Monte Carlo Based Mapping of Proton Microdosimetric Properties

Espen Johannessen Folkedal

Master Thesis in Medical Technology

Department of Physics and Technology

University of Bergen

June 2022

Acknowledgements

First of all, I would like to express my gratitude and appreciation for my supervisors, Kristian Smeland Ytre-Hauge and Andreas Tefre Samnøy. Their guidance, feedback and support throughout this thesis have been invaluable.

A special thank you goes to my fellow students, Trygve Birknes, Filip Bjurstrøm, and Karoline Lewinsen, who have been my closest friends and study partners for the last two years. I would also like to thank “lesesalgjengen” for countless lunch breaks and a lot of fun at room 534. I would also like to express my gratitude to all the PhD-candidates in the medical physics group at IFT, for helping and motivating me.

Finally, I wish to thank my family for all the support through the years. Without my family and their unconditional love and support, I would not be where I am today. Thank you to my mother, Marianne Johannessen, my father, Tore Folkedal, and my siblings, Ina & André.

Espen Johannessen Folkedal

Bergen, June 2022

Abstract

Purpose: This study aimed to investigate the microdosimetric properties of monoenergetic proton beams. As the primary goal of radiation therapy is to inactivate cancer cells by damaging their DNA, the energy deposition distributions at nano- and micrometric levels may help increase our understanding of radiobiology, as well as improve RBE-modelling.

Methods: In this project the GATE software was used to simulate particle tracks and energy depositions. The GEANT4 DNA physics list and the combination of Livermore and QGSP_BIC_HP were used and compared to one another. Monoenergetic proton beams of energies 1-100 MeV were simulated. Microdosimetric quantities were calculated for site sizes in the range 20 μm – 10 nm, and the results found in this work were compared to similar reported values in the literature.

Results: A mapping of the frequency- and dose-mean lineal energy, $\bar{y}_F$ and $\bar{y}_D$, as well as $yd(y)$-distributions for protons were calculated. The $\bar{y}_D$ and $\bar{y}_F$-values were found to increase with both decreasing proton energy and site size. The $yd(y)$-distributions was found to broaden with both increasing proton energy and decreasing site size. The GEANT4 DNA yielded systematically lower $\bar{y}_D$-values compared to the Livermore + QGSP_BIC_HP combination, with the difference increasing with smaller site sizes. The $\bar{y}_D$-values from this work corresponded well with similar research.

Conclusion: The results from this work could be used in investigating which site sizes that correlates best with the biological effectiveness of the proton beam. The field of radiobiology is complex and not understood fully, but microdosimetry could give a higher level of insight, and might improve modelling RBE.

Contents

1. Introduction

In 2020 there was 19.3 million new cases of cancer worldwide, as well as 10 million deaths [1]. Among many treatment types, radiotherapy is one of the most common, using high energy radiation-beams to deliver high doses to the tumor. The main goals of radiotherapy are to inactivate or kill the tumor cells, as well as sparing healthy tissue as much as possible. Many types of radiation are used in therapy, and among the most common is photon and charged particle-beams, including proton beams.

Protons have a beneficial energy deposition pattern compared to photons, as shown in Figure 1. Photons have their peak of energy depositions after traveling a short distance inside a phantom, following a falloff traveling further. On the other side, protons deposit almost all of the energy at the end of the track before stopping completely. This means the proton can deliver higher doses to the tumor and managing to further spare the healthy tissue simultaneously. In addition, charged particles have shown to be more effective in killing cancer cells at the same dose levels compared to photons. The interest in proton therapy is increasing, especially in Norway, as proton therapy centres are under construction in both Bergen and Oslo, expecting to treat their first patients in 2024 [2]. Proton therapy is a complex field, and though countries like USA and Japan has been performing proton therapy for several decades, much research is still needed to fully exploit the full potential of protons.

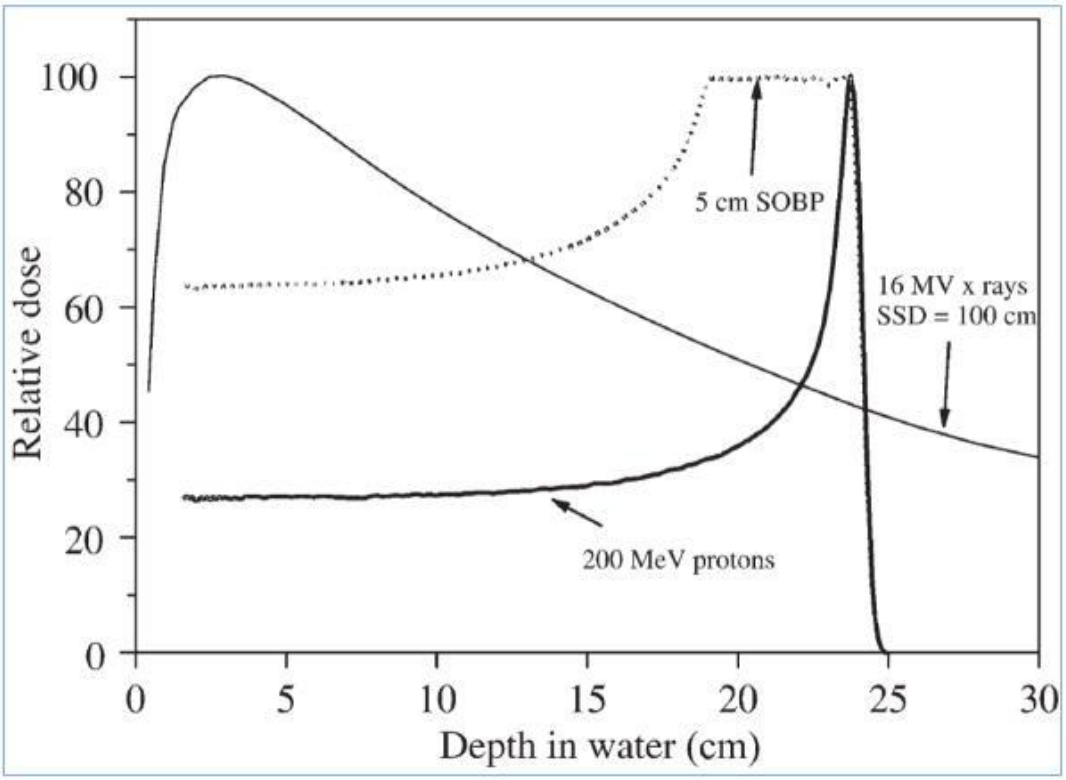

Figure 1: Shows dose-depth curves for a 200 MeV proton, a Spread Out Bragg Peak (SOBP) and 16MV photons in water. From [3]

1.1 Short history of radiotherapy

In the late 19th century, many important discoveries happened in the field of radiation, including Wilhelm Röntgen's discovery of x-rays and Marie & Pierre Curie's exploration of radiation. These findings accelerated the research on radiation therapy, and only 3 days after the Röntgen's announcement of the x-ray discovery, the first radiation cancer treatment took place [4]. During the next decades, the field experienced great progress, managing to somewhat reduce radiation to healthy tissue and developing new radiation sources. In the late 1940s, the thought of using charged particles, especially protons, started to form, and proton radiation experiments were performed on mice in the early 50s [5]. The first patient treated with proton therapy was in 1954, but it wasn't until the early 90s the first hospital-based proton therapy

center was opened in California [5]. As of May 2022, there are just over 100 proton therapy facilities worldwide, while 30 facilities are under construction [6].

1.2 Proton therapy and microdosimetry

As well as the beneficial physical attributes of protons, there are also some biological advantages compared to traditional photon therapy. Protons have shown to be more effective killing cancer cells given the same dose. This effect can be quantified with relative biological effectiveness (RBE), which is the ratio of dose for needed achieve the same biological effect as a reference radiation:

$$RBE = \frac{D_{reference}}{D_{Proton}} \tag{1.1}$$

The reference radiation is often set to photons. The RBE is a complex quantity depending on many factors, including dose and fractionation, ionization density of the radiation, tissue type, among others. In clinical practice, the RBE is set to a constant 1.1 with photons as a reference radiation, independent of proton energy and tissue [7]. However, radiobiological research gives reason to believe this is an oversimplification. The RBE of a proton beam is not constant, but seems to depend on a huge quantity of factors that are not fully mapped yet. A large amount of research is done to try and model RBE, and several models have been constructed, focusing on different aspects. One of the main factors that seems to affect RBE is the ionizing density of the radiation. As protons lose kinetic energy traveling through matter, the beam's linear energy transfer (LET) increases, leading to a more dense energy distribution [7]. A higher LET will lead to more damage to the tumor cells, making the probability of survival smaller.

The main target of proton therapy is to deactivate or kill the tumor cells by damaging the DNA. The cell nucleus, where the DNA is located, is a microscopic structure, while the size of DNA is in the nanometer range. LET is a quantity only describing

how a particle deposits its energy on average, and not where the energy is deposited in the absorbing material. Secondary electrons produced by the primary proton may travel several μm, depositing the energy far away from the protons initial track. On a microscopic level, the energy deposition may differ significantly even for two identical particles. To gain a deeper understanding of how the cell is damaged, the variation of energy depositions needs to be considered.

In microdosimetry, the energy depositions are measured in microscopical volumes called sites, where it is possible study the variations in energy depositions. However, the measurements are dependent on the site sizes. Studies have shown that analyzing smaller sites at the same scales as biological relevant structures may help model the biological effect for protons [8-10]. However, there is discussion on which volume sizes that are the most relevant for RBE modeling. Structures like the cell, the nucleus and the DNA itself has all been discussed as critical targets. Traditionally, the microdosimetric measurements have been performed with sites in the 1 μm-range, mainly because smaller detectors are harder to construct [10]. Due to the challenging dosimetry, it is of great value to use Monte Carlo-toolkits to simulate volumes down to the nanometer-scale, helping inspect energy depositions these ranges.

1.3 Project objectives/motivation

The proton RBE of 1.1 is widely seen as a simplification, and the use of microdosimetric measurements may help to better understand the complexity of radiobiology and RBE. If successful, this will lead to better dose calculations, which may help increase the dose to tumor and spare the healthy tissue.

The aim of this project is to map the microdosimetric properties of monoenergetic protons at different energies and site sizes with help of Monte Carlo simulations. While there is a lot of research in the microdosimetry field, no publications covering both a wide range of proton energies and site sizes were found at the time. Often, a wide range of proton energies were studied with 1 or 2 site sizes. To study how the

microdosimetric properties changes with different site sizes was one of the main goals for this work.

Using the GATE software, one of the main goals was to create a program simulating proton beams through water, collecting energy deposition data in a small sensitive volume. In GATE, there are several physics models with different levels of accuracy, and it was desirable to compare two of the models commonly used in medical applications, the Livermore and the GEANT4 DNA models and determine to which degree the more standard Livermore is valid for small site volumes. Protons with energy between 1-100 MeV were simulated, depositing energy in spherical sites with diameters in the range of 10 nm – 20 μm. With the raw data from the MC-simulations, microdosimetric quantities were calculated in python, with focus on the frequency- and dose-mean lineal energy $\bar{y}_F$ and $\bar{y}_D$, as well as the lineal energy distribution yd(y) for each energy and site size. The motivation behind this work is that these results could be used for input for RBE modelling as an alternative for LET, resulting in more accurate dose calculations.

2. Theory

2.1 is based on *The physics of proton therapy* [11]

2.2 and *2.3.1* is based on chapter 2 & 4 in *Microdosimetry - Experimental Methods and Applications* [12]

2.1 Physics of proton therapy

2.1.1 Proton interactions with matter

Charged particles typically loses their energy in many small interactions. Because of the particles' Coulomb field, the particles may interact with several atoms and their electrons while traveling through matter. There are several different possible interactions, including inelastic and elastic Coulomb scattering, nuclear reactions and bremsstrahlung. An illustration of different proton interactions are shown in Figure 2:

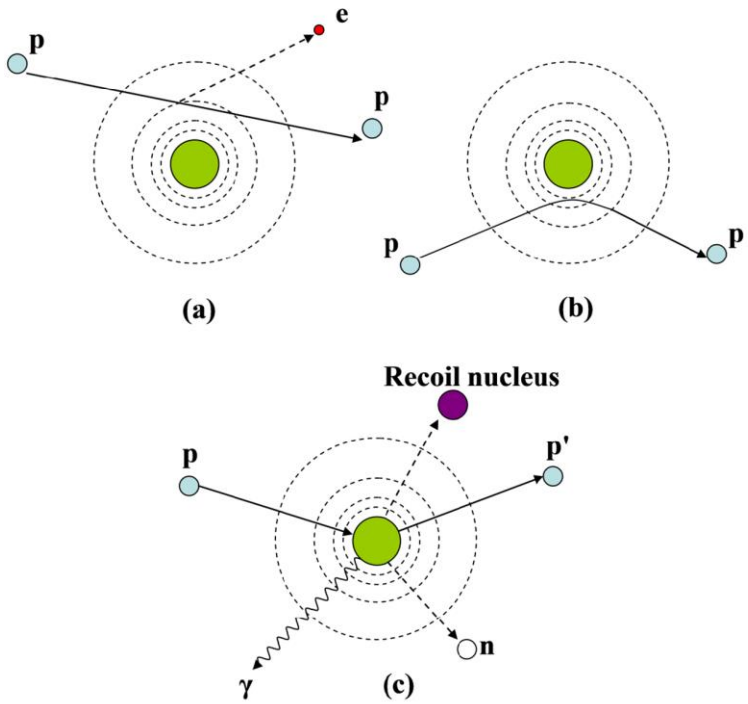

Figure 2: shows different types of proton interactions: (a) inelastic Coulomb scattering, (b) elastic Coulomb scattering, (c) non-eastic nuclear interaction. From [11].

Inelastic Coulomb scattering

Inelastic Coulomb scattering is the man factors of energy loss for protons at therapeutic energies. When a proton travels within close proximity to an atom, they interact with the atoms electrons. In these interactions, the proton relinquish some of its kinetic energy to the electrons, liberating them from the electron-shield. Because of the protons large mass relative to the electron, the proton continues in a straight line, while the electrons scatter in various directions. This interaction type is common, making it the main contributions to the protons energy loss. The high frequent energy loss from the inelastic Coulomb scattering can be seen as quasi-continuous.

Elastic Coulomb scattering

When the particle moves closer to the atomic nucleus, the proton may interact with the nucleus itself. In one of the interaction types the proton trajectory changes slightly due to a repulsive reaction from the nucleus. Because the nucleus is significantly heavier than the proton, no kinetic energy is transferred to the nucleus. When many of these small deflections happen, it is called Multiple Coulomb Scattering, which contributes to a lateral spread of the beam. The further the beam travels, the broader the beam becomes. However, when higher energy protons travels close to a nucleus, they can transfer some of their energy. In this situation, enough of the protons energy is transferred to the nucleus making it ionizing. The proton itself scatters in a large angle.

Non-elastic nuclear reactions

If a proton hits the nucleus, it may interact non-elastically. In a such interaction, the proton gets embraced by the nucleus. These interactions is rare, but have major impact on the proton that interacts. To enter the nucleus, the proton needs to have sufficient energy to overcome the Coulomb barrier, and in biological relevant material, the threshold is around 8 MeV. The nucleus that absorbs the proton gets

permanently transformed, and may eject several different secondary particles, including protons, neutrons, helium and others. Secondary protons from nuclear interactions can contribute up to 10% of the absorbed dose in a clinical, high-energy beam. If the nucleus doesn't absorb the proton during impact, the proton will transfer some of its energy, and then scatter out. The nucleus will then be left in an excited state, and will either release radiation or break up

2.1.2 Energy loss, stopping power

A particle's energy loss is stochastic, but it's still possible to calculate an expectation value for the energy loss along the trajectory. The energy loss rate by ions are defined by the mean amount of energy E the particle loses when traveling a distance x. The linear stopping power is often divided by the density of the matter it travels through. The stopping power is defined for a beam, not a single particle:

$$\frac{S}{\rho} = -\frac{dE}{\rho dx} \tag{2.1}$$

The Bethe-Bloch equation describes the stopping power for charged particles:

$$\frac{S}{\rho} = -\frac{dE}{\rho dx} = 4\pi N_A r_e^2 m_e c^2 \frac{Z}{A}\frac{z^2}{\beta^2}\left[\ln\frac{2m_e c^2\gamma^2\beta^2}{I} - \beta^2 - \frac{\delta}{2} - \frac{C}{Z}\right] \tag{2.2}$$

where N_A is Avogadro's number, r_e is the radius of an electron, m_e is the mass of an electron, c is the speed of light in vacuum, z is the charge of the traveling particle, Z is the charge of the absorbing medium, β is the relativistic speed v/c, $\gamma = \frac{1}{(1-\beta^2)^2}$, I is the mean ionization potential of the absorbing medium, δ is a density correction, and C is a shell correction factor.

There are some dependencies in this formula that is important to understand the stopping power of charged particles. Due to the ρ on the left side of the equation, the stopping power is proportional to the density of the absorbing medium. In a human body, the mass density, can differ extremely from lungs, to muscle tissue, to bones.

However, water is widely seen as a good tissue-equivalent material for proton therapy calculations, due to most of the body mass being water. What is also important to notice is the stopping powers dependency on the incident particles velocity. Outside the bracket, you can see that the stopping power is inversely proportional to β^2. This means that the stopping power decreases significantly with higher particle velocity. The β-terms inside the bracket contributes to a small rise at high energies (> 1 TeV).

The charge of the incident particle is also an important dependency. The stopping power is proportional to z^2, meaning a double in particle charge induces a 4 time increase in stopping power. The charge of protons is 1 e ($1.602 * 10^{-19}$ Coulombs). Carbon have a charge of 6 e, meaning carbons have a significantly higher stopping power than a proton of the same energy.

2.1.3 Range and straggling

The energy loss of charged particles is, as described over, a stochastic process. The randomness of the process makes it impossible to determine the exact range of a particle, and where it will stop. It is however, possible to estimate how long a particle, with energy E, will travel. When assuming the proton loses its energy continuously, an expectation value can be calculated, called range:

$$R(E) = \int_0^E \left(\frac{dE'}{dx}\right)^{-1} dE' \qquad 2.3$$

In practice, the range is the pathlength of the particle until it has lost all its energy. Another, closely related measure, is the projected range, defined as the longest penetration depth in one dimension of the charged particle. If the particle's direction changes slightly, due to for example elastic Coulomb scattering, the pathlength is longer than the projected range. In clinical settings, the pathlength of a proton is a good approximation of the penetration depth/projected range. The stopping power, along with the CSDA (continuous slowing down approximation) range is shown in Figure 3:

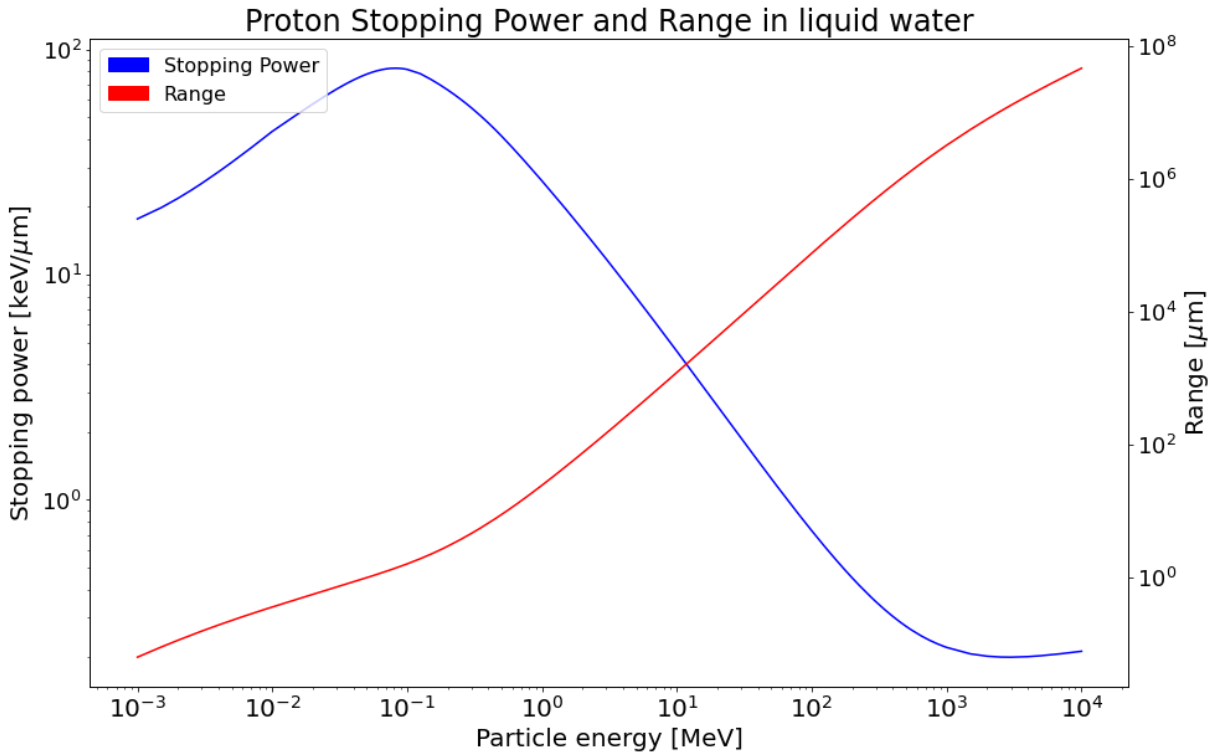

Figure 3: shows stopping power and range for protons in liquid water. Created with data from PSTAR [13]

As mentioned, the approximation of a continuous energy loss is needed to calculate the expected range of the proton. The processes of energy loss is stochastic, and the individual protons interact randomly with the matter. The difference in energy depositions between individual protons is called energy straggling. The energy straggling leads to a slightly different stopping depth for every proton. This effect is called range straggling. Figure 4 shows the fluence of a proton beam travelling through water and how the protons stops at different depths.

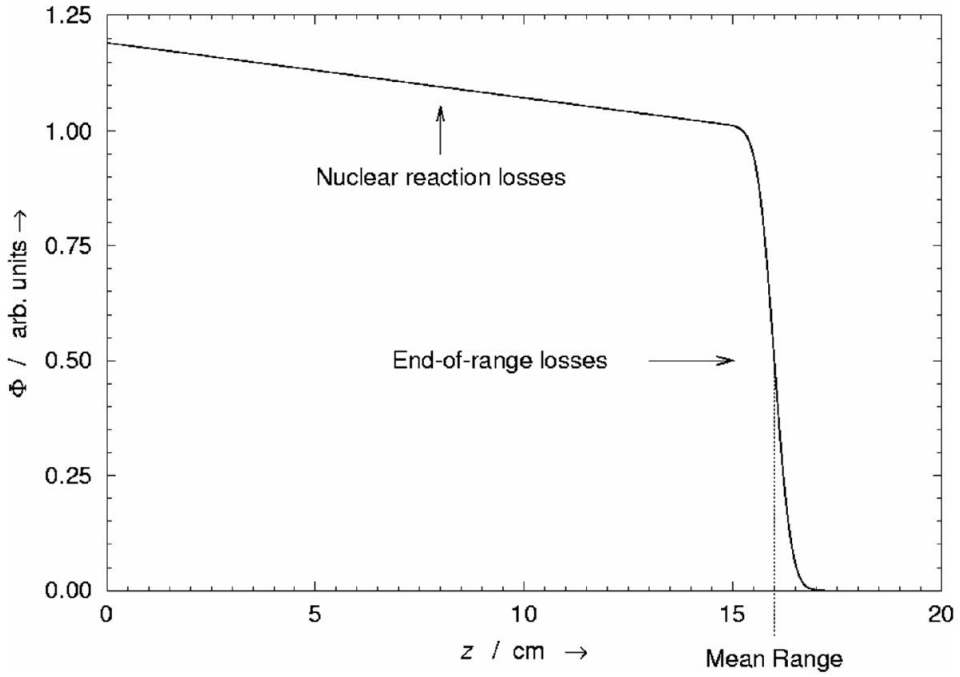

Figure 4: Shows relative fraction of the fluence of a proton beam traveling through water. From [11].

2.1.4 Depth-dose distributions

In radiotherapy, the main target of the radiation is often a tumor, and the tumor is regularly located several cm inside the body of the patient. The dose (energy deposit per mass) to the healthy tissue is desired to be as low as possible, though it is inevitable to irradiate the tissue in front of the tumor. For traditional photon radiation used in radiotherapy, the peak of the dose distribution happens only after a few cm of tissue. In addition, the radiation travels through the whole tumor, and irradiates the healthy tissue behind. Because of the stopping power of protons, the energy depositions (and thus the dose) is low for high energy protons, and rises with decreasing energy. This, including the fact that the protons stop when they have transferred all their kinetic energy to the absorbing matter, makes proton beams a better alternative than photon therapy for most cancer cases. Figure 1 shows an example of a depth-dose distribution for protons and photons.

The sharp rise of relative dose deposition at the end of the proton beam is called the Bragg Peak. After the Bragg Peak, the dose deposition decline quickly. Because of the range straggling of protons, the curve does not fall directly down to zero, but have a small "tail" at the end. The relative dose is significantly lower in the entrance area than in the Bragg peak. As equation 2.3 shows, the position of the Bragg Peak is energy dependent, and a series of several Bragg Peaks over a short distance can form a so called "Spread Out Bragg Peak" (SOBP), as shown with the dotted curve. This can be used to distribute the dose uniformly inside the tumor. The relative dose in the entrance area for SOBPs are higher than for a single, pristine Bragg Peak.

2.1.5 Secondary electron energy and range

The stopping power discussed over only accounts for the energy loss along the trajectory of the initial particle. The energy loss for protons are absorbed in a 3D-volume, much due to the secondary electrons produced by the inelastic Coulomb interactions. In fact, studies finds that secondary electrons in a 160 MeV proton beam deposits over 70% of the dose to the absorbing material [14]. The maximum energy a secondary electron can be transferred is depending on the initial protons energy:

$$E_e = 4\frac{E_p m_e}{m_e + m_P} \qquad 2.4$$

The maximum energies for secondary electrons from proton energies studied in this work are shown in Table 1, as well as the ranges for the electrons (gathered from [13]). The table shows that the energy the proton transfers is not necessarily deposited where the energy transfer happened, and that the secondary electrons can travel relatively far.

Table 1: Shows the maximum secondary electron range in liquid water for different Proton energies*. Ranges from [13] *rounded up to the next range in the table. **range was not in table, set to 1um for use in Methods.

PROTON ENERGY [MEV]	**1**	**2**	**5**	**10**	**20**	**50**	**100**
MAX δ-ELECTRON ENERGY [KEV]	2.2	4.4	10.9	21.8	43.5	108.9	217.7
MAXIMUM RANGE δ-ELECTRONS [μm]	1**	1**	8.6	12.7	35.9	208.3	637.2

When a proton is enters a medium (like water), the production of secondary electrons start. The region from the entry point to the maximum range of delta electrons are called the buildup region. When the proton has passed the point of the maximum range, there is now a maximum of secondary electrons. This is called charged particle equilibrium (CPE). Since electrons deposit a substantial part of the energy inside the absorbing material, and CPE is present in the patient during proton therapy treatment, it is also necessary to have CPE in the location of the measurements for the measurements to be relevant.

2.2 Microdosimetry

2.2.1 Dose related quantities

Energy

The most fundamental quantity in dosimetry in general is energy deposited. The energy ϵ deposited in a single interaction i is:

$$\epsilon_i = \epsilon_{in} - \epsilon_{out} + Q \qquad 2.5$$

Q is the change in rest energies, and ϵ_{in} and ϵ_{out} are the energies for the particles before and after the interaction, respectively. If Q is negative, there is an increase in

rest energies, and a positive Q indicates a decrease in rest energies. The energy imparted is the sum of all the energy deposits to the matter in a single volume/site. The unit is Joules, but can also be expressed in electron volts (eV):

$$\epsilon = \sum_i \epsilon_i \qquad 2.6$$

Specific energy, *z*

Specific energy, *z*, is the energy imparted in the volume divided by the mass of the volume. The unit for specific energy is Gray (J/kg):

$$z = \frac{\epsilon}{m} \qquad 2.7$$

As ϵ is stochastic, it follows that the specific energy is also stochastic. However, the mean specific energy is non-stochastic:

$$\bar{z} = \int_0^\infty z f(z)\, dz \qquad 2.8$$

f(z) is the frequency function (probability distribution). $\bar{z}$ equals the absorbed dose for the volume, and includes both single and multiple events. The single event frequency-mean specific energy is given as:

$$\bar{z_s} = \int_{z_{min}}^{z_{max}} z f_s(z)\, dz \qquad 2.9$$

where $f_s(z)$ is the probability density for *z* due to single events. The mean number of energy deposition events is defined as:

$$\bar{n} = \frac{\bar{z}}{\bar{z_s}} \qquad 2.10$$

The dose-mean specific energy can be calculated using the dose distribution function $z_d(s)$:

$$\bar{z}_{D,s} = \int_{z_{min}}^{z_{max}} z d_z(z)\, dz \qquad 2.11$$

Absorbed dose

Absorbed dose is given as energy deposited in a volume divided by the mass of the volume:

$$D = \frac{d\bar{\epsilon}}{dm} \qquad 2.12$$

In practice, the absorbed dose is measured in relative large volumes, which means the variation between measurements is small, and the difference between D and the mean $\overline{D}$ can be ignored. This means that $\overline{D} = D = \bar{z}$. The unit for absorbed dose is Gray (J/Kg). This is a macroscopic unit, and standard measurement unit in radiation therapy.

Linear energy transfer (LET)

LET stands for linear energy transfer. It describes the amount of energy an ionizing particle transfers into the matter per distance travelled. The unit is often expressed as keV/μm. It is in close relation to stopping power. LET quantifies the amount of energy that is deposited locally. We can separate between restricted and unrestricted LET. Restricted LET excludes energy lost by secondary electrons (delta rays) over a energy limit Δ. Restricted LET is more accurately displaying the local energy transfer, since the secondary electrons with higher energy have a longer range, and may deposit their kinetic energy non-locally. Restricted LET can be written as:

$$LET_{\Delta} = \left(\frac{dE}{dx}\right)_{\Delta} \qquad 2.13$$

Unrestricted LET equals LET_{∞}, and includes energy loss for all delta rays emitted, and is the same as the stopping power. When traveling through matter, the LET of the charged particle changes. The track-average LET is defined as:

$$LET_t = \int LET_{\Delta}\, t(LET_{\Delta}) dx \qquad 2.14$$

Where $t(LET_{\Delta})$ is the fraction of the track between the two values $LET_{\Delta,1}$ and $LET_{\Delta,2}$. When weighting the LET with the absorbed dose, LET_D can be calculated. $d(LET_{\Delta})$ is the dose fraction, analogous to $t(LET_{\Delta})$:

$$LET_D = \int LET_{\Delta}\, d(LET_{\Delta}) dx \qquad 2.15$$

When traveling through matter, the particle loses energy, subsequently changing the LET of the beam. The averaged LET-values are low for a proton beam from the entrance point until the Bragg Peak, and rises significantly at the peaks depth. It also rises after the peak, as seen in Figure 5.

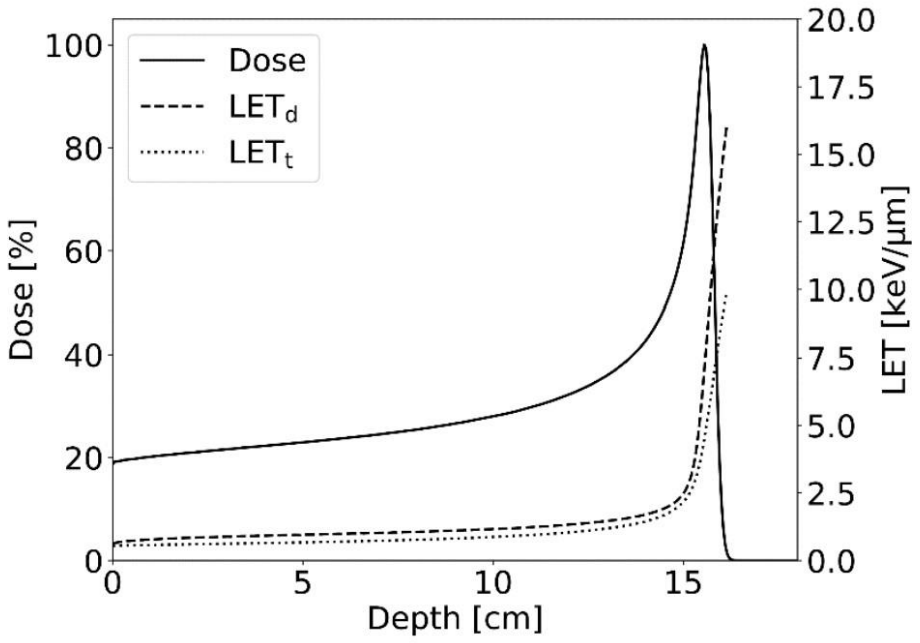

Figure 5: Shows dose from a 150 MeV proton beam in water, together with corresponding LET_t and LET_D values. Figure from [15].

LET is an average for a specific proton energy, and therefore also a deterministic quantity. This means LET does not describe the stochastic variation of energy depositions due to energy straggling. The statistical variations will have a larger influence on the energy depositions in smaller volumes. LET does not describe where in the absorbing medium is deposited, only the amount of energy is deposited by the particle per distance travels. This is not a problem when studying macroscopic structures, but when studying energy depositions on a micro- or nanoscopic scale, this becomes important. LET is also not possible to measure, but is calculated for each particle type, energy, and the medium the particle is traveling through. Therefore, LET is not a good metric quantifying the energy-depositions at a sub-macro level.

2.2.2 Microdosimetric quantities

As discussed over, the LET is a good metric to describe the energy depositions at macroscopic scales. At microscopic or nanoscopic scales, however, the LET does not accurately describe where the energy depositions take place. Microdosimetry was developed to solve some of the problems with using LET in the sub-macro field of dosimetry. Instead of describing energy depositions in a one dimensional matter, microdosimetry studies the energy depositions in a fixed volume. This makes the LET-equivalent microdosimetric quantity *lineal energy* measurable, as well as it describes the spatial energy distribution of a radiation beam.

Lineal energy *y*

In microdosimetry, the volumes that are investigated are called sites. It is important to know the size, shape and materials of the site [16]. The amount of energy a single event deposits is called ϵ_s. Lineal energy is defined as the energy a single event deposits to a volume divided by the mean chord length $\bar{l}$:

$$y = \frac{\epsilon_s}{\bar{l}} \tag{2.16}$$

The mean chord length is the average distance a particle travels inside the site. The mean chord length is dependent of the volume and surface area of the site [17]:

$$\bar{l} = \frac{4V}{S} \tag{2.17}$$

V is the volume of the site, and S is the surface. For spheres and cylinders, which is the two most common shapes of detectors/sites in microdosimetry, the mean chord lengths can be written as:

$$\bar{l}_{sphere} = \frac{4r}{3} \tag{2.18}$$

$$\bar{l}_{cylinder} = \frac{2rh}{r+h} \tag{2.19}$$

where r is the radius for the sphere/cylinder, and h is the height of the cylinder. Similar as the specific energy *z*, it is useful to use probability density distribution of *y* to find the frequency-mean lineal energy:

$$\bar{y}_F = \int_0^\infty yf(y)dy \tag{2.20}$$

An example of f(y) is shown in Figure 6, from a simulation a 10 MeV proton beam travelling through a spherical site with a diameter of 1 μm. It can be seen that most of the energy depositions happens from low-energy interactions:

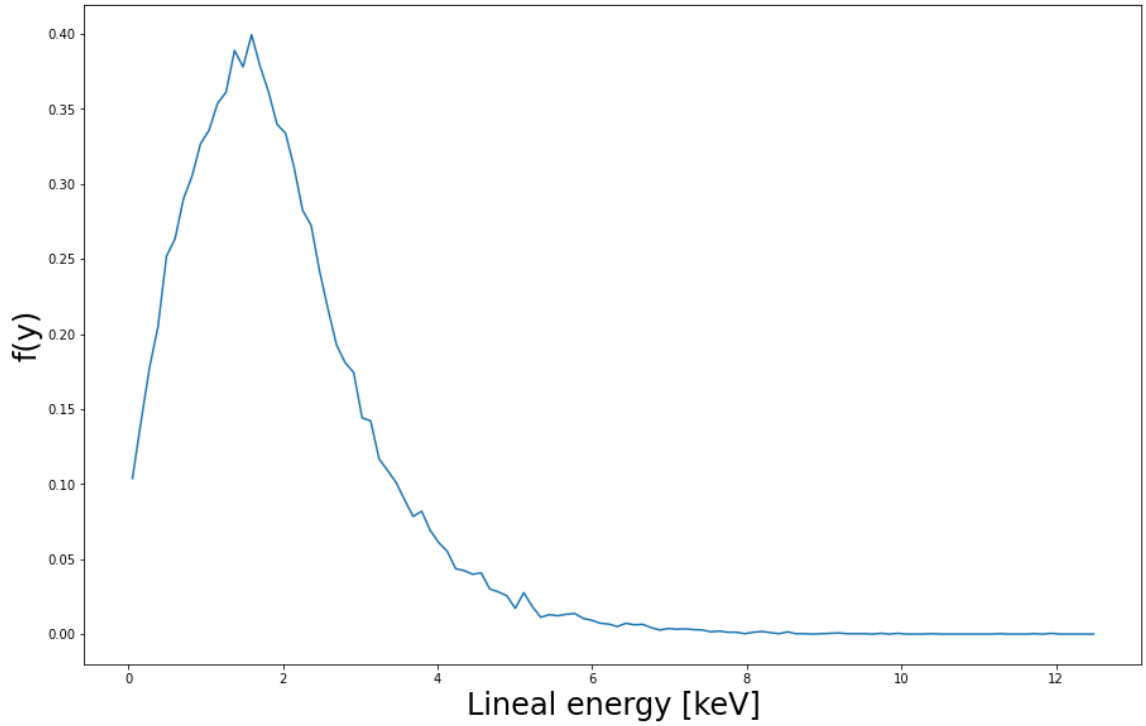

Figure 6: Shows the lineal energy distribution for 50 MeV protons through a 1 μm sphere site.

$\bar{y}_F$ was found to be 3.88 keV/µm in this. This graph will be looking different for different particle energies and particle types. Lineal energy y is stochastic, while $\bar{y}_F$ is a deterministic quantity. This lineal energy distribution only shows how frequently lineal energy values occur, and not which lineal energy contributes most to the dose. The dose-weighted lineal energy distribution is therefore introduced:

$$d(y) = \frac{y}{\bar{y}_F} f(y) \tag{2.21}$$

The dose-mean lineal energy derived from the dose-weighted distribution is written as:

$$\bar{y}_D = \int_0^\infty y d(y) dy = \frac{1}{\bar{y}_F} \int_0^\infty y^2 f(y) dy \tag{2.22}$$

Figure 7 shows the difference between the lineal energy distribution and the dose-weighted lineal energy distribution. $\bar{y}_D$ was found to be 5.43 keV/µm, which is

significantly higher than $\bar{y}_F$. This reflects that the higher lineal energy events contribute more to the total dose than the lower, and this is also shown in Figure 7.

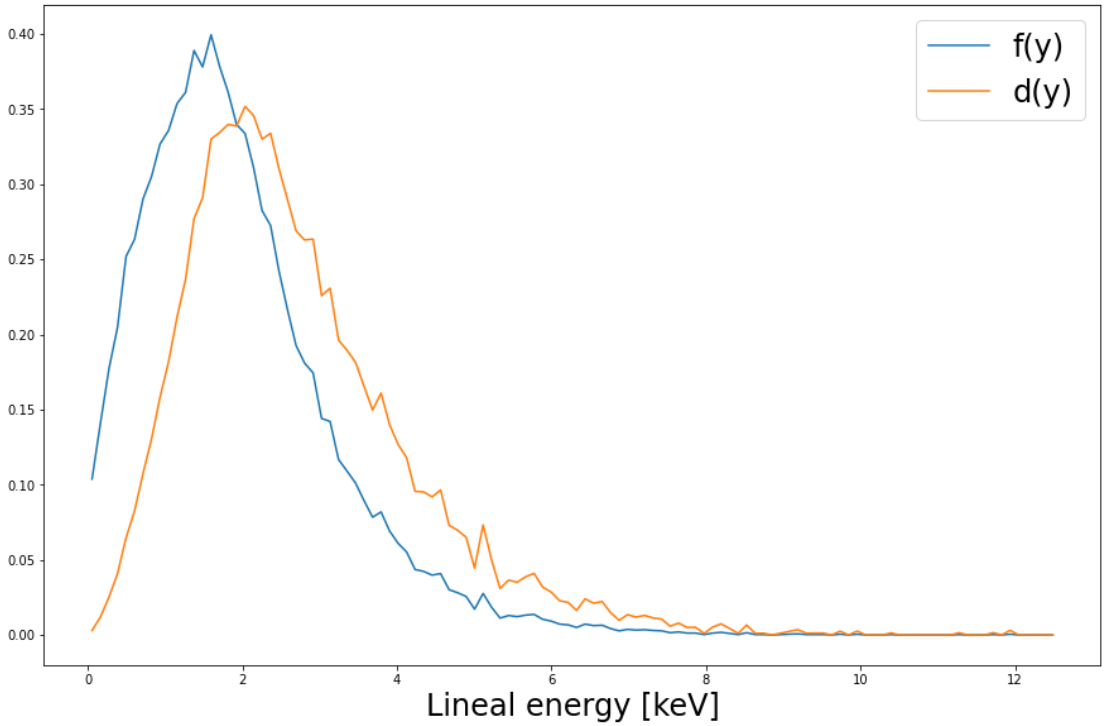

Figure 7: Shows both f(y) and d(y) for 10MeV protons travelling through a 1 µm sphere site.

The dose-mean lineal energy will of course change for different proton energies. Equation 2.2 shows that the energy loss of a particle is inversely proportional to β^2, indicating that the higher energy protons will have a lower stopping power, and accordingly a lower lineal energy. Because y is the energy deposit in the site divided by mean chord length, the radius of the site will also affect $\bar{y}_D$. Generally for protons, the dose-mean lineal energy increases with smaller site radius/diameter. Figure 8 below shows that while L_∞ is unaffected by the site diameter, $\bar{y}_D$ is inversely proportional.

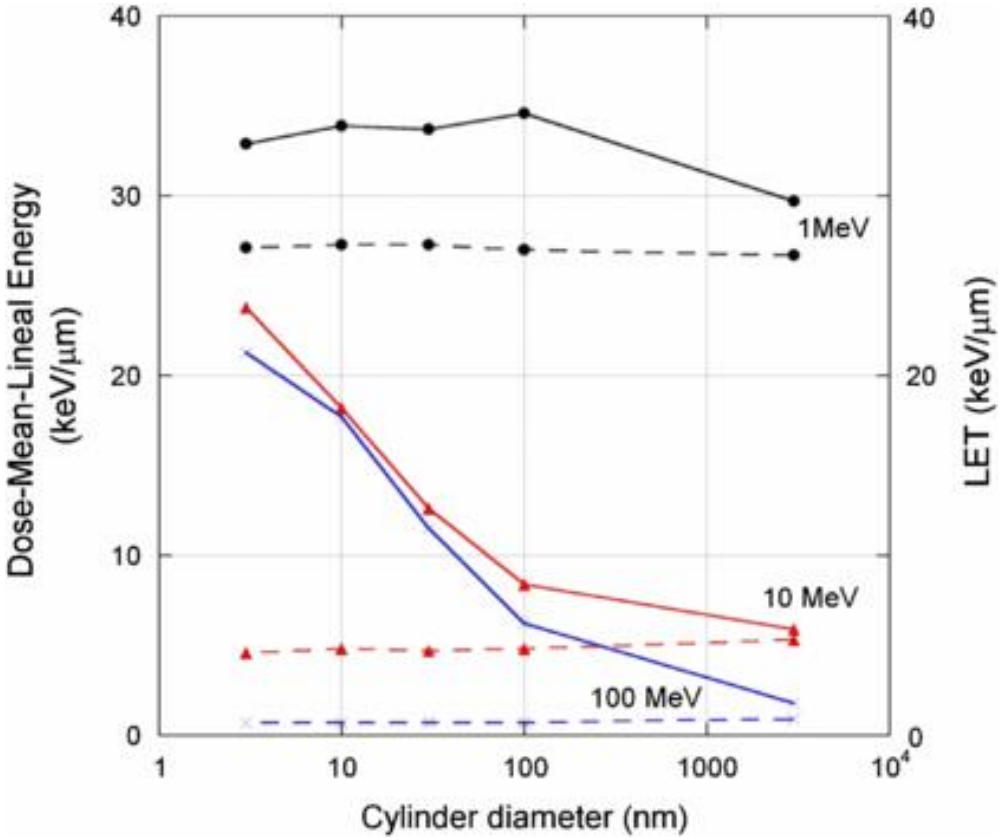

Figure 8: Comparison of LET_∞ and $\bar{y}_D$ for protons of different energies as a function of the diameter of cylindrical sites. Figure from [10]. The dotted lines are LET_∞, while the points with lines are $\bar{y}_D$.

2.2.3 Micrososimetry in practice

TEPC

In the field of microdosimetry, tissue equivalent proportional counter (TEPC) is seen as the gold standard for in-practice measurements [16]. The TEPC is a gaseous ionization detector consisting of a metal cylinder/sphere. The inside is filled with gas, and there is a wire (anode) going through the middle. If a voltage is applied, it creates an electric field around the wire. The electric field with the distance r from the wire can be described as [18]:

$$E(r) = -\frac{V_0}{r \ln\left(\frac{a}{b}\right)} \tag{2.23}$$

where a is the radius of the wire, and b is the radius of the cylinder/sphere. The electric field is therefore the strongest close to the wire. For proportional counters, the voltage must be held constant. When the gas volume gets radiated, an electron-ion pair is created. Because of the electric field, the electron travels towards the anode wire in the middle of the volume. The positive ion travels towards the metal wall. If the voltage is too low, there is a possibility that the ion-electron pair recombines before they are separated by the electric field. With high voltage, the electrons get accelerated enough to liberate other electrons close to the wire. This makes an avalanche-effect, which gets bigger the closer you get to the wire. This amplifies the signal, but the output signal is still proportional to the initial ionization. The concept of proportional counters is shown in Figure 9.

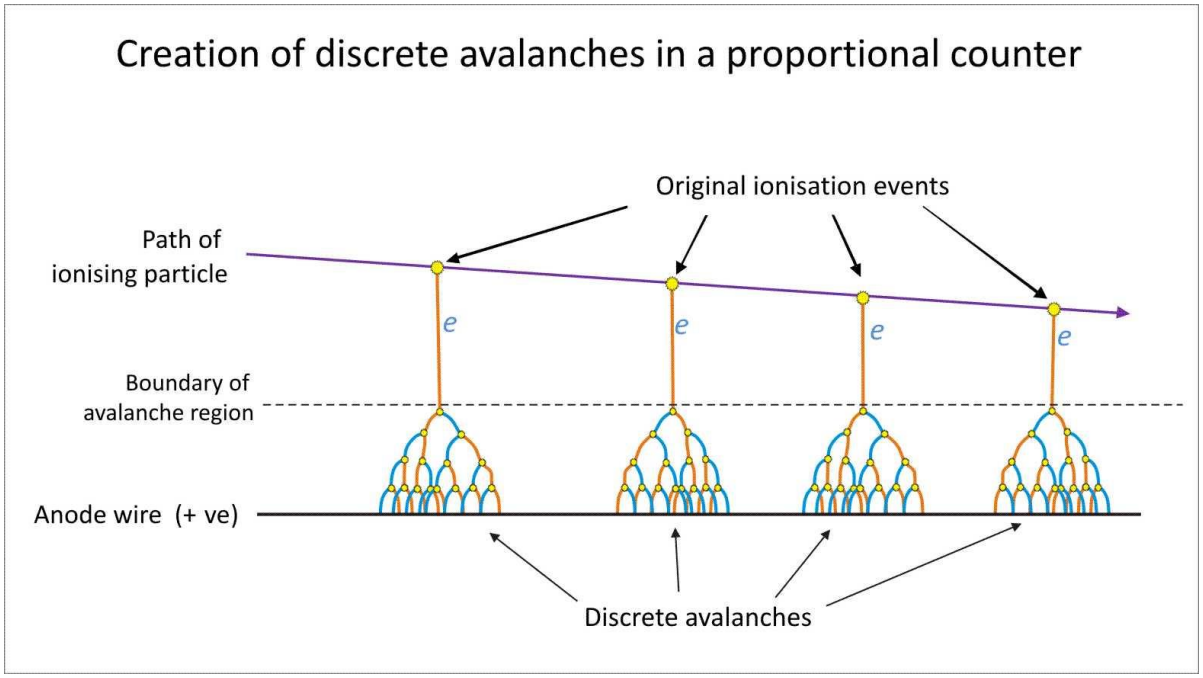

Figure 9: shows the concept of avalanches in a proportional counter. From [19].

The typical detector sizes ranges between 10-150mm in diameter. By lowering the pressure of the gas, smaller volumes can be simulated. By manipulating the pressure, the energy deposition for the simulated tissue and the detector gas can be made equal [16]:

$$\Delta E_t = (S/\rho)_{t\rho_t d_t} = (S/\rho)_{g\rho_g d_g} = \Delta E_g \quad 2.24$$

S/ρ is the mass stopping power, ρ is the density and d is the diameter. The shapes must be equal for the equation to be valid. The density of the gas can be set to:

$$\rho_{gas} = \frac{1}{d_g/d_t} \quad 2.25$$

This holds only if the atomic composition of the gas and the tissue is equal, and the mass stopping power is independent of the density. In addition, the tissue density must be set to unity. The Behte-Bloch formula shows that the mass stopping power is dependent of the density, but it can be ignored for particles with energies below the 1 GeV/u range [16].

The density of the walls in the TEPC are considerable higher than for the gas inside the cavity, which leads to some artifacts when measuring, called wall effects [16]. The delta-ray effect, as an example, happens when a secondary electron is produced in the wall before the primary particle enters the sensitive volume. Both particles are then going through the sensitive volume, creating an artificial high energy deposition that would not happen if the walls had the same density as the gas cavity.

2.3 Radiobiology

The main goal of radiotherapy is to kill cancer cells, and this is done by destroying the cell's DNA. DNA is formed as a double helix, and the main damage processes in radiation are divided into single-strand breaks (SSB) and double-strand breaks (DSB). DSBs are considered lethal events, though they can be fixed by the cell's repair mechanisms. When the cells cannot repair the DNA correctly, this leads to cell death [20]. The number of SSBs and DSBs is correlated to the number of ionizations that happens inside the cell nucleus, where the DNA is located. This indicates that the biological effect is dependent on ionization density, which can be described by for example LET or lineal energy. A higher LET will induce a higher number of

ionizations, which will result in a higher probability of cell death. Sørensen, et al. found an almost linear dependence between the relative biological effect (RBE) and LET in cellular experiments, as shown in Figure 10.

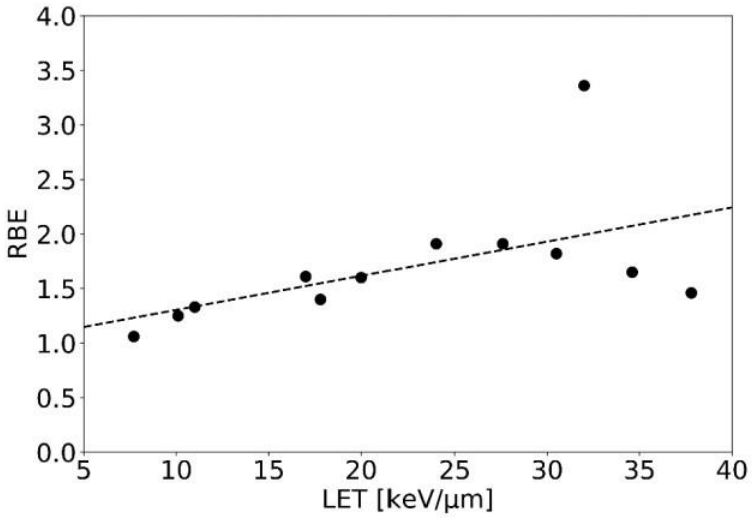

Figure 10: $RBE_{SF=0.1}$ for V79 cells irradiated with protons. From [15]

RBE describes the difference in biological effectiveness between two radiation types. But the biological effect is not only dependent on the ionization density of the radiation, but has several other dependencies, including cell type, beam type and tissue depth, among others. In clinical proton therapy, the proton RBE is set to 1.1, while studies seem to point towards that the RBE actually varies along the track. Therefore, multiple models have been constructed with the goal of quantifying the variation of the RBE. At the crossing point between biology and physics, it is almost impossible to find a universal model that works in every situation. Because of this, several models have been made, focusing on different aspects of the biological effect.

2.3.1 The Linear Quadratic model

The linear quadratic model is a relatively simple model, describing the fraction of cell survival when given a specific dose D. The survival fraction, SF, is given as:

$$SF(D) = e^{\alpha D + \beta D^2} \qquad 2.26$$

α & β are radiosensitivity parameters depending on the cell type. At low doses, the linear, α-dependent term dominates, while at higher doses the quadratic, β-dependent term takes over. It is often thought that the α-term indicates damage from single event-hits, while the β-term indicates damages where two events hits kill the cell [15]. High and low LET beams will have different survival fractions at the same dose. This is shown in the figure under:

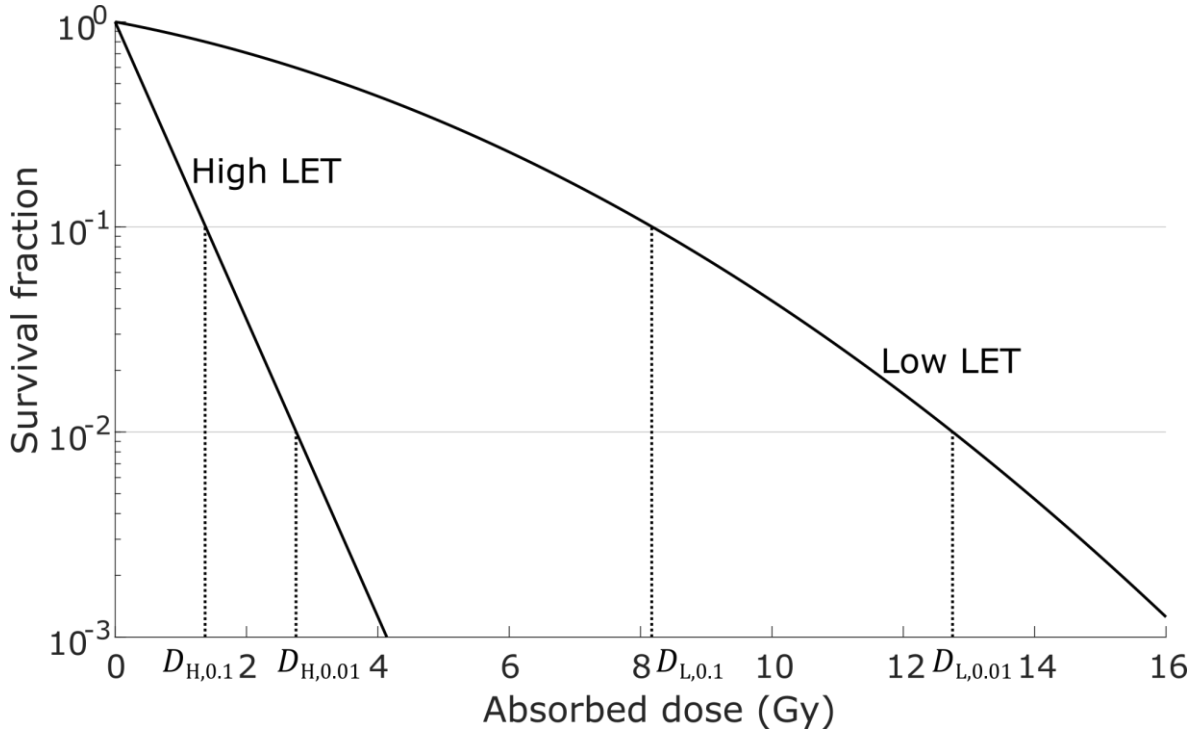

Figure 11: Survival fraction as a function of dose for low and high LET beams. From [16].

The RBE of the high LET-beam in figure Figure 11 can be calculated for the dose D_i, with the low LET-beam as the reference:

$$RBE = \frac{D_{L,D_i}}{D_{H,D_i}} \qquad 2.27$$

Where D_{L,D_i} is the dose given by the low LET beam, and D_{H,D_i} is the same for the high LET beam where both dose levels yield the same cell survival. It is important to understand that the RBE varies with the survival fraction, and thus $RBE_{SF=0.1}$ will be different than $RBE_{SF=0.01}$. In other words, the RBE is dose-dependent. Radiotherapy

is often given in multiple fractions, to let the healthy tissue recover. The biological effect of *n* dose fractions with fraction dose *d* can be written as [12]:

$$BE = \alpha nd + \beta nd^2 \tag{2.28}$$

which can be rewritten as (with $D = nd$):

$$BE = \alpha D\left(1 + \frac{d}{\alpha/\beta}\right) \tag{2.29}$$

α & β-values are often referred to as a ratio, called the α/β-ratio. Different cell types do have different α/β-ratios. This means that the different cell types will not have a similar response to the same beam/dose, which includes cell survival and regeneration. This can be exploited to find other fractionation schemes which will have a greater effect on the cancer cells, while letting the normal tissue around heal more. The biological effect formula can be used as well to find the weighting factor/RBE between a reference beam (γ) and another beam type:

$$W_b = \frac{D_\gamma}{D_b} = \frac{\alpha_b\left(1 + \frac{d_b}{(\alpha/\beta)_b}\right)}{\alpha_\gamma\left(1 + \frac{d_\gamma}{(\alpha/\beta)_\gamma}\right)} \tag{2.30}$$

Even though the equations 2.28 and 2.30 both use α and β, they are not necessarily equal [12]. Both alphas are however heavily dependent on radiation quality, and a common interpretation of this is that the alphas represent damage due to single events. β, on the other hand, is sometimes thought to be independent of radiation quality.

2.3.2 Proton RBE models

There are many different models trying to compute the variable RBE for proton therapy. Though the goal is the same, there are many ways to construct models,

focusing on different aspects of the radiation. The different models can be sorted into three categories: plan-based models, mechanistic models, and phenomenological models. The plan-based models are often simple, using quantities like dose and LET extracted from the treatment plan, instead of experimental data [15].

With a more biological approach, the mechanistic models try to quantify RBE basing the model on the biological damage processes that deactivate the cells, as well as the repair mechanisms of the cell. For these type of models, the energy depositions at micro- and nano-scales are of interest, and therefore microdosimetry is a convenient tool for these types of models. These models are hard to construct, as the cell's biological responses to radiation is yet to be understood fully. In addition, to connect a micro- or nanoscopic quantity to a macroscopic quantity is a hard and complex problem [21]. Examples of mechanistic models is the Microdosimetric Kinetic Model (MKM) and the Local Effect Model (LEM). The MKM is describing RBE with the help of microdosimetric measurements, using energy distributions inside sub-volumes of the cell to describe death. $\bar{z}_D$-measurements in sites in the range of 1 μm are most commonly used in the MKM. The LEM describes the cell death to be caused by lethal damaging events (lesions) inside small sub-volumes of the cell, and that the lesion types and event density are independent of radiation type and quality [12]. Both of these models studies biological effects at micro/nanometer levels.

Instead of focusing on the mechanical processes for radiation damage and repair, the phenomenological models are built by fitting a model to in vitro data from experiments on cell survival [15]. Most of the phenomenological models are based around the LQ model, and using α and β-constants for different cells for photon radiation as well as LET to predict the protons α and β-values [7]. Examples of phenomenological models include the McNamara-model and the Rørvik-model, both applicable for multiple cell-lines [22].

2.3.3 Microdosimetry and RBE

A large amount of the phenomenological models use LET to predict RBE for ionizing radiation, including protons and larger ions, like carbons. LET may be a relevant metric for modeling RBE for proton beams, with several models connecting LET to LQ-parameters like α and β [23]. However, LET is, as discussed in 2.2, a macroscopic quantity that only describes energy loss for a particle, not where the energy is deposited. The target of irradiation in cancer therapy is to inactivate the cells by damaging DNA, and there are therefore discussions if the energy depositions on a micro- and nanoscopic scale are more relevant for modelling RBE than LET. The experimental equivalent to LET, lineal energy, is dependent on the mean chord length, and thus also the site size. There are debates as to which site sizes are the most relevant to most accurately model the RBE of the proton beams. Typically, site diameters used in experimental microdosimetry are around 1-2 µm, but several studies point to smaller volumes to be more biologically relevant.

As an example, let's look at Lindborg et al. [10]. In 2013, they studied $\bar{y}_D$-values for several radiation types, including 60 Co photons, 100 KV X-ray, a proton SOBP , a carbon SOBP and a neutron beam. The site sizes used were ranging from 3-100 nm. Having the Co-60 beam as a reference, they calculated the $\bar{y}_D$-ratios for all radiation types and site sizes. They were compared with α-ratios from early-reacting tumors ($\alpha/\beta = 10$), and it was found that for all radiation types, the $\bar{y}_D$-ratios were similar to the α-ratio for the same radiation type with site sizes between 10-15 nm. The study concluded that these site sizes may be optimal for estimating radiation quality.

Villegas et al. [9] used microdosimetric distributions and quantities, including $f(y)$ and y_F, to try and fit a cell survival function to experimental LQ-values. This was done for multiple radiation types, including a Co-60 beam, low energy proton beams and a carbon beam. With site diameters ranging from 10 nm to 12 µm, they found that for 4.97 MeV proton beam, the optimal site size for their model was found to be between 1 and 0.5 µm. The optimal site sizes varied with both radiation type and

energy. They concluded that various models could benefit from microdosimetric estimates using site sizes that correlates with the size of the biological target in question.

These examples validate the belief that there are important biological targets both on cellular and sub-cellular level that are desirable to consider when constructing biological response-models.

2.4 Monte Carlo simulations

2.4.1 Basic principles of Monte Carlo simulations

As mentioned earlier, interactions between photons/charged particles and a medium are stochastic. This means that it's not possible to predict when and how every single interaction unfolds. To try and predict stochastic and random processes, a method called "Monte Carlo methods" is often used. The Monte Carlo (MC) method uses pseudo-random generated numbers and probability distributions for processes to estimate the behavior of the specific system you are investigating.

The most central part of the Monte Carlo method is the probability density function (PDF). The PDF is the function that describes the possibilities for every event. When generating a random number between A and B, the PDF ($p(x)$) is uniform [24]:

$$p(x) = \frac{1}{B-A}\ , A \leq x \leq B \tag{2.31}$$

$p(x)$ tells us the probability to generate the number x. The cumulative density function (CDF) can be derived by integrating the PDF:

$$P(x) = \int_A^x p(x)\ = \frac{x-A}{B-A}\ ,\ \ A \leq x \leq B \tag{2.32}$$

The CDF describes the probability of to draw every number up to x. The unform distribution is one of the simplest example of distribution functions, while depending

on the system you are study, the distributions may be more advanced. Some systems requires multiple PDF's for different processes that are examined.

2.4.2 Monte Carlo simulations in radiation physics

When studying radiation physics, there is often a high need for large amounts of data to acquire reliable results. To produce real life experiments with actual particle beams and detectors are also both costly and time consuming, if you have a particle accelerator available. In the field of microdosimetry, detectors are also difficult to construct. The randomness of Monte Carlo simulations suits the stochastic nature of radiations particle with matter, and is a great tool for both research and clinical dosimetry. The PDFs of the Monte Carlo models are based on data or models from real life data.

For particle physics, there are two main types of MC models, condensed-history and track structure models. The condensed-history (CH) models normally calculates the energy loss of a particle between each step of the simulation, assuming the particle undergoes a larger amount of interactions along the step [25]. This method is efficient, and is a good choice for high energy particle transport. Because of the time efficiency, the CH-models are often used in macroscopic dosimetry, and are found in clinical dose planning systems. However, the accuracy for lower energy electrons are one of the weaknesses for this type of models. For medical physics, there are a few CH-based models that have been developed, including GEANT4 and FLUKA.

The track structure (TS) models, on the other hand, simulates every interaction possible for all particles, having an analogous transport method [25]. The TS-models consists of cross sections for every possible interaction type, making the tracking of electrons down to about a few eV possible. During the simulation, each interaction type is considered individually. This simulation method increases the accuracy of low energy particles, which is necessary to imitate the energy depositions at micro- and nanoscopic levels. TS models are difficult to construct, due to the required detail in

each physics process. In addition, the computational power required to use TS models are much higher compared to CH-models, leading to much longer simulation time. The GEANT4-DNA project is one example of a high detail TS-model that are meant for medical applications at micro- and nano-scales.

3. Methods

3.1 GATE/Geant4

3.1.1 GEANT4

GEANT4 is a Monte Carlo toolkit for particles moving through matter [26]. It is used in many different areas, including particle and nuclear physics, space science and also medical physics. GEANT4 is highly flexible, making the users capable of reproducing their own real life experiments as accurate as possible. Depending on the users need, physics models with different levels of accuracy can be chosen. Particles of energies ranging from eV to several TeV can be simulated, depending on the settings. GEANT4 is written and executed in C++.

One of the fundamental bricks of the GEANT4 simulation is the geometry. The user can define the geometry of the simulation, using both simple and complex geometric shapes. The volumes are stored in a tree structure, where a volume inside another volume is called a daughter. The largest volume, called the world volume, needs to contain all the other volumes. Inside the world volume, many different volumes can be created, with completely different properties. Each volume can be assigned with its own material, their own list of physics processes, and other settings. One or more volumes can be assigned to be “sensitive”, where interactions and energy depositions can be recorded.

GEANT4 simulates the traveling particles in a step-by-step approach [26]. The particle travels a finite step, and the energy loss over that step length is then calculated. The possibilities for every interaction type to happen is calculated, such as inelastic or elastic scattering, secondary electron production, etc. The step length can be chosen by the user, and the lower the step length, the higher the accuracy. Shortening the step length leads to higher computational demands, increasing the time of the simulation. An evaluation of accuracy vs simulation speed is needed by

the user for each application. For microdosimetric purposes, accuracy in the micro- or even nano-meter scale may be desirable.

GEANT4 have a handful of physics lists to choose from, depending on the use case of the simulation. The physics lists are divided in to 3 main groups: particle physics, medical physics and space physics [27]. Each of these 3 areas of physics has different requirements for accuracy at different energy ranges. In particle physics for example, models with accuracy in high energy ranges for hadrons is important to simulate a real CERN-experiment, while studying a Bragg Peak in medical physics require a more refined low-energy model.

There are several techniques to improve the efficiency of the simulation and reduce computational time. One of these variance reduction techniques are production cuts. The user can define a production cut, meaning that the simulation will not produce secondary particles (such as electrons) that have an initial energy below the cut limit, and the energy is set to be deposited locally. If a secondary particle is supposed to be produced in a volume with no accuracy requirements, and it does not have the range to reach a high accuracy volume, there is no need to track that exact particle. Assigning different production cuts to different volumes, the simulations efficiency can be improved significantly without having to reduce the accuracy.

3.1.2 GATE

GATE [28] (GEANT4 Application for Tomographic Emission) is a general purpose Monte Carlo program based on GEANT4 for medical physics, in particular nuclear-based imaging systems [29]. It was made because of a need of a user-friendly program for Monte Carlo simulations in Positron Emission Tomography (PET). Using the GEANT4-toolkit as a base, the openGATE collaboration started to build applications on top of the MC toolkit. One of the main purposes of GATE is to make the applications of GEANT4 available to the medical physics community, without needing any C++ experience. GATE uses a simple macro-file structure to set up the

simulations. The user can either build the simulation line-by-line interactively, or write larger scripts to make the simulation repeatable.

GATE has inherited GEANT4's way of creating geometries, and also creating sensitive volumes (SV), where the output of the simulation is recorded. The output files from GATE can be set to either ASCII-files or ROOT-files. After each simulation, a run-file is made, containing how many primary particles/decays have been simulated. The two other raw file types that's relevant for microdosimetry are singles- and hits-files. A singles-file contain one line for each event that deposits energy inside the SV, containing information such as energy deposited, which volume it deposited its energy in (useful if you have more than one SV), and ID of the event. The singles-file does not include any positional information inside the SV. The hits-file is a larger file, where every step the particle moves inside the designated SV creates a line in the file. Each line includes the X, Y and Z position of the particles steps, how much energy it deposited through the step, the ID of the event, and more.

GATE also adds different tools, called actors, that can interact with the simulations. The actors can both gather information from the simulations, and adjust the simulations while they are running. There are numerous different actors. The actors that accumulate information, processes the raw data and may give energy spectrums, histograms, etc. An example of a modifying actor is the Kill Actor, which is able to stop the tracking of particles when entering or exiting a specified volume. This can be used to shorten simulation time.

In this work, GATE version 9.1 along with GEANT4 version 10.7.3 was used.

3.2 Simulation setup

In this project, GATE was used to simulate monoenergetic proton beams of energies 1, 2, 5, 10, 20, 50 and 100 MeV in water. The goal of this work was to inspect and

examine microdosimetric features for these proton beams at spherical sites with diameters in the range 10 nm - 20 μm, and see how the site size influences microdosimetric spectra at various energies.

The simulations are divided into three different setups, where the SV is set to 0.5, 2 and 20 μm. This is done to save time as there are differences in accuracy need for large site sizes like 20 μm and the smallest site size studied, 10 nm. If the same settings that were used to simulate the smallest SV should be used in the largest SV, the simulations would require several weeks, if not months, to finish. The actual settings will be discussed below. The specific site sizes studied is shown in Table 2, alongside the SV size used to simulate these site sizes. 7 different proton energies combined with 3 different SV setups, meaning 21 individual simulations was executed.

Table 2: SV size and corresponding site sizes studied in the different setups.

SV size	**Site sizes studied**
20 μm	20, 10, 5 μm
2 μm	2, 1 μm
0.5 μm	500, 200, 100, 50, 20, 10 nm

3.2.1 Geometry and variance reduction techniques

The geometry of the three different setups built the same way, where the main difference is the size of the SV. The geometry of the simulation is built up by 3 water cubes, as shown in Figure 12. The first and largest volume is the ‘world’ volume. It is at the front surface of the world volume that the primary protons are generated. The world volume acts as a build-up region for delta-electrons to ensure charged particle equilibrium (CPE) in the SV. The size of the world-volume depends on the maximum

range of delta electrons for the specific energy being simulated, thus it depends on the proton beam energy, as discussed in 2.1.5. Taking 5 MeV protons as an example, the maximum range for delta electrons is 8.6 μm. The world volume is then set to cover at least 8.6 μm on every single side of the SV. If the SV size is 2 μm, the world XYZ-sides is set to be at least $2 * 8.6\ \mu m + 2\ \mu m = 19.2$ μm. The rest of the secondary electron ranges used to calculate the world volume size for each proton energy is shown in Table 1. In the world volume, the production cuts are set to 5 μm for protons and electrons.

The second volume is called the 'low-cut region', where the production cuts is set to 0.1 μm. Since the production cuts for electrons are set to 5 μm in the world volume, distance between the SV and the world must be at least 5 μm, so that all possible secondary electrons produced in the world can reach the SV. In the 2 μm SV example, the low-cut region is set to have sides with length 12 μm. With these settings we avoid simulating many secondary electrons that could never reach the SV, while not excluding those who could. By dividing the build-up region in 2 volumes with different production cuts, the simulation time is cut drastically

The smallest volume of the simulation is the SV, where the three different sizes were used, 0.5, 2 and 20 μm. The SV is where the simulated particles interactions are recorded. Once entering this volume, every step a particle travels is written to a hits-file. The production cuts in the SV were set to 1 nm. It is also set an extra margin of 0.1 μm on each side of the SV, for the same reasons as for the low-cut region. For the 0.5 μm SV, the actual volume in the simulation is set to 0.7 μm. For the actual microdosimetric calculations, the 0.1 μm margin on the edge of the SV is not considered.

The source of the protons is a uniform beam, starting at the front edge of the world volume. The radius of the beam is set large enough to cover the whole world volume. The number of protons simulated in each simulation is chosen to have at least 20 000 registered events inside the SV.

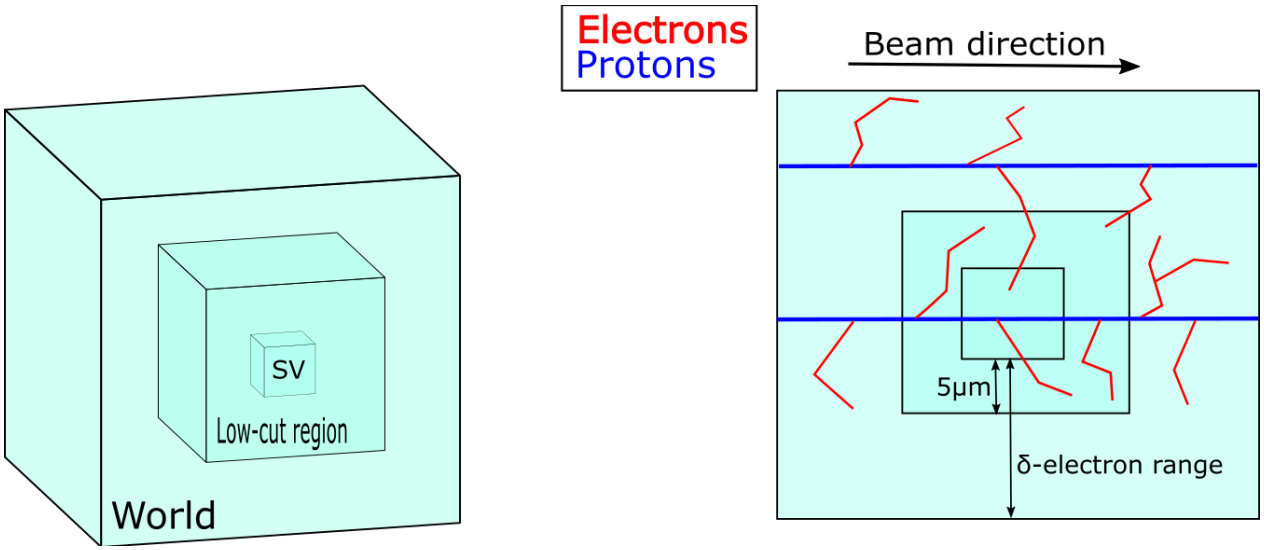

Figure 12: Left: a simple illustration of the simulation setup. The world and low-cut region is made to ensure CPE in the SV. Right: shows 2 proton tracks with corresponding secondary electrons.

3.2.2 Physics lists

For medical physics, there are a couple of physics lists/models that are recommended by the GEANT4 collaboration. In this work, the emlivermore, QGSP_BIC_HP and G4DNA is chosen. For a more detailed explanation of the 3 physics lists, the reader is referred to GEANT4s Guide for Physics list [30] and Physics Reference Manual [31].

The low energy Livermore Model (emlivermore in GATE) implements physics processes like photoelectric effect, Compton scattering, Rayleigh scattering, gamma conversion, bremsstrahlung and ionization for electrons and hadrons [31]. Livermore is a condensed-history model. The model is following the particles down to 10 eV, and is one of the recommended physics list for medical applications. The Livermore physics list implements higher accuracy for lower energy electrons than the standard GEANT4 physics list, which is desirable for microdosimetric purposes. However, reduced accuracy for electrons under 100 eV are to be expected for this physics list [25]. QGSB_BIC_HP is a physics list with higher accuracy for hadrons. The elastic, inelastic and capture processes for hadrons are built on top of sets of cross sections,

which gives a higher level of detail. This physics list uses the standard GEANT4 electromagnetic models.

The GEANT4-DNA project [32-35] started in 2009 with a goal to further develop the GEANT4 toolkit to improve accuracy for ionizing radiation at sub-cellular scale. The physics list G4DNA is meant to be used for simulation of liquid water, and has advanced models for both protons and electrons, as well for several other ions, including carbon and helium (alpha particles). The Track Structure model is based on sets of cross sections for each interaction type, and tracks electrons down to 0 eV. This list was therefore included, and is expected to give the best accuracy, but also the longest simulation time.

Choices of physics lists

For the 2 and 20 μm setup, the combination of Livermore and QGSP_BIC_HP physics models were chosen, as it fits the accuracy needs for site sizes ≥ 1 μm. The hadronic properties of the QGSP_BIC_HP and the accuracy for low-energy electrons from the Livermore list is thought to be a good fit for microdosimetric simulations in this range.

GATE has its own algorithm to calculate the step length of the simulated particles. When using this, it was found that the steps inside the SV was too long to accurately compute microdosimetric values for the smallest site sizes we wanted to study. Another thing to consider is the increase in data, because every step is equal to one line in the hits-file. The maximum step lengths inside the SV for the 20 and 2 μm setups was set to 200 and 50 nm, respectively, and was found to be a good balance between accuracy and file size. It was thought that the step length should be at least 1/10 of the site diameter in question. Any other settings are identical between these two setups.

For the 0.5 μm setup, the G4DNA physics list (called emDNAphysics in GATE) was chosen as the extra accuracy is needed at the sub-micro level. First trying with

emlivermore and QGSP_BIC_HP, it was observed that those physics model couldn't handle shorter step lengths, that is required studying site diameters of 10 nm. However, the computational power required to use G4DNA was found to be extremely high, and it was not possible to use this physics list in both the build-up region and the SV, as this would take several months. In order to cut down computational time, the simulations in this setup was split up in two, with different physics models.

In the first step of the simulation, the Livermore and QGSP_BIC_HP models was used in the build-up region ("world" and "low-cut region"). The build-up region does not require the same accuracy as the SV, as its sole purpose is to ensure CPE in the SV. Using a phase space actor in GATE, the particles energy, position, direction and type is stored in a file when it enters the SV. This file from the first step was used as a source to the next step of the simulation. The second step of starts at the edge of the SV, where the track of all the particles touching the edge of the SV in step one is continued, but with the G4DNA physics list. A 1 nm step length was chosen for this setup, such that the step length was 1/10 of the smallest site diameter studied (10 nm).

3.3 Data handling and analysis

For this work, the hits-file is used as the main data source. The hit-file records every step of every charged particle within the SV, with variables including event ID, X, Y and Z coordinates, energy deposited along the step, particle type, and more. Typical file sizes in this thesis ranges from 0.5GB up to almost 20GB, depending on simulation settings like step length and SV size.

To analyse the microdosimetric features of the proton beams, only a couple of the columns in the hit-file is needed. The "file compressor.ipynb" (Appendix A) script reduces the file size by removing every column but the step's event ID, coordinates, energy deposited and the particle type. In addition, every step that did not deposit any energy is removed, because these steps do not contribute to any energy depositions.

Finally, data from the build-up region for delta electrons (0.1 μm on each side of the cube) is removed. These steps made the largest hit-file shrink from 18.5GB down to under 700MB. This is crucial to computational speed in the next part of the data handling. A flowchart showing the different steps of the data handling is shown in Figure 13.

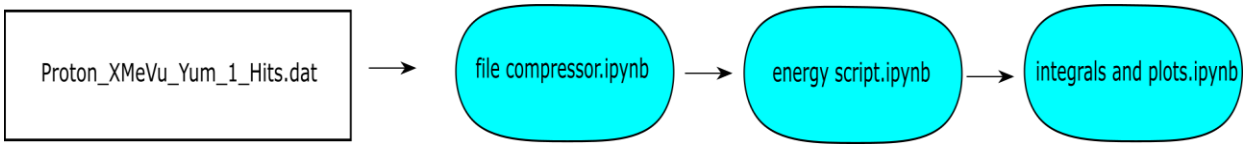

Figure 13: shows flow-chart of python scripts used to handle the raw simulation data for a simulation with proton energy X and simulation site Z.

3.3.1 Virtual site concept

In this work, we wanted to analyze spherical sites ranging from 20 μm to 10 nm. The simulations have a cubic SV where the data is recorded, having a side length of either 0.5, 2 or 20 μm. Instead of having one simulation for each energy and each site size, the concept of virtual sites was used to reduce simulation time and the total number of files. Taking the 2 μm SV setup as an example, analysing a 2 μm site, a virtual sphere with the diameter of 2 μm is put inside the cubical SV. For each event, the python script "energy script.ipynb" (Appendix B) then discards every part of the particle tracks that is outside of the sphere. The energy deposited by every single event (main proton + secondary particles) is then written down to a new file. Analysing 1 μm sites, the 2 μm cubic SV can fit 8 spheres with a diameter of 1 μm, as illustrated in Figure 14. The python script goes through every sphere the same way as previously described. One event can deposit energy inside more than one site, meaning the event can be used several times. Doing this, the amount of data is increased without increasing the number of protons simulated, meaning simulation time is drastically shortened.

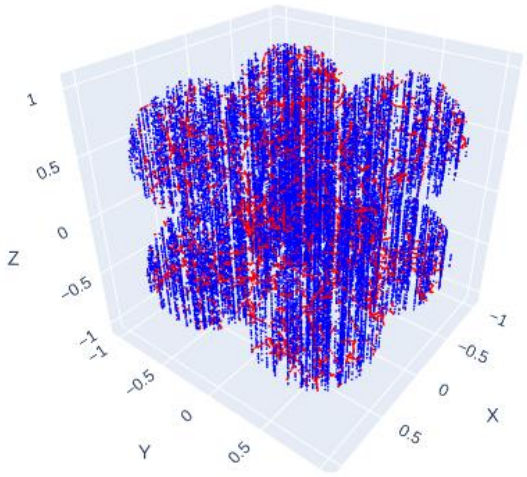

Figure 14: shows how virtual, spherical sites are made in a cubic SV to reuse data and save simulation time. Blue particles are protons, red particles are electrons. Here, a 2 µm cube can fit 8 cubes with a diameter of 1 µm.

This is done for every site size. For 10 nm sites in the 0.5 µm SV, the number of virtual sites is 125 000. This makes the simulation 125 000 times faster comparing to having 1 SV being a sphere with diameter 10 nm. The energy deposited per event in each site is then saved in a file. This file is then managed in the python script 'integrals and plots.ipynb' (Appendix C). In this script, the lineal energy y of each event in a site is then calculated. Further, the yd(y) distributions are computed, as well as the averages $\bar{y}_D$ and $\bar{y}_F$. The yd(y) distributions for each site diameter are then plotted together for each proton energy, and the $\bar{y}_D$ and $\bar{y}_F$ values are saved in a file.

4. Results

4.1 Simulation of proton tracks

Two proton tracks of energy 1 and 10 MeV, with belonging secondary electrons, can be seen in Figure 15. This illustrates that the 1 MeV proton produces significantly more secondary electrons than the 10 MeV proton. The stopping power of 1 MeV protons is higher than 10 MeV protons, resulting in a higher energy deposition per distance traveled. In this simulation, the total energy depositions (including secondary electrons) for the 1 and 10 MeV in the SV were 53.7 and 12.9 keV, respectively.

Figure 15 also illustrates the difference in positional energy depositions between protons and electrons. The protons often travels in a straight line, depositing their energy along the trajectory. The low energy electrons, however, travels in a tortuous track. The momentum of a proton is vastly greater than the secondary electrons, due to both the large mass and energy difference. This makes the protons harder to deflect off the initial track than the electrons, when they are colliding with the other electrons and nuclei. It is noticeable that the 1 MeV proton track is slightly bending, making a lateral deflection of ~16 nm, while the 10 MeV proton does not deflect at all.

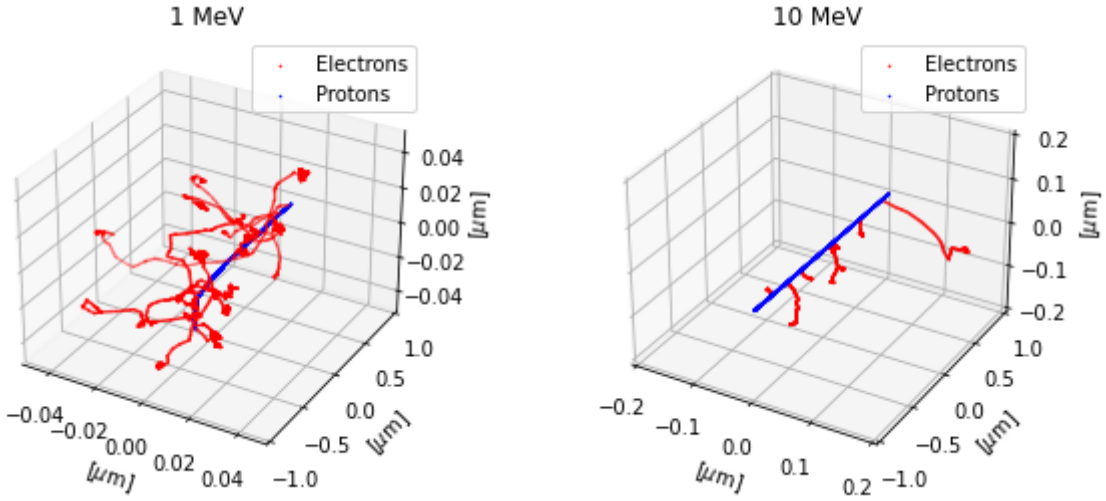

Figure 15: shows particle tracks for 1 and 10 MeV protons and their secondary electrons through a 2 µm cube of water. Note that the two figures have different axes.

4.2 Microdosimetric spectra

Figure 16 shows the lineal energy distributions yd(y) for monoenergetic protons ranging from 1 to 100 MeV, with spherical sites with diameters ranging from 20 μm to 10 nm. Note that the scales of each plot is different for better visibility. With increasing energy, the peak of the yd(y) distribution is shifting sligthly to a lower y, and the distributions generally becomes wider. This implies that the lower energy protons deposits more energy per event, and that there is higher variance in how much energy each event deposits for higher energy protons. For lower energies, the distributions have a narrow peak for the 20 μm sites. For the 5 MeV protons, the 20 μm peak is located at around 10 keV/μm, while its located at around 0.6 keV/μm for 100 MeV protons.

A shift in the lineal energy distributions with decreasing site diameter is also clearly visible. In general, reducing the site diameter results in a broader dose distribution. The distributions are also shifting towards higher y with lower site diameters. The $\bar{y}_D$ for each site size is also shown in Figure 16, and it shows that the $\bar{y}_D$ increases with smaller site sizes.

Another effect seen at higher energies in Figure 16 is big fluctuations for high y depositions. The interactions with large energy depositions happens infrequently for high energy, which creates the non-continuous, sporadic events at the tail of the distribution. The same phenomenon was found by Liamsuwan, et al. [36].

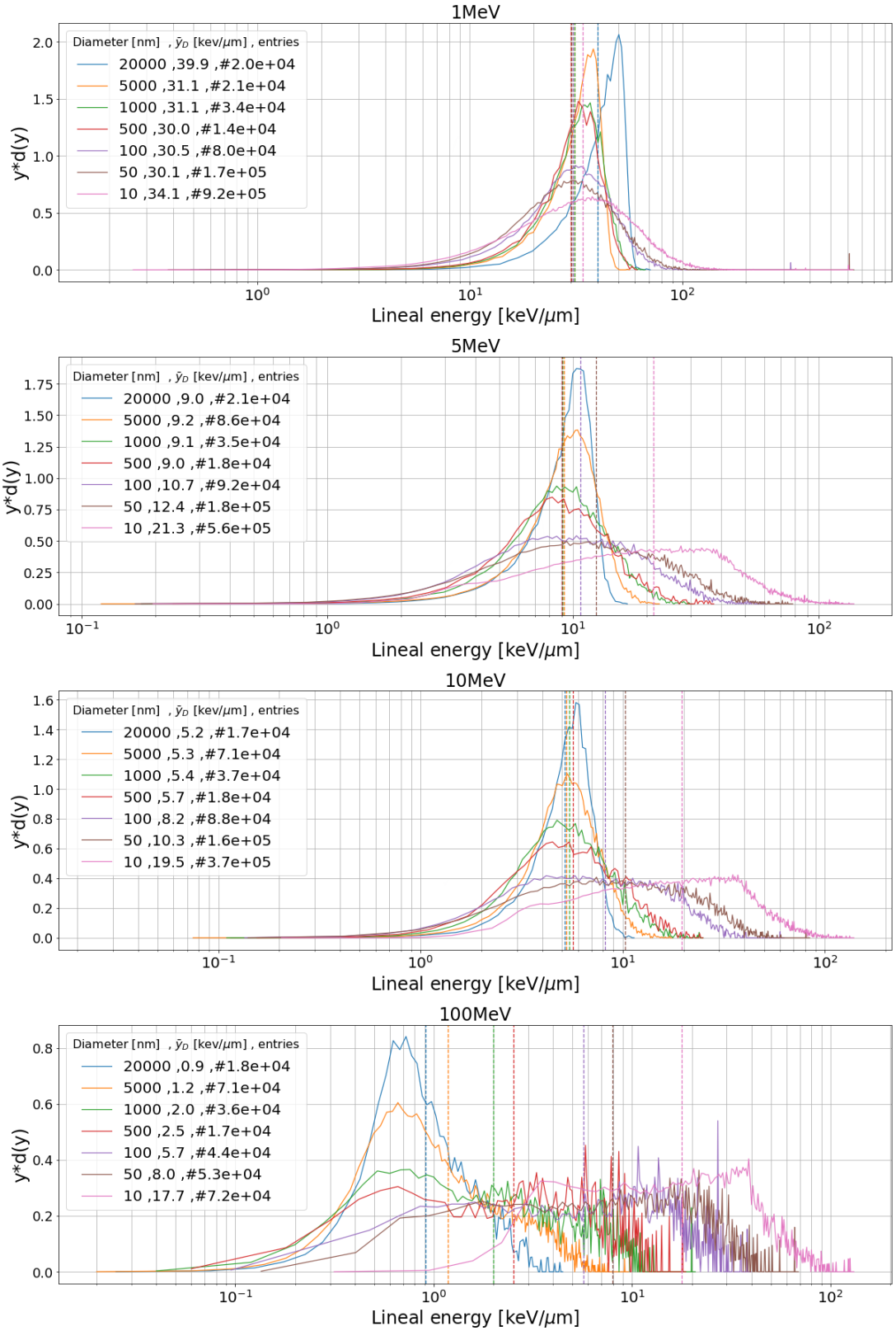

Figure 16: shows yd(y) distributions for 1-100 MeV monoenergetic protons with site diameter from 20 μm to 10 nm, simulated with. Number of events for each site size is shown in the legend, and $\bar{y}_D$ is shown as vertical, dashed lines.

4.3 Microdosimetric metrics

Table 3 & Table 4 shows the calculated $\bar{y}_D$ and $\bar{y}_F$ for protons for 1-100 MeV with site diameters ranging from 20 μm to 10 nm. $\bar{y}_D$ is also shown in Figure 17 and Figure 18, as a function of site diameter and proton energy, respectively. The same plots is shown for $\bar{y}_F$ in Figure 19 & Figure 20. The data points for d ≤ 0.5 μm was calculated from the G4DNA-simulation, while the data points for d > 0.5 μm was simulated using the emlivermore & QGSP_BIC_HP physics lists.

Comparing the dose mean lineal energy for the same proton energy in different site sizes in Figure 17, $\bar{y}_D$ is generally higher in lower site sizes than for larger sites. For 10 MeV protons, $\bar{y}_D$ changes from 5.4 to 19.5 keV/μm when changing the site diameter from 1 μm to 10 nm. $\bar{y}_D$ does not change as much for larger sites, as it's almost flat from 20 to 1 μm. Both these trends trend was seen for almost all energies.

An exception from the trends can be seen for lower energy protons in larger site sizes. For 1 MeV in particular, the $\bar{y}_D$ increases in the range 2-20 μm. The range of a 1 MeV protons in liquid water are roughly 24.6 μm [13]. This means that in large sites, the particles loses a significant amount of their initial energy inside the sites. The stopping power of protons increases with decreasing energy (down to ~0.08 MeV) [13], contributing to a higher energy deposit the longer the proton travels. This means that for larger site diameters, the $\bar{y}_D$ will rise up until the diameter is larger than the range of the 1 MeV proton.

Looking at Figure 18, the $\bar{y}_D$ is increasing with decreasing energy, independent of site size. The difference is bigger at the lower energies. The steep increase in $\bar{y}_D$ between 2 and 1 MeV is in the range 8-12 keV/μm at every site size, except for d > 5 μm. Between 100 and 50 MeV, the increase in $\bar{y}_D$ is in the range 0.1-0.5 keV/μm. This can be explained by the LET values in Table 3, showing that the

stopping power of protons in liquid water changes rapidly for low energies, while only small changes occur in the range between 50 and 100 MeV. Figure 18 also shows that the LET is lower than $\bar{y}_D$ for all site sizes explored in this work. $\bar{y}_D$ converges towards LET with larger site diameters. It is also noticeable that the difference between LET and $\bar{y}_D$ for larger site diameters increases with decreasing proton energy.

The frequency mean lineal energy $\bar{y}_F$ is also dependent on both proton energy and site diameter, as shown in Figure 19. For higher energies, the $\bar{y}_F$ is almost flat at larger site diameters, and increasing significantly with site diameters lower than 100 nm. For lower energies, however, $\bar{y}_F$ increases significantly with larger site diameters. This is especially noticeable for the 1 and 2 MeV protons. Looking at Figure 20, it is noticeable that $\bar{y}_F$ increases with lower proton energies. For higher energies, the smaller sites measures the highest $\bar{y}_F$, while this changes with lower energies. For 1 MeV, the 20 μm holds the highest $\bar{y}_F$, while the smaller sites has the lowest values. Comparing $\bar{y}_F$ to LET, Figure 20 shows that LET comparable to different site diameters for different proton energies. At 100 MeV, the LET is closest to $\bar{y}_F$ for site diameter between 0.02 and 0.05 μm. At 1 MeV, the LET value is close to the $\bar{y}_F$ values for 1 and 2 μm sites. $\bar{y}_F$ is generally lower than $\bar{y}_D$ for every proton energy. However, the increase in $\bar{y}_F$ for larger site sizes is steeper than for $\bar{y}_D$, especially for the lower energies, when comparing Figure 17 & Figure 19.

Table 3: shows $\bar{y}_D$ [kev/μm] values for protons with energy 1-100 MeV in spherical sites with diameter 20 μm to 10 nm. LET values are from [13].

	Proton energy [MeV]						
Site diameter [μm]	1	2	5	10	20	50	100
20.0	39.91	18.88	9.02	5.17	2.94	1.43	0.91
10.0	32.97	19.10	9.10	5.22	2.94	1.50	1.01
5.0	31.11	19.29	9.20	5.25	3.00	1.62	1.18
2.0	30.36	18.42	9.08	5.17	3.15	1.97	1.58
1.0	31.1	19.04	9.09	5.42	3.54	2.39	1.99
0.5	29.98	18.14	8.99	5.66	3.96	2.95	2.52
0.2	30.41	18.14	9.66	6.79	5.21	4.33	4.21
0.1	30.54	18.21	10.71	8.16	6.83	5.99	5.70
0.05	30.15	19.05	12.43	10.29	8.89	8.29	7.95
0.02	30.67	21.58	16.19	14.35	13.24	12.62	12.33
0.01	34.11	26.12	21.27	19.53	18.57	17.81	17.73
LET	26.08	15.86	7.91	4.57	2.61	1.25	0.73

Table 4: shows $\bar{y}_F$ [kev/μm] values for protons with energy 1-100 MeV in spherical sites with diameter 20 μm to 10 nm.

	Proton energy [MeV]						
Site diameter [μm]	1	2	5	10	20	50	100
20.0	34.90	16.57	7.74	4.31	2.37	1.09	0.63
10.0	28.93	16.65	7.69	4.25	2.32	1.08	0.63
5.0	27.31	16.51	7.59	4.18	2.28	1.06	0.63
2.0	26.01	15.03	7.17	3.91	2.20	1.04	0.61
1.0	25.94	14.84	6.93	3.86	2.18	1.04	0.62
0.5	24.18	13.71	6.45	3.65	2.07	1.02	0.61
0.2	22.99	13.00	6.24	3.60	2.12	1.13	0.78
0.1	22.07	12.52	6.15	3.69	2.32	1.47	1.15
0.05	20.76	12.14	6.31	4.11	2.92	2.20	1.90
0.02	19.47	12.28	7.59	5.88	4.95	4.32	4.01
0.01	20.08	14.13	10.33	8.94	8.14	7.54	7.25
LET	26.08	15.86	7.91	4.57	2.61	1.25	0.73

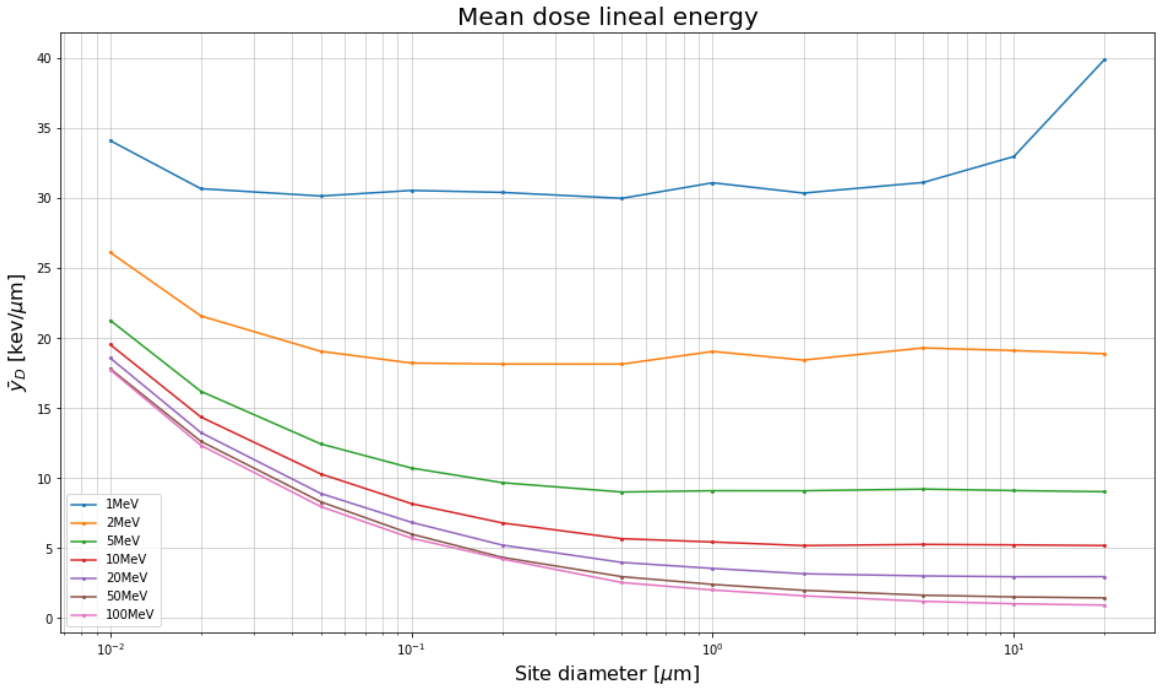

Figure 17: shows mean dose lineal energy $\bar{y}_D$ for 1-100MeV protons as a function of the site diameter.

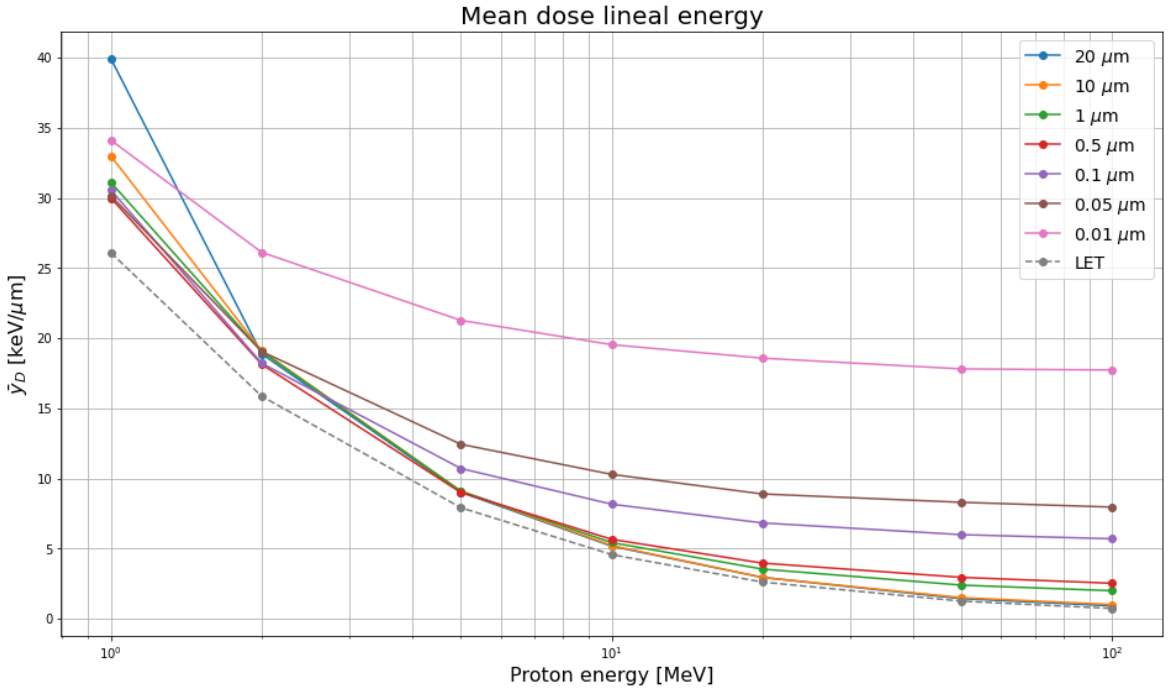

Figure 18: shows mean dose lineal energy $\bar{y}_D$ for 1-100 MeV protons in sites 20-0.01 µm, as a function of the proton energy. LET for protons in water is shown in stippled lines.

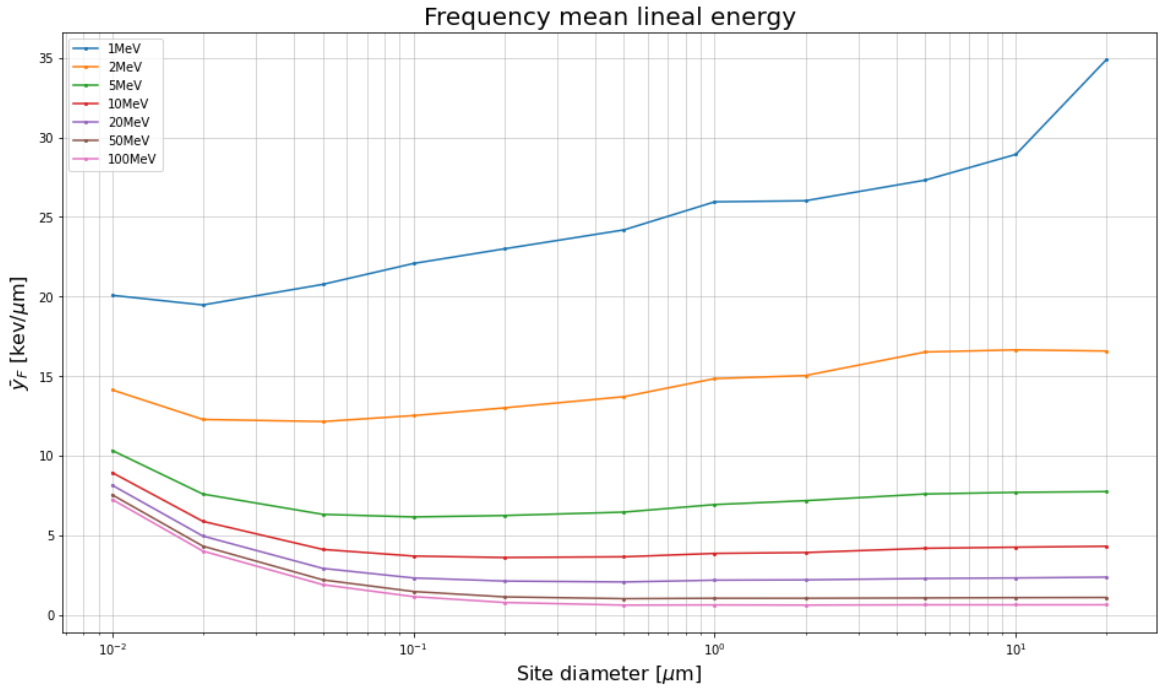

Figure 19: shows the frequency mean lineal energy $\bar{y}_F$ for 1-100 MeV protons in water as a function of site diameter.

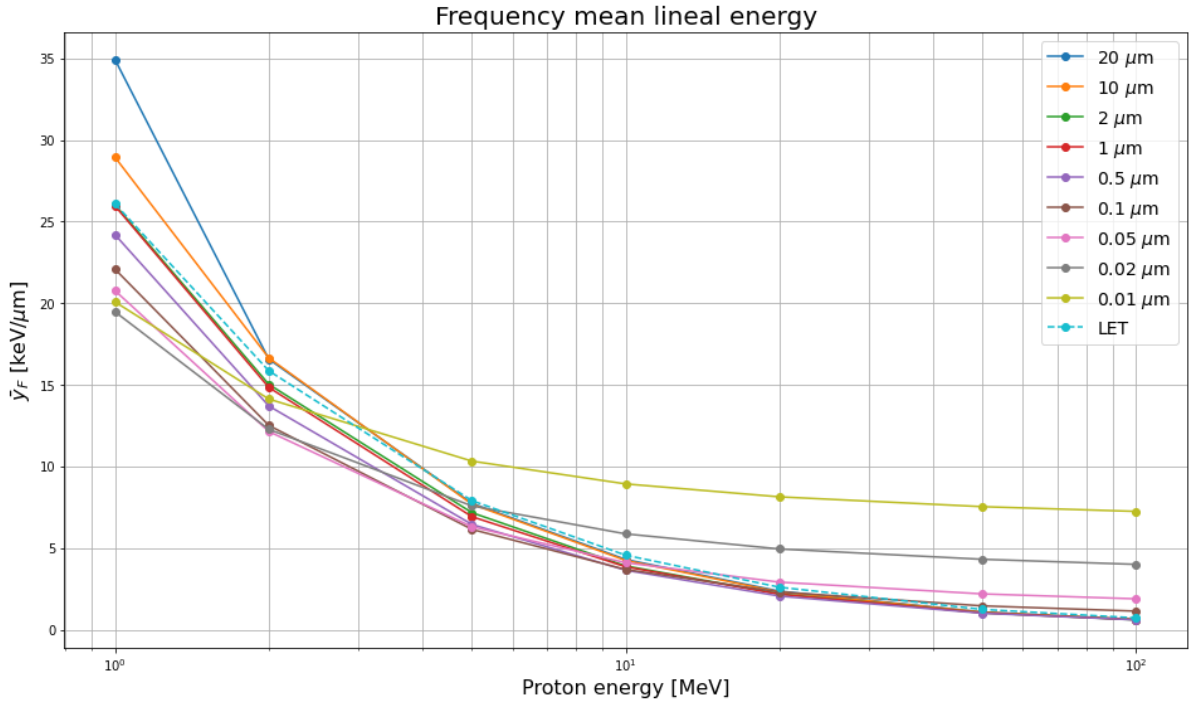

Figure 20: shows the frequency mean lineal energy $\bar{y}_F$ for 1-100 MeV protons in water as a function of proton energy. LET for protons in water is shown in stippled lines.

4.4 Comparison to other works

While no other publications covering a comparable range of site sizes and energies were found in the literature, several papers present relevant results for comparison. Similar work is done by Liamsuwan, et al. [36], using Monte Carlo methods to simulate monoenergetic proton beams in cylindrical sites. The paper includes energy ranging from 0.3 MeV to 300 MeV in 10 and 100 nm sites. Nijkoo, et al. [37] refers to calculations of microdosimetry parameters from an unpublished Monograph by Nijkoo himself. The table includes $\bar{y}_D$ values for protons with energies from 0.1-1000 MeV in a wall-less proportional counter with a diameter of 1 µm. Vassiliev, et al. used the GEANT4-DNA physics list to calculate microdosimetric values to model proton RBE [38]. $\bar{y}_D$ and $\bar{y}_F$ were calculated for protons from 0.1 to 100 MeV, for spherical sites with 2-1000 nm in diameter. The values gathered from these three papers were compared with the results of this work in Figure 21.

We can observe that for 10 nm, $\bar{y}_D$ from this work agrees well with values from Liamsuwan, et al. at low energies while the latter reports slightly lower values of $\bar{y}_D$-values above 5 MeV. Good agreement with results from Liamsuwan et al. was also seen for 100 nm. In comparison, results from Vassiliev et al. showed lower $\bar{y}_D$ for both 10 and 100 nm, while at 1 µm, results from this work matched both results from Vassiliev et al. and Nijkoo et al. relatively well. The differences are largest for higher energies. Comparing our results at 10 nm to Liamsuwan et al., the $\bar{y}_D$ calculated is about 2% higher at 1 MeV, while the relative difference increases to approximately 17% and 15 % for 50 and 100 MeV, respectively.

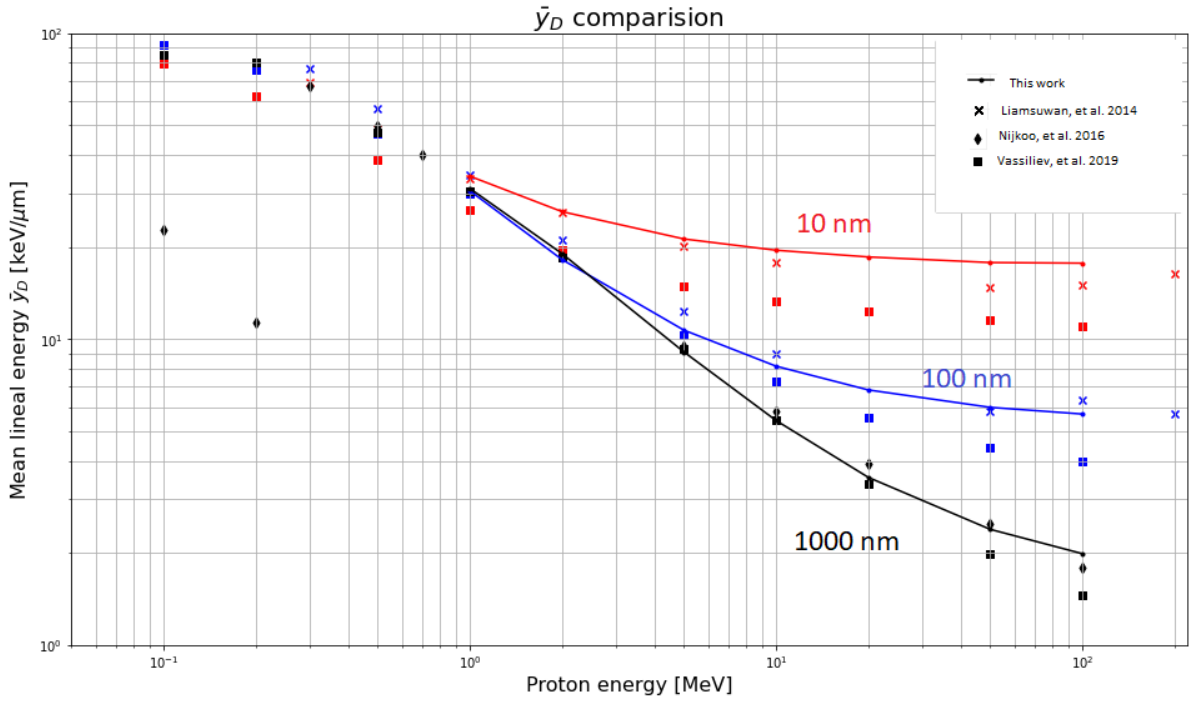

Figure 21: shows comparisons of $\bar{y}_D$ with other works for 10, 100 and 1000 nm and energies 0.1-200 MeV.

4.5 G4DNA vs emlivermore

The differences between the two physics lists used in the simulations was also explored, as it is unclear to what extent reliable results can be produced for small site sizes without applying the G4DNA physics. Using G4DNA physics list requires much higher amounts of computational power, and it is therefore of interest to use less CPU expensive physics lists whenever possible. The emlivermore combined with the QGSP_BIC_HP need a lot less computational power compared to the G4DNA. For the simulations with 2 μm SV, virtual sites were computed down to 100 nm to compare the $\bar{y}_D$ with the G4DNA simulation for site diameters 500-100 nm. The differences are shown in Table 5, and the relative difference ($\frac{\bar{y}_{D\,DNA} - \bar{y}_{D\,emlivermore}}{\bar{y}_{D\,DNA}}$) is shown in Figure 22.

Table 5: shows difference in $\bar{y}_D$ for the G4DNA and emlivermore + QGSP_BIC_HP physics lists (DNA-emlivermore) [keV/µm].

	Proton energy [MeV]						
Site diameter [µm]	1	2	5	10	20	50	100
	$\bar{y}_{D,DNA} - \bar{y}_{D,emlivermore}\ [keV/\mu m]$						
0.5	-1.70	-1.86	-0.34	-0.23	-0.23	-0.08	-0.13
0.2	-3.80	-4.37	-0.80	-0.52	-0.61	-0.43	-0.20
0.1	-7.21	-9.04	-1.83	-1.52	-1.51	-1.31	-1.54

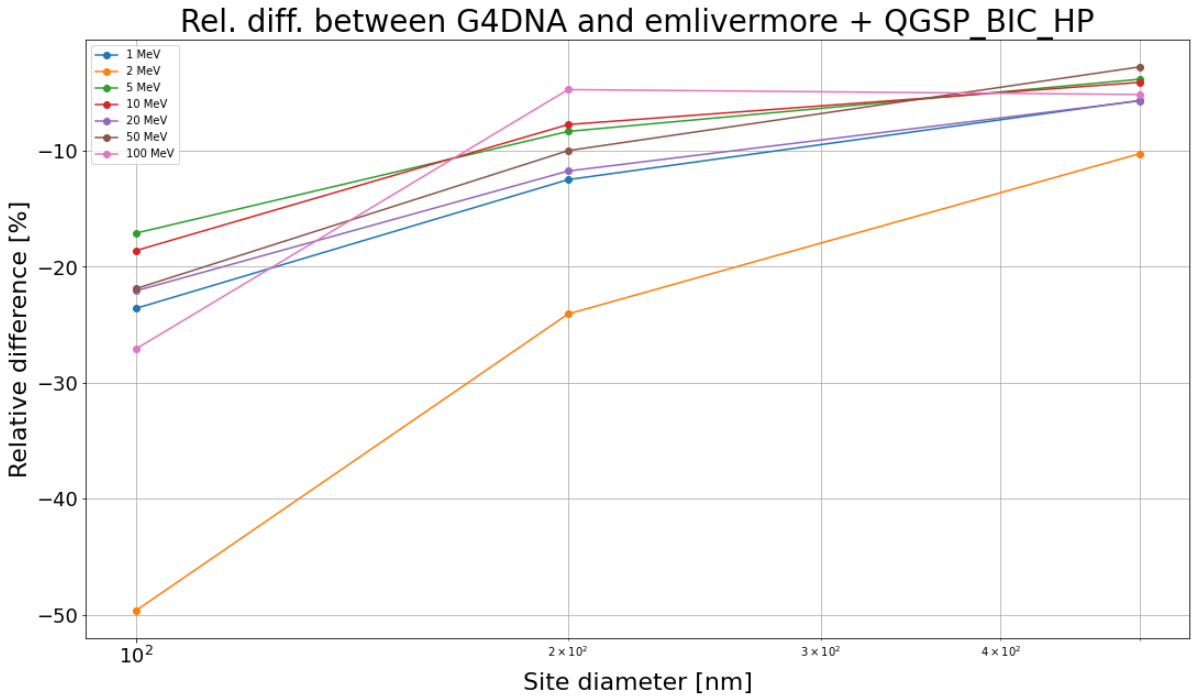

Figure 22: shows the relative difference between $\bar{y}_D$ values for the G4DNA and emlivermore + QGSP_BIC_HP physics lists. Negative values indicates higher $\bar{y}_D$ for the emlivermore + QGSP_BIC_HP combination.

For all energies, the emlivermore simulation predicts a higher $\bar{y}_D$ than G4DNA. For 500nm, the relative difference is in range 5-10%. The difference in the 2 physics list are largest for 2 MeV, while the other energies are quite similar in relative difference. It's unknown why the difference is so much larger for 2 MeV. It's also noticeable that the relative difference is increasing with smaller site diameter, indicating that the Livermore + QGSP_BIC_HP combination does not have sufficient accuracy to simulate low site sizes.

When setting up the simulation, it was shown that for high proton energies (≥ 50 MeV), the emlivermore and QGSP_BIC_HP combination had some problems with very short step lengths (10 nm). The yd(y) distributions for 100 MeV and 500 nm had a large peak around 0.4-0.7 keV/µm, before dropping to almost zero. For y higher than ~1 keV, the distribution appeared normal again. When this was observed, the troubleshooting quickly revealed that the 10 nm step length was too low for the physics list combination. When increasing the step length to 50 nm, the microdosimetric spectrum was as expected again, as seen in Figure 23. The two distributions were also compared to the G4DNA simulation with a 1 nm step length. The G4DNA physics list had no problems with such low step lengths.

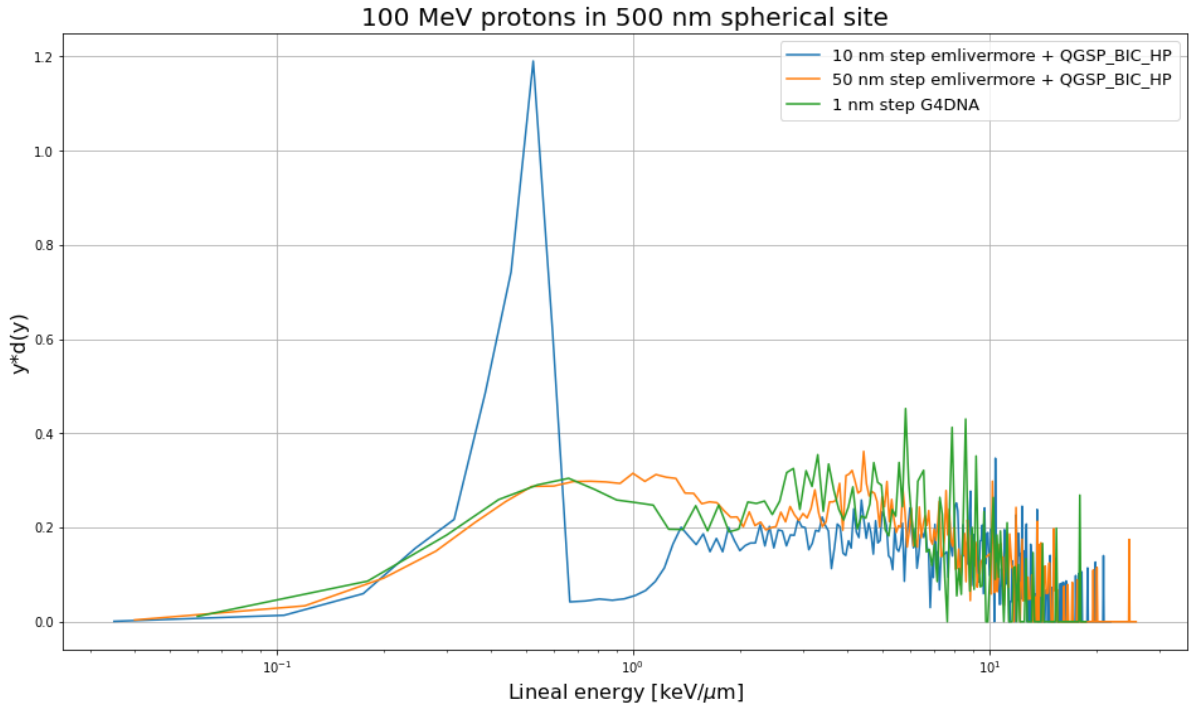

Figure 23: shows spectrum from two simulations with the emlivermore and QGSP_BIC_HP physics lists, with 50 and 10 nm step length, respectively, with the G4 DNA as a reference with 1 nm step.

5. Discussion

5.1 Simulation choices

The energies chosen to study was 1-100 MeV. The limiting factor at higher energies is the G4-DNA physics list, which at the moment has a limit at 100 MeV for protons. The GEANT4-DNA project is constantly updating and developing the physics list, and hopefully there will be possibilities for simulating even higher energy protons. However, the microdosimetric spectrums are changing the most for lower energy protons, while at > 20 MeV, the changes are relatively small. The maximum energy used in proton therapy is usually around 250 MeV [39], though only a small change in the microdosimetric spectrums from 100-250 MeV is expected.

For the lower energy protons, 1 MeV was chosen to be the lower limit. Protons with energies lower than 1 MeV will have a very short reach, as 1 MeV protons have a range of 24.6 μm. Cells vary a lot in size, but most have a diameter between 10-100 μm, meaning that proton with less than 1 MeV only could irradiate a couple of cells before stopping. And more importantly, for proton therapy, less than 1% of the dose will be deposited by protons of energies below 1 MeV in a typical clinical beam.

At first, test simulations producing 10 000 events inside the SV were made, but the shapes of both the f(y) and the yd(y) distributions were found to be rough and jagged. This was thought to be a result of a too low statistical foundation. The balance between statistical foundation and simulation time was evaluated, and the number of events inside the SV were increased a minimum of 20 000. This larger amount of events made the distributions smoother, increasing the confidence in the calculations of $\bar{y}_D$ and $\bar{y}_F$.

For all simulations, the beam radius was set to cover the whole world volume, to ensure the CPE inside the SV. However, for 100 MeV protons, the world volume was set to 1300 μm, meaning a beam radius of about 900 μm is needed for simulating

protons in the whole world volume. These simulations were found to be very time demanding, and the decision to reduce the beam radius to 450 μm was made. The possibilities of a secondary electron being produced at the edge of the world volume and hitting the SV was low, evaluated along with the time savings with the smaller beam radius. To create secondary electrons with maximum kinetic energy, a head on collision is needed, which in itself is rare. High energy transfers yield low scattering angles for electrons relative to the protons trajectory, meaning a maximum energy electron from the edge of the SV will not be able to travel perpendicular towards the SV. It is therefore assumed the difference in the microdosimetric distributions is negligible when decreasing the beam radius. Reducing the beam radius to half will quadruple the time efficiency of the simulation, since the number of protons needed to produce 20 000 events in the SV is proportional to the beam area. Each of the three 100 MeV simulations lasted for about a week, even with the reduced beam radius.

5.2 Virtual sites

In this work, the SV of the simulation was a square cube, while the spherical sites were created virtually inside the cube, as shown in Figure 14. This meant that one simulation could be used for several site diameters, instead of having one simulation per site diameter and proton energy. The fact that the SV could fit more spherical sites the smaller the diameter, means that each event could be reused multiple times. One primary proton track can travel through several sites, and the ejected secondary electrons may reach multiple sites as well. Looking at 5 MeV protons in Figure 16, it is noticeable that the number of events counted increased from about 18 000 to 560 000 when decreasing the site diameter from 500 to 10 nm. In that particular case, the 0.5 μm SV could fit 125 000 sites with diameters of 10 nm. If one simulation per energy per site diameter was chosen, the simulation time would be increased significantly. One drawback of this method is the differences in number of events may be an disadvantage when comparing the different site diameters. But as discussed in 4.2, large statistical foundation is most important for smaller site

diameters, due to the number of collisions per event in smaller sites is lower than for larger sites, yielding a higher variance between events and a broader spectrum. Larger amounts of data is also needed to smoothen out the broader spectrums.

Another disadvantage of this method of sampling sites can be seen at low proton energies, especially 1 MeV with site diameters ≥ 5 μm. The stopping power of a 1 MeV proton is about 26 keV/μm (see Table 3), and the range at this energy is roughly 24.6 μm. This means that after traveling a few μm, the proton has lost a significant amount of its initial energy. For the 20 μm setup, the SV can fit 64 spheres with diameter 5 μm, set up in 4 different layers relative to the protons direction. When the protons reach the second row of sites, their energy is expressively reduced, meaning we are no longer measuring 1 MeV protons. The stopping power increases drastically with decreasing energy in this range, meaning that the protons deposits more energy in the second row of sites than in the first, contributing to a higher $\bar{y}_D$ in these sites.

When using only the first layer of sites to compute $\bar{y}_D$ for 5 μm sites, it was measured to be 31.11 keV/μm. When using all 4 layers of sites, the value changed to 43.08 keV/μm, a substantial increase. The same effects between site layers was found at site diameters of 10 μm. Because of this effect, only the first layer of spherical sites was used when measuring the energy depositions for 1 MeV at 5 and 10 μm. This decreases the data foundation the values are calculated on, but it ensures that the microdosimetric calculations are based on protons with starting energy of 1 MeV.

Other site sampling methods has been used in other publications, like for example by Barrato-Roldan et al. [40]. In their work, their SV was defined as a slab in the middle of the world volume. An algorithm randomly selecting points $\vec{P}_C$ inside the SV was used to define centers of the spherical sites. This technique was found to be robust, but slow, because there would be created sites that would not record any hits. To improve efficiency, they used a weighted sampling-algorithm, where a random "energy transfer point" was chosen, and the site center point $\vec{P}_C$ was sampled randomly around that point. This made every site contain at least one hit, making the

site sampling considerably more efficient. However, this algorithm had an bias towards areas inside the SV with a high density of energy depositions, making the calculated microdosimetric quantities artificially high. A correction factor was then introduced to weigh up for the bias. While the method used in this work may be slower than the one of Barrato-Roldan, it does not have any bias.

5.3 Spherical vs sylindrical sites

In the field of microdosimetry, the typical shapes of detectors/sites have been either spherical or cylindrical [41]. Spherical sites is often used due their equal responses in both isotropic and unidirectional radiation fields. Cylindrical sites are easier to construct, though their response is dependent on the orientation relative to the beam. As discussed in 2.2.2, the mean chord lengths are defined by different formulas, due to their geometrical differences. However these formulas does not consider the chord length distributions (CLD) of the different shapes. Peter Bradley [41] discussed this and has shown how different the distributions for different shapes can be, even with the same mean chord length. This unevenness will influence both the microdosimetric spectrum, as well as the averaged values $\bar{y}_D$ and $\bar{y}_F$. For this work, and other work this was compared with, only spheres and cylinders with the same diameter and height is relevant.

For a spherical site, the maximum chord length of a track is restricted by the diameter d of the sphere. Because of the uniform shape of the sphere, the chord length distribution is uniform. For a cylinder with the same height and diameter d, the maximum chord length is larger than d, depending on the orientation of the cylinder relative to the beam. There is also a higher variance in the chord length distribution for cylinders than for spheres. The variances impacts both the lineal energy spectrums, as well as the mean-values for y. A correction factor η was found to find the effective chord length for different shapes, to find the equivalent sphere mean chord length. For cylinders with equal height and diameter, η was found to be 0.895

in an isotropic radiation field. To emulate a sphere with diameter 1 µm, the diameter and height of an equivalent cylinder needs to be 0.895 µm. This means when comparing $\bar{y}_D$ from spheres and cylinders with the same diameter/height d, the $\bar{y}_D$ values from the cylinder is expected to be $\frac{1}{0.895} = 1.12$ times larger than for the sphere.

In 4.4, the $\bar{y}_D$ values derived from this work was compared to other works. Liamsuwan, et al [36] simulated monoenergetic protons in cylindrical sites of 10 and 100 nm, where the diameter and height was equal. The orientation of the cylinders were randomized, emulating an isometric radiation field. For 100 nm sites, Liamsuwans values were about 10-16% higher than this work, with one exception for 50 MeV, where this work reported a 3% higher $\bar{y}_D$. This corresponds well with the correction factor. For 10 nm sites, the $\bar{y}_D$ from this work was 2-17% higher than reported from Liamsuwan, which is opposite of what could be expected.

5.4 Step length

In this work, the max step length was chosen to be 1/10 of the diameter of the smallest site investigated. For the G4DNA setup, this means the max step length was set to 1 nm. The mean chord length of a particle in a spherical site with a diameter of 10 nm is 6.67 nm. This means that the average particle travels around six or seven steps inside the site. In the hit-file, every step of a particle is written down, with the computed energy deposited along the step, meaning the "energy deposition point" is set to the end of the step.

When the particle enters or leaves the site, some border-artifacts may happen. Looking between step 1 and 2 in Figure 24, it is clear that most of the particle's track is outside the site. However, because step 2 is just inside the site, GATE records the energy deposition between the two steps to be inside the site. This creates an artificially high y-value for this track. The exact opposite phenomenon can be seen at

step 5 and 6, where the majority of the track length is inside the site, while the final step is outside. Both border-events may not happen for every track. For events with only a few steps inside the site, this may have an effect on the energy deposited. Because neither of these events have a higher probability of happening than the other, simulating a high number of particles will even out the number of steps for each event, making them cancelling each other. Another way to counter this problem is by lowering the step length, making the intervals between each "energy deposition point" smaller. This effect may have an influence at lower site sizes like the 10 nm sphere, but for the larger sites this effect would be negligible. In this work, the choice of 1 nm step length for the DNA-setup was made by weighing the accuracy up against simulation time and the amount of data made, as well as the errors are thought to be small.

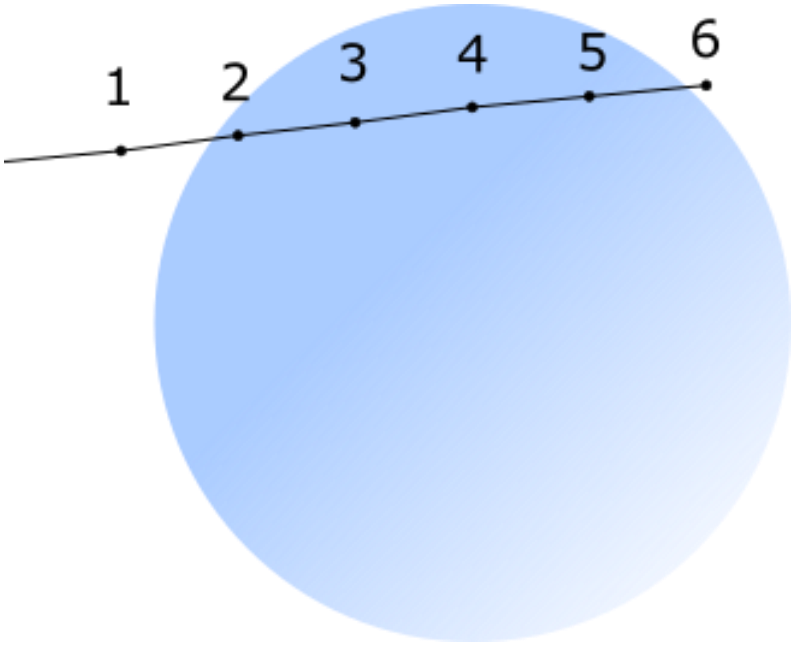

Figure 24: Shows an illustration for steps for a particle track in a site.

In GEANT4, there are possibilities of other output-files. There are ROOT-output files where every interaction is recorded, instead of every step for the particle. In this case, we know exactly where the interactions and energy depositions took place, instead of collecting the energy depositions along a step. This way of collecting the data output

would not have the same problem, though the whole post-processing part of the data handling would need change, since this work is based on ASCII-output (csv-files).

5.5 Water and tissue-equivalency and G4DNA limitations

In this work, the proton beams have been simulated traveling through liquid water. Water is widely seen as a good tissue-equivalent material, since a large percentage of the human body weight is water. However, the human body is made of inhomogeneous tissue, including muscle, fat, bones, air inside the lungs, etc. While bones have a higher density than water, adipose tissue (fat) have a slightly lower density than water [42]. The LET of proton is dependent of the density of the material it is traveling through, and changes in density will affect the energy depositions, and also range of the protons. When a tumor is located closely to an organ at risk (OAR), it is important to know the exact range of the protons, to make sure the dose to the OARs is as low as reasonably possible.

Traditionally, the physics lists available in GATE is made for use in several materials. There are possibilities for inserting CT-images as volumes, extracting the density with help of Hounsfield units and simulate proton beams through the patient. However, the GEANT4 DNA physics list is limited to liquid water, as well as a few DNA-related materials. This is due to the complexity of the track structure MC model, demanding cross sections for every interaction for the particles in the absorbing matter. A common way of calculating range for beam range through inhomogeneous tissue is to calculate a water equivalent thickness (WET) [5]. For every tissue type, the WET could be calculated and added together. The total accumulated WET could then be used to determine how much water should be set in front of the SV in a simulation using the GEANT4 DNA-physics list.

5.6 Connecting microdosimetry to RBE

As demonstrated in this work, microdosimetric quantities are strongly dependent on the site size they are measured in. As Figure 16 shows, the lineal energy spectra and $\bar{y}_D$ varies considerable with the site sizes. For 100 MeV, the $\bar{y}_D$ increases with a magnitude of 10 when decreasing the site diameter from 1 µm to 10 nm. When trying to connect microdosimetry to RBE, there is some debate on which site sizes that are most relevant [43]. Historically, the most used site diameter in physical measurements is 1 µm, due to technical difficulties in constructing the detectors [10]. A large spectrum of site sizes have been suggested to be relevant to model biological effectiveness. Different sizes representing sub-structures like the cell nucleus or the DNA itself has been thought to give the best yield when connecting the microdosimetric quantities to biological effect. The size of different cancer cells varies significantly, with measurements fluctuating from about 10 to several hundred µm [44], meaning the size of the biological targets also varies depending on different cancer types.

Newpower et al. [45] used GEANT4 to produce yd(y)-distributions and $\bar{y}_D$-values for 172 different proton energies. The results were used to fit the microdosimetric kinetic model to experimental data on multiple cell lines. They found that the optimal site diameter when calculating $\bar{y}_D$ was different for each cell line, ranging from ~400 to 900 nm. Several similar studies concluded with a site diameter around 500-1000 nm was optimal for fitting MKM to several cell lines. Other studies uses $\bar{y}_D$ to calculate α-values in LQ-based models, finding much smaller site diameters to be relevant. Lindborg et al. [10] found $\bar{y}_D$-ratios in site diameters in the range 10-15 nm to be proportional to α-ratios for multiple beam types. They assumed this to be a relation between the $\bar{y}_D$ at this volume size and complex double strand breaks for the DNA, leading to cell death. This shows that there is a large range of site sizes that is relevant for RBE-modelling, depending on cell types and model choice. Larger site

sizes seems to suit MKM based models well, while LQ-based models yields better results when using smaller site sizes in the nanometer range.

Also experiments without RBE-modelling points towards important targets in the sub-micrometer range. Zaider and Brenner [46] studied survival data from a ion beam experiment and how the probability of two energy transfers to create a lesion is dependent on the distance between them. They found that lesions were produced when the distance was under 20 nm, though sporadic lesion production was found up to a couple micrometers. Goodhead and Nikjoo [47] found that ultra soft x-rays with electron ranges down to 7 nm was still effective in inactivation of V79 cells, with an RBE of 3.4. Radiobiology is an extremely complicated field, and the understanding of the underlying mechanisms is poor. These examples shows how measurements at sub-micrometer range is needed to increase our understanding. This indicates a need for mapping of the microdosimetric properties of protons at a large range of site sizes.

LET have some similarities with lineal energy measured at ~1 μm. From 100 to 1 MeV, both LET and $\bar{y}_D$ varies the same amount, from about 1 to 30 keV, shown in Table 3. This indicates that the if critical biological targets is found in the micrometer range, LET may be as good as $\bar{y}_D$ for use as an input variable. However, at 10 nm, the $\bar{y}_D$ varies significantly less, and increases from 17.7 to 34.1 keV when changing the energy from 100 to 1 MeV. This shows that when studying smaller sites, the LET will not describe the energy deposition distribution as well as $\bar{y}_D$, and may yield worse RBE results if used as an input compared to $\bar{y}_D$.

Looking at Figure 17, $\bar{y}_D$-values for almost all the energies have the same trend when decreasing the site diameter. For the energies ≥ 5 MeV, the difference between 1 μm and 10 nm values is all in the range 12-16 keV/μm, with the difference systematically increasing with higher proton energies. This systematic pattern could be exploited to create a transformation function for $\bar{y}_D$-values, to convert the physical measurements of 1 μm sites to a smaller, more biological relevant size.

5.7 Future work

- Simulation of a clinical proton beam. In this work, only monoenergetic beams were studied. It would be of high interest to do the same work with a clinical beam, setting up the geometry in GATE as close as possible to a real-life beam. Calculating microdosimetric quantities at the entrance, the plateau and in the Bragg Peak. Both measurements pristine (single) Bragg Peaks as well as multiple points over the SOBP are of interest. This is crucial to RBE modelling, since a clinical proton beam is constructed to line up the Bragg Peak with the tumor.
- Using the microdosimetric quantities found in RBE model-fitting. As discussed in this work, both phenomenological and LQ-based model types have used microdosimetric quantities. With the large range of site sizes studied in this work, the optimal site size can be found while fitting the models.
- Simulation of other radiation types. It would be of interest to map the microdosimetric properties of a clinical photon beam (like a 6 MV beam). This could further be used in comparison to the microdosimetric values for protons for model fitting, like Lindborg et al.[10]. In addition, microdosimetric properties of heavier ions like carbon could also be calculated. Carbons have shown to have an much higher RBE than protons, as well as some other beneficial physical properties.
- Using this work as a foundation, a user interface can be programmed where the user choses the settings. Parameters like beam energy, the amount of water before the SV, the site size and amount of primary particles to be simulated could be set by the user, where a script is executing the MC simulation after. The program will output the wanted microdosimetric quantities, like $\bar{y}_D$, $\bar{y}_F$ and the lineal energy distribution yd(y). This would be a great tool for researchers fitting microdosimetric RBE models, where they can gather the quantities without having to construct the whole simulation from scratch.

- Examine the possibility of making a model for transformation of microdosimetric quantities between different site sizes. Transformation of yd(y)-distributions would be the most optimal quantity to transform, though models for this may be the hardest to construct. Mean values like $\bar{y}_D$ or $\bar{y}_F$ would be easier to transform. If this is proven to be possible, larger detectors that are easier to build could be used to derive microdosimetric results for smaller sites.

6. Conclusion

In this work, the microdosimetric properties of monoenergetic protons was studied. Creating a Monte Carlo simulation program in GATE, tracks and energy depositions were simulated for protons with energy 1-100 MeV in site with diameters ranging from 20 μm to 10 nm. With python-scrips post-processing the raw simulation data, the $\bar{y}_D$, $\bar{y}_F$ and $yd(y)$-distributions were calculated. It was found that $\bar{y}_D$ values increased systematically with smaller site diameters, and that the $yd(y)$-distributions became wider with both lower site diameters and higher proton energy. $\bar{y}_D$-values for small site diameters ($\geq$ 1 μm) were comparable to earlier reported values in the literature.

Different physics lists in GATE/GEANT4 were also compared at small site diameters ($\geq$ 500 nm). The GEANT4 DNA physics list yielded systematically lower $\bar{y}_D$ than the combination of the Livermore and QGSP_BIC_HP. The GEANT4 DNA physics list was also found to be able to process shorter step length than the mentioned combination.

The results in this thesis could be used to investigate which site sizes correlates best with the biological effect of proton therapy. Combined with experimental data from cell irradiations, the results could contribute to improved modelling of RBE in proton therapy.

Bibliography

1. Sung, H., et al., Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 2021. 71(3): p. 209-249.DOI: https://doi.org/10.3322/caac.21660.
2. Bergen, H. *Innføring av protonbehandling til Norge*. 2020 24.01.2022 22.05.2022]; Available from: https://helse-bergen.no/om-oss/protonsenter-i-bergen/innforing-av-protonbehandling-til-norge#fakta-proton-til-norge.
3. Mohan, R., and David Grosshans, Proton therapy - Present and future. Advanced drug delivery reviews, 2017. 109.DOI: https://doi.org/10.1016/j.addr.2016.11.006.
4. Huh, H.D. and S. Kim, History of Radiation Therapy Technology. Progress in Medical Physics, 2020. 31(3).DOI: https://doi.org/10.14316/pmp.2020.31.3.124.
5. Paganetti, H., *Proton Therapy Physics*. 2012.
6. *Particle Therapy Co-Operative Group*. Available from: https://www.ptcog.ch/index.php.
7. McMahon, S.J., Proton RBE models: commonalities and differences. Physics in Medicine & Biology, 2021. 66(4).DOI: https://doi.org/10.1088/1361-6560/abda98.
8. Mark Newpower, et al., Using the Proton Energy Spectrum and Microdosimetry to Model Proton Relative Biological Effectiveness. International Journal of Radiation Oncology*Biology*Physics, 2019. 104(2).DOI: https://doi.org/10.1016/j.ijrobp.2019.01.094.
9. Villegas, F., N. Tilly, and A. Ahnesjö, Target Size Variation in Microdosimetric Distributions and its Impact on the Linear-Quadratic Parameterization of Cell Survival. RADIATION RESEARCH, 2018. 190.DOI: 10.1667/RR15089.1.
10. Lindborg, L., et al. , Lineal energy and radiation quality in radiation therapy: model calculations and comparison with experiment. Physics in Medicine & Biology, 2013. 58.DOI: https://doi.org/10.1088/0031-9155/58/10/3089.
11. Wayne D Newhauser, R.Z., The physics of proton therapy. Physics in Medicine & Biology, 2015. 60(8).DOI: https://doi.org/10.1088/0031-9155/60/8/R155.
12. Lindborg, L. and A. Walker, *Microdosimetry: Experimental Methods and Applications*. 2017: CRC Press.
13. M.J. Berger, J.S.C., M.A. Zucker, J. Chang. *Stopping-Power & Range Tables for Electrons, Protons, and Helium Ions*. 2017 July 2017; Available from: https://dx.doi.org/10.18434/T4NC7P.
14. Liamsuwan, T., et al., Physical and biophysical properties of proton tracks of energies 1 keV to 300 MeV in water. International Journal of Radiation Biology, 2011. 87(2).DOI: https://doi.org/10.3109/09553002.2010.518204.
15. Fjæra, L.F., *Studies of the linear energy transfer and relative biological effectiveness in proton therapy of pediatric brain tumors*. 2021, University of Bergen.
16. Samnøy, A.T., *Characterization and application of 3D silicon microdosimeters*. 2020, University of Bergen.
17. Sjøstrand, N.G., What is the average chord lenght? Annals of Nuclear Energy, 2002. 29(13).DOI: https://doi.org/10.1016/S0306-4549(02)00003-8.
18. Harald H. Rossi and M. Zaider, *Microdosimetry and Its Applications*. 1996: Springer Berlin, Heidelberg.
19. Commons, W. *File:Proportional counter avalanches.jpg --- Wikimedia Commons, the free media repository*. 2020; Available from:

https://commons.wikimedia.org/w/index.php?title=File:Proportional_counter_avalanches.jpg&oldid=504800467.
20. Giuseppe Minniti, C.G., Michael Brada, *Chapter 16 - Radiotherapy*, in *Handbook of Clinical Neurology*, F.B. Michael J. Aminoff, Dick F. Swaab, Editor. 2012, Elsevier.
21. McNamara, A., L, Willers, H., & Paganetti, H, Modelling variable proton relative biological effectiveness for treatment planning. The British journal of radiology, 2020. 93(107).DOI: https://doi.org/10.1259/bjr.20190334.
22. Eivind Rørvik, L.F.F., Tordis J Dahle, Jon Espen Dale, Grete May Engeseth, Camilla H Stokkevåg, Sara Thörnqvist, Kristian S Ytre-Hauge, Exploration and application of phenomenological RBE models for proton therapy. Phys. Med. Biol, 2018. 63(18).DOI: https://doi.org/10.1088/1361-6560/aad9db.
23. Bertolet, A., M.A. Cortés-Giraldo, and A. Carabe-Fernandez, Implementation of the microdosimetric kinetic model using analytical microdosimetry in a treatment planning system for proton therapy. Physica Medica, 2021. 81.DOI: https://doi.org/10.1016/j.ejmp.2020.11.024.
24. Harrison, R.L., Introduction To Monte Carlo Simulation. AIP conference proceedings, 2010. 1204.DOI: 10.1063/1.3295638.
25. I. Kyriakou, V.I., D. Sakata, M.C. Bordage, S. Guatelli, S. Incerti, D. Emfietzoglou, Influence of track structure and condensed history physics models of Geant4 to nanoscale electron transport in liquid water. Physica Medica, 2019. 58.DOI: https://doi.org/10.1016/j.ejmp.2019.01.001.
26. S. Agostinelli, e.a., Geant4—a simulation toolkit. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2003. 506(3): p. 250-303.DOI: https://doi.org/10.1016/S0168-9002(03)01368-8.
27. GEANT4. *Physics lists*. Available from: https://geant4.web.cern.ch/support/physics_lists.
28. Jan S, S.G., Strul D, et al., GATE: a simulation toolkit for PET and SPECT. Phys Med Biol, 2004. 49(19).DOI: doi:10.1088/0031-9155/49/19/007.
29. David Sarrut , M.B., Manuel Bardiès, Julien Bert, Maxime Chauvin ,, et al., Advanced Monte Carlo simulations of emission tomography imaging systems with GATE. Physics in Medicine & Biology, 2021. 66(10).DOI: https://doi.org/10.1088/1361-6560/abf276.
30. GEANT4. *Guide For Physics Lists*. 2020; 10.7:[Available from: https://geant4-userdoc.web.cern.ch/UsersGuides/PhysicsListGuide/BackupVersions/V10.7/fo/PhysicsListGuide.pdf.
31. Collaboration, G. *Physics reference manual, Release 10.7*. 2021; Available from: https://geant4-userdoc.web.cern.ch/UsersGuides/PhysicsReferenceManual/BackupVersions/V10.7/fo/PhysicsReferenceManual.pdf.
32. Incerti, S., et al., Geant4-DNA example applications for track structure simulations in liquid water: A report from the Geant4-DNA Project. Medical Physics, 2018. 45(8): p. e722-e739.DOI: https://doi.org/10.1002/mp.13048.
33. Bernal, M.A., et al., Track structure modeling in liquid water: A review of the Geant4-DNA very low energy extension of the Geant4 Monte Carlo simulation toolkit. Physica Medica, 2015. 38(8).DOI: https://doi.org/10.1016/j.ejmp.2015.10.087.

34. Incerti, S., et al., Comparison of GEANT4 very low energy cross section models with experimental data in water. Medical Physics, 2010. 37(9): p. 4692-4708.DOI: https://doi.org/10.1118/1.3476457.
35. INCERTI, S., et al., THE GEANT4-DNA PROJECT. International Journal of Modeling, Simulation, and Scientific Computing, 2010. 01(02): p. 157-178.DOI: 10.1142/s1793962310000122.
36. Thiansin Liamsuwan, et al., Microdosimetry of proton and carbon ions. Medical Physics, 2014. 41(8).DOI: http://dx.doi.org/10.1118/1.4888338.
37. H Nikjoo, e.a., Radiation track, DNA damage and response—a review. Reports on Progress in Physics, 2016 79.DOI: https://doi.org/10.1088/0034-4885/79/11/116601.
38. Vassiliev, O.N., Peterson, C. B., Cao, W., Grosshans, D. R., & Mohan, R., Systematic microdosimetric data for protons of therapeutic energies calculated with Geant4-DNA. Physics in medicine and biology, 2019. 64(21).DOI: https://doi.org/10.1088/1361-6560/ab47cc.
39. Sengbusch E, et al., Maximum proton kinetic energy and patient-generated neutron fluence considerations in proton beam arc delivery radiation therapy. Medical Physics., 2009. 36(2).DOI: 10.1118/1.3049787.
40. Baratto-Roldán, A., et al., Microdosimetry and Dose-Averaged LET Calculations of Protons in Liquid Water: A Novel Geant4-DNA Application. Frontiers in Physics, 2021. 9.DOI: 10.3389/fphy.2021.726787.
41. Bradley, P.D., *The Development of a Novel Silicon Microdosimeter for High LET Radiation Therapy*, in *Department of Engineering Physics*. 2000, University of Wollongong.
42. Shirmardi, S.P., et al., Tissue inhomogeneity in proton therapy and investigation of its effects on BRAGG peak by using MCNPX code. International Journal of Radiation Research, 2014. 12(4): p. 335-341.
43. Goodhead, D.T., Energy deposition stochastics and track structure: what about the target? Radiation Protection Dosimetry, 2007. 122(1-4): p. 3-15.DOI: 10.1093/rpd/ncl498.
44. Si-Jie Hao, Y.W., Yi-Qiu Xia, Xin Zou, Si-Yang Zheng, Size-based separation methods of circulating tumor cells. Advanced Drug Delivery Reviews, 2018. 125.DOI: https://doi.org/10.1016/j.addr.2018.01.002.
45. Newpower, M., et al., Using the Proton Energy Spectrum and Microdosimetry to Model Proton Relative Biological Effectiveness. International Journal of Radiation Oncology*Biology*Physics, 2019. 104(2): p. 316-324.DOI: https://doi.org/10.1016/j.ijrobp.2019.01.094.
46. Zaider, M. and D.J. Brenner, The Application of Track Calculations to Radiobiology: III. Analysis of the Molecular Beam Experiment Results. Radiation Research, 1984. 100(2): p. 213-221.DOI: 10.2307/3576342.
47. Goodhead, D.T. and H. Nikjoo, Track structure analysis of ultrasoft X-rays compared to high- and low-LET radiations. Int J Radiat Biol, 1989. 55(4): p. 513-29.DOI: 10.1080/09553008914550571.

Appendix A - file compressor.ipynb

In [ ]:
```
import pandas as pd
import numpy as np
```

In [ ]:
```
#Choosing the parameters e (energy) and d (simulation size)
e=1
d=2.2
d_new=(d-0.2).__round__(1)
r_new=d_new/2
d_new=int(d_new)
```

In [ ]:
```
#Loading the file and convert units

df = pd.read_csv(f'../DNA/{d_new}um/results/{e}MeVu/Proton_{e}MeVu_{d}um_1_Hits.dat', sep='\s+',header=None,
                 usecols=[1,12,14,15,16,17])


df =df.rename(columns={1:'ID event',12:'Energy', 14:'X',15:'Y',16:'Z', 17:'Particle generated'})

# converting coordinates from mm to um
df['X'] = df['X'].multiply(1000)
df['Y']= df['Y'].multiply(1000)
df['Z'] = df['Z'].multiply(1000)

df['Energy'] = df['Energy'].multiply(1000) # from MeV til KeV
#df.info()
```

In [ ]:
```
#discarding all hits that deposits 0 energy
#+ the 0.1 um build up region

df = df.loc[(df['Energy']!=0) & (abs(df['X'])<=r_new) &
            (abs(df['Y'])<=r_new) & (abs(df['Z'])<=r_new)]
```

In [ ]:
```
#saving the hits in a new file
df.to_csv(f'hits/{d_new}um/Proton_{e}MeVu_{d_new}um_hits.csv')
```

Appendix B - energy script.ipynb

In [ ]:

```
import pandas as pd
import numpy as np
import warnings

warnings.simplefilter(action='ignore', category=FutureWarning)
```

In [ ]:

```
def site_energy(e,sim_d, site_d):
    #load the file
    df = pd.read_csv(f'hits/{sim_d}um/Proton_{e}MeVu_{sim_d}um_hits.csv',
                     usecols=['ID event','Energy','X','Y','Z','Particle generated'])

    #creating a matrix of site centers for the chosen
    #simulation diameter and site diameter
    for d in site_d:
        r = d/2
        nm=int(d*1000)

        edge=sim_d/2

        x = np.arange((-edge+r), edge, d)
        y = np.arange((-edge+r), edge, d)
        #z = np.arange((-edge+r), edge, d)
        z = [7.5]

        #MESHGRID FOR SITES

        xx, yy, zz =np.meshgrid(x, y, z)
        xx = xx.flatten()
        yy = yy.flatten()
        zz = zz.flatten()
        test = pd.DataFrame({'X':xx,
                     'Y':yy,
                     'Z':zz}
                     )

        #LOOPING THROUGH the site centers
        #saving energy deposited in site
        #sum energy depositions for each event

        sjekk = pd.DataFrame()

        energy1 = pd.Series()
        n_detectors = 0
        for row in test.iterrows():
            x = df.loc[np.sqrt((df['X']-row[1]['X'])**2 + (df['Y']-row[1]['Y'])**2 + (df['Z']-row[1]['Z'])**2 ) < r]
            lineal = x.groupby('ID event')['Energy'].sum()
            #print(lineal.head(15))
            energy1 = energy1.append(lineal)
            sjekk = sjekk.append(x)
            n_detectors=n_detectors+1
            if row[0] % 1000 ==0:
                print(row[0])
        sjekk['Partile generated'] = sjekk['Particle generated'].astype(str)
        print(n_detectors)
        print(len(energy1))

        #saves the energy depositions per event inside sites
        energy1.to_csv(f'energy/{sim_d}um/energy_{e}MeV_{nm}nm.csv')
```

Appendix C - integrals and plots.ipynb

In [ ]:

```
import pandas as pd
import plotly.express as px
import numpy as np
pd.options.display.max_columns = None
import matplotlib.pyplot as plt
import pickle
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
```

In [ ]:

```
def integrals(e, nm, sim_d):

    # a function to calculate y_D, y_F and yd(y)
    #for cosen energy {e} and site diameter {nm} from simulation size {sim_d}
    #also returns list of bins to used for plotting in later functions

    energy = pd.read_csv(f'energy/{sim_d}um/energy_{e}MeV_{nm}nm.csv', header=0)
    energies = energy['0']
    entries=len(energies)
    #print('no of events:', entries)

    #calculation of lineal energy

    r=nm/(2*1000);
    l = (4*r)/3
    y = energies/l  #kev/um

    maxy = max(y)*1.05



    binsize = round(3*np.std(y)*(len(y)**(-1/3)),2)
    #print('bin size is: ', binsize)
    Ebinsize = round(2*np.std(y)*(len(y)**(-1/3)),2)
    #print(Ebinsize)


    # I N T E G R A L S & P L O T S

    ybins = np.arange(binsize/2,maxy,binsize).tolist()

    plt.figure(0)
    h, bins1, patches = plt.hist(y,ybins, histtype = 'step')


    #normalising f(y) to 1
    f_y = [x/(sum(h)*binsize) for x in h]
```

```
    #calculation of y_F
    integral_f_y = 0
    yf = 0

    for i in range(len(f_y)-1):
        fydy = f_y[i]*binsize
        integral_f_y += fydy #f(y)*dy sum, skal være 1

        y_i = ybins[i] #lineal energy for bin i

        yf += y_i*fydy      #integral (y*f(y)dy)

    #print(integral_f_y)
    #print('yf is ', yf)


    # Calculcation of y_D
    dy = []
    yd = 0
    for i in range(len(f_y)):
        y_i = ybins[i]
        fydy = f_y[i]*binsize

        dy.append((y_i*f_y[i])/yf)
        yd += (y_i**2*fydy)/yf



    #print('sum av dy = ',sum(dy))
    #print('y_d = ', yd)


    # calculation of lineal energy distribution yd(y)
    yd_y = []
    yd = 0
    for i in range(len(dy)):
        y_i = ybins[i]

        yd_y.append(dy[i]*y_i)

        yd += y_i*dy[i]*binsize  #integral y*d(y)*dy
```

```
    plt.figure(3)
    plt.plot(ybins[:-1],yd_y, linewidth=0.0001)
    plt.xlabel(r'Lineal energy [keV/$\mu$m]', fontsize=24)
    plt.ylabel('y*d(y)', fontsize=24)
    plt.xscale('log')
    plt.grid(which='both', linewidth=0.8)


    print('y_d = ', yd, 'for ', nm, 'nm')


    return(ybins, yd_y, entries, yd, yf)
```

```
def ydyplots500nm(e, lst):

    #calculating quantities for the sub-500nm sites
    #returning lists of y_D and y_F for the chosen energy and site sizes
    plt.figure(1)
    plt.xlabel(r'Lineal energy [keV/$\mu$m]', fontsize=14)
    plt.ylabel('y*d(y)', fontsize=14)
    plt.xscale('log')
    sim_d=0.5

    yd_list = []
    yf_list = []
    colors = ['C0', 'C1', 'C2', 'C3', 'C4', 'C5', 'C6']
    col=0
    plt.figure(1, figsize=(20,10), dpi=500)

    for nm in lst:

        ybins, yd_y, entries, yd, yf = integrals(e,nm,sim_d);
        scientific_entries = "{:.1e}".format(entries)
        plt.figure(1)
        plt.plot(ybins[:-1],yd_y, linewidth=1.5, label = f'{nm} ,{yd.round(2)} ,#{scientific_entries}', color=colors[col] )
        plt.axvline(x=yd,color=colors[col], linestyle='--', lw=1.2 )
        yd_list.append(yd)
        yf_list.append(yf)
        col+=1


    plt.figure(1, figsize=(30,20), dpi=200)
    plt.legend(loc="upper left", fontsize=14, title = r'Diameter [nm]  , $\bar{y}_D$ , entries', title_fontsize=16)
    plt.title(f'Dose distributions for {e}MeV Protons')
    plt.grid(which='both', linewidth=0.8)

    plt.show()

    return yd_list, yf_list
```

```
ydyplots500nm(100,[500,200,100, 50, 20, 10])
```

```
yd_df = pd.DataFrame(columns=[0.5, 0.2, 0.1, 0.05, 0.02, 0.01])
yf_df = pd.DataFrame(columns=[0.5, 0.2, 0.1, 0.05, 0.02, 0.01])
for e in [1, 2, 5, 10, 20, 50, 100]:
    ydy_list, yf_list = ydyplots500nm(e, [500, 200, 100, 50, 20, 10]);
    print(ydy_list)
    yd_df.loc[e]=ydy_list
    yf_df.loc[e]=yf_list
yd_df.transpose().to_csv('yd_dfs/0.5um.csv')
yf_df.transpose().to_csv('yd_dfs/yf_0.5um.csv')
```

```
def ydyplots2um(e, lst):
    #calculating microdosimetric quantities for 1 and 2 um

    plt.figure(1)
    plt.xlabel('Lineal energy [KeV]')
    plt.ylabel('y*d(y)')
    plt.xscale('log')
    sim_d=2
    yd_list2um=[]
    yf_list2um= []
    colors = ['C0', 'C1', 'C2', 'C3', 'C4', 'C5', 'C6']
    col=0

    for nm in lst:

        ybins, yd_y, entries, yd, yf = integrals(e,nm,sim_d);
        scientific_entries = "{:.1e}".format(entries)
        plt.figure(1)
        plt.plot(ybins[:-1],yd_y, linewidth=1.5, label = f'{nm} ,{yd.round(2)} ,#{scientific_entries}', color=colors[col] )
        plt.axvline(x=yd,color=colors[col], linestyle='--', lw=1.2 )
        yd_list2um.append(yd)
        yf_list2um.append(yf)
        col+=1

    plt.figure(1)
    plt.legend(loc="upper left", fontsize=8)
    plt.title(f'yd(y) spectrum {e}MeV')

    plt.show()
    return yd_list2um, yf_list2um
```

```
yd_df2 = pd.DataFrame(columns=[2, 1, 0.5, 0.2, 0.1])
yf_df2 = pd.DataFrame(columns=[2, 1, 0.5, 0.2, 0.1])
for e in [1, 2, 5, 10, 20, 50, 100]:
    ydy_list2um,yf_list2um = ydyplots2um(e, [2000,1000, 500, 200, 100]);
    print(ydy_list2um)
    yd_df2.loc[e]=ydy_list2um
    yf_df2.loc[e]=yf_list2um

yd_df2.transpose().to_csv('yd_dfs/2um.csv')
yf_df2.transpose().to_csv('yd_dfs/yf_2um.csv')
```

In [ ]:

```
def ydyplots20um(e, lst):

    #calculating y_D and y_F for 2, 5, 10 and 20 um, saving it in a file
    plt.figure(1)
    plt.xlabel('Lineal energy [KeV]')
    plt.ylabel('y*d(y)')
    plt.xscale('log')
    sim_d=20
    yd_list20um=[]
    yf_list20um=[]
    colors = ['C0', 'C1', 'C2', 'C3', 'C4', 'C5', 'C6']
    col=0

    for nm in lst:

        ybins, yd_y, entries, yd, yf = integrals(e,nm,sim_d);
        scientific_entries = "{:.1e}".format(entries)
        plt.figure(1)
        plt.plot(ybins[:-1],yd_y, linewidth=1.5, label = f'{nm} ,{yd.round(2)} ,#{scientific_entries}', color=colors[col] )
        plt.axvline(x=yd,color=colors[col], linestyle='--', lw=1.2 )
        yd_list20um.append(yd)
        yf_list20um.append(yf)
        col+=1

    plt.figure(1)
    plt.legend(loc="upper left", fontsize=8)
    plt.title(f'yd(y) spectrum {e}MeV')

    plt.show()
    return yd_list20um, yf_list20um
```

In [ ]:

```
yd_df20 = pd.DataFrame(columns=[20, 10, 5])
yf_df20 = pd.DataFrame(columns=[20, 10, 5])
for e in [1, 2, 5, 10, 20, 50, 100]:
    ydy_list20um,yf_list20um = ydyplots20um(e, [20000,10000, 5000]);

    print(ydy_list2um)
    yd_df20.loc[e]=ydy_list20um
    yf_df20.loc[e]=yf_list20um
yd_df20.transpose().to_csv('yd_dfs/20um.csv')
yf_df20.transpose().to_csv('yd_dfs/yf_20um.csv')
```

In [ ]:

```
# adding all y_D and y_F-dataframes to one, and saving them as files
yd_total = pd.concat([yd_df20.transpose(), yd_df2.transpose().iloc[[0,1]], yd_df.transpose()], axis=0)
yf_total = pd.concat([yf_df20.transpose(), yf_df2.transpose().iloc[[0,1]], yf_df.transpose()], axis=0)

yd_total.to_csv('yd_dfs/yd_total.csv')
yf_total.to_csv('yd_dfs/yf_total.csv')
```

In [ ]:

```
def ydyplotsall(energies, lst):

    #using the integrals function to plot all energies and site sizes

    index=1
    colors = ['C0', 'C1', 'C2', 'C3', 'C4', 'C5', 'C6']
    for e in energies:
        col=0
        for nm in lst:
            if nm>=5000:
                sim_d=20
            if nm <=2000:
                sim_d=2
            if nm <=500:
                sim_d=0.5
            plt.figure(index)
            ybins, yd_y, entries, yd, yf = integrals(e,nm,sim_d);
            scientific_entries = "{:.1e}".format(entries)
            plt.subplot(4,1,index)
            plt.plot(ybins[:-1],yd_y, linewidth=1.2, label = f'{nm} ,{yd.round(1)} ,#{scientific_entries}', color=colors[col] )
            plt.axvline(x=yd,color=colors[col], linestyle='--', lw=1.3 )
            col+=1
        plt.title(f'{e}MeV', fontsize=24)
        plt.legend(loc="upper left", fontsize=20,title = r'Diameter [nm]  , $\bar{y}_D$ , entries', title_fontsize=16)
        plt.tight_layout(pad=1.5)

        index+=1


    plt.show()
```

In [ ]:

```
plt.rcParams['figure.figsize'] = [17,25]
ydyplotsall([1,5,10,100],[20000, 5000,1000,500,100,50,10])
```