additional degree of freedom with an upper limit of 1.50 m integral length. So the optimal length is to be found as well. In total, the optimisation has now 9 degrees of freedom.

The result is again a solution with 93 % transmission. Figure 5.19 shows the resulting lattice. The second triplet has been moved further back and is now 24 cm away from the gap centre. The resulting total length is 1.39 m and the quadrupole strengths (Table 5.11) are all at least 10 % below their limits. It would therefore actually be possible to make them somewhat shorter again, especially since the field component outside the quadrupoles increases the effective quadrupole length anyway. Figure 5.20 shows the output particle distribution and Figure 5.21 the particle traces of the corresponding particle tracking simulation. While transverse focusing is possible without any problems, there are
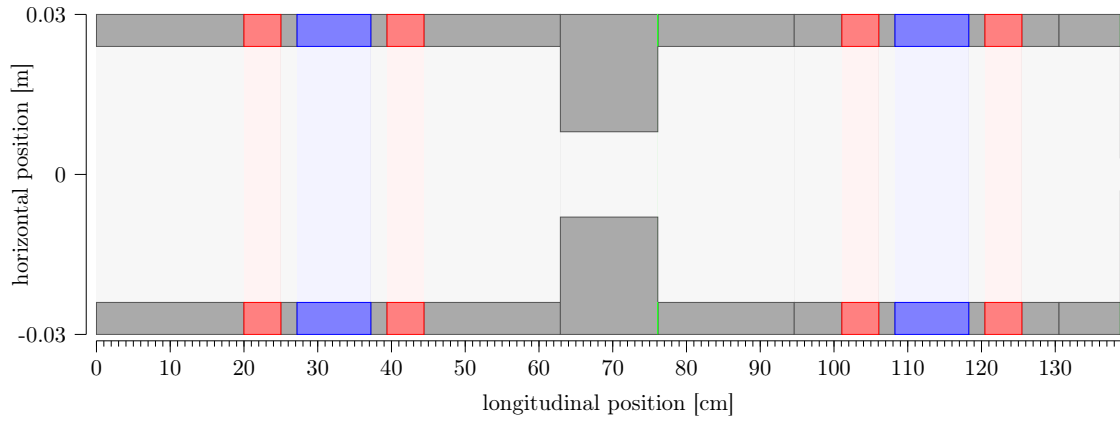
Figure 5.19.: Optimised lattice for the realistic scenario.

Table 5.11.: Parameter values of the solution for the realistic case, found by GLC. The transmission is 93% and the optimal integral length of the matching section is found to be 1.39 m.

parameter	value	parameter	value	parameter	value
k_{t11}	$-205.7/\mathrm{m}^2$	k_{t21}	$-152.4/\mathrm{m}^2$	U_b	$92\,\mathrm{kV}$
k_{t12}	$136.0/\mathrm{m}^2$	k_{t22}	$146.7/\mathrm{m}^2$	L_total	$1.39\,\mathrm{m}$
k_{t13}	$-140.4/\mathrm{m}^2$	k_{t23}	$-222.3/\mathrm{m}^2$	s_d	$24\,\mathrm{cm}$
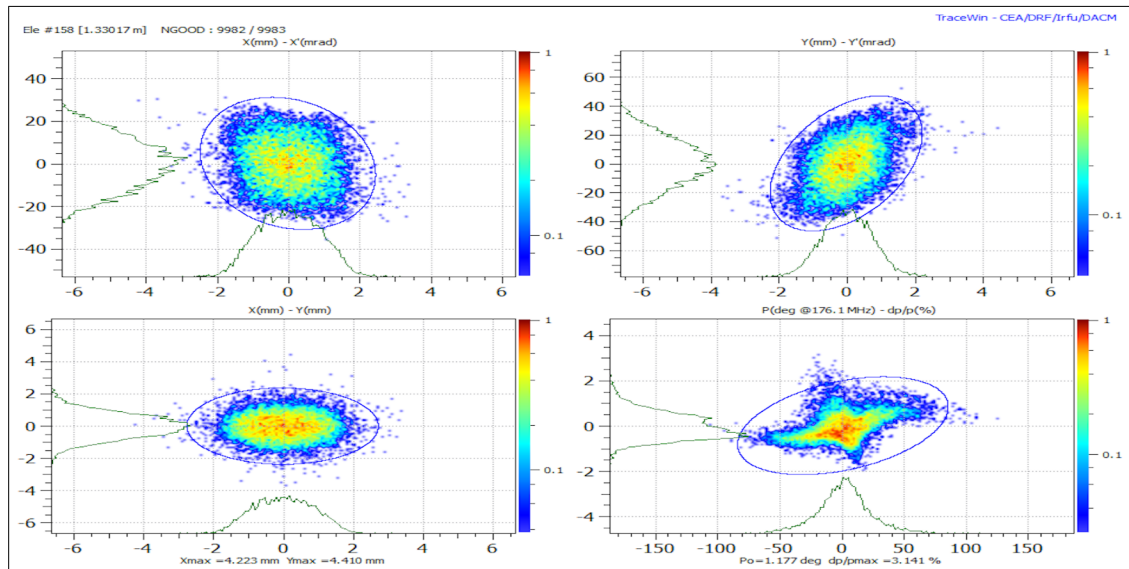
Figure 5.20.: Output particle distribution for the realistic scenario.

major difficulties in focusing all the particles longitudinally. Here only the core is well focused. This is because the position of the buncher was fixed and it cannot be moved further forward due to the given space restrictions. Given the input distribution, some particles already have a phase deviation of more than 90 degrees at the position of the buncher. Due to the asymmetry of the input distribution, particles with a positive phase
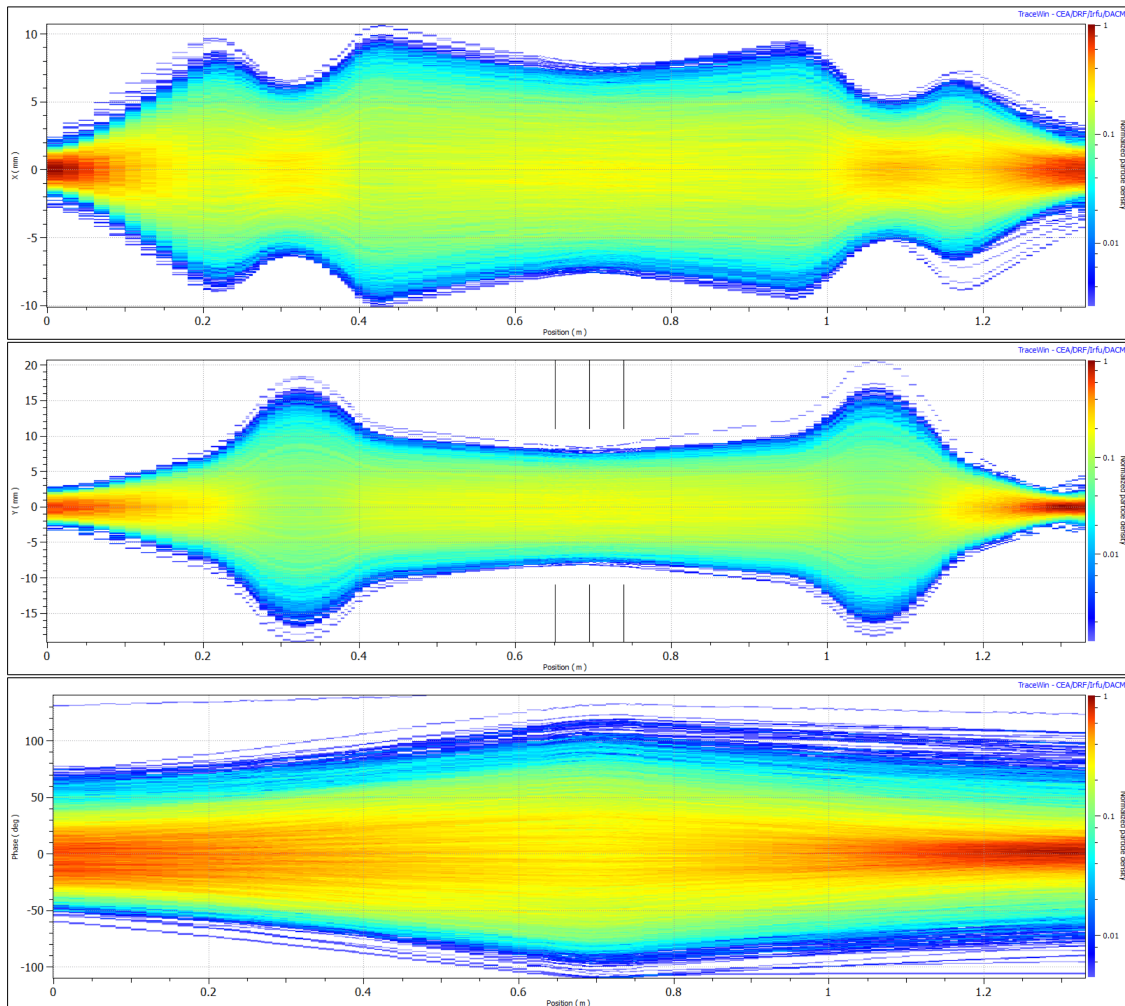
Figure 5.21.: Result of the particle tracking simulation for the realistic scenario, converted to Tracewin [74] (top: horizontal motion, centre: vertical motion, bottom: longitudinal motion (phase)).

exceed this limit first. This means that these particles can no longer be focused and are lost.

It would therefore be imperative to save space at the front of the section, to be able to shift the buncher further to the front. One option is to install the complete diagnostics behind the buncher and to minimise the space between triplet and buncher. Smaller distances between the quadrupoles are also desirable. Every spared centimetre can make a difference, since longitudinal phase space filamentation only starts in the second gap. If possible, one should try to design the RFQ 1 in such a way that the phase focus is shifted backwards. Another option is to operate RFQ 1 with half the frequency, in which case the phase would also be only half the size. The RFQ 2 would then have to run at twice the frequency and only every second RF pulse would be filled with beam, which is probably not desired. Another alternative variant would be to install the quadrupoles inside the drift tubes e.g., to build an Alvarez-type buncher. In this case a drift tube spacing of $\beta\lambda$ would have to be used instead of $\frac{\beta\lambda}{2}$. But it would be possible to focus in all three dimensions simultaneously over a short distance. But this solution has other disadvantages. Experience shows that small internal quadrupole lenses are difficult to build, to align and

to maintain in the long term. Frequent tank openings can become necessary, especially at older accelerators, where drift tubes with water leaks have to be shut down and replaced regularly [75].

5.3.3. Conclusion

While these are only preliminary studies in an early design phase and the results are therefore likely to be subject to changes until the final design, it has been clearly demonstrated that GLC provides persuading results and can support the design process as it progresses. If, for example a new beam distribution would follow from the design of the preceding RFQ1, or if the acceptance of the second RFQ would have to be re-defined, then it is sufficient to perform the latest optimisation again with the changed boundary conditions. The same is true in case the general geometry would have to be changed due to new findings or construction restrictions.

It should be mentioned, that the results of the previous section were also successful in particular because a lot of limiting constraints were known and integrated from the beginning. For example, the maximum length, spacing, number and basic design of the elements were predefined and in the last examples there were additional field limitations. As already mentioned in the chapter about the genetic algorithms, the more known information is given as boundary conditions for the optimisation, the better results can be achieved. The background is that each piece of information reduces the size of the parameter space and may additionally simplify its structure or even reduce the number of dimensions. It is therefore not surprising that it is generally easier to find an optimum under these conditions.

6. The Accelerator Construction Set

As part of this work, another particle tracking application (the **A**ccelerator **C**onstruction **S**et) was developed, mainly to provide a flexible and fast interface for testing the optimisation algorithms with many particles while having access to the element handling on source code level. The tracking core is an s-Code, based on the matrix formalism, presented in chapter 1 and contains the reference implementations for both, the General Autotune and the Lattice Construction algorithm. The application is written in C++ [76] and was mainly used as a command line tool on a Unix environment. For better visualisation it is published with a QT5 user interface (Figure 6.1) to provide a low entry barrier. The
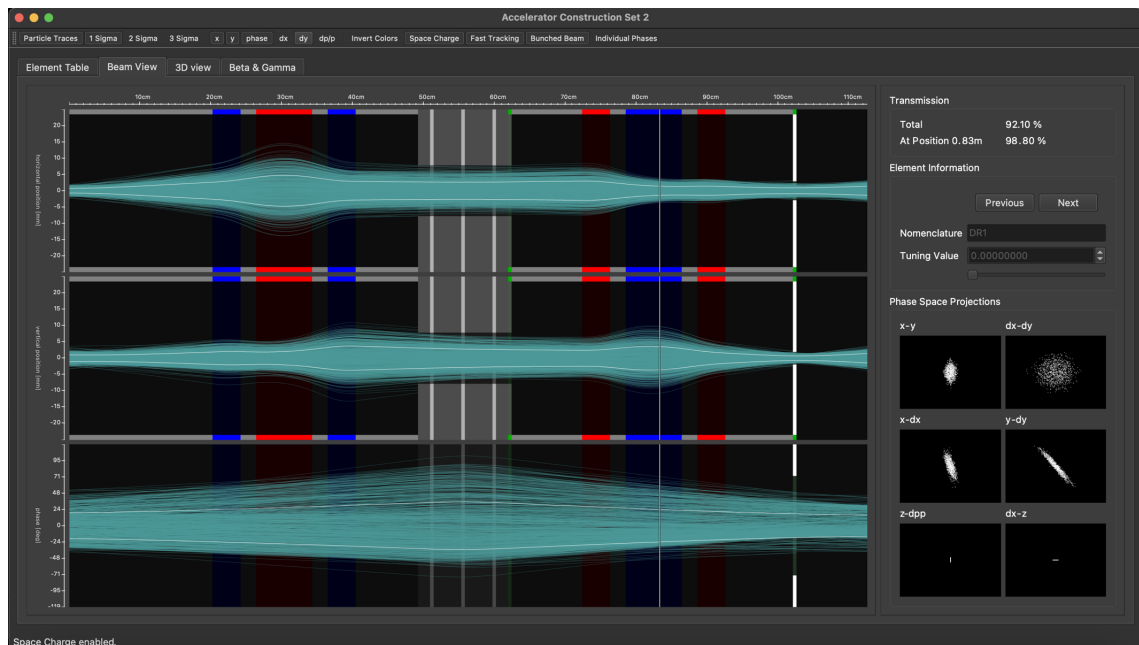
Figure 6.1.: Standard beam view of Accelerator Construction Set 2. There are 3 views of the development of the beam (top: horizontal, centre: vertical, bottom: longitudinal). The respective momenta can also be switched visible. On the bottom right, selected 2 dimensional projections of the phase space are located.

user interface makes it easy to compose beam lines from a set of pre-defined elements. All elements have specific properties, according to their type, one of which can be selected for optimisation.

All properties can be edited directly in the element table. The program is fully interactive. Setting changes are applied directly and the effects on the beam are calculated and updated instantaneously. Different beam distributions can be generated, loaded and stored.

To perform an autotune optimisation, the corresponding lattice has to be constructed or loaded. There must be at least one beam transformer in the path to evaluate the transmission, which is usually attached at the end. For the lattice construction, all areas

in which the elements can be positioned must additionally be defined. This is solved by 'Free Space' elements of the desired length. After that, the elements to be placed have to be defined. For both Autotune and GLC, the underlying algorithm can then be selected and its respective hyper parameters can be set. However, the default parameters are a good starting point. At the time of submitting the work, the following algorithms are available

- gradient descend
- genetic algorithm
- simulated annealing
- particle swarm
- BOBYQA

and the following accelerator elements are implemented

- field free drift sections
- quadrupole magnets
- solenoids
- Gabor-lenses
- thin lenses
- dipoles (sector type and rectangular)
- local angular kicks (steerer)
- accelerating gaps
- buncher cavities
- slit elements
- profile grids
- beam transformer
- 6D acceptance limitation elements

The treatments of all ACS elements used in this work were successfully tested for equivalence with Tracewin results.

7. Conclusion and outlook

7.1. Autotune

Within the scope of this work at GSI and in the meantime also at various accelerator facilities (for example at CERN and SLAC), it could be shown that accelerator settings can be automatically optimised online very well under certain boundary conditions. In particular, this applies to initial commissioning with low intensity pilot beams, for systems that provide a high repetition rate or for continuous wave linear accelerators. Optimisation of the injection parameters for fast cycling synchrotrons is also conceivable. For cycle times of less than a second, one could also optimise an entire synchrotron cycle. However, it is important in this case that there is already a model in which physical parameters such as radial beam position, septum- and kicker angles, local bumps or the synchrotron tune can be manipulated directly.

Since it is only possible to optimise for observables that can also be measured with the necessary precision, it is essential for automatic online optimisation that the section to be optimised has suitable beam instrumentation installed. Since most optimisation problems can be reduced to transmission optimisation, beam transformers and Faraday cups play a special role in this context. High-precision, stable measurement of beam current over several orders of magnitude is essential. Regarding beam diagnostics and the settings management system, a good integration into the control system and well defined software interfaces are fundamental. Automatic tuning works best under stable beam conditions. So a stable source current, precise power supplies or RF-transmitters can help, to increase its performance.

Automatic tuning could be further developed towards a theory assisted optimisation, in which an offline accelerator simulation evaluates settings in parallel and excludes settings with total beam losses from the online evaluation. This could avoid the incidence of unwanted activation and thus expand the application field and the performance of automatic optimisation in the future.

Although the successes are compelling, machine learning remains a double-edged sword. The understanding of the system does not necessarily improve and if not explicitly excluded, an automatic algorithm also sometimes find solutions that do not make much sense. In the end, an understanding of the matter is essential. An automatic optimisation can give hints, lead out of dead ends in special cases or even replace tedious manual tuning. However, a well understood and stable model of the machine remains to be the gold standard as this allows direct optimisation using optical online models or beam based feedback systems.

But the use of the suggested theory assisted optimisation would at least open the possibility to automatically adapt the model more and more to the real measured data and thus have to check less and less settings on the real machine. So, as a side effect, a more accurate understanding of the machine could still be obtained.

7.1.1. Considerations of utility for state-of-the-art accelerator designs

From the equations (2.109) and (2.110) follows a dependence of the focusing terms on the particle velocity in accelerating gaps

$$\frac{1}{f} \propto \frac{E_0}{\beta^3} \, . \tag{7.1}$$

Especially for small β at the beginning of the acceleration, this results in a rather high value and thus in more emittance growth. For $\beta \to 1$, the effect is increasingly less important. It is therefore advantageous if the acceleration gradient is less strong in the first gaps and then successively increases. This principle can only be achieved with a construction of shorter tank structures, where the voltage in the first structures can be set differently (lower) than in the following ones. In long monolithic structures the use of this principle is not possible. As a result, the RF phases between the different tanks must be very finely tuned within a narrow parameter range. This is especially important because the acceleration, for which the RF phase is usually set to values of $\phi_s \approx -30°$, always results in phase asymmetry in the bunch due to the asymmetry of the sine around this value. However, the necessary accuracy cannot be calculated practically, since the necessary design tolerances are impossible to realise. The use of GAA provides the ideal solution for operating such devices, with tuning these phases automatically online. Autotune therefore theoretically enables the construction and efficient use of state-of-the-art accelerators with previously unattainable beam qualities.

7.2. Genetic Lattice Construction

Very good results were also obtained for the automatic design of beam transport lines. The GLC algorithm is applicable to a large class of design problems and, together with precise particle tracking tools, GLC provides an excellent tool for the ion-optical design of future linear accelerators or for the optimisation of beam transport lines of any kind. The investigations concerning the incorporation of position errors during phenotype construction suggest that GLC finds more robust solutions with respect to error studies. Further research should be conducted on this topic, as in the best case the possibility exists to make additional error studies obsolete.

Both the dimensionality of the parameter space and its structure have an influence on the performance of GLC. A high number of quadrupoles leads to many local minima and lowers the reliability of the algorithm significantly. To some extent, this can be compensated with more optimisation cycles. Therefore, it should always be tried to combine parameters and to reduce the number of degrees of freedom.

From the set of optimisation algorithms presented, the genetic algorithm proved to be particularly reliable in both areas of application, but the topic of machine learning is currently very dynamic and it can be expected that new algorithms and methods will bring further improvements. Since the GLC method is generally independent of the underlying optimisation algorithm, the algorithm can easily be replaced to directly benefit from future developments.

Bibliography

[1]   Stephan Reimann, Martin Droba, Oliver Meusel and Holger Podlech. 'An algorithm for automated lattice design of transfer lines'. In: *Journal of Physics: Conference Series* 1350 (Nov. 2019), p. 012110. DOI: 10.1088/1742-6596/1350/1/012110.

[2]   K. L. Brown. 'First- and second-order matrix theory for the design of beam transport systems and charged particle spectrometers.' In: *Adv. Part. Phys.* 1 (Jan. 1971), pp. 71–134. DOI: 10.2172/4742148.

[3]   Frank Hinterberger. *Physik der Teilchenbeschleuniger und Ionenoptik*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. ISBN: 978-3-540-75282-0. DOI: 10.1007/978-3-540-75282-0.

[4]   L. C. Teng. 'Expanded form of magnetic field with median plane'. In: *ANL-LCT-28* (Dec. 1962). DOI: 10.2172/7364341.

[5]   Klaus Wille. *The physics of particle accelerators : an introduction*. Oxford, New York: Oxford University Press, 2000. ISBN: 978-0198505495.

[6]   Martin Droba. *Teilchendynamik in gefalteten Schwerionenlinearbeschleunigern*. Aachen: Shaker, 2002. ISBN: 3-8322-0188-2.

[7]   Karl L. Brown, David C Carey, F Christoph Iselin and F Rothacker. *TRANSPORT: a computer program for designing charged-particle beam transport systems; 2nd ed.* CERN Yellow Reports: Monographs. Also publ. as SLAC and FERMILAB. Geneva: CERN, 1973. DOI: 10.5170/CERN-1973-016.

[8]   B. Franczak. 'MIRKO — an interactive program for beam lines and synchrotrons'. In: *Computing in Accelerator Design and Operation*. Ed. by Winfried Busse and Roman Zelazny. Berlin, Heidelberg: Springer Berlin Heidelberg, 1984, pp. 170–175. ISBN: 978-3-540-39130-2. DOI: 10.1007/3540139095_102.

[9]   H. Grote and F. Schmidt. 'Mad-X - an upgrade from mad8'. In: *Proceedings of the 2003 Particle Accelerator Conference*. Vol. 5. IEEE, 2003, pp. 3497–3499. DOI: 10.1109/pac.2003.1289960.

[10]   Didier Uriot and N. Pichoff. *TraceWin*. Paris, France: CEA Saclay, May 2020. eprint: http://irfu.cea.fr/dacm/logiciels/codesdacm/tracewin/tracewin.pdf.

[11]   P. J. Bryant. 'AGILE, A Tool for Interactive Lattice Design'. In: *Proc. 7th European Particle Accelerator Conf. (EPAC'00)* (Vienna, Austria). July 2000, pp. 1357–1359. eprint: https://accelconf.web.cern.ch/e00/PAPERS/TUP6B03.pdf.

[12]   David Ondreka, Ralph Bär and Andreas Redelbach. *Steuerdatenversorgung für den Heidelberger Therapiebeschleuniger HICAT*. Darmstadt, Germany: GSI Helmholtzzentrum für Schwerionenforschung GmbH, 2008.

[13]   E. D. Courant and H. S. Snyder. 'Theory of the Alternating-Gradient Synchrotron'. In: *Annals of Physics* 281.1 (2000), pp. 360–408. ISSN: 0003-4916. DOI: 10.1006/aphy.2000.6012.

[14] Ernest D. Courant, M. Stanley Livingston and Hartland S. Snyder. 'The Strong-Focusing Synchroton—A New High Energy Accelerator'. In: *Phys. Rev.* 88 (5 Dec. 1952), pp. 1190–1196. DOI: 10.1103/PhysRev.88.1190.

[15] David Ondreka, Peteris Apse-Apsitis, Kathrin Schulte and Peter Spiller. 'Overcoming the Space Charge Limit: Development of an Electron Lens for SIS18'. In: *8th International Particle Accelerator Conference*. May 2017. DOI: 10.18429/JACoW-IPAC2017-TUPVA059.

[16] D. Gabor. 'A Space-Charge Lens for the Focusing of Ion Beams'. In: *Nature* 160.4055 (July 1947), pp. 89–90. DOI: 10.1038/160089b0.

[17] O. Meusel et al. 'Low-energy beam transport using space-charge lenses'. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 544.1 (2005). Proceedings of the 15th International Symposium on Heavy Ion Inertial Fusion, pp. 447–453. ISSN: 0168-9002. DOI: 10.1016/j.nima.2005.01.276.

[18] Oliver Meusel. 'Fokussierung und Transport von Ionenstrahlen mit Raumladungslinsen'. PhD thesis. Frankfurt am Main, Germany: Goethe-University, 2006.

[19] M. Reiser. 'Comparison of Gabor lens, gas focusing, and electrostatic quadrupole focusing for low-energy ion beams'. In: *Proceedings of the 1989 IEEE Particle Accelerator Conference, 'Accelerator Science and Technology'*. 1989, 1744–1747 vol.3. DOI: 0.1109/PAC.1989.72912.

[20] A. Bouwers. *Elektrische Höchstspannungen*. Berlin: J. Springer, 1939. ISBN: 978-3-642-90748-7.

[21] J. De Mascureau. 'Induction linacs'. In: (1996). DOI: 10.5170/CERN-1996-002.79.

[22] Manuel Heilmann. personal communication. 15th Apr. 2021.

[23] P. M. Lapostolle. *Proton linear accelerators: A theoretical and historical introduction*. Los Alamos, United States, July 1989. DOI: 10.2172/6038195.

[24] Thomas P. Wangler. *RF linear accelerators*. Weinheim, Germany: Wiley-VCH, Mar. 2008. ISBN: 978-3-527-40680-7.

[25] W. Panofsky. *Beam dynamics of the linear accelerator*. 1951. eprint: https://escholarship.org/uc/item/8p16g51t.

[26] Charles Birdsall. *Plasma physics via computer simulation*. New York: Taylor & Francis, 2005. ISBN: 0-7503-1025-1.

[27] M. Reiser. *Theory and design of charged particle beams*. Weinheim: Wiley-VCH, 2008. ISBN: 978-3-527-40741-5.

[28] I. M. Kapchinskij and V. V. Vladimirskij. 'Limitations of proton beam current in a strong focusing linear accelerator associated with the beam space charge'. In: *2nd International Conference on High-Energy Accelerators*. 1959, pp. 274–287. eprint: https://inspirehep.net/files/526fe2f2ff6f05063bcb9a9b08da12f9.

[29] F. J. Sacherer and T. R. Sherwood. 'The Effect of Space Charge in Beam Transport Lines'. In: *IEEE Transactions on Nuclear Science* 18.3 (1971), pp. 1066–1067. DOI: 10.1109/TNS.1971.4326279.

[30]   P. M. Lapostolle. 'Possible Emittance Increase through Filamentation Due to Space Charge in Continuous Beams'. In: *IEEE Transactions on Nuclear Science* 18.3 (1971), pp. 1101–1104. DOI: `10.1109/TNS.1971.4326292`.

[31]   John Holland. *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence.* Cambridge, Mass: MIT Press, 1992. ISBN: 978-0-262-58111-0.

[32]   David Goldberg. *Genetic algorithms in search, optimization, and machine learning.* Reading, Mass: Addison-Wesley Publishing Company, 1989. ISBN: 0-201-15767-5.

[33]   R. Eberhart and J. Kennedy. 'A new optimizer using particle swarm theory'. In: *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science.* 1995, pp. 39–43. DOI: `10.1109/MHS.1995.494215`.

[34]   Nicholas Metropolis et al. 'Equation of State Calculations by Fast Computing Machines'. In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. DOI: `10.1063/1.1699114`.

[35]   Martin Pincus. 'Letter to the Editor—A Monte Carlo Method for the Approximate Solution of Certain Types of Constrained Optimization Problems'. In: *Operations Research* 18.6 (1970), pp. 1225–1228. DOI: `10.1287/opre.18.6.1225`.

[36]   S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi. 'Optimization by Simulated Annealing'. In: *Science* 220.4598 (1983), pp. 671–680. ISSN: 0036-8075. DOI: `10.1126/science.220.4598.671`.

[37]   A.B. Finnila et al. 'Quantum annealing: A new method for minimizing multidimensional functions'. In: *Chemical Physics Letters* 219.5 (1994), pp. 343–348. ISSN: 0009-2614. DOI: `10.1016/0009-2614(94)00117-0`.

[38]   Christopher De Sa. *Principles of Large-Scale Machine Learning Systems.* 2019. URL: `https://www.cs.cornell.edu/courses/cs4787/2019sp/`.

[39]   Guido Walz. *Lexikon der Mathematik.* Heidelberg: Spektrum, Akad. Verl, 2003. ISBN: 3-8274-0439-8.

[40]   Tinu Theckel Joy, Santu Rana, Sunil Gupta and Svetha Venkatesh. 'A Flexible Transfer Learning Framework for Bayesian optimization with Convergence Guarantee'. In: *Expert Systems with Applications* 115 (Aug. 2018). DOI: `10.1016/j.eswa.2018.08.023`.

[41]   M. Powell. 'The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives'. In: *Technical Report, Department of Applied Mathematics and Theoretical Physics* (June 2009). eprint: `http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf`.

[42]   M.J.D. Powell. 'The NEWUOA software for unconstrained optimization without derivatives'. In: *Di Pillo G., Roma M. (eds) Large-Scale Optimization. Nonconvex Optimization and Its Applications* 83 (2006), pp. 255–297. DOI: `10.1007/0-387-30065-1_16`. eprint: `http://bipop.inrialpes.fr/people/guilbert/newuoa/newuoa.pdf`.

[43]   S. Appel and S. Reimann. 'Beam Line Optimization Using Derivative-Free Algorithms'. In: *Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019* (Melbourne, Australia). International Particle Accelerator Conference 10. Geneva, Switzerland: JACoW Publishing, June 2019, pp. 2307–2310. ISBN: 978-3-95450-208-0. DOI: `10.18429/JACoW-IPAC2019-WEPMP005`.

[44] Marco Dorigo and Luca Maria Gambardella. 'Ant colonies for the travelling salesman problem'. In: *Biosystems* 43.2 (1997), pp. 73–81. ISSN: 0303-2647. DOI: `10.1016/S0303-2647(97)01708-5`.

[45] W. Vent. 'Rechenberg, Ingo, Evolutionsstrategie — Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. 170 S. mit 36 Abb. Frommann-Holzboog-Verlag. Stuttgart 1973. Broschiert'. In: *Feddes Repertorium* 86.5 (1975), pp. 337–337. DOI: `10.1002/fedr.19750860506`.

[46] G. Cybenko. 'Approximation by superpositions of a sigmoidal function'. In: *Mathematics of Control, Signals, and Systems* (1989), pp. 303–314. DOI: `10.1007/bf02551274`.

[47] David Kriesel. *A Brief Introduction to Neural Networks.* 2007. URL: `https://www.dkriesel.com/science/neural_networks`.

[48] Daniel Rivero, Julian Dorado, Enrique Fernández-Blanco and Alejandro Pazos. 'A Genetic Algorithm for ANN Design, Training and Simplification'. In: *Bio-Inspired Systems: Computational and Ambient Intelligence.* Ed. by Joan Cabestany, Francisco Sandoval, Alberto Prieto and Juan M. Corchado. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 391–398. ISBN: 978-3-642-02478-8.

[49] Marco Cavazzuti. *Optimization methods : from theory to design scientific and technological aspects in mechanics.* Berlin, New York: Springer, 2013. ISBN: 978-3-642-31187-1.

[50] D. H. Wolpert and W. G. Macready. 'No free lunch theorems for optimization'. In: *IEEE Transactions on Evolutionary Computation* 1.1 (Apr. 1997), pp. 67–82. DOI: `10.1109/4235.585893`.

[51] W. Geithner et al. 'Status and outlook of the CRYRING@ESR project'. In: *Hyperfine Interactions* 238.1 (Jan. 2017). DOI: `10.1007/s10751-016-1383-5`.

[52] Geithner, Wolfgang. *Cryring Wiki – YRT1 Injector Details.* (Visited on 23/04/2021).

[53] Petra Schuett, Stephan Reimann and Markus Vossberg. 'Planning and Tracing a 2-year Long Shutdown Period'. In: *ARW 2017 Book of Abstracts* (Versailles, France). Accelerator Reliability Workshop. Oct. 2017, p. 9.

[54] S. Appel et al. 'Automatized Optimization of Beam Lines Using Evolutionary Algorithms'. In: *Proc. of International Particle Accelerator Conference (IPAC'17), Copenhagen, Denmark, 14-19 May, 2017* (Copenhagen, Denmark). International Particle Accelerator Conference 8. Geneva, Switzerland: JACoW, May 2017, pp. 3941–3944. ISBN: 978-3-95450-182-3. DOI: `10.18429/JACoW-IPAC2017-THPAB096`.

[55] W. Geithner et al. 'Genetic Algorithms for Machine Optimization in the Fair Control System Environment'. In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018* (Vancouver, BC, Canada). International Particle Accelerator Conference 9. Geneva, Switzerland: JACoW Publishing, June 2018, pp. 4712–4715. ISBN: 978-3-95450-184-7. DOI: `10.18429/JACoW-IPAC2018-THPML028`.

[56] Stephan Reimann and Oksana Geithner. 'User Beam Time Report 2019'. In: *GSI-FAIR Scientific Report 2019* (2020), pp. 110–113. DOI: `10.15120/GSI-2020-00416`.

[57] P. J. Spiller et al. 'Status of the FAIR Project'. In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018* (Vancouver, BC, Canada). International Particle Accelerator Conference 9. Geneva, Switzerland: JACoW Publishing, June 2018, pp. 63–68. ISBN: 978-3-95450-184-7. DOI: `10.18429/JACoW-IPAC2018-MOZGBF2`.

[58] Youssef El-Hayek et al. 'Initial Beam Loss and Control of Dynamic Vacuum Effects in SIS18'. In: *Proc. 4th International Particle Accelerator Conference (IPAC'13)* (Shanghai, China). May 2013, pp. 300–302. ISBN: 978-3-95450-122-9.

[59] M. Bai et al. 'Challenges of FAIR Phase 0'. In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018* (Vancouver, BC, Canada). International Particle Accelerator Conference 9. Geneva, Switzerland: JACoW Publishing, June 2018, pp. 2947–2949. ISBN: 978-3-95450-184-7. DOI: `10.18429/JACoW-IPAC2018-THYGBF3`.

[60] D. Ondreka, J. Fitzek, H. Liebermann and R. Müller. 'Generic Settings Generation for FAIR: First Experience at SIS18'. In: *Proc. 6th International Particle Accelerator Conference (IPAC'15), Richmond, VA, USA, May 3-8, 2015* (Richmond, VA, USA). International Particle Accelerator Conference 6. Geneva, Switzerland: JACoW, June 2015, pp. 156–158. ISBN: 978-3-95450-168-7. DOI: `10.18429/JACoW-IPAC2015-MOPWA027`.

[61] S. Appel and O. Boine-Frankenheim. 'Optimization of Multi-turn Injection into a Heavy-Ion Synchrotron using Genetic Algorithms'. In: *Proc. 6th International Particle Accelerator Conference (IPAC'15), Richmond, VA, USA, May 3-8, 2015* (Richmond, VA, USA). International Particle Accelerator Conference 6. Geneva, Switzerland: JACoW, June 2015, pp. 3689–3692. ISBN: 978-3-95450-168-7. DOI: `10.18429/JACoW-IPAC2015-THPF007`.

[62] Johannes Kirschner et al. *Adaptive and Safe Bayesian Optimization in High Dimensions via One-Dimensional Subspaces*. 2019. arXiv: `1902.03229`.

[63] G. Agakichiev et al. 'The high-acceptance dielectron spectrometer HADES'. In: *The European Physical Journal A* 41.2 (July 2009), pp. 243–277. ISSN: 1434-601X. DOI: `10.1140/epja/i2009-10807-5`.

[64] M. Sapinski et al. 'Measurements of the GSI Transfer Beam Lines Ion Optics'. In: *Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019* (Melbourne, Australia). International Particle Accelerator Conference 10. Geneva, Switzerland: JACoW Publishing, June 2019, pp. 131–134. ISBN: 978-3-95450-208-0. DOI: `10.18429/JACoW-IPAC2019-MOPGW024`.

[65] M. Sapinski et al. 'Upgrade of GSI HADES Beamline in Preparation for High Intensity Runs'. In: *Proc. of International Particle Accelerator Conference (IPAC'17), Copenhagen, Denmark, 14-19 May, 2017* (Copenhagen, Denmark). International Particle Accelerator Conference 8. Geneva, Switzerland: JACoW, May 2017, pp. 2214–2216. ISBN: 978-3-95450-182-3. DOI: `10.18429/JACoW-IPAC2017-TUPVA060`.

[66] J. Pietraszko. 'Slow extraction, input from HADES at SIS18'. HIC4FAIR Workshop. Darmstadt, Germany, Feb. 2016.

[67]   A. Rubin et al. 'Status of the Beam Dynamics Design of the New Post-Stripper DTL for GSI-FAIR'. In: *Proc. of International Particle Accelerator Conference (IPAC'17), Copenhagen, Denmark, 14-19 May, 2017* (Copenhagen, Denmark). International Particle Accelerator Conference 8. Geneva, Switzerland: JACoW, May 2017, pp. 4414–4416. ISBN: 978-3-95450-182-3. DOI: `10.18429/JACoW-IPAC2017-THPVA003`.

[68]   N. Abgrall et al. 'NA61/SHINE facility at the CERN SPS: beams and detector system'. In: *Journal of Instrumentation* 9.06 (June 2014), P06005–P06005. DOI: `10.1088/1748-0221/9/06/p06005`.

[69]   CERN. *Schematic of North Area Beamlines.* URL: `https://na62.web.cern.ch/ExperimentalArea/schematic.png` (visited on 19/04/2021).

[70]   G Burton, M Hanney and P Strolin. *BEATCH: a Fortran programme for the particle optics of beam transfer channels.* Tech. rep. Geneva: CERN, May 1969. eprint: `http://cds.cern.ch/record/310542`.

[71]   CERN. *H4 filter mode.* URL: `http://sba.web.cern.ch/sba/BeamsAndAreas/H4/H4Fmode.pdf` (visited on 31/01/2021).

[72]   T. Brückel et al. *Conceptual Design Report Jülich High Brilliance Neutron Source (HBS).* Vol. 8. Jülich, Germany: FZ Jülich, 2020. ISBN: 978-3-95806-501-7.

[73]   H. Podlech et al. 'General Layout of the 17 MeV Injector for MYRRHA'. In: *Proc. 2nd Int. Particle Accelerator Conf. (IPAC'11)* (San Sebastian, Spain). JACoW Publishing, Sept. 2011, pp. 2574–2576. eprint: `http://accelconf.web.cern.ch/IPAC2011/papers/weps039.pdf`.

[74]   Martin Droba. personal communication. 23rd Apr. 2021.

[75]   Peter Gerhard. 'UNILAC Status report'. In: *GSI Scientific Report 2015 - GSI Report 2016-1* (2016). DOI: `10.15120/GR-2016-1`.

[76]   Bjarne Stroustrup. *The C++ programming language.* Reading, Mass: Addison-Wesley, 1997. ISBN: 0-201-88954-4.

My Publications

[1] Stephan Reimann, Martin Droba, Oliver Meusel and Holger Podlech. 'An algorithm for automated lattice design of transfer lines'. In: *Journal of Physics: Conference Series* 1350 (Nov. 2019), p. 012110. DOI: 10.1088/1742-6596/1350/1/012110.

[43] S. Appel and S. Reimann. 'Beam Line Optimization Using Derivative-Free Algorithms'. In: *Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019* (Melbourne, Australia). International Particle Accelerator Conference 10. Geneva, Switzerland: JACoW Publishing, June 2019, pp. 2307–2310. ISBN: 978-3-95450-208-0. DOI: 10.18429/JACoW-IPAC2019-WEPMP005.

[53] Petra Schuett, Stephan Reimann and Markus Vossberg. 'Planning and Tracing a 2-year Long Shutdown Period'. In: *ARW 2017 Book of Abstracts* (Versailles, France). Accelerator Reliability Workshop. Oct. 2017, p. 9.

[54] S. Appel et al. 'Automatized Optimization of Beam Lines Using Evolutionary Algorithms'. In: *Proc. of International Particle Accelerator Conference (IPAC'17), Copenhagen, Denmark, 14-19 May, 2017* (Copenhagen, Denmark). International Particle Accelerator Conference 8. Geneva, Switzerland: JACoW, May 2017, pp. 3941–3944. ISBN: 978-3-95450-182-3. DOI: 10.18429/JACoW-IPAC2017-THPAB096.

[55] W. Geithner et al. 'Genetic Algorithms for Machine Optimization in the Fair Control System Environment'. In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018* (Vancouver, BC, Canada). International Particle Accelerator Conference 9. Geneva, Switzerland: JACoW Publishing, June 2018, pp. 4712–4715. ISBN: 978-3-95450-184-7. DOI: 10.18429/JACoW-IPAC2018-THPML028.

[56] Stephan Reimann and Oksana Geithner. 'User Beam Time Report 2019'. In: *GSI-FAIR Scientific Report 2019* (2020), pp. 110–113. DOI: 10.15120/GSI-2020-00416.

[59] M. Bai et al. 'Challenges of FAIR Phase 0'. In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018* (Vancouver, BC, Canada). International Particle Accelerator Conference 9. Geneva, Switzerland: JACoW Publishing, June 2018, pp. 2947–2949. ISBN: 978-3-95450-184-7. DOI: 10.18429/JACoW-IPAC2018-THYGBF3.

[64] M. Sapinski et al. 'Measurements of the GSI Transfer Beam Lines Ion Optics'. In: *Proc. 10th International Particle Accelerator Conference (IPAC'19), Melbourne, Australia, 19-24 May 2019* (Melbourne, Australia). International Particle Accelerator Conference 10. Geneva, Switzerland: JACoW Publishing, June 2019, pp. 131–134. ISBN: 978-3-95450-208-0. DOI: 10.18429/JACoW-IPAC2019-MOPGW024.

[65] M. Sapinski et al. 'Upgrade of GSI HADES Beamline in Preparation for High Intensity Runs'. In: *Proc. of International Particle Accelerator Conference (IPAC'17), Copenhagen, Denmark, 14-19 May, 2017* (Copenhagen, Denmark). International Particle Accelerator Conference 8. Geneva, Switzerland: JACoW, May 2017, pp. 2214–2216. ISBN: 978-3-95450-182-3. DOI: 10.18429/JACoW-IPAC2017-TUPVA060.

List of Figures

List of Tables

A. Appendix

A.1. Taylor-series expansion of the magnetic field in the magnetic mid-plane

The Taylor-series expansion of the magnetic field $\boldsymbol{B} = (B_x, B_y, B_s)$ is shown, following the approach of L. C. Teng [4]. Using Maxwell's equations for the vacuum and the fact that there is no time-varying electric field, Ampère's circuital law is reduced to

$$\nabla \times \boldsymbol{B} = 0\,. \tag{A.1}$$

The rotation of a scalar gradient disappears, therefore in this case the magnetic field can be written as a gradient of a scalar potential $\boldsymbol{B} = \nabla\phi$. For the potential $\phi(x, y, s)$ a Taylor expansion can be performed at the position $\boldsymbol{0}$.

$$
\begin{aligned}
\phi(x, y, s) =&\phi(\boldsymbol{0}) + \frac{\partial \phi}{\partial x}(\boldsymbol{0})x + \frac{\partial \phi}{\partial y}(\boldsymbol{0})y + \frac{\partial \phi}{\partial s}(\boldsymbol{0})s + \\
&\frac{1}{2!}\frac{\partial^2 \phi}{\partial x^2}(\boldsymbol{0})x^2 + \frac{1}{2!}\frac{\partial^2 \phi}{\partial y^2}(\boldsymbol{0})y^2 + \frac{1}{2!}\frac{\partial^2 \phi}{\partial s^2}(\boldsymbol{0})s^2 + \\
&\frac{\partial^2 \phi}{\partial x \partial y}(\boldsymbol{0})xy + \frac{\partial^2 \phi}{\partial x \partial s}(\boldsymbol{0})xs + \frac{\partial^2 \phi}{\partial y \partial s}(\boldsymbol{0})ys + \cdots
\end{aligned}
\tag{A.2}
$$

Due to the mid plan symmetry (Figure 2.2), ϕ is an odd function regarding y.

$$\phi(x, y, s) = -\phi(x, -y, s) \tag{A.3}$$

This eliminates all terms that do not contain y with odd exponents. All terms dependent on s are shifted to the coefficients $A_{2m+1,n}(s)$ and the potential can be noted as follows

$$
\begin{aligned}
\phi(x, y, s) =&(A_{10} + A_{11}x + A_{12}\frac{x^2}{2!} + A_{13}\frac{x^3}{3!} + \cdots)y + \\
&(A_{30} + A_{31}x + A_{32}\frac{x^2}{2!} + A_{33}\frac{x^3}{3!} + \cdots)\frac{y^3}{3!} + \cdots.
\end{aligned}
\tag{A.4}
$$

For the development of the magnetic field $\boldsymbol{B} = \nabla\phi$ up to the 2nd order, only coefficients up to the 3rd order must be considered and so the magnetic field has the components

$$
\begin{aligned}
B_x(x, y, s) &= \frac{\partial \phi}{\partial x} = A_{11}y + A_{12}xy + \cdots, \\
B_y(x, y, s) &= \frac{\partial \phi}{\partial y} = A_{10} + A_{11}x + \frac{1}{2!}(A_{12}x^2 + A_{30}y^2) + \cdots, \\
B_s(x, y, s) &= \frac{1}{1 + hx}\frac{\partial \phi}{\partial s} = \frac{1}{1 + hx}(A'_{10}y + A'_{11}xy + \cdots).
\end{aligned}
\tag{A.5}
$$

From the Laplace equation $\nabla^2 \phi = 0$ follows an expression for A_{30}, where only terms have to be considered which are independent of both x and y.

$$A_{30} = -A_{10}'' - A_{12} - hA_{11} + \cdots \tag{A.6}$$

The coefficients of B_y from equation (A.5) can be directly assigned to multipoles from a field expansion $B_y(x, y)$ around (x=0,y=0).

$$B_y(x, y) = B_y(\mathbf{0}) + \frac{dB_y}{dx}(\mathbf{0})x + \frac{1}{2!}\frac{d^2 B_y}{dx^2}(\mathbf{0})x^2 + ... \tag{A.7}$$

So $A_{10} = B_y(\mathbf{0}) = B_0$, $A_{11} = \frac{dB_y}{dx}(\mathbf{0})$ and $A_{12} = \frac{d^2 B_y}{dx^2}(\mathbf{0})$. Usually the terms $h(s)$, $n(s)$ and $\beta(s)$ are defined to characterise the dipole, quadrupole and sextupole strengths.

$$h = \frac{q}{p_s}B_0$$
$$n = -\frac{1}{hB_0}\frac{\partial B_y}{\partial x}(\mathbf{0}) \tag{A.8}$$
$$\beta = \frac{1}{2!h^2 B_0}\frac{\partial^2 B_y}{\partial x^2}(\mathbf{0})$$

This results directly in the coefficients

$$A_{10} = \frac{p_s}{q}h\,,$$
$$A_{11} = \frac{p_s}{q}(-nh^2)\,,$$
$$A_{12} = \frac{p_s}{q}2\beta h^3\,, \tag{A.9}$$
$$A_{10}' = \frac{p_s}{q}h'\,,$$
$$A_{11}' = \frac{p_s}{q}(-n'h^2 - 2nhh')\,,$$
$$A_{30} = \frac{p_s}{q}(-h'' - 2\beta h^3 + nh^3)\,.$$

Using the Taylor expansion at Position $\mathbf{0}$ for

$$\frac{1}{1 + hx} = 1 - h + 2h^2 + \cdots, \tag{A.10}$$

the development of the magnetic field up to the 2nd order can be written in the general form.

$$\begin{aligned}
B_x(x, y, s) &= \frac{p_s}{q}\left(-nh^2 y + 2\beta h^3 xy\right) \\
B_y(x, y, s) &= \frac{p_s}{q}\left(h - nh^2 x + \beta h^3 x^2 - \frac{1}{2}(h'' + 2\beta h^3 - nh^3)y^2\right) \\
B_s(x, y, s) &= \frac{p_s}{q}\left(h'y - (n'h^2 + 2nhh' + hh')xy\right)
\end{aligned} \tag{A.11}$$

A.2. Derivation of the change of the radial momentum within a gap

With values for the synchronous particle in the centre of the gap has gained half of the energy $\beta_s = \frac{\beta_f - \beta_i}{2}$ and $\gamma_s = \frac{\gamma_f - \gamma_i}{2}$ and using equations (2.98) and (2.102), the radial Lorentz force can be written as

$$\frac{dp_r}{dt} = q(E_r + \beta c B_\theta) = -q\gamma_s(1 - \beta\beta_s)I_1(Kr)E_0T\sin\phi \, . \tag{A.12}$$

With the radial momentum near axis is $p_r = mc\beta\gamma r'$ and $r' = \frac{dr}{dz}$ follows

$$mc\beta\gamma\frac{dr}{dt} = -q\gamma_s(1 - \beta\beta_s)I_1(Kr)E_0T\sin\phi \, ,$$
$$mc\beta\gamma\frac{dr}{dt}\frac{dt}{dz} = -q\gamma_s(1 - \beta\beta_s)I_1(Kr)E_0T\sin\phi\frac{1}{v_z} \, , \tag{A.13}$$
$$\beta\gamma r' = -\frac{1}{mc}q\gamma_s(1 - \beta\beta_s)I_1(Kr)E_0T\sin\phi\frac{1}{\beta c}$$

and so the change of the radial momentum in the gap is

$$\Delta(\beta\gamma r') = \frac{1}{mc}\int_0^L -q\gamma_s(1 - \beta\beta_s)E_0TI_1(Kr)\sin\phi\frac{dz}{\beta c}$$
$$= -\frac{q\gamma_s(1 - \beta\beta_s)E_0TLI_1(Kr)\sin\phi}{mc^2\beta} \, . \tag{A.14}$$

Since $\beta \approx \beta_s$ and using the relativistic relation $\gamma^2 = \frac{1}{1-\beta^2}$

$$\Delta(\beta\gamma r') = -\frac{qE_0TLI_1(Kr)\sin\phi}{mc^2\gamma_s\beta_s} \, . \tag{A.15}$$

The value of modified Bessel function can be approximated to $I_1(Kr) \approx \frac{\pi r_i}{\gamma_s\beta_s\lambda}$

$$\Delta(\beta\gamma r') = \beta_f\gamma_f r'_f - \beta_i\gamma_i r'_i = -\frac{\pi qE_0TL\sin\phi}{mc^2\gamma_s^2\beta_s^2\lambda}r_i \, . \tag{A.16}$$

The index i marks the initial values before the gap and the index f the final values after the acceleration. Thus the radial momentum change can be noted down

$$r'_f = -\frac{\pi qE_0TL\sin\phi}{mc^2\gamma_s^2\beta_s^2\lambda}\frac{1}{\beta_f\gamma_f}r_i + \frac{\beta_i\gamma_i}{\beta_f\gamma_f}r'_i \, . \tag{A.17}$$

A.3. Algorithms

A.3.1. Genetic algorithm

Algorithm 1 genetic algorithm

1: **procedure** $\text{GENETIC}(K, F_{\text{end}}, s)$ $\qquad\qquad\qquad \triangleright K, F_{\text{end}}, s$ as defined in text
2: $\quad \boldsymbol{G} \leftarrow$ initialise set of K initial Genomes G_k with n random Genes each
3: $\quad \boldsymbol{x}_{\min} \leftarrow X(G_{\min})$ where $F(X(G_{\min})) \leq F(X(G_k)), \forall G_k \in \boldsymbol{G}$
4: $\quad$ **while** $F(\boldsymbol{x}_{\min}) > F_{\text{end}}$ **do**
5: $\qquad$ remove s elements from $\boldsymbol{G}$ with highest $F(X(G_k))$
6: $\qquad \boldsymbol{G}' \leftarrow$ create empty set
7: $\qquad$ **for** $k \leftarrow 1, K$ **do**
8: $\qquad\quad G_a \leftarrow$ select random $G \in \boldsymbol{G}$
9: $\qquad\quad G_b \leftarrow$ select random $G \in \boldsymbol{G}$
10: $\qquad\quad \hat{O} \leftarrow$ one or combination of $\hat{C}_u, \hat{C}_n, \hat{C}_h, \hat{A}, \cdots$ $\qquad \triangleright$ recombination
11: $\qquad\quad G_c \leftarrow \hat{O}(G_a, G_b)$ $\qquad\qquad\qquad\qquad\qquad \triangleright$ child genome
12: $\qquad\quad G'_c \leftarrow$ one of $\hat{M}_\sigma(G_c), \hat{M}_j(G_c)$ $\qquad\qquad\qquad \triangleright$ mutation
13: $\qquad\quad$ add G'_c to set $\boldsymbol{G}'$
14: $\qquad \boldsymbol{G} \leftarrow \boldsymbol{G}'$
15: $\qquad \boldsymbol{x}_{\min} \leftarrow X(G_k)$ where $F(X(G_k))$
16: $\quad$ **return** $\boldsymbol{x}_{\min}$

A.3.2. Particle swarm algorithm

Algorithm 2 particle swarm algorithm

1: **procedure** $\text{PARTICLESWARM}(K, m, w_c, w_s, F_{\text{end}})$ $\quad \triangleright K, m, w_c, w_s, F_{\text{end}}$ as defined in text
2: $\quad \boldsymbol{X} \leftarrow$ initialise set of K initial positions $\boldsymbol{x}_k$
3: $\quad \boldsymbol{V} \leftarrow$ initialise set of K initial velocities $\boldsymbol{v}_k$
4: $\quad \boldsymbol{x}_{\min} \leftarrow \text{argmin}_{\boldsymbol{x} \in \boldsymbol{X}}(F(\boldsymbol{x}))$
5: $\quad \boldsymbol{x}_{k,\min} \leftarrow \text{argmin}_{\boldsymbol{x} \in \boldsymbol{X}}(F(\boldsymbol{x}))$
6: $\quad$ **while** $F(\boldsymbol{x}_{\min}) > F_{\text{end}}$ **do**
7: $\qquad$ **for** $k \leftarrow 1, K$ **do**
8: $\qquad\quad r_1 \leftarrow \text{rnd}(0, 1)$
9: $\qquad\quad r_2 \leftarrow \text{rnd}(0, 1)$
10: $\qquad\quad \boldsymbol{v}_k \leftarrow m\boldsymbol{v}_k + w_c r_1(\boldsymbol{x}_{k,min} - \boldsymbol{x}_k) + w_s r_2(\boldsymbol{x}_{min} - \boldsymbol{x}_k)$
11: $\qquad\quad \boldsymbol{x}_k \leftarrow \boldsymbol{x}_k + \boldsymbol{v}_k$
12: $\qquad\quad$ **if** $F(\boldsymbol{x}_k) < F(\boldsymbol{x}_{\min})$ **then** $\boldsymbol{x}_{\min} \leftarrow \boldsymbol{x}_k$
13: $\qquad\quad$ **if** $F(\boldsymbol{x}_k) < F(\boldsymbol{x}_{k,\min})$ **then** $\boldsymbol{x}_{k,\min} \leftarrow \boldsymbol{x}_k$
14: $\quad$ **return** $\boldsymbol{x}_{\min}$

A.3.3. Simulated annealing

Algorithm 3 simulated annealing

1: **procedure** SIMULATEDANNEALING(r, T_t, F_{end}) $\triangleright r, T_t, F_{\text{end}}$ as defined in text
2:     $t \leftarrow 0$
3:     $\boldsymbol{x} \leftarrow$ random initial state
4:     $\boldsymbol{x}_{\text{min}} \leftarrow \boldsymbol{x}$
5:     **while** $F(\boldsymbol{x}_{\text{min}}) > F_{\text{end}}$ **do**
6:         $\boldsymbol{x}' \leftarrow \boldsymbol{x} + \Delta\boldsymbol{x}$ with random $|\Delta\boldsymbol{x}| < r$
7:         **if** $F(\boldsymbol{x}') \leq F(\boldsymbol{x})$ **then** $\boldsymbol{x} \leftarrow \boldsymbol{x}'$
8:         **else**
9:             $p \leftarrow e^{\frac{F(\boldsymbol{x}') - F(\boldsymbol{x})}{T_t}}$
10:             **if** $\text{rnd}(0, 1) < p$ **then** $\boldsymbol{x} \leftarrow \boldsymbol{x}'$
11:         **if** $F(\boldsymbol{x}) < F(\boldsymbol{x}_{\text{min}})$ **then** $\boldsymbol{x}_{\text{min}} \leftarrow \boldsymbol{x}$
12:         $t \leftarrow t + 1$
13:     **return** $\boldsymbol{x}_{\text{min}}$

A.3.4. Bayesian optimisation

Algorithm 4 bayesian optimisation

1: **procedure** BAYOPT(function f, kernel K, acquisition function a, N_{warmup}, N_{cycles})
2:     $y_{\text{best}} \leftarrow \infty$
3:     **for** $i = 1$ to N_{warmup} **do**
4:         randomly select $\boldsymbol{x}_i$
5:         compute loss function $y_i \leftarrow F(\boldsymbol{x}_i)$
6:         **if** $y_i \leq y_{\text{best}}$ **then**
7:             $\boldsymbol{x}_{\text{best}} \leftarrow \boldsymbol{x}_i$
8:             $y_{\text{best}} \leftarrow y_i$
9:     **for** $i = N_{\text{warmup}} + 1$ to N_{cycles} **do**
10:         update kernel matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{i \times i}$ according to equation (3.13)
11:         condition the Gaussian process model on all $F(\boldsymbol{x}_i) = y_i$, calculate $\boldsymbol{\mu}(\boldsymbol{x}')$ and $\sigma(\boldsymbol{x}')$ of $F(\boldsymbol{x}')$
12:         $\boldsymbol{x}_i \leftarrow \text{argmin}_x a(\boldsymbol{\mu}(\boldsymbol{x}'), \sigma(\boldsymbol{x}'), y_{\text{best}})$
13:         compute loss function $y_i \leftarrow F(\boldsymbol{x}_i)$
14:         **if** $y_i \leq y_{\text{best}}$ **then**
15:             $\boldsymbol{x}_{\text{best}} \leftarrow \boldsymbol{x}_i$
16:             $y_{\text{best}} \leftarrow y_i$
17:     **return** $\boldsymbol{x}_{\text{best}}$