TOF-PET scanner geometries for proton therapy quality assurance: a simulation study

H.J.T. (Tom) Buitenhuis
S1719254
December 2013
MSc Applied Physics thesis
University of Groningen
Kernfysisch Versneller Instituut
Faculty of Mathematics and Natural Sciences
Supervisor: dr. P.G. Dendooven
Supervisor: prof. dr. S. Brandenburg
1. Summary

A recent development in the treatment of cancer tumors is the increase in availability of facilities offering proton-beam therapy for tumor irradiation. Proton therapy offers numerous advantages over conventional photon and electron beam radiotherapy, such as higher dose-conformity and precision due to the inverted depth-dose-profile, i.e., the Bragg peak. Due to the high dose in the Bragg peak and the finite range of the protons, the proton dose profile is highly sensitive to errors which have an impact on the range of the beam. To translate the favourable properties of proton therapy into a clinical benefit, a method of verifying the dose delivery is mandatory. Direct measurements on the protons are not possible since they are stopped inside the body, meaning only secondary gamma radiation can be measured. The most technologically advanced treatment verification system foresees the use of a PET scanner to measure the radioactive isotopes that have been created during the irradiation. The distal edge is imaged to obtain information on range shifts with respect to the treatment planning.

In this study, we investigate the clinical benefit of using a full ring PET scanner, a limited angle PET scanner and a dual-head setup PET configuration using either an in-situ or in-room scanning protocol. Also the effect of TOF resolution is investigated. To this end, a prototype quality assurance framework was developed. This framework uses Monte Carlo simulation software to simulate the proton treatment and to obtain the isotope distributions using a convolution with experimental production cross sections. The resulting radioactive isotope distributions are then simulated for different detector designs, scanning protocols and TOF resolutions. A ML-EM algorithm was used to reconstruct these PET coincidence data. The effect of scanner design and scanning protocol on the total number of counts and the image quality was investigated. A distal edge detection algorithm was developed to quantify the ability of the scanners to measure the effect of artificially induced proton range changes due to spherical inserts of 5-10 mm diameter.

The scanner systems and scanning protocols were evaluated regarding coincidence counts, image quality, and distal edge detection ability. The proposed scanner designs differ in angular coverage as well as positioning. This will translate into a difference in total number of counts. The full-ring in situ scanner (0 s delay) was used as a hypothetical reference. This protocol delivered the highest number of counts, and has full tomographic coverage. Since a full-ring in situ scanner is not possible, tradeoffs have to be made. One can choose to place the full-ring scanner separate from the gantry (in-room setup). However, this will introduce some delay and a subsequent drop in total counts. Another tradeoff could be made, by keeping the scanning system in an in-situ position (0 s delay), but reducing angular coverage. This can be done by installing a partial-ring system or a dual-head system. In terms of count rate, the dual-head system performs better than the 2/3 limited-angle scanner, and about equal to a full-ring clinical in-room scanner with a delay of 45 seconds.

Regarding the image quality evaluated for different TOF resolutions, a clear improvement of image quality is seen when comparing the 600 ps resolution images to the 150 ps resolution images. All 150 ps images of different scanner systems seem to be of equal quality when doing a visual inspection. This indicates that state-of-the-art TOF performance can eliminate most limited-angle reconstruction artifacts that might be present at worse TOF resolutions. A quantitative characterization of different scanning protocols is in progress.

The detection of artificially induced range shifts in the PET images was not successful. This might be explained by several factors, such as that the size of the insert was too small to be measured, the fact that the spherical geometry introduces a non-uniform range shift, and a possible unforeseen lateral spread on the proton beams.
2. **Table of Contents**

1. Summary......................................................................................................................... 2
2. Table of Contents ............................................................................................................. 3
3. Introduction ...................................................................................................................... 5
   3.1. Problem and solution strategy ................................................................................. 8
   3.2. Solution approach ..................................................................................................... 9
4. The quality assurance framework.................................................................................... 10
   4.1. Treatment planning ................................................................................................. 11
   4.2. Proton transport simulation ..................................................................................... 11
       4.2.1. Preparation of input ....................................................................................... 11
       4.2.2. Translation of HU to material composition ..................................................... 12
       4.2.3. Fluence-based approach for calculating isotope production ......................... 16
       4.2.4. Timing structure of the beam delivery ............................................................. 17
       4.2.5. Geometry of the treatment environment ........................................................ 19
   4.3. Scaling to clinical values ......................................................................................... 21
4.4. Simulate PET with best parameters using GATE ..................................................... 22
4.5. Post-processing .......................................................................................................... 22
4.6. PET image reconstruction ......................................................................................... 23
4.7. Distal edge detection ................................................................................................. 24
5. Results ........................................................................................................................... 26
   5.1. Validation of physics list ......................................................................................... 26
   5.2. Planning CT used as phantom ............................................................................... 28
   5.3. Dose and isotope production .................................................................................. 29
   5.4. Scanner designs ....................................................................................................... 30
   5.5. Total number of counts ......................................................................................... 32
   5.6. Image quality .......................................................................................................... 34
   5.7. Distal edge detection ............................................................................................. 36
6. Discussion and conclusions .......................................................................................... 37
7. Acknowledgements ........................................................................................................ 39
8. References ..................................................................................................................... 40
9. Appendix ....................................................................................................................... 43
   9.1. Production cross sections ..................................................................................... 43
3. Introduction

A recent development in the treatment of cancer tumors is the widespread use of ion-beam therapy for tumor irradiation. The ions that are often discussed in the literature are $^1$H and $^{12}$C, i.e. proton therapy and carbon ion therapy. In this paper, we will limit our investigation to proton therapy. These protons offer numerous advantages over regular photon radiotherapy, such as higher dose-conformity and precision due to the inverted depth-dose-profile (Figure 1). Photons reach maximum dose after a slight build-up region of the order of a centimeter. After this maximum, the delivered dose drops exponentially. Protons and heavier ions, having non-zero mass and charge, lose energy in an entirely different way (Bethe & Ashkin, 1953). The characteristic depth-dose distribution of charged ions is called the Bragg-peak, which shows a plateau region upon entering the tissue followed by an increase of the delivered dose as the ion slows down. Due to increasing stopping power at low energy, the proton dose deposition shows a sharp peak at the end of the proton range after which the delivered dose drops to zero almost immediately. This dose-profile allows for more precise targeting of a tumor while sparing the surrounding tissue. Since the dose of one Bragg-peak often is not enough to fully irradiate a tumor site, the dose profile is extended to 3D. Along the beam direction, the energy of the beam can be modulated to allow different penetration depths. This will create a so-called Spread Out Bragg-Peak (SOBP). Perpendicular to the beam direction, the field can be extended by passive scattering or rasterization/spot-scanning. Passive scattering uses a scatter foil to extend the energy layer perpendicular to the beam. A special collimator can then be used to tune the lateral beam profile. In spot-scanning, different pencil beams on a fixed grid are used to “paint” a special profile tuned to the tumor dimensions for each energy layer. An example of such a grid can be seen in Figure 2. Using different weights for each modulated beam, an almost uniform dose can be delivered to the tumor site. Due to the finite range of the protons, tissue behind the end-point of the distal beam will not receive any dose. This can lead to effective treatment plans with only a few fields or even just one field.

Due to the high dose in the Bragg-peak and the finite range of the protons, the proton dose profile is highly sensitive to errors which have an impact on the range of the beam. To translate the favorable properties of ion-beam therapy into a clinical benefit, a method of verifying the dose delivery is necessary. Some treatment facilities have developed such a method, such as the BASTEI detector at GSI (Crespo, 2005), however, all such methods are still experimental and not yet routinely available.

In a recent review article, Knopf and Lomax give an overview of the state of the art in in vivo dose delivery verification methods that are currently in use or are being developed (Knopf & Lomax, 2013). They categorize each method using measurement technique (direct, indirect), timing (online, offline), and dimension (1D, 2D, 3D). Online imaging can give real-time feedback during the irradiation, while offline imaging provides information after irradiation has completed. The direct measurement methods discussed are the proton-range probe (online, 1D), and proton radiography and tomography (online, 2D). These methods are both based on the same principle, i.e. shooting high-energy protons through the patient and determining the residual range of the protons. These methods give direct information on the stopping power for protons, but due to multiple coulomb scattering the spatial resolution is rather poor compared to x-ray tomography. These methods are still experimental and are not used in the clinical practice.
Figure 1: Illustration of delivered dose as a function of penetration depth inside the body, comparing photon and proton beams. The dose of several proton beams are added to form the SOBP region. The physical dose deposition benefit of protons with respect to photons is indicated by the red region.¹

Figure 2: Rasterscan / spot-scanning method for extending the SOBP perpendicular to the beam. For each energy layer, different spots (the grid in the top right) are painted with pencil beams of different weights to obtain a uniform dose distribution that is highly conformal the tumor shape

¹ Image taken from:
http://appliedradiationoncology.com/wp-content/uploads/2013/02/Buchsbaum_figure01.jpg
The discussed indirect methods are **prompt-gamma imaging** (online) and **PET imaging** (offline, 3D). During dose delivery to the patient, all protons are stopped inside the body. This gives rise to the favorable sharp edge in the dose distribution, but this also means that proton transmission cannot be measured for dose delivery verification. However, several types of gamma radiation are produced, which can be used to extract information on the proton range. The first type of prompt gamma radiation is produced when the protons induce an excited state in tissue atoms, which decay back to their ground state. The second type is the radiation that is produced when the protons induce nuclear reactions which produce unstable isotopes, which then decay using gamma radiation. The third type of gamma radiation that can be measured is the bremsstrahlung caused by the deceleration and deflection of protons in the electro-magnetic field of atomic nuclei. Bremsstrahlung is not generally considered prompt-gamma radiation; however it can provide equally valuable information. The detector systems to image this **prompt gamma radiation** consist of a wide range of camera's, most notably the Compton camera, electron-tracking Compton camera's, and a gamma camera in combination with sliding single collimators, multi slit collimators and knife-edge slit collimators. (Bom, Jouliaezadeh, & Beekman, 2012; Cambria Lopez et al., 2012; Kormoll et al., 2011; Kurosawa, 2012; Min, Kim, Youn, & Kim, 2006; Min, Lee, Kim, & Lee, 2012; Smeets, 2012) The image dimensionality of prompt-gamma imaging depends on the camera design. For example, a slit camera gives a 1D image, but combined with the irradiation's spot scanning sequence (knowledge of the position of the spot at any one time), a 3D image can be obtained. A gamma-camera gives a 2D projected image, but when used in SPECT mode, one gets a 3D image. Compton camera's also deliver 3D images. The main advantage of prompt-gamma imaging is the ability to provide real-time feedback on the proton range, since the lifetime of excited states is of the order of nanoseconds. These prompt-gamma imaging systems are still in development and not optimized enough to use in the clinical practice.

The most promising technique to verify dose delivery is **positron emission tomography (PET)**. During the irradiation, the protons induce nuclear reactions in the tissue atoms which produce $\beta^+$ decaying nuclei. These $\beta^+$ particles (positrons) travel a certain distance before encountering their antiparticle, i.e. an electron. The positron and the electron recombine, and because of energy and momentum conservation, this produces two back-to-back 511-keV photons. These resulting photons can be detected using a PET-scanner and the resulting PET image can be correlated to the delivered dose. PET images always give 3D distributions, since this is inherent to the measurement technique.

There is no one-to-one correspondence between the measured PET image and the delivered dose due to the fact that the cross sections of the isotope production are dependent on the incident energy of the proton and that the production of positron emitting isotopes depends on tissue composition. However, because a substantial amount of the total dose (~25%) in a SOBP is delivered by the most distal energy-plane, in most cases it will suffice to verify that the distal edge of the PET-image is where it is expected. Deviations from the treatment plan that cause a range shift of the proton beams can cause overdosage of healthy tissue and underdosage of the tumor. Such range shifts will be visible in the distal edge of the PET image.
3.1. Problem and solution strategy

To obtain the maximum clinical benefit from proton therapy, some dose delivery verification is necessary. Currently PET is the most advanced technology that can be used to this end. Several types of PET dose delivery verification protocols have been proposed. Firstly there is offline PET, used for example in studies at Heidelberg Ion-Beam Therapy Center (HIT). (Bauer, Unholtz, Kurz, & Parodi, 2013) After proton radiation treatment, the patient is transported to a clinical PET scanner somewhere in the treatment facility. This introduces a delay of about 5-20 minutes before the start of the scan which lowers the PET activity, but has the advantage of full angular coverage and using a clinical scanner. Secondly there is in-room PET, which consists of a full ring clinical PET scanner in the same room as the treatment gantry. This will reduce transportation time to about 1-2 minutes and still has the advantage of full angular coverage. The last option is a dedicated PET system to be used in the treatment position, often called in-situ or in-beam PET. This protocol has the advantage of being able to start data acquisition immediately after the treatment has ended, eliminating isotope decay and minimizing biological washout. The downside is that regular PET scanners are unsuitable for this kind of application, since space has to be made to allow access for the treatment gantry and beam. This has given rise to research in alternative PET geometries, such as OpenPET (Yamaya et al., 2008), slanted full ring PET scanners (Crespo, 2005) and several limited angle designs, where sectors from a full ring PET system are taken out.

In this study, we investigate the clinical benefit of using a full ring PET scanner, a limited angle PET scanner and a dual-head setup in an in-situ PET configuration. This dual-head setup consists of two identical panels which can be positioned more freely than a clinical system, thus making it possible to position the detectors as close to the patient as possible, constrained by the geometry of the gantry and constrained by the shielding of the detector from the radiation during beam delivery. We compare these systems in terms of count rate, image quality and their ability to detect proton range shifts in the PET image.

The full ring system will deliver most likely the highest count rate and image quality, however it is also the most difficult to integrate into a treatment gantry and will be the most expensive option regarding hardware costs. The limited angle system is equivalent to a full ring system with some sectors taken out. This makes integration into the gantry easier and it will be cheaper since less hardware is involved. Reconstruction of the PET image will be more difficult, since fewer counts will be detected, and image reconstruction will suffer from limited-angle artifacts. The dual-head system will likely be the cheapest system to produce and the easiest to integrate into a treatment facility. The performance of this system will be subject to this investigation. Limited-angle artifacts will also be present in the dual-head system, since the reconstructor will only have partial tomographic coverage. However, when we introduce additional tomographic information for the reconstruction, such as Time Of Flight (TOF) information, these artifacts can be compensated for. (Surti, Zou, Daube-Witherspoon, McDonough, & Karp, 2011) In this thesis, we will investigate the effect of TOF information on the image quality for different systems, as well as the effect of different scanning protocols (in-situ, in-room), which we translate to increasing waiting time between the end of the irradiation and the start of PET data acquisition.
3.2. Solution approach

In order to investigate the effects of different scanner geometries, TOF resolutions, and scanning protocols on PET image quality for proton therapy dose delivery verification, we will explain the simulation framework that was used in Chapter 4. The results can be found in Chapter 5. Discussion and conclusions can be read in Chapter 6.
4. The quality assurance framework

The proton therapy quality assurance framework is based on a custom application for proton dose delivery, using the Geant4.9.6 toolkit\textsuperscript{2} for the Monte Carlo simulation of the passage of particles through matter. The framework also includes the GATE 6.1 application for imaging simulations (Jan et al., 2011; Jan, Santin, Strul, Staelens, & Assi, 2004), which itself is based on Geant4.6.5. In order to obtain clinically relevant results, all simulations are based on a real patient case for a head-and-neck irradiation. This patient was part of a study at the University Medical Centre in Groningen (UMCG) into the clinical benefit of proton therapy. The patient was treated with conventional photon radiotherapy, but a treatment plan was also made for proton irradiation. The goal of the framework is to be able to generate dose distributions and β+ decaying isotope distributions using a treatment plan from a treatment planning system (TPS), to introduce range modification into these distributions, and to use a PET scanner to image these distributions. This PET data will be reconstructed using a maximum-likelihood expectation-maximization (ML-EM) algorithm, and the resulting PET image will be used to detect the distal edge of the produced β+ distribution. We use this framework to investigate the effect of detector geometry, TOF resolution and scanning protocol on the total number of counts, the image quality and the ability to detect proton range modifications. A schematic representation of the workflow is depicted in Figure 3.

\textsuperscript{2}http://geant4.web.cern.ch/geant4/

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{simulation_workflow.png}
\caption{A schematic representation of the workflow of the simulation framework used in this thesis. The connection of the framework to a real scanner at a treatment facility can be seen in the left green box. Simulated and measured data can be reconstructed and compared with each other using the distal edge verification algorithm. Detailed explanations are given in the text.}
\end{figure}
4.1. Treatment planning

The first phase of the framework is the treatment planning phase. For the case study in this thesis, a treatment plan was made for a head-and-neck case based on a clinical CT. The plan was made using the XiO proton planning software by Elekta for a spot scanning beam delivery. The plan consists of a full irradiation of three fields, one from the back of the patient (i.e. along the sagittal axis), one from the left front, and one from the right front (both at 50 degrees from the sagittal axis). We will assume that the field from the back will be delivered first. After this first field, a PET scan will be taken. The other fields will be subsequently delivered, but are not simulated here. In this thesis, we will only regard the field from the back in order to have a clear distal edge image and we will use this to investigate the detection of the distal edge.

The treatment plan energy layers range from 24 – 177 MeV, and the field consists of 39 x 31 spots with a spot-to-spot separation of 5 mm. There are 7227 non-empty spots with a total cumulative beam delivery of 709112 monitor units (MU). These monitor units are what is measured by ionization chambers in the beam line at proton facilities to measure the dose delivered by the beam. Monitor units scale linearly with the number of proton. The resulting total planned dose distribution for all three fields can be seen in Figure 4.

![Figure 4: Total planned dose for all 3 fields of the head-and-neck patient using Elekta’s XiO treatment planning software for protons. The sagittal view depicts an irradiation of the mouth and neck area with the patient looking to the right. The field is limited by the neck (on the left) and the throat (on the right).](image)

4.2. Proton transport simulation

The treatment plan will be simulated using a custom Geant4.9.6 application, which was developed in-house at KVI. This program takes as input the translated treatment plan, the planning CT data, a conversion table from CT data to Geant4 materials and the timing information of the beam delivery. This program will simulate proton transport and it will generate a 3D delivered dose map, as well as isotope production.

4.2.1. Preparation of input

To be able to simulate the dose delivery, all input files need to be prepared in the right way. The planning CT, which is used as the Geant4 phantom, is in general delivered in DICOM format. Geant4 cannot natively load DICOM data, so Matlab is used to translate the DICOM files into one 3D binary file where each voxel represents the radiodensity of the patient in Hounsfield units (HU) at that
specific point. The planning CT for our head-and-neck patient used voxels of 0.975 x 0.975 x 2 mm³, where the 2mm slices are in the axial (head-to-toe) direction, so a linear interpolation was used to generate a 3D 1 x 1 x 1 mm³ voxel binary.

The treatment plan data from the XiO system was extracted to a plain text file. This file is formatted according to a proprietary and obfuscated template. Using trial and error, the specific relevant properties of the plan were extracted from this text file. Most TP systems can also export the treatment planning data to a DICOM file. For future simulations using different TP systems, an interface with DICOM’s treatment planning abilities will be implemented. The data extracted from the TPS and converted to a simulation macro includes: gantry radius, isocenter position and z-position of gantry, gantry angle, total number of rows and columns of spots, spot separation, and spot-size. For each spot in the treatment plan, also the spot position, energy, and amount of MU are extracted to the simulation macro.

4.2.2. Translation of HU to material composition

The treatment planning CT is used as a phantom for the proton transport calculations in Geant4. To be able to simulate proton dose delivery in Geant4, the phantom CT data in HU needs to be converted to Geant4 materials with a specific density and elemental composition. This is not an easy task, since there is no one-to-one correspondence between HU and stopping power of tissues. Different tissues can have the same HU value in the CT, and materials with identical stopping powers for protons can correspond to different radio densities in the CT. (Paganetti, 2012) The most sophisticated method to correlate HU to human tissue is based on (Schneider, Pedroni, & Lomax, 1996). Schneider et al. measured different materials in a CT scanner and from this data constructed a conversion table between HU and elemental composition and density of tissue samples. Our method builds on this work by interpolating between these elemental composition values to obtain smooth transitions between materials (see Figure 6). A total of 537 different materials are defined. The elemental composition of these materials is taken from the interpolated data extracted from Schneider’s paper. The mass density of the materials is calculated using the elemental composition and the electron density calibration curve of the scanner that was used for the planning CT. The electron density calibration data is displayed in Table 1.

To calculate the measured electron density from the relative electron density to water, the following formula is used

\[ \rho_e(HU) = \frac{\rho_{e_{rel}}(HU) \times n_{e_{water}} \times N_A}{m_{water}} \]

where \( \rho_e \) is the absolute electron density, \( \rho_{e_{rel}} \) is the electron density relative to water, \( n_{e_{water}} \) is the number of electrons in a water molecule (10), \( N_A \) is Avogadro’s constant, and \( m_{water} \) is the mass of a water molecule (18.01528 g mol⁻¹). This electron density is the electron density that was measured using the CT scan.
<table>
<thead>
<tr>
<th>HU</th>
<th>Relative electron density (to water)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1002</td>
<td>0</td>
</tr>
<tr>
<td>-722</td>
<td>0.28</td>
</tr>
<tr>
<td>-553</td>
<td>0.4</td>
</tr>
<tr>
<td>-87</td>
<td>0.9</td>
</tr>
<tr>
<td>-46</td>
<td>0.96</td>
</tr>
<tr>
<td>-7</td>
<td>0.99</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>1.05</td>
</tr>
<tr>
<td>87</td>
<td>1.07</td>
</tr>
<tr>
<td>222</td>
<td>1.09</td>
</tr>
<tr>
<td>219</td>
<td>1.11</td>
</tr>
<tr>
<td>451</td>
<td>1.28</td>
</tr>
<tr>
<td>808</td>
<td>1.47</td>
</tr>
<tr>
<td>1209</td>
<td>1.69</td>
</tr>
</tbody>
</table>

*Table 1: Calibration data taken from the scanner that was used to generate the planning CT. This data is used to translate the HU values from the scan into mass density.*

From the average atomic composition of the material, the average number of electrons per molecule can be calculated by

\[
\langle n_e \rangle(HU) = \sum_i z_i \varepsilon_i(HU)
\]

where \(\langle n_e \rangle\) is the average number of electrons per molecule, \(i\) loops over all atoms present in the tissue, \(z_i\) is the charge number of that atom, i.e. how many electrons it has, and \(\varepsilon_i\) is the relative abundance of that element in the tissue. This average number of electrons is calculated purely from the data by Schneider. From the CT measured electron density and the average electron density per molecule in Schneider’s tissue composition data, we can estimate the average molecular density using

\[
\langle \rho_{mol}\rangle(HU) = \frac{\rho_e(HU)}{\langle n_e \rangle(HU)}
\]

The mass density of the material is then calculated by multiplying the average molecular density with the average molecular mass

\[
\rho(HU) = \langle \rho_{mol}\rangle(HU) \sum_i m_i \varepsilon_i(HU)
\]

where \(m_i\) is the atomic mass of the atom. Using this method, the density of the material in the planning CT is calculated using the atomic composition from Schneider and the electron density.
from the CT calibration data. The resulting density as a function of radiodensity in HU is displayed in Figure 5. The tissue composition that was used in this calculation is displayed in Figure 6.

**Figure 5**: The calculated mass density using elemental composition from data by Schneider and the HU calibration of the scanner that was used to make the planning CT. “Measurement density” is the density of the materials that were used in the calibration of the CT scanner. “Calculated density” is the calculated tissue density used in the Geant4 simulation software.
Figure 6: Elemental composition of tissue as a function of radiodensity. These values are the fractions of atoms of the given type relative to the total number of atoms. This must not be confused with mass fractions.
4.2.3. Fluence-based approach for calculating isotope production

Geant4 offers built-in models to keep track of radioactive isotope production. It does this by adding a process that simulates when an inelastic collision with a nucleus produces another isotope. There are two downsides to this method. The first is that this will lead to poor statistics. The production cross sections are of the order of 10-100 mbarn. This means that the likelihood of an isotope being produced is relatively small, and a lot of primary particles are needed to provide sufficient statistics, since isotope production in Geant4 is a discrete process. The second problem is that the cross sections in Geant4 are known to differ substantially from experimental cross sections, which will lead to an isotope distribution that has a poor correlation with experimental data. To remedy these problems, a fluence-based approach was implemented which can be convolved with experimental production cross sections to obtain isotope production.

In order to be able to accurately simulate the isotope production, a Geant4 scorer was developed. (**ProtontherapyPSFluence**) This scorer can be applied to a specific volume, in this case the CT phantom, after which it keeps track of interactions within that volume. This scorer works alongside the analysis manager (**ProtontherapyAnalysisManager**) to keep track of a 4-D fluence matrix $fluence(x, y, z, E)$. This is not a regular fluence matrix, but it keeps count of the path length of protons through a voxel at a specific kinetic energy.

After a certain number of protons have been processed, the fluence scorer will calculate isotope productions from these protons. Currently, the program incorporates 10 different reaction channels which are relevant for production in biological tissues. These cross sections are taken from the EXFOR database. ("EXFOR," 2013) These cross sections are fitted with constant, linear, quadratic and exponential functions to best match with the experimental data. The cross sections of the channels that are currently incorporated into the framework are: O16(p,pn)O15, O16(p,p2n)O14, C12(p,pn)C11, N14(p,2p2n)C11, O16(p,pn)C11, C12(p,p2n)C10, N14(p,pn)N13, O16(p,p2n)N13, P31(p,pn)P30, and Ca40(p,2pn)K38. An overview of these cross-sections is given in the appendix. Of these cross sections, the most important produced isotopes are O15 and C11. Both of these are produced in soft tissue with a high probability, and have a good correlation with delivered dose. P30 and K38 are mostly produced on bone structures and as such do not play an important role in distal edge detection and correlate poorly with the dose distribution.

In the production calculation stage, the program will loop through all elements in the fluence matrix. For each element, it will calculate the material composition by using the Geant4 defined materials to obtain the relevant mass fractions and density of that material. From this data, we can calculate the atomic density in each voxel by using for example for O16

$$O16density = \frac{\rho \times massFracO}{16.00 \times 1.66 \times 10^{-24}}$$

where $1.66 \times 10^{-24}$ is the conversion factor from atomic mass units to grams and the density $\rho$ is given in g/mm$^3$. For other elements, the atomic weight will differ from 16.00. When the number of atoms in each voxel is known, the analysis manager uses the convolution with the experimental cross sections to obtain isotope production in the following way, e.g. for O15 production on O16
where $\sigma(E)$ is the cross section in mbarn at a specific energy, fluence is given in mm/voxel and $10^{-25}$ is the conversion factor from mbarn and mm/voxel to production/voxel.

**4.2.4. Timing structure of the beam delivery**

In most treatment facilities, the delivery structure protocol of the beam is fixed. The treatment starts with the highest energy layer of the most important field, i.e. often the field with the highest dose. The rationale behind this is that at the start of the treatment, the patient positioning error is the smallest. Over the course of the treatment, the patient may move and cause some misalignment between the desired position and his actual position. Since the highest energy layer delivers the most dose (roughly 25% of the total dose), it is beneficial to deliver the highest energy layer first, and subsequently the lower energy layers.

When the treatment facility offers PET imaging after beam delivery, the image can be improved by reversing the energy layers, i.e. to start with the lowest energy layer and end with the highest energy layer. This protocol is also known as *distal edge last* dose delivery. Radioactive isotopes are produced during the irradiation. These isotopes start to decay immediately after they are produced. To obtain the highest count rate at the distal edge, i.e. the most important part of the PET image, it is important to produce isotopes there right before imaging starts. When the distal edge is irradiated first, and the total dose delivery takes approximately 2 minutes, a substantial amount of the produced isotopes has already decayed before imaging starts. A list of the half-life's of the isotopes in question is displayed in Table 2. Isotopes produced at the start of a 2 minute irradiation have already existed for 1 lifetime for O15, 2 lifetimes for O14, and 6 lifetimes for C10. For C11 this is much less important, since its half-life is 20 minutes. Only a small fraction of the C11 will have decayed during irradiation. This means the ratio of O15/C11 will dramatically change over time. To obtain the maximum count rate at the distal edge, the treatment plan must be executed using the *distal edge last* protocol.

The decay of radioactive isotopes during beam delivery plays an important factor in the final $\beta^+$ distribution that can be measured with a PET scanner. In our simulation framework, we implemented a specific timing structure for treatment delivery that is based on a spot scanning mechanism. The time it takes to switch spots was set to 5 ms and the time it takes to switch to a different energy was set to 50 ms. These values are comparable to the timing structure at the *Paul Scherrer Institut* (PSI) treatment facility is Switzerland. Continuous beam delivery was simulated without any spill structure. This corresponds to beam generation in a cyclotron. A synchrotron would amount to a specific spill structure, since in a synchrotron the accelerator must be loaded with a batch of protons which are all at once accelerated to the target energy. This means that during the batch acceleration of the protons, no beam extraction is possible. The spill structure of the synchrotron would then amount to a couple of seconds where the protons are accelerated to the target energy, followed by an extraction period, which is also typically in the order of seconds. The spill duration and the time in between spill is much shorter than the half-life's of the main PET isotopes, so the simulation outcome is valid for both cyclotrons and synchrotrons, given that the total irradiation time is comparable. The total beam-on time was set to 2 minutes, which leads to a
total time of about 160 seconds for the entire delivery of the first field as the total time to change from one spot to the next and from one energy layer to the next is about 40 s.

The simulation time slice was set to 6 seconds. After each time slice $t_s$, the analysis manager first multiplies the already produced isotope production maps with a factor of $\left(\frac{1}{2}\right)^{t_s/t_1/2}$ to account for the decay that the isotopes have experienced that were produced before the time slice. Then the analysis manager loops through the fluence matrix to calculate isotope production within this time slice. The analysis manager multiplies this production with a factor $\left(\frac{1}{2}\right)^{t_s/t_2/2}$ to account for the average decay of the isotopes that were produced during the time slice. Then the analysis manager adds the production in the current time slice to the cumulative production, and the process repeats after the next time slice.

<table>
<thead>
<tr>
<th>Isotope</th>
<th>$t_{1/2}$</th>
<th>$\lambda$ (1/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O15</td>
<td>122.24 s</td>
<td>5.67 x 10^{-3}</td>
</tr>
<tr>
<td>O14</td>
<td>70.598 s</td>
<td>9.82 x 10^{-3}</td>
</tr>
<tr>
<td>C11</td>
<td>20.334 min</td>
<td>5.68 x 10^{-4}</td>
</tr>
<tr>
<td>C10</td>
<td>19.29 s</td>
<td>3.59 x 10^{-2}</td>
</tr>
<tr>
<td>K38</td>
<td>7.636 min</td>
<td>1.51 x 10^{-3}</td>
</tr>
<tr>
<td>N13</td>
<td>9.965 min</td>
<td>1.16 x 10^{-3}</td>
</tr>
<tr>
<td>P30</td>
<td>2.498 min</td>
<td>4.62 x 10^{-3}</td>
</tr>
</tbody>
</table>

*Table 2: An overview of the half-life and the decay constant of each isotope of which the production is simulated. The most important isotopes for tumors in soft tissue are O15 and C11.*
4.2.5. Geometry of the treatment environment

The geometry of the treatment environment plays an important role in the final produced dose distribution. A typical treatment environment is shown in Figure 7. For instance, the type of treatment bed, the applied immobilization device, the distance from the nozzle to the patient, the radius of curvature of the gantry, and the energy and position spread of the beam all have an effect on the dose delivery. In the planning CT that was used in this study, a bed was used that is different from the final bed that is installed in the treatment room. For the treatment planning, this bed was removed from the planning calculations, since treatment beds are often built from thin carbon fiber, which has little effect on the range of the protons. For our simulation of the dose delivery, the CT was modified to set the material of the bed and everything below it to air.

The immobilization device that was used for the final treatment was also used for the planning CT. This ensures the highest conformity between planning position and actual treatment position. However, this device is made from a material that is different from human tissue, as it in most cases is some kind of plastic. The automatic translation from CT data in HU to Geant4 material definition only incorporates human tissue and air. This means that the immobilization device will be translated to the equivalent human tissue material in the list. Since the device has a low density, this will not have a great impact on the proton range.

From the XiO treatment planning software, some parameters on the expected gantry geometry were extracted. XiO used a radius of curvature of approximately 2 m to the isocenter of the beam in the planning phase. This has also been implemented in the simulation software to increase the conformity of the simulations to the treatment planning reference. In the Geant4 proton therapy simulation software, all primary protons were generated at a axial-position that was equal to the axial-position of the isocenter (the axial axis is from head-to-toe). A distance of 2 m of the primary protons to the isocenter is ensured by generating the primary protons on a circle of radius 2 m around the isocenter. The position on this circle is defined by the field angle specified in the treatment plan. For one field, all the primary protons are generated at the same position. The direction of the protons is calculated from the spot separation and the spot position. The spot-size of 4 mm at the depth of the isocenter was created by applying a Gaussian spread of $2\sigma = 4\,\text{mm}$ perpendicular to the beam direction. No vacuum beam tubing or nozzle geometry was simulated. The nozzle geometry will depend strongly on the manufacturer design choices used in actual proton therapy facilities. For our simulation study, the effect of the nozzle was disregarded.

Energy and position spread of the beam, as well as divergent or convergent properties all depend on the specifics of the cyclo- or synchrotron and the further hardware that is involved. For the simulation, a realistic Gaussian energy spread of $\sigma = 1.5\,\text{MeV}$ was chosen, while no position spread was specified.
Figure 7: A typical proton treatment gantry. In this image, the gantry can be rotated to allow proton fields from different directions. The beam nozzle is also indicated. The Patient Positioning System (PPS), i.e. the treatment bed, is indicated with PPS and can generally be rotated along 6 axes.\footnote{Image taken from: http://www.scielo.br/img/revistas/ca/v15n1/a09fig01.gif}
4.3. Scaling to clinical values

When the treatment delivery simulation has finished, the software will produce eight binary files: one dose file and seven isotope production files (O15, O14, C11, C10, K38, N13, and P30) which hold the production of these isotopes at the end of the radiation, taking into account the decay during the irradiation. The production of these isotopes and the delivered dose scales with the total number of primary particles that were simulated, which in general is less than the actual amount of particles that are used in a real-world therapeutic environment. The dose and production binary files are then scaled up to compensate for this. This scaling will not only amplify the signal, but also the statistical noise present in the data.

The spot weights of each spot in XiO are given in monitor units, as is the standard for treatment planning software. The number of protons per monitor unit depends on the energy of the protons, and this energy dependence is in general not constant over time or easy to predict since it depends on atmospheric conditions at the ionization chamber such as the temperature and the humidity. At treatment facilities, this dependence is calibrated daily. For simulation purposes, we simulated 10 protons per monitor unit, using the approximation that there is no energy dependence between monitor units and protons. This dependence is generally not very large, so the approximation will yield realistic results.

To scale to the actual delivered dose and production values, it is important to know how many protons per monitor unit XiO assumes to be delivered. In order to calculate this scaling factor, a calibration simulation was done. Using the treatment planning software, a treatment plan for a uniform SOBP in water of 50 Gy was generated. This plan was simulated using 1 proton per monitor unit. The final dose distribution had to be scaled with a factor of $2.5 \cdot 10^6$ for the dose map to arrive at a SOBP dose of 50 Gy. From this data we can state that, if we assume an energy-independent relation between protons and monitor units, the scaling factor used in XiO is $2.5 \cdot 10^6$ protons per monitor unit.

The total delivered dose to the tumor in our head-and-neck case is set to 70 Gy by the physician in charge. This dose will be delivered over 35 equal fractions. Each 3 field fraction is thus set to deliver a dose of 2 Gy to the patient. In our simulation study, we simulate one field of one of these fractions with 10 protons per monitor unit to increase statistics and obtain smoother dose and production distributions. This means the scaling factor for one fraction simulated with 10 protons per monitor unit will be equal to $2.5 \cdot 10^6 / (35 \cdot 10) = 7142$. The maximum dose delivered to the tumor site for the simulated field from the back is then calculated to be 0.46 Gy.

In the next simulation step, GATE expects the isotope distribution to be given in activity, not in production, so we must scale these production binaries with their decay rate. The activity of the isotope is then defined by the decay rate times the production

$$A = \lambda N = \frac{ln(2)}{t_{1/2}} \cdot Production$$

where $A$ denotes the activity in a certain voxel and $N$ is the amount of isotopes produced. The value of the decay rate of each isotope is displayed in Table 2.
4.4. Simulate PET with best parameters using GATE

When the isotope production maps have been converted to β+ activity maps, these distributions are ready for imaging. This Monte Carlo imaging simulation is done in GATE 6.1. Different type of PET geometries have been simulated, i.e. a full-ring clinical scanner, a partial-ring clinical scanner and a dual head scanner. An image of these geometries can be seen in Figure 13. GATE simulates β+ decay using positron primary particles originating in the CT phantom translated to material properties using the same material definitions as were used in the treatment delivery simulation. It includes the positron energy spectrum for each isotope, taken from The Lund/LBNL Nuclear Data Search database, as well as the resulting accolinearity of the opposing 511 keV photons. Simulation time for all scanner configurations was set to 3 minutes, however only 2 minute simulation sections are converted to list mode files, as will be explained in section 4.5.

The GATE 6.1 PET simulation uses a digitizer to convert photon interactions in crystals to coincidence pulses. The digitizer is set to an energy resolution of 11%, typical for the LSO scintillation crystal considered (see section 5.4). The coincidence energy window is selected from 435 keV to 650 keV, with a coincidence timing window of 4.1 ns. No dead time was added. This is justified because of the low positron emitter activities and thus low count rates involved. These scanner parameters are taken from the Siemens Biograph PET scanner that is used at the UMCG hospital and are typical for all modern scanners. In the case of the full-ring scanner, coincidences are only registered if the two 511 keV annihilation photons are detected with an angular difference of at least 60 degrees. All coincidences were selected on the basis of the takeWinnerOfGoods protocol present in the GATE software. This protocol only selects the coincidence with the highest energy, if there is more than one coincidence present in the timing window. For this framework, the digitizer is set to operate with perfect TOF resolution. It will produce a ROOT file where the interaction position and timing difference of the coincidence is known with a precision far beyond what the scanner in question can register. TOF resolution is added afterwards, to allow the same coincidence dataset to be processed with different TOF resolutions. In this way, the effect of the change in TOF resolution can be studied in detail, without any other physical or statistical interference.

4.5. Post-processing

The ROOT file that was created by GATE contains practically infinite resolution on coincidence interaction position and timing difference. A random number generator is used to add a Gaussian blurring to the timing difference, equal to the resolution of the scanner. The post-processing ROOT script is also able to discard coincidences based on the acquisition time. This way the software is able to select the coincidences that were acquired in the separate intervals [0, 120] s, [30, 150] s, and [60, 180] s. Using these three timing intervals of 2 minutes each, the effect of delay between the end of the treatment and the start of the data acquisition can be investigated. For instance, the delay when using an in-situ scanner will be almost negligible, while the delay for an in-room scanner will be up to 2 minutes due to the need to move the patient from the irradiation position into the scanner.

The post-processing software is also able to reject coincidences from a full-ring scanner on the basis of the modules in which the coincidence was detected. This way the full-ring geometry
simulation can be used to generate a dataset for a partial-ring geometry. All coincidences that use a module that is not present in the partial-ring geometry will be rejected.

When all the coincidence data is selected, the data is formatted for use in the image reconstructor. This resulting coincidence information will be written to a list mode data file, which contains position and timing difference of all coincidences, and can be used in the PET image reconstruction, using a ML-EM algorithm. The position is spread uniformly random over the cross section of the scintillation crystals and, in order to minimize the positioning error of the coincidence in the crystal, at 9 mm depth inside the 22 mm crystal.

### 4.6. PET image reconstruction

Image reconstruction based on the generated list mode data is executed at the MEDISIP group of the Department of Electronics and Information Systems at the University of Gent in Belgium. The reconstruction is done using a maximum-likelihood expectation-maximization (ML-EM) iterative reconstruction algorithm. (Shepp & Vardi, 1982) Ray-tracing is based on the Siddon algorithm, and sensitivity and attenuation correction is being applied. Attenuation correction uses the translated CT data as used in the dose delivery and the PET imaging simulation. Sensitivity correction is calculated from a calibration simulation using GATE 6.1 in which 511 keV back-to-back gamma's were distributed uniformly along the field of view (FOV) of the scanner. From the detected coincidences as a function of the position on which they were generated, the sensitivity map can be calculated.
4.7. Distal edge detection

To be able to detect range shifts in the treatment delivery which corresponds to range shifts in the PET image, we employ a distal edge detection algorithm akin to the algorithm described in (Helmbrecht, Santiago, Enghardt, Kuess, & Fiedler, 2012). The PET images are first smoothed using a 3D median filter, which preserves edges while smoothing sharp spikes in the data. This algorithm determines the end-point of the PET distribution in the direction of the beam, so it will generate a 2D map (the plane perpendicular to the proton field) where each pixel corresponds to the distal edge depth on that position.

A visual representation of the algorithm is shown in Figure 8. The distal edge is determined in three steps. For each column in the beam direction:

i) Determine the global maximum in that column. This maximum serves as a reference value.

ii) Determine the top of the distal edge, which is defined as the last local maximum that is still larger than or equal to $LocalMax = fracOfGlobalMax \times GlobalMax$. In a typical analysis, the $fracOfGlobalMax$ parameter is set to 0.5 or higher.

iii) Determine the distal edge depth. The distal edge is defined as the first point of the distribution in the direction of the beam after the top of the distal edge, where the PET activity falls below a fraction of $LocalMax$. This is described by $DistalEdge = fracDistalEdge \times LocalMax$. Because protons have a sharp falloff at the distal edge, the $fracDistalEdge$ parameter can be set to 0.3 or lower.

When the distal edges are determined, the distal edge maps of two different distributions can be compared by subtracting the maps. Three different applications for this algorithm can be thought of:

1. **Compare simulated data with experimental data.** In the case of verifying dose delivery to a real patient, this would mean comparing the distal edge map of the measured PET image with the distal edge map of the simulated PET image. Any range deviations from the original treatment plan that have occurred during the treatment will be visible in this way.

2. **Compare experimental data with other experimental data.** Another case could be made for comparing the distal edge map of a first fraction PET scan as a reference to compare with the distal edge map of further fractions. Most radiation treatments are spread over approximately 30 fractions which can be given over the course of a month. During this month, some anatomical changes can take place, as well as occasional misalignments of the patient. When this introduces a range deviation from the first fraction, this will show up in the distal edge map.

3. **Compare simulated data with other simulated data.** In our study to investigate the effects of scanner geometry, scanning protocol, and TOF resolution, we compare the distal edge map of an unmodified patient phantom to the distal edge image belonging to a patient phantom that was modified with different inserts to induce range deviations from the original plan. We compare the two distal edge maps for each scanner configuration and protocol to determine the ability of the PET scanner to detect these range deviations.
Figure 8: Distal edge detection on a PET distribution. The blue and yellow graphs show PET activity in a 1D column in the beam direction. The sharp dip in the activity between Global max and Local max is caused by the nasal air cavity, in which very few radioactive isotopes are produced.
5. Results

5.1. Validation of physics list

The Geant4 software framework uses a numerical Monte Carlo approach to calculate the trajectory and energy loss of the protons through matter. Several numerical physics models are available in the Geant4 code, and each model has a different energy range in which it is valid. A lot of different combinations of physical processes are possible, and some are a better fit to experimental data than others. For example, in high energy physics the relevant numerical models are widely different from the models that are accurate for medical particle therapy. Geant4 was set up as a framework, and the idea of its creators was that users could swap different physics lists with each other. This has proven to be more inconvenient as Geant4’s user base grew. A recent development in the use of the Geant4 framework is the introduction of standard physics lists. These physics lists are lists of physics packages, which contain different numerical models of physical processes. For use in therapeutic particle therapy simulations, the appropriate physics list is the QGSP_BIC_EMY physics list, which is also used in the C11-beam HadronTherapy example packaged with Geant4. This physics list contains the following packages: G4EmStandardPhysics_option3, HadronPhysicsQGSP_BIC, G4EmExtraPhysics, G4HadronElasticPhysics, G4StoppingPhysics, G4IonBinaryCascadePhysics, G4NeutronTrackingCut.

Besides the relevant processes, also parameters such as the tracking cut and the maximum step size play a role. The tracking cut has effect on the amount of secondary particles that are generated. Geant4 uses an algorithm to decide if secondary particles are created or not. If the physics processes indicate that a secondary particle can be created, Geant4 first checks the energy of this particle. If the energy is too low, so that the average range of the particle is smaller than the tracking cut that was set by the user, the particle will not be generated and its energy will be deposited locally. In this way, the tracking cut plays a role in the resulting dose distribution. A smaller tracking cut will generate more accurate results at the expense of increasing computation time. The maximum step size also plays a role in the accuracy of the program. Geant4 uses an internal equation to generate the step size of the proton trajectory. Among the constraints is the fact that the stopping power should not change much along the step. Smaller steps will create more smooth and accurate trajectories, but the step size has a great impact on computation time, so a compromise should again be made. The physics settings optimized for hadron therapy by (Zacharatou Jarlskog & Paganetti, 2008) and (Grevillot et al., 2010) were implemented, which lead to a tracking cut of 10 μm and a maximum step size of 100 μm.

To validate the selected physics processes and parameter settings, a comparison was made between simulated proton range using our software and measured proton range at the MD Anderson facility described in (Grillin, 2010). For this validation, 1D proton beams of varying energies were simulated, using the same set of energies that was measured at the MD Anderson facility. In each plane perpendicular to the beam direction, the dose was calculated and compared to the experimental data. A graph of these proton ranges can be seen in Figure 9. From this experiment, we can conclude that the proton ranges are accurately simulated using our physics settings.
Figure 9: Comparison of experimental proton dose profiles with simulated dose profiles. Experimental data is taken from (Grillen, 2010). All dose profiles are normalized using their maximum. The simulated dose profiles are superimposed on the experimental data, where the data markers represent the measurements, and the solid lines represent simulation data. We can see good agreement for the proton range and the position of the Bragg peak. Some deviation occurs for the plateau region.

Besides the 1D dose of a single Bragg-peak, also the behavior of the proton beams when used in a SOBP is of interest. To see the effects of physics in a SOBP, we compare simulated dose profiles with an analytically calculated 3D SOBP dose profile. The analytical profile was calculated by (Yagi, Oxley, Dendooven, & Goethem, n.d.) in an unpublished paper. For this comparison, a SOBP region in water of 10 x 10 x 10 cm³ was filled with 21 x 21 x 23 spots, each separated by 5 mm in the direction perpendicular to the beam, and 4.5 mm in the beam direction. The maximum proton energy was set to 150 MeV to obtain a range of about 16 cm. The dose profile was calculated analytically as well as using the Geant4 Monte Carlo method. To obtain the 1D dose profile, the SOBP dose was averaged along the beam axis. Results of this comparison can be seen in Figure 10. The periodic behavior in the SOBP region of the Monte Carlo code is due to the fact that no energy uncertainty on the primary beam was defined. This way the Bragg peaks of each individual beam are still visible in the SOBP. The average profile of the SOBP however is comparable.
Figure 10: Dose comparison of a SOBP region of $10 \times 10 \times 10 \text{ cm}^3$ with $21 \times 21 \times 23$ spots, each separated by $5 \text{ mm}$ in the direction perpendicular to the beam, and $4.5 \text{ mm}$ in the beam direction. The analytical calculation is indicated by the solid black line, and the Monte Carlo calculation is indicated by the orange line. The periodic behavior in the SOBP region of the Monte Carlo code is due to the fact that no energy uncertainty on the primary beam was defined. This way the Bragg peaks of each individual beam are visible in the SOBP. The average profile of the SOBP however is comparable.

5.2. Planning CT used as phantom

The planning CT of the head-and-neck case that was used for this study consisted of 186 2-mm thick slices with a voxel size of $0.975 \times 0.975 \times 2 \text{ mm}^3$ for a total size of $480 \times 480 \times 374 \text{ mm}^3$. This CT was translated to a binary file where each voxel’s value indicated the radiodensity in HU, which can be seen in Figure 11. The material inside the region of interest (ROI) for this study, as indicated by the red circle, consists of soft brain tissue with a HU value in the range of $[0,100]$. When looking at the translation of this tissue into Geant4 materials, as can be seen in Figure 5 and Figure 6, it can be seen that the tissue density in this region is around $1 \text{ g cm}^{-3}$ and the PET production in this region will be dominated by $^{15}$O produced on $^{16}$O. Another, less important, PET image contribution will be made by $^{11}$C based on $^{12}$C. This estimation however is only valid for this phantom and this treatment plan. It cannot be generalized to other cases.
5.3. Dose and isotope production

The treatment plan is executed at the Millipede cluster of the University of Groningen (RUG). Using the custom Geant4 proton therapy software, the delivered dose map and the produced isotope maps are calculated in a simulation that uses 10 protons per monitor unit. The resulting dose distribution and O15 distribution, which is the most important isotope in the ROI, is displayed in Figure 12. From the dose distribution, we can clearly see that all protons stop inside the body, about 2 cm before the end of the jaw line. In the dose field, there is some dose visible outside at the back of the patient as well as inside the nasal cavities. Delivered dose is defined as deposited energy per mass density. In the air regions, delivered energy drops by about a factor of 1000, but this is accompanied by a similar density change, meaning the delivered dose will be equivalent to tissue dose. Air-filled cavities are thus basically invisible in the dose distribution. In the O15 distribution, we can clearly see the effects of the different materials. Air contains oxygen atoms, but with a density that is about 1000 times lower than that of human tissue. This means that the isotope production will also scale with such a value.
5.4. Scanner designs

The isotope distributions (O15, C11, O14, C10, N13, K39, and P30) were converted to positron activity distributions using the scaling as explained in section 0. In GATE 6.1, the activity distributions were imaged using different scanner designs. A graphical representation of the scanner designs can be seen in Figure 13. The first scanner design is the full-ring clinical scanner. This scanner consists of 36 modules, each containing 18 x 45 LSO crystals. The dimensions of these crystals are 4 x 4 x 22 mm$^3$ in a 4 mm pitch, and the radius of the scanner to the front-face of the LSO crystals is set to 42.24 cm. Scanner electronics were not simulated. The scanner is positioned along the z-axis (sagittal axis) such that the distal edge corresponding to our region of interest is at the center of the field of view (FOV). At the center of the FOV, the scanner sensitivity is highest, which will lead to the best image quality. This scanner design is based on typical designs of diagnostic PET scanners such as the Siemens Biograph and the Philips Ingenuity TF. Because there is no entry point for the proton beam, this scanner design is only applicable using an in-room scanning protocol.

The second design is the partial-ring scanner. This scanner design uses the same coincidence data set as the full-ring scanner. In post-processing, some modules are excluded from the dataset, and coincidences that are detected in these modules will be discarded. This way we are able to use the same simulation as for the full-ring scanner to obtain coincidence data for a scanner with 2/3 angular coverage as well as 1/2 angular coverage. For the 2/3 angular coverage scanner, 6 consecutive modules which face each other are discarded on each side of the scanner. For the 1/2 angular coverage scanner, 9 modules on each side are discarded. The modules are removed along the sagittal axis at the point where the beam would enter. This design makes it possible to use this scanner using an in-situ scanning protocol.

The last design which was included in this study is the dual-head scanner. This scanner consists of two identical panels of 180x220 mm$^2$, which can be positioned freely and as close to the patient as possible. Because of the freedom in positioning, the scanner is ideally suited for in-situ scanning. Because the scanner is positioned close to the patient, spatial resolution might increase by using smaller crystals. To investigate this effect, LSO crystals of 4 x 4 x 22 mm$^3$ as well as 2 x 2 x 22 mm$^3$ were used.

These scanner geometries were translated to GATE macros and used to image the positron distributions that were generated by the proton dose delivery application. In the post-processing stage, TOF resolutions of 600 ps, 300 ps, and 150 ps FWHM CRT (coincidence resolving time) were applied. Total scan-time was fixed at 2 minutes, but delays between the end of the radiation and the start of the acquisition of 0 s, 30 s, and 60 s were introduced to investigate the effect of an in-situ option (0 s delay) and the in-room option (30 s and 60 s delay).
Figure 13: Overview of the two different scanner geometries: full-ring clinical scanner and a dual-head two panel scanner. The position of the scanners is chosen such that the scanners have the highest sensitivity at the distal edge. The scanner geometries are overlayed on the planning CT and calculated O15 distribution (color wash distribution).
5.5. Total number of counts

An overview of the number of detected counts is displayed in Table 3. The proposed scanner designs differ in angular coverage as well as positioning. This will translate into a difference in total number of counts. For example, the 1/2 angular coverage scan has half the detector surface area of the full-ring clinical scanner. This means that of all the gamma particles that are emitted from the patient, the likelihood of detection is decreased by a factor of about 2 for the limited-angle scanner. The PET distribution is not symmetrical around the center of the FOV, which influences the total number of detected counts in each module. If the PET source were a point source at the center of the FOV, or at least cylindrical symmetrical around the central axis, the difference in total counts would be expected to be exactly equal to 2. The decrease in likelihood of detection will lower the total number of counts. The limited-angle scanners will have fewer counts that the full-ring clinical scanner, which means that image reconstruction will be more difficult.

The same goes for the difference in scanning protocol. *In-situ* scanning can start right after the irradiation has ended, i.e. when the activity and equivalently the count rate is highest. Longer delays between the end of the irradiation and the beginning of the scan will mean that some of the produced isotopes have already decayed and that the count rate will be lower. Fewer counts will again lead to worse image quality for the PET image reconstructions. In this sense, the total detected counts can provide an indication of the performance of the scanners.

Contributions from different isotopes are compared in Table 4. Since the ROI where the scanner is focused on consists of mostly soft brain tissue, which is rich in oxygen, production on O16 will contribute most. On O16, the isotope with the highest production cross section is O15. Also C11, N13 and O14 are produced, but with a smaller cross section. Since realistic scan-times are in the order of minutes, this favors the isotopes with a shorter half-life, since more of the produced isotopes will have decayed in the same time. This effect is can also be seen in the number of coincidence counts. The brain tissue also has some carbon in it, on which C11 and C10 are produced. C10 has a small cross section, but also a short half life (20 s), so most of the produced C10 will have decayed at the end of the scan-time. C11 has a much higher production cross section than C10, but its half-life is in the order of 20 minutes. In the 2 minute scan-time, only a small fraction of the produced C11 will have decayed. This effect will cause that, for realistic scan times in the order of a couple of minutes, while C11 and O15 may be produced in the same quantity, the main contribution in the PET image from an oxygen and carbon rich environment will originate from the O15 decay. In our case study, the contribution in the coincidence counts of C11 was 5 times lower than the contribution from O15. However, these numbers are not generalizable to other cases.
<table>
<thead>
<tr>
<th>Geometry and protocol</th>
<th>Total number of counts</th>
<th>Relative number of counts (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In-situ</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-ring</td>
<td>3363027</td>
<td>100</td>
</tr>
<tr>
<td>2/3 ring</td>
<td>2300727</td>
<td>68.4</td>
</tr>
<tr>
<td>1/2 ring</td>
<td>1729472</td>
<td>51.4</td>
</tr>
<tr>
<td>Dual-head</td>
<td>2589530</td>
<td>77.0</td>
</tr>
<tr>
<td><strong>In-room full ring</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 s delay</td>
<td>2778983</td>
<td>82.6</td>
</tr>
<tr>
<td>60 s delay</td>
<td>2353231</td>
<td>69.9</td>
</tr>
</tbody>
</table>

Table 3: An overview of the total number of counts of different geometries and protocols, with the total number of detected counts, as well as the number of detected counts relative to the full-ring in-situ scanner.

<table>
<thead>
<tr>
<th>Isotope</th>
<th>Total number of counts</th>
<th>Relative number of counts (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O15</td>
<td>2924221</td>
<td>100</td>
</tr>
<tr>
<td>N13</td>
<td>119025</td>
<td>4.1</td>
</tr>
<tr>
<td>C11</td>
<td>630492</td>
<td>21.6</td>
</tr>
<tr>
<td>O14</td>
<td>476868</td>
<td>16.3</td>
</tr>
<tr>
<td>C10</td>
<td>167240</td>
<td>5.7</td>
</tr>
<tr>
<td>K38</td>
<td>20345</td>
<td>0.7</td>
</tr>
<tr>
<td>P30</td>
<td>20485</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Table 4: Total number of coincidence counts specified per isotope which generated the 511-keV photons. Since most coincidences were generated by O15 isotopes, this isotope was used as a reference to compare with the contributions of other isotopes.
5.6. Image quality

All protocols displayed in Table 3 have been post-processed for a TOF resolution of 600 ps, 300 ps, and 150 ps FWHM CRT. The list mode files produced were reconstructed at the University of Ghent using an iterative ML-EM algorithm on a 4 x 4 x 4 mm³ reconstruction grid. Better timing resolutions will constrain the reconstruction algorithm in the possible distributions. This will lead to a faster approximation of the correct PET image in less iterations, which will reduce noise and improve image quality. Finding the optimal reconstruction iteration depends on qualitative factors, and as such, a visual inspection of iterations was needed to select the optimal iteration. After the optimal iteration was selected, image quality for the different TOF resolutions can be compared.

The effect of TOF resolution on image quality can be seen in Figure 14.

In total, 21 different distributions were reconstructed, i.e. 7 protocols and 3 TOF resolutions each. This is too much to display in this thesis, so 4 protocols were selected to showcase the effect of increasing TOF resolution. The first protocol is the full-ring in-room scanner, which has a delay of 60 seconds between the end of the irradiation and the start of the measurement. This would mean that the patient was transported from the treatment position to the scanner in 60 seconds. When the system is separated from the gantry, it makes sense to install a full-ring clinical scanner which will give the best image quality for that protocol. This in-room protocol is compared to the in-situ protocol. In the in-situ protocol, space has to be made for the beam to enter, so only the partial-ring and the dual-head geometry are feasible. Because of the in-situ positioning, the system is able to start data acquisition as soon as the irradiation ends, meaning a delay of 0 s.
Figure 14: An overview of the effect of TOF resolution on image quality. For four protocols, the image quality for 600 ps, 300 ps, and 150 ps is displayed.
5.7. Distal edge detection

The ultimate goal of PET quality assurance for proton therapy is the ability to detect proton range changes by analyzing the PET image. To assess the performance of these different scanner designs and protocols in detecting these range changes, the CT phantom was modified by inserting into the ROI a sphere of air or a sphere of “dense water” in the path of the protons. This dense water is a fictional material with the same chemical composition as water, but with a density of 1.5 g cm\(^{-3}\). These modifications to the phantom cause a change in the range of the protons. For the air insert, the range is prolonged, while for the dense water, the range is shortened, since the unmodified material has a density of around 1 g cm\(^{-3}\). Because the range of the protons is different, the positron activity distributions will also have a range shift at the distal edge. These range shifts in the positron distributions will be measured by the PET scanners. We can then use the distal edge detection algorithm on the modified PET image and use the unmodified PET image as a reference. The resulting 2D distal edge maps are then subtracted from each other to measure the artificially induced range change. The performance of the scanner can then be measured by its ability to detect these range changes. For this study, spherical inserts of 5 mm and 10 mm were introduced to the phantom in the ROI.

The distal edge detection algorithm as described in section 4.7 was applied to all PET images. Different values for the parameters were used: \(fracOfLocalMax\) ranging from 0.7 to 0.5, and \(fracDistalEdge\) from 0.3 to 0.1. The generated 2D distal edge maps were subtracted and in this dataset a ROI was defined where the range change should be visible in the distal edge. The mean and standard deviation of this range change was calculated. However for all distributions and for all parameter values, the calculated range change did not yield any statistically relevant value. All measured range changes were in the order of the surrounding noise, and when the mean and standard deviation was calculated, the measured range change was not statistically relevant.
6. Discussion and conclusions

The three main results in this thesis are regarding count rate, image quality and distal edge performance of different scanning systems and protocols. From the comparison of the count-rate of different protocols, which can be seen in Table 3, some conclusions can be drawn regarding scanner performance. The full-ring in situ scanner (0 s delay) was used as a hypothetical reference, since such a setup is not practically possible. This protocol delivered the highest number of counts, and has full tomographic coverage. Because of this, the full-ring in situ scanner has the best image quality. Since a full-ring in situ scanner is not possible, tradeoffs have to be made. One can choose to place the full-ring scanner separate from the gantry (in-room setup). However, this will introduce some delay and a subsequent drop in total counts. This effect can be mitigated by speeding up the process of transporting the patient from the treatment position to the scanning position, for example by using an on-rail system, where the patient or the scanning setup is placed on a rail to automatically move the patient into the scanner. Transportation of the patient or the scanner will also introduce the possibility of misalignment because the patient is moved between the end of the treatment and the start of the scan.

Another tradeoff could be made, by keeping the scanning system in an in-situ position (0 s delay), but reducing angular coverage. This can be done by installing a partial-ring system or a dual-head system. In terms of count rate, the dual-head system performs better than the 2/3 limited-angle scanner, and about equal to a full-ring clinical in-room scanner with a delay of 45 seconds. Also, any in-situ scanning setup will reduce the effect of biological washout. Since biological processes such as the movement of fluids transport radioactive isotopes away from the place where they were generated, correlation of PET with delivered dose will become worse for longer waiting times as in the off-line and in-room scanning systems. A dual-head system will likely be cheaper to manufacture, since less hardware is involved than in a full-ring or partial-ring system. The dual-head system also has more freedom in positioning than the other scanners. One possible disadvantage of the dual-head system is radiation damage. To obtain a sufficient count rate, the detector panels must be placed as close to the patient as possible. This will likely introduce neutron radiation damage to the electronics, as well as activate the LSO crystals, which will lead to more noise and possible rapid degradation of the scanning system. This might be counteracted by a good shielding solution, which is good at shielding the setup from neutrons but does not reduce the sensitivity to 511 keV photons too much.

The effect of introducing better TOF resolution is displayed in Figure 14. A clear improvement of image quality is seen when comparing the 600 ps resolution images to the 150 ps resolution images. The change in quality is especially apparent when the nasal cavity regions are compared. For the 600 ps in-room scanner with 60 s delay, the nasal cavity is closed at the top, while for the 150 ps resolution, the nasal cavity is sharply separated from the surrounding activity. For 600 ps, the in-room scanner seems to be much worse than the other scanners. This might be explained by the fact that the surrounding tissue is rich in oxygen. Most activity in that region will be based on O15 decay, which has a lifetime of 2 minutes (see Table 2). When there has been a delay of 60 s, half a lifetime for O15 has passed, which means that around 70% of original activity is left. This indicates the importance of using an in-situ setup. For irradiation done in a more carbon-rich environment, this might not pose such a problem, since the lifetime of carbon is 20 minutes. However, to see the same fraction of the decays of C11 as we see with O15, measurement time has
to be extended 10x longer. For realistic measurement times, measurements on O15 will give most counts. All 150 ps images seem to be of equal quality when doing a visual inspection. This indicates that state-of-the-art TOF performance can eliminate most limited-angle reconstruction artifacts that might be present at worse TOF resolutions. A quantitative characterization of different scanning protocols is in progress.

The detection of artificially induced range shifts in the PET images was not successful. This might be explained by several factors. The shape of the inserts, a sphere, was chosen to obtain clinically relevant results. Few, if any, simple geometrical objects can be discovered in the human body. There are no structures inside the body that have the shape of a simple box or cylinder, while a sphere seems much more realistic. However this means that the introduced range-shift is not constant over the surface of the insert, but it shows a peak at the distal edge. For the 5 mm inserts, the effective range shift is only 5 mm at the center of the insert, less than 5 at the rest of the surface, while going to 0 on the borders. The effective range shift is less than the diameter of the insert. A second problem with these inserts is that they are relatively small. PET reconstruction must be done on a 4 x 4 x 4 mm$^3$ voxel grid. Smaller grid-sizes have been tried, but in that case noise explodes due to insufficient statistics.

In a 4 mm grid, the 5 mm diameter range shift would show up as a difference of one voxel in each direction in the distal edge map. In an otherwise noisy environment, such a change is too small to statistically discern from its surroundings. Comparable studies use different type of shifts. For example: a study by (Robert et al., 2013) uses PET verification in ion beam therapy for detecting patient misalignment errors, which reduces to a change in the entire proton-field perpendicular to the beam direction. Such a change is easier to detect, since the whole proton field plays a role. Another study by (Helmbrecht, Santiago, Enghardt, Kuess, & Fiedler, 2012) does introduce a range change for the proton field, but the size of this change is a fulcrum (sawed off pyramid) of 4 x 4 cm$^2$ with a range shift of up to 1 cm. The area to detect this range shift is more than 64 times as large as the 5 mm diameter sphere, and more than 16 times as large as the 10 mm sphere. Such a geometry is much easier to detect, since the surface area is larger and more reconstruction voxels are involved. The distal edge detection algorithm has been verified at Philips Research in The Netherlands to perform accurately for 2 x 2 cm$^2$ inserts.

When the distal edge detection algorithm is executed on the O15 distribution, a range change of on average 2 mm is detected for the 10 mm air insert, which should yield an average range shift of somewhat less than 10 mm. This distribution is based on 1 x 1 x 1 mm$^3$ voxels, so voxel size plays less of a role. The fact that the 10 mm insert is not detected in the O15 distribution is curious, since the brain is oxygen rich and the effect should be visible. Another contributing factor might be the way the treatment plan is executed geometrically. All protons originate in the same spot for each field, and the spot is selected by changing the starting direction of the protons. This point is 2 m away from the isocenter. This means that the protons travel around 1.7 m through air before they enter the patient. This might cause an additional unforeseen lateral spread of the beam, where protons from other spots spill into the spot where the insert is placed, essentially eliminating the effect of the distal edge. It might also introduce additional range straggling which is not usually present at a treatment facility.
7. Acknowledgements

I would like to acknowledge and thank the following persons. First of all, I would like to thank Peter Dendooven for his supervision of the thesis and for sharing his knowledge and experience on proton therapy. Thanks is also due to David Oxley for his work on setting up the simulation framework and for his support of the software, even though he had left to work at PSI. I would also like to thank Sytze Brandenburg for allowing me to do the thesis in the Proton Therapy and Imaging (PT&I) group and an internship at Philips Research. At Philips Research in The Nederlands, I would like to thank Pedro Rodrigues, André Salomon, and Torsten Solf for supervising the internship in the same field as this thesis and providing valuable information regarding PET systems and reconstruction. I would also like to acknowledge the Millipede cluster staff for their support in running the software on a high performance computing cluster. Finally, I would like to thank the whole PT&I group for their discussions.
8. References


9. Appendix

9.1. Production cross sections

These isotope production cross sections are taken from the EXFOR database, and fitted with constant, linear, quadratic, and exponential functions to best account for the shape of the data. Important factors are the production threshold as well as the position of the resonance peak.

![O15 production cross sections](image)

*Figure 15: Production cross section of O16(p,pn)O15*
Figure 16: Production cross section of $^{12}$C$(p,pn)^{11}$C, $^{14}$N$(p,2p2n)^{11}$C, and $^{16}$O$(p,p\alpha n)^{11}$C.

Figure 17: Production cross section of $^{16}$O$(p,p2n)^{14}$O. In the EXFOR database, no data was available on this reaction, but since the mechanism is the same as for $^{12}$C to $^{10}$C, the cross-section is estimated to be equivalent to $^{12}$C$(p,p2n)^{10}$C.
Figure 18: Production cross section of $^{12}\text{C}(p,p_{2n})^{10}\text{C}$

Figure 19: Production cross section of $^{14}\text{N}(p,pn)^{13}\text{N}$, and $^{16}\text{O}(p,2p_{2n})^{13}\text{N}$

C10 production cross sections

![Graph showing C10 production cross sections](image)

N13 production cross sections

![Graph showing N13 production cross sections](image)
Figure 20: Production cross section of $P^{31}(p,pn)P^{30}$

Figure 21: Production cross section of $Ca^{40}(p,2pn)K^{38}$