Designing patient-specific ridge filters using Monte Carlo simulations and a genetic optimization algorithm

Duke Bouwer[1,2]
Supervisors: Danny Lathouwers[2], Martin van Gijzen[1]

[1] EEMCS, Delft University of Technology, The Netherlands
[2] AS, Delft University of Technology, The Netherlands

February 14, 2025

Student name: Duke Alain Bouwer
Thesis committee: Danny Lathouwers, Martin van Gijzen, Johan Dubbeldam, Zoltán Perkó

Preface

This paper is written for my Bachelor end project, to finalize my Bachelor of Applied Mathematics and Bachelor of Applied Physics at Delft University of Technology (TU Delft). This project has been completed within the Radiation Science & Technology Department of the Applied Science faculty of the TU Delft, and the Numerical Analysis Section of the Faculty of Electrical Engineering, Mathematics and Computer Science of the TU Delft.

First of all, I would like to thank both my Thesis supervisors Danny Lathouwers and Martin van Gijzen for all their support throughout this project. I am very grateful to Danny for the valuable feedback and the critical questions during the progress meetings and for his concern for my mental well-being. To Martin, I am very grateful for his expertise, availability to questions, helping to keep the project grounded and goals realistic, and his understanding of the situation.

To Lisa Keurentjes, I express many thanks for their help, patience, and understanding while learning C++. As well for her continued availability in responding to my many questions at often sporadic times, and the company during coffee.

I want to thank my rowing team, for their help in clearing my head when needed, and at the same for their continued support when I could not be there physically and oftentimes mentally. Gijs, Maurits, Max, Sam, Sem, and Tjeu: Thank you for being such good friends throughout this period! Moreover many thanks to all the people I have coached throughout this time, for allowing me to vent my heart and grievances over coffee.

Finally, I want to thank my family for being there when needed, for their continued interest, and for their help and input! Hanne, you were always willing to help, were there when I needed it, and motivated me to remain focused on achieving my goals.

Everyone thank you very much for all the help!

Duke Bouwer
Delft, December 2024

Abstract

Proton therapy is a form of radiation therapy, that leverages the unique properties of protons to maximize dose deposition in treatment volumes. The usefulness of proton therapy treatment, in sparing healthy tissue, becomes even more evident with the incorporation of the FLASH effect. FLASH delivers ultra-high dose rates with minimal treatment time while maintaining therapeutic efficiency in eradicating tumours. However, due to practical challenges such as energy layer switching in pencil beam scanning systems the clinical applications are limited.

This thesis researched the development of patient-specific ridge filters (RFs) for proton therapy using an optimization algorithm. Ridge filters are energy modulators that help enable dose delivery without energy layer switching, making the FLASH effect feasible as was shown in 2018 [1]. Previous studies used static and dynamic methods [2] and fluence-based optimization [3] to construct patient-specific RFs. This study presents a novel framework for optimizing patient-specific RFs using a combination of TOPAS, a Monte Carlo particle simulation software that accurately models particle transport, and PyGAD, a Python-based genetic algorithm (GA) module. This class of optimizers is effective in cases where no derivatives are available or are very difficult to compute. GAs do this by testing various simulations and evaluating these using a fitness function. The methodology involves simulating dose distribution in a scoring volume, optimizing ridge pin geometry, and evaluating performance using fitness functions.

The results demonstrate that the proposed framework effectively generates patient-specific RFs with minimal deviation from the desired dose distribution in simple cases, with a maximum dose difference of 2.66 % and mean dose of 99.21 % over the region of interest. Comparative analysis with prior approaches shows that the framework achieves similar results. However, applying the framework to cases with obstructions in the scoring volume requires further refinement of the algorithm. The findings provide a basis for using GAs for constructing patient-specific RFs for FLASH proton therapy. Future work should be aimed at refining the GA and Monte Carlo simulation and assessing the viability of producing the generated RFs.

Contents

1 Introduction **1**

2 Theory **3**
2.1 Proton Therapy . . . 3
2.1.1 FLASH Radiation Therapy . . . 4
2.1.2 Spread-out Bragg Peak . . . 4
2.1.3 Pencil Beam Scanning . . . 6
2.2 Ridge Filters . . . 7
2.3 Monte Carlo Simulation . . . 9
2.4 Genetic Algorithms . . . 10

3 Methodology **11**
3.1 Experimental set-up . . . 11
3.1.1 Ridge filter . . . 12
3.2 Monte Carlo Simulation . . . 17
3.2.1 Settings . . . 17
3.2.1.1 Physics settings . . . 17
3.2.1.2 Beam settings . . . 17
3.2.1.3 General settings . . . 18
3.2.2 Simulation setup . . . 20
3.2.2.1 Scorer . . . 20
3.2.2.2 Scorer modifications . . . 21
3.2.2.3 Ridge filter . . . 22
3.2.3 Beam source . . . 23
3.3 Data processing . . . 24
3.4 Genetic Algorithm . . . 25
3.4.1 Genetic Algorithm settings . . . 25
3.4.2 Initial population . . . 25
3.4.3 Fitness function . . . 25
3.4.4 Parent Selection . . . 28
3.4.5 Crossover function . . . 28
3.4.6 Mutation function . . . 28
3.4.7 Convergence criteria . . . 29
3.5 Code . . . 29

4 Results and Discussion **30**
4.1 Results . . . 30
4.1.1 Base case . . . 30
4.1.2 Benchmark case . . . 33
4.1.3 Obstructed Scorer . . . 34
4.2 Discussion . . . 36

5 Conclusions and Future Work **37**

List of Figures

2.1 Depth-dose curves for radiotherapy. Dose distributions as a function of depth in water and energy level are shown. For electrons (20 MeV), photons (18 MeV), Carbon Ions (300 MeV), and Protons (130 MeV). Showcasing different dose-depth distributions based on particle type. [4]. 3
2.2 Bragg peak for different energies. (a) The deposition of the dose varies according to the energy of the proton. The maximum of the Bragg peak varies according to the energy; (b) The relationship Range - Energy for protons in water is shown. Edited to have depth on the z-axes instead of range. [5]. 4
2.3 Depth-dose curves for photon and proton beams. The proton Bragg peak allows for precise dose delivery to the tumour target and a relative dose elimination compared with photon exit dose [6]. 5
2.4 Schematic view of pencil beam scanning, here two pairs of dipole magnets are illustrated. The first magnet is responsible for horizontal scanning and the second magnet for vertical scanning. By, adjusting the magnets strengths the entire tumour volume is treated. Image from Maasland [7]. 6
2.5 Cross-section of a single ridge pin, without the base. 7
2.6 Ridge filter produced by GSI, optimized for a SOBP with a width of 2.5 cm, with an initial beam energy of 150 MeV [8]. 8
2.7 Flowchart explaining how a GA works in pseudo-code 10
3.1 The beam source positioned at z_1, the RF positioned at z_3, and the scorer, as water box, positioned at z_5. The distance between the proton source and RF is denoted by d_1, 25 cm, and the distance between the RF and the scorer by d_2, 25 cm. The distance between z_2 and z_3 is the height of the RF, 35 mm, and the distance between z_4 and z_5 is the depth of the water box, 20 cm. 11
3.2 Front and top view of a simplified ridge filter design with a 2 by 2 grid of ridge pins, with the central axis shown in the front view. With values corresponding to the experimental setup listed, ridge pin height, baseplate height, RF width, and RF depth. 12
3.3 Schematic of an RF with the ridge pin height (30 mm), maximum ridge pin width (6 mm), base plate height (5 mm), and pitch-to-pitch distance (6 mm) between the ridge pins 13
3.4 Ridge pin schematic with the ridge pin height of 30 mm, step height of 1 mm, and the maximum ridge pin width of 6 mm shown. Number of steps not to scale. 14
3.5 Impact of ridge filter width on protons scattering around the ridge filter. 15
3.6 Boundaries of the region of interest when simulating limit cases. 16
3.7 Simulation results with varying the number of particles (”NumberOfHistoriesInRun”) 19
3.8 2-Dimensional illustration of the water box scored across the z-axis, with 21 bins. The points z_4 and z_5 are the same as in figure 3.1, the depth of the water box is 20 cm 20
3.9 2-Dimensional illustration of the water box scored across the z-axis, with 21 bins, actual simulation uses 100 bins. In the middle of the water box a piece of ”bone” is inserted halfway along the y axis into the water box. The points z_4 and z_5 are the same as in figure 3.1, depth of the scorer is 20 cm. 21
3.10 Front and top view of a simplified ridge filter design with a 2 by 2 grid of ridge pins, with the central axis shown in the front view. With values corresponding to the experimental setup listed, ridge pin height (30 mm), baseplate height (5 mm), RF minimum width (40 mm), and RF minimum depth (40 mm). 22
3.11 Average convergence of a generation containing a population of 40 for the first 10 generations 27

3.12 Visualized example of TOPAS output with 3 bins in the region of interest (ROI) between the proximal and distal edge. The desired output in the ROI is $\overline{y_1} = \overline{y_2} = \overline{y_3} = 1$, the example output is $y_1 = 0.8$, $y_2 = 0.9$, and $y_3 = 0.7$ 28
3.13 QR Code to the GitHub repository containing the code. 29
4.1 Heatmap of ridge pin coordinates, with navy denoting a low frequency, and yellow denoting a high frequency. Illustrates the convergence to the final optimal ridge pin design for the ridge filter, this is plotted in red. 31
4.2 The IDD of a SOBP generated by the optimized ridge pin, as illustrated in figure 4.5. With the distal edge at 14.9 cm and the proximal edge at 10.9 cm, and a fitness value of 0.0109. . 32
4.3 Average fitness value plotted against generation in blue, with the best fitness value plotted in red 32
4.4 IDD showing the SOBP received from simulating using anonymously provided RF, cropped to show interval between 10 cm and 18 cm. Proximal edge at 13.1 cm and distal edge at 15.1 cm 33
4.5 Heatmap of ridge pin coordinates, with navy denoting a low frequency, and yellow denoting a high frequency. Illustrates the convergence to the final optimal ridge pin design for the ridge filter, this is plotted in red. 34
4.6 The IDD of a SOBP generated by the optimized ridge pin, as illustrated in figure 4.5. With the distal edge at 14.9 cm and the proximal edge at 10.9 cm, and a fitness value of 0.23819. . 35
4.7 Average fitness value plotted against generation in blue, with the best fitness value plotted in red 35

List of Tables

1 List of physics modules and their function in TOPAS [9] 17
2 Beam parameters and their values. 17
3 Elemental composition of bone. 21
4 Simulation results for optimized ridge filter 30
5 Simulation results for benchmark ridge filter 33
6 Simulation results for optimized ridge filter with a bone plate inserted halfway into the scoring volume. 34

Nomenclature

Abbreviations

Abbreviations	Definition
BEP	Bachelor End Project
ENDF	Evaluated Nuclear Data File
FLASH-RT	FLASH radiation therapy
GA	Genetic algorithm
HPTC	Holland Proton Therapy Centre
IMPT	Intensity-modulated proton therapy
IDD	Integral depth dose
LANL	Los Almos National Laboratory
PBS	Pencil beam scanning
PBS-IMPT	Pencil beam scanning intensity modulated proton therapy
PS	Passive scattering
RMAE	Root maximum absolute error
RMSE	Root mean square error
RF	Ridge filter
ROI	Region of interest
RT	Radiation therapy
SOBP	Spread-out Bragg peak
SSS	Steady state selection
TOPAS	TOol for PArticle Simulation
TU Delft	Delft University of Technology
UDHR	Ultra-high dose rate

Symbols

Symbol	Definition	Unit
D	Dose	[Gy]
	Dose rate	$\frac{[Gy]}{[s]}$
s	Seconds	[s]
eV	Electron-volt	[J]
ρ	Density	$\frac{[Kg]}{[m^3]}$

1 Introduction

Cancer accounts for almost 15% of global deaths, this constitutes cancer as being the second most common worldwide cause of death [10]. Despite the significant progress in the most common treatment methods for cancer patients surgery, radiotherapy, chemotherapy, immunotherapy, and hormone therapy, further progress remains imperative to reduce the loss of life as a result of cancer. Radiation therapy is employed in over half of cancer treatments [11], often alongside surgery and chemotherapy [8].

Radiation therapy is used to treat malignant tumours. In this process, high doses of radiation are delivered to patients. Radiation damages the DNA of cancerous tumours to the point that DNA repair and division in the tumour is no longer possible, ultimately leading to cell death of the malignant tumour [8]. The challenge of this process is maximizing the damage to cancerous tissue so that cell death occurs while at the same time sparing as much healthy tissue as possible. Proton therapy achieves this goal by using the unique physical and radiobiological properties of protons. Recent advancements in radiotherapy with ultra-high dose rates (UDHR), most notably the discovery of the FLASH effect by Favaudon et al. in 2014 [12], have shown potential. The FLASH effect has demonstrated equivalent therapeutic efficiency for killing malignant tumours while sparing an increased amount of healthy tissue [13]. The discovery of the FLASH effect has further increased the research potential of proton therapy and this field has gained increasing attention since [2, 13, 14].

The clinical implementation of FLASH in combination with proton therapy still presents challenges. First among these is the reduction of time needed to deliver the dose in clinical settings using pencil beam scanning [2]. To tackle this Patriarca et al. [1] proposed using ridge filters in 2018. Ridge filters remove the need to switch energy layers, thereby significantly reducing treatment times in systems using pencil beam scanning (PBS) while still achieving the ultra-high dose rates needed for FLASH.

Given the established need for ridge filters, it raises the question of how to produce these. Nominally, every patient's malignant tumours are unique. This uniqueness implies that achieving precise dosing to only the tumour requires patient-specific ridge filters tailored to the specific shape, size, and location of the tumour. Currently, there are multiple methods for optimizing ridge filters for patient tumours. For example, Zhang et al. used a unique dose-influence matrix in 2022 [2]. However, the two methods tried by Zhang et al., static and dynamic, had a couple of issues. Foremost among them are limited reusability in the procedure for the static case and the increased complexity in the dynamic case. Another paper by Liu et al. used a matrix-based approach using linear energy transfer influence matrices in 2023 [3]. However, the treatment planning software used in this paper, MatRad, optimizes based on fluences instead of a dose distribution.

This paper researches the possibility of constructing patient-specific RFs based on a dose distribution using an optimizer. Both of these papers used one type of optimizer, however other optimization methods are available. One class of such methods is that of the Genetic Algorithms (GA). GAs are very useful in so-called black-box optimization, otherwise called derivative-free optimization [15]. The problem of optimizing ridge filters to patient-specific tumours falls into the category of not having a gradient. This makes genetic algorithms particularly useful in this scenario due to their effectiveness in black-box optimization.

In this paper the following question will be researched: ”Can patient-specific ridge filters for proton therapy be constructed based on desired dosage distributions through an optimization procedure? And how does this approach compare to previously established research in terms of achieving target dosage distributions?”. The report is structured as follows, it consists of five sections: in section 2 the theoretical background will be established introducing topics such as proton therapy, FLASH, ridge filters, Monte Carlo simulations, and genetic algorithms. After that in section 3 the implementation of the Monte Carlo simulation and the genetic algorithm will be shown. Moreover, a description is given of how the simulation data has been processed. Continuing in section 4 the results are presented and discussed. The conclusion and future research recommendations are in section 5.

2 Theory

2.1 Proton Therapy

The physical and radiobiological properties of particles like protons enable the delivery of a high dose to a tumour, with the energy level determining the specific distal edge of the dose deposition. This behaviour is described by the Bethe-Bloch equation, where the energy and dose deposited are inversely proportional to the square velocity. As a consequence most of the energy is deposited right before the particle stops [16]. This characteristic allows proton therapy to deliver more dose to the tumour volume while minimizing radiation beyond the distal edge, particularly useful for tumours near critical organs like the lungs and heart [17–19].

This property makes charged particles unique, as conventional radiation therapy with photons irradiate past the tumour, this is illustrated in figure 2.1. In this figure, the electron and photon deposit a substantial dose across most of the depth. Whereas charged particles such as protons and Carbon ions exhibit a sharp dose drop-off, which depends on the particle's energy level. The sharp dose deposition is known as the Bragg peak and the entire depth dose curve as the Bragg curve [20].

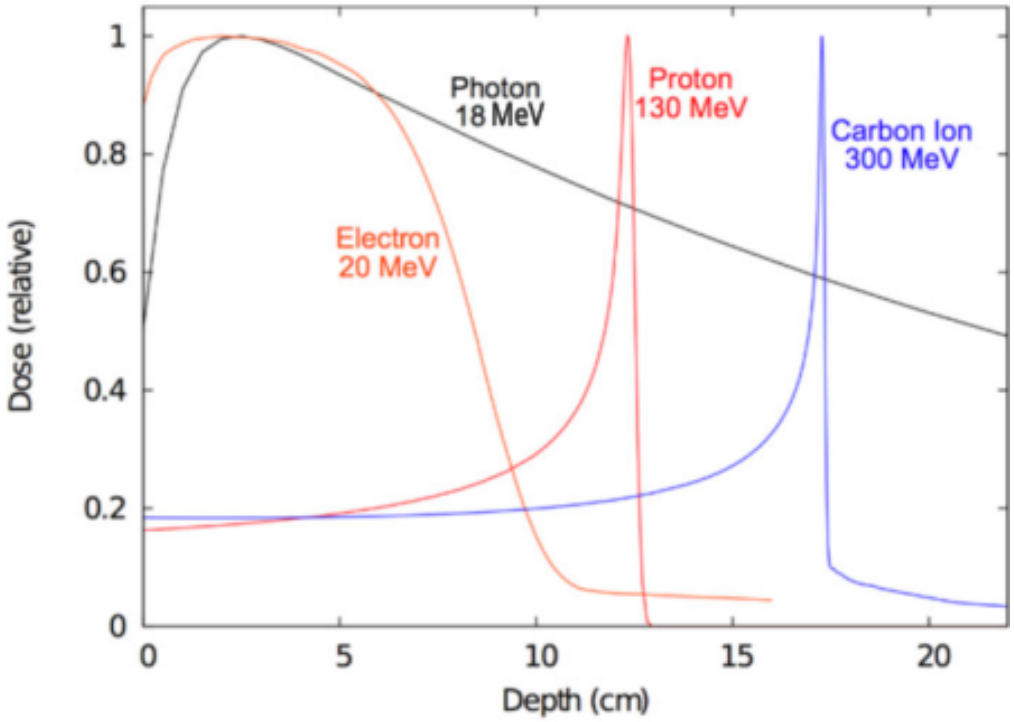


Figure 2.1: Depth-dose curves for radiotherapy. Dose distributions as a function of depth in water and energy level are shown. For electrons (20 MeV), photons (18 MeV), Carbon Ions (300 MeV), and Protons (130 MeV). Showcasing different dose-depth distributions based on particle type. [4].

Figure 2.2 further illustrates the relation between the Bragg peak and the energy level. Higher energy levels allow for the Bragg peak to extend further.

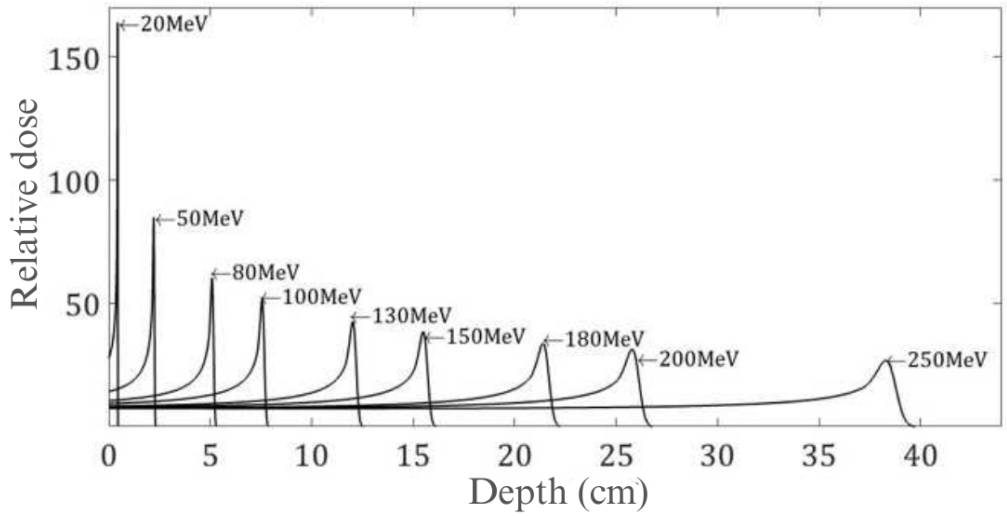


Figure 2.2: Bragg peak for different energies. (a) The deposition of the dose varies according to the energy of the proton. The maximum of the Bragg peak varies according to the energy; (b) The relationship Range - Energy for protons in water is shown. Edited to have depth on the z-axes instead of range. [5].

The Bragg peak characteristic makes proton therapy particularly useful for treating deep-seated tumours and those close to organs at risk. This is due to minimal dosage being deposited beyond the distal edge of the Bragg peak thereby sparing healthy tissues and organs at risk.

2.1.1 FLASH Radiation Therapy

FLASH radiation therapy (FLASH-RT) is a promising approach that, contrary to most current radiation therapy, delivers an ultra-high dose rate (> 40 Gy/s) in a short time frame (< 200 ms [21, 22] or ~ 500 ms [12, 13] depending on the study). To provide a frame of reference, conventional therapy currently in use delivers a standard dose of up to 1 Gy/min, with treatment times of over a minute [12]. Recent studies have shown that FLASH-RT shows equivalent efficiency for killing tumours as traditional methods, while also offering more sparing of normal tissue [13]. This is due to FLASH-enabled proton therapy sparing healthy tissues both before the proximal edge [13] and beyond the distal edge, thanks to the Bragg peak, while maintaining the same therapeutic efficiency [7].

The initial proposition for the FLASH effect was based on transient oxygen depletion [23], but recent studies have shown that oxygen depletion might not fully describe the underlying mechanism [24, 25]. Other hypotheses now include the kinetics of radicals and an intrinsic differentiation in oxygen delivery between tumours and healthy tissue micro-environment [14]. However, a definitive explanation of the FLASH effect remains elusive.

2.1.2 Spread-out Bragg Peak

As explained in section 2.1, the location of the Bragg peak is dependent on the particle's energy. This means that if the proton beam consists of one energy level, also known as mono-energetic, the majority of the dose will irradiate only one part of the tumour. To address this it becomes necessary to create a spread-out Bragg peak (SOBP) to cover the entire tumour. The production of a SOBP through the use of a weighted collection of mono-energetic proton beams has been

studied before by Bortfeld in 1996 [26]. The SOBP allows for the treatment of larger tumours and larger three-dimensional volumes. Figure 2.3 illustrates that the SOBP covers the entire tumour target in comparison, the Bragg peak of the mono-energetic beam only covers a fraction of the total tumour target. The blue line represents a photon with its integral depth doses (IDD), which deposits considerable doses beyond the distal edge of the tumour target.

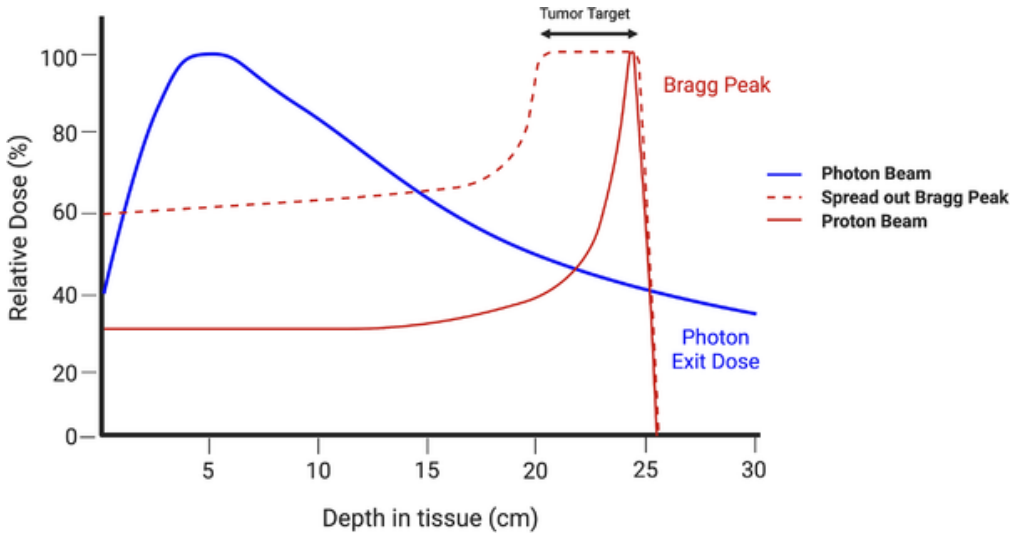


Figure 2.3: Depth-dose curves for photon and proton beams. The proton Bragg peak allows for precise dose delivery to the tumour target and a relative dose elimination compared with photon exit dose [6].

2.1.3 Pencil Beam Scanning

Pencil beam scanning (PBS) is a technique in proton therapy to precisely direct a proton to the tumour [7]. This technique offers advantages, such as high dose rates and good dose conformity. Making it suitable for FLASH-RT, as described in section 2.1.1. PBS is commonly used in intensity-modulated proton therapy (IMPT). In a standard PBS-IMPT delivery, the system switches between energy layers to shift the Bragg peak. However, this takes ~ 1 s per energy layer leading to total treatment time taking over 20 s [2]. Multiple beam angles, which are often needed, further extend the treatment time.

A schematic overview of PBS is illustrated in figure 2.4. Typically, a narrow proton beam (diameter 1 cm) [27] is used to fully irradiate a tumour. The PBS starts with the first layer, closest to the distal edge, and then irradiates spot by spot [28]. The energy level of the beam is adjusted to move through the layers, shifting the Bragg peak to the correct depth. Dipole magnets steer the proton beam within each layer so that the entire layer can be covered spot by spot ensuring full tumour coverage [7].

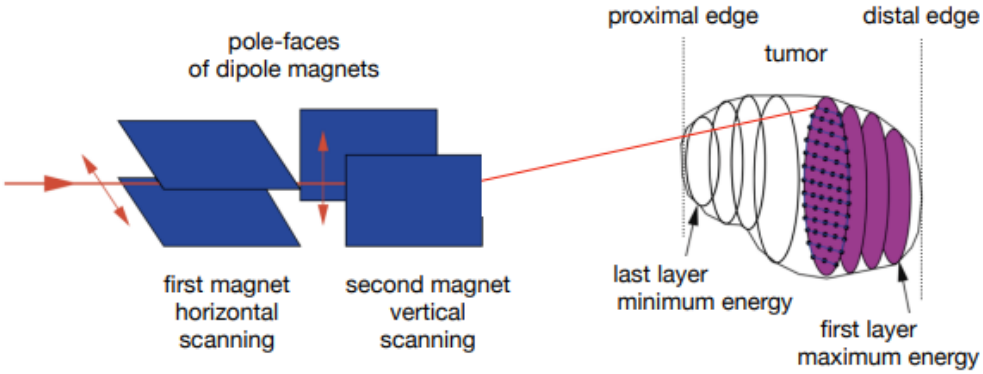


Figure 2.4: Schematic view of pencil beam scanning, here two pairs of dipole magnets are illustrated. The first magnet is responsible for horizontal scanning and the second magnet for vertical scanning. By, adjusting the magnets strengths the entire tumour volume is treated. Image from Maasland [7].

To reduce treatment times to FLASH parameters, static ridge filters (RFs) have been used. Although knowledge of RFs has been around since the 1960s, applying RFs for achieving the FLASH effect was first demonstrated by Patriarca et al. in 2018 [1]. RFs modulate energy to form an SOBP (see section 2.1.2) and reduce treatment times by eliminating the time-consuming steps of energy/layer switching and modulation wheel rotation [2], enabling the use of PBS for FLASH-RT.

Other methods, such as passive scattering (PS), are less suitable for FLASH-RT. As PS distributes the dose over a larger volume, not all protons reach the target. This reduces the dose rate and falls below the threshold needed for FLASH [7, 13].

2.2 Ridge Filters

As discussed in section 2.1.3, ridge filters (RFs) function as energy modulators. They do so by altering the energy of the protons as they pass through the material. When protons interact with matter three primary interactions occur, Coulomb collisions (interactions between electrons and nuclei), bremsstrahlung radiation loss, and nuclear interactions [29]. Among these, the Coulomb collisions contribute the most significant effect, as the continuous energy loss from collisions with electrons is the primary mechanism of energy dissipation [30]. In simpler terms, the more matter a proton passes through the more energy it loses, resulting in the construction of the Bragg peak.

RFs exploit this property when sending a proton beam through an RF all the individual protons experience different degrees of interaction. Some protons interact with a substantial amount of material and some of them almost none. Thus creating a weighted collection of mono-energetic protons and constructing an SOBP, as explained in section 2.1.2.

Illustrated in figure 2.5 is a schematic of a single ridge pin. Depending on the height of the ridge pin the amount of material varies. Resulting in some protons having significant interactions and losing a considerable amount of energy, while some pass through minimal material, losing less energy. This process produces a weighted set of mono-energetic beams giving rise to an SOBP. Tweaking the design of the ridge pin allows for precise shaping of the SOBP. To achieve a uniform dose multiple ridge pins are used [31].

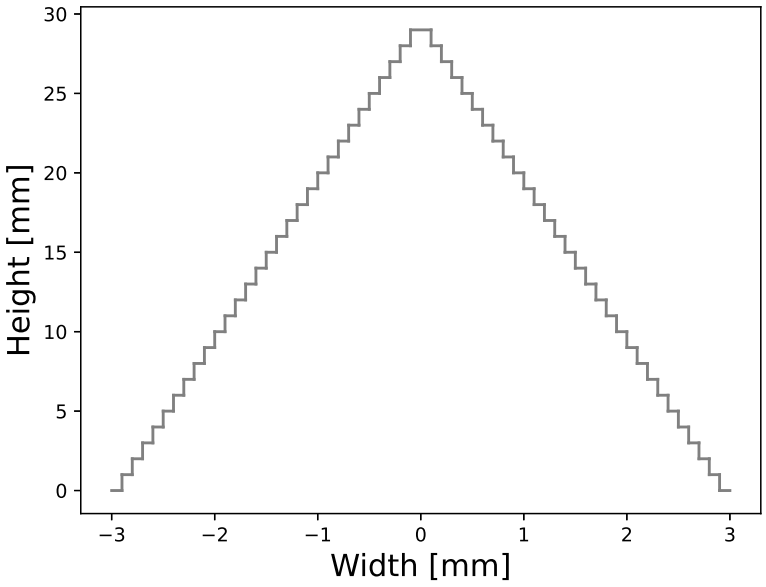


Figure 2.5: Cross-section of a single ridge pin, without the base.

The complete package of a RF includes ridge pins, a base plate, and additional supports for alignment. Figure 2.6, illustrates a RF that is optimized to produce a SOBP with a width of 2.5 cm, based on an initial beam energy of 150 MeV. And includes additional supports for alignment. It was developed by GSI and tested at the Holland Proton Therapy Centre (HPTC) [8].

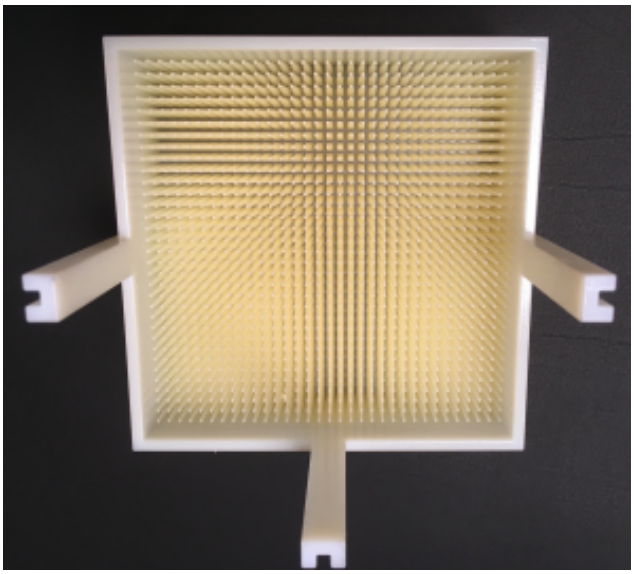


Figure 2.6: Ridge filter produced by GSI, optimized for a SOBP with a width of 2.5 cm, with an initial beam energy of 150 MeV [8].

2.3 Monte Carlo Simulation

The impact of changes in ridge filter design on the spread-out Bragg peak (SOBP) is evaluated through Monte Carlo Simulations. Simulation is favoured over physical production, as it facilitates iterative design adjustments efficiently. A Monte Carlo simulation is a type of simulation that uses random sampling and statistical modelling to estimate the result of a mathematical equation [32]. For example, the chance of a coin toss resulting in head or tails can be estimated by flipping the coin many times and observing the average outcome. Eventually, through the use of randomness, you get a deterministic answer. This concept can be traced back to Georges Louis Le Clerc, Comte de Buffon (1707-1788), an influential French scientist who used randomness to estimate probabilities [32]. However, this method used previously understood deterministic methods. Modern-day Monte Carlo methods invert this idea by defining the sampling distribution of the inputs a priori. Then the Monte Carlo simulation repeatedly simulates the model, each time taking a different set of inputs from the sampling distribution, eventually giving a set of outputs [33].

The use of Monte Carlo simulations in particle transport problems started at Los Almost National Laboratory (LANL), during the Manhattan project [34]. Since then almost 50 years have passed, with LANL still being a main source of Monte Carlo codes. They have continuously improved their Monte Carlo code, with the most recent version being MCNP6. Throughout all the years of development, improvements have been made to the nuclear physics parameters. This was done by using the Evaluated Nuclear Data File (ENDF), with MCNP6 using ENDF/B-VIII [34]. Alongside MCNP6 other particle transport Monte Carlo simulations have been developed. In particular, for the support of nuclear physics experiments, GEANT4 has been developed. GEANT4 was created for simulating the passage of particles through matter for use in among others medical- and radiation physics [35]. To make particle transport Monte Carlo simulations more accessible TOPAS was built upon GEANT4 [36]. In comparison to GEANT4, TOPAS does not need extensive knowledge of C++, and most simulations can be done using a text-based instruction file.

However, Monte Carlo simulations do have a couple of disadvantages, first of all, the amount of simulation runs must be defined. The amount of simulation runs is the number of times you repeat an entire simulation. This can be quite tricky as the number of simulations varies based on the goal. Secondly, as the Monte Carlo simulation is dependent on a large number of runs, this can result in very long computation times. Finally, pseudo-randomness is something you must be aware of in Monte Carlo simulations. This is because no computer is truly random. Say you use an algorithm to generate a sequence of random numbers, if you later start the algorithm at the exact same place the sequence of 'random' numbers will repeat [33]. According to Ripley in 1990 [37] and Wichmann in 2006 [38] you need at least a period of $2^{50} \simeq 10^{15}$ before pseudo-randomness is no longer a problem and that the simulations can handle highly parallel applications. Most operating systems such as Windows 11 are 64 bits [39], so there is a period of $2^{64} - 1 \simeq 10^{19}$. Therefore the period in the pseudo-randomness is not a concern in modern-day Monte Carlo particle transport simulation.

2.4 Genetic Algorithms

To automatically adjust the ridge filter design an optimizer is employed. Given the derivative-free nature of the problem Genetic Algorithms (GAs) are chosen as the class of optimizers. GAs are a family of powerful optimization techniques inspired by the idea of survival of the fittest. They do so by iteratively refining the solutions from previous generations until a near-optimal solution is reached [40]. One of the advantages of GAs is the fact that they do not necessarily need a continuous probability space to work. In cases where no continuous probability spaces are available, derivatives are typically also unavailable unless interpolated. A majority of optimization methods use derivatives to find an optimal value, as the derivative can give information on where you are concerning the desired optimum [40]. However, GAs are capable of using discrete probability space, making them suitable for situations where derivatives are not available. This is because GAs are probabilistic instead of deterministic. So GAs do not work with a singular solution, but rather a solution set and then evaluate the set of solutions using fitness functions [40]. Not needing derivatives is beneficial, as not all real-life applications involve continuous probability spaces or if they do the derivates may not be easy to find [15].

Figure 2.7 illustrates a flowchart of how a GA works. The algorithm begins by initializing a population in a probability space. The next step is evaluating which of the inputs of the initial population is 'good'. This is done through the use of a fitness function, otherwise called an optimization function. These solutions, called parents, are selected to generate the next generation's population. To produce new generations, a crossover function is applied, which combines two parents and produces new offspring. The idea is that the new offspring may outperform its parents. After this the new population is mutated, this is done to explore other parts of the probability space, thus avoiding local minima and reducing the chance of premature convergence of the GA. This process continues until the stopping criteria are satisfied, at which the algorithm is considered to have 'converged'. Choice in convergence criteria will have a significant impact on the location of the found optimum [40].

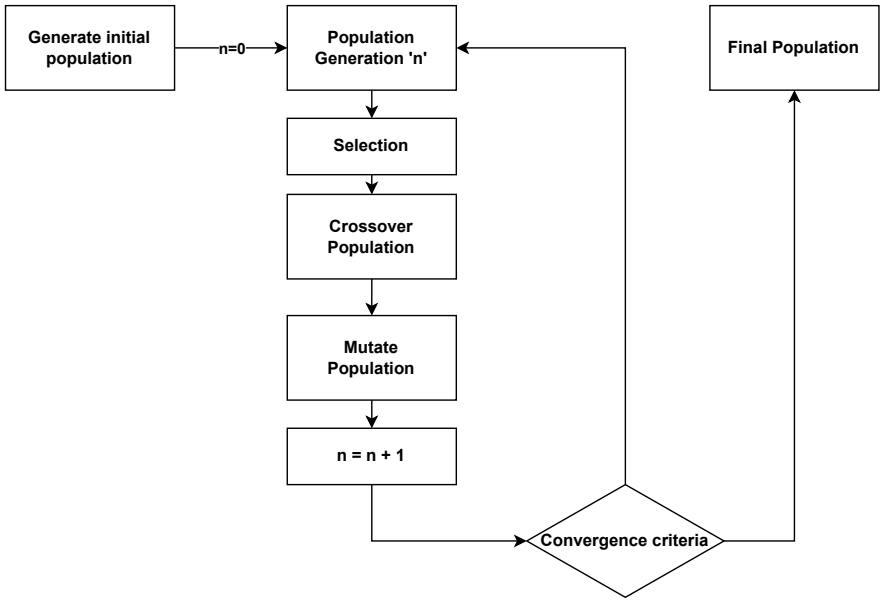


Figure 2.7: Flowchart explaining how a GA works in pseudo-code

3 Methodology

3.1 Experimental set-up

The research problem addressed in this paper is: ”the construction of RFs that produce an SOBP that matches a desired dose distribution using an optimizer”. For this an experimental setup was devised, the experimental setup is illustrated schematically in figure 3.1.

Here the beam source, ridge filter (RF), and the scorer (water box) are shown. All these objects are centred around the z-axis, and placed in an environment filled with air. The distance between the beam source and RF, D_1 in the schematic (equivalent to $|z_1 - z_2|$), is 25 cm. The distance between the RF and the water box, D_2 (equivalent to $|z_3 - z_4|$), is 25 cm too. The height of the RF ($|z_3 - z_2|$) is 35 mm. The depth, width, and height of the scoring volume ($|z_5 - z_4|$) are 20 cm. The material the scoring volume consists of is water as the example RF was tested in a water box.

When the simulation starts the proton source sends a proton beam of 150 MeV along the positive z-axis. This beam will then hit the RF and scatter into a spread-out Bragg peak (SOBP). This SOBP will then be registered in the scorer.

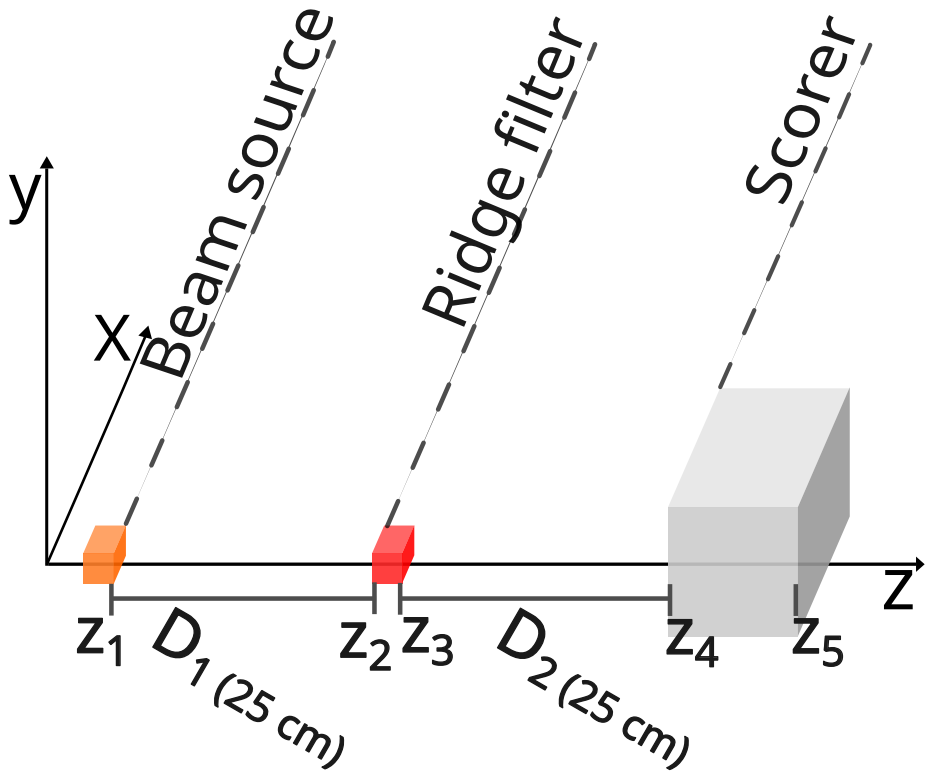


Figure 3.1: The beam source positioned at z_1, the RF positioned at z_3, and the scorer, as water box, positioned at z_5. The distance between the proton source and RF is denoted by d_1, 25 cm, and the distance between the RF and the scorer by d_2, 25 cm. The distance between z_2 and z_3 is the height of the RF, 35 mm, and the distance between z_4 and z_5 is the depth of the water box, 20 cm.

3.1.1 Ridge filter

A schematic of the ridge filter is illustrated in figure 3.2. In this figure, the top and front views of a ridge filter consisting of a 2 by 2 grid of ridge pins are shown. Also, the central axis, the ridge pin height, the baseplate height, the RF width, and the RF depth are indicated. As can be seen in the figure each ridge pin is essentially a number of rectangular prisms with uniform heights stacked on top of each other and centred around a central axis. In the simulation setup the RF, as illustrated in figure 3.2, is rotated by 90 degrees. The top of the RF then corresponds to z_2 and the bottom of the RF to z_3 in figure 3.1. As explained in section 2.1.2, the different thicknesses of the material cause different weights of mono-energetic beams thus generating an SOBP.

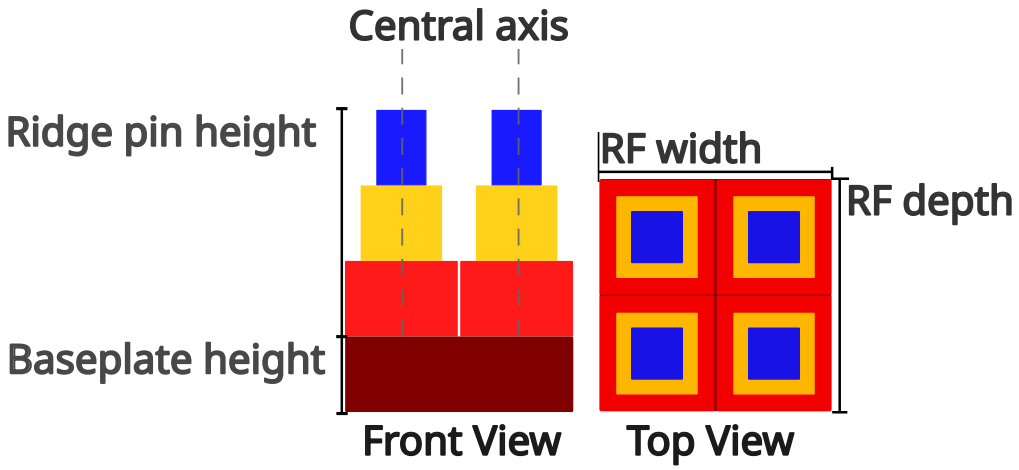


Figure 3.2: Front and top view of a simplified ridge filter design with a 2 by 2 grid of ridge pins, with the central axis shown in the front view. With values corresponding to the experimental setup listed, ridge pin height, baseplate height, RF width, and RF depth.

To reduce the scope of this research a couple of parameters are fixed, these parameters are illustrated in figure 3.3. Here the fixed ridge pin measurements are shown in a schematic. The fixed base plate height of 5 mm and the ridge pin height of 30 mm are shown. Furthermore, the maximum width that a ridge pin layer can have, 6 mm, and the pitch-to-pitch distance fixed at 6 mm are indicated.

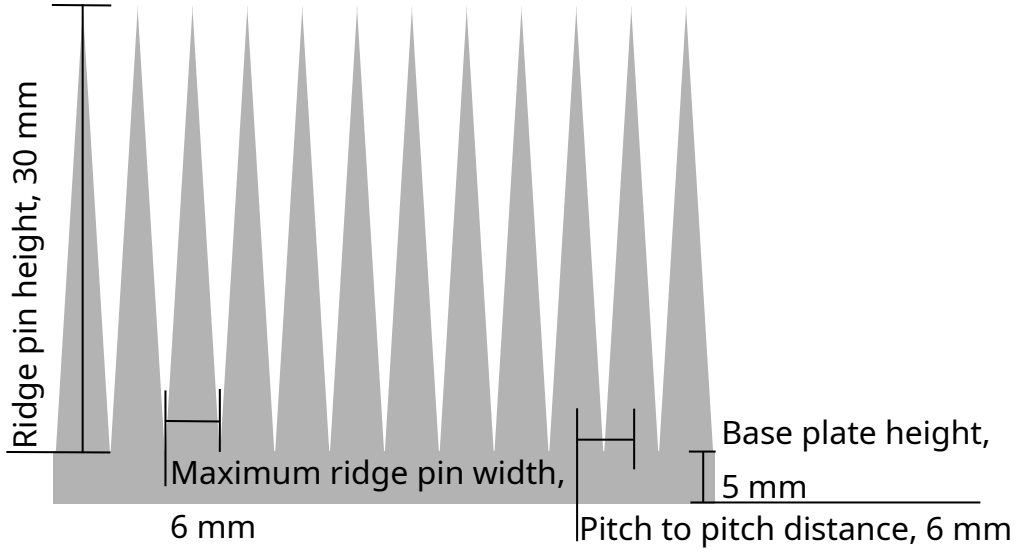


Figure 3.3: Schematic of an RF with the ridge pin height (30 mm), maximum ridge pin width (6 mm), base plate height (5 mm), and pitch-to-pitch distance (6 mm) between the ridge pins

Fixed parameters chosen for individual ridge pins are illustrated in figure 3.4. Here the ridge pin height is 30 mm and the maximum ridge pin width of 6 mm is identical to those in figure 3.3. However, to clearly illustrate that it is a maximum ridge pin layer width, the base layer is smaller than the maximum allowed width. Alongside that, a fixed step height is introduced, this step height is calculated by dividing the ridge pin height by the number of layers desired. For this project, the amount of layers chosen is 30, resulting in a step height of 1 mm. Another constraint placed upon the ridge pin structure is that the width of the layers must be monotone decreasing with respect to the z axis, which means that each layer must be the same size or smaller than the layer below it in the z axis. This is done so that there are no air gaps in the ridge filter and to make the ridge pin producible.

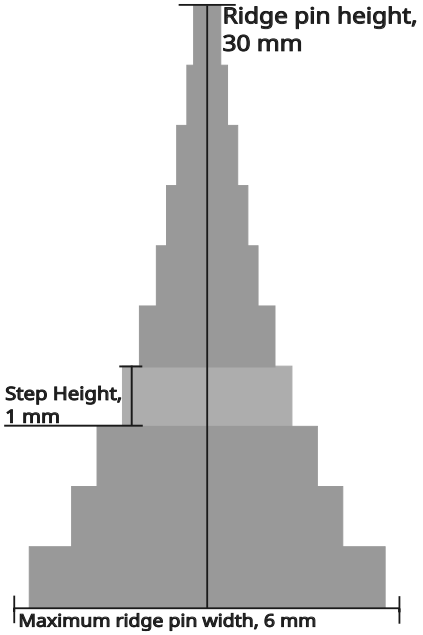
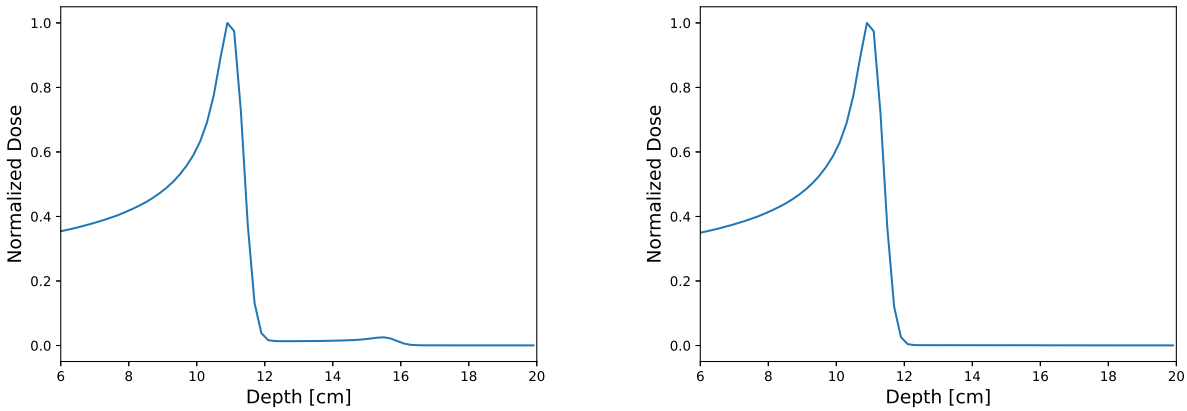


Figure 3.4: Ridge pin schematic with the ridge pin height of 30 mm, step height of 1 mm, and the maximum ridge pin width of 6 mm shown. Number of steps not to scale.

This leaves the following variables free: number of ridge pins in the x direction, number of ridge pins in the y direction, RF width, RF depth, and the ridge pin layer width depending on the ridge pin height. Here the number of ridge pins in the x and y direction correlate with the RF width and depth.

When doing trial runs and trying to achieve the characteristic Bragg peak shape, as described in section 2.1, and taking the fixed RF parameters a curious phenomenon happened. The resulting dose distribution in the scorer is illustrated in figure 3.5a. In this figure, a small bump appears at the trailing end of the dose distribution, the presumed cause is that the RF width and depth are too narrow and protons reach the water box by scattering around the RF. Using an iterative process it was found that if the RF has a minimum width and depth of 40 mm the bump at the trailing end disappears. It is presumed that an RF with a width and depth greater than 40 mm does not allow for scattering around the RF so all the protons that reach the scorer pass through the RF. The resulting depth dose distribution is illustrated in figure 3.5b. To make certain that scattering can not occur around the RF, 33 ridge pins are placed in the x and y axis, resulting in an RF with a width of 200 mm.



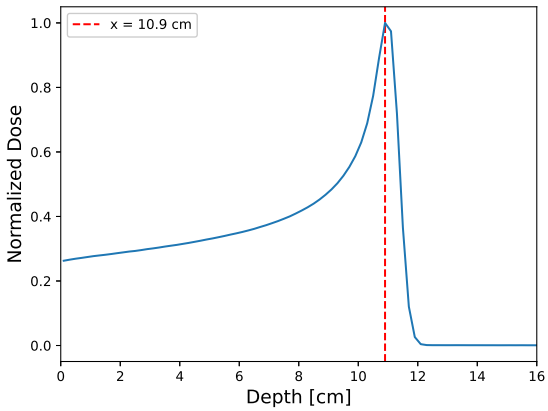
(a) Dose distribution for a ridge filter with a width and depth of 30 mm and height of 5 mm, here protons scatter around the RF and cause a small bump at the trailing end of the dose distribution.

(b) Simulation with a RF satisfying the minimum width and depth requirement of 40 mm, protons do not scatter around the RF so the trailing end of the dose distribution remains flat.

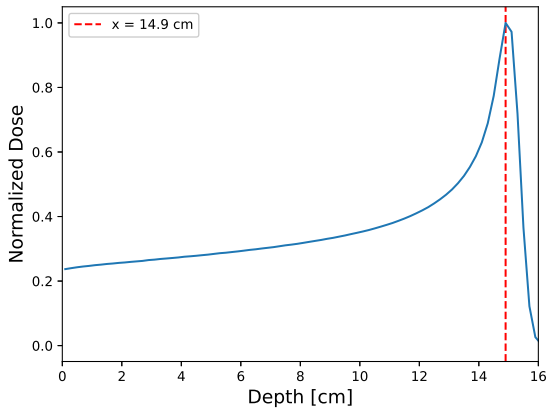
Figure 3.5: Impact of ridge filter width on protons scattering around the ridge filter.

Fixing the number of ridge pins in the x and y axis also fixes the RF depth and width. This leaves only one free variable the ridge pin layer width depending on the ridge pin height. This is the free variable that the optimizer uses. So the optimizer varies the ridge pin width based on height, with all ridge pins being identical, to achieve the desired SOBP.

From the choice of fixing the base plate height to 5 mm and the total RF height to 35 mm (30 mm of ridge pin height and 5 mm of base plate height) a region of interest (ROI) in the dose distribution provided by the score appears. The ROI appears when simulating the limit cases of the RF. The dose distributions belonging to the limit cases can be seen in figure 3.6. The distal- and proximal edges are defined as the location on the z axis of the scorer where the maximum dose value of the Bragg peak is achieved when simulating limit cases of the RF. In figure 3.6b the proximal edge is located at 10.9 cm while simulating only the base plate, in figure 3.6a the distal edge is located at 14.9 cm found by simulating a cube of total RF height and RF width and depth.



(a) Proximal edge of the region at 10.9 cm achieved when simulating an RF with height of 35 mm (fixed base thickness of 5 mm and fixed maximal height of 30 mm) and fixed RF width and depth of 200 mm.



(b) Distal edge of the region at 14.9 cm achieved when simulating an RF with only the chosen base thickness of 5 mm and fixed RF width and depth of 200 mm.

Figure 3.6: Boundaries of the region of interest when simulating limit cases.

3.2 Monte Carlo Simulation

Section 2.3 described the use of Monte Carlo simulations in particle transport experiments. While many different Monte Carlo simulations are available for particle transport, for this project, TOol for PArticle Simulation (TOPAS) is used [36]. Other popular options for simulating therapeutic high-energy photon and proton beams are PENH, FLUKA, and GEANT4. According to research by Baumann et al., the maximum error between the different programs is 0.5% [41], therefore this was not a deciding factor. TOPAS was chosen due to its open-source nature and that simulations do not require extensive knowledge of C++ compared to programs such as GEANT4, on which TOPAS is built.

In section 3.2.1, the settings used throughout the simulation are discussed, this is subdivided into the physics modules, the beam settings, and general simulation settings. After that in section 3.2.2, the implementation of the experimental setup into TOPAS is discussed.

3.2.1 Settings

3.2.1.1 Physics settings

Throughout TOPAS, the standard physics modules were used, listed in table 1.

Table 1: List of physics modules and their function in TOPAS [9]

Module name	Function
g4em-standard_opt4	GEANT4 standard physics option 4
g4h-phy_QGSP_BIC_HP	GEANT4 quark gluon string model binary cascade mode high precision neutron model
g4decay	GEANT4 decay physics
g4ion-binarycascade	GEANT4 binary cascade
g4h-elastic_HP	GEANT4 elastic physics high precision neutron
g4stopping	GEANT4 stopping physicss

3.2.1.2 Beam settings

The beam settings used in this paper are shown in table 2.

Table 2: Beam parameters and their values.

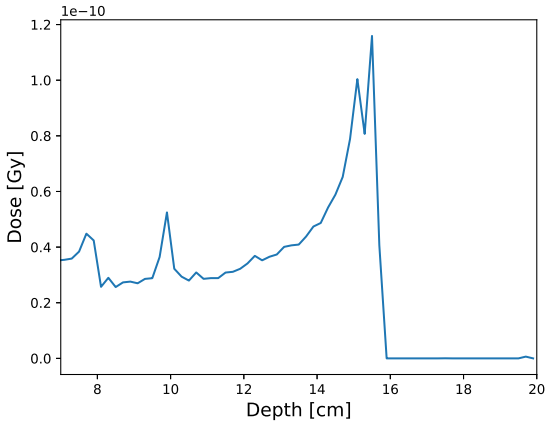
Parameter in TOPAS	Value
BeamEnergy	150 MeV
BeamEnergySpread	0.757504
BeamPositionDistribution	”Flat”
BeamPositionCutoffShape	”Ellipse”
BeamPositionCutoffX	3 cm
BeamPositionCutoffY	3 cm
BeamAngularDistribution	”Gaussian”
BeamAngularCutoffX	90°
BeamAngularCutoffY	90°
BeamAngularSpreadX	0.0032 rad
BeamAngularSpreadY	0.0032 rad

3.2.1.3 General settings

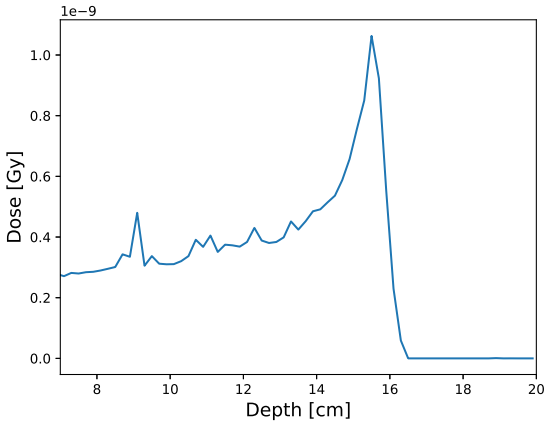
Another key parameter for simulating is what TOPAS calls: ”NumberOfHistoriesInRun”. To translate this it means the amount of particles used in the simulation. For this paper, each run in the simulation uses 10^5 protons as particles.

To determine this number, test runs were done in incrementing magnitudes from 10^1 to 10^6. First, the data was processed as described in section 3.5, except normalizing the data. Illustrated in figure 3.7, is the corresponding data, a Bragg peak appears very neatly when increasing the amount of particles. Note that in the top left of the individual images the scale of the dose changes, depending on the number of particles. Normalization of the IDDs removes this variable in the optimization, data processing is talked about in section 3.3.

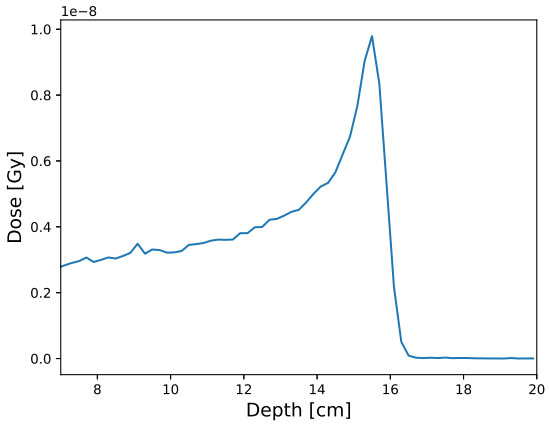
The number of particles (”NumberOfHistoriesInRun”) that is used throughout the remaining simulation is 10^5. There are no apparent differences between the simulation of 10^5 and 10^6 particles, as illustrated in figures 3.7e and 3.7f. So the reduction in simulation time is a deciding factor.



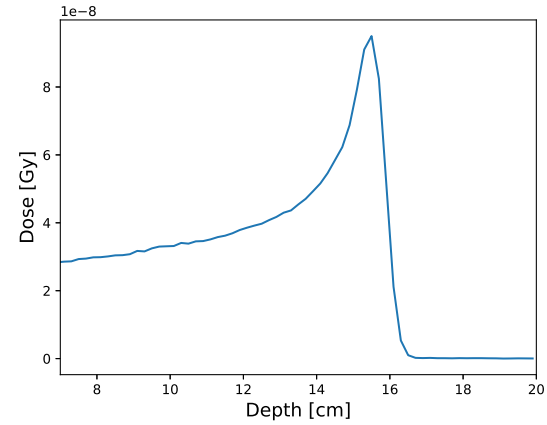
(a) ”NumberOfHistoriesInRun” equal to 10^1



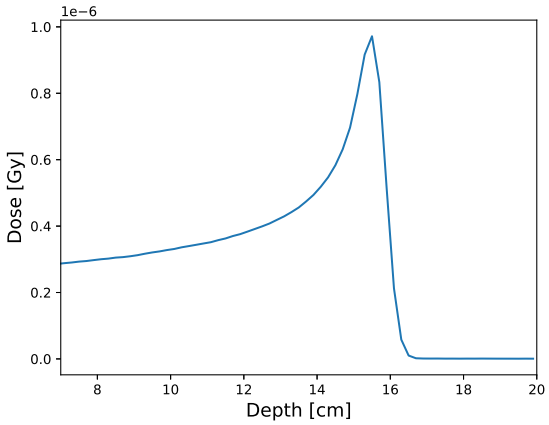
(b) ”NumberOfHistoriesInRun” equal to 10^2



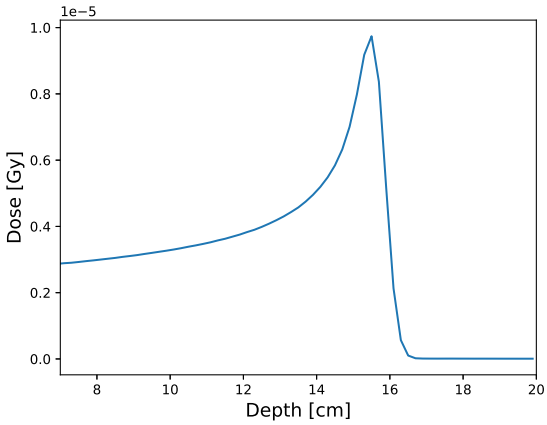
(c) ”NumberOfHistoriesInRun” equal to 10^3



(d) ”NumberOfHistoriesInRun” equal to 10^4



(e) ”NumberOfHistoriesInRun” equal to 10^5



(f) ”NumberOfHistoriesInRun” equal to 10^6

Figure 3.7: Simulation results with varying the number of particles (”NumberOfHistoriesInRun”)

3.2.2 Simulation setup

The experimental setup was introduced in section 3.1 and illustrated schematically in figure 3.1. The implementation of these features into TOPAS is discussed in this section.

3.2.2.1 Scorer

The type of scorer used in this experiment is called a volume scorer. For this experiment, a 'DoseToMedium' scorer included in TOPAS was used. This is a volume in the simulation over which you calculate the dose received in each volume region. For this paper the scorer takes the shape of a box with a width, height, and depth of 20 cm, the material it is made of is water.

The scorer can be further subdivided to provide more precise dose measurements in an area, these subdivisions are called bins, and they can be in the x, y, and z directions. For this experimental setup, the only axis of interest is along the z-axis. The bin settings, along the x and y axes, are set to 1 (1 is the minimal number of bins). Along the z axis, the number of bins is 100. With the scorer having a depth of 20 cm, this allows a precision of 2 mm in the z direction. A 2-dimensional schematic of the scorer along the z axis is illustrated in figure 3.8, the depth of the water box is as explained in section 3.1, with z_4 and z_5 referring to the same points in the setup. In the figure the water box is scored along the z-axis with 21 bins, in the actual simulation there are 100 bins.

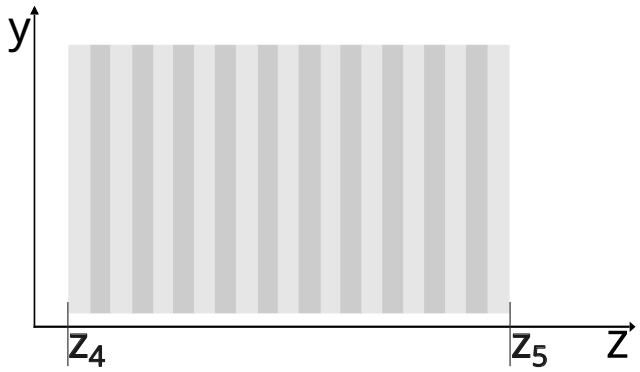


Figure 3.8: 2-Dimensional illustration of the water box scored across the z-axis, with 21 bins. The points z_4 and z_5 are the same as in figure 3.1, the depth of the water box is 20 cm

3.2.2.2 Scorer modifications

Modification to the scorer as discussed in section 3.2.2.1, consists of the introduction of an obstruction inserted halfway in the middle of the water box, illustrated in figure 3.9. The chosen material for the obstruction is bone. This was done as bone is a material that is often present in treatment volumes in the human body. The bone plate has a width of 2 bins (0.4 cm) and a height of 10 cm, this replaces the top half of the water in bins 50 and 51. The introduction of an obstruction halfway in the scoring volume across the y axis necessitates changing the number of bins in the y axis to 2.

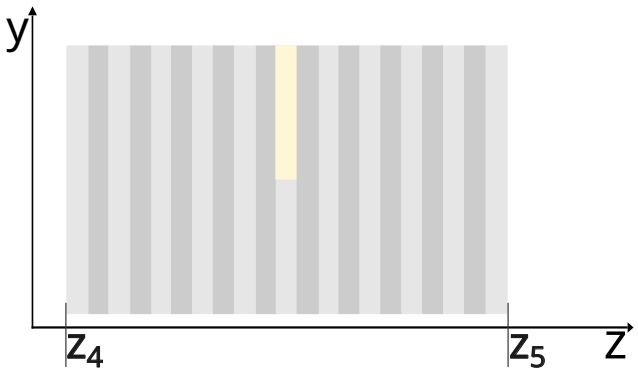


Figure 3.9: 2-Dimensional illustration of the water box scored across the z-axis, with 21 bins, actual simulation uses 100 bins. In the middle of the water box a piece of "bone" is inserted halfway along the y axis into the water box. The points z_4 and z_5 are the same as in figure 3.1, depth of the scorer is 20 cm.

Bone is not part of the standard GEANT4 material list, TOPAS uses this list for materials, therefore the material must be added manually. The make-up of the bone was taken from Adult Reference Computational Phantoms published in 2009 [42], more specifically from ICRP110PhantomMaterial_Male. The composition of bone as used in the simulation can be found in table 3. Aside from the bone composition, the following parameters were used: a density of $1.920 \frac{g}{cm^3}$ and a mean excitation energy of 106.40 eV were used.

Table 3: Elemental composition of bone.

Element	**Fraction of element per volume**
Hydrogen	0.036
Carbon	0.159
Nitrogen	0.042
Oxygen	0.448
Sodium	0.003
Magnesium	0.002
Phosphorus	0.094
Sulfur	0.003
Calcium	0.213

3.2.2.3 Ridge filter

Ridge filters (RFs), as described in this paper in section 3.1.1, are not a part of standard geometry components available in TOPAS. Therefore implementing RFs into TOPAS needs additional work. Illustrated in figure 3.10, is the top and front view of a RF consisting of a 2 by 2 grid of ridge pins, the central axis, ridge pin height (30 mm), baseplate height (5 mm), minimum RF width and depth (40 mm).

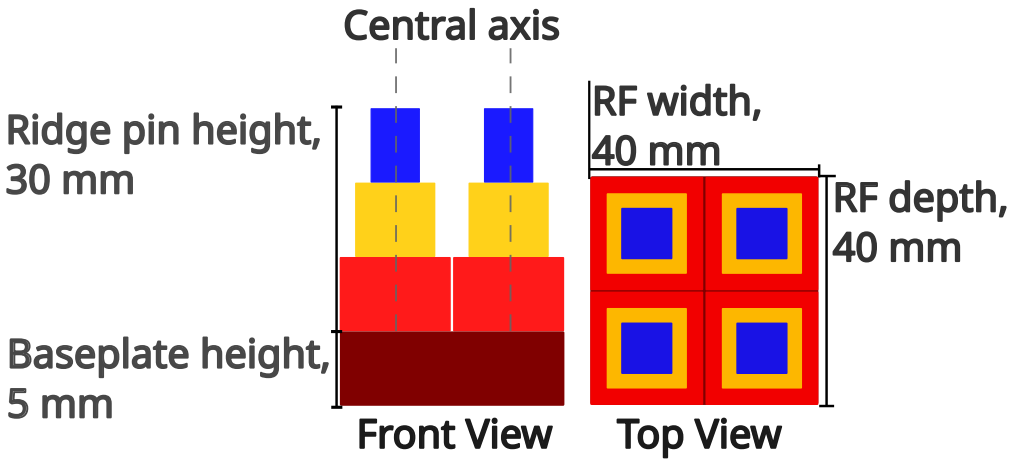


Figure 3.10: Front and top view of a simplified ridge filter design with a 2 by 2 grid of ridge pins, with the central axis shown in the front view. With values corresponding to the experimental setup listed, ridge pin height (30 mm), baseplate height (5 mm), RF minimum width (40 mm), and RF minimum depth (40 mm).

Because the geometry of ridge pins consists of stacked rectangular prism it was first tried to build this in TOPAS using instruction files. However, this approach was not feasible as it resulted in long computation times. Partly, due to this approach generating approximately 500.000 lines of instruction code. Moreover, these all needed to be collision checked during start-up and due to floating point errors these inevitably happened.

Therefore, a geometry extension for TOPAS was written, in C++. The geometry extension is drastically faster because you only need the text file to give the widths of the ridge pin layers and the number of layers. After that, the extension takes care of generating the ridge pin geometry and the RF. Due to the geometry being generated in TOPAS instead of in the instruction file, no overlap checks are needed. Also, the reduction of the instruction file by about 500.000 lines of code reduces parsing times. This all speeds up simulation times significantly.

Using the instruction file the average computation time for a single solution in a generation was upward of 8 minutes, using the extension reduces this to $\sim$ 40 seconds when the remaining parameters remained unchanged. This is mainly due to significant reductions in the pre-processing step time.

3.2.3 Beam source

The beam source sends a proton beam at a mono-energetic value of 150 MeV. This value was chosen as the example RF was generated for a proton beam with an energy level of 150 MeV. The ”BeamPositionDistribution” was set to ”flat” to simplify the simulation. The remaining parameters are from the example file RidgeFilter.txt[1] provided in TOPAS [43].

[1] https://topas.readthedocs.io/en/3.7.0/examples-docs/SpecialComponents/RidgeFilter.html

3.3 Data processing

Before the output data of TOPAS can be used the data needs to be processed. The chosen output format for the experimental data from TOPAS is a '.csv' file. This file contains the coordinates of the bins in the scoring volume and the sum of the dose in each bin.

Using the recommended package, topas2numpy, from the TOPAS documentation, the '.csv' files are imported into Python. For the base case, the scoring volume is only scored in the depth direction, this is illustrated in figure 3.8. In the data from TOPAS the location of x and y bins are not needed as there is only one bin in those directions. Thus the location data of the bins in the x and y directions is removed this is done by squeezing the data. This removes the height and width dimensions from the array. Now only the information on the depth and the corresponding dosis per bin remains. This data is used to generate integral depth dose (IDD) plots, the integral depth dose corresponds to the dose distribution.

In the case with an obstruction, the number of bins in the y axis is increased. As the obstruction goes exactly halfway, the number of bins in the y axis becomes two. In this case, the y axis does provide valuable information so it can not simply be discarded. However, the final data that is of interest is the influence of the obstruction on the dosis in the depth direction. For this, the two bins in the y direction can be summed, as this does not throw away any necessary information. After this, the data format is equivalent to that in the base case, and the steps as written above can be repeated.

Now that the data has been reduced to an integral depth dose the data is normalized, this is done to remove the absolute dose as a variable. The normalization preserves the shape of the IDD. The normalization is achieved by dividing by the maximum value of the IDD. This preserves the shape of the distribution and discards the dose by setting the maximum value of the simulated values to 1.

Furthermore, for each simulation per generation and solution, the following information is stored in a JSON file: generation number, solution number in the generation, solution values, fitness value of the simulation, time of simulation, and '.csv' file location.

3.4 Genetic Algorithm

The implementation of the genetic algorithm (GA) in this paper was done in Python, for this, the library PyGAD was used [44]. PyGAD offers a broad area of support while allowing great user flexibility in the mutator and crossover functions. This was needed as the constraints in place during this experiment do not allow standard mutator and crossover functions to be used. This is because none of the standard mutator- and crossover functions allow for monotonicity to be preserved.

3.4.1 Genetic Algorithm settings

Key parameters of a GA include the number of genes and population size. The size of the population is important because it reduces the chance you miss a globally optimal solution, due to being stuck in a locally optimal solution. Moreover, the number of genes is significant as it allows for a greater genetic diversity to be present. For this paper, the number of genes was chosen to be 30 as this would result in a step size of 1 mm. For the scope of this research, this was deemed enough.

According to Vrajitoru in 1999 [45], it is preferable to have a larger population, compared to the number of generations. However, they provide no definitive ratio. Further research on the topic has not been conclusive on how much bigger the population should be compared to the number of genes. But it seems to be somewhere between $2 - 3$, the number of genes has been chosen to be 30 which gives rise to a population size of 80.

3.4.2 Initial population

Due to the constraints, listed in section 3.1.1, an initial population needs to be created that complies with the monotonicity constraint. In section 3.4.1, a population of 80, each having 30 genes was established. Call the number of genes, $n = 30$, and the chosen size of the population, $m = 80$.

Then for the initial population, a set is generated with m subsets, and each of these m subsets contains n items, for each set it holds: $\forall i, j \in [0, n]$ with $i < j$ then $m_i \leq m_j$ and $0 \leq m_i \leq 6$ mm. To clarify it means that there are m lists with n items and all the items are sorted so that the biggest number is at the beginning and the smallest number is at the end. For the problem, this results that the initial population will be of size 80, each containing 30 values sorted in monotone increasing order.

3.4.3 Fitness function

The ROI has a proximal edge at 10.9 cm and a distal edge at 14.9 cm, as discussed in section 3.1.1. To evaluate the effect of a change in the width of a ridge pin layer on the SOBP, a fitness function is used on this region. The fitness function returns a fitness value that PyGAD can use. PyGAD is a minimizer so the goal of the fitness function is to get as close to 0 as possible. To start, the data is processed according to section 3.3. The processing returns only the shape of the SOBP, which can be compared to the shape of the desired dose distribution in the region of interest.

The shape of the desired dose distribution in this experimental setup is as flat as possible. This corresponds to returning values near 1 after normalizing. So the fitness function needs to return 0 as fitness value when this is achieved. Therefore the fitness function pointwise compares the desired

dose distribution and the dose distribution from TOPAS. For the remainder of this explanation, the desired dose distribution is that of the base case, however, the procedure holds for other desired dose distributions as well.

To compare the simulated results and the desired results two different error functions were tried out.

$$RMSE = \sqrt{\frac{\sum_{i=0}^{n}(y_i - \overline{y_i})^2}{n}} \tag{1}$$

The first potential error is the root mean squared error (RMSE), defined in equation 1. In this equation, n is the number of values in the region of interest, $y_1, y_2, ..., y_n$ the simulated dose values and $\overline{y_1}, \overline{y_2}, ..., \overline{y_n}$ the desired dose values.

$$RMAE = \sqrt{\max_{i \in [0,...,n]} |y_i - \overline{y_i}|} \tag{2}$$

The second contender is the root maximum absolute error (RMAE), defined in equation 2. This function takes the square root of the maximum absolute error, with $[0, ..., n]$ corresponding to the number of values in the ROI, and as described for equation 1, $y_1, y_2, ..., y_n$ the simulated dose values and $\overline{y_1}, \overline{y_2}, ..., \overline{y_n}$ the desired dose values. The norms of these two error functions are equivalent, the RMAE corresponds to an l_∞ norm and the RMSE to a $\frac{l_2}{\sqrt{n}}$ norm.

$$\alpha\| \cdot \|_p \leq \| \cdot \|_q \leq \beta\| \cdot \|_p, \text{ for } p, q, \alpha, \beta, \in \mathbb{R} \tag{3}$$

Two norms are equivalent when equation 3 is satisfied.

$$\alpha\|x\|_\infty \leq \frac{\|x\|_2}{\sqrt{n}} \leq \beta\|x\|_\infty, \text{ for } \alpha, \beta, \in \mathbb{R} \text{ and } x \in \mathbb{R}^n \tag{4}$$

Filling in the $\frac{l_2}{\sqrt{n}}$ and the l_∞ norm, gives equation 4. With $\|x\|_2 = \sqrt{\sum_i x_i^2}$ and $\|x\|_\infty = \max_j |x_j|$.

$$\alpha\|x\|_\infty = \alpha \max_j |x_j| \leq \alpha\sqrt{\sum_i x_i^2} = \alpha\|x\|_2 = \frac{1}{\sqrt{n}}\|x\|_2 = \|x\|_{RMSE} \tag{5}$$

The left-hand side holds for $\alpha = \frac{1}{\sqrt{n}}$ as seen in equation 5.

$$\|x\|_{RMSE} = \frac{\|x\|_2}{\sqrt{n}} = \frac{\sqrt{\sum_i x_i^2}}{\sqrt{n}} \leq \frac{\sqrt{\sum_i \|x\|_\infty^2}}{\sqrt{n}} = \frac{\sqrt{n\|x\|_\infty^2}}{\sqrt{n}} = \frac{\sqrt{n}\|x\|_\infty}{\sqrt{n}} = \|x\|_\infty \tag{6}$$

The right-hand side follows from equation 6, with $\beta = \sqrt{n}$.

Now that both the left and right-hand sides of equation 3 hold, the norms of both error functions are equivalent. So in the infinite case, both norms would go to 0. However, a decision between the error functions must still be made. Illustrated in figure 3.11, is the average convergence per generation containing a population of 40 for the first 10 generations. The results indicate that RMSE consistently achieves lower fitness values than RMAE. PyGAD is a minimizer so fitness values closer to 0 are desirable, hence in the remainder of the simulations RMSE is used. Furthermore,

RMSE is more sensitive to outliers than RMAE. As outliers can have severe consequences when performing radiation treatment, such as not fully irradiating a tumour thus reducing treatment effectivity RMSE was chosen. Giving the objective function as in equation 1 with the region of interest between 10.9 cm and 14.9 cm.

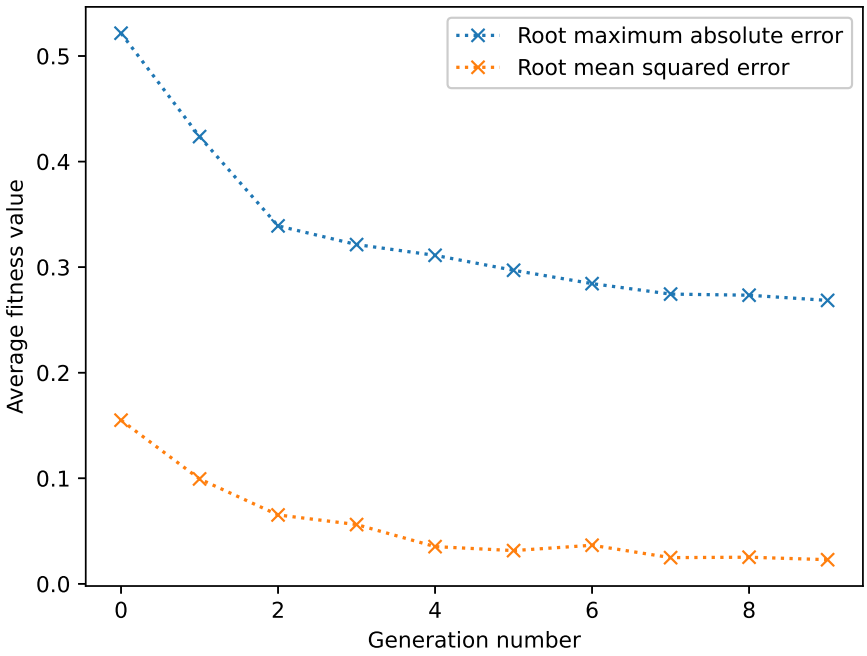


Figure 3.11: Average convergence of a generation containing a population of 40 for the first 10 generations

With the fitness function established, a simplified case can be examined. Figure 3.12 presents a simplified output from TOPAS. The region of interest boundaries are plotted in red, with PE standing for the proximal edge and DE for the distal edge. The blue lines represent values obtained from TOPAS and the desired dose is plotted by the green lines.

$$\sqrt{\frac{(0.8-1)^2+(0.9-1)^2+(0.7-1)^2}{3}} = 0.216 \tag{7}$$

Using the fitness function, as described in equation 1, and filling in the values from figure 3.12 results in equation 7 this gives a fitness value of 0.216.

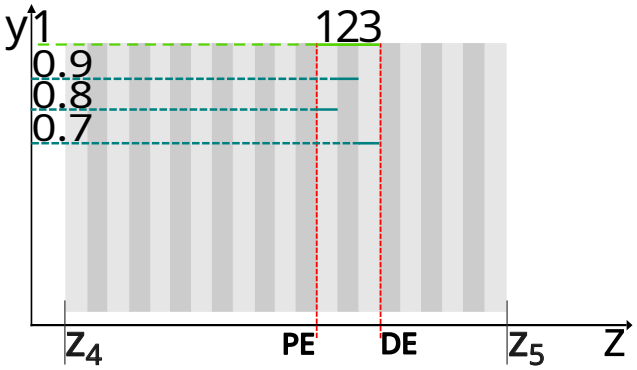


Figure 3.12: Visualized example of TOPAS output with 3 bins in the region of interest (ROI) between the proximal and distal edge. The desired output in the ROI is $\overline{y_1} = \overline{y_2} = \overline{y_3} = 1$, the example output is $y_1 = 0.8$, $y_2 = 0.9$, and $y_3 = 0.7$

3.4.4 Parent Selection

In this experiment, the parent selection method steady state selection (SSS) is used. Every generation SSS uses a number of 'good' solutions, ranked by the fitness value returned from the fitness function, defined in section 3.4.3. This number has a corresponding parameter that PyGAD calls 'keep_elitism', for this paper 'keep_elitsm' was set to 10. So SSS uses the 10 best solutions to create new offspring, replacing the 10 worst solutions. This method was chosen due to it preserving diversity, thus giving solutions time to propagate.

3.4.5 Crossover function

Due to the monotonicity restriction on the genes, explained in section 3.1.1, the standard crossover functions available in PyGAD can not be used. The custom crossover function uses the 10 parents provided as described in section 3.4.4. Then it creates five pairs of two parents from the 10 available. Following this, the crossover function selects a random crossover point for each of the pairs. After the crossover point, the remaining genes in the parents get swapped. Then the function sorts the two resulting genes to preserve the monotonicity.

3.4.6 Mutation function

Due to the monotonicity restriction, discussed in section 3.1.1, the standard mutation functions available in PyGAD can not be used. Therefore a custom mutation function has been written. The custom mutator works as follows, it randomly chooses 10% of the population to mutate (giving 8 solutions to mutate as the total population is 80). This was done by setting the PyGAD parameter 'mutation_percent_genes' to 10. Then for each selected parent one gene is randomly selected to mutate. The mutation introduces the following change, the gene is swapped with a random value within a specified lower- and upper bound. For the first gene in the parent, it randomly selects a value between 3 and that of the gene above it, for the last gene in the parent it randomly selects a value between the gene below it and 0, for all the genes in between it takes a random value between the value of the gene above and below it.

3.4.7 Convergence criteria

The convergence criteria used in this paper is 'saturate' which is a standard stopping criterion available in PyGAD. The 'saturate' stopping criterion works as follows, the GA stops after the best fitness value available does not change for x amount of generations.

The reason this experiment uses the "saturate" method is that it allows more time for latent genes to diversify and propagate. If the best fitness value available does not change for 7 generations, the GA will stop.

3.5 Code

All the code for processing the data, the GA, the TOPAS extension, and the instruction files for TOPAS are located on an open repository on Github[2], figure 3.13 contains a QR code to the GitHub repository.

Figure 3.13: QR Code to the GitHub repository containing the code.

[2]https://github.com/Dux02/BEP

4 Results and Discussion

4.1 Results

4.1.1 Base case

This simulation was done according to the setup explained in section 3.1. The convergence of the ridge pin design is illustrated in figure 4.5. The initial spread of points that the ridge pin used in the algorithm is shown in navy. Then the change in colour from navy to yellow indicates a convergence to a ridge pin shape optimized by the GA. The SOBP that corresponds to the RF generated by this ridge pin shape is illustrated in figure 4.2, this SOBP returned a fitness value of 0.0109. Illustrated in figure 4.3, is the average fitness value per generation plotted against the generation number. Furthermore, the plot has the best fitness value found, and the simulation stops after seven generations if a better fitness value has not been found illustrated. The total time taken for the optimization was 1.5 day. Further relevant values are present in table 4.

Table 4: Simulation results for optimized ridge filter

Variable	**Value**
Proximal edge	10.9 cm
Distal edge	14.9 cm
Simulation time	~ 1.5 days
Fitness value	0.0109
Maximum dose (ROI)	100%
Minimum dose (ROI)	97.3350%
Mean dose (ROI)	99.2072%
Maximum dose difference (ROI)	2.66495%

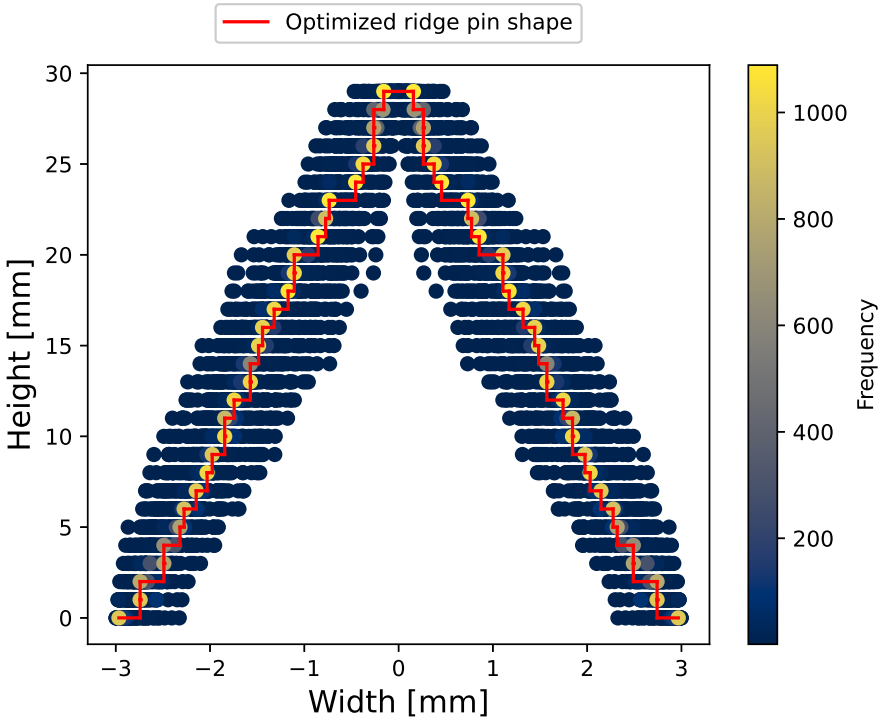


Figure 4.1: Heatmap of ridge pin coordinates, with navy denoting a low frequency, and yellow denoting a high frequency. Illustrates the convergence to the final optimal ridge pin design for the ridge filter, this is plotted in red.

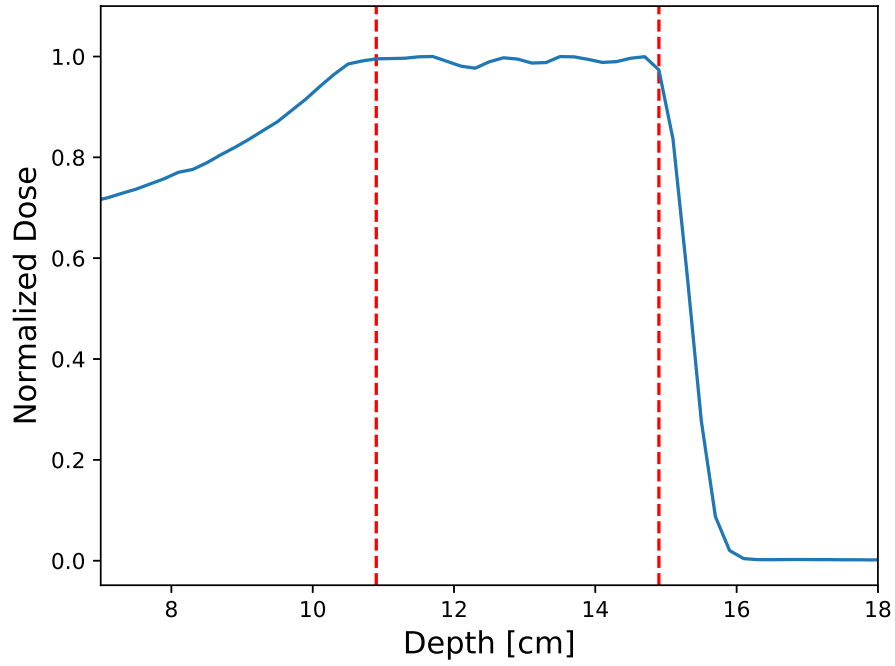


Figure 4.2: The IDD of a SOBP generated by the optimized ridge pin, as illustrated in figure 4.5. With the distal edge at 14.9 cm and the proximal edge at 10.9 cm, and a fitness value of 0.0109.

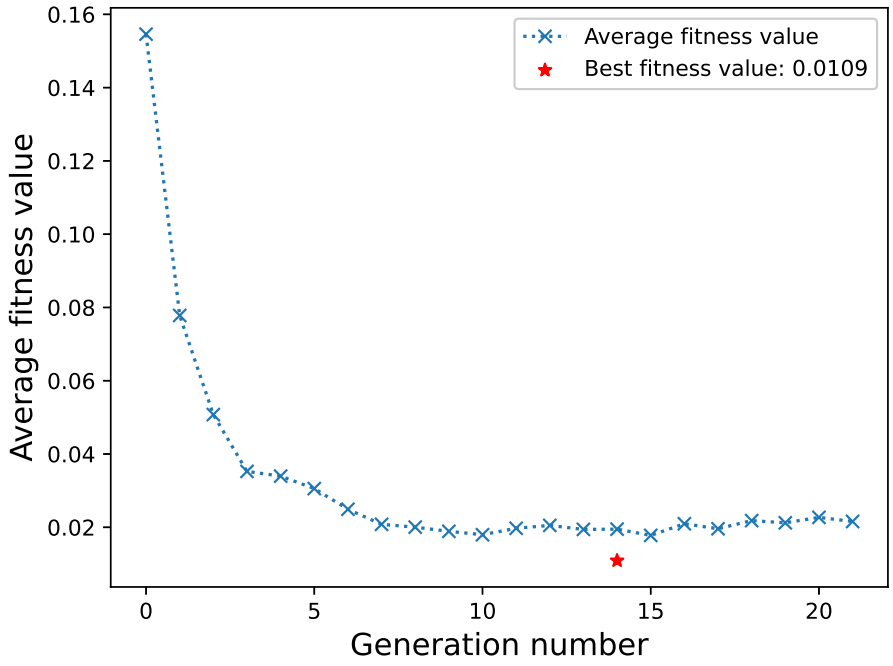


Figure 4.3: Average fitness value plotted against generation in blue, with the best fitness value plotted in red

4.1.2 Benchmark case

The fitness function was adjusted to work over a region of interest with a proximal edge of 13.1 cm and a distal edge of 15.1 cm. Pertinent values are documented in table 5. The resulting SOBP is illustrated in figure 4.4.

Table 5: Simulation results for benchmark ridge filter

Variable	Value
Proximal edge	13.1 cm
Distal edge	15.1 cm
Simulation time	Unknown
Fitness value	0.011045
Maximum dose (ROI)	100%
Minimum dose (ROI)	97.5876%
Mean dose (ROI)	99.1463%
Maximum dose difference (ROI)	2.41236%

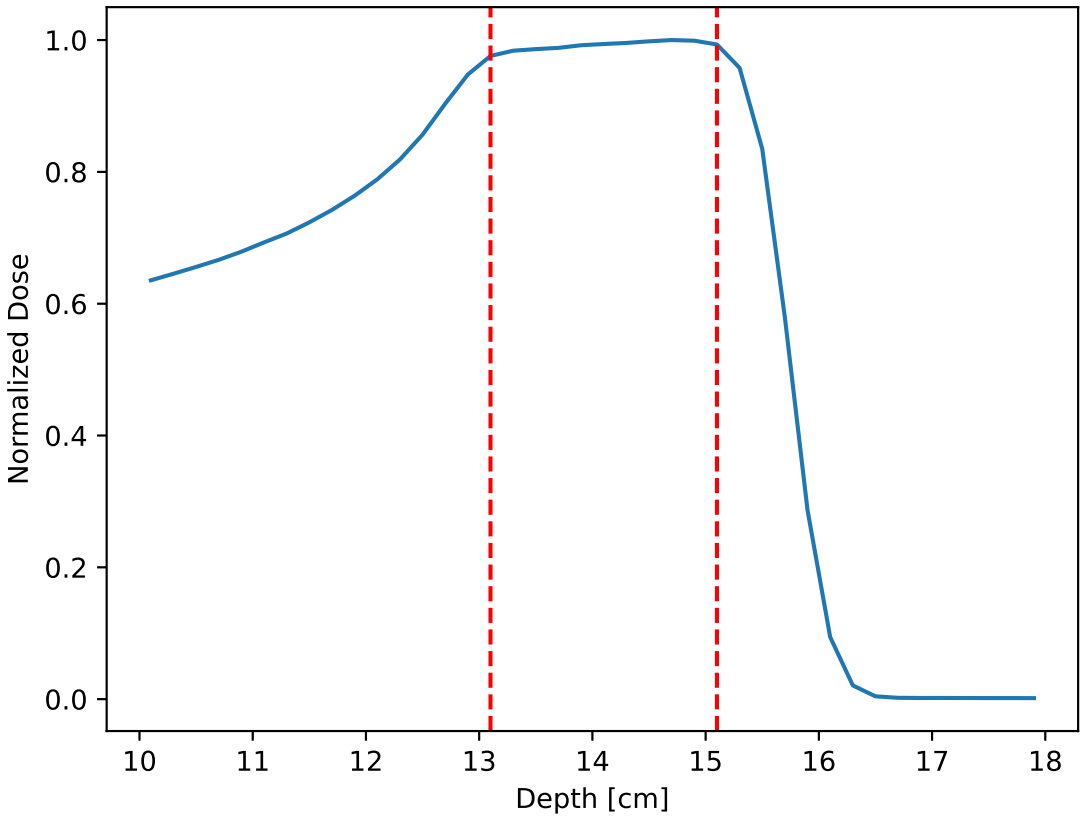


Figure 4.4: IDD showing the SOBP received from simulating using anonymously provided RF, cropped to show interval between 10 cm and 18 cm. Proximal edge at 13.1 cm and distal edge at 15.1 cm

4.1.3 Obstructed Scorer

For this simulation, the simulation set-up was as described in section 3.1, with the addition of a bone plate in the scoring volume as described in section 3.2.2.2. Relevant numbers are located in table 6.

Table 6: Simulation results for optimized ridge filter with a bone plate inserted halfway into the scoring volume.

Variable	**Value**
Proximal edge	10.9 cm
Distal edge	14.9 cm
Simulation time	$\sim$ 1 days
Fitness value	0.238185
Maximum dose (ROI)	100%
Minimum dose (ROI)	54.5468%
Mean dose (ROI)	83.40838%,
Maximum dose difference (ROI)	45.4532%

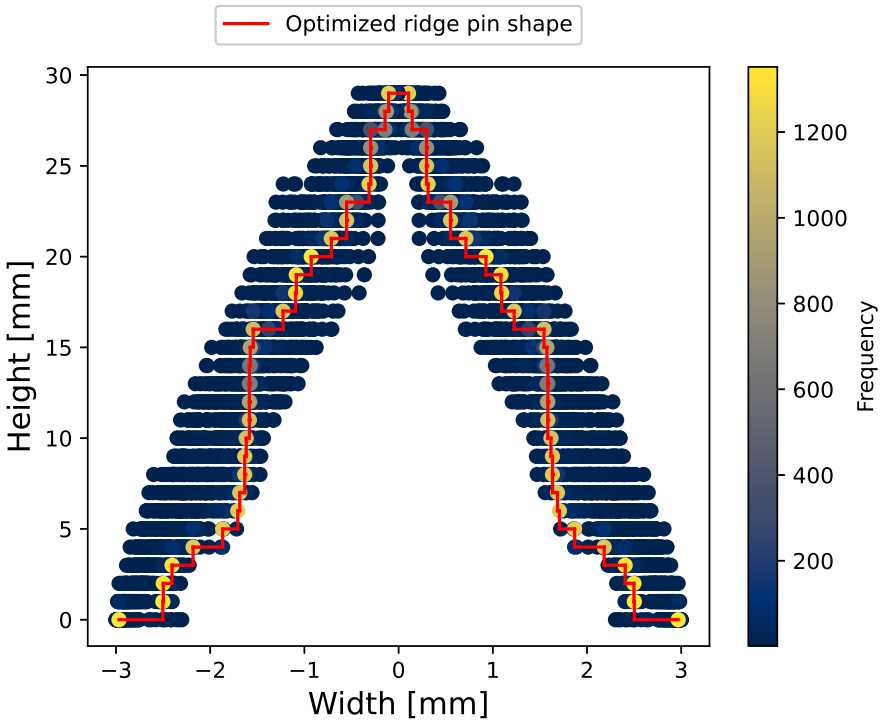


Figure 4.5: Heatmap of ridge pin coordinates, with navy denoting a low frequency, and yellow denoting a high frequency. Illustrates the convergence to the final optimal ridge pin design for the ridge filter, this is plotted in red.

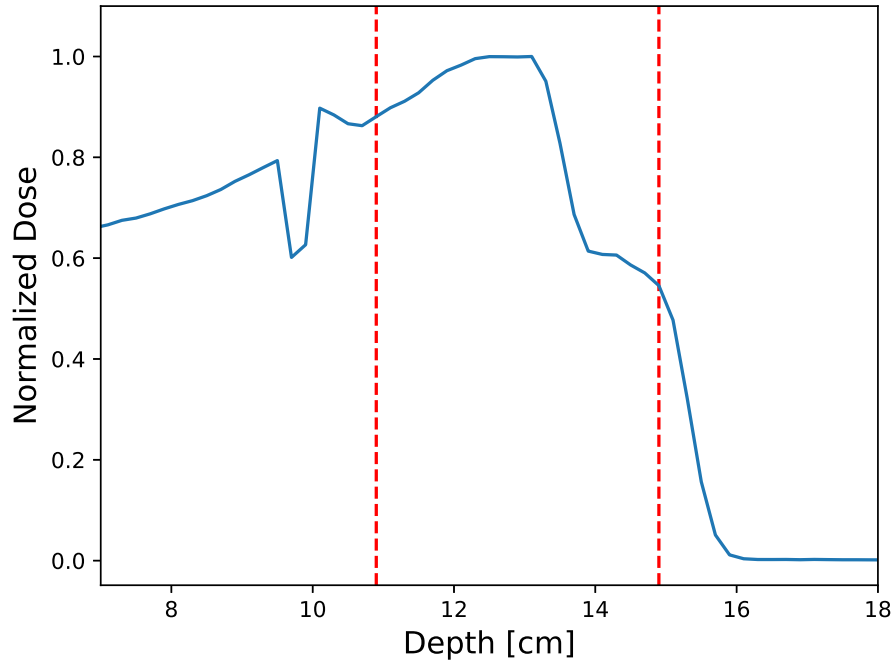


Figure 4.6: The IDD of a SOBP generated by the optimized ridge pin, as illustrated in figure 4.5. With the distal edge at 14.9 cm and the proximal edge at 10.9 cm, and a fitness value of 0.23819.

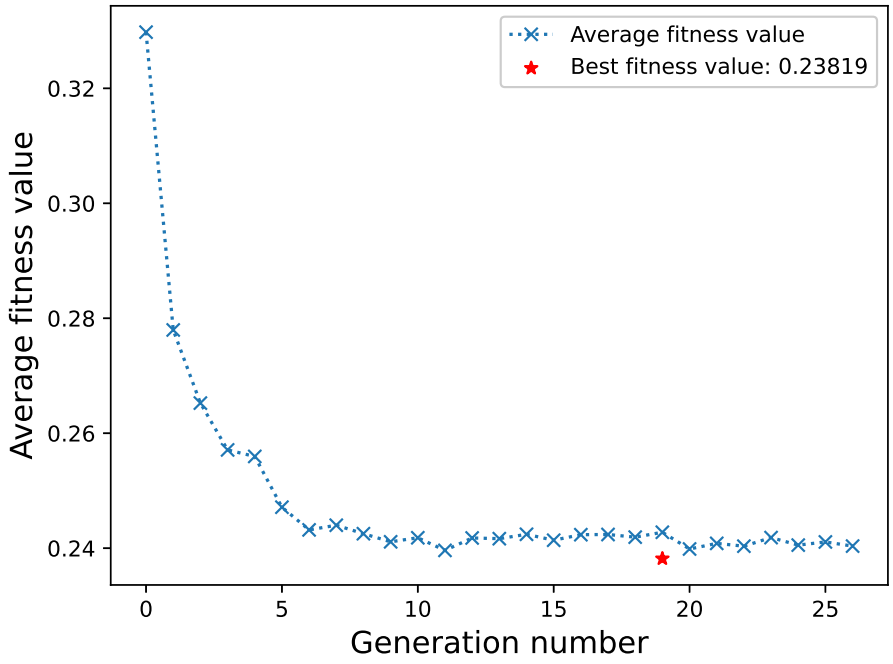


Figure 4.7: Average fitness value plotted against generation in blue, with the best fitness value plotted in red

4.2 Discussion

The research questions posed in this paper were: ”Can patient-specific ridge filters for proton therapy be constructed based on desired dosage distributions through the use of genetic algorithms? And how does this approach compare to previously established research in terms of achieving target dosage distributions?” To address these questions, a base case was established to validate the method, which was then applied to a more complex situation.

In subsection 4.1.1, the optimized ridge pin design achieved a fitness value of 0.0109. In comparison, in section 4.1.2, the benchmark case results in a fitness value of 0.0110. These results support the claim that the combination of a Monte Carlo simulation and a GA is capable of achieving results comparable to those of previous studies. However, the optimized RF design does not achieve as flat of an SOBP in the ROI. This limitation is likely due to some constraints in the fitness function. Despite this, the maximum dose difference over the region of interest differs minimally, $\sim 2.66\%$ (optimized RF) vs. $\sim 2.41\%$ (benchmark RF). Furthermore, the mean dose is $\sim 99.2\%$ (optimized RF) vs. $\sim 99.1\%$ (benchmark RF). The clinical implications of these differences remain unknown due to the limited scope of this paper.

When a bone plate was introduced into the scoring volume, the effectiveness of the optimization significantly decreased. The fitness value rose to 0.2381, an order of magnitude greater than that of the base case. Furthermore, the ROI exhibits a maximum dose difference of $\sim 45.5\%$ and a mean dose of $\sim 83.4\%$. These results suggest that the results of the included obstructions were limited by the convergence criterion and population size. Increasing both parameters could allow for more genetic diversity and latent genes to propagate, potentially improving results.

Overall, the results of this study demonstrate that a GA can generate patient-specific ridge filters effectively for simple cases. However, for more complex situations, additional research is needed for relevant parameters of the genetic algorithm.

5 Conclusions and Future Work

This research explored the use of a genetic algorithm (GA) and a Monte Carlo simulation for optimizing patient-specific ridge filters (RFs) for proton therapy. The results indicate that the optimized ridge filter achieves comparable results to the benchmark case. The optimized RF design for an SOBP of 4 cm has a maximal dose difference of $\sim 2.7\%$ and a normalized mean dose of $\sim 99.2\%$ in the region of interest (ROI). The benchmark ridge filter achieved $\sim 2.4\%$ and $\sim 99.1\%$ respectively. These findings validate the feasibility of using GAs for RF optimization.

However, significant challenges arose when obstructions were introduced into the scoring volume. In the obstructed case, the optimization resulted in an SOBP with a maximum dose difference of $\sim 45\%$, and a mean dose of $\sim 83\%$. These results highlight the need for further refinement of the methodology for more complex situations.

While the use of a simplified 2D model and a fixed pre-chosen dose distribution limit the generalizability of the method, the results serve as a proof of concept for patient-specific RF optimization. Future studies have several potential avenues to enhance the framework's practical applications.

- **Patient-specific dose distributions**: incorporating desired dose distributions based on DICOM files to add to clinical relevance.
- **3D Printing**: Investigating the possibility of producing optimized RFs.
- **Simulation Parameters**: The exploration of alternative settings and libraries to improve efficiency without compromise in accuracy.
- **Genetic algorithm improvements**: Adjusting convergence criteria, population size, and gene size to enhance optimization for complex cases.
- **Expand free variables**: Introducing additional free variables such as ridge pin heights and the introduction of range modulators to refine SOBP positioning and accuracy.

This study demonstrates the potential of combining Monte Carlo simulations and a genetic algorithm to produce patient-specific RFs for simple scenarios. However, further development is needed to extend the framework to more complex situations. Optimized RFs could significantly reduce treatment times, allowing the FLASH effect to be possible in pencil beam scanning systems.

References

[1] A. Patriarca, C. Fouillade, M. Auger, F. Martin, F. Pouzoulet, C. Nauraye, S. Heinrich, V. Favaudon, S. Meyroneinc, R. Dendale, A. Mazal, P. Poortmans, P. Verrelle, and L. De Marzi. Experimental set-up for flash proton irradiation of small animals using a clinical system. *International Journal of Radiation Oncology, Biology, Physics*, 102(3):619–626, 2018.

[2] Guoliang Zhang, Wenchao Gao, and Hao Peng. Design of static and dynamic ridge filters for flash–impt: A simulation study. *Medical Physics*, 49, 06 2022.

[3] Ruirui Liu, Serdar Charyyev, Niklas Wahl, Wei Liu, Minglei Kang, Jun Zhou, Xiaofeng Yang, Filipa Baltazar, Martina Palkowitsch, Kristin Higgins, William Dynan, Jeffrey Bradley, and Liyong Lin. An integrated physical optimization framework for proton stereotactic body radiation therapy flash treatment planning allows dose, dose rate, and linear energy transfer optimization using patient-specific ridge filters. *International Journal of Radiation Oncology, Biology, Physics*, 115(2):364–373, 2023. Received Jul 22, 2022; Accepted for publication Jan 21, 2023.

[4] Adeel Kaiser, John Eley, Nasarachi Onyeuku, Stephanie Rice, Carleen Wright, Nathan McGovern, Megan Sank, Mingyao Zhu, Zeljko Vujaskovic, Charles Simone, and Arif Hussain. Proton therapy delivery and its clinical application in select solid tumor malignancies. *Journal of Visualized Experiments*, 02 2019.

[5] Jorge Otero, Ivan Felis, Miguel Ardid, and Alicia Herrero. Acoustic localization of bragg peak proton beams for hadrontherapy monitoring. *Sensors*, 19:1971, 04 2019.

[6] Nader Mohamed, Anna Lee, and Nahyun Lee. Proton beam radiation therapy treatment for head and neck cancer. *Precision Radiation Oncology*, 6, 12 2021.

[7] Anouk Maasland. Master thesis. Master's thesis, Delft University of Technology, 2022.

[8] A. Ibrahimi. Characterization of the proton beam line in the experimental room of hollandptc. Master's thesis, Delft University of Technology, 2020.

[9] Giuliana Milluzzo. Physics and physics list in geant4, 2024.

[10] IHME, Global Burden of Disease. Alcohol use disorders [dataset], 2024.

[11] Dieter Schardt, Thilo Elsässer, and Daniela Schulz-Ertner. Heavy-ion tumor therapy: Physical and radiobiological benefits. *Rev. Mod. Phys.*, 82:383–425, Feb 2010.

[12] V. Favaudon, L. Caplier, V. Monceau, and et al. Ultrahigh dose-rate flash irradiation increases the differential response between normal and tumor tissue in mice. *Science Translational Medicine*, 6(245):9, 2014.

[13] S. Van De Water, S. Safai, J. M. Schippers, D. C. Weber, and A. J. Lomax. Towards flash proton therapy: the impact of treatment planning and machine characteristics on achievable dose rates. *Acta Oncologica*, 58(10):1463–1469, 2019.

[14] E. S. Diffenderfer, B. S. Sørensen, A. Mazal, and D. J. Carlson. The current status of preclinical proton flash radiation and future directions. *Medical Physics*, 49(3):2039–2054, 2022.

[15] Siddhartha Bhattacharyya and Marvin D. Troutt. Genetic search over probability spaces. *European Journal of Operational Research*, 144(2):333–347, 2003.

[16] Brookhaven National Laboratory. Bragg curves and peaks, 2024. Accessed: 2024-12-19.

[17] J. D. Bradley, R. Paulus, R. Komaki, and et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage iiia or iiib non-small-cell lung cancer (rtog 0617): A randomised, two-by-two factorial phase 3 study. *The Lancet Oncology*, 16:187–199, 2015.

[18] A. Modh, A. Rimner, E. Williams, and et al. Local control and toxicity in a large cohort of central lung tumors treated with stereotactic body radiation therapy. *International Journal of Radiation Oncology, Biology, Physics*, 90:1168–1176, 2014.

[19] P. van Marlen, M. Dahele, M. Folkerts, E. Abel, B. J. Slotman, and W. F. A. R. Verbakel. Bringing flash to the clinic: Treatment planning considerations for ultrahigh dose-rate proton beams. *International Journal of Radiation Oncology, Biology, Physics*, 106(3):621–629, 2020.

[20] W. H. Bragg. Lxxiii. on the absorption of rays, and on the classification of the rays from radium. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 8(48):719–725, 1904.

[21] M.-C. Vozenin, J. H. Hendry, and C. L. Limoli. Biological benefits of ultra-high dose rate flash radiotherapy: sleeping beauty awoken. *Clinical Oncology*, 31(7):407–415, 2019.

[22] J. Bourhis, P. Montay-Gruel, P. Gonçalves Jorge, and et al. Clinical translation of flash radiotherapy: why and how? *Radiotherapy and Oncology*, 139:11–17, 2019.

[23] V. Favaudon, R. Labarbe, and C. L. Limoli. Model studies of the role of oxygen in the flash effect. *Medical Physics*, 49(3):2068–2081, 2022.

[24] X. Cao, R. Zhang, T. V. Esipova, S. R. Allu, R. Ashraf, M. Rahman, J. R. Gunn, P. Bruza, D. J. Gladstone, B. B. Williams, and et al. Quantification of oxygen depletion during flash irradiation in vitro and in vivo. *International Journal of Radiation Oncology, Biology, Physics*, 111(1):240–248, 2021.

[25] J. Jansen, J. Knoll, E. Beyreuther, J. Pawelke, R. Skuza, R. Hanley, S. Brons, F. Pagliari, and J. Seco. Does flash deplete oxygen? experimental evaluation for photons, protons, and carbon ions. *Medical Physics*, 48(7):3982–3990, 2021.

[26] T. Bortfeld and W. Schlegel. An analytical approximation of depth-dose distributions for therapeutic proton beams. *Physics in Medicine and Biology*, 41(8):1331–1339, 1996.

[27] Ph.D. Richard L. Maughan. Proton therapy delivery: The equipment, 2025. Accessed: 2025-02-01.

[28] Guoliang Zhang, Junliang Wang, Yuenan Wang, and Hao Peng. Proton flash: passive scattering or pencil beam scanning? *Physics in Medicine Biology*, 66, 01 2021.

[29] Zafar Riazi, Hossein Afarideh, and Rasoul Sadighi-Bonabi. Influence of ridge filter material on the beam efficiency and secondary neutron production in a proton therapy system. *Zeitschrift für Medizinische Physik*, 22(3):231–240, 2012.

[30] Daniel W Miller. A review of proton beam radiation therapy. *Medical physics*, 22(11):1943–1954, 1995.

[31] David Roddy, Camille Bélanger-Champagne, Sebastian Tattenberg, Stanley Yen, Michael Trinczek, and Cornelia Höhr. Design, optimization, and testing of ridge filters for proton flash radiotherapy at triumf: The hedgehog. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1063:169284, 06 2024.

[32] R. L. Harrison. Introduction to monte carlo simulation. In *AIP Conference Proceedings*, volume 1204, pages 17–21, 2010.

[33] Peter Bonate. A brief introduction to monte carlo simulation. *Clinical pharmacokinetics*, 40:15–22, 02 2001.

[34] Alireza Haghighat. *Monte Carlo Methods for Particle Transport*. CRC Press, 07 2020.

[35] John Allison, Katsuya Amako, J E A Apostolakis, HAAH Araujo, P Arce Dubois, MAAM Asai, GABG Barrand, RACR Capra, SACS Chauvie, and RACR Chytracek. Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53(1):270–278, 2006.

[36] J. Perl, J. Shin, J. Schumann, B. Faddegon, and H. Paganetti. TOPAS: An innovative proton Monte Carlo platform for research and clinical applications. *Medical Physics*, 39:6818, 2012.

[37] B.D. Ripley. Thoughts on pseudorandom number generators. *Journal of Computational and Applied Mathematics*, 31(1):153–163, 1990.

[38] B.A. Wichmann and I.D. Hill. Generating good pseudo-random numbers. *Computational Statistics Data Analysis*, 51(3):1614–1622, 2006.

[39] Andreas Baumgarten. Does windows 11 exist in both bit versions (32 and also 64 bit)? https://learn.microsoft.com/en-us/answers/questions/1479466/does-windows-11-exist-in-both-bit-versions-(32-and, 2024. Accessed: 2024-01-05.

[40] Itissam Abuiziah and N. Shakarneh. A review of genetic algorithm optimization: Operations and applications to water pipeline systems. *International Journal of Physical, Natural Science and Engineering*, 7:341–347, 12 2013.

[41] K. S. Baumann, F. Horst, K. Zink, and C. Gomà. Comparison of penh, fluka, and geant4/topas for absorbed dose calculations in air cavities representing ionization chambers in high-energy photon and proton beams. *Medical Physics*, 46(10):4639–4653, 2019.

[42] International Commission on Radiological Protection (ICRP). *Adult Reference Computational Phantoms*, volume 39 of *ICRP Publication 110*. Annals of the ICRP, 2009. Prepared jointly with ICRU. Accompanied by CD-ROM containing numerical data representing the phantoms.

[43] B. Faddegon, J. Ramos-Mendez, J. Schuemann, A. McNamara, J. Shin, J. Perl, and Paganetti H. The TOPAS Tool for Particle Simulation, a Monte Carlo Simulation Tool for Physics, Biology and Clinical Research. *Physica Medica*, 2020.

[44] Ahmed Fawzy Gad. Pygad: An intuitive genetic algorithm python library. *Multimedia Tools and Applications*, pages 1–14, 2023.

[45] Dana Vrajitoru. Large population or many generations for genetic algorithms? implications in information retrieval. *Studies in Fuzziness and Soft Computing*, 12 1999.