University of Trento

Department of Physics
Ph.D. in Physics
XXXIV cycle

Expanding microdosimetry from radiation field characterization to radiobiological damage modeling

FINAL THESIS

Supervisor:
Prof. Chiara La Tessa
Co-Supervisor:
Dr. Maurizio Boscardin

Candidate:
Marta Missiaggia

Contents

"Costanzia: non chi comincia, ma quel che persevera."
– Leonardo Da Vinci

"It takes a fool to remain sane."
– The Ark

Introduction

Around 50% of all patients with localized malignant tumors undergo treatments using ionizing radiation, mostly in combination with tumor resection and/or chemotherapy [Delaney et al., 2005]. High energy photons are by far the most used radiation type in this field, but in the past decades the application of accelerated particles, especially protons and carbon ions, has gained popularity in cancer therapy. By the end of 2020, more than 290000 patients were treated worldwide with particle therapy, of which around 250000 with protons, close to 40000 with carbon ions and about 3500 with helium, pions or other ions [PTCOG, 2020].
The main reason for using charged particles in radiotherapy is their favorable depth-dose profile, i.e. the "Bragg curve", named after Sir William Henri Bragg[1] who investigated the slowing down of α particles in air [Bragg and Kleeman, 1905].
After many years from Bragg's studies, Robert R. Wilson[2] spent most of 1946 at the Berkeley Radiation Laboratory (LBL) (Berkeley, CA, USA) for his work on the design of a new 150 MeV cyclotron for Harvard University, to replace the cyclotron that had been relocated to Los Alamos during the war, under the supervision of Ernest Lawrence[3]. Wilson found that high-energy protons produced by the 150 MeV cyclotron could be used for cancer treatment [Endo, 2018]. In the same year of 1946, he published an article in the medical journal Radiology entitled "Radiological Use of Fast Protons", based on the ideas that he had developed at Berkeley [Wilson, 1946]. This study was the first to propose the use of particle beams for cancer treatment.
Radiotherapy with ions has come a long way since its beginning in the Berkeley Lab, and it is now a well-established strategy against cancer. However, not all the ions feasible for this application has gained the same popularity. Scientists focused on establishing, tuning and optimizing the treatment procedure for protons and carbon beams only, while the use of other particles was overshadowed. Once the community gained knowledge and confidence in this methodology, it started to open up to new challenges and once again begun exploring the possibility to perform radiotherapy with other ions.
Physical and biological experimental data pointed to ^{4}He and ^{16}O ions as the most promising species and several heavy ions centers have been designed to offer the capability to deliver these particles in addition to protons and carbon ions [Rovituso and La Tessa, 2017]. In July 2021, the first patient has been treated with a helium beam at the Heidelberg Ion-Beam Therapy Center (HIT) since the pioneering studies carried out at the Lawrence Berkeley Laboratory between 1975 and 1993.

Although ion therapy is improving both in precision and in efficiency, numerous treatment uncertainties remain a major obstacle to the full exploitation of this type of treatment. For each ion type in fact, a dedicated beam model has to be developed and implemented in the Treatment Planning System software, which provides the dose calculation and beam delivery for each treatment. To achieve this goal, the physical processes of interest have to be characterized in terms of production of secondary fragments from nuclear reactions, as well as energy loss and lateral scattering of primary and secondary ions. These interactions modify the radiation field composition, energy and direction, and thus influence the depth-dose and lateral profile.
One of the major uncertainties of ion therapy is related to the knowledge of the energy released from the radiation field, from which the dose deposition and radiobiological ef-

[1] William H. Bragg (1862 – 1942) was an English scientist.
[2] Robert R. Wilson (1914 – 2000) was an American physicist.
[3] Ernest O. Lawrence (1901 – 1958) was an American nuclear scientist.

fects are estimated. The pioneering work of Zirkle[4] and many other following studies have pointed out that, together with the amount of energy absorbed, the biological effectiveness of ionizing radiation depends on the spatial distribution of energy deposition [Rossi, 1959], which is one of the ingredients of the so-called *radiation quality* information.

Radiation field quality

Thee radiation field characterization is well recognized to play a fundamental role in ion therapy, however there is no quantitative and uniquely accepted definition of the concept of "radiation quality".
The *Publication 60 recommendations of ICRP* (1991) reports that "*the probability of stochastic effects is found to depend, not only on the absorbed dose, but also on the type and energy of the radiation causing the dose. This is taken into account by weighting the absorbed dose by a factor related to the quality of the radiation*". However, the definition of *quality of the radiation* is itself non unique.
Dr. Dudley Goodhead[5] in his speech at the *DoReMi Radiation Quality workshop* in Brussels (9-10 July 2013) [Goodhead, 2013] tried to answer to the question "What is radiation quality?". He started pointing out that the biological effects of ionizing radiation depend on: i) the quantity of radiation (dose, fluence, ...), ii) the temporal pattern (dose rate, fractionation), and finally on iii) the spatial pattern (tracks structures). Only the latter is actually related to the quality of radiation, and the variety of radiation tracks at the location of interest is determined (stochastically) by the types and energies of charged particles at that location. Radiation quality is then defined by the fluence spectrum of radiation particles at the locations of interest in the target material. This spectrum specifies the relative numbers of particles according to type and energy, it includes charged and neutral components and it depends on the radiation source features and the intervening material. Finally, differences in radiation quality can lead to differences in biological effectiveness for the same quantity of radiation, and to qualitative differences in biological effects. The study of the radiation quality also provides an excellent tool for probing underlying mechanisms of radiobiological effects.
However, the spatial pattern of the energy deposition of radiation in matter depends on many different electromagnetic and nuclear interactions, that is worth summarizing in order to understand the topic complexity. In addition, to grasp how the quality of radiation relates into biological damage, the basic information on the biological effects of radiation will be illustrated.

Physical and biological background

When charged particles pass through a medium, two main types of physical interactions have to be considered[Schardt, 2016]:

- *Electromagnetic interactions* which rule the energy deposition in matter. Coulomb forces between the positively charged ion and the negatively charged orbital electrons of the medium atoms are established. These interactions concur also in the lateral spread of particles, also known as *scattering*

[4] Raymond E. Zirkle (1902 – 1988) was an American radiation biologist.
[5] Dudley T. Goodhead is an American radiation biophysicist

- *Nuclear interactions*, which can cause wider lateral spread with respect to the electromagnetic contribution as well as a loss of the primary ions and production of secondary radiation.

Energy deposition in matter

Ions interact inelastically with the shell electrons of the medium nuclei, resulting in an energy loss dE/dx well described by the Bethe-Bloch formula [Bethe, 1953]

$$\frac{dE}{dx} = \frac{4\pi e^4 Z_t Z_p^2}{m_c v^2}\left[\ln\frac{2m_e v^2}{\langle I\rangle} - \ln\left(1-\beta^2\right) - \beta^2 - \frac{C}{Z_t} - \frac{\delta}{2}\right], \tag{1}$$

where Z_p and Z_t denote the nuclear charge of the projectile and the target, respectively, $< I >$ is the mean ionization energy of the target atom or molecule, and m_e and e are the mass and the charge of the electron, respectively. The formula includes the relativistic corrections by Fano et al. [Fano, 1963] with a dependence from β^2, and two additional terms for the shell (C/Z_t) and the density effect ($\delta/2$) corrections. The shell correction accounts for the effects which arise when the the incident particle velocity is comparable or smaller than the orbital velocity of the bound electrons. This correction affects the stopping power up to 6%. The density term corrects for polarization effects in the target. The mean ionization $< I >$, instead, corrects for the quantum mechanical energy levels available for the transfer of energy to the target electrons. For liquid water, accurate Bragg curve measurements for protons and different heavier ions show a values of 78 eV [Steidl et al., 2008].
A further correction to the Bethe-Bloch formula takes into account the dependence of the projectile effective charge on its velocity. At high velocities, the atomic electrons are completely stripped off and the ion effective charge is equal to its atomic charge number Z_p. As the ion slow down, the mean charge state decreases due to the interplay of ionization and recombination processes, and Z_p has to be replaced by the effective charge Z_{eff} described by the Barkas' empirical formula [Barkas, 1973]:

$$Z_{\text{eff}} = Z_p\left[1 - \exp\left(125\beta Z_p^{\frac{2}{3}}\right)\right].$$

The maximum energy transfer, corresponding to the Bragg peak, is reached at a projectile velocity of

$$v_p \approx Z_p^{-\frac{2}{3}} v_0\,,$$

where $v_0 = e^2\hbar$ is the Bohr velocity and the value of $e^2\hbar = 1/137$ represents the fine-structure constant.
An example of the energy loss curve as a function of the penetration depth, named "Bragg curve" for ions, can be seen in Figure 1, where a comparison between the different behaviours of carbons and of both X-rays and megavolt photon beams is given. In X-rays radiotherapy, the highest doses occur at small depths, so just below the skin, and deep seated tumors can only be treated safely by focusing beams on the tumor from many angles using a rotating linac [Schardt et al., 2010].
In contrast to photons, in those small depths regions also referred to as the entrance channel or plateau region, the ions loose only a small fraction of their kinetic energy by ionization. Instead, towards the end of their path, the energy transfer increases drastically and reach a maximum value indicated as Bragg peak, which is followed by a sharp

fall-off. The position of this peak can be precisely adjusted to the desired depth in tissue by changing the incident ion kinetic energy.

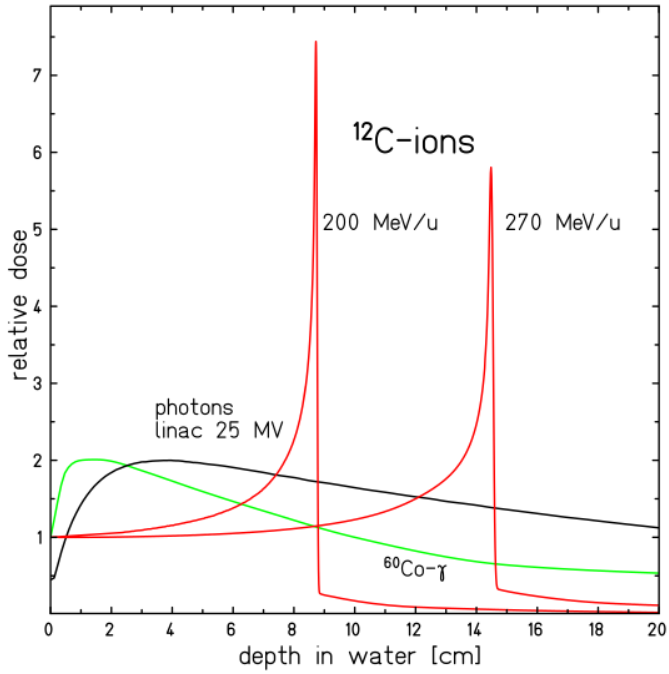

Figure 1: Depth-dose profiles of X-rays, megavolt photon beams, and ^{12}C ions in water. Figure taken from [Schardt et al., 2010]

Mean range and energy straggling

The mean range R is defined as the average thickness of material traversed by the charged particle before it comes to rest:

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

It can be assumed to be the same as the total path length of the particle trajectory in an absorber. At a given ion kinetic energy, the range in a medium scales as A/Z^2, as shown in Figure 2 for protons, ^{12}C, ^{16}O and ^{20}Ne in water.

According to Equation 1, the energy loss of a single ion plotted as a function of the penetration depth would result in a very sharp peak near its stopping point. However, the energy loss is a stochastic process and thus the statistical fluctuations will cause a broadening of the Bragg peak for a monoenergetic beam [Comfort et al., 1966].
These fluctuations result in the so called "energy-loss straggling", well described by the Landau or the Vavilov distributions, for very thin and thin absorbers respectively. In the limit of many collisions (or a thick absorber), the Vavilov distribution approaches a Gaussian form [Ahlen, 1980] with a σ expressed as

$$\sigma_E = 4\pi Z_{eff} Z_t e^4 N \Delta x \left[\frac{1-\beta^2/2}{1-\beta^2}\right] .$$

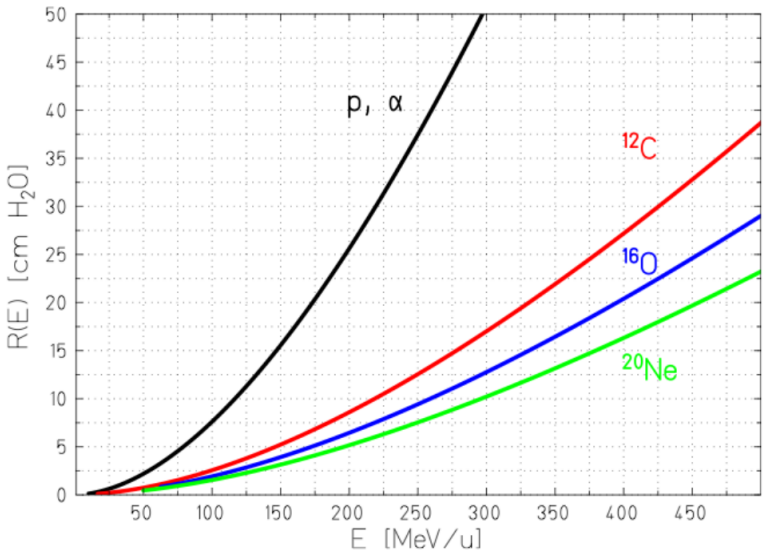

Figure 2: Mean range of several ion species in water (protons, ^{12}C, ^{16}O and ^{20}Ne). Figure taken from [Schardt et al., 2010].

The range straggling variance σ_R^2 is related to the energy-loss straggling variance σ_E^2, and the ratio between σ_R^2 and the mean range R is proportional to $\frac{1}{\sqrt{M}}$, with M being the ion mass. Thus, the range straggling gets smaller for heavier ions, being maximum for protons and, for example, a factor 3.5 larger compared to the ^{12}C ions' one [Schardt et al., 2010].

Lateral beam spread

The lateral spreading of charged particles passing through matter is mainly caused by their elastic Coulomb interactions with the target nuclei, and it is well described by the Molière-Theory [Moliere, 1948].
For small angles, the higher-order terms in Molière's solution can be neglected, and the angular distribution can be described by a Gaussian function with standard deviation given by the Highland approximation:

$$\sigma_\theta[rad] = \frac{14.1\,MeV}{\beta\,p\,c} Z_p \sqrt{\frac{d}{L_{rad}}} \left[1 + \frac{1}{9}\log\left(\frac{d}{L_{rad}}\right)\right] . \tag{2}$$

where Z_p and p are the charge and the momentum of the projectile, respectively. The absorber material is characterized by its thickness d and radiation length L_{rad}, which, at a given thickness, results in a larger angular spread for materials containing heavier elements. The lateral spread calculated with Equation (2) yields a smaller value for heavy ions than for protons at a certain depth, coming from the factor βpc.

Nuclear interactions

Compared to the electromagnetic interactions with the medium atomic electrons, the probability for an ion to undergo nuclear reactions is much smaller, but leads to significant effects as the penetration depths increases. At energies of several hundred of MeV/u,

violent nuclear spallation reactions may result in a partial or complete disintegration of both the projectile and target nuclei (e.g., in central head-on collisions). For geometrical reasons, peripheral collisions, where the beam particle loses one or few nucleons, are the most frequent nuclear reactions occurring along the stopping path of the ions. They can be well described as a two-step process (Figure 3) with the abrasion-ablation model [Serber, 1947].

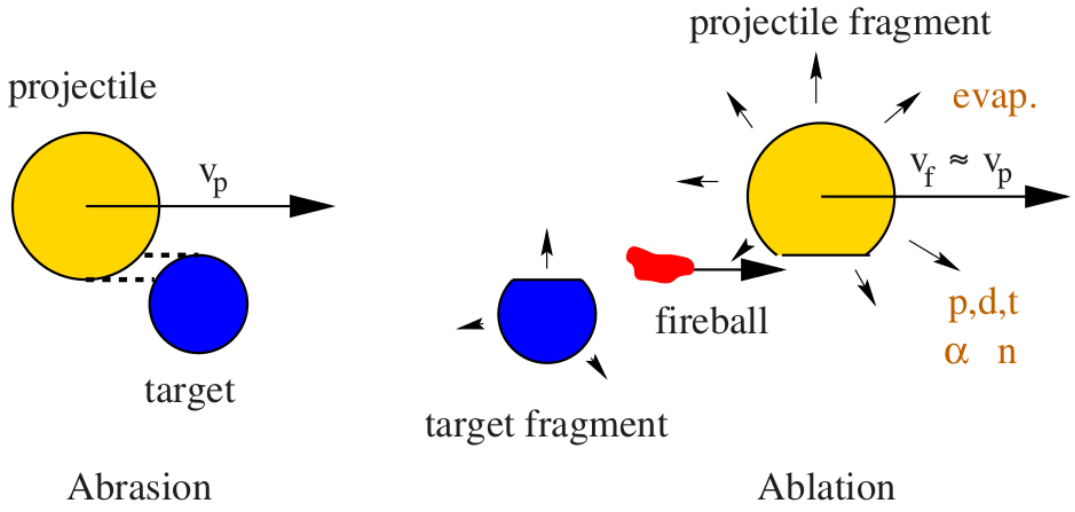

Figure 3: Illustration of the abrasion-ablation model of peripheral collisions at high energies, according to [Serber, 1947]. Figure from [Gunzert-Marx et al., 2008]

In the first step (abrasion), nucleons contained in the overlapping reaction zone (the "fireball") are abraded, while the outer nucleons ("spectators") are only slightly affected. The remaining projectile and target fragments then de-excite by evaporation of nucleons or clusters in the second step (ablation).
Fragmentation reactions have been extensively studied in nuclear physics [Goldhaber and Heckman, 1978, Hüfner et al., 1975] and experimental data are available for many projectile-target combinations and for a wide range of beam energies [Otten, 1989, Norbury et al., 2012].
Nuclear fragmentation yield some important effects relevant for radiotherapy with high-energy ion beams:

- nuclear reactions cause a loss of primary beam particles and a buildup of lower-Z fragments. These effects become more and more important with increasing penetration depth.

- the secondary (or higher-order) projectile-like fragments are moving with approximately the same velocity as the primary ions. In general, they have a bigger range than the primary beam and lead to a tail behind the Bragg peak.

The impact of nuclear fragmentation on the Bragg curve is shown in Figure 4. With increasing penetration depth, the peak-to-entrance dose ratio becomes gradually smaller, mainly caused by the exponentially diminishing flux of primary ions. The buildup of lower-Z fragments is clearly visible in the dose tail behind the Bragg peak at larger depths.

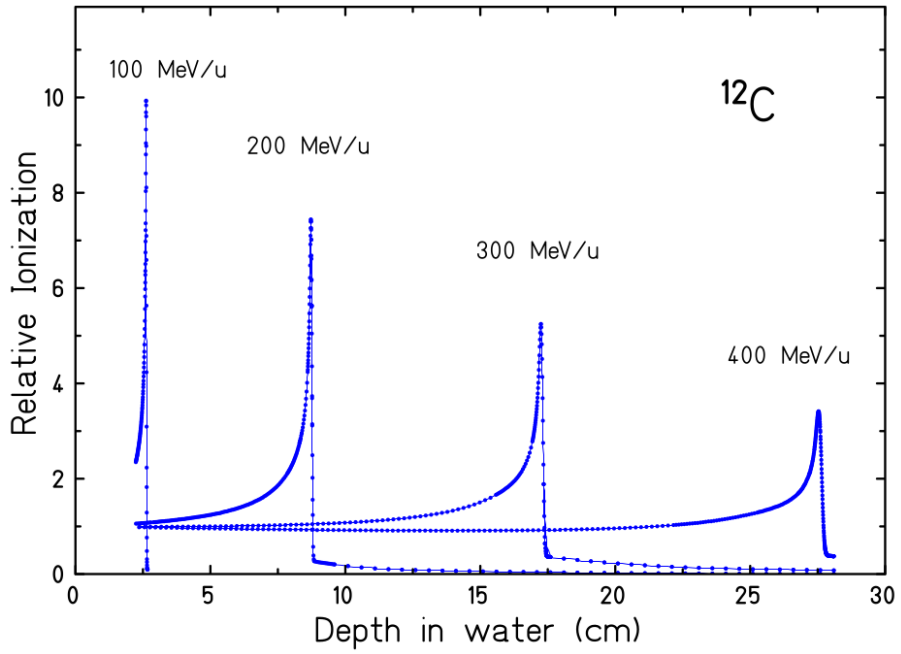

Figure 4: Measured Bragg curves of ^{12}C ions stopping in water. The dose tail behind the Bragg peak increases at larger depths due to nuclear fragmentation. Figure taken from [Schardt et al., 2008].

Fragmentation processes have an impact not only on the physical dose distribution but also on the biological effect of ions. Different ions will create different fragments species, each characterized by a different biological effect. When dealing with protons one can, to a good approximation, fully rely on the physical dose released in the Bragg peak, which can quite easily be measured, as the dose is caused essentially by the same radiation species. For heavier ions, this approximation is not valid anymore, as by just measuring the physical dose one cannot separate the contribution of different ions species.

Biological effect of radiation

The goal of curative cancer radiotherapy treatments is to inactivate all malignant cells while keeping the normal tissue complications, caused by unavoidable out-of-target dose, at an acceptable level.
Tumor cells have to be sterilized so that they cannot transfer their genetic defects to daughter cells. An effective way to achieve this result is to damage the DeoxyriboNucleic Acid (DNA), that contains all the entire set of genetic information.
The DNA of all cells has multiple ways to repair itself. Under the assumption that only unrepaired or misrepaired DNA damage leads to cell death, the cell repair capability can be described by the probability S to survive after being exposed to a radiation dose D. S is described by a linear-quadratic (LQ) model [McMahon, 2018a],

$$S(D) = exp(-\alpha D - \beta D^2)\,.$$

The parameters α and β and are cell type specific factors. The repair capacity (i.e. the radiosensitivity) of individual tissues is reflected in the quadratic term β and is commonly reported using the ratio $\frac{\alpha}{\beta}$. Hence, cells with a large α/β ratio show small repair capability

and viceversa [Fowler, 1989].
The most common radiation-induced DNA damages are single strand breaks, double strand breaks and the loss of a base (Figure 5).

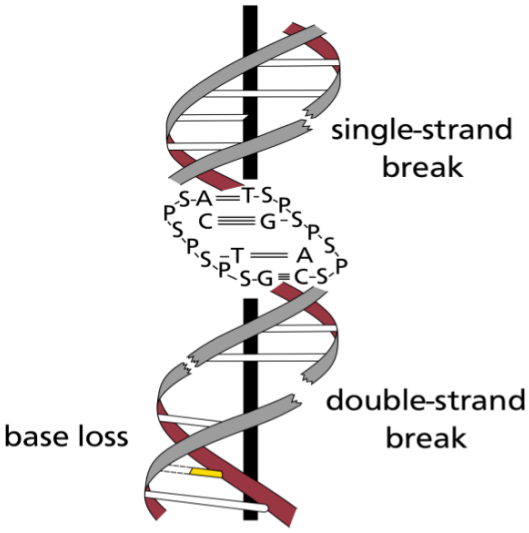

Figure 5: The backbone of a DNA molecule is composed of two phosphate-sugar strands in opposite directions. Between compatible bases hydrogen bounds are created, that keep two strands together, giving the typical double helix shape. Radiation induced damages on the DNA are mainly single or double strand breaks [Richter, 2012].

Depending on how far the primary ionization events are separated in space, radiation is characterized as sparsely ionizing (X-rays) or densely ionizing (charged particles); heavier particles produce higher ionization density.
In Figure 6 the different spatial distributions of energy deposition between photons (X-rays) and charged particles (carbon ions) is clearly visible, together with the effect of this depositions on the DNA strands: sparsely ionizing radiation leads to a random spatial distribution of energy deposition events in the cell volume. Conversely, ions deposit their energy by the emission of secondary electrons leading a very localized energy deposition along the track of primaries [Scholz, 2006].

It is well known that the features of the radiation track structure at the nanometer level have important implications in terms of radiation effects in biological targets [Nikjoo, 2003]. This is especially true for energetic ions, which have complex track structures characterized by energy depositions not only along the primary particle path, but also projected out radially with respect to the track "core". This is due to the so-called "delta rays", namely high-energy secondary electrons which can travel distances of the order of tens of micrometers in biological targets.
Due to this track structure, densely ionizing radiation, like charged particles, induces more complex DNA damage and this complexity makes the DNA repair process more difficult [Schardt et al., 2010].
DNA can be damaged mainly in two ways: radiation can directly ionize its atoms (*Direct Action*) or it can interact with other molecules, in particular water molecules forming free radicals (*Indirect Action*). The latter are molecules with an unpaired electron highly reactive and diffuse to short distances interacting with DNA [Hall and Giaccia, 2006].

For what concerns a quantitative description of the effectiveness of ions in causing a

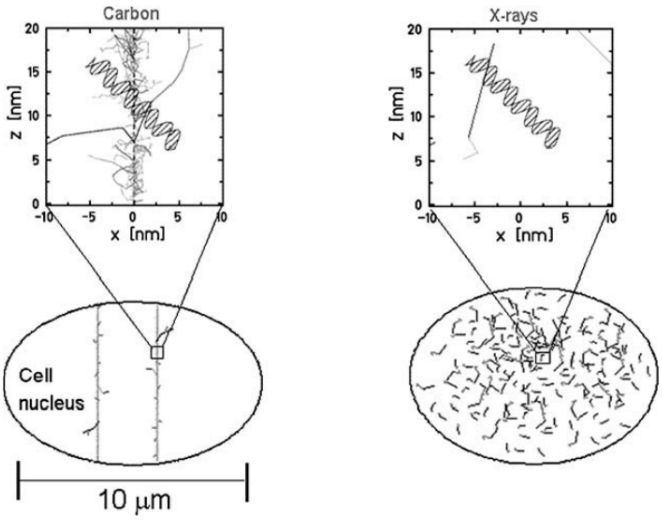

Figure 6: (Left) Carbon track structure in cell nucleus: the electrons density is highly localized around the ion path producing a large number of ionization to the DNA. (Right) Uniform energy depositions of X-rays in the cell nucleus [Scholz, 2006].

biological damage, the most widely used variable is the *Relative Biological Effectiveness* (RBE). The RBE is defined as the ratio of the physical absorbed dose of a reference radiation (250 keV photons) with respect to the dose administrated by ions yielding the same biological effect:

$$RBE \equiv \left.\frac{D_{ref}}{D_{ion}}\right|_{\text{isoeffect}}$$

RBE is the most relevant quantity in heavy ion therapy for describing cell killing and normal tissue complications [Dieter Schardt, 2010].

A quantitative description of radiation quality: dosimetry vs microdosimetry

The radiation field quality is described using either a macroscopic or a microscopic approach. The widely used macroscopic description is based on the *linear energy transfer* (LET = dE/dx energy released by radiation per unit length). Since the energy is imparted in or near the charged particles tracks, it is convenient to express the heterogeneity of energy deposition in terms of the linear density of energy loss in these tracks. For this purpose, the term linear energy transfer (LET) has been coined by Zirkle et al. (1952) [Zirkle et al., 1951]. Using this concept, radiation quality can be described as a distribution of LET, specifying the fraction of the dose deposited in each LET interval [Rossi, 1959].
As defined in [Kalholm et al., 2021], two different ways of averaging LET spectrum are typically used: (i) averaging over the fluence Φ_i, or number of tracks, that each charged particle produces, called *fluence-averaged LET* (also called *track-averaged LET* and denoted by LET_F and (ii) averaging over the microscopic dose d_i given in an infinitesimally small volume contributed to by the track of a single particle, called *dose-averaged LET*

and denoted by LET_D. The two quantities are thus defined as

$$\mathrm{LET}_F := \frac{\sum_i \Phi_i \mathrm{LET}_i}{\sum_i \Phi_i},$$
$$\mathrm{LET}_D := \frac{\sum_i d_i \mathrm{LET}_i}{\sum_i d_i} = \frac{\sum_i \Phi_i \mathrm{LET}_i^2}{\sum_i \Phi_i \mathrm{LET}_i}.$$

In order to grasp how the radiation quality affects the biological damage, it is useful to study the behaviour of the RBE as a function of the LET of the radiation. The link between LET and RBE represents in fact the bridge between the physical characteristics of radiation and the biological outcomes. In this view, LET variations represent the main source of RBE variability, and thus an accurate RBE assessment calls for a detailed knowledge of the LET spectrum. Treatments with ions are characterized by LET changes between the entry channel and the tumor, especially if deep-seated. This feature is due to the nuclear and electromagnetic interactions of the primary ions with the patient's bod,y that cause both a change in the energy of the primaries, and the production of secondary fragments emitted at all angles and with large range of energies. The resulting mixed radiation field causes a spatial variation of RBE, both inside and outside the tumor.
Figure7 illustrates the RBE dependence on the LET for different ion types. The RBE increases for increasing LET up to a maximum value, which is ion-specific. For particles with high LET the ionization density is higher, and hence they induce more severe damage, thus resulting in an increased RBE. However, empirical evidences indicate that over certain LET values the RBE decreases. This phenomenon is often explained with the so-called *overkilling* and *saturation* effects, and is treated in details from a modeling point of view in 2.2. In few words, at high LET, the energy deposited into the cell nucleus is extremely high and in particular is greater than the energy needed to inactivate the cell. Therefore, any additional energy deposited results in the same biological effect leading to the cell death. In addition, as the LET increases the probability that a particle hits the nucleus decreases. The joint effect of an excessive energy deposition together with a sparing effect results in a reduction of the biological effectiveness and an RBE decrease at high LET. Figure7 also indicates that in the region upstream of the peak, protons have a higher RBE than Helium ions of the same LET. This is because at this LET, helium ions are faster than the protons with a broader track that reduces the ionization density within the track core. Therefore, at the same LET the cell killing of helium is lower than for protons. Despite being the most used approach to link radiation quality to its biological effectiveness, the LET is measured in macroscopic volumes, i.e. on a scale much larger than the cellular level where the actual damage occurs. This approach neglects the stochastic nature of energy deposition, which results in cell-to-cell variations, and LET is considered constant throughout the material. Furthermore, it carries several approximations, for instance track structure is not taken into account, and the distribution of the delta ray energy and its link to the spatial dose are not considered.

Microdosimetry has instead grown to overcome the limitations of describing radiation quality with a single mean value of LET. With this approach, the effect of radiation on cells is investigated in a region comparable to the structures of interest, where the stochastic nature of energy deposition is not negligible. The microscopic description accounts for this stochasticity exploiting the whole distributions of quantities related to the energy depositions in a volume equivalent to a micrometer region of tissue.

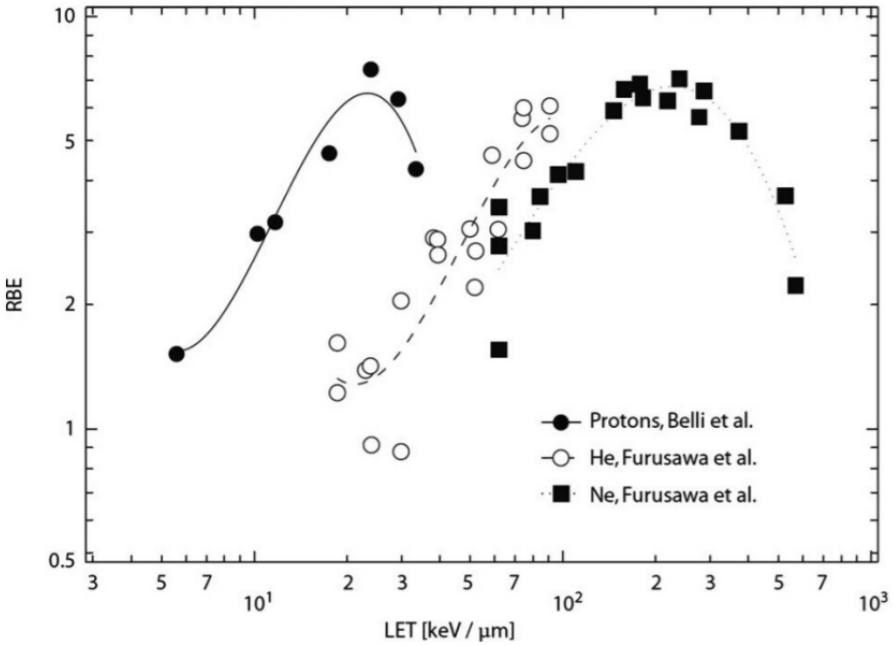

Figure 7: Dependence of the RBE on LET for protons, Carbon and Neon ions for experiments with V79 hamster cells. Figure taken from [Scholz, 2003].

The contribution of the present work to the field

Besides the general introduction given above, the main features and quantities of the common thread of the whole work, namely microdosimetry, are given in Part I.

Moreover, an overview of the most useful details of the experiments, Monte Carlo simulations, data analysis and machine learning techniques used throughout the whole work are presented in Part II.

Finally, the core of this thesis can be divided into three main parts, focusing on different aspects of microdosimetry. A summary of the three parts is given below.

Microdosimetric radiation field characterization

Several data-taking campaigns using a commercial TEPC have been carried out, performing experiments to characterize the radiation quality in different positions inside a water phantom, both in- and off-beam.

In the experiments, we mainly aimed at studying the radiation fields coming from secondary particles and scattered primary particles in the so called *out-of-field* region, that would correspond to the normal tissues around the irradiated tumour in a clinical scenario.

A set of measurements have been taken at the Trento Protontherapy Center with a monoenergetic therapeutic proton beam and a *Spread Out Bragg Peak* (SOBP). In both cases, we evaluated the radiation quality and assessed in addition the potential radiobiological damage in the different regions.

Analogous measurements have been also performed with ions heavier than protons, namely helium and oxygen.

Finally, microdosimetry finds also an application in space radioprotection. In this perspective, two experimental data campaigns have been carried out at GSI *Helmholtz Centre for Heavy Ion Research* in Darmstadt (Germany), with the goal of assessing microdosimetric energy depositions spectra of neutrons.

This part is based on the following original works:

- **Missiaggia, M.**, et al "Microdosimetric measurements as a tool to assess potential in-and out-of-field toxicity regions in proton therapy." *Physics in Medicine & Biology* (2020) [Missiaggia et al., 2020]
- **Missiaggia, M.**, et al "In- and off-beam microdosimetric characterization of a 148-MeV proton spread-out Bragg peak: TEPC measurements and TOPAS benchmark." *Submitted to International Journal of Radiation Oncology, Biology, Physics* (2022) [Missiaggia et al., 2022b]
- **Missiaggia, M.**, et al "Helium and oxygen beams microdosimetric in- and out-of-field characterization." *In preparation* (2022) [Missiaggia et al., 2022a]
- Horst, F., Boscolo, D., Cartechini, G., Durante, M., Hartel, C., Kozlova, E., La Tessa, C., **Missiaggia, M.**, et al "A multi-detector experimental setup for the study of space radiation shielding materials: measurement of secondary radiation behind thick shielding and assessment of its radiobiological effect" *European Physical Journal – Web of Conferences* (2021) [Horst et al., 2022]

Hybrid Detector for Microdosimetry (HDM): a new tool for extending the microdosimetric information

A feature shared by every microdosimeter is that the *lineal energy* y (microdosimetric counterpart of the LET) of each particle is obtained by dividing the energy deposition ϵ with the mean chord length l traversed in the detector. While ϵ is directly measured, the value of l is calculated as the average path travelled by the particle inside the detector, and thus depends both on the detector geometry and on specific assumptions on the radiation field (typically considered isotropic and uniform).

To investigate the validity of this approximation, we performed Monte Carlo calculations using GEANT4 toolkit. We analyzed the track length distributions of all particles traversing the TEPC and found that the mean chord was not always a good representative of the whole population.

In order to experimentally achieve the real track length information without relying on the mean chord length approximation, we designed a new hybrid 2-stage microdosimeter (HDM: hybrid detector for microdosimetry) composed of a spherical TEPC followed by four Low Gain Avalanche Detectors (LGADs).

HDM provides lineal energy spectra in tissue-equivalent with an event-by-event measurement of the path length and a submillimetric spatial resolution. To assess the detector performances, in the feasibility study we tested different configurations (distance between detectors, number of strips in a single LGAD) and studied HDM response when irradiated with protons and carbon ions at different water depths.

We found out that the detection efficiency is the most critical issue. To improve it, we exploited modern *Machine Learning* (ML) techniques, and developed a model composed by two modules: the first one aims at improving the detector efficiency, filling the missing

spatial point values on the LGADs; the second one reconstructs the tracks of the particles to calculate microdosimetric spectra using the real track length.

This part is based on the following original works:

- **Missiaggia, M.**, et al. "A novel hybrid microdosimeter for radiation field characterization based on TEPC detector and LGADs tracker: a feasibility study." *Frontiers in Physics* (2021) [Missiaggia et al., 2021]

- **Missiaggia, M.**, et al "Machine learning techniques applied to therapeutic energies particle tracking with a novel Hybrid Detector for Microdosimetry (HDM)" *Submitted to Physics in Medicine & Biology* (2022) [Missiaggia et al., 2022d]

The Generalized Stochastic Microdosimetric Model (GSM^2)

Currently, the only two radiobiological models used in clinical applications are the Microdosimetric Kinetic Model (MKM) and the Local Effect Model (LEM). The main limitation shared by both models is the assumption that all physical and biological variables follow a Poisson distribution. This hypothesis neglects stochastic fluctuations of energy deposition both from cell to cell and within dose fractions. Although some generalizations to overcome the Poissonian assumption have been developed, [Bellinzona et al., 2021], a comprehensive stochastic description of the radiation-induced DNA damage formation and dynamics accounting for both spatial and temporal features of the dose deposition are still missing.
To overcome this limitation, we have developed the generalized stochastic microdosimetry model (GSM^2) [Cordoni et al., 2021]. By modeling the probability distribution of DNA damages, GSM^2 provides a general probabilistic framework to describe the damage formation and evolution.
One of the most relevant strengths of GSM^2 is the capability to efficiently treat the different levels of spatio-temporal stochasticity for an irradiation. An extensive study of the cell survival probability for acute irradiation conditions (as it is the case in particle therapy) has been carried out in this part, showing GSM^2 potentiality to provide a better ground for the mechanistic interpretation of cell killing compared to the existing models. As a relevant consequence, we showed how GSM^2 provides a generalization to the multi-hit model, that accounts for non-Poissonian effects and damage repair. In addition, a separate work focusing on the different level of stochasticities and their effect on the cell survival curve has been carried out.
Finally, GSM^2 provided an ideal mathematical framework for the information provided by HDM, being able to account for the whole microdosimetric distribution to predict survival curves, instead of using just mean values like the majority of existing RBE models.

This part is based on the following original works:

- Cordoni, F., **Missiaggia, M.**, et al "Generalized Stochastic Microdosimetric Model: the main formulation." *Physical Review E* (2021) [Cordoni et al., 2021]

- Cordoni, F., **Missiaggia, M.**, et al "Cell survival computation via the Generalized Stochastic Microdosimetric Model (GSM^2) - Part I: the theoretical framework" *Radiation Research* (2022) [Cordoni et al., 2022b]

- **Missiaggia, M.**, et al "Cell survival computation via the Generalized Stochastic

Microdosimetric Model (GSM^2) - Part II: numerical results via TOPAS microdosimetric extension" *Submitted to Radiation Research* (2022) [Missiaggia et al., 2022c]

- Cordoni, F., **Missiaggia, M.**, et al "Multiple Levels of Stochasticity accounted in different Radiation Biophysical Models: from physics to biology" *Submitted to International Journal of Radiation Biology* [Cordoni et al., 2022a] (2022)

Part I

The framework: Microdosimetry

Microdosimetry, in its present sense, was founded entirely on an original approach introduced by Harald H. Rossi[6], when he recognized the fundamental difference between macroscopic absorbed dose and the energy deposition in microscopic structures. He realized that the important quantities that described the problem at microscopic scale were inherently "stochastic variables". He and his colleagues proceeded then to develop sophisticated techniques for measuring the random fluctuations of energy deposition, to construct a novel conceptional and mathematical framework, and to apply the new concepts and methods to radiobiology [Santa Cruz, 2016].

Albrecht M. Kellerer[7] told the story of the beginning of microdosimetry in his paper "Microdosimetry: reflections on Harald Rossi" (2002) [M. Kellerer, 2002], where he reports:

> When Harald began to work for Giachino Failla[8] it was the attraction of two similar minds. They both liked refined instrumentation and sophisticated physics and they combined this with a passion for mathematics — real mathematics, not mere computing. So it seemed natural that Failla handed a rather intricate problem to his young co-worker, he asked him to measure LET distributions in various radiation fields. The LET distributions were, of course, to serve as a suitable quantification of radiation quality, as they still do in the quality factor for radiation protection.
> Harald loved the problem. It gave him a perfect opportunity to engage in the construction of fancy proportional counters, to use refined calibration procedures, and to subject the measured data to sophisticated mathematical unfolding procedures. Such procedures were required because LET is a statistical average of energy loss by a particle of specified type and energy. Energy loss straggling would wash out the LET spectra, or — if the pressure in the counter was sufficiently increased to reduce straggling — the ranges of some of the charged particles would be too short, which again would distort the spectra. A true challenge to measurement and analysis.
> A lesser mind would have immersed himself cheerfully into the challenge and might have got lost in it permanently. Harald did start to get immersed by doing fine work on energy loss straggling and he could have continued. But he had the wisdom to step back at the right moment and to reconsider the problem. Then he realised that it was pointless to unfold the measured data. The data were a poor representation of LET, but they were more meaningful than the LET distributions. They represented the actual energy distributions and these — not the somewhat academic LET distributions — determined the effect of the radiation on the cell. This was the birth of the new concepts that were preannounced in the paper by Rossi and Roesch, and it was the birth of the new science of stochastic dosimetry. The expression stochastic dosimetry is here used, because it took a few years before the new branch of dosimetry received its current name. When initially — and somewhat before the first Symposium in Stresa — a name was sought for the newly born branch of radiation research Harald came up, tentatively, with the term micrology which expressed his notion that the microgeometry of the physics of energy deposi-

[6] Harald H. Rossi (1917–2000) was an American physicist.
[7] Albrecht M. Kellerer is a German physicist.
[8] Gioacchino Failla (1891 – 1961) was an Italian-born American physicist.

> tion and of the critical cell structures were equally important in the analysis. Fortunately an ancient dictionary was at hand at the occasion, and micrology happened to be defined there as the pursuit of unimportant matters. That settled the issue and the more orthodox name microdosimetry was then adopted.

The first formulation of microdosimetry was based on the concept of sites that are regions of specified dimensions in which the energy absorbed from ionizing radiations is considered without regard to its microscopic distribution within a site. This approach, termed *regional microdosimetry*, received major attention because it involves quantities that can more easily be related to radiation effects and that often are subject to accurate measurement.
An alternative, and more advanced form, namely *structural microdosimetry* was originated by Kellerer. Structural microdosimetry permits a detailed description of the microscopic pattern of energy absorption and it is of basic importance because the immediate effect of radiation is essentially determined by the intersection of this pattern with that of sensitive components in irradiated matter [Zaider et al., 1996].
Microdosimetry has both experimental and theoretical aspects. The former deals with the measurement of microdosimetric quantities and the latter is concerned with relations between these, and other more general, physical quantities. The experimental approach is especially suited to regional microdosimetry although the results obtained can be analyzed to provide information on structural microdosimetry. The theoretical approach is on the contrary of primary importance in structural microdosimetry but it can be extended to regional microdosimetry.

Chapter 1

Microdosimetry: quantities and detection

1.1 Quantities and distributions

The formalism on the measured and derived quantities in microdosimetry can be found extensively in [Booz et al., 1983] and in [Zaider et al., 1996]. Here, we report the definitions and the most interesting properties of the microdosimetric quantities relevant for the present work.

1.1.1 Energy deposited and imparted

Energy deposited

The quantity ϵ_i is the *energy deposited* in a single interaction i:

$$\epsilon_i = T_{in} + T_{out} + Q_{\Delta m}$$

where:

- T_{in} represents the kinetic energy of the incident ionizing particle (exclusive of the rest mass);
- T_{out} represents the sum of the kinetic energies of all ionizing particles leaving the interaction (exclusive of the rest mass);
- $Q_{\Delta m}$ represents the changes of the rest mass energy of the atoms and all particles involved in the interaction ($Q_{\Delta m} > 0$ implies a rest mass decrease , $Q_{\Delta m} < 0$ implies a rest mass increase).

It is important to notice that:

- ϵ_i is a stochastic quantity;
- ϵ_i may be considered as the energy deposited at the point of interaction (which is also called transfer point);

In addition, it is useful to consider a particular summation of the energy deposit in the track j of a charged particle of a given type and energy:

$$\tilde{T}_j(x) = \frac{\sum_i \sum_k \epsilon_i \epsilon_k}{\sum_i \epsilon_i}$$

where i runs over all energy transfer points of the track and k runs over all transfer points within a distance up to x from the transfer point.
$\tilde{T}_j(x)$ is a function of the distance x, which reflects the proximity between the transfer points of the track j, where j runs over all n tracks in the selected volume.

Energy imparted

The energy imparted ϵ to matter in a specific volume is:

$$\epsilon = \sum_i \epsilon_i$$

where the summation is performed over all energy depositions ϵ_i in that volume. It can be noticed that:

- ϵ is a stochastic quantity;
- the energy depositions over which the summation is performed may be related to one or more events, i.e. due to one or more statistically independent particle tracks.

1.1.2 Specific energy

The specific energy, z, is the quotient of ϵ by m, where ϵ is the energy imparted by ionizing radiation to matter of mass m:

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

Further, it is worth noticing that:

- z is a stochastic quantity. The value of the distribution function $F(z)$ is the probability that the specific energy is equal to or less than z. The probability density $f(z)$ is the derivative of $F(z)$ with respect to z:

  $$f(z) = \frac{dF(z)}{dz}$$

  The probability density $f(z)$ includes a discrete component (a Dirac delta function) at $z = 0$ for the probability of no energy deposition.
  The expectation value:

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

  is called mean specific energy and is a non-stochastic quantity.

- The specific energy may be caused by one or more energy deposition events. The distribution function of the specific energy deposited in a single event, $F_1(z)$, is the conditional probability that a specific energy less than or equal to z is deposited if one event has occurred. The probability density, $f_1(z)$, is the derivative of $F_1(z)$ with respect to z:

$$f_1(z) = \frac{dF_1(z)}{dz}$$

and it is also called the single event distribution of z. The expectation value:

$$\bar{z_F} = \int_0^\infty z f_1(z) dz$$

is called *frequency-mean specific energy per event*, and is a non-stochastic quantity.

- The number, n, of energy deposition events which have contributed to a particular specific energy, z, is, in general, randomly distributed, and thus can be described by a Poisson distribution. The mean events number, $\bar{n}$, is the ratio between the mean specific energy, $\bar{z}$, and the frequency-mean specific energy per event, $\bar{z}_F$,

$$\bar{n} = \frac{\bar{z}}{\bar{z}_F}$$

- The mean number of energy deposition events per unit absorbed dose is called the *event frequency*, $\phi^*(0)$

$$\phi^*(0) = \frac{\bar{n}}{D} = \frac{\bar{z}}{\bar{z}_F D}$$

$\phi^*(0) \approx 1/\bar{z}_F$, if $z \approx D$, as it is usually the case.

- It is also useful to consider the dose distribution of z per energy deposition event. Let $D_1(z)$ be the fraction of absorbed dose per event delivered by energy deposition events of specific energy less than or equal to z. The dose probability density, $d_1(z)$, is the derivative of $D_1(z)$ with respect to z:

$$d_1(z) = \frac{dD_1(z)}{dz}\,.$$

The expectation value is

$$\bar{z}_D = \int_0^\infty z d_1(z) dz\,,$$

it is called *dose-mean specific energy per event*, and is a non-stochastic quantity. The relations between $f_1(z)$ and $d_1(z)$, and between $\bar{z}_F$ and $\bar{z}_D$ are

$$d_1(z) = \frac{z}{\bar{z}_F} f_1(z)\,,$$
$$\bar{z}_D = \frac{1}{\bar{z}_F} \int_0^\infty z^2 f_1(z) dz\,.$$

1.1.3 LET limitations and lineal energy definition

Over the past years, it has become more and more evident that there are many and great limitations in the concept of LET as the key-parameter that takes into account the quality of the radiation. In spite of its complexities, the LET concept can, at best, provide a crude characterization of the charged particle tracks that occur in the exposed medium [Kellerer et al., 1985]. Moreover, since the biological damage induced by radiation is directly related to the energy absorbed by the target (such as cells, tissue, etc.), the energy loss measured as an average in a macroscopic volume becomes meaningless in the majority of the cases.
The size of most biological targets is so small that the statistical fluctuations of the amount of energy deposited can be very large. Therefore, in many cases of importance in radiation biology, the hypothesis of exploiting averaging may be inaccurate.
Another important limitation of LET is that this concept is never adequate for electrons. In fact, there are no volumes of interest sufficiently small to disregard the finite range of electrons, and at the same time sufficiently large to discount the energy-loss straggling and lateral escape of δ rays. For heavy ions, on the other hand, there are site sizes and particle energies for which the LET concept predicts adequately the energy deposition. However, even in this case the LET is of limited value since it cannot provide the energy distribution within the sites, although the situation can differ substantially for particles that have the same LET but different velocities.
Rossi suggested that the energy deposited by charged particles and their secondaries in volumes of specified size should replace the use of linear energy transfer as a measure of the radiation quality.
Microdosimetry itself originated when he recognized that the actual energy concentrations determine the biological effect and not the values of LET; the former need then to be analyzed.

The *lineal energy*, y, is then defined as the ratio between the energy imparted to the matter in a volume by a single energy-deposition event (ϵ) and the mean chord length in that volume (l)

$$y = \frac{\epsilon}{l}\,.$$

With this definition, y, unlike LET, has no ambiguity.
It is important to notice that:

- the mean chord length in a volume is the mean length of randomly oriented chords in that volume. For a convex body:

  $$\bar{l} = \frac{4V}{S}\,,$$

  where V is the volume and S is the surface area of this body. The mean chord length of different convex volume is given in Table 1.1 using the following parameters:

  - r is the radius of sphere or cylinder;
  - h is the cylinder length;

	Sphere	**Cylinder**
$\bar{l}$	$\frac{4}{3}\,r$	$\frac{2rh}{r+h}$

Table 1.1: Mean chord lengths of the most used active volumes of gas based microdosimeters.

The lineal energy is a stochastic quantity and is useful to consider its distribution function, $F(y)$, which represents the probability that the lineal energy is equal to or less than y.
The probability density, $f(y)$, is the derivative of $F(y)$ with respect to y:

$$f(y) = \frac{dF(y)}{dy}\,,$$

and it is also called lineal energy distribution. It is important to keep in mind that y is defined for single energy-deposition events only.

- The lineal energy distribution, $f(y)$, is independent of the absorbed dose or dose rate.

- The expectation value:

$$\bar{y}_F = \int_0^\infty y f(y) dy\,,$$

is called *frequency-mean lineal energy*, and it is a non stochastic quantity.

- It is also useful to consider the *dose distribution* of y. Let $D(y)$ be the fraction of absorbed dose delivered with lineal energy less than or equal to y. The *dose probability density*, $d(y)$, of y is the derivative of $D(y)$ with respect to y:

$$d(y) = \frac{dD(y)}{dy}\,.$$

- The distribution $d(y)$ is independent of the absorbed dose or dose rate.

- The expectation value:

$$\bar{y}_D = \int_0^\infty y d(y) dy\,,$$

is called *dose-mean lineal energy*, and it is a non stochastic quantity.

- The relations between $d(y)$ and $f(y)$ and between $\bar{y}_D$ and $\bar{y}_F$ are:

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

and

$$\bar{y}_D = \frac{1}{\bar{y}_F} \int_0^\infty y^2 f(y) dy\,.$$

Quantity	Unit of measurement
ϵ_i	keV
ϵ	keV
y	keV/μm
z	Gy
V	μm^3
S	μm^2
$\bar{l}$	μm
d	g/cm^3

Table 1.2: Units of measurement of the principal microdosimetric parameters and of the parameters used in calculations.

1.1.4 Relationships between microdosimetric quantities

There are several relationships between microdosimetric quantities and their macroscopic analogues. Some of these are of a general nature and follow directly from the definition of the quantity, or can be derived mathematically. Others are restricted to a particular geometry or make use of a specific approximation.

Relationships between ϵ, y and z

The relation between the quantities ϵ, y and z is:

$$\epsilon = y\bar{l} = zm\,.$$

The same relationships hold for the frequency and dose means. Numerical constants in the following formulas are chosen for the set of units reported in Table 1.2.

Then

- $z = ky$;
- $z = \frac{k}{\bar{l}}\epsilon$;
- $z = \frac{\epsilon}{m}$.

With some calculations:

$$\begin{aligned} z[Gy] = z[\frac{J}{kg}] = k\frac{\epsilon}{\bar{l}}[\frac{keV}{\mu\, m}] = \\ = k\frac{\epsilon}{\bar{l}}[\frac{J}{\mu\, m}]\frac{1}{\rho}[\frac{cm^3}{g}] = \\ = k\frac{1.6\,10^{-16}}{\bar{l}}[\frac{J}{\mu\, m}]10^{15}[\frac{\mu\, m^3}{kg}]\,, \end{aligned}$$

from which we obtain that $k = 0.16\,\frac{\bar{l}}{\bar{V}}$.
For the important special case of a spherical volume:

$$k = 0.16\frac{\bar{l}}{V} = 0.16\frac{\frac{4}{3}\frac{d}{2}}{\frac{4}{2}\pi\left(\frac{d}{2}\right)^3} = \frac{0.204}{d^2}\,.$$

Then, since in this work it has been always used a spherical microdosimeter, the absorbed dose in Gy is simply given by the following relationship

$$z[Gy] = \frac{0.204}{d^2\,[\mu\,\mathrm{m}^2]}y\,[\mathrm{KeV}/\mu\,\mathrm{m}]\,. \tag{1.1}$$

1.2 Experimental microdosimetry

Detectors for microdosimetry are based on on gas counters or on solid-state counters and they aim at providing the energy depositions of each event in their active volume.
The first microdosimeters were gas proportional counters made with tissue-equivalent plastic and filled with tissue-equivalent gas mixtures. Because of that, they were called tissue-equivalent proportional counters (TEPCs).
In the early 1950s, H.H. Rossi and W. Rosenzweig devised a spherical TEPC to evaluate the LET in complex radiation fields. Since the LET is defined as an average quantity at a point, they proposed to operate the proportional counter at low pressure so that the product of the gas density and the detector diameter was of the order of 10^{-4} g/cm^2, which was equivalent to 1 micrometer of material of unit density [Rossi and Rosenzweig, 1955]. Measurements of the energy deposited in that small simulated site when exposed to neutron irradiation showed a much greater variation in the distribution of the recoil protons path lengths in the sphere, compared to the ones measured at standard pressures. This was a strong demonstration of the stochastic nature of energy deposited by charged particles, a phenomenon that was well known from theoretical considerations but that had been observed primarily as the qualitative variation between traces made by identical ionising particles in cloud chambers. Rossi recognised that the range of energy deposits observed in the simulated micrometer volume was characteristic of the energy deposited in biological cells; this was pointed out to be significant in initiating biological damage [Braby, 2015].

TEPCs are characterized by an optimum tissue-equivalence, and their response to primary and secondary charged particles is accurate over a wide energy depositions range. In addition, they have a high detection efficiency, thanks to the electrons multiplication in the filling gas.
For these and other reasons, different types of TEPC have been built and used fro radiation field characterizations in the last decades. Miniaturized TEPCs (mini-TEPCs) with sensitive volumes smaller than 1 mm have been manufactured in order to operate at therapeutic intensities [De Nardo et al., 2004], and they operate in the range from about 0.3 to 2 μm of equivalent tissue diameter.
Other prototypes of TEPCs have been studied and developed to be used in ion-beam therapy, such as the low-pressure avalanche-confinement TEPC [Mazzucconi et al., 2019] or the gas electron multiplier (GEM) TEPC [Farahmand et al., 2004].

The study of semiconductor detectors for microdosimetry begins instead in 1980 [Dicello et al., 1980]. Since then, several devices coupled to tissue-equivalent (TE) converters have been employed for silicon microdosimetry [Orlic et al., 1989]. More recently, a silicon

microdosimeter consisting of an array of microscopic p–n junctions based on the silicon-on-insulator (SOI) technology was fabricated and tested with various radiation fields for hadron therapy [Rosenfeld et al., 2000]. In addition, a silicon telescope has also been proposed for microdosimetry [Agosteo et al., 2006].
The sensitive volumes of silicon detectors can be micrometric, and thus of great interest for microdosimetry. They can be used for measuring the quality of radiation therapy beams, coupling them to TE converters. The use of micrometric volumes avoids the contribution of wall effects to the measured spectra. These effects represent one of the main issues when using TEPCs: events can deposit their energy in the TEPC sensitive volume (such as charged particles created in the TEPC walls), which would be partially acquired in a micrometric site of tissue.
Further, in the TEPCs the electrodes are biased at several hundreds of volts, they need an accurate gas pressure control and they have limited capabilities in high-intensity radiation fields because the geometrical cross-sectional area of the sensitive volume cannot be reduced below 1 mm^3. Solid-state microdosimeters, instead, are biased at low voltage and the thickness of the sensitive volume can be as small as 1 μm, making them suitable for measurements even in very intense radiation fields. Moreover, pixelated arrays of detectors can be constructed, allowing a simultaneous two-dimensional mapping of the radiation field quality [Colautti et al., 2020]. Further advantages of such detectors are its compactness, cheapness, transportability, low power consumption and low sensitivity to vibrations. However, the following problems need to be solved when using a silicon device for microdosimetry: (i) the sensitive volume has to be confined in a region of well-known dimensions; (ii) the electric noise limits the minimum detectable energy, due to the high-capacitance of a very thin sensitive zone; (iii) corrections for tissue equivalency should be made for the silicon-sensitive zone; (iv) correction for shape equivalency should be made when referring to a spherical site, since the sensitive zone of a silicon device can be approximated to a parallelepiped; (v) the angular response should be evaluated carefully; and (vi) the efficiency of a single detector of micrometric dimensions is very poor and detector arrays should be considered [Agosteo and Pola, 2011].

Based on the pros and cons of existing microdosimeters, we selected a commercial TEPC model LET-1/2 from Far West Technology for all experimental and simulated measurements related to this work.

1.2.1 TEPC LET-1/2

The TEPC model LET-1/2 from Far West Technology is a proportional counter that employs a spherical chamber with tissue equivalent walls and a tissue equivalent filling gas. The spherical shape guarantees that its response would be independent of the direction from which the radiation comes.

The detector is a spherical cavity with a 12.7 mm internal diameter. An aluminum shell of 1.8 mm thickness is used as a vacuum tight container for the sphere and is mounted on an aluminum stem. This allows the sphere to be placed below the surface of a tissue-equivalent fluid phantom. The inner diameter of the TEPC gas filled sphere is 12.7 mm and the wall thickness is 1.27 mm.
Table 1.3 summarize the TEPC LET-1/2 specifications.

Aluminium stem	
Length	254 mm
Diameter	6 mm
Aluminium shell	
Length	6.35 cm
Wall thickness	1.8 mm
Diameter	2 cm
Sphere	
Wall thickness	1.27 mm
Internal diameter	12.7 mm

Table 1.3: TEPC LET-1/2 specifications.

H	**C**	**N**	**O**	**F**	**Na**	**Mg**	**Si**	**P**	**S**	**K**	**Ca**
ICRU tissue, muscle (ICRU, 1964)											
10.2	12.3	3.5	72.9	-	0.08	0.02	-	0.2	0.5	0.3	0.007
Muscle-equivalent plastic A-150 (Smathers et. al., 1997)											
10.1	77.6	3.5	5.2	1.7	-	-	-	-	-	-	1.8
Muscle-equivalent gas, with propane (Srdoc, 1970)											
10.3	56.9	3.5	29.3	-	-	-	-	-	-	-	-

Table 1.4: Elemental composition of compounds and mixtures in percentage by weight of the so-called *A-150* muscle-equivalent plastic and the muscle-equivalent gas propane-based, compared to the ICRU muscle definition.

The simulation of biological tissue

In order to achieve the simulation of biological tissue, both the filling gas and the sphere wall are required to be tissue-equivalent.
The most commonly used gases are methane-based or propane-based. From the production company, TEPC LET-1/2 was filled with a mixture of gas, propane-based, called *TE-propane* gas. Besides using this compound as the gas filling the TEPC in a part of the current work (7), a refill of the gas was needed after a long period of use. Being easily available and being demonstrate its equivalence in [Chiriotti et al., 2015a], we chose the pure propane gas to fill the TEPC. For this reason, in the majority of the experiments and Monte Carlo simulations presented here, we used the detector with this mixture.

The most widely used material in microdosimetry, and the one composing the TEPC LET-1/2 sphere wall is the muscle-equivalent plastic formulation, called *A-150*. Table 1.4 shows the elemental composition of muscle-equivalent compounds and mixtures in percentage by weight, together with the muscle-equivalent propane-based gas, compared to the ICRU 1964 muscle definition [ICRU, 1964].

A-150 is a mixture of polyethylene, nylon, carbon and calcium fluoride. The elemental composition of *A-150* is very close to standard muscle tissue with regard to hydrogen and nitrogen. There is, however, a substitution of carbon for oxygen; in fact, no practically usable solid has an oxygen content equal to that of tissue.

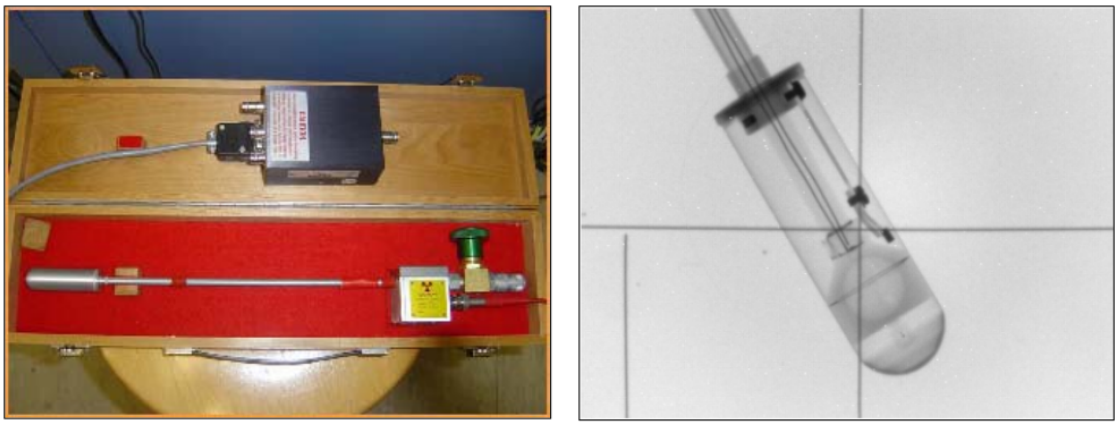

Figure 1.1: The TEPC and its X-rays image (Model LET-1/2, Far West Technology, Inc., Goleta, CA, USA).

Simulating a microscopic volume

In order to relate the energy deposition to a microscopic volume of biological tissue, the latter is replaced by a much larger cavity filled with a tissue-equivalent gas of much lower density. To achieve the simulation of a microscopic volume, the physical size of the detector and the density of the gas are used as variables, as they form a conjugate pair in controlling the site size simulation.
In order to be equivalent, it is required that the energy loss of the charged particles traversing the tissue sphere is the same as in the gas sphere, for equivalent trajectories. Let ΔE_t and ΔE_g be the mean energy losses of charged particles in tissue and gas, respectively

$$\Delta E_t = \left(\frac{S}{\rho}\right)_t \rho_t d_t \,,$$
$$\Delta E_g = \left(\frac{S}{\rho}\right)_g \rho_g d_g \,,$$

where $\left(\frac{S}{\rho}\right)_t$ and $\left(\frac{S}{\rho}\right)_g$ are the mass stopping powers, ρ_t and ρ_g are densities and d_t and $d_g = d_t \kappa_{gt}$ are the diameters of the tissue and gas spheres, respectively.

So,

$$\Delta E_t = \Delta E_g \,.$$

By definition of tissue equivalence, the mass stopping power of the biological tissue and the tissue-equivalent gas are the same. Then, the condition that has to be verified to achieve simulation of a tissue microscopic volume is

$$\rho_g = \frac{\rho_t}{\kappa_{gt}} = \rho_t \frac{d_t}{d_g} \,. \tag{1.2}$$

By minimizing the detector volume, a higher gas density is obtained for a given simulated site size. Substituting ρ_t and d_g in (1.2), the gas density for different site sizes is calculated. By using the gas density values in the ideal gas equation, the pressure required to simulate different site sizes can be obtained as

$$PV = nRT\,,$$

or

$$PV = \frac{\rho}{M}RT\,,$$

where ρ is the gas density (in g/cm^3), M is the gas molecular weight, R is the gas constant, T is assumed to be the room temperature of 293.15 K.
In our case, the filled propane gas is at a pressure of 60 mbar in order to be equivalent to 2 μm of tissue, and the corresponding density is:

$$\rho = \frac{P[atm] \cdot M[g/mol]}{R[\frac{m^3 \cdot atm}{K \cdot mol}] \cdot T[k]} = 1.08 \cdot 10^{-4} g/cm^3 \tag{1.3}$$

where $R = 8.206 \cdot 10^{-5} \frac{m^3 \cdot atm}{K \cdot mol}$, $M = 44.1 g/mol$.

Chapter 2

Microdosimetry-based models for cell survival prediction

An accurate prediction of the relative biological effectiveness (RBE) of the irradiating field is a fundamental requirement to correctly estimate treatment responses [Durante and Loeffler, 2010]. The RBE value depends on several factors of different nature, both biological and patient specific, because of the complexity of the mechanisms of action underlying tumour and normal tissue responses in radiation therapy.
A numbers of models have been developed to predict RBE, attempting to account for such effects. They can be divided into four main categories:

1. purely phenomenological models: NIRS[1] mixed beam approach [Kanai et al., 1997, Kanai et al., 1999, Gueulette et al., 2007, Tsujii et al., 2004];

2. linear LET_d-based models (developed mainly for protons) [Carabe et al., 2012, Wedenberg and Toma-Dasu, 2014, Jones, 2015, McNamara et al., 2015];

3. Local Effect Model (LEM)-based models [Scholz et al., 1997, Scholz and Kraft, 1992, Scholz and Kraft, 1996, Elsässer and Scholz, 2007, Elsässer et al., 2008, Elsässer et al., 2010, Friedrich et al., 2012];

4. models based on microdosimetry concepts:

   (a) models based on the Microdosimetric Kinetic Model (MKM), initially proposed by Hawkins in 1994 [Hawkins, 1994] and then explored and extended [Hawkins, 2003, Kase et al., 2006, Kase et al., 2007, Sato and Furusawa, 2012, Manganaro et al., 2017];

   (b) other models, such as the repair-misrepair-fixation RMF model [Carlson et al., 2008, Frese et al., 2012, Stewart et al., 2018], and phenomenological models based on RBE-weighting functions [Pihet et al., 1990, Menzel et al., 1990, Wambersie, 1994, Wambersie et al., 1990].

All the different models present different advantages and limitations. Even if RBE is not physically measurable, models belonging to category 4 provide a strong link with microdosimetric physics measures and they will be therefore the focus of the present work's theoretical part.

[1] National Institute of Radiological Sciences (NIRS, Chiba, Japan)

2.1 Single- and multi-event microdosimetric distributions

When dealing with *specific energy spectra*, it is important to distinguish between the *single-event distribution* and the *multi-event distribution*. Although the experimental microdosimetry achieves single event quantities such as the ε_1 or the *lineal energy* y, the starting point for the models are multi-event quantities such as the *specific energy* z and its distribution.
The *single–event distribution*, denoted by $f_1(z)$, is the probability distribution of z conditioned to the fact that exactly a single-event happened. The single–event distribution is the building block to define the more-general $n-$event distribution $f_n(z)$ and the *multi–event distribution* $f(z)$.
The $n-$event distribution $f_n(z)$, that is the probability distribution conditioned to the fact that precisely n events occurred, can be computed as the $n-$fold convolution of the single-event distribution $f_1(d)$, as follows [Zaider et al., 1996]

$$\begin{aligned} f_2(z) &:= \int_0^\infty f_1(\bar{z}) f_1(z-\bar{z}) d\bar{z}\,, \\ &\dots\,, \\ f_n(z) &:= \int_0^\infty f_1(\bar{z}) f_{n-1}(z-\bar{z}) d\bar{z}\,, \end{aligned} \tag{2.1}$$

Using the $n-$event distributions defined above, we can define the general *multi–event distribution* as

$$f(z;\lambda_n) := \sum_{n=0}^{\infty} p(n;\lambda_n) f_n(z)\,, \tag{2.2}$$

with $p(n;\lambda_n)$ an integer valued probability distribution with average λ_n, meaning that

$$\lambda_n := \sum_{n=0}^{\infty} n p(n;\lambda_n)\,.$$

The *multi–event distribution* $f(z;\lambda_n)$ plays a crucial role in the development of microdosimetric-based radiobiological models. The quantity $f(z;\lambda_n)$ depends on the number of events n only through $p(n;\lambda_n)$, which is independent of the specific energy z. Furthermore, given $p(n;\lambda_n)$, the single–event distribution f_1 completely determines the multi-event distribution $f(z;\lambda_n)$.
Typically, since events are statistically independent, $p(n;\lambda_n)$ is assumed to be a Poisson distribution with mean value λ_n, so that Equation (2.2) becomes

$$f(z;\lambda_n) := \sum_{n=0}^{\infty} e^{-\lambda_n} \frac{\lambda_n^n}{n!} f_n(z)\,. \tag{2.3}$$

Denoting by $\langle z \rangle$ the first moment of the distribution $f(z;\lambda_n)$, formally

$$\langle z \rangle := \int_0^\infty z f(z;\lambda_n) dz\,, \tag{2.4}$$

from which it follows that the subsequent relation holds true,

$$\langle z \rangle = \lambda_n z_F\,, \tag{2.5}$$

In microdosimetry, $\bar{z}$ is often identified with the absorbed dose D, and we will adopt this identification in the chapter.
Above argument, with particular reference to Equation (2.5), leads to the average value λ_n of the *multi–event distribution* to be [Zaider et al., 1996, Booz et al., 1983]

$$\lambda_n = \frac{D}{z_F}\,, \tag{2.6}$$

From now on, we will consider λ_n to be defined as in Equation (2.6), unless specified otherwise.
Further computations [Zaider et al., 1996, Booz et al., 1983] show that, for what concerns the second moment, it holds that

$$\int_0^\infty z^2 f(z;\lambda_n)dz = D^2 + z_D D\,, \tag{2.7}$$

with z_D the *dose average* of the single-event specific energy

$$z_D := \frac{1}{z_F}\int_0^\infty z^2 f_1(z)dz = \frac{\int_0^\infty z^2 f_1(z)dz}{\int_0^\infty z f_1(z)dz}\,. \tag{2.8}$$

The notation and the computations presented in the current section will be extensively used thorough the work to formally derive analytical solutions for some relevant biological endpoints, typically the cell–survival probability, starting from a mathematical model for the DNA damage.
In the following, we assume that a cell nucleus is divided into N_d domains, so that the microdosimetric distributions will be used both on single domains and on the whole cell nucleus. In particular, the superscript (c,d) will denote that the corresponding quantity, such as as a microdosimetric distribution or the corresponding average value, is considered on the domain d of the cell c. Further, the subscript n denotes that the microdosimetric distributions are calculated on the cell-nucleus, whereas if no subindex is specified, it is assumed that the corresponding distribution is derived on the domain.
In addition, whenever we say that we average a function $g(z)$ over all domains of a cell nucleus, denoted for short by $\langle g\rangle_d^{(c)}$, it formally means

$$\langle g\rangle_d^{(c)} := \frac{1}{N_d}\sum_{d=1}^{N_d}\int_0^\infty g(z) f^{(c,d)}(z;z_n)\,dz\,. \tag{2.9}$$

where $f^{(c,d)}(z;z_n)$ denotes the probability density of z in a domain for cell with nucleus specific energy z_n.
Similarly, by averaging a function $g_n(z)$ over all cell population denoted by $\langle g_n\rangle_c$, we mean

$$\langle g_n\rangle_c := \frac{1}{N_c}\sum_{c=1}^{N_c}\int_0^\infty g_n(z) f_n^{(c)}(z;D))\,dz\,. \tag{2.10}$$

where N_c is the total number of the considered cells and $f^{(c)}(z;D)$ indicates the probability density of z in a nucleus for a population of cells irradiated with a macroscopic dose D. A classical assumption in the computations of the survival probability for cell population, the microdosimetric spectra are typically reasonably considered equals among different cells

and domains. In this case, we will drop the indexes c and d, and the sums in Equations 2.9 and 2.10 can be carried out implicitly:

$$\langle g \rangle_d = \int_0^\infty g(z) f(z; z_n)\, dz \tag{2.11}$$

$$\langle g_n \rangle_c = \int_0^\infty g_n(z) f_n(z; D))\, dz\,. \tag{2.12}$$

2.2 Microdosimetric Kinetic Model

The Microdosimetric Kinetic (KM) model was developed by Roland B. Hawkins [2] [Hawkins, 1994] by taking inspiration from the theory of dual radiation action (TDRA) [Kellerer and Rossi, 1972, Kellerer and Rossi, 1978], the repair-misrepair model [Tobias, 1980, Tobias, 1985] and the lethal-potentially lethal (LPL) model [Curtis, 1986, Curtis, 1988].
Figure 2.1 represents a conceptual scheme of the main MKM formulations and extensions.

Historical bases

The theory of dual radiation action (TDRA) [Kellerer and Rossi, 1972, Kellerer and Rossi, 1978] assumes that, after the cell irradiation, the number of lethal lesions in a small volume of the cell nucleus, defined *site*, is proportional to the square of the specific energy z deposited in that site:

$$\varepsilon(z) = Kz^2 \tag{2.13}$$

where the K factor expresses average yield of lethal lesion for a imparted dose of z. In the MKM development [Hawkins, 1994], this postulate is generalized by adding the quadratic proportionality to the lethal damages, and therefore implying a linear-quadratic dependence of the survival probability on z.
The MKM inherits the concept of damage time evolution for the repair or conversion into a lethal non reparable lesion (chromosome aberration) [Van Houten et al., 2018, Schürmann et al., 2018] of the primary potentially lethal radiation induced lesions in DNA from the *repair-misrepair (RMR) model*, developed by Tobias *et al.* to interpret radiobiological experiments with heavy ions [Tobias, 1980, Tobias, 1985]. The RMR model considers that the amount of DSBs in the DNA, $U(t)$, is linearly proportional to the radiation dose rate. A number of DSBs evolve in lethal lesions, $L(t)$, while most breaks are successfully repaired with a first-order process. The model includes also the possibility of a misrepair as a second-order process since it involves two broken DNA strands to form a chromosomal aberration. The idea of misrepair was initially applied by Lea and Catcheside [Lea and Catcheside, 1942] to describe the formation of chromosome aberrations in tradescantia.
These assumptions yield the following kinetic equations:

$$\begin{aligned} \frac{dU}{dt} &= \underbrace{\delta \dot{D}}_{\text{damage}} - \underbrace{\lambda U}_{\text{repair}} - \underbrace{\kappa U^2}_{\text{misrepair}} \\ \frac{dL}{dt} &= \underbrace{(1-\phi)\lambda U}_{\text{unsuccessful repair}} + \underbrace{\sigma \kappa U^2}_{\text{lethal misrepair}} \end{aligned} \tag{2.14}$$

[2] Roland B. Hawkins is an American physicist.

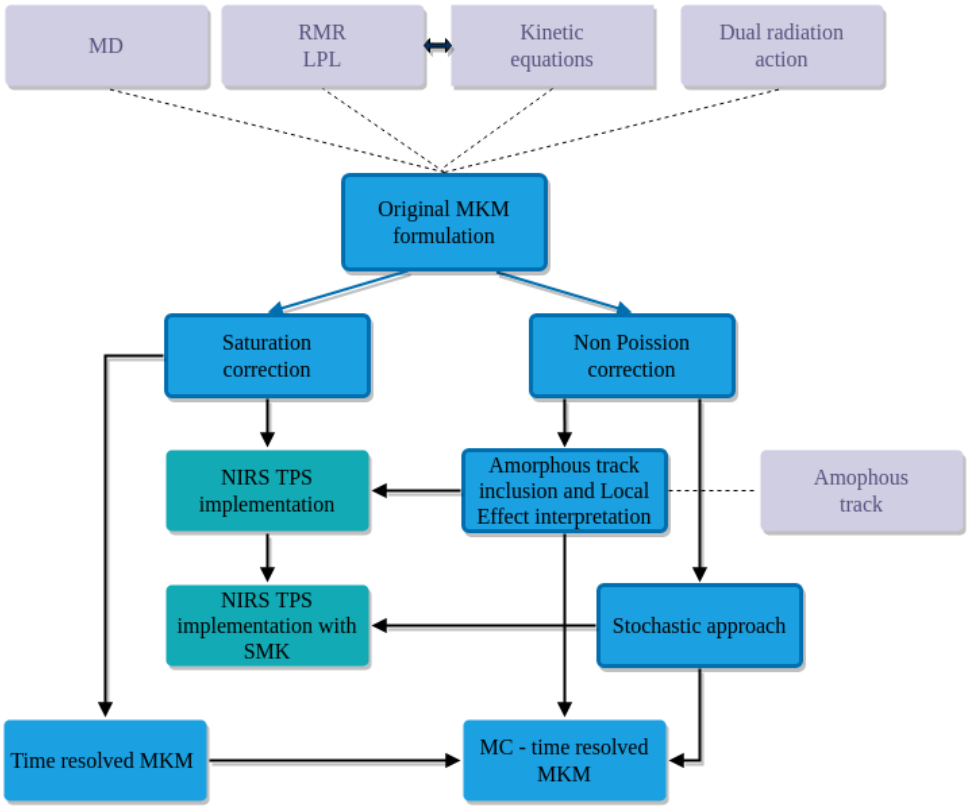

Figure 2.1: Conceptual map of the evolution of some of the microdosimetric kinetic models (blue) developed over the years. Some of these models are currently used for RBE and RBE-weighted dose evaluations in TPS applications (dark cyan). The dotted lines mark the theoretical bases of the considered formulation *(light gray color)*. In the figure, MD stands for *MicroDosimetry*, RMR stands for *Repair-MissRepair* model, LPL stands for *Lethal-Potentially Lethal* model, SMK stands for *Stochastic Microdosimetric Kinetic* model and MC stands for *Monte Carlo*.

where δ is the number of DSBs induced per Gy of radiation, λ is the rate at which the DSBs are repaired, κ is the rate constant for second-order DSB interaction, and ϕ is the fraction of simple repairs that are successful. The fraction of misrepairs that result in a lethal lesion is σ.

Like the RMR model, the *lethal-potentially lethal (LPL)* model [Curtis, 1986, Curtis, 1988] accounts that the damage caused by ionizing radiation at the molecular level to cell death can be separated into two broad classes: damages that have the potential of being lethal, $P(t)$ (by fixing or binary misrepair) but also can be repaired correctly and that which is lethal *ab initio* and cannot be repaired correctly, $L(t)$. Both lesions are linearly proportional to the radiation dose-rate [Kuang et al., 2016], and after a prescribed time, the remaining potentially lethal lesions become lethal as described in the following equations:

$$\begin{aligned} \frac{dP}{dt} &= \underbrace{\delta\eta\dot{D}}_{\text{reparable damage}} - \underbrace{\lambda P}_{\text{repair}} - \underbrace{\kappa P^2}_{\text{misrepair}} \\ \frac{dL}{dt} &= \underbrace{\delta(1-\eta)\dot{D}}_{\text{irreparable damage}} + \underbrace{\kappa P^2}_{\text{lethal misrepair}} \end{aligned} \tag{2.15}$$

where η is the amount of radiation induced DSBs that are repairable, while all the other parameters are the same as Equation (2.14).
The solution of the model equations are similar in form to those for the RMR model and the LPL. However, unlike the RMR, the LPL predicts that the probability of the interaction between potentially lethal lesions is strongly dependent to the dose rate and becomes negligible for low dose-rates, where only direct creation of lethal events through λ dominate.

Original formulation and general considerations

The MK model is based on the following funding assumptions [Hawkins, 1994, Hawkins, 1996, Hawkins, 1998]:

1. the cell nucleus is the sensitive target and it is divided into N_d sub-units, called domains, similar to the sites of the TDRA. In general, the domains have a variety of shapes that fit together to fill the nucleus. In the case of mammalian cells, the domain diameter is usually considered to be in the range $0.5 \leq d_d \leq 1.0\,\mu$m and the number of domains per nucleus is in the order of few hundred;

2. radiation can create two different types of DNA damages, called of type I and II;

3. type I lesions represents a damage that cannot be repaired, for this reason it will be also called lethal lesion. On the contrary type II lesions, also called sublethal or potentially-lethal lesions, can repair or convert into a lethal lesion either by spontaneous conversion or by binary combination with another sublethal lesion;

4. type I and II lesions are confined to the domain in which they are created. This assumption defines a sub-nuclear correlation length among lesions, because the interaction of two lesions can happen only if they are in close spatial proximity. Specifically, a pair of type II lesions can combine to form a type I lesion only if they are created in the same domain;

   The idea behind the division of a cell into subvolumes arises because couples of type

II lesions are all likely to happen in a short time period, even for lesions that are far away in the cell-nucleus. In order to overcome such a problem, a possible approach is to divide the nucleus into smaller subdomains so that interactions might happen only inside a single volume, as it is assumed in the MKM. Choosing the domains size plays a key role in the predicted cell-survival: if the domains are too big, then lesions that are created far away can interact, while on the contrary, if the domains are too small, then that the overall number of lesions inside a single domain is so small that couple interactions is less likely to happen. Therefore, different choices of domains can in principle lead to different model predictions. A possible solution to reduce the model sensibility from the arbitrary choice of the domains, it is to assume that interactions are possible also within different domains. This hypothesis allows lesions to move from one domain to another, or pairs of lesions to interact if they are formed in adjacent domains.

5. the initial number of type I and II lesions in a single domain d is proportional to the specific energy z deposited in the domain.

If above assumptions hold then the following further assumptions are made:

6. if at least one domain suffers a lethal lesion, then it is considered "dead";

7. if at least one domain is dead, then the whole cell is "dead".

While the MK assumptions reported in this section are general, in many studies [Matsuya et al., 2014, Chen et al., 2017] the lethal lesions are intended to represent a specific complex DNA damage (e.g. lethal chromosome aberrations) that cannot be repaired, whereas the creation of sublethal lesions are explicitly associated to the induction of DSB that can be repaired.
Following the MK notation, we denote by $x_I^{(c,d,z)}(t)$ and $x_I^{(c,d,z)}(t)$ the time-dependent average number of type I and type II lesions for a cell-domain (c,d) caused by an acute dose $z^{(c,d)}$ deposited in the cell c and domain d at $t=0$. Starting from the concept introduced in the TDRA that a cell experiences a randomly varying dose in a microscopic volume [Booz et al., 1983, Rossi and Zaider, 1991], the microscopic specific energy $z^{(c,d)}$ is considered as a random variable with $\langle\langle z^{(c,d)}\rangle_d\rangle_c = D$ being the macroscopic dose experienced by the cell population.
Type II lesions are assumed to be repaired with a constant repairing rate r , or converted to irreparable lesions through a first order process with a constant rate a, or at the second order, representing pairwise combinations, with a constant rate b. The average number of type I and II lesions at $t=0$ is proportional to the amount of specific energy $z^{(c,d)}$ with factors λ and κ. These assumptions formally define the following set of coupled Ordinary Differential Equations (ODE) similar in concept to Equations (2.14)

$$\begin{cases} \dot{x}_I^{(c,d,z)} = a x_{II}^{(c,d,z)} + b\left(x_{II}^{(c,d,z)}\right)^2 , \\ \dot{x}_{II}^{(c,d,z)} = -(a+r) x_{II}^{(c,d,z)} - 2b\left(x_{II}^{(c,d,z)}\right)^2 , \end{cases} \tag{2.16}$$

subject to the initial average number of lesions

$$x_I^{(c,d,z)}(0) = \lambda z^{(c,d)} , \quad x_{II}^{(c,d,z)}(0) = \kappa z^{(c,d)} . \tag{2.17}$$

For charged particles, typically the rate of pairwise combinations between type II lesions is negligible with respect to the first order evaluation of x_{II} for low does [Hawkins, 1996], that is,

$$2b\left(x_{II}^{(c,d,z)}\right)^2 \ll (a+r)x_{II}^{(c,d,z)} \tag{2.18}$$

so that the time-evolution of the average number of type II lesion can be rewritten as

$$\dot{x}_{II}^{(c,d,z)} = -(a+r)x_{II}^{(c,d,z)}\,. \tag{2.19}$$

The solution to Equation (2.19) is

$$x_{II}^{(c,d,z)}(t) = \kappa z^{(c,d)} e^{-(a+r)t}\,. \tag{2.20}$$

Substituting Equation (2.20) into the kinetic Equations (2.16) and integrating $x_{II}^{(c,d,z)}$ with respect to time, it follows that

$$x_I^{(c,d,z)}(t) = \lambda z^{(c,d)} + a\kappa z^{(c,d)}\left(\frac{1-e^{-(a+r)t}}{a+r}\right) + b\kappa^2\left(z^{(c,d)}\right)^2\left(\frac{1-e^{-2(a+r)t}}{a+r}\right)\,. \tag{2.21}$$

An example of the temporal evolution of the lesions in a cell is depicted in Figure 2.2.

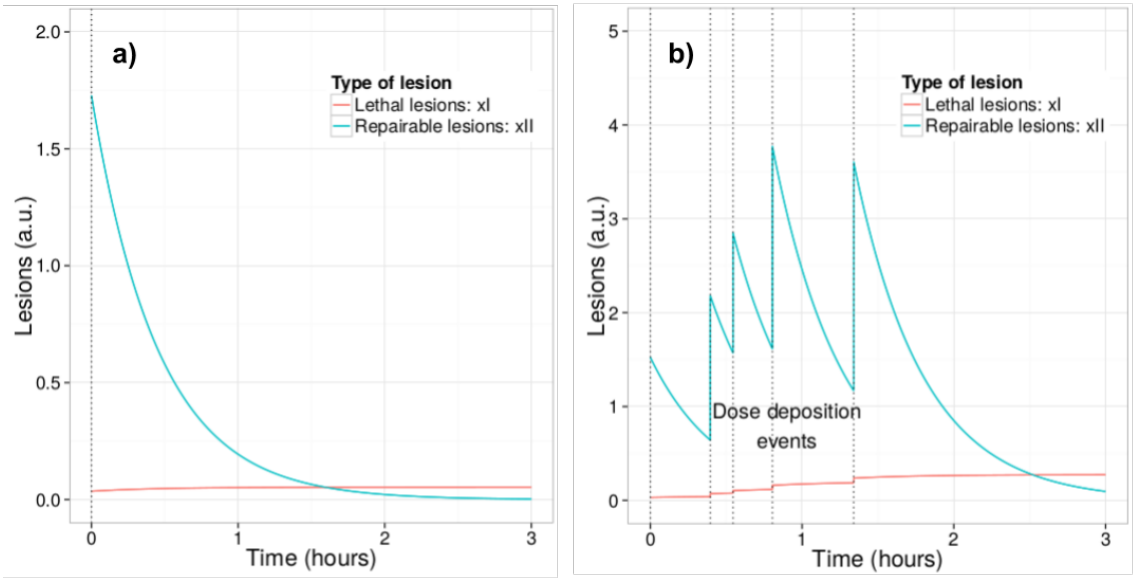

Figure 2.2: Time evolution of x_I and x_{II} damages for a single instantaneous irradiation as described by Equations (2.21) and (2.20), respectively. Left panel: temporal evolution for a single irradiation. The dotted vertical lines represent the energy deposition events in the cell nucleus due to the passage of ionizing particles. Right panel: generalization of the temporal evolution for any time-structured irradiation. The dotted vertical lines represent the energy deposition events in the cell nucleus due to the passage of ionizing particles; figure taken from [Manganaro, 2018].

The exponential decay in Equation (2.21) derives from the assumption; empirical evidences show how it could likely represents an approximation of more complex repair kinetics present in the real cell [Dikomey and Franzke, 1986, Fowler, 1999, Dale

et al., 1999, Carabe-Fernandez et al., 2011]. Postulating that the total number of lesions $x_I(t) + x_{II}(t) \sim N_{\mathrm{DSB}}(t)$ describes the number of DSBs in the DNA, the repair kinetics represented in Equation (2.21) can be verified through H2AX phosphorylation mapping experiments (γ-H2AX) [Schettino et al., 2011, Mariotti et al., 2013]. For high-LET particles such as carbon ions, the presence of a plateau (offset) in the observed $N_{\mathrm{DSB}}(t)$ [Asaithamby et al., 2008, Carabe-Fernandez et al., 2011, Asaithamby et al., 2011] suggests the presence of an irreparable complex clustered damage that can be related directly to the parameter λ of the kinetic equations, and hence to the linear parameter α_0 of the macroscopic cell survival LQ formulation.
In order to connect above explicit solution of Equation (2.21), *i.e.* the average number of type I and II lesions generated by a certain energy deposition $z^{(c,d)}$, to the survival probability, one more fundamental assumption must be made:

8. given a specific energy z, the lethal lesion distribution follows a Poisson distribution.

Under this assumptions, the probability that the domain d survives at time $t \to \infty$ when exposed to the specific energy $z^{(c,d)}$, denoted by $s^{(c,d)}(z^{(c,d)})$, can be computed as the probability that the random outcome of a Poisson random variable is null. Therefore, $s^{(c,d)}$ is given by

$$s^{(c,d)}(z^{(c,d)}) = e^{-\lim_{t\to\infty} x_I^{(c,d,z)}(t)} \,. \tag{2.22}$$

Using Equation (2.21), the average number of lethal lesion given $z^{(c,d)}$ as $t \to \infty$ can be computed as

$$\lim_{t\to\infty} x_I^{(c,d,z)}(t) = \left(\lambda + \frac{a\kappa}{a+r}\right) z^{(c,d)} + \frac{b\kappa^2}{2(a+r)} \left(z^{(c,d)}\right)^2 , \tag{2.23}$$

so that the log-survival for the domain d is given by

$$\log s^{(c,d)}(z^{(c,d)}) = -Az^{(c,d)} - B\left(z^{(c,d)}\right)^2 , \tag{2.24}$$

with A and B defined as

$$A = \left(\lambda + \frac{a\kappa}{a+r}\right) , \quad B = \frac{b\kappa^2}{2(a+r)} \,. \tag{2.25}$$

These constants are independent of the domain d and specific energy $z^{(c,d)}$ deposited in the domain d.

Indicating with $S_n^{(c)}(z_n^{(c)})$ the survival probability of the cell c that has received exactly a specific energy $z_n^{(c)}$ in the nucleus, the log-survival of this quantity, $-\log S_n^{(c)}(z_n) = x_{I,n}^{(c)}(z_n)$, represents the total number pf lesions in the entire cell nucleus, and thus can be evaluated by summing the single–domain log–survival $-\log s^{(c,d)}(z) = x_I^{(c,d)}(z)$ over all the cell domains, or, equivalently, by formally using this quantity average over the domains. Assuming that the probability density functions are the same over all the domains and cells, we can drop the index c and d and use Equations (2.11) to write

$$\begin{aligned} \log S_n(z_n) &:= -x_{I,n}(z_n) \\ &= -N_d \langle x_I(z)\rangle_d = -N_d \langle \log s(z)\rangle_d \\ &= -N_d \left(A\langle z\rangle_d + B\langle z^2\rangle_d\right) \\ &= -N_d A \int_0^\infty z f(z;z_n)\, dz - N_d B \int_0^\infty z^2 f(z;z_n)\, dz \,, \end{aligned} \tag{2.26}$$

where $f(z; z_n)$ is the probability density of z in a domain for a cell with a mean specific energy z_n deposited the nucleus. In particular, as shown in Section 2.1, the following holds

$$z_n = \langle z \rangle_d = \int_0^\infty z f(z; z_n) dz \,. \tag{2.27}$$

Using Equations (2.7) and (2.8) derived in Section 2.1, the log survival in Equation (2.26) can be written as

$$\log S_n(z_n) = -(\alpha_0 + z_D \beta_0) z_n - \beta_0 \left(z_n\right)^2 \,, \tag{2.28}$$

with $\alpha_0 := N_d A$ and $\beta_0 := N_d B$. Also, z_D is the dose average z per event in a domain, obtained by applying Equation (2.8) to the domain.
In Equation (2.26), we have used the notation $f(z; z_n)$ to denote the *multi–event distribution*, rather than $f(z; \lambda_n)$ as done in Section 2.1. This is due to the fact that, since the following relation holds true

$$\lambda_n = \frac{z_n}{z_F} \,, \tag{2.29}$$

we have preferred to specify the dependence upon the *multi–event distribution* average.
In order to obtain the cell survival $S(D)$ for a population of cells irradiated with a macroscopic dose D, the quantity $S_n(z_n)$ defined in Equation 2.26 should be averaged accounting for the distribution of the specific energy z_n over the cell population. In terms of $\log S$ (the logarithm of the cell population survival), and under the assumption that all the cells have the same probability distribution of specific energy z_n, the average of the log-survival probability over the entire cell population can be written as:

$$\begin{aligned} \log S(D) &:= \log \langle S_n(z_n) \rangle_c \\ &= \log \left(\int_0^\infty S_n(z_n) f_n(z_n; D) \, dz_n \right) \,, \end{aligned} \tag{2.30}$$

where $f_n(z_n; D)$ is the probability density of z_n for a macroscopic absorbed dose D delivered to the cell population, *i.e.*

$$D = \langle z_n \rangle_c = \int_0^\infty z_n f_n(z_n; D) dz_n \,. \tag{2.31}$$

Equation (2.30) is fundamentally different from Equation (2.26) since the it considers the average of the logarithm argument, whereas in Equation (2.26) the logarithm of the average has been taken. This indicates that, due to the stochastic nature of z_n, the distribution of lethal lesions $\log S_n(z_n)$ over the cell population is in general non-Poisson, and hence the log of the survival fraction cannot be directly related to the average number of lethal lesions per cell, $\log S(D) \neq -\langle x_{I,n}(z_n) \rangle_c$. However, provided that the variance of z_n is small, a Poission approximation can be assumed and the same procedure used to obtain Equation (2.26) can be applied in this case. With the Poisson approximation, Equation (2.30) can be written as follows,

$$\begin{aligned} \log S(D) &= \log \langle S_n(z_n) \rangle_c \\ &\approx -\langle x_{I,n}(z_n) \rangle_c = \langle \log S_n(z_n) \rangle_c \\ &= \int_0^\infty \log \left(S_n(z_n) \right) f_n(z_n; D) \, dz_n \\ &= -(\alpha_0 + (z_D - z_{n,D}) \beta_0) D - \beta_0 D^2 \,, \end{aligned} \tag{2.32}$$

with $z_{n,D}$ being the dose in the nucleus per event. All the quantities $z_{n,D}$ z_D and $z_n \approx \langle z_n \rangle_c = D$ are assumed to be the same for each cell or domain. All other notations are used as previously introduced. Since the domain size is usually much smaller than the nucleus, it holds that $z_{n,D} \ll z_D$ [Hawkins, 1996]), so that we obtain

$$\log S = -\alpha_{\mathrm{P}} D - \beta D^2 \,, \tag{2.33}$$

with

$$\alpha_{\mathrm{P}} := \alpha_0 + z_D \beta_0 \,, \quad \beta := \beta_0 \,. \tag{2.34}$$

where the subscript P indicates that the relationships hold when the assumption of Poisson distribution of lethal lesions among the irradiated cell population is reasonable, *i.e.* for low-LET radiation.

A further refinement of the MKM kinetic equations involves a fourth type of possible interaction, that happens at time t_r. The following is assumed:

9. after a time $t_r > 0$, all sub–lethal lesions that are not either dead or repaired, automatically transform into lethal lesions.

The mathematical formulation of the main kinetic equations remain the same as in Equations (2.16)–(2.20)–(2.21) in the time interval $t \in [0, t_r)$. As soon t_r is reached, all type II lesions that have not been either repaired or died, will be converted into type I lesions, meaning

$$x_{II}(t) = 0 \,, \quad t > t_r \,. \tag{2.35}$$

The solution for the average number of type I lesion can be now explicitly found for $t > t_r$, adding all type II lesions that persisted for $t > t_r$, that is

$$x_I^{(c,d,z)}(t) = x_I^{(c,d,z)}(t) + x_{II}^{(c,d,z)}(t_r) \,, \tag{2.36}$$

so that we obtain

$$\begin{aligned} \lim_{t\to\infty} x_I^{(c,d,z)}(t) &= \lim_{t\to\infty} x_I^{(c,d,z)}(t) + x_{II}^{(c,d,z)}(t_r) = \\ &= \left(\lambda + \frac{a\kappa}{(a+r)} + \frac{\kappa r}{(a+r)} e^{-(a+r)t_t}\right) z^{(c,d)} + \\ &+ \frac{b\kappa^2}{2(a+r)} \left(1 - e^{-2(a+r)t_r}\right) \left(z^{(c,d)}\right)^2 \,. \end{aligned} \tag{2.37}$$

Taking the average of the number of lesions over all cell domains and cell population, we obtain the generalization of Equation (2.33)

$$\log S = -\alpha D - \beta D^2 \,, \tag{2.38}$$

with

$$\begin{cases} \alpha & := \bar{\alpha}_0 + z_D^{(c,d)} \bar{\beta}_0 \,, \quad \beta := \bar{\beta}_0 \\ \bar{\alpha}_0 & := N_d \left(\lambda + \frac{a\kappa}{(a+r)} + \frac{\kappa r}{(a+r)} e^{-(a+r)t_t}\right) \,, \\ \bar{\beta}_0 & := N_d \frac{b\kappa^2}{2(a+r)} \left(1 - e^{-2(a+r)t_r}\right) \,. \end{cases} \tag{2.39}$$

The non-Poisson correction

In the approximation introduced in Equation (2.32), it is assumed that the variance of the absorbed specific energy z_n among cells is sufficiently small. In this case, the number of lethal events follows the same Poisson distribution in each cells, with average $x_{n,I}$. However, in general, the specific energy received by the cell is a stochastic quantity that varies from cell to cell, with a deviation from the Poisson distribution when considering the whole population of irradiated cells. This deviation is present even if the radiation is perfectly mono-energetic. In this case, the specific energy z_n variance arises from the fluctuation of the number of particles that hit the cells. These variations are particularly relevant when the particle LET is relatively high, because, at a given macroscopic dose D, the average number of high-LET particles interacting with the cell is lower than the number of low-LET particles. To account for the non-Poisson distribution of the lethal events, Hawkins introduced a correction to the MK model [Hawkins, 2003], with a deviation from the linear behaviour of the RBE vs. LET in the high-LET region.

The effect of the lethal lesions non-Poisson behavior is considered by explicitly evaluating the fraction of hit and non-hit cell nuclei. Considering a very low dose high-LET irradiation, $D \ll 1$, the probability for a cell to interact with more than one particle is negligible. In this case, the population of cells can be divided into a fraction Φ of cells that suffer a single particle interaction and a fraction $1-\Phi$ of cell with zero interactions. We call $x_{I,n}(z_{n,D})$ the average number of type I lethal lesions in the fraction Φ of cells whose sensitive nucleus has been hit by a single particle, imparting exactly a specific energy $z_{n,D}$Then, using Equations (2.32)– (2.34), we obtain

$$x_{I,n}(z_{n,D}) = -\log S(z_{n,D}) = (\alpha_0 + z_D\beta_0)z_{n,D} + \beta_0 z_{n,D}^2 \,. \tag{2.40}$$

It is possible to explicitly write the global cell surval fraction (including both hit and non-hit nuclei) as:

$$S(D) = (1-\Phi) + \Phi e^{-x_{I,n}(z_{n,D})} \,. \tag{2.41}$$

This corresponds to consider a probability density function $f_n(z_n; D) = (1-\Phi)\delta(z_n) + \Phi\delta(z_n - z_{n,D})$ in Equation (2.30). The number of lethal lesions per cell averaged over the whole cell population (including both hit and non-hit nuclei) exposed to the macroscopic dose D can be directly evaluated as

$$\langle x_{I,n}(z_n)\rangle_c = \Phi x_{I,n}(z_{n,D}) \,, \tag{2.42}$$

and then Equation 2.41 can be rewritten as

$$\begin{aligned} S(D) &= 1 + \frac{\langle x_{I,n}(z_n)\rangle_c}{x_{I,n}(z_{n,D})}(e^{-x_{I,n}(z_{n,D})} - 1) \\ &= 1 + \left[\frac{e^{-(\alpha_0+z_D\beta_0)z_{n,D}-\beta_0 z_{n,D}^2} - 1}{(\alpha_0 + z_D\beta_0)z_{n,D} + \beta_0 z_{n,D}^2}\right]\left((\alpha_0 + \beta_0 z_D)D + \beta_0 D^2\right) \end{aligned} \tag{2.43}$$

If we expand the log of S around $D = 0$, and drop the terms D^2 or higher powers, then the linear term of $\log S(D)$ can be written as

$$\begin{aligned} -\log S(D)|_{D\to 0} &\approx (\alpha_0 + z_D\beta_0) \times \left(\frac{1 - e^{-(\alpha_0+z_D\beta_0)z_{n,D}-\beta_0 z_{n,D}}}{(\alpha_0 + z_D\beta_0)z_{n,D} + \beta_0 z_{n,D}^2}\right) \times D \\ &\approx \alpha_{\mathrm{P}} \times \left(\frac{1 - e^{-\alpha_P z_{n,D}}}{\alpha_P z_{n,D}}\right) \times D \\ &= \alpha_{\mathrm{NP}} \times D \end{aligned} \tag{2.44}$$

where α_P is the Poisson α coefficient defined in Equation (2.34), while the subscript NP indicates the Non-Poisson corrected α coefficient. Following Hawkins' original formulation [Hawkins, 2003], the quadratic term $z_{n,D}^2 \ll z_{n,D}$ in Equation (2.44) was also neglected.

In Equation (2.44), a subtle approximation is assumed in order to have a Poisson distribution in the lethal events, with only a single defined value of $z_n = z_{n,D}$ when the particle hits the cell. Generally, this is not the case and the specific energy can also vary as a function of the particle impact parameter with respect to cell nucleus. Still, $x_{I,n}(z_{n,D})$ is used as an estimation of the lethal lesions average number in those cells that have registered a single event after exposure to a dose D.

The saturation correction

Kase *et al.* [Kase et al., 2006] introduced a correction factor in the MK model to account for the RBE decrease due to the overkill effect observed in high-LET radiation (see for example Figure 2.3). The correction factor was applied to the dose-averaged saturation-corrected specific energy per event, z_1^*, for mixed radiation field with wide-ranging energy spectra.

In terms of lineal energy, the new value of y_D (and hence z_D) was obtained by applying a correction for each lineal energy component of the lineal energy spectrum. This correction was obtained by using an empirical *saturation parameter* y_0 based on the method introduced by [Powers et al., 1968] and then used in the TDRA [Kellerer and Rossi, 1978]

$$y_D^* = \frac{y_0^2 \int \left[1 - \exp\left(-y^2/y_0^2\right)\right] f(y)\ \mathrm{d}y}{\int y f(y)\ \mathrm{d}y} \tag{2.45}$$

The saturation parameter indicates the lineal energy above which the correction due to the overkill effects became important. The correction to cell survival is then obtained by evaluating the corrected dose-averaged saturation specific energy per event z_D^* in the domain, which can derived from the corrected dose-averaged lineal energy (2.45) using the relationships reported in Equation (2.8):

$$z_D^* = \frac{\bar{l}_d}{m_d} y_D^* = \frac{y_D^*}{\rho \pi r_d^2} \tag{2.46}$$

where ρ, r_d, $\bar{l}_d$ and m_d are the density, radius, mean cord length, and mass of the domain, respectively. The equation for the cell survival (Equation (2.33)) is then modified as follows:

$$-\ln(S) = (\alpha_0 + \beta_0 z_D^*)\, D + \beta_0 D^2 \tag{2.47}$$

Considering the linear term in the macroscopic dose D, the corrected α^* coefficient is hence:

$$\alpha^* = (\alpha_0 + \beta_0 z_D^*) \tag{2.48}$$

No correction is considered for the β coefficient and it is still assumed to be independent of the energy spectrum.

A comparison between the MKM prediction including the saturation correction and the experimental data is shown in Figure 2.3, where the α vs. y_D for HSG cells irradiated with carbon ions is plotted. Comparing Equations (2.44) and (2.48), the saturation correction can be considered as an alternative way to describe the non-Poisson correction defined in section 2.2, since both factors modulate the behaviour of RBE_α in similar ways. In particular, it was shown that for monoenergetic spectra, Equations (2.44) and (2.48) are

functionally equivalent for $y < 500$ keV/μm [Kase et al., 2006]. Thus, by matching these equations in the limit of low LET ($y_D \to 0$), and defining $z_{n,D} = y_D/\rho\pi R_n$, with R_n the radius of the nucleus, it is possible to link the saturation correction parameter y_0 with the other parameters

$$y_0 = \frac{\rho\pi r_d R_n^2}{\sqrt{\beta_0(r_d^2 + R_n^2)}} \tag{2.49}$$

A typical value selected for the saturation parameter is $y_0 = 150$ keV/μm [Kase et al., 2006, Inaniwa et al., 2010].

In the saturation-corrected MKM, other quantities necessary to estimate the RBE are the lineal energy spectra, obtainable with a microdosimeter detector such as the TEPC [Lindborg and Waker, 2017], and the values of α_0, r_d and R_N from which the correction to the y_D^* is calculated. The α_0 and r_d coefficients can be experimentally extrapolated from the the survival curve initial slope (Equation 2.48) for low-LET irradiation (i.e. in the limit of $y_D \to 0$ and $D \to 0$)

$$\begin{aligned} r_d &= \sqrt{\frac{\beta\,(y_D - (y_D)_X)}{\rho\pi\,(\alpha - \alpha_X)}} \\ \lim_{y_D \to 0} \alpha \equiv \alpha_0 &= \alpha_x - \left(\frac{\alpha - \alpha_x}{y_D - (y_D)_X}\right)(y_D)_X \end{aligned} \tag{2.50}$$

where $\rho = 1.0$ g/cm^3, α_X is the LQ parameter of the X-Ray, and $(y_D)_X$ is the dose mean lineal energy for X-ray irradiation.

The saturation-corrected formulation of the MK model is one of the most widely used approaches to estimate the RBE from microdosimetric measurements. Many published studies compare RBE predictions with values measured along the Bragg curves of monoenergetic beams, as well more complex mixed field [Rosenfeld, 2016, Guardiola et al., 2015, Kase et al., 2011, Bianchi et al., 2020].

2.3 Biological weighting functions

2.3.1 RBE weighting functions

The microdosimetric RBE-weighting function approach was initially proposed by Menzel, Pihet, Wambersie *et al.* [Pihet et al., 1990, Menzel et al., 1990, Wambersie, 1994, Wambersie et al., 1990] to compare the beam quality of different neutron [Pihet et al., 1990] and proton [Robertson et al., 1994, Coutrakon et al., 1997] therapeutic installations using measured microdosimetric distributions of lineal energy. Based on previous studies on proton beams [Kliauga et al., 1978, Hall et al., 1978], this approach combines microdosimetric spectra of y with an experimental derived *biological weighting function*, for specific cell line and endpoints, $r(y)$, to evaluate the RBE.

Let $P(y)$ be the cellular response function for a population suffering the fraction of dose $d(y)dy$ corresponding to the lineal energy y. $d(y)$ is the dose probability density of y and can be evaluated as $d(y) = \frac{y}{\bar{y}_F} f(y)$ [Booz et al., 1983]. The linear α parameter, interpreted as the biological effect E per unit dose, is expressed as:

$$\alpha = E/D = \int \frac{P(y)}{y} d(y)\, dy \equiv \int r(y) d(y)\, dy \tag{2.51}$$

where $r(y)$ is defined as the response function.

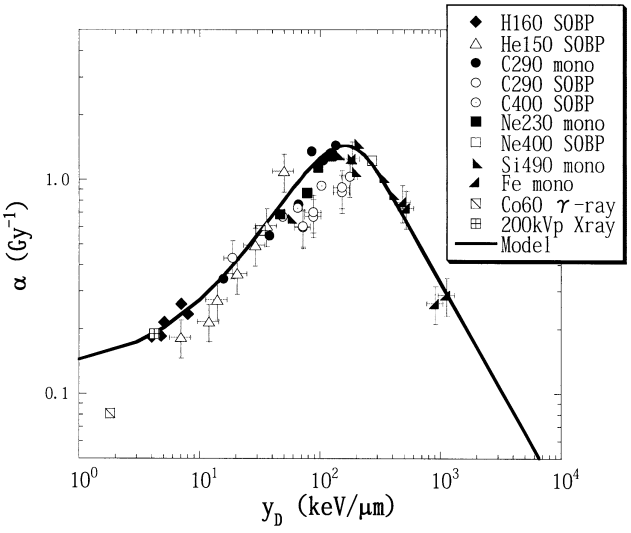

Figure 2.3: Experimental α values of HSG cells as a function of the dose mean lineal energy y_D. The y_D were measured by a TEPC with a simulated diameter of 1.0 μm. The solid line indicates the curve calculated with Equation (2.48) and the following model parameters: $r_d = 0.42$ μm, $R_n = 4.1$ μm, $\alpha_0 = 0.13$ Gy^{-1}, and $\beta_0 = 0.05$ Gy^{-2}. Plot taken from [Kase et al., 2006]. α parameters have been fitted by the linear-quadratic model from the survival curves of HSG cells with a fixed $\beta_0 = 0.05$ Gy^{-2}, as a function of the dose mean lineal energy, y_D.

Usually either $P(y)$ or $r(y)$ are derived from experimental measurements. A formulation for $r(y)$ is [Paganetti et al., 1997, Morstin et al., 1989]

$$r(y) = \sigma_E(1 - \exp\left(-a_1 y - a_2 y^2 - a_3 y^3\right)/y\,, \tag{2.52}$$

where σ_E, a_1, a_2 and a_3 are parameters specific of the radiobiological end points and are independent of the radiation quality. These parameters are determined experimentally by fitting a set of different measurements of α_i or $\mathrm{RBE}_{\alpha,i} = \alpha_i/\alpha_X$, with α_X to be defined as related to irradiation with 250kVp X-ray, using several irradiation modalities with various radiation qualities $i = 1, 2, 3, \ldots, N$.

[Paganetti et al., 1997, Morstin et al., 1989]. The set of relations that have to be fitted to data is hence

$$\mathrm{RBE}_{\alpha,i} = \int r(y) d_i(y)\; dy;\;\; i = 1, \ldots N \tag{2.53}$$

The solution of Equation (2.53) can be obtained with different methods, such as the non-parametric multi-objective optimization [Olko, 1989] or iterative procedures like the Loncol [Loncol et al., 1994] through which an initial guess function $r(y)$ is iteratively updated to best fit Equation (2.53).

2.3.2 Quality factor from microdosimetric distributions

The microdosimetric spectra can also be used to estimate the quality factor Q and equivalent dose H [Zaider and Brenner, 1985].

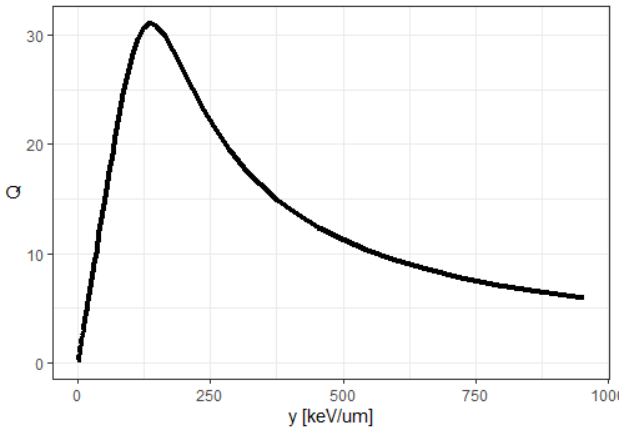

Figure 2.4: Quality factor as a function of the lineal energy in the ICRU40 approach.

The quality factor Q can be calculated from the microdosimetric absorbed dose distribution $d(y)$ according to the *ICRP Report 40* [on Radiological Protection. Committee 4 and on Radiological Protection, 1984]:

$$Q = \int_0^\infty Q(y)d(y)dy \tag{2.54}$$

where $Q(y)$ is the quality factor as a function of the lineal energy, and can be calculated according to the definition of *ICRU 40* or exploiting the Kellerer-Hahn approximation. In particular, Q is a quantity defined in radiation protection to weigh the absorbed dose with regard to its presumed biological effectiveness, from which the equivalent dose H and the health risks related to radiation exposure can be assessed. Radiation with higher Q factors will cause greater damage to tissue. The quantities Q and H were introduced to determine the effect of unwanted low dose-bath on healthy tissues, and for this reasons they are particularly useful for radioprotection applications.
The formal connection of microdosimetry with radiation quality was made in the *ICRU Report 40* (1986), which defined it as a continuous function of the lineal energy y. *ICRU Report 36* [International Commission on Radiation Units and Measurements, 1983] defines the equivalent dose as:

$$H = Q \cdot D \tag{2.55}$$

where D is the physical dose and Q is the average quality factor.

The ICRU40 approach In the ICRU Report40 (1986) the quality factor (Q) is defined as:

$$Q(y) = \frac{5510}{y}(1 - e^{-5\cdot10^{-5}y^2 - 2\cdot10^{-7}y^3}) \tag{2.56}$$

with y the lineal energy expressed in keV/um.

Figure 2.4 shows the quality factor dependence from the lineal energy in this approach, which is the one used in all calculations with both simulated and experimental measurements in this work.

The Kellerer-Hahn approximation Kellerer and Hahn in 1988 proposed an alternative expression for $Q(y)$ with more convenient analytic properties than Equation (2.56):

$$Q_{KH}(y) = 0.3y \left[1 + (y/137)^5\right]^{-0.4} \tag{2.57}$$

with y expressed in keV/μm.

In an effort to retain LET in the definition of Q, Kellerer and Hahn suggested an approximate relation $Q(L)$ by replacing y with the expression $y = \left(\frac{9}{8}\right) L + 0.75$ with both y and L expressed in keV/μm. The resulting microdosimetry-based $Q(L)$ expression can be considered as an alternative to the one suggested by the *ICRP Report 60* (1991).

Part II

Boring-but-useful things

Chapter 3

Facilities of the experimental campaigns

All the experimental characterizations of the TEPC and the characterization of the different proton radiation fields presented in this work have been performed at the Proton-therapy Center (PTC), in Trento, Italy. Helium and oxygen characterizations presented have been instead carried out at the NASA Space Radiation Laboratory, Upton (NY), in US. Finally, neutrons experiments for space radioprotection applications have been performed at the GSI Helmholtz Centre for Heavy Ion Research in Darmstadt, Germany.
These laboratories main characteristics are reported below in order to better understand the framework of the experimental data campaigns.

3.1 Trento Proton Therapy Center

The Trento Protontherapy facility [PTC, 2022], which is part of the Trentino Healthcare Agency (Azienda Provinciale per i Servizi Sanitari — APSS, Italy), started clinical operations in October 2014. A cyclotron (IBA, Proteus 235) serves two medical treatment rooms both equipped with rotating gantries [Tommasino and Durante, 2015].
The facility is also equipped with an experimental area where the beam line is split in two branches, both dedicated to a large spectrum of scientific applications, including medical physics, detector testing, radiation hardness measurements, space research and radiobiology.
Following an institutional agreement with APSS, the beam is available in the experimental room outside clinical hours and all activities are managed and supervised by the Trento Institute for Fundamental Physics and Applications (TIFPA), which is part of the Italian National Institute for Nuclear Physics (INFN).
Proton beam production and transport in the Trento facility are under the responsibility of the IBA company (Ion Beam Applications, Louvain-La-Neuve, Belgium), which produced and installed the related infrastructure. IBA is also responsible for beam operations and has a resident staff in the facility. The cyclotron accelerates the beam up to a maximum energy of 228 MeV. Shortly after the cyclotron exit, a coarse energy selection is carried-out by a rotating degrader of different thicknesses and materials in order to reduce the beam energy down to its minimum value of 70 MeV. This is part of an Energy Selection System (ESS) that allows the fine selection of the desired energy to be transported downstream. Two branches of the main line transport the beam to the gantries, while a third branch connects it to the experimental room. The beam cannot be shared simultaneously among

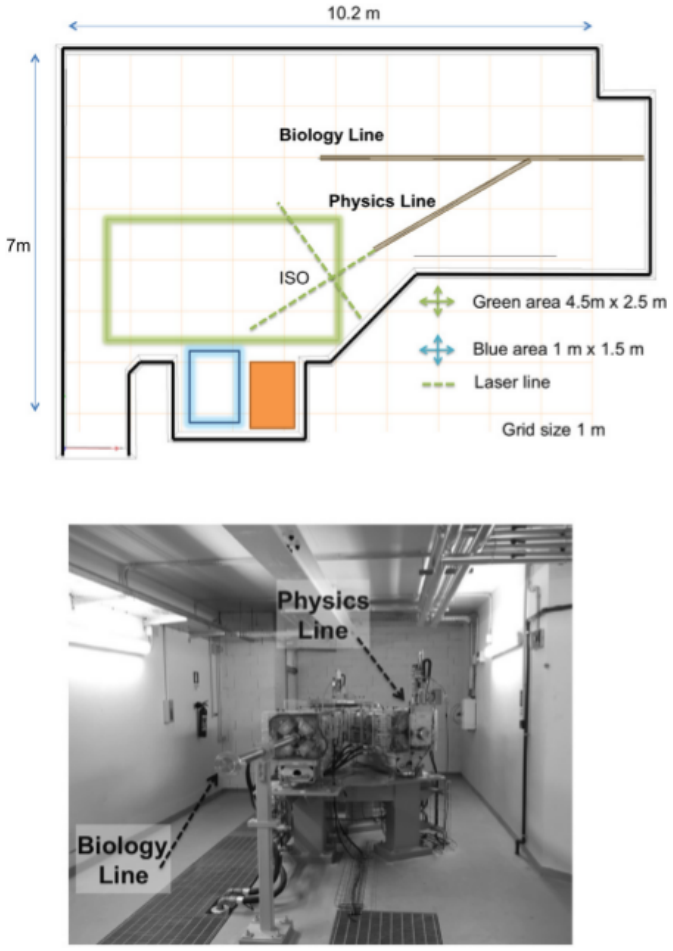

Figure 3.1: A picture (upper panel) and a schematic view (lower panel) of the experimental room at the Trento Protontherapy centre. The Biology (0° branch) and Physics (30° branch) beam lines are indicated. In the map the space available for experimental setup at the Physics line is indicated (green area), together with an additional area for equipment storage inside the cave during irradiation (blue area) [Tommasino and Durante, 2015]

the different rooms and can only be requested alternately in either the gantries or the research area.
Different beam intensities can be requested at the exit of the cyclotron, in a range spanning between 1 and 320 nA.
The experimental area consists of two different spaces: a multifunctional preparation room and the irradiation cave. The former is equipped with a control station for monitoring the activities inside the cave via remote control cameras and alignment lasers. The main beam line is split into two additional sub-branches at 0° and 30° with respect to its initial direction by a dipole magnet 3.1.
This allows the simultaneous setup of two different experiments if necessary but the beam cannot be transported along the two branches at the same time. 0° and 30° lines are referred as the "Biology" and "Physics" beam lines, respectively (see Figure 3.1), since they are intended for different purposes.
A fixed pencil beam is available at the Physics line and it has an energy range between

70 and 228 MeV and a spot size in-air between 6.9 and 2.7 mm sigma at the lowest and highest energies, respectively. Lower energies can be obtained by passive degraders added after the exit window.
Lasers are available for target alignment at 1.25 m from the exit window, which is defined as "Isocenter" in analogy to the treatment rooms.
In addition, the flux dynamic range in air ranges between 10^1-10^{10} particles per second. This guarantees the possibility to perform a great variety of experiments from nuclear and particle physics, detector testing, test of shielding materials and radiation hardness.

3.2 NASA Space Radiation Laboratory

The increasing permanence of humans in Low Earth Orbit (LEO) and, in the future, in extended deep space missions beyond Earth, requires studies for assessing the risk from space radiation exposure and finding effective countermeasures. This has been the motivation for developing ground-based programs where the space radiation environment could be simulated.
The NASA Space Radiation Laboratory (NSRL) is a multidisciplinary accelerator-based center that provides charged particles for space radiation research. Commissioned in 2003, the facility is funded by NASA and managed in collaboration with Brookhaven National Laboratory (BNL), Upton NY, where it is located [La Tessa et al., 2016].
A schematic representation of the transport line from the ion source to NSRL target room is shown in Figure 3.2.
Three sources can provide ions to the Booster and ultimately to NSRL. These are the LINAC (protons only), the Tandem Van de Graaff (certain ions and protons), and the Electron Beam Ion Source (EBIS) (any ion except protons). The particle beams produced by any of the sources mentioned above are injected into the Booster synchrotron where they are further accelerated to reach the final energy and delivered to NSRL via a single line branching off from the Booster (R-Line).
Ion species from protons to gold are presently available, at energies ranging from 50 MeV/u to 2500 MeV for protons and 1500 MeV/u for ^{56}Fe. Lower energies can be achieved with passive absorbers.
Intensities available at NSRL are ion and energy dependent. Typical dose rates for the NASA work do not exceed 1 Gy/min independent of the beam type. However, higher intensities can easily be achieved either by increasing the ion flux directly at the source or by reducing the field size. Lower values, down to approximately 100 particles cm^{-2}, can be achieved for all ion species combining collimation and beam optics.

3.3 GSI Helmholtz Centre for Heavy Ion Research

The GSI Helmholtz Centre for Heavy Ion Research (German: GSI Helmholtzzentrum für Schwerionenforschung) is a federally and state co-funded heavy ion research center in the Wixhausen suburb of Darmstadt, Germany. It was founded in 1969 as the Society for Heavy Ion Research (German: Gesellschaft für Schwerionenforschung), abbreviated GSI, to conduct research on and with heavy-ion accelerators. It is the only major user research center in the State of Hesse. The laboratory performs basic and applied research in physics and related natural science disciplines.

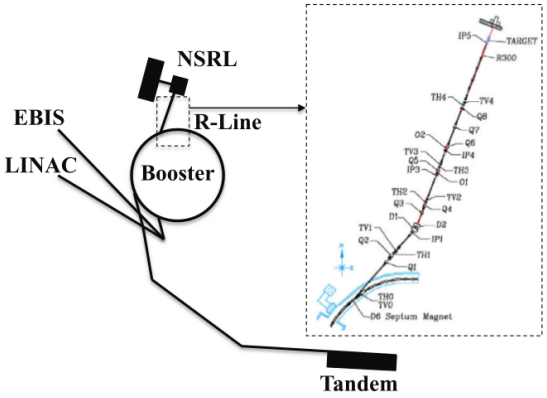

Figure 3.2: Layout of the accelerators for NSRL: sources (EBIS, Tandem and LINAC), synchrotron (Booster) and transport segment between the ring and the target room (R-Line)

In particular, neutron measurements presented in this work have been performed in the high-energy radiation facility "Cave A". Since 1991, this facility has been used by international research groups for radiation experiments in the areas of radiation biology, physical investigations for heavy ion therapy, space exploration and materials research.

Available ion types and energies include many types of particles, from protons to uranium, in the energy range from 50 to 2000 MeV/u. About the particle fluence, it can span from 10^0 to 10^{12} ions/cm^2, allowing a big variety of experiments to be performed.

Chapter 4

TEPC characterization and preliminary tests

4.1 TEPC electronic readout

The ionizations due to the primary beam and its secondaries in the TEPC generate signals. As the signals generated by the TEPC are relatively small, they are sent to a preamplifier, model A422A (CAEN), to be increased. In this way, the detection range can be extended to particles that generate a small signal.
The charge-sensitive preamplifier integrates the charge on a feedback capacitor, and delivers an output voltage with an amplitude proportional to the charge released by the counter and inversely proportional to the feedback capacitance. The preamplifier is a source of electronic noise which should be kept as low as possible so that the minimum measurable energy deposition is low. Therefore, the length of the cable connecting the TEPC and the preamplifier was chosen as small as possible.
After that, the signal is fed to three different shaping amplifiers, in order to achieve a big dynamic range with a good resolution: two model N968 from CAEN with different gains ($\sim$ 100 and $\sim$ 1000, referred to as medium and high gain, respectively) and one model 7243E by Intertechnique with gain $\sim$ 10 (called low gain). Finally, each signal is analysed by Analog to Digital Converters: the pulse coming from medium and high gain is fed to an ADC model 927 (ORTEC), while the low gain signal is fed to an ADC model 926 (ORTEC).
This large dynamic range is essential in order both to collect very high-y events, that can be the more biologically effective ones, and low-y events with a good resolution, being for lighter ions as protons, the most populated region.
Finally, the output signals were digitized by self-triggering *Analog to Digital Converter* (ADC) modules and the raw spectra were visualized and recorded with the data acquisition system MAESTRO (ORTEC) [ORTEC, 2019].
The voltage to the TEPC detector was provided by a high voltage power supply model NHQ 206L (Iseg), and set to a value of +700 V.
The gains of the amplifiers were chosen to have overlap regions in the ADC between high-medium and medium-low spectra, to achieve an "intercalibration" of the three spectra, which could then be merged into one single distribution.
A scheme of the electronic setup used for all the measurements presented in this work is shown in Figure 4.1 and its details are given in Table 4.1.

In this chapter an analysis of the TEPC performances, including the lineal energy

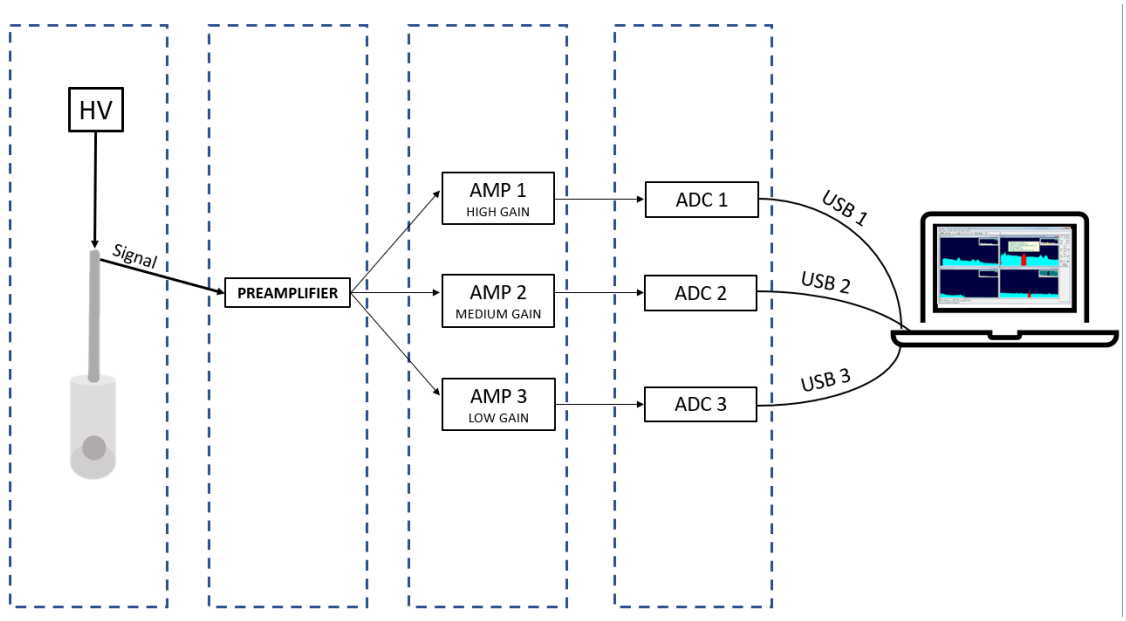

Figure 4.1: Scheme of the electronics setup used in this work. The voltage (HV) is provided to the TEPC and the raw signal is passed to a preamplifier in order to be amplified. Then, three amplifiers with different gains (low, medium and high) collect the amplified signal, and their outputs are sent to the ADCs. Finally, through USB cables, the ADCs signals are recorded via the software MAESTRO (ORTEC).

HV		NHQ 206L	
Preamp		CAEN A4224	
Counter		ORTEC 871	
AMP 1	INTERTECHNIQUE 7243E	High Gain: 1000	input pos, output uni
AMP 2	CAEN N968	Medium Gain: 100	input pos, output uni
AMP 3	CAEN N968	Low Gain: 10	input pos, output uni
ADC 1	ORTEC 927	INPUT 1	4096 channels
ADC 2	ORTEC 927	INPUT 2	4096 channels
ADC 3	ORTEC 926	INPUT 3	1024 channels

Table 4.1: Details of the modules used in the TEPC acquisition chain for this work.

calibration procedure and a study on the pile-up effect, is illustrated. In addition, an overview of the electronic readout system is presented.

4.2 TEPC lineal energy calibration

The detector has to be calibrated in order to assess the relation between the electronic readout output, namely the histograms of the analog to digital converter channels, and the energy deposited in the gas.
The lineal energy calibration of a spherical TEPC in particular can be performed with either an internal built-in alpha-particle source, or an external radiation able to penetrate the detector walls. The radiation field can be used for calibration if a particular marker point of known lineal energy is found in the measured spectrum. This point is often identi-

fied with the proton edge, which corresponds to the maximum energy deposited by protons in the given volume. A detailed description of this methodology applied to a spherical TEPC can be found in [Moro et al., 2015].
In order to achieve the calibration, we decided to follow another procedure, due to the absence of both internal and external particle sources. We exposed the TEPC to proton beams of different energies, and measured the microdosimetric distributions. The maximum values of these peaks have been related to their corresponding microdosimetric distributions peaks obtained exploiting Monte Carlo simulation of the same setup. In the spectra analysis, we also evaluated he proton edge. In this way we obtained a more comprehensive calibration curve, based on a fit of more than two points, as it instead happens in the standard proton edge procedure, thus being more robust. Figure 4.2 shows the linear quadratic fit performed on the data to obtain the calibration curve. Figure 4.3 illustrates the same calibration but with the signal fed to the TEPC instead of ADC channels. The final equation form is:

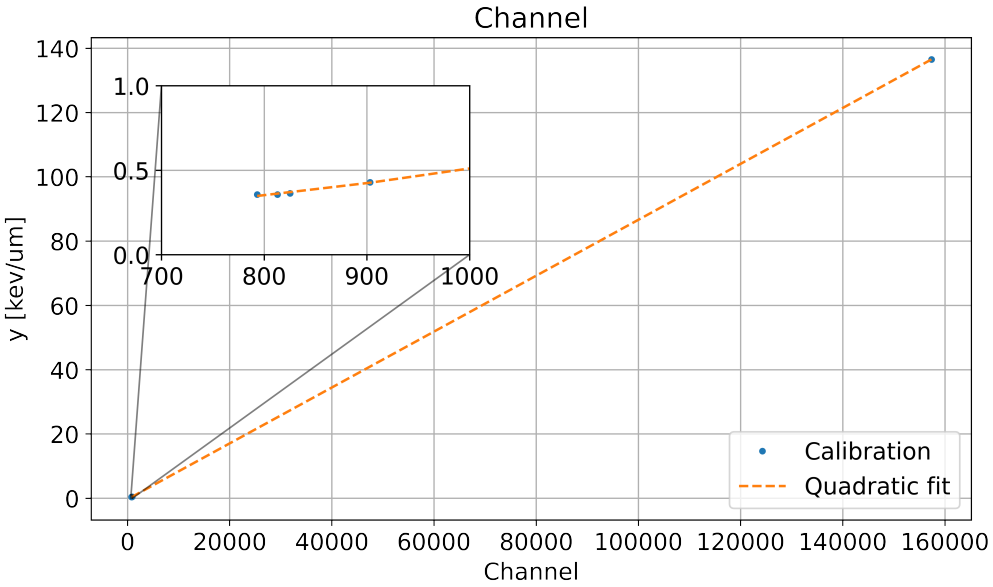

Figure 4.2: Linear quadratic calibration from ADC channel to lineal energy values y for the TEPC. The curve (Quadratic fit) is a fit of the experimental and Monte Carlo peak values obtained from the microdosimetric $f(y)$ distributions (Calibration).

$$y[keV/\mu m] =$$
$$= (9.875 \pm 7) \cdot 10^{-10}\ {ADC_{channel}}^2 + (7.1 \pm 1.1) \cdot 10^{-4} \cdot ADC_{channel} - (2.2 \pm 0.9) \cdot 10^{-1}\,.$$

4.3 Saturation rate and pile-up effect study

In order to study the detector response to the beam intensity, we acquired measurements with monoenergetic proton beams with a variable rate of protons per second. We then acquired different microdosimetric spectra at different energies of the incident proton beam, changing the rate of the particles delivered. Above a certain particle rate, the readout is not able to collect separately every event, and thus two or more events are recorded

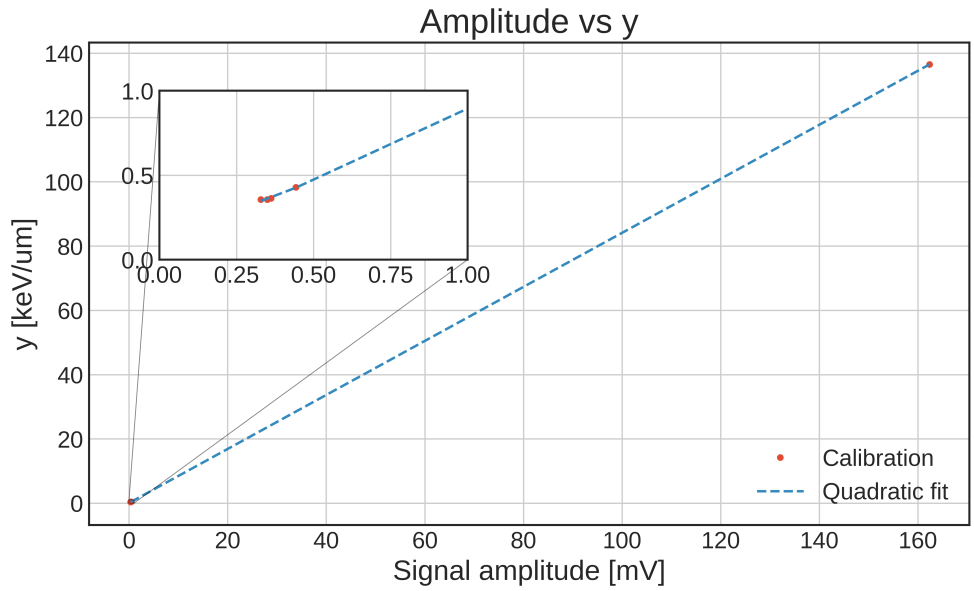

Figure 4.3: Linear quadratic calibration of the signal amplitude [mV] to lineal energy values y for the TEPC. The curve (Quadratic fit) is a fit of the experimental and Monte Carlo peak values obtained from the microdosimetric $f(y)$ distributions (Calibration).

as a single event of higher energy deposition, namely the sum of the energy depositions of the single events: this phenomenon is referred to as pile-up. Together with Monte Carlo simulations, we were able to estimate the probability of pile-up events contained in the acquired spectrum. To achieve this goal, we implemented an *ad-hoc* code that recalculates the simulated Monte Carlo spectrum adding a given pile-up. In this way, we can obtain the microdosimetric distributions and mean values with a given percentage of pile-up probability from the energy depositions event scored by the Monte Carlo simulation. Comparing the experimental data with the simulated spectra associated to different pile-up probabilities, we can estimate if the data were affected by this issue and to quantify it. Figure 4.4 shows an example of a comparison between experimental and simulated microdosimetric yd(y) spectra with pile-up probability of 0 or 33 %, when changing the delivered particle rate from 1286 ± 35 Hz to 48152 ± 219 Hz.

The pile-up study is critical to test the feasibility of performing measurements with a TEPC in a standard treatment irradiation condition, where the beam intensity is $\sim 10^8$. This value is indeed much higher than the value that the detector can sustain ($\sim 10^4$) without giving a distorted distribution due to pile-up.

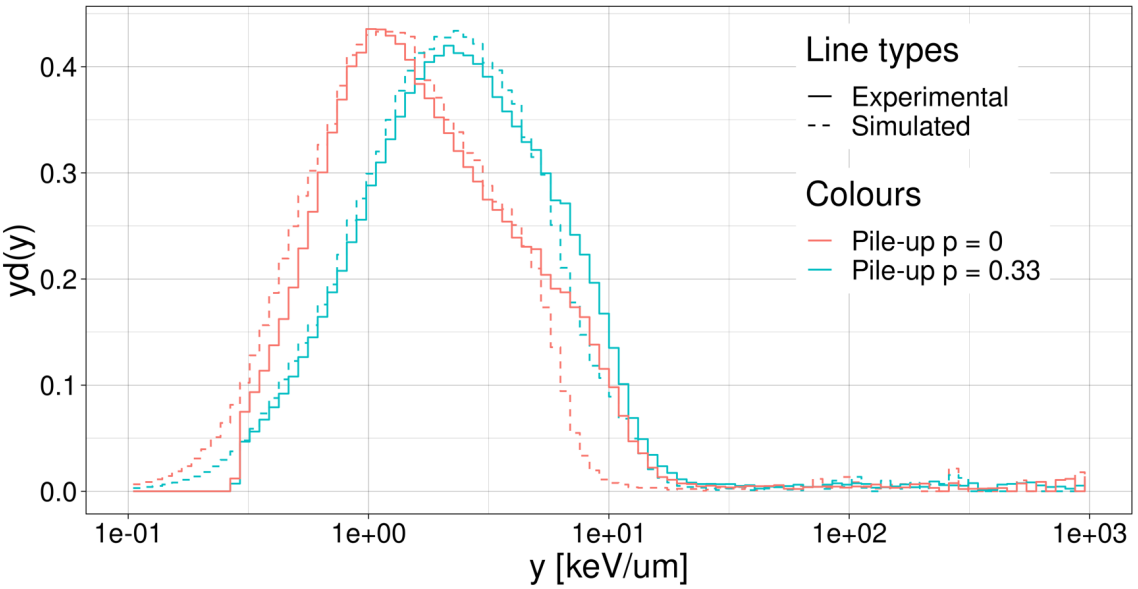

Figure 4.4: Pile-up effect for a 70 MeV proton beam traversing 3 cm of water. The microdosimetric spectra have been acquired both experimentally (solid lines) using two different proton rates affecting the overall distributions.The comparison with Monte Carlo simulations (dashed lines) at different pile-up probabilities is also shown. At the lower rate of particles (pink data), the experimental spectrum agrees with the simulation calculated with $\sim 0\%$ pile-up probability. At the higher rate (blue data), the experimental distribution matches the spectrum simulated with the $\sim 33\%$ pile-up.

Chapter 5

Computational tools

5.1 Data analysis of the microdosimetric spectra from TEPC

An *ad-hoc* code has been written in the R language [Team, 2000] to analyze the spectra measured by the TEPC and acquired with MAESTRO. The code has been also developed to manipulate the microdosimetric data obtained from Monte Carlo softwares, such as Geant4 [Agostinelli et al., 2003] and TOPAS [Perl et al., 2012]. In particular, in these cases, a file containing all the energy deposition events is taken as input.
In Figure 5.1, a scheme of the code features is illustrated for both the experimental and Monte Carlo input types.

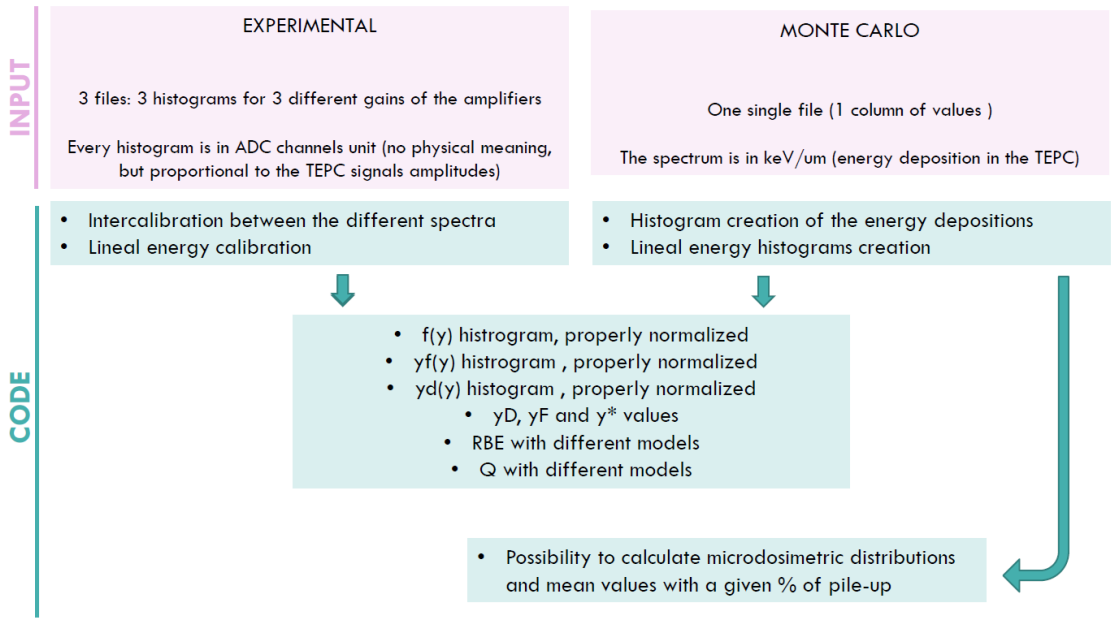

Figure 5.1: Scheme of the R language code for analyzing the microdosimetric data. The input type, depending if it has been acquired experimentally or simulated via Monte Carlo, is imported and the data are processed with different functions implemented in the *ad-hoc* code.

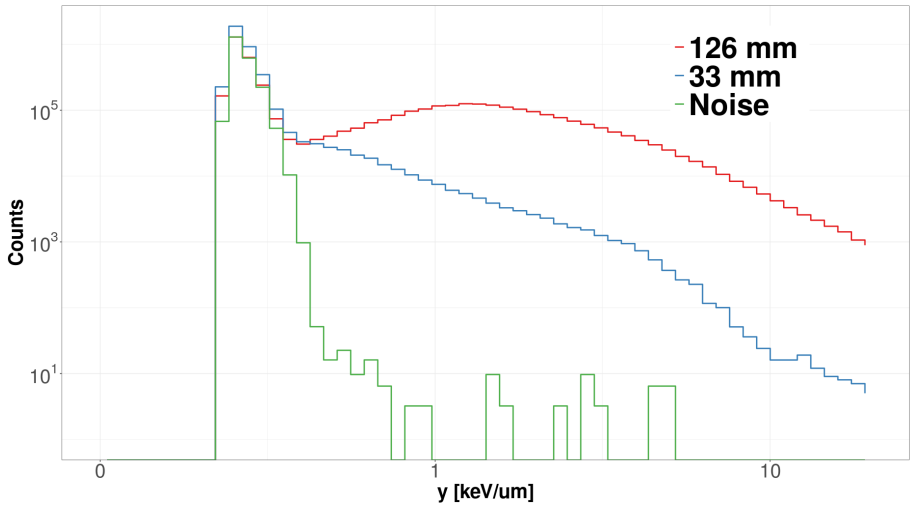

Figure 5.2: Microdosimetric protons raw spectra at different depths in-field with their noise spectra (green) in two different depth in water position along the in–field direction (33 mm and 126 mm, marked in blue and red respectively), yielding to different primary beam energy depositions: in the 33 mm spectrum the primary protons peak below the noise region, while in the 126 mm the peak appears above the lineal energies populated by the noise.

5.1.1 Noise filtering

In addition to all these features, the background noise manipulation in the data analysis has to be taken into account for all the measurements and the first y region of the microdosimetric spectrum, where the noise has its major impact, had to be recovered.
To quantify the background noise, mainly produced by environmental sources, spectra were acquired with the beam switched off and subtracted from the raw microdosimetric spectra. The subtraction method worked efficiently only for the spectra that peaked above the noise region. When the distribution overlapped with the noise, as at the entry channel in-field, the signal-to-noise ratio could not be estimated and the subtraction method potentially produced artifacts in the resulting spectrum. In these cases, we did not perform noise subtraction, and included this uncertainty in the overall error estimate.
The two cases are shown in Figure5.1.1 where the green line represent the noise spectrum, while the blue and red lines are the raw spectra achieved at 33 and 126 mm depth in water, respectively, along the beam direction. The latter demonstrate a clear peak outside the region populated by the noise, while the 33 mm spectrum's peak of primary protons fall in the same region of the noise, making them indistinguishable. In addition to the noise subtraction, the calibrated y spectra have been cut below 0.3 keV/μm, which represent our typical electronic detection limit, where the noise is so high that we could not perform any noise filtering. In order to extend the distributions to the standard initial value of 0.1 keV/μm, the missing y region was reconstructed using an interpolation procedure based on the Stineman algorithm [Stineman, 1980].

5.2 Monte Carlo simulations toolkits: Geant4 and TOPAS

Monte Carlo (MC) method provides numerical solutions to both deterministic and non-deterministic problems. The physical process that models the behaviour of a complex system is described in terms of probability distributions. (Pseudo) random numbers are used to sample the probability distributions and to score the quantity of interest. In particle physics, MC is used to simulate the interaction of radiation with matter. In fact, although the physics of radiation interactions in matter is well understood, in general it is impossible to develop an analytic expression to describe particle transport in a medium. A properly benchmarked Monte Carlo system can thus be used to study the feasibility of a setup or to reproduce an experiment and then be compared with experimental data.
In this work, Geant4 toolkit[Agostinelli et al., 2003] (version 10.7) has been used together with its user-friendly interface TOPAS MC toolkit (versions 2.2.1 and 3.5.1).
Geant4 is a toolkit for the simulation of the passage of particles through matter. Its areas of application include high energy, nuclear and accelerator physics, as well as studies in medical and space science.
This toolkit has been used in particular to investigate the detector performances of the *Hybrid Detector for Microdosimetry*. Several physics lists are available in Geant4 for different energy ranges of interest. For electromagnetic interactions, the high accuracy list *G4EmLivermorePhysics* based on Livermore physics model has been used while hadronic interactions were managed by *QGSP BIC*.
TOPAS (TOol for PArticle Simulation) was instead introduced in 2012 as an extension of the Geant4 Monte Carlo toolkit that facilitates particle therapy simulations [Perl et al., 2012]. Also this software is used for simulating the passage of particles through matter and offers the possibility of custom geometry implementation and material definition. This toolkit has instead been used to simulate both dosimetric (dose and LET_D values) and microdosimetric data of the different radiation field characterizations. In particular, the microdosimetric extension recently presented in [Zhu et al., 2019] ha been exploited for all the microdosimetric assessments, focusing on the spherical TEPC geometry implemented. In addition, it has been used in the cell survival assessment of the *Generalized Stochastic Microdosimetric Model*. Also in this case, being TOPAS a branch of Geant4, the physics list implemented are the same as in Geant4; for what concerns this work, in the protons characterizations we exploited the electromagnetic list ($g4em - standard\_opt4$) and the *Binary Cascade* (BIC) ($g4h - phy\_QGSP\_BIC\_HP$). For heavy ions, instead, we employed also a different hadronic interaction descriptions, namely the *Quantum Molecular Dynamics* (QMD) ($g4ion - QMD$) together with the *Binary Cascade* (BIC); both have been paired with the same electromagnetic list as before, the option4 standard list.

5.3 Machine Learning

Machine Learning (ML) has emerged in last decades as an extremely flexible tool that allows to construct accurate and robust predictive models from data. Recently, the research community developed ML algorithms in several directions, proposing on one side more accurate models to explain the data and on the other side faster optimization algorithms to train the models, [Bishop and Nasrabadi, 2006]. Nowadays ML models are extensively used in many research as well as industrial scenarios. Due to the above mentioned reasons, ML approaches have seen a constant increasing attention as a new paradigm to tackle and solve data-driven and poorly structured problems.

There is no unanimously accepted definition of ML, but broadly speaking ML can be defined as an algorithm that can learn a task from a given dataset without being explicitly taught to. A good definition of machine learning is the one given by Yoshua Bengio[1], one of the worldwide experts and main innovator in the world of ML: "Machine Learning research is part of research on artificial intelligence, seeking to provide knowledge to computers through data, observations and interacting with the world. That acquired knowledge allows computers to correctly generalize to new settings."

Therefore, the development of a ML algorithm must begin with a precise identification of the specific task the ML should learn to solve. Thus, through experience extracted on the dataset the model is trained to solce the task in the most accurate way possible.

The most common ML tasks can be divided into three main categories: classification, regression and clustering, [Bishop and Nasrabadi, 2006]. In this work, only *regression* task has been exploited. The first two classes, namely classification and regression, are typically method of *supervised learning* and differs from the fact that the variable that the ML must learn to predict is discrete in the classification and continuous in the regression case, whereas the last, namely *clustering*, is an *unsupervised learning* task.

Supervised learning consists of ML algorithms optimized using labeled data, meaning that in the learning process the ML model is shown the true outcome that the algorith must learn. More formally, to the ML is given a set of input/output pairs $(x_i, y_i) \in X \times Y$ and the ML must learn the function $f : X \to Y$ that maps the input x into the corresponding output y.

Unsupervised learning consists of tasks with the goal of identifying the underlying patterns from unlabelled data. Classical approaches within unsupervised learning are clustering, density estimation, dimensionality reduction, and probability distribution modeling.

5.3.1 Decision trees: bagging and boosting

Several ML models have been developed over the years; decision trees (DT) are among the most successful. Nonetheless, typically a single DT is not a robust model, so that usually DT are trained either in parallel or in series to boost their performances. An ensemble of DT trained in series is called *boosted* whereas when the ensemble is trained in parallel is called *bagged*. Typical models within the former class are XGBoost, LightGBM and CATBOOST whilst in the latter the most common algorithm is Random Forest. In the current thesis Random Forest will be used. Bagging allows to create from an ensemble of weak learners, single decision trees, a powerful and robust model, as illustrated in Figure 5.3 and it will be treated in mode details in Chapter 10.

[1] Yoshua Bengio is a Canadian computer scientist

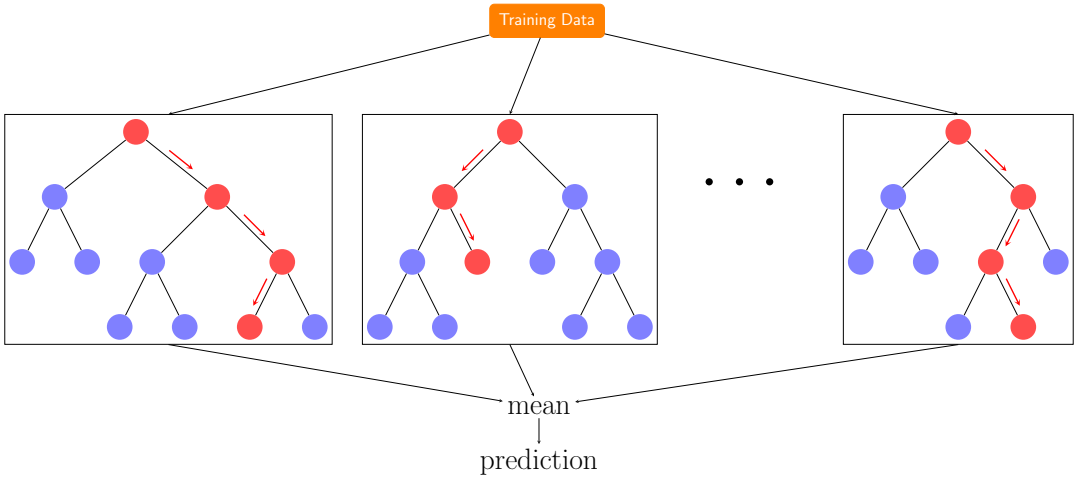

Figure 5.3: Classical structure of a Random Forest algorithm: different decisional trees are trained and the final prediction is derived as the average between the predictions of the single decisions trees.

Part III

Microdosimetric characterization of homogeneous and complex radiation fields

Using a Tissue Equivalent Proportional Counter LET-1/2 1.2.1, we characterized the radiation fields generated by the interaction of different ion-target combinations: i) 154 MeV monoenergetic and 148 MeV passive Spread Out Bragg Peak proton beam in water, iii) 160 MeV/u monoenergetic helium beam in water, iii) 360 MeV/u monoenergetic oxygen beam in water, and iv) 1 GeV/u iron beam in aluminum. We performed measurements (i)-(iii) using a water phantom, and assessed the radiation quality both in- and out-of-field. In the experiment (iv), we focused on the out-of-field region to evaluate the quality of secondary neutrons.

Chapter 6

Monoenergetic and SOBP therapeutic proton beams microdosimetric characterization

The measurements presented in this chapter were carried out at the experimental room of the Trento Protontherapy Center 3.1. All microdosimetry spectra have been measured with the TEPC described in 1.2.1.
Two experimental campaigns are described here, one with a 152 MeV monoenergetic proton beam and one with a 148 MeV proton Spread Out Bragg Peak.
The main goal of both experimental campaigns was to characterize the radiation quality in- and out-of-field. These outcomes were used to study the variability of the RBE along the beam direction, that in the clinical practice is currently considered fixed at 1.1. In addition, for the out-of-field regions, the objective was to combine the radiation quality and physical dose to assess the probability of inducing toxicities in the normal tissue surrounding the target region.

6.1 152 MeV proton beam in- and out-of-field characterization and RBE assessment

The setup included the TEPC placed inside a water phantom (model Blue phantom, IBA) and exposed to a 152 MeV proton pencil beam. The elements between the exit window and the water phantom (including the air gaps) degraded the protons initial energy and enlarge the beam spot size. These parameters were measured at the phantom entrance using the MLIC (MultiLayer Ionization Chamber) and Lynx detectors [Tommasino et al., 2017], respectively. The results indicated that the protons residual energy was 152 MeV (158 mm range in water) and that the FWHM of the beam spot was 13.3 mm on both planes perpendicular to the propagation direction. The beam spot after traversing the water phantom walls was large enough to irradiate the TEPC with a homogeneous field. The number of primary ions impinging on the water phantom was monitored with a 3 mm plastic scintillator and an ion chamber (0.5 mm water equivalent thickness). To avoid pile up effects, the proton rate was set to have an event rate at the TEPC around 5 kHz. A minimum of 10^6 events were acquired for each measurement to reduce statistical fluctuations of rare events [Conte et al., 2019]. A scheme of the experimental setup is shown in Figure 6.1 (top panel).

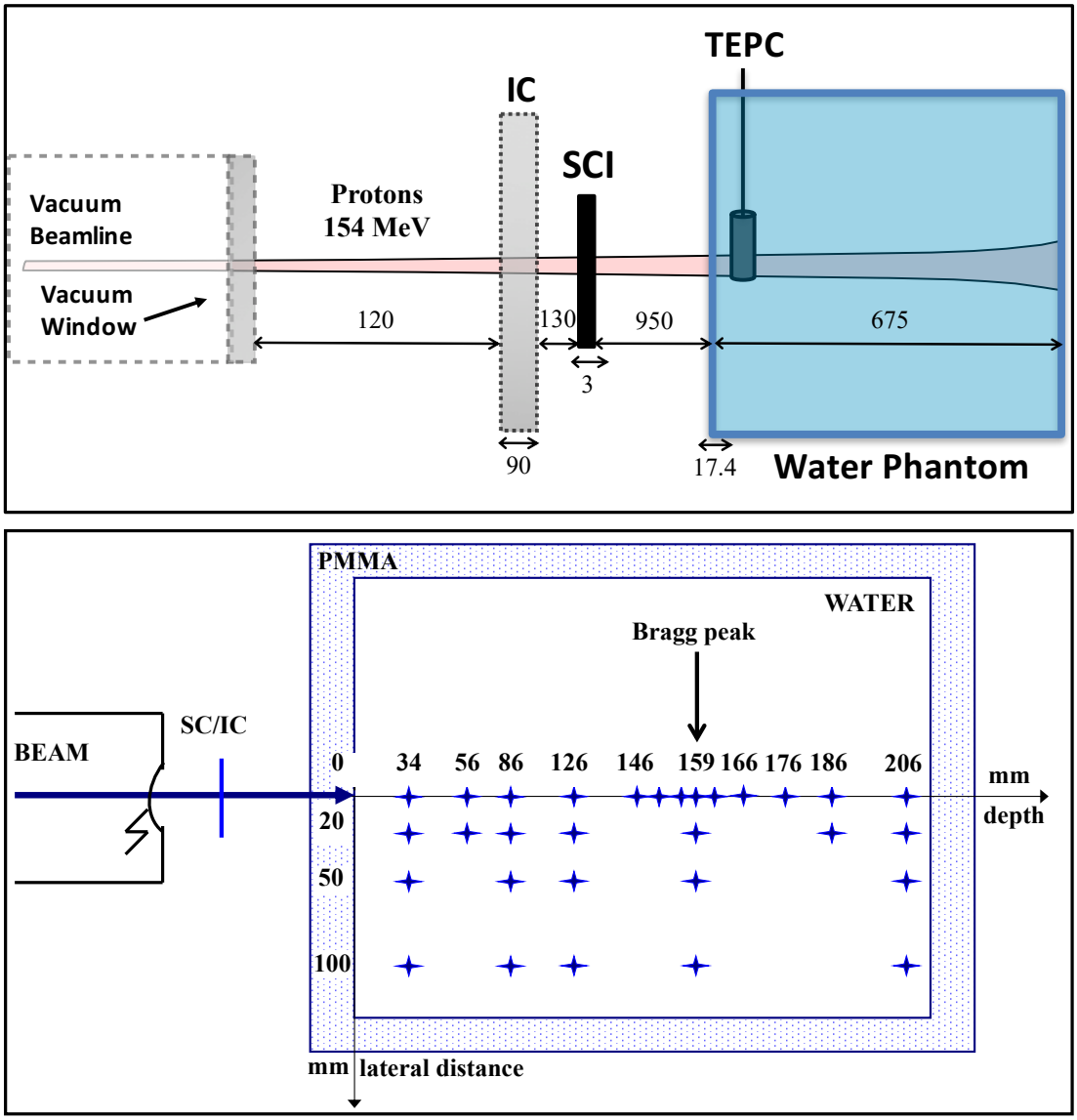

Figure 6.1: Scheme of the experimental setup (top panel) and of the TEPC measurement positions inside the water phantom (bottom panel). All distances and thicknesses are in mm. The depth takes into account also the water phantom wall, whose thickness is expressed in water equivalent. The expected Bragg peak position at 158 mm depth is marked with an arrow.

To select the measurements positions, the TEPC was centered on the beam axis using a laser system. Taking into account the beam width at the entry channel and the detector size, four different regions were identified as of interest for this study:

- *in-beam region*, where the detector was fully and homogeneously irradiated by the primary beam;
- *beam-edge region*, where the detector was moved 20 mm laterally from the beam axis (i.e. just outside the primary irradiation field);
- *close out-of-field region*, where the detector was moved 50 mm laterally from the beam axis;
- *far out-of-field region*, where the detector was moved 100 mm laterally from the beam axis.

For each region, the microdosimetric spectra were acquired at different water depths both upstream and downstream of the Bragg peak. A scheme of the measurement positions inside the water phantom is shown in Figure 6.1 (bottom panel). The depths includes also the PMMA wall (17.4 mm expressed in water equivalent).

6.1.1 Analysis of the microdosimetry spectra

The raw experimental microdosimetric spectra have been converted into lineal energy spectra by applying the linear quadratic calibration described in Section 4.2.

To characterize the radiation field quality, we considered the following quantities:

- frequency-mean lineal energy y_F (Equation (1.1.3))
- dose-mean lineal energy y_D (1.1.3)
- saturation-corrected dose-mean lineal energy $y*$ (2.45)

The quantity y_0 represents the saturation parameter to correct for the overkilling effect of high-LET radiation, theoretically explained in 2.2. Here, a value of 150 keV/μm has been assigned to y_0 [Kase et al., 2013]. From the calibrated spectra, the total absorbed dose D can be obtained using the microdosimetric version of the standard formula for a mixed radiation field [International Commission on Radiation Units and Measurements, 1983]:

$$D = \frac{k}{d^2} y_F \tag{6.1}$$

where $k = 0.204$ for spherical volumes, d is the TEPC simulated diameter and y_F the frequency-mean lineal energy calculated according to Equation((1.1.3)).

By coupling the microdosimetry spectra and dose profiles with biological data, the RBE for cell survival can be assessed. This methodology relies on the Linear Quadratic (LQ) model and on a modified version of the MKM model [Kase et al., 2013]. The first step is to calculate the survival fraction S of cells at each measurement position:

$$S = exp\left[-\left(\alpha_0 + \frac{\beta}{\rho\pi r_d^2} y^*\right) D + \beta D^2\right] \tag{6.2}$$

The dose D and the saturation-corrected dose-mean lineal energy y^* can be estimated from the lineal energy spectra while ρ is the density of tissue assumed to be 1 g cm^{-3}.

The cell radius r_d as well as the α_0 and β parameters of the LQ model depend on the cell line. For this study, the Human Salivary Glands (HSG) tumor cells have been selected (r_d=0.42 μm, α_0=0.13 Gy^{-1} , β=0.05 Gy^{-2} and α/β=2.6 Gy) [Kase et al., 2006]. From the survival fraction, the RBE can be calculated as:

$$RBE_S = \frac{D_{S,R}}{D_S} = \frac{\sqrt{\alpha_R^2 - 4\beta ln(S)} - \alpha_R}{2\beta D_S} \tag{6.3}$$

where D_S and $D_{S,R}$ are the doses required by the radiation of interest and the reference radiation, respectively, to yield the same cell survival fraction S. Following the work of Kase et al. [Kase et al., 2006], X-rays at 200 keV were chosen as reference radiation with an α_R value of 0.19 Gy^{-1} and a β_R value of 0.05 Gy^{-2}.

It is important to remark that in this study the cell survival from which the RBE is estimated is not a fixed value (e.g. 10%), but depends on the dose measured in that position and is calculated using Equation ((6.2)).

6.1.2 Simulations with TOPAS

To obtain a 2D dose map with high spatial resolution as well as to investigate how radiation quality changed if a macroscopic- or microscopic-based approach was adopted, we performed Monte Carlo simulations with the GEANT4-based Simulation Toolkit TOPAS version 3.2.2 [Perl et al., 2012]. The geometry used for all simulations reproduces the experimental setup described above and illustrated in Figure 6.1. TOPAS default physics settings were set to score the absorbed dose and dose-averaged LET$_D$ distribution *in-beam* as a function of water depth in a mesh of 1×1×1 mm^3 voxels. The dose includes the contribution of both primaries and secondaries and was scored delivering 2*10^7 primary protons to the tumor. The LET$_D$ has been calculated with the ProtonLET scorer, that consider the energy deposited by primary and secondary protons (including the associated secondary electrons) but not by heavy fragments [Cortés-Giraldo and Carabe, 2015, Granville and Sawakuchi, 2015]. Taking advantage of the microdosimetric exstension implemented in TOPAS (5.2), we also calculated the y_D in a spherical TEPC, whose geometry and physical properties matched those of the detector used for the experiment.

6.1.3 Data uncertainty

The overall error is the sum of the systematic component stemming from the lineal energy calibration of the microdosimetry spectra, the reproducibility component coming from the setup procedure (e.g. the detector alignment) and the statistical uncertainties.

To estimate the systematic error, we varied the calibration parameters within their error bars and applied them to the raw spectra. We then assessed how different calibrations effected the microdosimetry parameters. Using this method, we obtained an uncertainty of 8% on y_F, of 15% on y_D and of 12% on y^*. Reproducibility was assessed by acquiring the same data point in different moments. The results yielded an error of 3% on y_F, of 16% on y_D and of 8% on y^*. The statistical uncertainties proved to be negligible compared to the other two contributors because of the large number of events collected for each measurement.

The dose errors were obtained propagating the uncertainties on y_F according to the standard theory. The uncertainties on the calculated RBE stem from two sources: physical (dose and y^*) and radiobiological (α and β) parameters. While the radiobiological

quantities depends on the cell type, the physical quantities are related to the irradiation characteristics (e.g. dose, beam quality) and change within the field. In this work, we considered only the statistical fluctuations arising from the dose and y^*, and assumed α and β to be errorless. In this way, we can disentangle the two contributions and investigate RBE variations only caused by the beam quality and absorbed dose.

6.1.4 Results

To understand the features of the spectra it is useful to divide the radiation field into its three main components: primary protons, secondary ions and secondary neutrons. The latter two are generated by nuclear interactions of the proton beam with water, and are also referred to as target fragments. A detailed description of the relevant nuclear physics concepts can be found in [Tommasino and Durante, 2015, Durante and Paganetti, 2016, Rovituso and La Tessa, 2017]. The number of primary protons that undergo fragmentation, and thus produce at least a secondary particle, has been estimated to be around 1% per cm of water traversed [Tommasino and Durante, 2015]. The primary beam is fairly monoenergetic at the entrance channel, and thus has a specific lineal energy, but its spectrum becomes wider when it approaches the Bragg peak region due to the energy straggling. Instead, target fragments can be neutrons and charged particles of different species and energies, producing a broad y spectrum at any depth. The majority of secondary ions are generated with lower energy than the primary protons, and because they have the same or higher charge, their stopping point will be upstream or around the Bragg peak depth. Because of their physical properties, neutrons are the only fragment type that do not follow this behavior and can reach sites very far from the main beam. For this reason, they become an important source of dose both downstream of the Bragg peak and at increasing lateral distance from the beam axis. Results presented by [Dewey et al., 2017] provide a summary of all these considerations, showing the contribution to the dose of each fragments species produced by a primary proton beam in water, when a y-based weight is applied.

All data presented here can be interpreted on this basis.

Lineal energy spectra

The microdosimetric lineal energy spectra, $yd(y)$, in their standard semi-log representation, are presented *in-beam* and off-beam at several depths in water. All spectra range from 0.3 to 700 keV/μm and are normalized so that the integral of the $yd(y)$ distribution is equal to 1.

At the entrance channel, the spectrum measured *in-beam* has a relatively defined maximum around 0.5 keV/μm, with a shoulder on the right-end side which extends up to 10 keV/μm. Very few events can be observed at higher lineal energies. At the *beam-edge*, we observed that the peak and the shoulder merge together to form a broad flat region ranging from 0.3 to 10 keV/μm. Moving out-of-field, the distributions become even flatter and wider, eventually stretching across the whole lineal energy range. The low-y peak corresponds to the primary beam, which can reach also the out-of-field region because of lateral scattering in water while the remaining contribution is given by secondary electrons and fragments.

At increasing water depth, the peak of primary protons moves to higher y and gets wider, eventually incorporating the right shoulder of the fragments. As protons slow down, their lineal energy increases and reaches the maximum in the Bragg peak region, where

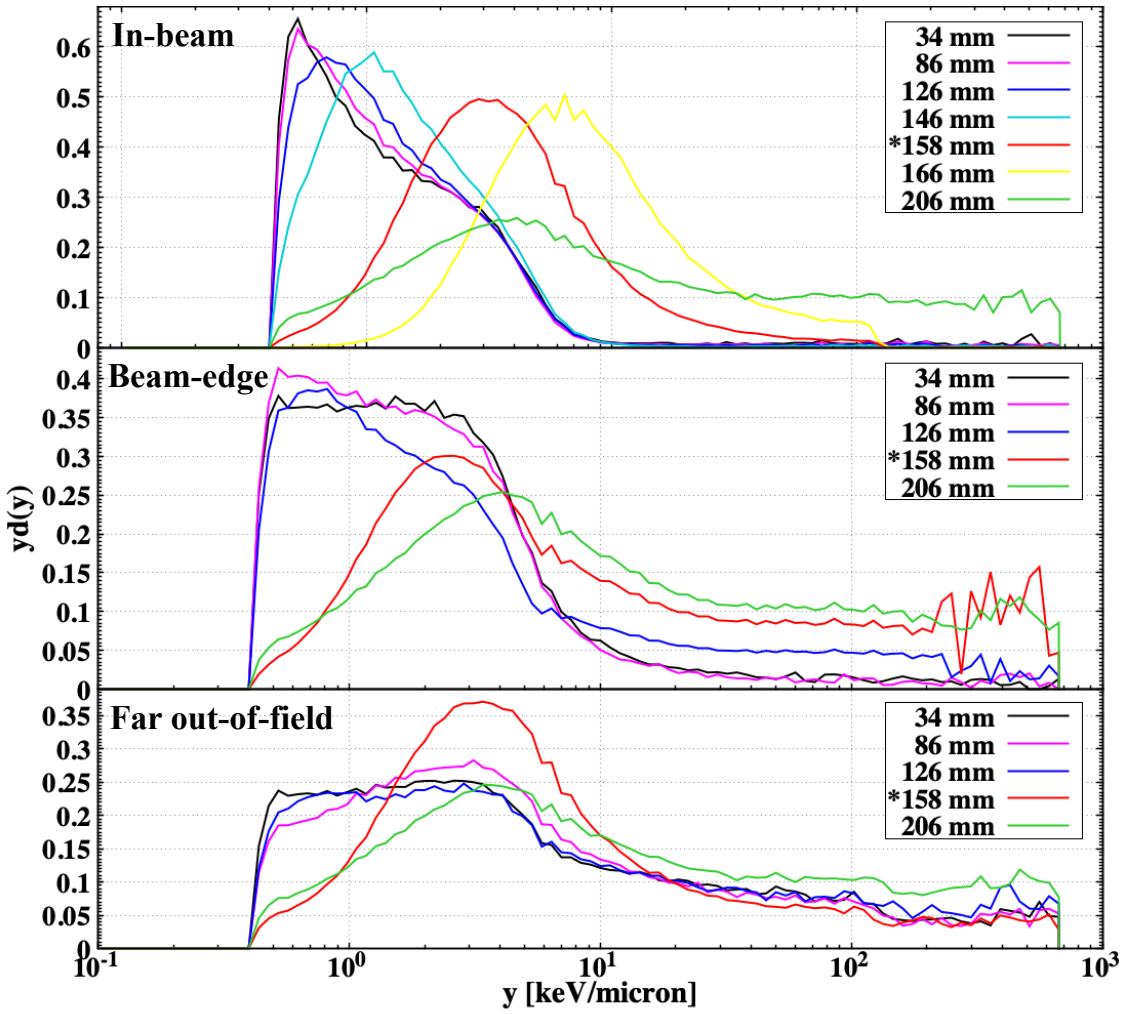

Figure 6.2: In- and off-beam microdosimetric lineal energy spectra $yd(y)$ measured at several depths. The off-beam regions are defined as *beam-edge* (20 mm from the beam axis) and far out-of-field (100 mm from the beam axis). The asterisk indicates the spectra measured at the Bragg peak depth. All spectra are normalized so that the integral of the $yd(y)$ distribution is equal to 1.

the average y is around 5 keV/μm. The distribution at the *beam-edge* show a remarkable change in shape as the depth approaches the Bragg peak. The number of primary ions reaching this area grows with water depth and eventually the spectra resemble those measured *in-beam* with the addition of a longer tail at high-y.

All far out-of-field spectra measured upstream of the Bragg peak appear very similar, indicating that the radiation quality does not change with depth as at the field-edge. At the Bragg peak (158 mm), the distribution resembles the spectrum *in-beam*, indicating that a large portion of particles that reaches this region are stopping around this depth.

In the tail (or distal) region, all primary protons as well as most charged fragments have ranged out. Both in- and off-beam spectra loose almost completely the low-y tail (below 1 keV/μm) and extend to very high-y (above 100 keV/μm).

Independently of the lateral distance, the high-y channels of all spectra become more populated at increasing depth. The reason for this trend is that the probability for the primary protons to either undergo multiple scattering or produce a fragment increases with increasing depth. Secondary particles can be produced at large angles with respect to the primary beam direction and can also suffer lateral scattering being deflected even further from the beam axis. In addition, they can yield further generations of fragments, whose production point can be already out-of-field. All the above potentially leads to high-y events, thus populating the high-y right tail of the distribution.

Radiation quality and dose

The frequency-mean lineal energy y_F (Equation (1.1.3)), dose-mean lineal energy y_D (Equation (1.1.3)) and saturation-corrected dose-mean lineal energy y^* (Equation (2.45)) obtained from the $yf(y)$ and $yd(y)$ spectra are plotted in Figure 6.3.

Independently of the lateral position, the y_F, y_D and y^* values appear relatively flat in the entrance channel and plateau. Measurements *in-beam* and at the *beam-edge* show the same trend, with a rapid growth as the depth approaches the Bragg peak, while off-beam the radiation quality does not show large variations with depth, as already observed in the $yd(y)$ spectra. At the *beam-edge*, the highest values are reached for all quantities at the Bragg peak depth and then stay constant in the tail. In-beam, the y_F has a maximum at the Bragg peak and then decreases in the distal region, while y_D and y^* keep growing up to 2 cm downstream of the peak where they plateau. With the exception of the Bragg peak region *in-beam*, the values of y_D are always larger than the y^*, ranging from a factor 1.5 at the entrance channel and growing up to a factor 3 in the tail region. This result indicates a non negligible contribution to the spectra from particles with $y > 150$ keV/μm.

The microdosimetric quantities y_F, y_D and y^* describe the quality of the radiation field and thus they are affected if its composition changes both in terms of particle type and energy. In-beam and at the *beam-edge*, the field is a mix of primary and secondary radiation, whose composition changes significantly with depth as more particle undergo fragmentation, slow down or scatter. At increasing lateral depth, most charged particles range out and eventually only neutrons remain.

To investigate the influence of the scoring methodology on the radiation quality, we simulated the y_D values measured *in-beam* with TOPAS and compared them with the macroscopic $\overline{LET}_D$. The results are illustrated in Figure 6.4, where the macroscopic physical dose calculated *in-beam* is also plotted.

TOPAS predictions of y_D agree with the experimental data within the error bars while the $\overline{LET}_D$ is approximately a factor 4 lower up to the Bragg peak, after which becomes consistent with the microdosimetric data. It should be remarked here that the $\overline{LET}_D$

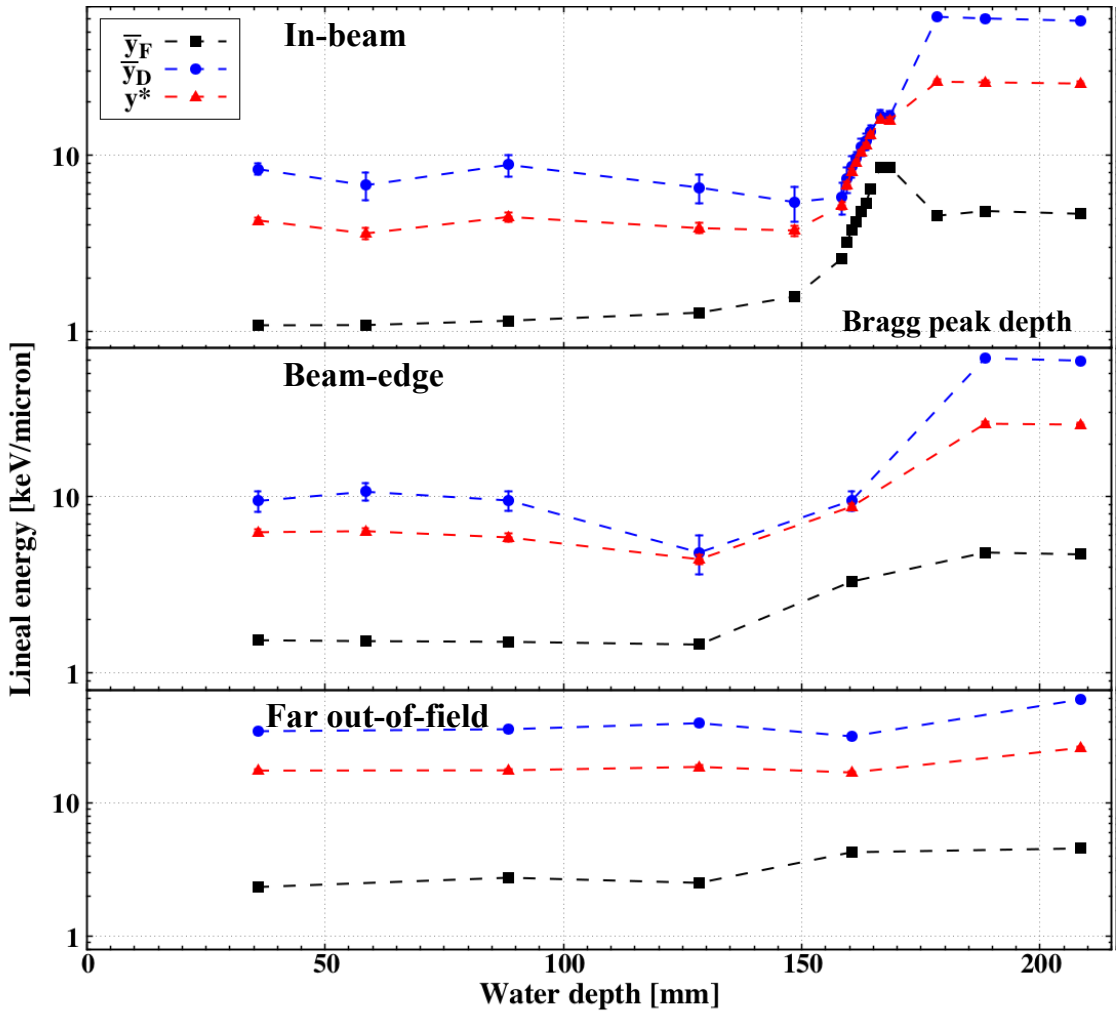

Figure 6.3: Frequency-mean lineal energy y_F (Equation ((1.1.3))), dose-mean lineal energy y_D (Equation ((1.1.3))) and saturation-corrected dose-mean lineal energy y^* (Equation ((2.45))) plotted as a function of the depth in water. The off-beam regions are defined as *beam-edge* (20 mm from the beam axis) and far out-of-field (100 mm from the beam axis). The dotted line marks the Bragg peak position (158 mm).

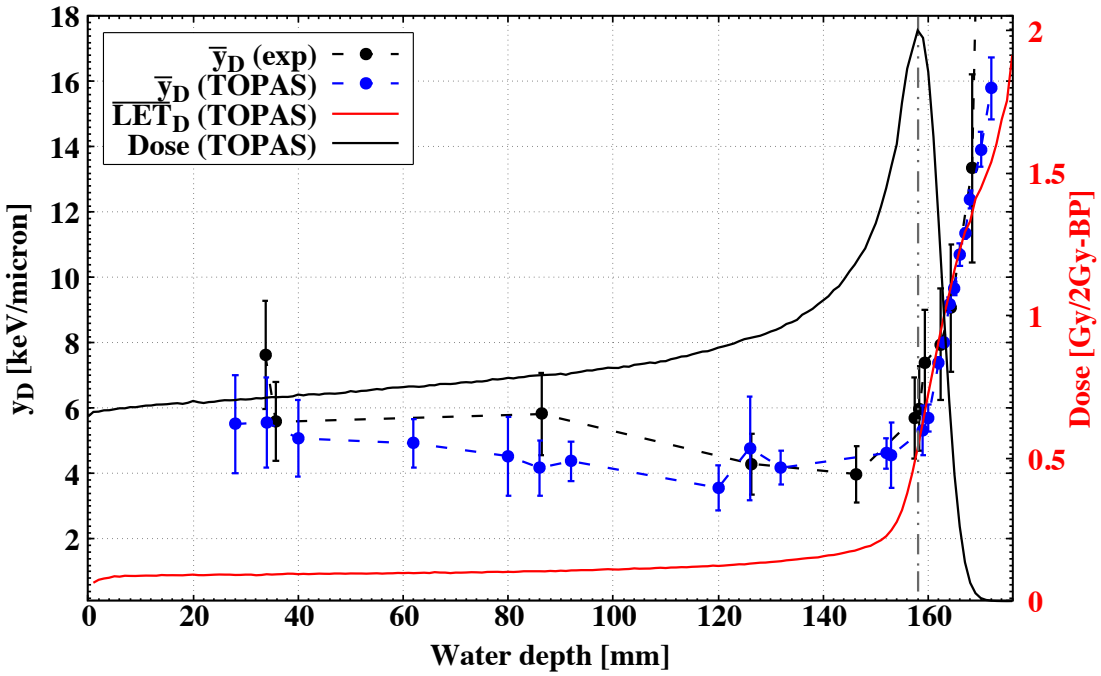

Figure 6.4: Comparison of experimental (black dots) and simulated (blue dots) y_D *in-beam* with the macroscopic $\overline{LET}_D$ (red curve). The depth-dose curve scored in a 1×1×1 mm^3 voxels is also plotted. All simulations were performed with TOPAS (Perl et al. 2012).

calculated with TOPAS includes only primary and secondary protons as well as secondary electrons but not the contribution from heavy ions.

Using Equation ((6.1)), the dose can be estimated from the microdosimetry spectra. To allow for a direct comparison, all measurements have been scaled to deliver a dose of 2 Gy at the Bragg peak position and thus they will be reported in Gy/2Gy-BP. The results are plotted in Figure 6.5 and summarized in Table 6.1.

To investigate the dose gradient inside the water phantom as well as to disentangle the contribution of primary protons from secondary particles, we generated 2D color maps with TOPAS. The outcomes are presented in Figure 6.6. In this case, the dose was scored with a macroscopic approach using 1×1×1 mm ^{3} volume due to to calculation time constrains. The values are scaled to deliver 2 Gy to the Bragg peak position.

The dose measured *in-beam* slowly rises from the entrance channel up to the Bragg peak at 158 mm depth, in agreement with the range of 152 MeV protons in water measured independently with the MLIC detector. After the peak, the dose drops to around 5 10^{-4} Gy/2Gy-BP within less than 20 mm. TOPAS indicates that the overall dose *in-beam* is dominated by primary protons (Figure 6.6, panel a and c), as expected. The contribution from secondary particles is relatively flat along the whole curve and around 5 10^{-2} Gy/2Gy-BP. Then, it drops sharply several orders of magnitudes within few mm from the Bragg peak position, indicating that the range of most fragments do not exceed that of the primary protons.

We observed a relatively large build-up at the *beam-edge*, proving that a non negligible amount of dose is deposited in this area. The maximum value of (1.01±0.12) 10^{-1} Gy/2Gy-BP is observed at the Bragg peak depth. The 2D maps show a steep gradient in this region, with a dose drop of almost 2 orders of magnitudes within less than 30 mm from the field

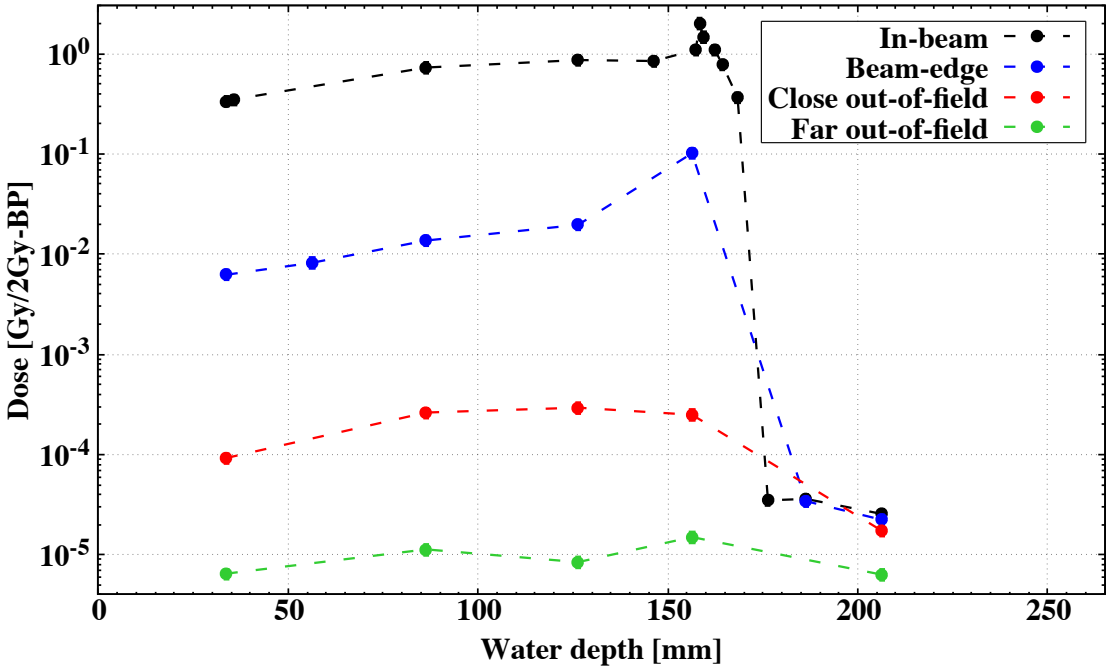

Figure 6.5: Dose profiles in- and off-beam when 2 Gy are delivered to the Bragg peak position. The off-beam regions are defined as *beam-edge* (20 mm from the beam axis), close out-of-field (50 mm from the beam axis) and far out-of-field (100 mm from the beam axis).

center. In agreement with the measurements, TOPAS predicts that the highest dose at the *beam-edge* is delivered around the Bragg peak depth. Secondary particles deposit most of the dose, while the contribution from primary protons scattered in this region increases with increasing depth and reaches the maximum around the Bragg peak.

At larger lateral distances, the experimental curves are rather flat and present a drop just downstream of the Bragg peak (more pronounced in the close-out-of-field). The maximum dose in the close-out-of-field region is one order of magnitude higher than in the far-out-of-field area (10^{-4} versus 10^{-5} Gy/2Gy-BP). TOPAS predictions indicate that the dose is around 10^{-4} Gy/2Gy-BP at 50 mm off-beam and decreases to 10^{-5} Gy/2Gy-BP at 100 mm off-beam, with secondary particles being the only contributors to the dose in these areas.

It is interesting to notice that the dose measured at 250 mm depth drops to around 5 10^{-5} Gy/2Gy-BP independently of the region. At the same depth, the simulated dose is of the same order of magnitude and is entirely delivered by the fragments and neutrons.

In summary, the measured and simulated dose profiles suggest that there is a non negligible number of primary ions that deviate enough from their initial path to deposits some dose out-of-field, especially at the *beam-edge*. However, as the distance from the field increases, the secondary particles become the major and eventually the only source of dose. Exploiting the results of Figure 6.6, we can give an estimate of the dose deposited off-beam when a dose comparable to a full treatment is delivered at the Bragg peak. Assuming that a dose of 60 Gy is delivered at the peak, the *beam-edge* receives up to 6 Gy, while the close and far out-of-field regions receive on the order of 10^{-3} Gy and 10^{-4} Gy, respectively.

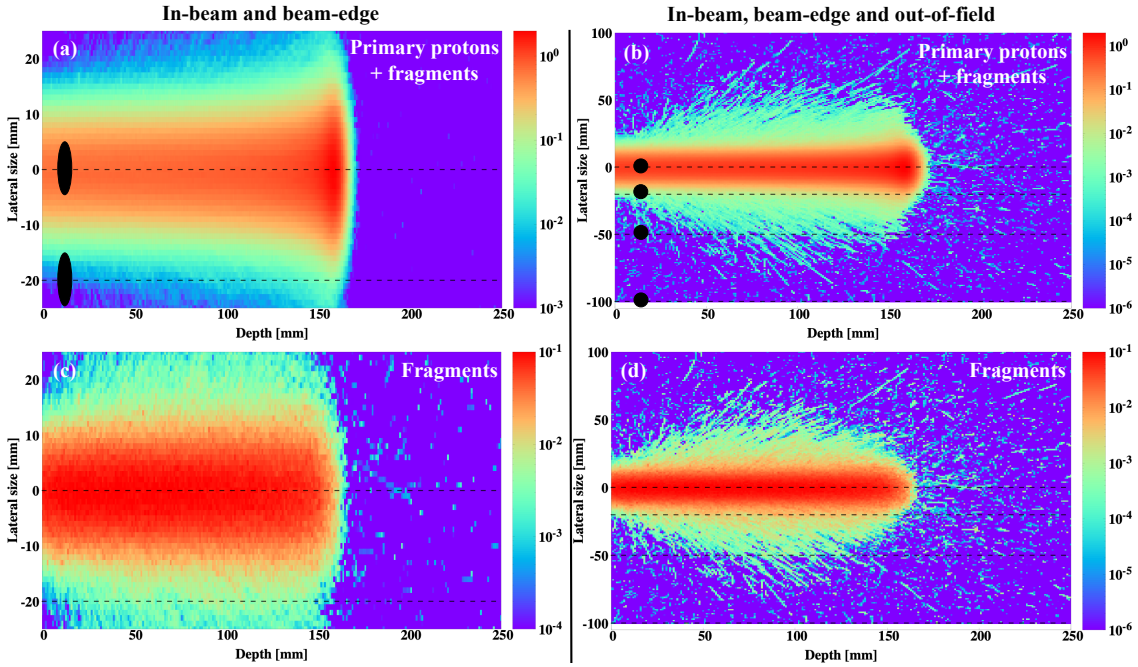

Figure 6.6: 2D color plots of the dose deposition inside the water phantom simulated with TOPAS (Perl et al. 2012). The values have been rescaled to deliver 2 Gy at the Bragg peak position. The contribution from primary protons and secondary fragments is shown in panel (a) inside the entire water phantom and panel (b) as a zoom around the beam axis. The contribution only from secondaries is displayed in panel (c) inside the entire water phantom and panel (d) as a zoom around the beam axis. In all plots, the off-beam regions defined as *beam-edge* (20 mm from the beam axis), close out-of-field (50 mm from the beam axis) and far out-of-field (100 mm from the beam axis) are marked with dashed lines. The TEPC size is marked with black ellipses or circles, depending on the axes scale.

Table 6.1: Dose (Equation ((6.1))), RBE (Equation ((6.3))) and RBE*Dose measured *in-beam* and out-of-field. For each position, the lateral distance from the beam axis and the water depth are reported in parentheses. All doses are measured when 2 Gy are delivered to the Bragg peak and thus are reported as Gy/2Gy-BP.

Position (mm)		Dose (Gy/2Gy-BP)	RBE	RBE*Dose (Gy/2Gy-BP)
In-beam (0)	Entrance (34)	$(3.5\pm0.4)10^{-1}$	1.0±0.2	$(3.2\pm0.8)10^{-1}$
	Plateau (86)	$(7.3\pm0.9)10^{-1}$	0.94±0.19	$(6.8\pm1.6)10^{-1}$
	Rise (126)	$(8.6\pm0.10)10^{-1}$	0.92±0.19	$(7.9\pm1.9)10^{-1}$
	Bragg peak (158)	2.0±0.2	1.05±0.15	2.1±0.4
	Tail (176)	$(3.5\pm0.4)10^{-5}$	2.58±0.15	$(9.0\pm1.2)10^{-5}$
	Tail (206)	$(2.5\pm0.3)10^{-5}$	2.51±0.15	$(6.4\pm0.9)10^{-5}$
Beam-edge (20)	Entrance (34)	$(6.3\pm0.8)10^{-3}$	1.1±0.2	$(6.7\pm1.6)10^{-3}$
	Rise(126)	$(2.0\pm0.2)10^{-2}$	1.0±0.2	$(1.2\pm0.3)10^{-2}$
	Bragg peak (158)	$(1.01\pm0.12)10^{-1}$	2.3±0.16	$(2.3\pm0.3)10^{-1}$
	Tail (206)	$(2.2\pm0.3)10^{-5}$	2.56±0.15	$(5.7\pm0.8)10^{-5}$
Close-out (50)	Entrance (36)	$(9.3\pm1.1)10^{-5}$	1.3±0.2	$(1.2\pm0.2)10^{-4}$
	Rise(126)	$(2.9\pm0.4)10^{-4}$	1.2±0.2	$(3.5\pm0.8)10^{-4}$
	Bragg peak (158)	$(2.5\pm0.3)10^{-4}$	1.3±0.2	$(3.4\pm0.7)10^{-4}$
	Tail (206)	$(1.7\pm0.2)10^{-5}$	2.47±0.16	$(4.2\pm0.6)10^{-5}$
Far-out (100)	Entrance (36)	$(5.6\pm0.7)10^{-6}$	1.90±0.18	$(1.06\pm0.16)10^{-5}$
	Rise(126)	$(8.5\pm1.0)10^{-6}$	2.0±0.18	$(1.7\pm0.3)10^{-5}$
	Bragg peak (158)	$(1.50\pm0.18)10^{-5}$	1.79±0.19	$(2.7\pm0.4)10^{-5}$
	Tail (206)	$(5.5\pm0.6)10^{-6}$	2.54±0.15	$(1.4\pm1.9)10^{-5}$

RBE and RBE-weighted dose

The microdosimetry spectra and dose profiles have been combined to obtain an estimate of the RBE for cell death. It is important to underline that the RBE presented in this work was not calculated for a fixed survival fraction (e.g. 10%) because such a value would never be observed out-of-field. Instead, we considered the dose received at each position and calculate the RBE according to Equations (6.2) and (6.3). The results are reported in Table 6.1 and plotted in Figure 6.7 (top panel). A full line at 1.1 marks the reference value for protons while a dotted line indicates the Bragg peak depth (158 mm).

The datasets *in-beam*, at the *beam-edge* and close out-of-field show the same trend, with an extended plateau followed by a rise near the Bragg peak depth while in the far-out-field region the calculated RBE increases in the tail. In-field, the values measured upstream of the peak are either consistent with or lower than the reference value of 1.1. The RBE found at the Bragg peak is 1.05±0.15 and increases up to 2.58±0.15 at a depth of 176 mm. The values estimated at the field-edge are constant and consistent with 1.1 in the entry channel, begin to grow at a depth of 126 mm and then plateau again around 2.5 in the distal region. The RBE at the Bragg peak depth is 2.27±0.17, i.e. 2 times higher than the *in-beam*.

Out-of-field, the calculated RBE is always significantly higher than 1.1 and increases with increasing lateral distance. Downstream of the Bragg peak, all regions are characterized by a rather constant and similar RBE around 2.5. Among all data points, the highest

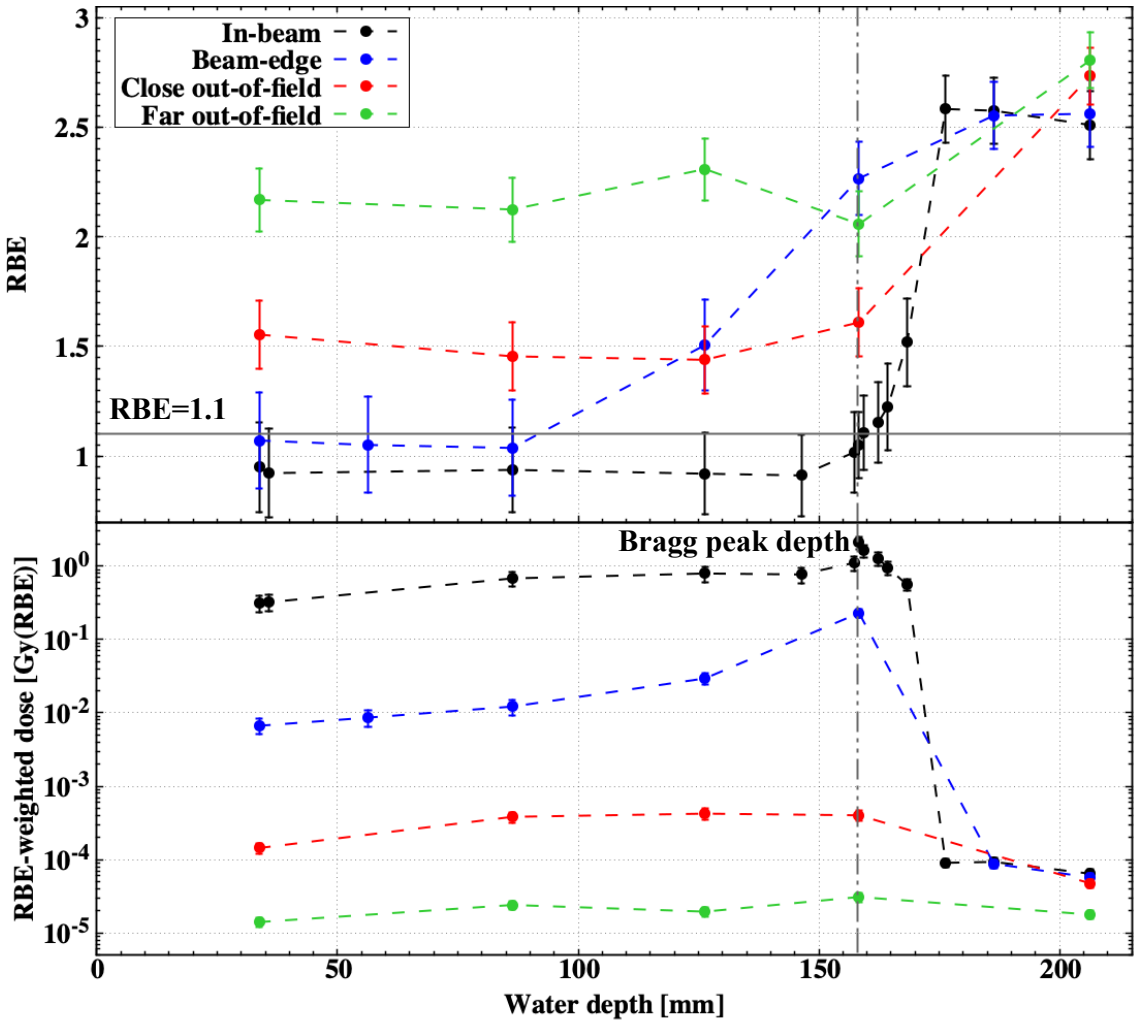

Figure 6.7: *Top panel*: In- and off-beam RBE values as a function of water depth. The off-beam regions are defined as *beam-edge* (20 mm from the beam axis), close out-of-field (50 mm from the beam axis) and far out-of-field (100 mm from the beam axis). All RBE values have been calculated according to Equation ((6.3)) when 2 Gy are delivered to the Bragg peak. The reference protons value of 1.1 is marked with a full line, while the Bragg peak depth (158 mm) is marked with a dotted line. *Bottom panel*: RBE-weighted dose in- and off-beam as a function of water depth. The off-beam regions are defined as *beam-edge* (20 mm from the beam axis), close out-of-field (50 mm from the beam axis) and far out-of-field (100 mm from the beam axis). The results have been obtained multiplying the dose (Figure 6.5) with the corresponding RBE value for each position.

value of 3.4±0.3 was found far out-of-field at a depth of 206 mm.

The rather large RBE variations are due to the fragments, whose lineal energy can be substantially higher than the primary protons as demonstrated by the $yd(y)$ spectra (Figure 6.2). In the entrance channel, the estimated RBE is always rather constant, reflecting that the $y*$, and thus the radiation quality do not change much. In-beam and at the field-edge, most particles are either primary protons or fast fragments with low y, and thus the RBE is close to the nominal value. The out-of-field regions, instead, are mostly populated by slower secondary particles, whose y is higher, and thus the calculated RBE is well above 1.1. At increasing depth, both primary ions and fragments slow down, the lineal energy increases and so does the RBE.

Although the RBE represents an indicator of the radiation effectiveness, it has to be combined with the dose to assess the biological outcomes. Thus, the data from Figure 6.6 (top panel) have been multiplied with the corresponding dose measurements (Figure 6.6) to obtain the RBE-weighted dose curves plotted in Figure 6.6 (bottom panel).

The high RBE observed outside the target area is in most cases heavily moderated by the dose, which drops below 10^{-3} Gy out-of-field. However, fairly high values are found at the *beam-edge* and at the end-of-range *in-beam*, pointing at these areas as at the highest risk of potential toxicities.

6.2 In- and out-of-field microdosimetric characterization of a 148 MeV proton spread-out Bragg peak: measurements and TOPAS benchmark

6.2.1 Experimental setup

All microdosimetric spectra were acquired with the LET-1/2 tissue-equivalent proportional counter TEPC as described in Section 1.2.1. The TEPC was positioned inside a water phantom (Blue Phantom, IBA Dosimetry) and exposed to a 3 cm radius SOBP produced with a passive modulator [Tommasino et al., 2019]. The in-field SOBP, corresponding to a 2 Gy uniform dose region, spreads between 105 mm and 132 mm water depth and its uniformity was measured to be at least 98%; the distal R90 was found instead equal to 134 mm.

We identified three different regions of interest for this study (Figure 6.8), and defined them with respect to the dose measured at the SOBP center (i.e. the maximum dose):

- *in-beam region* where the maximum dose is released. This area extended up to 30 mm laterally from the beam axis center;
- *beam-edge region* where the dose gradually decreases at increasing distance from the beam center, with an average of about 50% of the maximum dose. This region ranges from 40 mm to 60 mm from the beam axis center;
- *far out-of-field region* where the dose drops below 1%, extending from 60 mm to 150 mm laterally from beam axis center.

The exact positions where the TEPC was placed to measure the spectra are also marked in the Figure. During the acquisitions, the primary beam impinging on the passive modulator was monitored with an ionization chamber. In order to avoid pile-up effects described in Section 4.4, the particle rate at the TEPC was kept below 10^4 p/s for all measurements.

Microdosimetric spectra, RBE, H and Q factor

For each data point, we collected a statistics of about one million events. The raw spectra were then converted into lineal energy y distributions with a standard calibration procedure, whose details can be found elsewhere [Missiaggia et al., 2020].

From the y distributions, we obtained the $yf(y)$ and $yd(y)$ microdosimetric spectra. In addition, we calculated the frequency-mean lineal energy y_F and the dose-mean lineal energy y_D.

By combining the microdosimetric spectra with the biological α and β parameters of the human large cell lung carcinoma (*H460*), we calculated the RBE for cell survival both *in-* and *out-of-field*. We used the *Microdosimetric Kinetic Model (MKM)*, as extensively depicted in Section 2.2 and as already done in Section 6.1, but with the physical and cell model parameters described in [Bertolet et al., 2021]. The RBE formula is

$$\mathrm{RBE} = \frac{1}{2d}\left[\sqrt{\left(\frac{\alpha_0}{\beta_0}\right)^2 + 4d\left(\frac{\alpha_0}{\beta_0}\right)\mathrm{RBE}_{\max} + 4d^2\mathrm{RBE}_{\min}^2} - \frac{\alpha_0}{\beta_0}\right] .$$

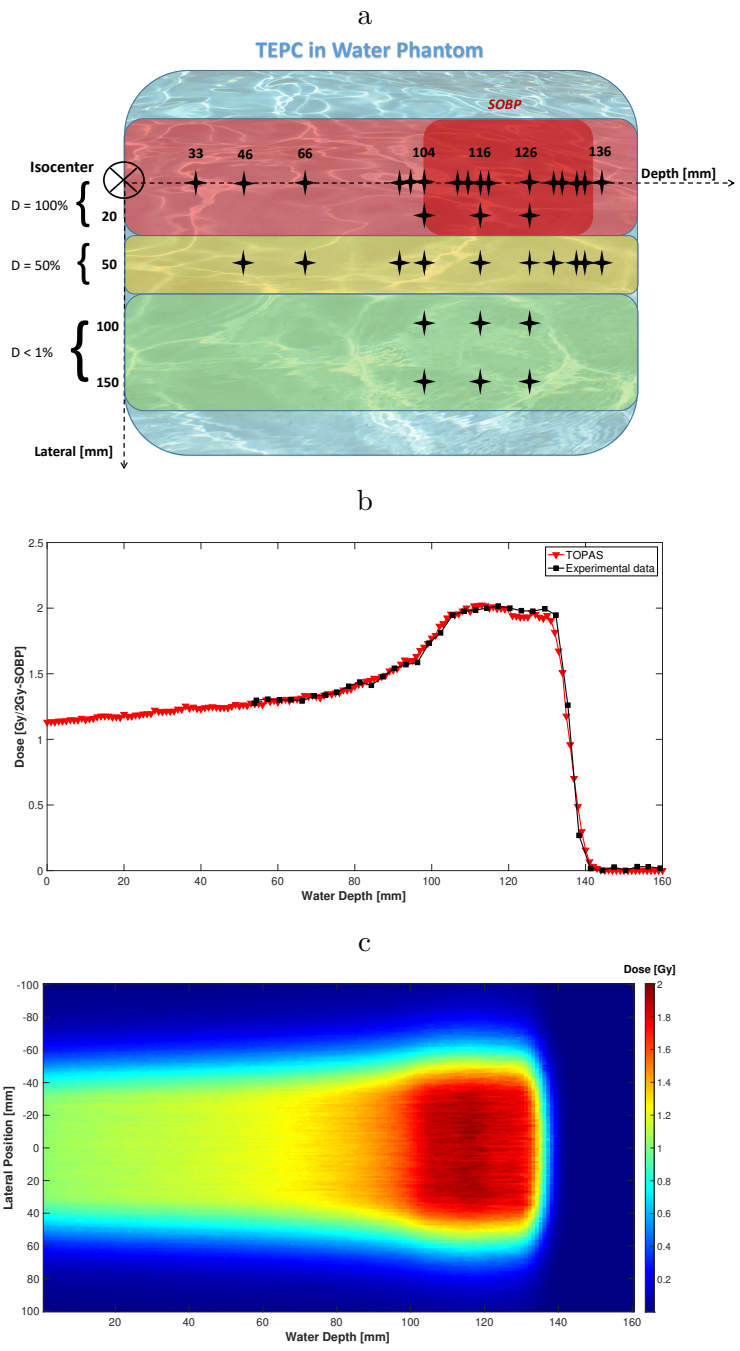

Figure 6.8: Panel (a): Positions measured in the water phantom with the TEPC. Three main regions have been considered depending on the dose measured at the entrance channel: *in-beam* region (D=100%), *beam-edge* region (D=50%), far *out-of-field* region (D<1%). Panel (b): 1D depth-dose profile measured in water with a CC01 ionization chamber and compared with TOPAS prediction. Panel (c): 2D dose distribution calculated with TOPAS in water.

where d is the dose per fraction, while $RBEmax = \frac{\alpha_P}{\alpha_0}$ (with $\alpha_P = \alpha_0 + \beta_0 \frac{y_D}{\rho\pi(r_d)^2}$) and $RBEmin = \sqrt{\frac{\beta_P}{\beta_0}}$ are the RBE asymptotic values for $d = 0$ and $d \to \infty$, respectively. In the MKM, RBEmin = 1 since $\beta_P = \beta_0$ is assumed.
For *H460* cells, α_0=0.29 ± 0.06 Gy^{-1}, β_0=0.032 ± 0.009 Gy^{-2} and r_d=370 ± 50 nm. The choice of using this formulation of the MKM that uses y_D values instead of the one used in Section 6.1 that exploits $y*$ values for the RBE assessment is due to the fact that we wanted to confront them, in the in-field direction, with the calculated RBE using the McNamara phenomenological model [McNamara et al., 2015], which exploits the LET_d:

$$\mathrm{RBE}\left[D_p, \left(\frac{\alpha}{\beta}\right)_x, \mathrm{LET}_d\right] = \\ = \frac{1}{2D_p}\left[\sqrt{\left(\frac{\alpha}{\beta}\right)_x^2 + 4D_p\left(\frac{\alpha}{\beta}\right)_x\left(p_0 + \frac{p_1}{(\alpha/\beta)_x}\mathrm{LET}_d\right) + 4D_p^2\left(p_2 + p_3\sqrt{\left(\frac{\alpha}{\beta}\right)_x}\mathrm{LET}_d\right)^2} - \frac{\alpha}{\beta}\right].$$

Here D_P is the proton dose, while LET_d is the dose average calculated with TOPAS considering both primary and secondary protons.

The radiation quality was either estimated from the experimental microdosimetric spectra (for MKM) or calculated with TOPAS (for McNamara). We then evaluated the RBE for the corresponding survival fraction.

In addition, following the approach presented in [Conte et al., 2020], we also calculated the RBE_{micro} in-field:

$$\mathrm{RBE}_{\mathrm{micro}} = \int r(y)d(y)dy\,.$$

$d(y)$ is the microdosimetric dose distribution measured at a given water depth. The quantity $r(y)$ is the Loncol's biological weighting function [Loncol et al., 1994] as discussed in 2.3.1 and relates to early intestinal intolerance assessed by crypt regeneration in mice. Thus, the RBE_{micro} is calculated for a different biological endpoint, cannot be directly compared to the MKM and McNamara RBE values. Loncol's method, however, is the only one that uses the whole microscopic $d(y)$ distribution instead of the average values y_D, as for the MKM, and thus is more sensitive to differences in the radiation field quality. For this reason, we decided to include it in this work.

The microdosimetric spectra were also used to estimate the quality factor Q and equivalent dose H [Zaider and Brenner, 1985] in the *beam-edge* and *out-of-field* regions, as discussed in 2.3.2, using the ICRU 40 approach.

Monte Carlo simulations

All simulations were performed with TOPAS Monte Carlo toolkit (version 3.2.2 [Perl et al., 2012]). We used the default TOPAS physics list recommended for protontherapy applications [Jarlskog and Paganetti, 2008], and the range cut values (i.e. the particle production and tracking thresholds) proposed in [Zhu et al., 2019]. To simulate the microdosimeter, we employed the spherical TEPC geometry available in the code, and selected the Trento beam line geometry already implemented in Geant4 [Tommasino et al., 2019].

First, we reproduced the 1- and 2D depth-dose profiles delivered in the water phantom with the passive modulation system (Figure 6.8 panels (b) and (c)). The 1D curve was verified against measurements acquired with a CC01 ionization chamber (IBA Dosimetry).

With the same geometry, we selected a $1 \times 1 \times 1\ mm^3$ mesh grid to obtain the dose and LET_D values at the positions where the TEPC was placed and the microdosimetric spectra were measured (Figure 6.8 panel (a)).
To simulate these spectra, we employed the TOPAS lineal energy scorer associated to the spherical TEPC. With this method, the total energy deposition inside the detector sensitive volume of both primary and secondary particles is recorded event-by-event, and then divided by the detector mean chord length to obtain the lineal energy y. All events are then stored into an histogram of the lineal energy frequency distribution $f(y)$. The number of primaries was set to match the experimental statistics ($\sim 10^6$ events scored by the detector) *in-beam* and at the *beam-edge*, but it was decreased to $\sim 10^5$ for the farthest *out-of-field* positions, due to computational time constraints.

The $f(y)$ histograms were scored for each measurement positions, recording the y distribution as well as the contributions coming from different particles (i.e. protons, heavy fragments and secondary electrons). Applying the same analysis described in Sec. 6.2.1, we obtained the $yf(y)$ and $yd(y)$ spectra, from which the microdosimetric quantities y_F and y_D were calculated.

6.2.2 Results

Microdosimetric spectra

A selection of the measured and simulated microdosimetric $yd(y)$ spectra at several depths in- and off-beam are plotted in Figure 6.9 in their standard semi-log representation (the area under the curve indicates the relative contribution to the total number of events).

Panel (a) shows the experimental distributions for six different positions between 96 and 136 mm depth, all acquired in-field. At increasing depth, we observe a gradual shift of the spectra towards higher y values. They are centered around 0.9 keV/μm in the entry channel, between 1 and 4 keV/μm in the SOBP, and around 7 keV/μm in the distal penumbra. We observe the same behavior in the spectra at the field edge, as well as at different lateral distances from the beam center and a fixed depth in the middle of the SOBP (Figure 6.9, panels (b) and (e)). Together with the mean value, also the distribution width is not constant, but increases at increasing depth and lateral distance from the field. The most significant difference is visible in the right tail, which is populated by high y events. The spectra change very rapidly *in-beam* and at *beam-edge*, while they appear to be more constant at increasing lateral distance from the field, where the mean value range from 1.3 keV/μm *in-beam* to about 2.5 keV/μm far-*out-of-field* (150 mm lateral). The sharp edge at around 120 keV/μm, especially visible in the distal position, namely the purple line at 136 mm of water depth, is known as "the proton edge", and represents the maximum energy that primary protons can deposit in the TEPC [Chiriotti et al., 2015b]. The microdosimetric spectra behavior stems from the combination of three effects: i) the energy straggling, which increases at decreasing particle energy, ii) the heterogeneity of the radiation field composition both in terms of particle type and energy, which increases with depth and lateral distance from the primary beam path, and iii) the coulomb scattering, which increases with decreasing particle charge and energy. In the entrance channel, the radiation field is mainly composed by monoenergetic primary protons, and thus the microdosimetry spectrum is relatively narrow. As the beam penetrates water, it both slows down and create secondary particle via nuclear interactions, whose y is equal or higher than the primary protons. The increasing field heterogeneity, combined with the energy straggling, generate broader spectra centered at higher y. Secondary particles

have a low abundance compared to primary protons in-field, but become dominant at the end-of-range, and eventually the only contributor to the field in the penumbra.

The comparison between measured and simulated $yd(y)$ spectra (Figure 6.9, panels (c), (d), (e), (f)) indicate an overall good agreement, as both the peak positions and distribution shapes are reproduced. The only exceptions are observed at the entrance channel and in the plateau region up to about 90 mm depth, e.g. the spectra *in-beam* at 33 mm depth (panel c) and at the *beam-edge* at 46 mm depth (panel e). This discrepancy is caused by the noise contamination discussed in Section 5.1.1. Also the *out-of-field* spectra (150 mm lateral distance from the beam center) shown in panel (f) indicate a deviation between measurements and simulations. This region is entirely populated by secondary particles, and the result can point to a limited accuracy of the nuclear models implemented in TOPAS.

Microdosimetric quantities

Using Equations (1.1.3) and (1.1.3), we calculated the frequency-mean lineal energy y_F and the dose-mean lineal energy y_D. The experimental values and TOPAS predictions at several depths in-field and at the *beam-edge* are shown in Figure 6.10. The y_F profiles appear very similar, indicating that the radiation field in those regions has the same quality. The only difference is seen at the SOBP, where the curve *in-beam* has a steeper gradient than at the field edge. Also for the y_D, we see the largest difference along the SOBP and in particular in the distal position: the *in-beam* value at the latter position reach 14.16 $\pm$ 1.31 keV/μm, while the *beam-edge* y_D at the same depth is 23.24 $\pm$ 2.10 keV/μm.
The comparison between the measurements and TOPAS predictions indicate an overall good agreement for the y_F, but a larger discrepancy for the y_D. The latter values are very sensitive to rare secondary events belonging to the high y region, and thus are expected to demonstrate larger fluctuations.

Biological and equivalent dose

The experimental data presented in this work, as well as the microdosimetric spectra and LET_d values calculated with TOPAS have been combined with the simulated physical dose profiles (Figure 6.8, panel (c)) to estimate the RBE with the MKM, Loncol and McNamara models as described in Section 6.2.1.
The results *in-beam* are shown as a function of the water depth in Figure 6.11 (panel (a)). For the Loncol and MKM, we reported two datasets, because we used both the experimental and simulated microdosimetric distributions for calculating the RBE. The McNamara describes the radiation field quality with the LET_d, which could not be obtained from our measurements, but only estimated with TOPAS. The RBE profiles refer to two different biological endpoints: the survival of H460 cells for the MKM and McNamara, and the early intestinal intolerance for the Loncol. To assess the RBE with the MKM and McNamara, we did not fix a cell survival percentage, but calculated it on the basis of the dose delivered in each position. This approach allowed us to assess the RBE also *out-of-field* using realistic doses and to calculate the biological dose in each position. The data show that independently of the model, the RBE values have similar trend: they are flat in the plateau region, they increase slightly in the SOBP, and they have a steep growth between the SOBP far end and the penumbra. At the entry channel, both the Loncol and the McNamara predict an RBE below 1.1, while the MKM is slightly higher

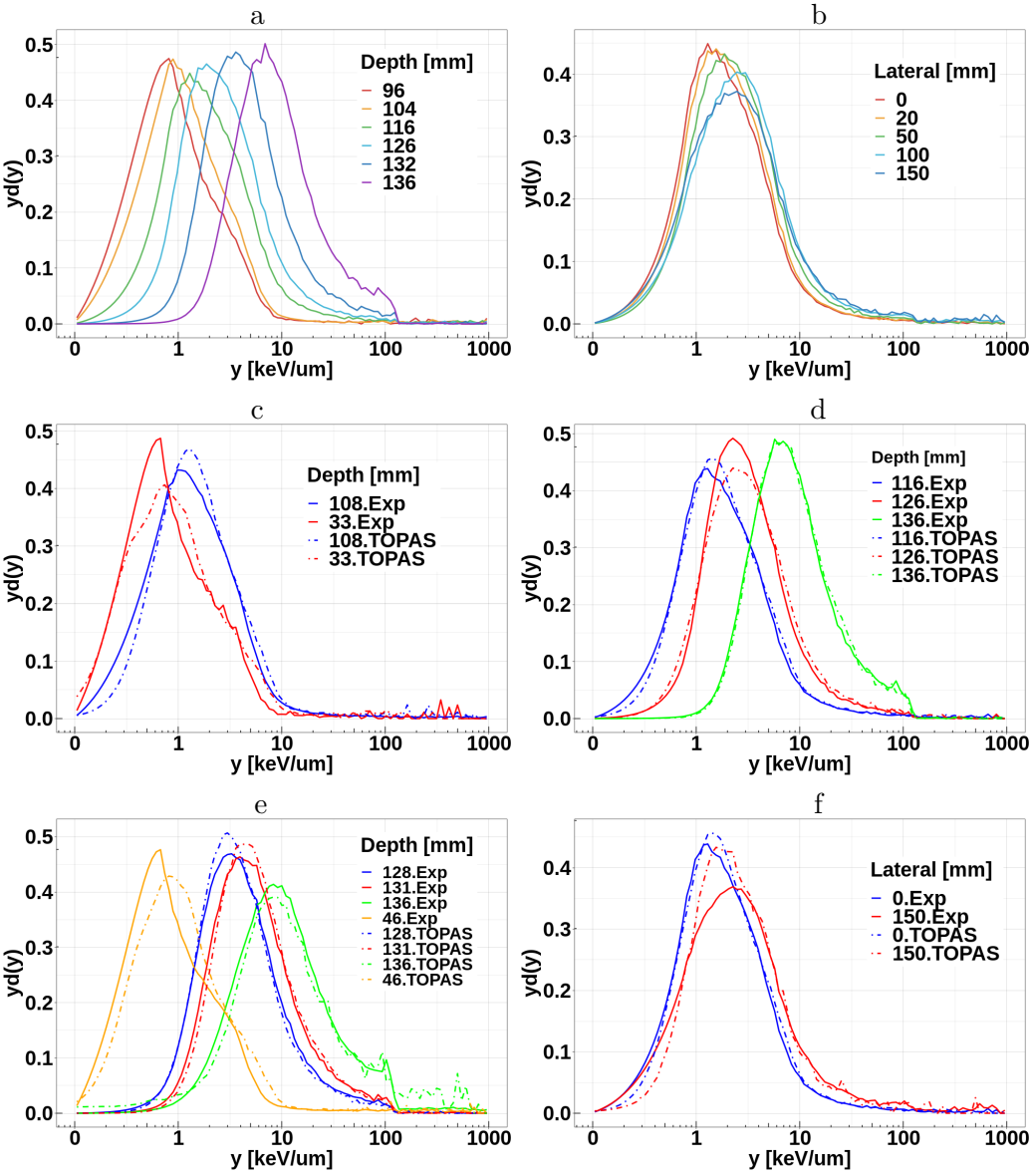

Figure 6.9: Experimental (solid lines) and simulated (dashed lines) microdosimetric yd(y) spectra at different positions. The positions in the SOBP region are marked with an asterisk. The distributions measured *in-beam* at several depths are shown in panel (a), while those acquired at a fixed depth of 126 mm and increasing lateral distance from the beam center are plotted in panel (b). Panels (a), (b), (c), (d) report the comparisons between experimental data (Exp) and simulated data (TOPAS) with the TOPAS Monte Carlo toolkit (only some depths are shown for visualization purposes). In particular, panels (c) and (d) refers to the *in-beam* region at different depths in water reported in the figure's labels; in particular, panel (c) refers to the entrance and plateau region while panel (d) shows comparisons at the SOBP and distal penumbra zone. Panel (e) contains instead the comparisons at different depths (reported in figure's legend) at the *beam-edge* while panel (f) illustrates two selected positions at a fixed depth of 126 mm, namely in the SOBP region (0 mm) and in the *out-of-field* region (150 mm).

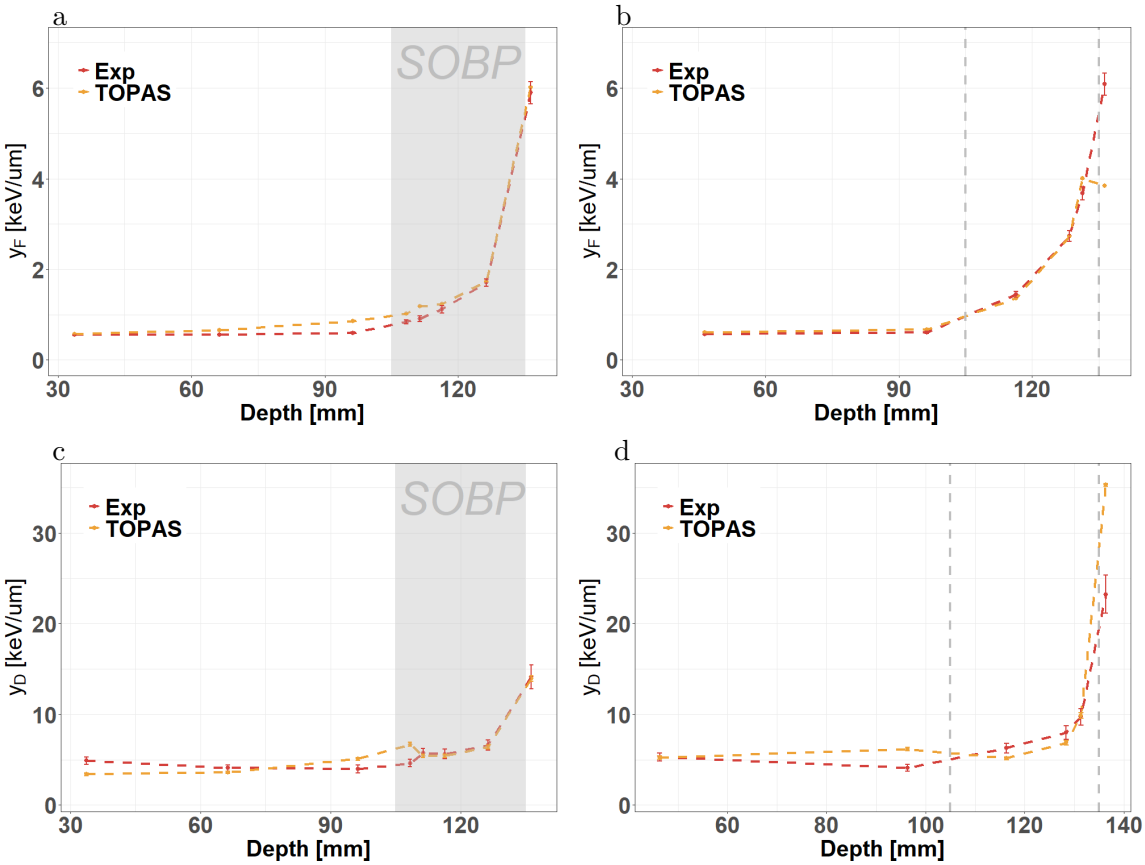

Figure 6.10: Simulated (TOPAS) and experimental (Exp) y_F (panels (a) and (b)) and y_D (panels (c) and (d)) values *in-field* and at the *beam-edge* (50 mm from field center). The grey area marks the SOBP region, while the grey dotted lines indicate the depths corresponding to the SOBP region. The experimental error bars include the contribution from both the energy calibration procedure and reproducibility, while TOPAS bars represent the statistical error.

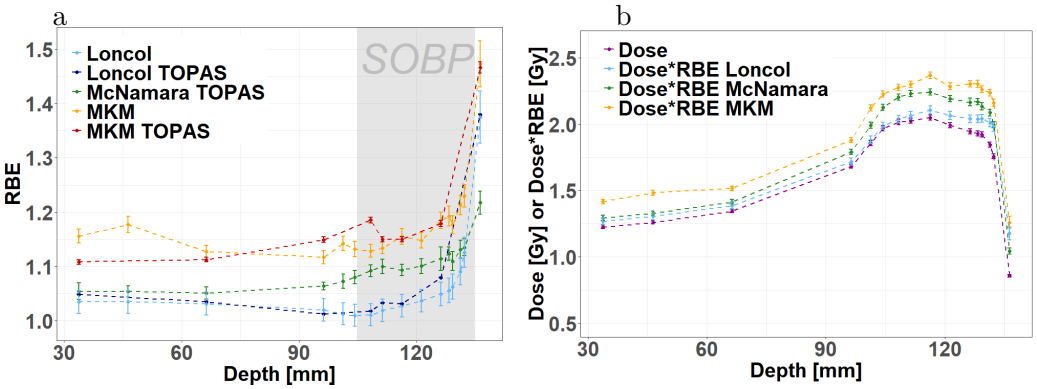

Figure 6.11: Panel (a) shows the RBE profiles in the *in-field* region calculated with the MKM, Loncol and McNamara models described in Section 6.2.1 for two different biological endpoints. The MKM and McNamara refer to the survival fraction of *H460* cells calculated from the dose value at each position, while the Loncol predicts RBE for early intestinal intolerance. For the McNamara, the LET_d was calculated with TOPAS, while for the MKM and Loncol the radiation quality was either obtained from the experimental microdosimetric spectra, or from TOPAS. The grey area marks the SOBP area. Panel (b) illustrates the physical dose calculated with TOPAS and the biological dose (Dose*RBE) estimated with all models. For the MKM and Loncol, we used the microdosimetric measurements presented here to describe the radiation field quality.

(1.15 ± 0.01). From the beginning of the SOBP to the penumbra, the McNamara ranges from 1.09 ± 0.01 to 1.21 ± 0.02, while the Loncol increases more rapidly (from 1.01 ± 0.02 at the beginning of the SOBP to 1.27 ± 0.03 at the penumbra), getting closer to the MKM (from 1.12 ± 0.01 to 1.47 ± 0.04, at the same positions). The RBE assessed with the measurements agree with the values entirely obtained from TOPAS for the Loncol, but show a larger discrepancy for the MKM. This behavior is caused by the fact that the MKM employs y_D values, whose simulated and measured results presented in Figure 6.10 show relatively large variations. The Loncol instead relies on the entire microdosimetric spectra, for which the overall agreements between experiments and calculations is more robust.

The RBE values from McNamara, as well as those from the Loncol and the MKM based on the microdosimetric measurements, have been used to calculate the biological dose, i.e. the physical dose multiplied by the RBE (Figure 6.11, panel (b)). As we employed the same physical dose curve for all calculations (Figure 6.8, panel (b)), the biological dose reflects the RBE trend. The highest values are obtained with the MKM, ranging from 1.41 ± 0.02 at the entry channel to 2.26 ± 0.03 in the SOBP. In the entry region, the McNamara and the Loncol are close to the physical dose, as both models predict an RBE around 1.05. At increasing depth, the RBE from McNamara grows more rapidly, and the biological dose reaches a value of 2.24 ± 0.02 at the SOBP, approaching the MKM, while the Loncol remains very close to the physical dose profile.

Using both our experimental data and TOPAS outputs, we computed the quality factor Q according to Eqs (2.54) and (2.56), as well as the equivalent dose H according to Equation (2.55). The results as a function of water depth are shown in Figure 6.12 (panels (a), (b) (c) and (d)) for the *beam-edge* region and for the lateral profile at a fixed depth of 116 mm (middle of the SOBP). The RBE calculated with the MKM and the corresponding

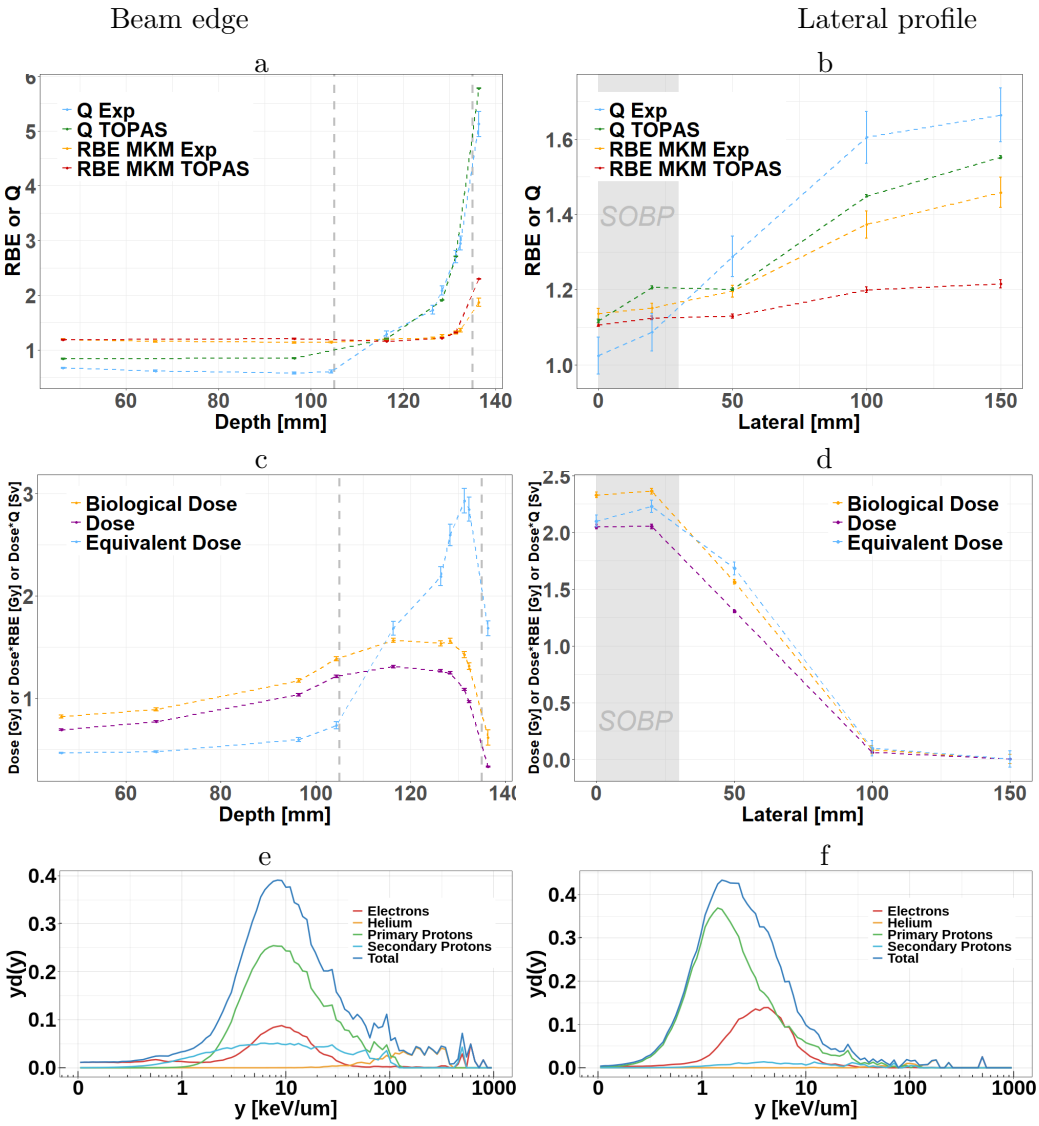

Figure 6.12: Panel (a) and (b) shows both the RBE (calculated exploiting the MKM model for H460 cells) and the Q values depth profiles for the *beam-edge* region and for the lateral profile at a fixed depth of 116 mm, respectively, comparing experimental microdosimetric data (Exp) to TOPAS simulated spectra (TOPAS).
In panels (c) and (d), instead, the physical, RBE-weighted and equivalent dose for the *beam-edge* region and for the lateral profile are illustrated, in both cases calculated from experimental data.
In addition, particles contribution to the yd(y) spectra are plotted for the higher RBE and Q values of the *beam-edge* and of the lateral profile (respectively in (e) for the 136 mm in depth at 50 mm in the lateral direction and in (f) for the 116 mm in depth at 150 mm in the lateral direction).

biological dose are reported for comparison. We chose the MKM as reference as it describes radiation quality with microdosimetry and allows to calculated RBE values for different biological endpoints. In both regions, the Q and RBE curves follow a similar trend, but with different values. At the *beam-edge*, Q is lower than the RBE in the plateau (Q=0.07 ± 0.01, RBE=1.18 ± 0.02). Then, it steadily increases in the SOBP region, reaching a value of 2.95 ± 0.13 and overcoming the RBE (1.36 ± 0.03). Also in the penumbra, the growth is steeper for Q than for the RBE, with maximum values of 5.12 ± 0.23 and 1.87 ± 0.08, respectively. In the lateral profile, Q increases rapidly, ranging from 1.02 ± 0.05 at the beam center to 1.66 ± 0.07 *out-of-field*, while the RBE is relatively flat (1.13 ± 0.01 to 1.46 ± 0.04). The equivalent dose H reflects the Q behavior, remaining constant around 0.5 Sv at the entrance, and then rapidly increasing up to 2.93 ± 0.13 Sv at the SOBP far edge. In the lateral profile, H is constant *in-beam* (2.10 ± 0.10) and then drops at increasing distance from the SOBP down to 0.0038 ± 0.0002. H appear to be higher than the RBE at the SOBP far-end and *beam-edge*, but then the two quantities converge on similar values *out-of-field*.

Off-beam, we selected two positions with the highest Q and RBE values (*beam-edge* at 136 mm depth, lateral profile at 150 mm distance from the beam center and 116 mm depth), and investigated the contributions of different particles to the yd(y) spectrum as well as to the total Q value. In both cases, for each bin of the histograms we evaluated the contribution of every single particle type. To obtain the Q value of each component, in addition, the integrations 2.54 have been evaluated. The results were obtained using TOPAS and are plotted in panels (e) and (f) of Figure 6.12. Due to CPU constrains, the simulations have a lower statistics (10^5 events) with respect to the standard value (10^6 events). This issue might especially affects rare events, e.g. heavy fragments such as helium, that mostly contribute to the high y region (right tail). We scored as primary protons all source protons which did not undergo inelastic collisions. Secondary protons are defined as particles produced from nuclear interactions of the primary beam with the target.
At the *beam-edge*, the major contribution comes from the primary protons, whose profile peaks around 8 keV/mum and matches the total distribution. Electrons have the same trend of the primary beam, while secondary protons are relatively flat, extending mainly from 2 keV/mum to 100 keV/mum, with an additional contribution to the very high y tail (around 500 keV/mum). Helium ions are only observed in the high y region starting around 100 keV/mum until 400 keV/mum, where they dominate this portion of spectrum. The corresponding contributions to the total quality factor are: 47% primary protons, 24% helium, 18% secondary protons and 11% electrons.
In the lateral position *out-of-field*, primary protons become ever more dominant and their distribution is centered around 1.2 keV/mum. Also electrons peak at a smaller value (4 keV/mum), which is higher than primary protons and result in broadening of the total spectrum on the right side. The contributions to the total Q are: 46% electron, 31% primary protons, 7% secondary protons and 16% helium.

To conclude, overall the microdosimetric spectra indicate that radiation quality changes more rapidly with depth than with lateral distance from the SOBP. in-field, the y_D rises from $\sim$ 4 keV/μm at the entry channel to $\sim$ 8 keV/μm at the far-end SOBP, and reaches the highest value of $\sim$ 15 keV/μm at the penumbra. At the *beam-edge*, the y_D are very similar to the values in-field, mostly because many primary protons reach this region due to Coulomb scattering and become the dominant component of the field. Moving lat-

erally from the SOBP center, the y_D changes from 5.62 ± 0.26 to 11.14 ± 0.35 at 150 mm depth. In addition, the comparison between y_D and LET_d values suggests that the estimate of radiation quality is largely influenced by the methodology used for describing it, especially outside the SOBP. Our hypothesis is that this difference is mostly caused by short-range heavy fragments, which have an extremely lower abundance as compared to protons, but deposit significantly more energy. As microdosimetry is more sensitive in detecting fluctuations of energy deposition, it appears to be more accurate in characterizing radiation quality at the cellular level. The combination of y_D values and physical dose identifies two regions of potential elevated risk of toxicity: the *beam-edge* (dose 0.7 to 1.3 Gy/2Gy-treatment) and the distal penumbra in-field (∼ 0.5 Gy/2Gy-treatment).

For what concerns the comparison between RBE and Q profiles out-of-field, the two quantities do not always agree with each other, not showing a clear trend. It has to be considered also that, by definition, Q predicts the radiation biological effectiveness at very low doses, but the accuracy and the precise extent of the dose range validity has not been extensively verified at the out-of-field doses due to therapeutic irradiation. On the other hand, the precision of RBE for cell survival at low doses might be limited. Our findings highlight the limitations of both the RBE and Q in describing the biological effectiveness at the field-edge and out-of-field.

6.3 Discussion on the monoenergetic and SOBP proton beam characterizations

Finding a solid bridge between the physical quantities describing a radiation field and the corresponding biological effects, and translating them into medically relevant results represents one of the most difficult challenges in radiotherapy. To address this issue, the National Cancer Institute created a dedicated panel whose task was to identify the most suitable set of physical parameters for characterizing a particle beam [Durante et al., 2019]. When applied to clinical research, the recommendations would improve the intercomparison of different biological experiments and ultimately help understanding the relationship between the physics parameters and the clinical outcomes. The report produced by the panel included also a list of standard measurements recommended for obtaining a detailed beam characterization.
Microdosimetry was listed among the techniques for investigating beam purity as well as for assessing the radiation field quality. The absorbed dose alone does not carry enough information to assess the biological and, ultimately, the clinical effects of particle beams. A full characterization of the radiation field in terms of particle species and kinetic energy (or LET), as well the information on the radiation track structure are necessary tools for predicting the biological damage [Conte et al., 2017]. Aside from providing a better understanding of the particle effectiveness in tumor killing, these data can help predict the risk of undesired effects following the treatment.

In literature, there are very few microdosimetric studies of therapeutic proton and heavy ion beams including out-of-field data [Wroe et al., 2007, Wroe et al., 2009, Martino et al., 2010, Tran et al., 2017, Chartier et al., 2018, James et al., 2021]. These data are crucial to characterize the radiation quality received by the normal tissue, which is a key parameter for assessing the risk of toxicity. In addition, to the best of our knowledge, only [Martino et al., 2010] employ a Tissue Equivalent Proportional Counter (TEPC), which is the reference detector in microdosimetry.
In this chapters, we have presented a microdosimetry-based methodology for characteriz-

ing the beam quality, which can be combined with biological parameters to estimate the RBE for cell survival, both for a 152 MeV monoenergetic proton beam and a 148 MeV Spread Out Bragg Peak proton radiation field. All microdosimetry spectra indicate that the number of high lineal energy particles increases with increasing depth both in-field and out-of-field. In particular, the radiation quality changes more rapidly with depth than with lateral distance from the field *in-beam*. The change in the high-y population is caused by the production of secondary particles, mostly with higher charge and lower energy than the primaries, and translates into a RBE growth along both the distal and lateral directions.

To understand the biological significance of the calculated RBE values, they have to be combined with the absorbed dose. In the monoenergetic beam study, we evaluated the dose directly from the microdosimetric spectra, while for the SOBP we assessed the dose from TOPAS simulations. Both methodologies can be used, but they represent different information: the microdosimetric-based assessment of the dose is evaluated in the volume of the TEPC active region (micrometric dimension), while the simulated dose is evaluated in a standard volume (voxel) of millimetric dimension. Only for the voxel approach, stochasticities of energy deposition can be neglected. In addition, the microdosimetric method can be used only if the number of primaries delivered is known while, for the TOPAS-based approach the number of delivered primaries is easily known (corresponding to the statistics of the simulation performed by the user). On the other hand, the Monte Carlo method requires a validation in order to assess its accuracy, and this can not be the case for certain radiation fields, e.g. the one produced by ions heavier than protons where experimental cross section are still not completely reliable, especially in the out-of-field regions [Muraro et al., 2020]. On the contrary, the microdosimetry approach relies only on the experimental measurements, thus being truthful for every radiation field.

The results on the biological doses point to the *beam-edges* (lateral as well as distal) as the off-beam regions exposed to the highest combination of dose and RBE (or Q), and thus having the highest risk of potentially developing toxicities. Although the dose received out-of-field is relatively low, recent studies have demonstrated that protons in the therapeutic energy range can induce neurocognitive defects in mice with doses in the cGy range [Cekanaviciute et al., 2018, Kiffer et al., 2019].

In addition, assessing the biological dose in the out-of-field region is relevant since an incorrect estimate can have two potential outcomes: it can cause toxicity if underestimated, and it can limit dose escalation if overestimated. Thus, increasing the correctness of the RBE estimate both inside and outside the tumor region will result in a direct improvement of treatment.

Another important finding of the microdosimetric study on the SOBP, is that the microdosimetric spectra are in good agreement with the TOPAS Monte Carlo simulated distribution, both in- and out-of-field. This study represents, to the best of our knowledge, the first extensive benchmark for protons (especially out-of-field) of the microdosimetric extension recently introduced in TOPAS [Zhu et al., 2019].

Conclusions

In conclusion, we showed that variations in the radiation quality both in depth and in lateral distances from the beam axis result in RBE values diverging from the reference value of 1.1. The extent of these variations depend on the approach used for describing the radiation quality and on the radiobiological model used to assess the RBE. We need to evaluate if these differences have a quantitative impact on the treatment planning and

to assess the potential clinical consequences both in terms of tumor control and normal tissue toxicity. The achievement of this goal calls for accurate radiobiological data to validate the RBE models, and to investigate the validity of assuming a fixed value, as it is currently done in the clinical practice. Implementing a variable-RBE based optimization in the treatment planning system represents an opportunity to decrease the patient-specific biological dose for organs-at-risk despite the RBE uncertainties. Furthermore, we proved that TOPAS microdosimetric extension is a robust tool for predicting distributions both in- and out-of-field in a clinical proton irradiation scenario.

Chapter 7

Microdosimetric characterization of ^{4}He and ^{16}O therapeutic beams

In the present chapter, we will present a comprehensive microdosimetric characterization of the radiation generated in- and out-of field by 160 MeV/u ^{4}He and 365 MeV/u ^{16}O beams penetrating water. This study lays the groundwork for the future clinical application of oxygen in radiotherapy, as well as a more accurate knowledge of the helium behavior.

7.1 Material and Methods

Experimental setup

All measurements were performed at the NASA Space Radiation Laboratory [La Tessa et al., 2016], using 160 MeV/u ^{4}He and 365 MeV/u ^{16}O pencil-like beams. The energies have been chosen to have the same range in water for both ions and to be therapeutically relevant (175 mm). The oxygen beam had circular profile with approximately 10 mm diameter, while the helium beam was much broader (30 mm diameter), and had to be collimated to produce a pencil beam. The collimator, placed just downstream of the exit window, was made of 2 pieces: a 102 mm squared brass piece of 14 mm radius aperture and 51 mm thick, and a copper piece of 8 mm radius aperture.

All microdosimetry spectra have been measured with the LET-1/2 I. Unlike the other studies presented in this work, where the filling gas was pure propane, for this experiment the detector sensitive volume was filled with a propane-based gas (55% C_3H_8, 39.6% CO_2 and 5.4% N_2 in volume) whose pressure was adjusted to simulates a tissue-equivalent sphere of 2.7 μm diameter. The reason of this difference is explained in Section 1.2.1.

The TEPC was placed inside a 30×30×30 cm^3 water phantom, whose PMMA walls were 20 mm thick (23.6 mm water equivalent). The number of primary ions impinging on the water phantom was monitored with an ion chamber (0.7 mm water equivalent thickness) and a 2 mm plastic scintillator. A scheme of the experimental setup is shown in Figure 7.1 (top panel).

To select the measurements positions, the TEPC was centered on the beam axis using a laser system. Following the previous work described in 6, four different regions were identified as of interest for this study:

- *in-beam region*, where the detector was fully and homogeneously irradiated by the primary beam;

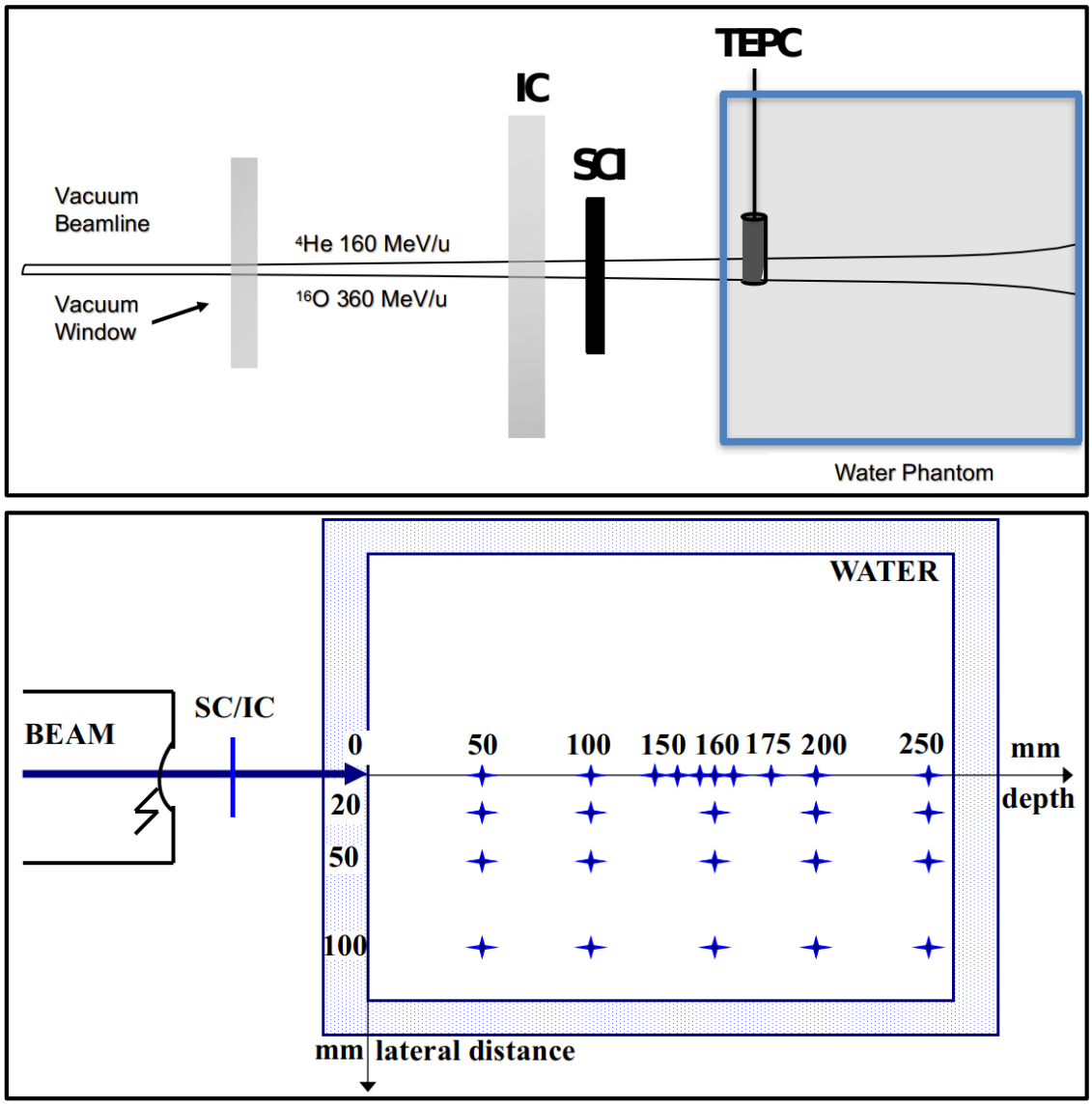

Figure 7.1: Scheme of the experimental setup for both helium and oxygen experimental campaigns (top panel) and of the TEPC measurement positions inside the water phantom (bottom panel). The number of primary ions impinging on the water phantom was monitored with an ion chamber (IC) (0.7 mm water equivalent thickness) and a 2 mm plastic scintillator (SCI). All distances and thicknesses are in mm. The depth also takes into account the water phantom wall, whose thickness is expressed in water equivalent. The Bragg peak of both ions is at 175 mm depth.

- *beam-edge region*, where the detector was moved 20 mm laterally from the beam axis (i.e. just outside the primary irradiation field);
- *close out-of-field region*, where the detector was moved 50 mm laterally from the beam axis;
- *far out-of-field region*, where the detector was moved 100 mm laterally from the beam axis.

For each region, the microdosimetric spectra were acquired at different water depths both upstream and downstream of the Bragg peak. A scheme of the measurement positions inside the water phantom is shown in Figure 7.1 (bottom panel).

Monte Carlo simulations with TOPAS

Monte Carlo simulations of the measurements have been performed exploiting the microdosimetric extension of TOPAS [Zhu et al., 2019], as previously stated in Section 5.2. The geometries implemented for helium and for oxygen are truthful to the experimental ones described in 7.1.
Following the state-of-the-art available in literature [Muraro et al., 2020, Arce et al., 2021, Bolst et al., 2020] we employed two of the different hadronic interaction descriptions both suitable for heavy ions, namely the *Quantum Molecular Dynamics* (QMD) and the *Binary Cascade* (BIC), and paired with the same electromagnetic list ($g4em - standard\_opt4$).

Pile-up study

In the entrance channel region, both helium and oxygen ions microdosimetric spectra have been found to be influenced by the pile-up effect. Following the methodology described in Chapter 4.4, we compared the experimental spectra to TOPAS simulations calculated with and without a given pile-up probability, and were able to quantify this effect. This can be therefore considered in the mean lineal energy quantities error bars.

7.2 Results

The behavior of the microdosimetry spectra is ruled by electromagnetic and nuclear interactions of the ions with water, and can be interpreted with the help of the measures discussed in this work in Chapter 6, as well as of experimental [Martino et al., 2010, Zeitlin and La Tessa, 2016, Rovituso and La Tessa, 2017] and of Monte Carlo [Zhu et al., 2019, Burigo et al., 2015] studies available in literature. The main aspects that have to be considered are summarized below.

- At the entrance channel, the field is mostly composed by monoenergetic primary ions and hence has a rather uniform lineal energy y. At increasing water depth, the electromagnetic interactions cause the beam to loose energy and thus to increase the y;
- helium suffers a higher energy straggling than oxygen;
- the probability to undergo nuclear interactions increases with increasing depth. Thus, when moving both along the beam path and in the lateral direction, the number of fragments increases and eventually they becomes the only contributors

to the spectra. Published studies report a fragmentation cross sections measured in water of approximately 700 mb for helium and 1600 mb for oxygen [Rovituso and La Tessa, 2017] in the energy range of interest. These results indicate that, at a given depth, the probability for oxygen to change charge is over two times higher than for helium;

- nuclear fragmentation yields charged and neutral fragments emitted at different angles with a wide range of energies (and y) that can exceed that of the primary beam;
- although both electromagnetic and nuclear interactions can lead to a deviation of the incoming radiation from the initial path, Multiple Coulomb Scattering dominates. At a given energy, the scattering angle increases with the decreasing of particle charge and the decreasing of kinetic energy; for both reasons, helium suffers more lateral scattering that oxygen.

On the basis of these points we can interpret the measured data.

Microdosimetric spectra

Figure 7.2 presents $yd(y)$ microdosimetric spectra for helium and for oxygen ions at different depths, both along the in-field direction, and off-beam at the *beam-edge* and in the far-out-of-field region. Details on the depths considered can be found in the figure's legend.

In-beam, oxygen peaks are narrower than the helium ones. The general trend is however for both ions similar, along the in-field direction: the spectra peak at higher y values with increasing water depth until they reach the Bragg peak position, after which the distributions becomes larger and the peaks move back to lower y values. In detail, helium distributions are centered around $\sim$ 4 keV/μm at the entrance channel, grow up to 15 keV/μm at the Bragg peak, and then decrease to $\sim$ 5 keV/μm in the distal region. For oxygen, the spectra peak $\sim$ 20 keV/μm at the entrance, increase to $\sim$ 150 keV/μm at the Bragg peak and then move back to $\sim$ 60 keV/mum in the distal region.

This is due to the fact that after the Bragg peak, the radiation field is mainly populated by fragments lighter than the primaries. For both ions, these are mostly particles with a Fermi momentum transferred, which imparts an additional energy component that leads these particles to deposit dose after the Bragg peak.

Another feature displayed only by oxygen is that the spectrum at the Bragg peak consists of both a main peak centered around 150 keV/μm mainly populated by primaries and a second much smaller peak at around 80 keV/μm of secondaries. For what concerns helium instead, a similar characteristic behaviour arise at 180 mm depth in water: here two peaks of nearly the same height are visible. The first peak at $\sim$ 2 keV/μm is populated mainly of secondary protons, while the second peak at $\sim$ 50 keV/μm is due to helium ions.

At the *beam-edge* region, helium show two different behaviour, with the fist positions in water peaked at $\sim$ 10 keV/μm and the last two positions at $\sim$ 5 keV/μm. Helium ions can be scattered at large angles, so that at 50, 100 and 150 mm in the *beam-edge* distributions at higher values of y are shown due to slower α particles.

The distributions given at 200 and 250 mm depth instead, are similar to the far out-of-field ones, since they are due to the secondary fragments produced by the helium nuclear interactions. At large angles the spectra are mainly populated protons, deuterons and neutrons, and the radiation field is very similar in all the depths.

Also the oxygen distributions in all the off-beam positions are populated by light secondary

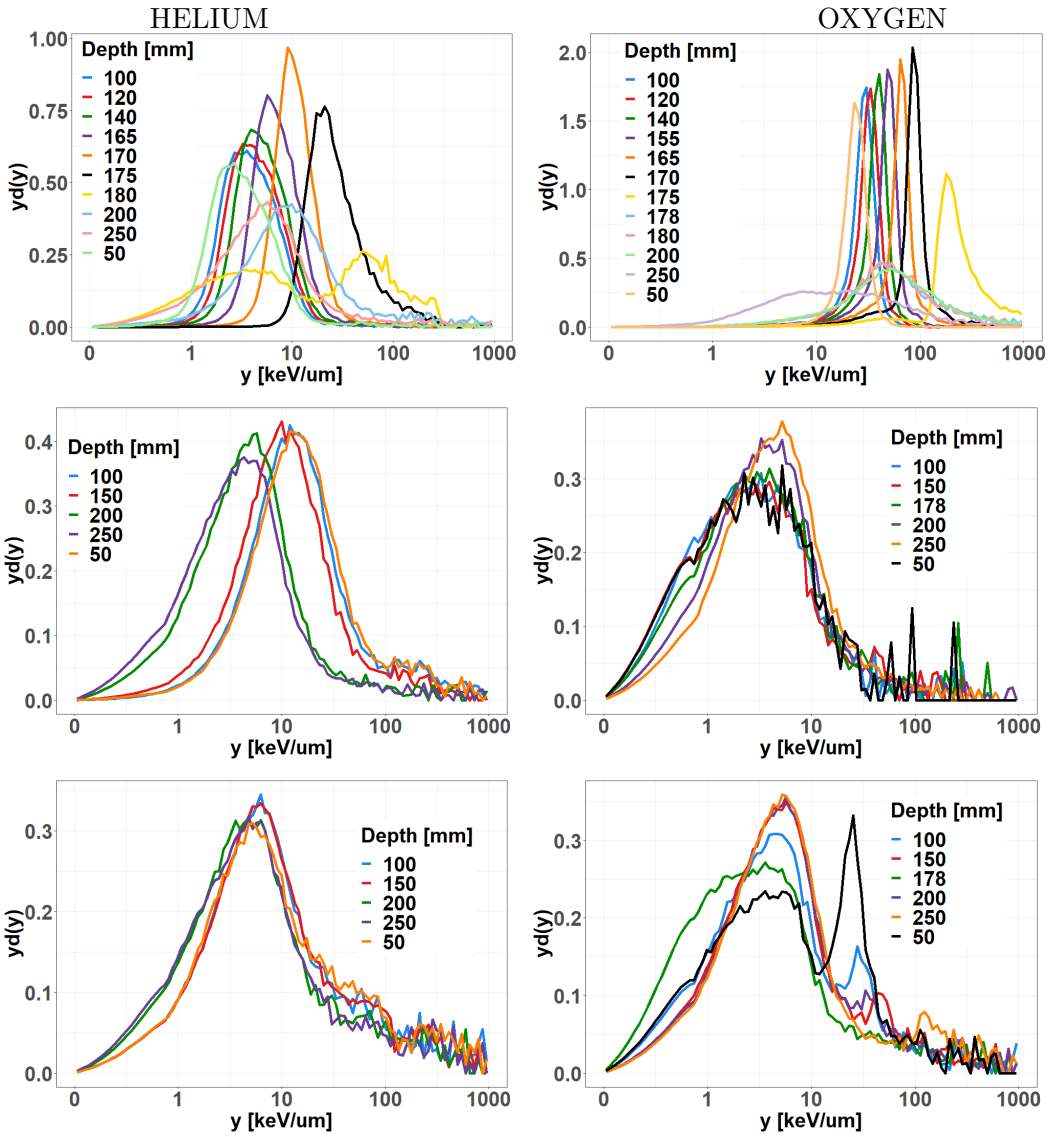

Figure 7.2: In- and off-beam microdosimetric lineal energy spectra $yd(y)$ measured at several depths for helium (left panels) and for oxygen (right panels). The off-beam regions reported in the central and in bottom panels are defined as *beam-edge* (20 mm from the beam axis) and far out-of-field (100 mm from the beam axis). The Bragg peak position is at 175 mm depth for both ions.

fragments (protons, deuterons, neutrons and α particles) with lower lineal energy values and resemble the distributions of helium in the same regions.
In the far out-of-field region however, oxygen shows a clear peak at $\sim$ 20 keV/mum at the entrance channel, that rapidly decreases at deeper positions in water, where just the lower y peak is present (bottom right panel of the Figure 7.2).

Furthermore, the in-and out-of-field spectra have been simulated with TOPAS. The comparison between experimental and simulated data is shown in Figures 7.3 (helium) and 7.4 (oxygen). Independently of the primary ion, the agreement *in-beam* is extremely good in the entrance and plateau channel (top panels), while it worsen in the Bragg peak and in distal positions (bottom panels, right and left panels respectively).

Using TOPAS, we also estimated the the pile-up probability, whose values are reported in Figures 7.3 and 7.4. For helium, a pile-up probability of 8% was found at the entrance and plateau regions *in-beam*, which decreased to 0 around the Bragg peak. For the spectra affected by pile-up, we also reported the simulations calculated with pile-up zero. For oxygen, instead, the entrance and plateau distributions have a 2% pile-up probability, which decreases to 0 around the Bragg peak position.

For both helium and oxygen ions, we calculated the microdosimetric spectra using the two hadronic models QMD and BIC. The results along the *in-beam* direction until 180 mm in water showed no appreciable differences. For this reason, in all these cases we chose the BIC hadronic model in order to be consistent with what was previously done for protons in Chapter 6.

On the contrary, out-of-field spectra have not been shown due to the overall disagreement between the simulated spectra for both hadronic model implemented (BIC and QMD). An example of this is given in Figure 7.5, where helium in the *beam-edge* region at 100 mm depth is not correctly described by neither QMD nor BIC hadronic list.

Radiation quality

Radiation quality is assessed by the frequency-mean lineal energy y_F, dose-mean lineal energy y_D and the saturation-corrected dose-mean lineal energy y^*, which can all be estimated from the microdosimetric spectra as described in the Section I. In Figure 7.6, the three mean quantities have been reported both for helium (left panels) and oxygen (right panels) along the beam direction at increasing depth in water. We also reported the values obtained from TOPAS simulated spectra, but only for the positions *in-beam* where the agreement with the measurements was found good. When the match was obtained with a certain pile-up probability, we plotted the microdosimetric quantities calculated both with and without pile-up.

For both ions, the values of y_F, y_D and y^* *in-beam* appear rather flat in the entrance channel and plateau. For oxygen, all quantities reach the maximum at the Bragg peak position, while for helium the y_D peaks in the tail region at 175 mm depth. This behavior is a direct consequence of nuclear fragmentation. Up to the Bragg peak, the primary ions represent the main component of the radiation field and thus heavily influence the quality. The tail, instead, is populated by fragments of very different y (as seen in Figure 7.6), and thus y_F, y_D and y^* represent an average of the quality of each component of the spectrum. Oxygen can only produce lower-Z fragments and thus the highest y is reached at the Bragg peak position where the primaries stops.

The difference between y_D and y^* is almost negligible in the plateau but becomes very significant at the Bragg peak and in the tail region. The saturation correction has two very different effects when applied to helium and oxygen. In the former case, it shifts

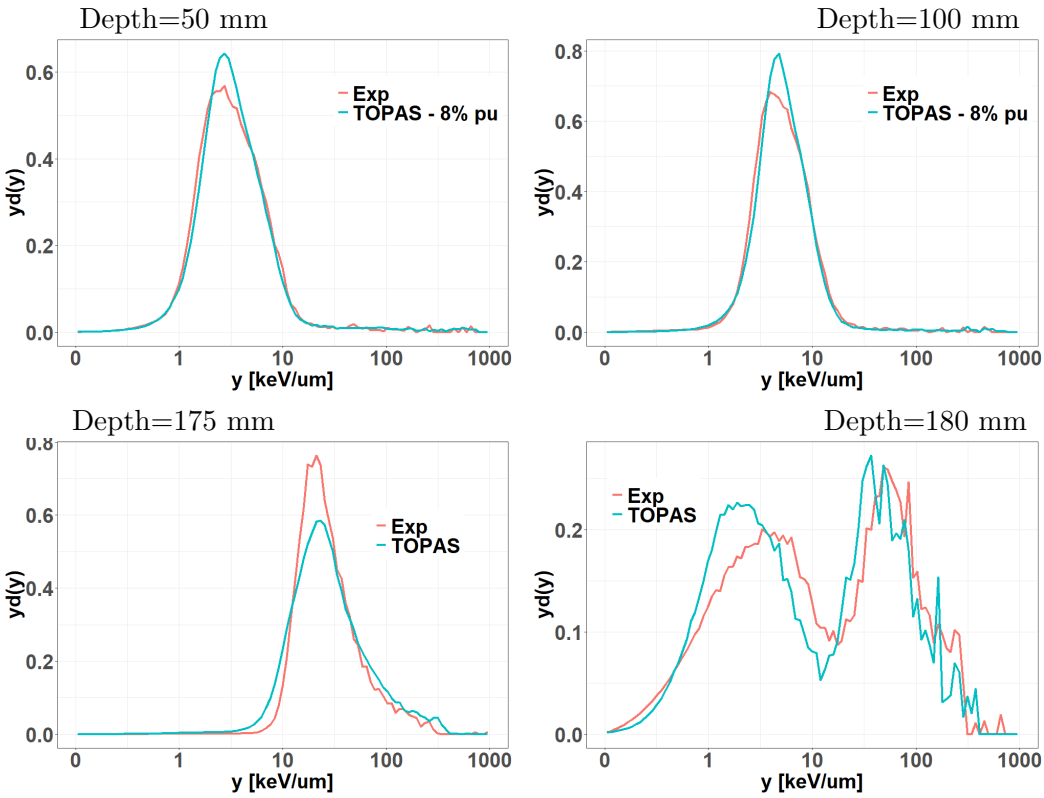

Figure 7.3: Comparison between experimental (Exp) and simulated (TOPAS) $yd(y)$ microdosimetric spectra measured at several depths for helium. For the simulated distributions, we used different pile-up probabilities (pu) and selected the value that provided the best agreement with the measured spectrum (reported in the legends).

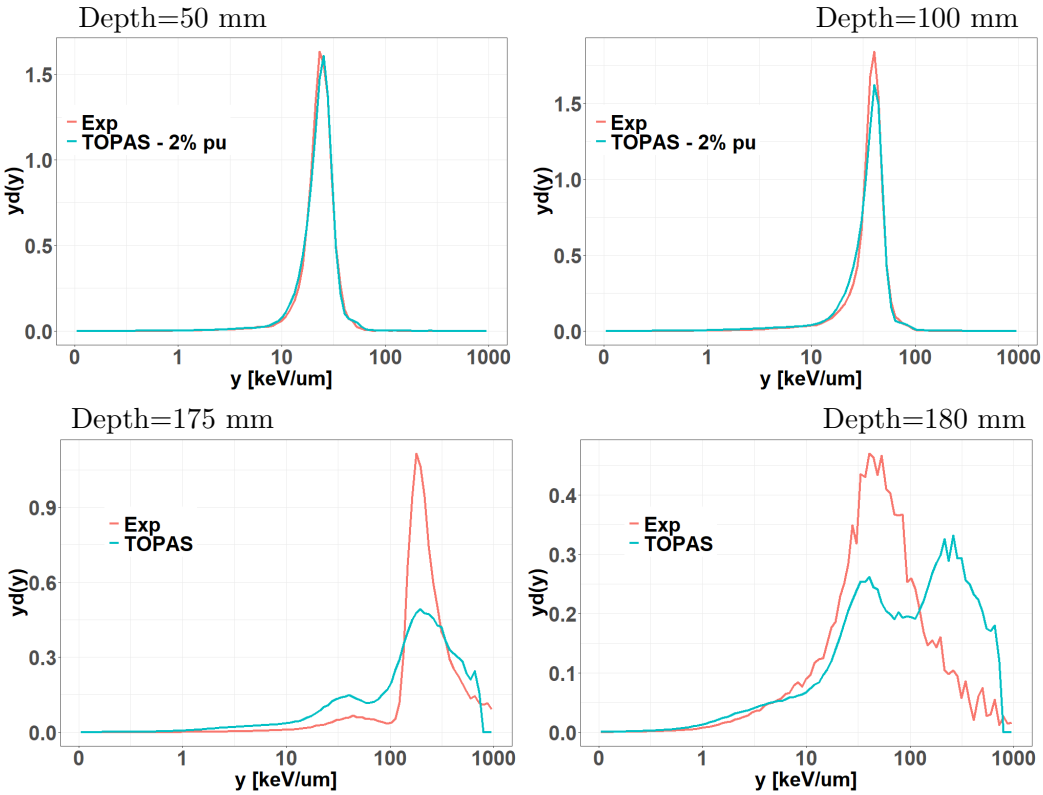

Figure 7.4: Comparison between experimental (Exp) and simulated (TOPAS) $yd(y)$ microdosimetric spectra measured at several depths for oxygen. For the simulated distributions, we used different pile-up probabilities (pu) and selected the value that provided the best agreement with the measured spectrum (reported in the legends).

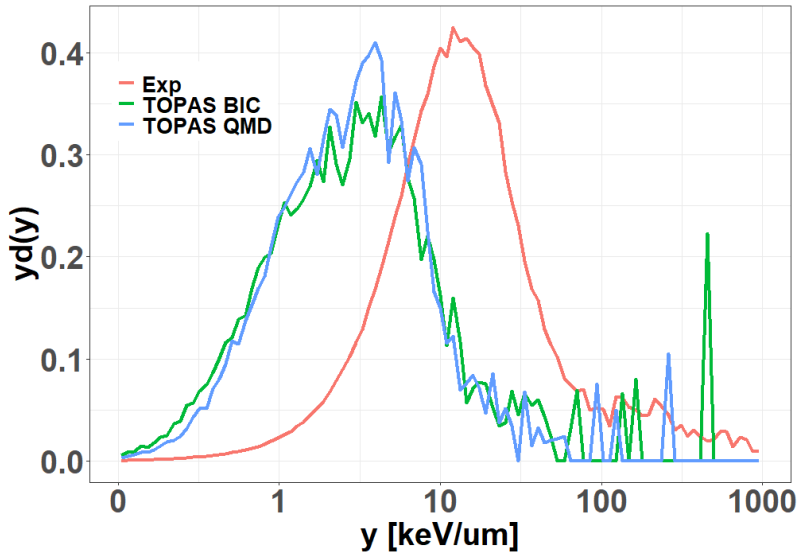

Figure 7.5: An example of inaccurate description of the experimental microdosimetric distribution in the *beam-edge* region, both using Quantum Molecular Dynamics (QMD) and Binary Cascade (BIC) hadronic interaction descriptions. The experimental and the simulated spectra have been measured at 100 mm depth in water.

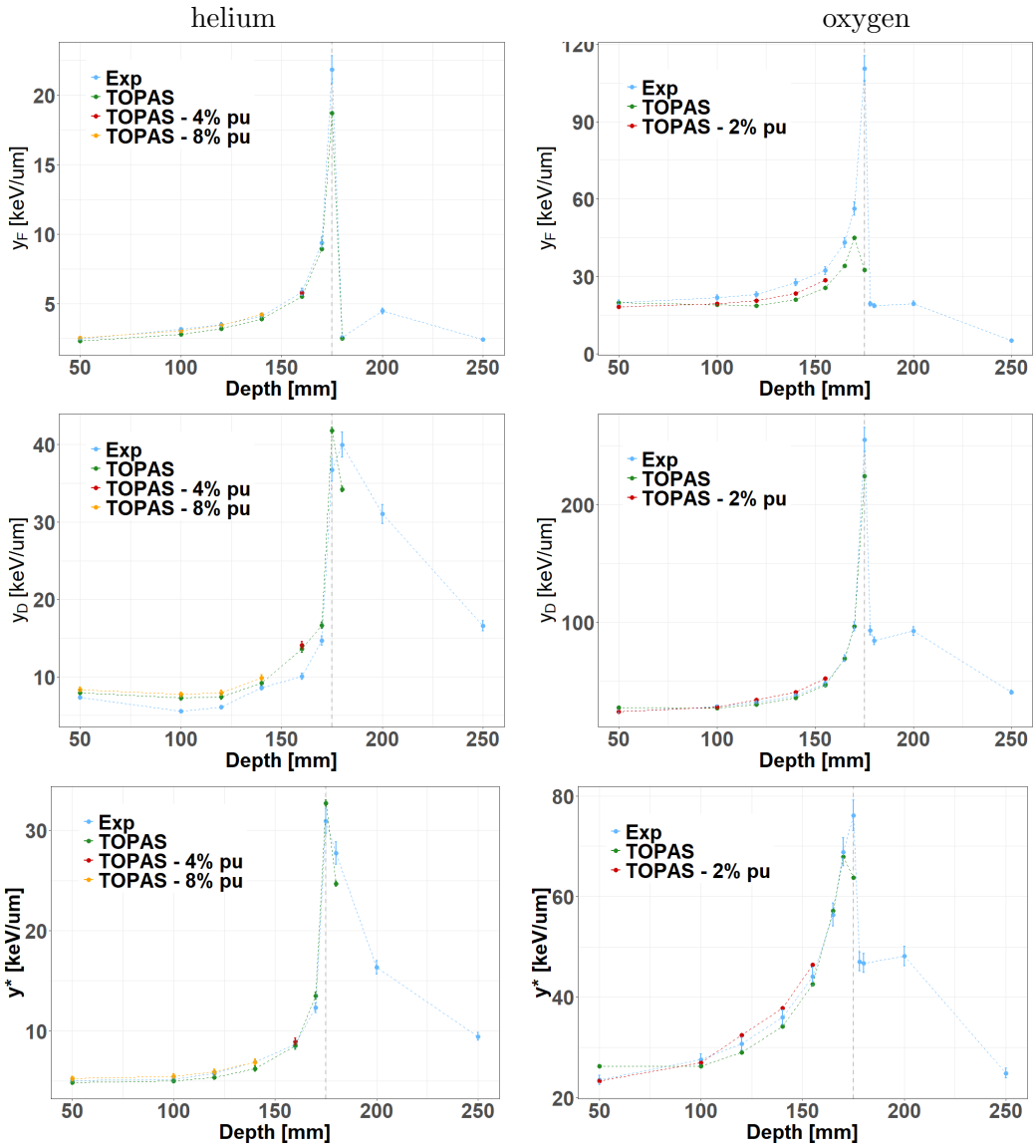

Figure 7.6: Experimental (Exp) and simulated (TOPAS) frequency-mean lineal energy y_F (Equations. (1.1.3)), dose-mean lineal energy y_D (Equations (1.1.3)) and saturation-corrected dose-mean lineal energy y^* (Equations (2.45)) plotted as a function of the depth in water for helium (reported in the legend), measured along the beam direction. Simulated values have been plotted only for good agreement between experiments and simulations (from 50 to 180 mm for helium and fom 50 to 175 mm for oxygen ions.)

the y_D maximum from the tail to the Bragg peak because only the fragments can reach a y greater than y_0. For oxygen, instead, the correction has the main consequence of drastically reducing the y_D at the Bragg peak, where the highest y can be achieved.

In Figure 7.7, the same microdosimetric quantities profiles have been plotted at the *beam-edge* (top panels), near out-of-field (central panels) and far out-of-field (bottom panels) for helium (left panels) and for oxygen (right panels).

At the *beam-edge* region, helium shows a decrease of all microdosimetric quantities as a function of depth, whereas oxygen is rather flat, with only a small increase of the y_D values compared to the Bragg peak depth.
As shown by the y spectra of Figure 7.2, helium ions suffer significant lateral scattering even close to the entrance channel and contribute heavily to the *beam-edge* spectrum. The fragments build-up increases at increasing depth, their y is on average lower than the primaries and thus y_F, y_D and y^* tend to decrease. On the other hand, the *beam-edge* for oxygen is mostly populated by fragments, as the primary ions remain focused. The fragments energy decreases with increasing water depth, and thus their y gets higher, increasing the overall mean microdosimetric values. The radiation quality near and far out-of-field is then rather constant for both ions. These regions are mostly populated by light ions, whose lineal energy appear to be overall uniform.

Absorbed dose

The absorbed dose can be estimated using Equation (6.1). The values for both ions as a function of water depth are plotted in Figure 7.8. In order to compare the results, all measurements have been normalized to deliver a dose of 2 Gy at the Bragg peak position, which are located at a depth of 175 mm, in agreement with the theoretical calculations. In addition, to better visualize the dose profiles at different lateral positions, the y axis has been plotted in a logarithmic scale.

At the selected energies, the dose delivered at the entrance channel is around 0.5 Gy/2Gy-BP for both ions.

In the tail region, the helium curve drops from $\sim 10^{-2}$ Gy/2Gy-BP 15 mm after the peak to 10^{-4} Gy/2Gy-BP at the deepest measured position (250 mm). Oxygen, instead has a much more pronounced tail due to the secondary fragments. The dose stays around 10^{-1} Gy/2Gy-BP up to 20 mm from the peak and decrease only to 10^{-2} Gy/2Gy-BP at the furthest depth (250 mm).

The dose at the *beam-edge* is in the order of 10^{-4} Gy/2Gy-BP for helium and 10^{-3} Gy/2Gy-BP for oxygen, and for both ions it drops to 10^{-5} Gy/2Gy-BP far out-of-field. At increasing lateral distances, the contribution from heavier fragments decrease, either because they range out or because they are not produced at such a large angle, and the only contributors to the dose are light particles, which are common for both primary ions, thus leading to a similar radiation quality, even if the doses differ.

7.3 Discussion and Conclusions

In-beam, both ions show a pretty well defined peak, whose center shifts towards higher y at increasing water depth. The maximum value is reached at the Bragg peak, after which the spectra move back to lower y and get broader. The main peaks of the $yd(y)$ distributions are mostly populated by the primary beam while fragments fill up the rest of

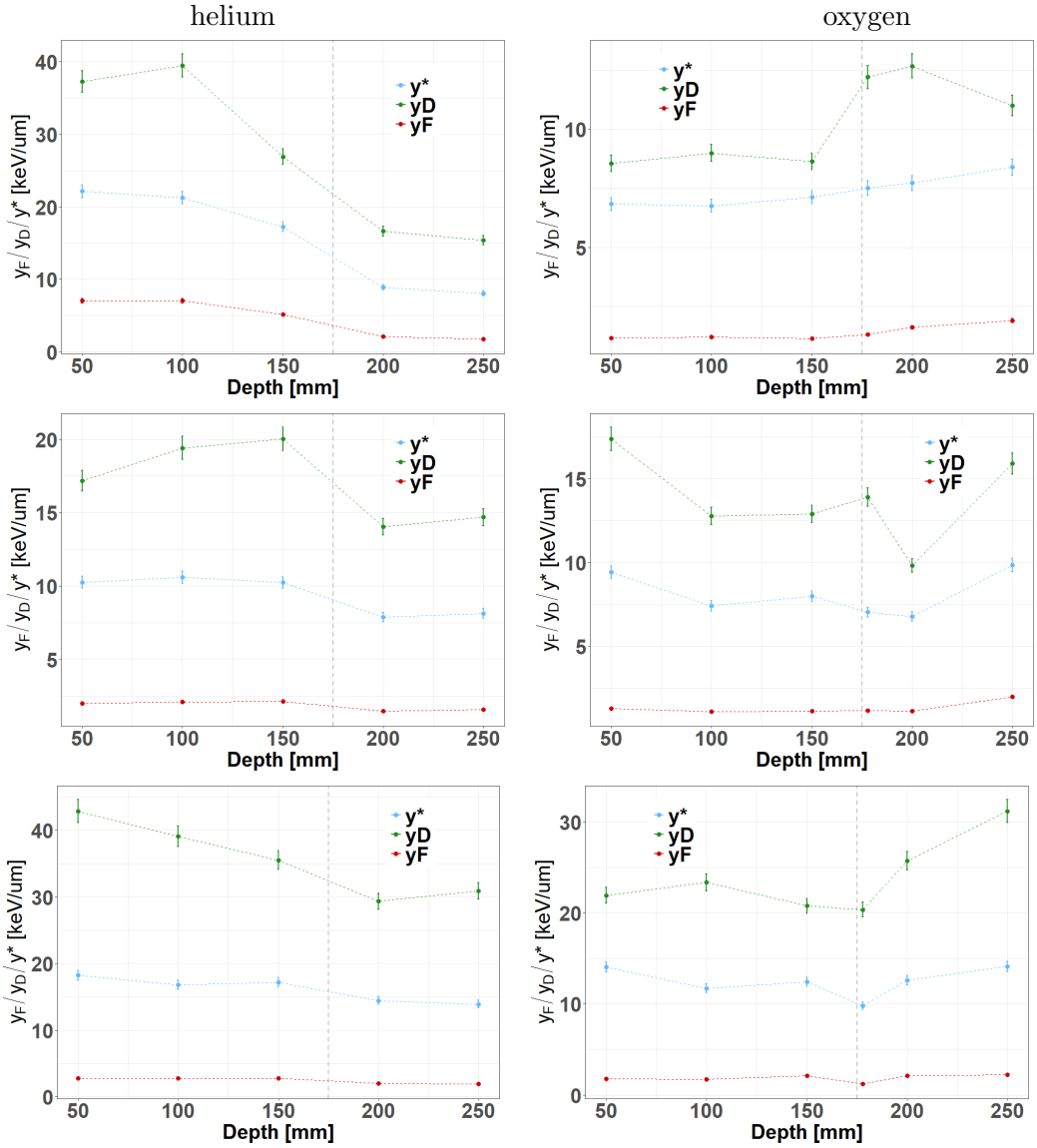

Figure 7.7: Frequency-mean lineal energy y_F (Equation (1.1.3)), dose-mean lineal energy y_D (Equation (1.1.3)) and saturation-corrected dose-mean lineal energy y^* (Equation (2.45)) plotted as a function of the depth in water for helium at different lateral positions: *beam-edge* (top panel), out-of-field (middle panel) and far out-of-field (bottom panel).

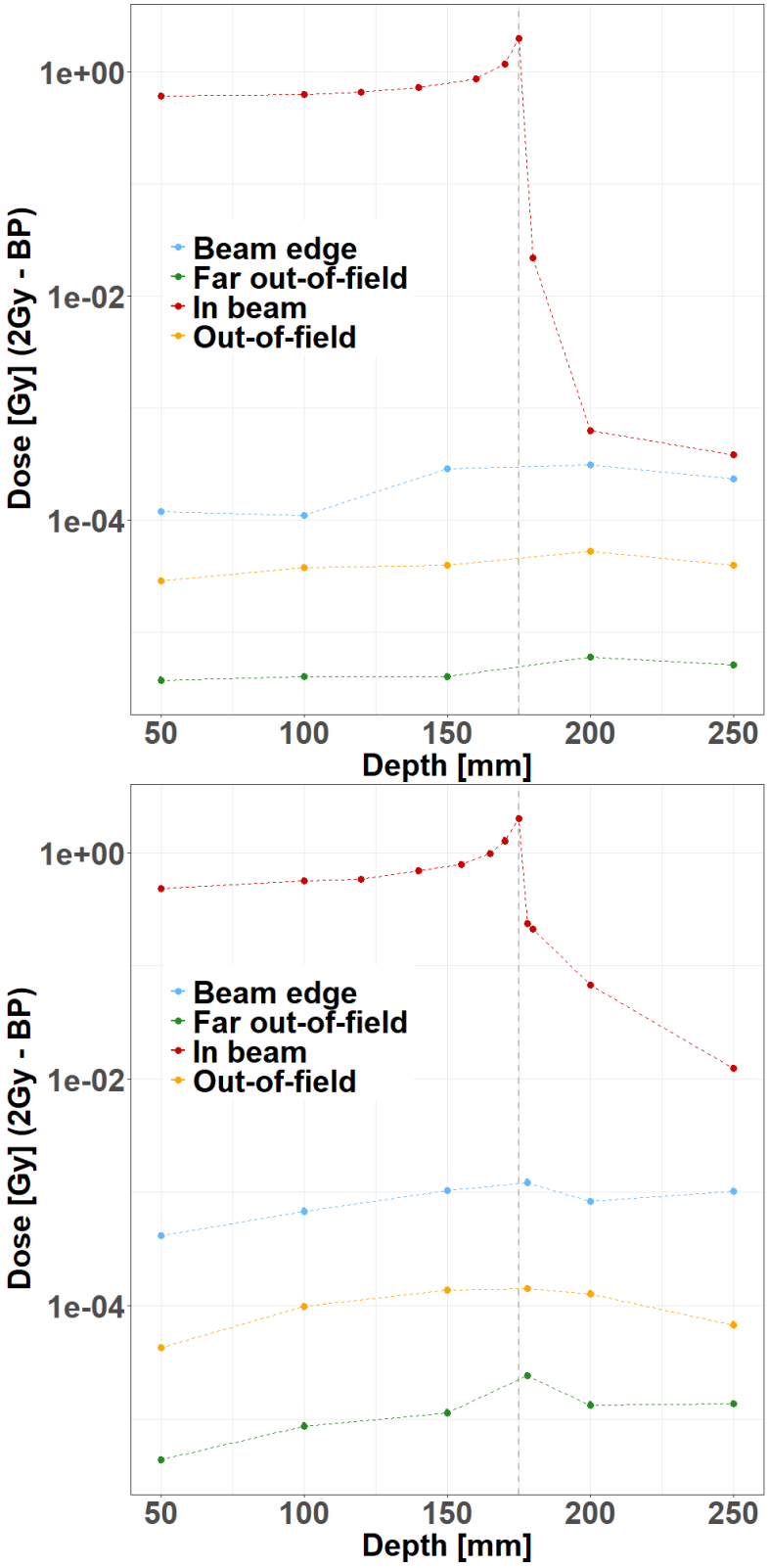

Figure 7.8: Dose profiles as a function of water depth calculated according to Equation (6.1) for helium (top panel) and oxygen (bottom panel). All values have been normalized to deliver 2 Gy to the Bragg peak position (2Gy - BP). The off-beam regions are defined as *beam-edge* (20 mm from the beam axis) and far out-of-field (100 mm from the beam axis). The dotted line marks the Bragg peak depth.

the spectrum. The peaks measured with oxygen *in-beam* are always narrower than those measured with helium, due to the fact that the latter suffers more energy straggling.

The trend of the *beam-edge* spectra measured with helium is different than oxygen in the entrance channel. The former undergoes more lateral scattering, and hence more primary ions reach this region. This hypothesis is supported by the fact that the spectra are very similar to those measured *in-beam*. Oxygen, instead, remains very focused and the *beam-edge* is mostly populated by secondary particles at all depths.

In the far-out-of-field region, both helium and oxygen distributions are rather flat (helium is still a bit more peaked than oxygen) and cover the whole y range. Here only light secondary ions and neutrons contribute to the spectrum. The absorbed doses reflects this behaviour, but the values observed for oxygen are on average a factor 5 higher then helium, due to the fact that oxygen produce more fragments. Even if using oxygen for clinical application is of more concern for potential toxicity both in the distal and in the lateral out-of-field regions, showing higher doses compared to lighter ions (protons, helium and carbon ions), the rational for using ions heavier than carbon in radiotherapy is motivated by hypoxic tumours. In fact, at increasing size of the central hypoxic region of the tumour, the benefits deriving from the usage of oxygen ion start to be clear and may become the best compromise between treatment efficacy and possible side effects in healthy tissue [Sokol et al., 2017]. Although the physical characterization of the radiation field is a key ingredient for assessing the radiobiological damage, it is not the end of the story. Future work will include an RBE estimation and a more comprehensive study for both helium and oxygen beams, to find a more direct link between the physical characteristics of the radiation field and the biological effect both inside and outside the target region. In addition, a more detailed study of the different phenomena concurring to the radiation field in each measured position will be done, exploiting also the Monte Carlo simulation when the agreement with data is good, being an invaluable tool to assess the different particles contributions to the final distribution.

Chapter 8

Neutron microdosimetric characterization for space radioprotection application

Within the ESA *AO-2019-IBER* funded project entitled *A novel approach to physical and biological countermeasures against radiation for deep space exploration* (2020-present), we performed neutron microdosimetric measurements for space radioprotection applications [Horst et al., 2022].
Space agencies have in fact recognized the risks of astronauts' exposure to space radiation and are developing complex model-based risk mitigation strategies. All current models have a limited accuracy in predicting the yield of neutrons and light ion produced by nuclear fragmentation of heavy ions, which are an important component of galactic cosmic radiation (GCR). A research collaboration has been set up to characterize the secondary radiation field produced by GCR-like radiation produced by a particle accelerator in thick shielding. The aim is to develop a novel method for producing high-quality experimental data of neutrons and light ions production in shielding materials relevant for space radiation protection. These kind of experiments are performed using a ground-based particle accelerator; in this case in particular, they have been performed at the heavy ion synchrotron SIS18 at the GSI Helmholtzzentrum für Schwerionenforschung in Darmstadt, Germany, in Cave A.

Microdosimetry in particular, and especially TEPC detectors, are invaluable tools to characterize also the radiation field of the complex radiation environments in space. For example, TEPCs have been placed both on space shuttles and on the International Space Station (ISS), to serve as the primary active dosimeters in space for nearly two decades [Braby, 2015].

In this work, we employed the TEPC LET 1/2 1.2.1 to characterize the radiation field of neutrons produced when ^{56}Fe ions is stopped in an aluminum target and we assessed the quality factor in two different out-of-field positions.

8.1 Experimental setup

The full setup used during the experimental campaign can be found in Figure 8.1. The primary 1 GeV/u ^{56}Fe ion beam exits the vacuum beam line through a thin exit window, and traverses a beam monitor ionization chamber. Optionally, for low intensity

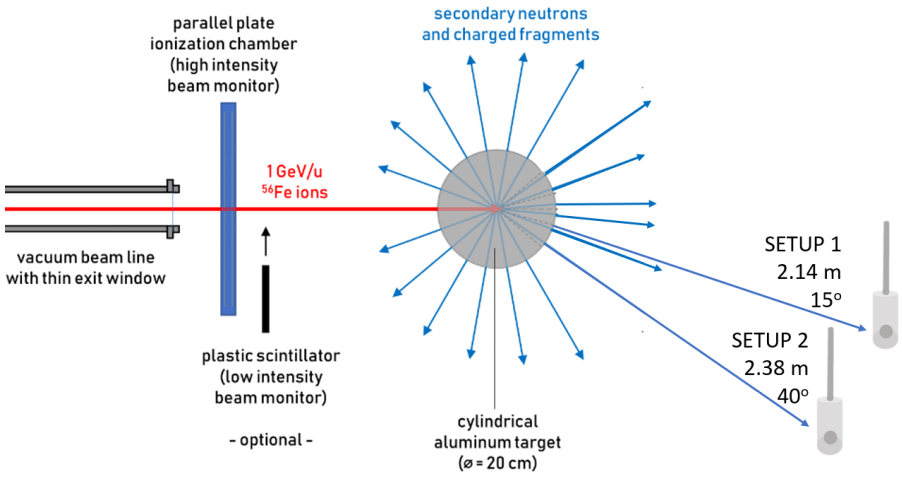

Figure 8.1: Scheme of the characterization of the secondary neutron and light ion radiation field produced by a high-energy heavy ion beam in a thick aluminum target.

measurements, also a plastic scintillator can be inserted. The beam intensity is adjusted to match the optimal measurement conditions of each detector system. After passing the beam monitors, the primary ions are stopped in a thick cylindrical aluminum target (20 cm width), producing different secondary particle species which the most abundant are neutrons and protons.
Focusing on microdosimetry experiments, they were conducted using two different setups: the first one (experimental setup 1) is characterized by a 15° angle from the beam axes and a distance of 2.14 m from the aluminium target; the second experimental setup (experimental setup 2) is instead characterized by a 40° angle from the beam axes and a distance of 2.38 m from the aluminium target.

8.1.1 Results

During the data taking, few channel counts of the high-gain ADC suddenly jumped to unrealistically high values as shown in Figure 8.2. This phenomenon can occur when the ADCs electronics is directly exposed to radiation, that causes single event upsets [O'Bryan et al., 2000]. To correct for this effect, we performed a linear fit to predict the counts of the affected channels. By defining an interval around the peak, and removing the peak itself, it was possible to recover the real energy deposition distributions. This method was applied in both setup 1 and 2 and the first setup is shown in Figure as an example 8.2

Figures 8.3 and 8.4 show the microdosimetric $yf(y)$ and $yd(y)$ spectra acquired with the TEPC placed at 15° with respect to the beam axis (setup 1). A total of 2023495 events were acquired.
Figures 8.5 and 8.6 reports instead the results acquired with the setup 2 configuration, where the TEPC was placed at 40° lateral to the beam axis. A total of 1005389 events were acquired in this case.
The microdosimetric distributions $yd(y)$ measured with the two setups show some similarities, both peaking around 3 keV/μm and demonstrating a large tail on the higher lineal energies (from $\sim$ 10 keV/mum to $\sim$ 1000 keV/mum).

A peak can be observed in the $yd(y)$ plot around 6 keV/μm. This peak was also present in the background noise spectrum acquired with the iron beam off, i.e. the spectrum ac-

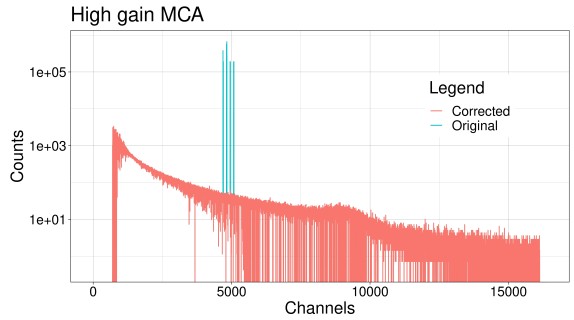

Figure 8.2: (a) Example of the unrealistic peak in the raw spectrum of the high-gain ADC channels for the experimental setup 1, probably due to single event upsets.

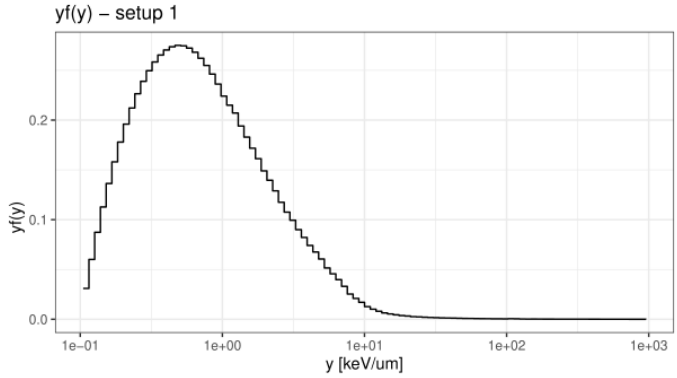

Figure 8.3: $yf(y)$ spectrum measured in the experimental setup at 40° lateral to the beam axes.

quired with the beam off. For this reason, we hypothesized that this peak is generated by the activation of material in the experimental room. The lower statistics of total events collected in the second setup geometry has therefore highlights this activation peak in this configuration, since the proportion between counts of the activation peak and the total counts due to neutrons from the iron beam is higher.

The y_F, y_D, y^* and Q are reported for both setup in Table 8.1

Experimental setup	y_F [keV/µm]	y_D [keV/µm]	y^* [keV/µm]	Q
1	1.37 ± 0.06	25.01 ± 1.01	10.63 ± 0.42	3.02 ± 0.15
2	1.51 ± 0.06	26.56 ± 1.18	11.66 ± 0.46	3.53 ± 0.17

Table 8.1: Microdosimetric y_F, y_D and y^* measured with the two experimental setups of Figure 8.1.

The microdosimetric mean values y_F, y_D and y^* measured at the two positions are very close, with values slightly higher for the larger angle. This result indicates that the radiation field quality at the two positions is similar. Instead, the quality factors exhibit a greater discrepancy, with a relative difference of about 15%. Unlike the microdosimetric mean values, the quality factors are calculated using the whole microdosimetric spectra (2.3.2).

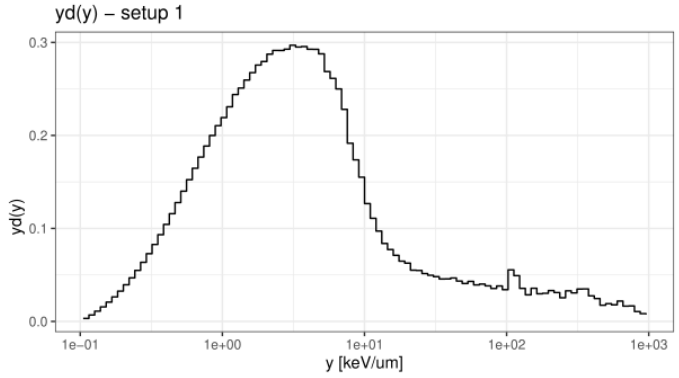

Figure 8.4: $yd(y)$ spectrum measured in the experimental setup at 40° lateral to the beam axes.

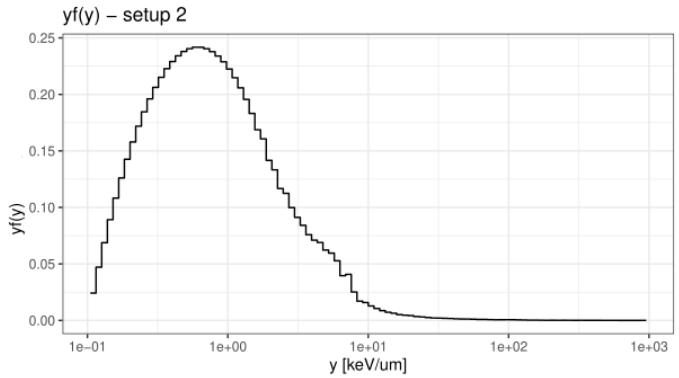

Figure 8.5: $yf(y)$ spectrum measured with the experimental setup 2 at 40° lateral to the beam axes.

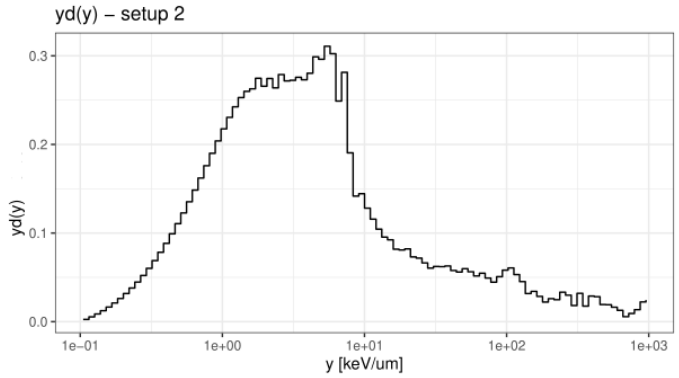

Figure 8.6: $yd(y)$ spectrum measured with the experimental setup 2 at 40° lateral to the beam axes.

At higher angles, neutrons are produced with a lower kinetic energy, thus depositing more in the TEPC; this reflects into higher values of microdosimetric lineal energies and a slightly more higher quality factor Q in the setup 2.

Part IV

Hybrid Detector for Microdosimetry (HDM): a new tool for extending the microdosimetric information

Chapter 9

Hybrid Detector for Microdosimetry (HDM)

A limitation shared by all microdosimeters is that while ϵ is directly measured, the value of $\bar{l}$ has to be theoretically estimated as the mean path travelled by a particle inside the detector, and thus it depends on the detector geometry. In addition, $\bar{l}$ values calculated for standard geometries can be used only if the microdosimeter is exposed to a homogeneous and isotropic field (also called uniform isotropic randomness) [Kellerer et al., 1985] and a different $\bar{l}$ value will be obtained under different irradiation conditions, i.e. for other types of randomness. Some attempts have been made to overcome the limited accuracy of the the mean chord length concept, e.g. in unidirectional particle field when the isotropicity assumption drops. An example is SOI microdosimeters [Bolst et al., 2017], which are composed of 3D sensitive volumes (SVs) arrays with a well defined thickness, and thus path length. Furthermore, few theoretical studies focused on finding a formula of the mean path length for both uniform [Cruz et al., 2001] and non uniform [Santa Cruz et al., 2001] radiation fields. So far, only the calculation for a uniform isotropic randomness could be successfully applied to experimental methodologies. Estimating the path length l is a critical parameter in microdosimetry that will influence the accuracy of the radiation field quality characterization [Abolfath et al., 2020]. In fact, for a given energy ϵ deposited in the detector, the resulting y value can assume a wide range of values depending on the l. For example, if ϵ=10 keV in a 2 μm diameter sphere made of tissue, y can varies from 5 keV/μm to 1000 keV/μm just considering l values ranging from the sphere diameter to 0.01 μm.

For this reason, since the quantity y is traditionally intended as the ϵ over the mean chord length value $\bar{l}$, we introduce a new quantity y_T, defined as ϵ divided by the particle real track length l.

Hybrid Detector for Microdosimetry (HDM) is indeed designed to measure the y_T. This detector have been specifically intended for particle therapy application, where a knowledge of the y_T yields a more direct link to the biological damage.
Together with providing a direct measurement of the track length l, this design also improve the lateral spatial resolution of existing TEPCs. HDM is composed of a spherical TEPC followed by four layers of Low Gain Avalanche Detectors (LGADs) [Pellegrini et al., 2014]. LGAD is a recent technology in silicon systems featuring detection of particles in a wide energy range with improved accuracy for timing and tracking measurements [Pellegrini et al., 2014]. The LGAD application in particle therapy has been also recently investigated [Vignati et al., 2017]. In the proposed setup, the TEPC will provide the

energy deposition ϵ directly in a tissue-equivalent medium while the LGADs will offer information about particle spatial distribution with a precision of about 200 or 300 μm, depending on the chosen configuration.

Details of the detector components, geometrical configurations as well as read-out solutions are illustrated here. Using GEANT4 toolkit, we investigated HDM performances when exposed to protons and carbon ions in the therapeutic energy range. The influence on all microdosimetric quantities when the real l is used instead of the mean track length approximation is discussed. Detection efficiency and tracking precision are also reported.

9.1 HDM feasibility study

Together with the TEPC that provides energy depositions, LGADs constitute the tracker stage of HDM. An introduction on these silicon detectors is provided here.

LGAD

LGAD is a recent technology in silicon detection system. It was first fabricated at CNM-IMB clean room facilities by diffusing a p-type layer just below the n+ electrode [Pellegrini et al., 2014]. From then it has been used for particle timing and tracking and, more recently, its application in radiotherapy has been explored [Vignati et al., 2017]. LGADs, using n-in-p silicon diodes, differ from standard Avalanche Photodiodes (APDs) due to their low and controlled internal multiplication mechanism for detecting charged particles. This technique allows also to produce thinner sensors with the same output signal of standard thick substrates.

The main features of the LGADs used for the HDM prototype can be found in [Sola et al., 2019]. In particular, the active region is 50 μm thick while the substrate is 300 μm and can be thinned down to 100 μm postproduction.

An additional LGAD production for HDM is under development at FBK and will include sensors with alternative geometries and active layer doping in order to obtain different spatial resolutions and gains.

A constraint on the detector geometry is that the optimal active area of one strip is $\sim$ 2 mm^2, which correspond to 5pF of capacitance; in fact, the read-out chip have been designed for this value. For what concerns the 71-strips configuration, the area is already optimized, while the 34-strips configuration has a larger area. However, previous experiments with similar area have shown the feasibility of this capacitance also.

Furthermore, the dead area between two strips must be 66 μm wide independently of the strip width. Thus, narrower strips result into a higher spatial resolution but also a decreased detection efficiency due to a larger dead area and a resulting lower fill factor. In addition, to cover the same area more strips are needed, which translates into a larger number of channels to be read-out.

To find the optimal detector geometry for our application, we simulated three configurations: i) 34 strips, each 294 μm wide and 12.5 mm high (sensor height 13.8 mm and width 13.4 mm); ii) 71 strips, each 114 μm wide and 12.5 mm high (sensor width 14 mm and height 13.8 mm) and iii) 288 strips, each 114 μm wide and 50.22 mm high (sensor height 51.52 mm and width 51.84 mm). An image of the design project of this configuration of the complete sensor is given in Figure 9.1 (left panel). While the first two configurations are now being produced, configuration (iii) is not currently feasible and

was tested to investigate the tracking efficiency for a larger detector with the same spatial resolution of the 71 strips detector (ii).

9.1.1 HMD geometry

The LGAD position with respect to the TEPC determine the detector performances and the optimal configuration depends on the goal of the specific measurement. In this work, we investigated the configuration with the TEPC upstream of the 4 LGAD layers. This setup has been chosen because we wanted to characterize the radiation field with standard microdosimetric measurements, without possible artifacts due to the LGADs in front. The distances between the detectors can be found in Figure9.1 (right panel). In particular, the first LGAD have been placed as close as possible to the TEPC to minimize lateral scattering and energy loss of particles exiting the microdosimeter.

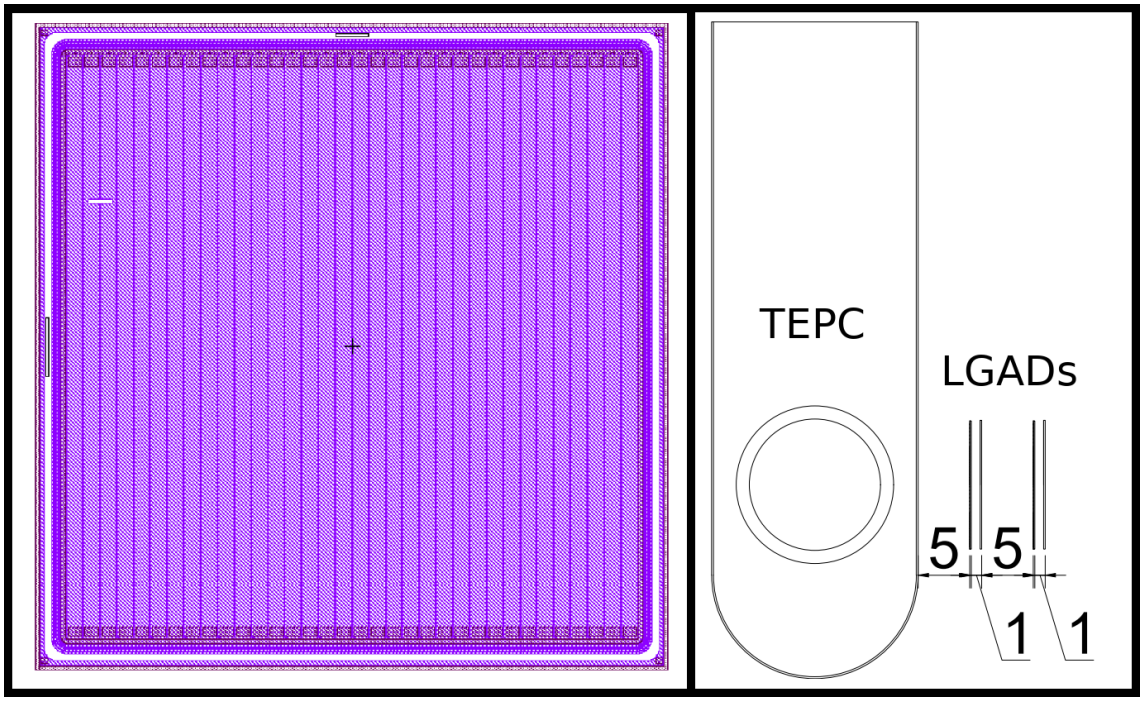

Figure 9.1: Left panel: design of one LGAD sensor with 34 active strips. Right panel: Scheme of the HDM setup, showing the TEPC followed by four LGAD layers. Distances between detectors are reported in millimeters.

9.1.2 Geant4 simulations of the HDM detector

To investigate the detector performances, we run Monte Carlo calculations using Geant4 toolkit (5.2). As the HDM design is optimized for applications in particle therapy, we focused the study on the response to protons and carbon ions at therapeutic energies. All calculations were run to acquired a minimum of 10^6 events on the TEPC, which is considered an adequate statistics for experimental measurements [Missiaggia et al., 2020].

In addition, since microdosimetry deals with patterns of single energy deposition in tissue at the micrometer scale, we computed the energy deposition ϵ of a particle traversing the TEPC as the sum of the energy deposited by the primary event and all the related secondary particles that entered the detector.

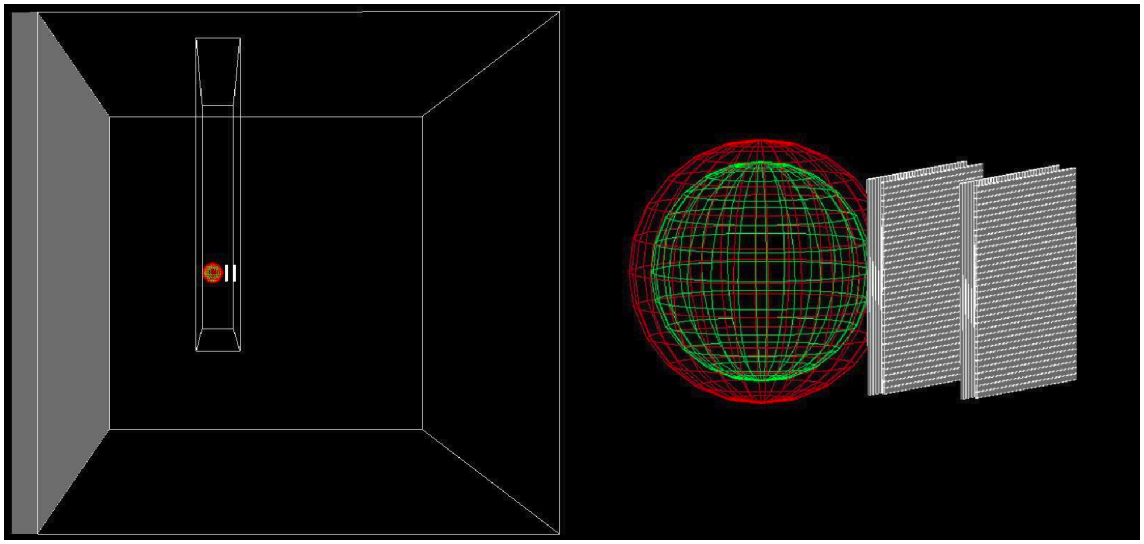

Figure 9.2: 3D scheme of the geometry used for all Geant4 simulations. Both the TEPC and the four 24-strips LGADs are contained in PMMA box filled with air. The box is placed inside a water phantom, whose walls are made of PMMA. A broader view is show in left panel, while a zoom on HDM is illustrated in right panel.

The simulation geometry consisted of a water phantom with PMMA walls (1.74 cm water equivalent thickness) where the hybrid system was placed. To reproduce a realistic setup, HDM was contained in an additional air box 2.8×20×2.8 cm thick. A 3D view of the setup is shown in Figure 9.2.

The water phantom was irradiated with 290 MeV/u carbon ions and 150 MeV protons, which have the same range in water ($\sim$ 160 mm). The beam spots were circular with a 3 cm radius to ensure that the detectors were fully immersed in a homogeneous and isotropic radiation field. The detector box was placed at 10.74 cm in water along the beam direction. This depth represented a good compromise to assess HDM performances in a relatively mixed field in terms of particle species and energies, but upstream of the Bragg peak, where most particles have a low energy and thus might stop inside the TEPC.

9.1.3 Tracking

Tracking algorithm

To measure a particle track, the LGADs were positioned to have the strips in different directions, two horizontals (x plane) and two verticals (y plane). By coupling two sensors with different orientations, a spatial position for a particle can be measured. Thus, two pairs of sensors are the minimum requirement for reconstructing a particle track. To reproduce a realistic experimental scenario, in the simulation we scored only the position of the strip hit by the particle. Then, we used a lineal interpolation to reconstruct the particle path inside the TEPC, from which we could estimate the real track length.

Tracking efficiency

Using Geant4 simulations, we studied the HDM tracking efficiency. As a first step, we focused on identifying the lost events and divided them into three categories:

1. particles that reach all the detectors, but traverse an inter-strip dead zone in at least one of the LGADs;

2. particles that range out before reaching the fourth LGAD;

3. particles that undergo lateral scattering and are deflected outside the solid angle covered by all detectors.

Category 1 is related to the probability to hit a dead region and thus depends on the LGAD geometry. Assuming a uniform radiation field, the probability to reach an active strip is given by A_{act}/A_{tot}, where A_{act} is the total area covered by active strips and A_{tot} the total area of the sensor, including both active strips and dead inter-strips. As the probabilities of hitting the active region of two sensors are independent, the overall probability of the joint event is the product of the single probabilities. To test the validity of these assumptions, in the simulation we also scored the particles traversing the inter-strip regions.

For category 2, we investigated the minimum detectable kinetic energy for each ion type, i.e. the minimum energy that a particle must have to pass through all detectors. The values for all particle species of interest have been estimated with LISE++ toolkit version 10.0.6a [Tarasov and Bazin, 2008]. These kinetic energy cutoffs depends only electromagnetic interactions in the detector layers and do not take into account additional losses due to multiple Coulomb scattering (MCS). To estimate a realistic kinetic energy detection threshold, we performed simulations of HDM exposed to a given particle species and decreases the initial energy until we found the minimum value required to traverse all detectors. We then repeated the test for the ion types of most interest.

The percentage of particles deflected outside the solid angle covered by all detectors (category 3) dependent on the LGADs size. To assess this value and its dependence on the LAGDs geometry, we performed simulations for every configuration described in 9.1.

For the events seen by the TEPC and by an active zone of each of the 4 silicon layers (i.e. the trackable particles), we investigated the tracking accuracy using the algorithm described in 9.1.3. From the simulations, we could extract the real particle track and compare it to that reconstructed with the tracking algorithm, estimating a mean discrepancy between the predicted and actual values. The tests were repeated for all LGADs configurations taken into consideration.

9.1.4 Results

Radiation field characterization in the TEPC

The composition of the radiation field entering the TEPC was investigated at a depth of 10.74 cm in beam. The results include kinetic energy spectra of all particle species, track length distributions and microdosimetric spectra $yd(y)$ obtained with both the real track length and the mean chord length. The results are shown in Figs. 9.3 and 9.5 for protons and carbon ions, respectively. In detail: panels **A, B** of Figure 9.3 and panels **A, B, C, D** of Figure 9.5 illustrate the kinetic energy distributions of all particles entering the TEPC, with and without the contribution from the primary ions (in both cases the energy distributions of the single components are normalized to one); the track distributions of all the particles are plotted in panel **C** for protons and in panel **E** for carbons, with the mean chord length of 8.47 mm marked with a dashed red line; panels **D** for protons and **F** for carbons contain a comparison between the microdosimetric spectra calculated with the

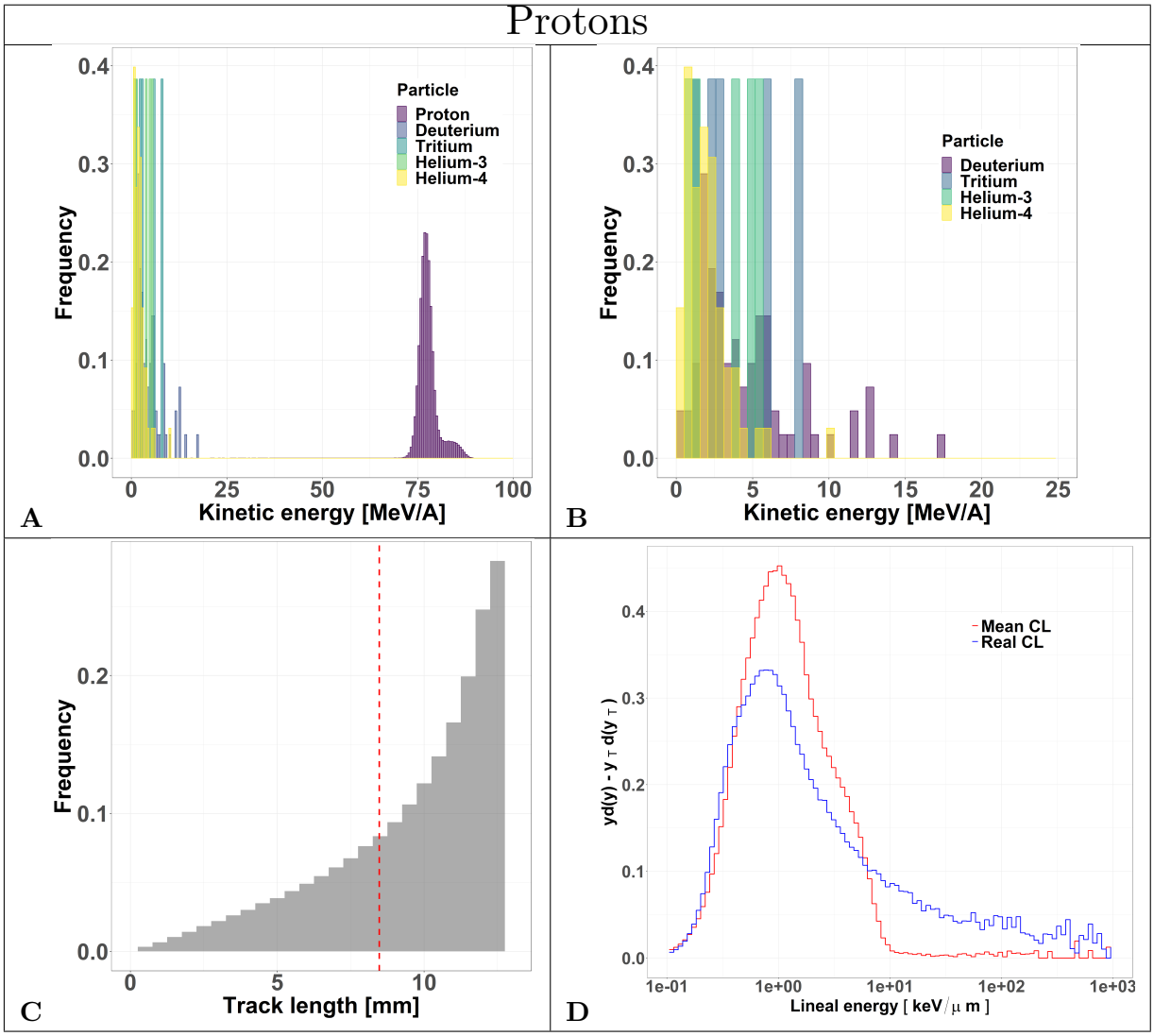

Figure 9.3: Characterization of the radiation field generated by 150 MeV protons after traversing 10.74 cm of water and seen by the TEPC. Panels **A** and **B**: kinetic energy spectra of the most abundant components of the radiation field including and excluding the primary ions. Panel **C**: track length distribution of all the particles detected by the TEPC. The mean chord length at 8.47 mm is marked with a red dotted line. Panel **D**: microdosimetric yd(y) spectra obtained with the mean chord length approximation (red line) and microdosimetric $y_T d(y_T)$ spectra obtained using the real chord length values (blue line) .

mean chord length approximation ($yd(y)$) or the real track length ($y_T d(y_T)$). Furthermore, the mean values and standard deviations of the track length distributions are also reported in Table 9.1 for both ions of interest.

Secondaries produced by protons, are mostly low-energy (below 10 MeV) and the distribution does not have a peak. For carbon ions, the energy of all fragments species peaks around 170 MeV/u, which is the residual primary beam energy (9.3, panel **A**). Protons can only generate fragments from the target nuclei, and thus their energy will be relatively low [Tommasino and Durante, 2015]. Carbon ions, instead, can produce both

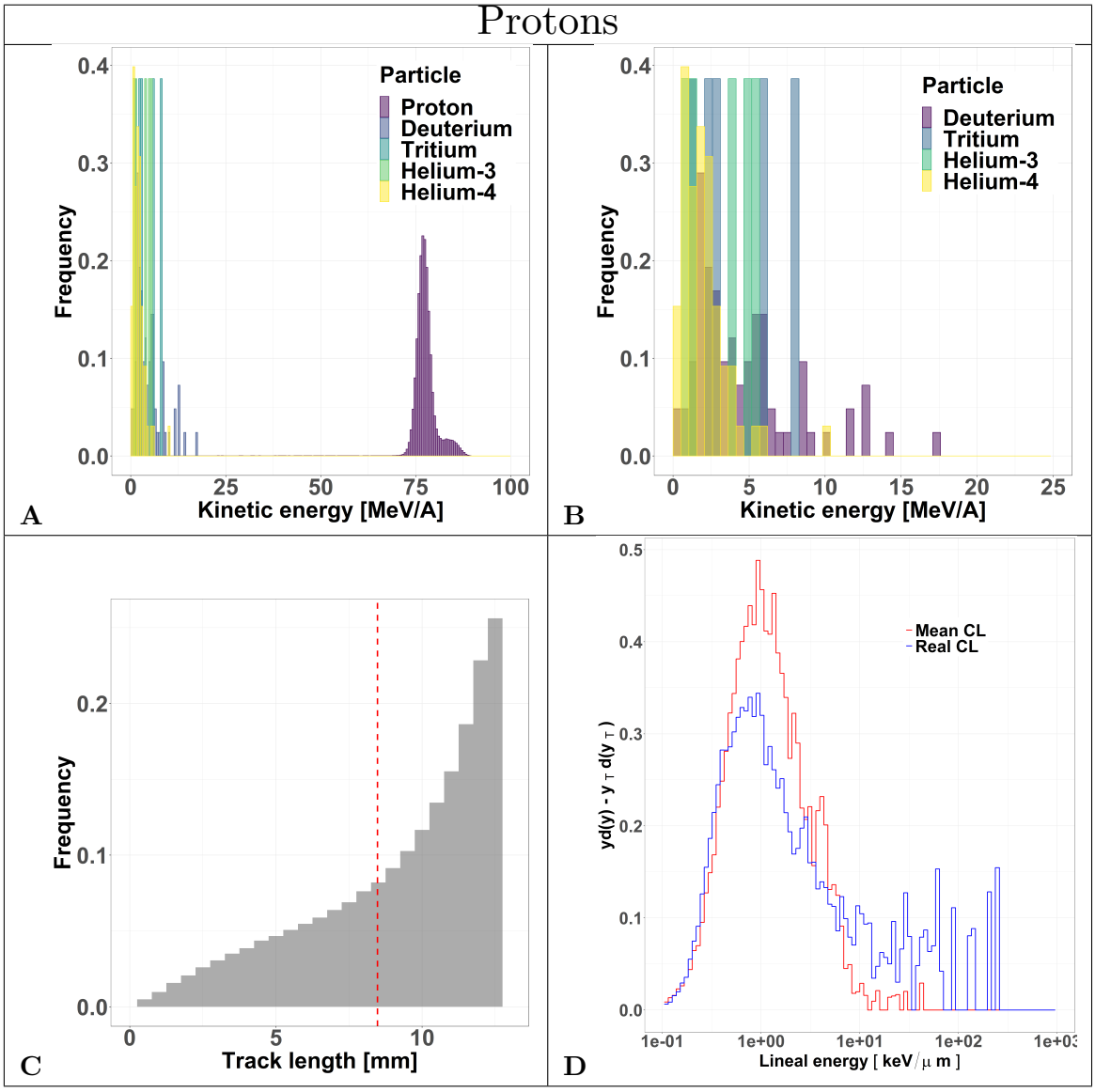

Figure 9.4: Characterization of the particles lost by HDM when irradiated with 150 MeV protons at a depth or 10.74 cm in water. Panels **A** and **B**: kinetic energy spectra of the most abundant components of the radiation field including and excluding the primary ions. Panel **C**: track length distribution of all the particles detected by the TEPC. The mean chord length at 8.47 mm is marked with a red dotted line. Panel **D**: microdosimetric yd(y) spectra obtained with the mean chord length approximation (red line) and microdosimetric $y_T d(y_T)$ spectra obtained using the real chord length values (blue line).

projectile and target fragments, whose kinetic energies have a much wider range, peaking at the same value as the primary ions [Mohamad et al., 2018, Tommasino et al., 2015].

The track length distributions of both protons and carbon ions are very broad and do not present a peak. Furthermore, the mean track length calculated for both protons and carbon ions is higher than the mean chord length, indicating that the latter does not provide an accurate description of the system. The limitation of the mean chord length approximation can be further investigated by comparing the standard microdosimetric $yd(y)$

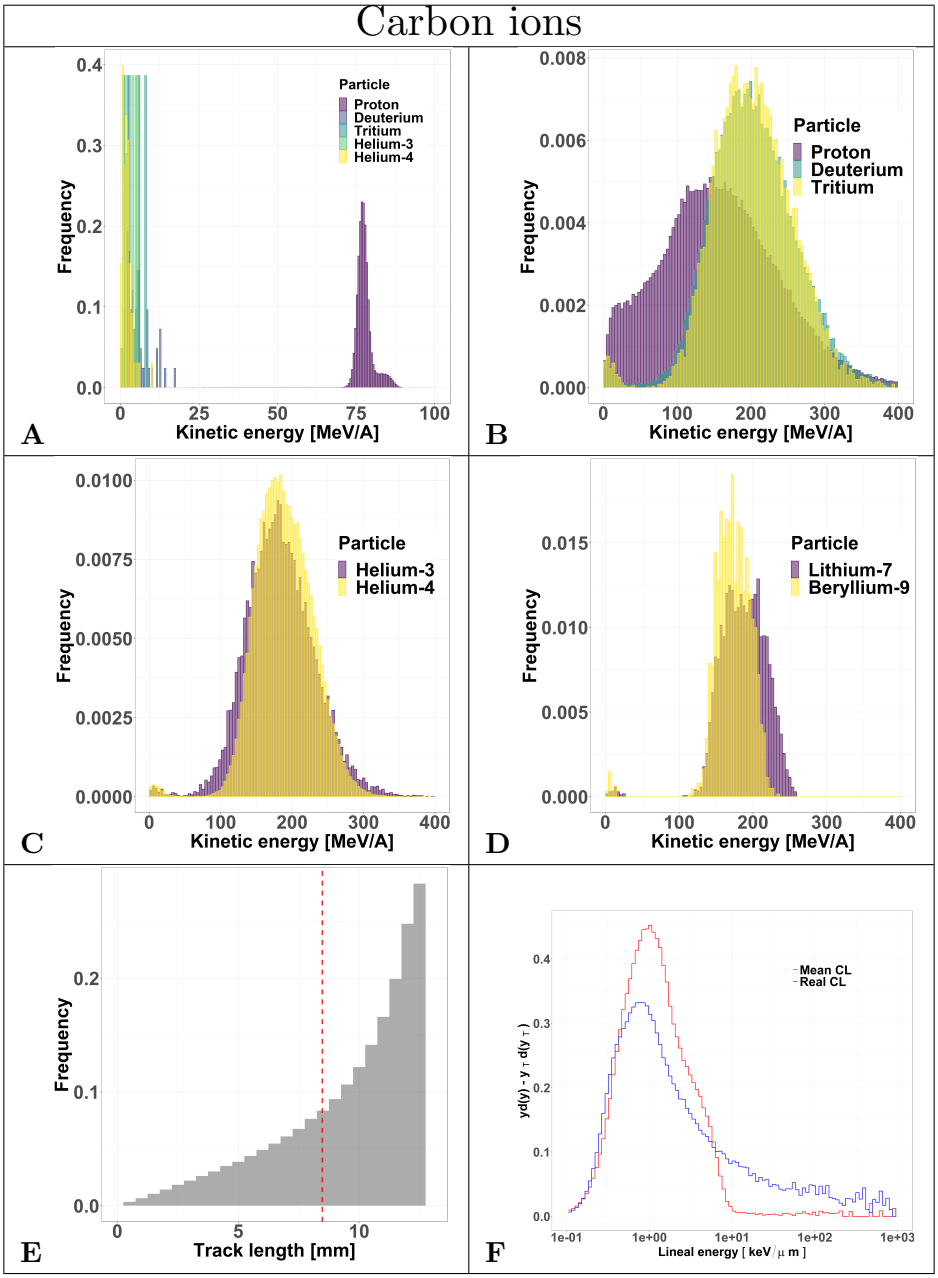

Figure 9.5: Characterization of the radiation field generated by 290 MeV/u carbon ions after traversing 10.74 cm of water and seen by the TEPC. Panels **A** to **D**: kinetic energy spectra of the most abundant components of the radiation field including and excluding the primary ions. Panel **E**: track length distribution of all the particles detected by the TEPC. The mean chord length at 8.47 mm is marked with a red dotted line. Panel **F**: microdosimetric yd(y) spectra obtained with the mean chord length approximation (red line) and microdosimetric $y_T d(y_T)$ spectra obtained using the real chord length values (blue line).

spectra with those obtained with the real track length ($y_T d(y_T)$). The latter distributions show a non negligible contribution in the high y_T region. Those contributions are due to events that deposit energy along a small chord length and they are underestimated in the

Ion	Mean track length [mm]	Standard deviation [mm]
Carbon	9.17	3.03
Proton	9.53	2.8

Table 9.1: Mean track length values and standard deviations of protons and carbon ions traversing a spherical TEPC of 12.6 mm diameter. The mean chord length of this detector exposed to a uniform and isotropic radiation field is 8.47 mm.

Ion	Configuration	Tracked particles [%]	Mean track length of tracked particles [mm]		Standard deviation [mm]		Mean absolute tracking error [mm]
			Real	Reconstructed	Real	Reconstructed	
Carbon	*34 strips*	31.4	10.10	10.09	2.42	2.43	0.38
	71 strips	12.1	9.99	10.00	2.53	2.52	0.20
	288 strips	14.6	9.53	9.55	2.81	2.79	0.25
Proton	*34 strips*	45.8	10.06	9.89	2.37	2.50	0.91
	71 strips	15.3	9.91	9.91	2.46	2.47	0.24
	288 strips	16.6	9.63	9.64	2.68	2.67	0.28

Table 9.2: Percentage of particles tracked by HDM, including their mean track length, standard deviations and the absolute values of the mean tracking error of the algorithm with respect to the actual value. The results are reported for both protons and carbon ions and for three LGAD configurations (34, 71 and 288 strips).

$yd(y)$ spectra where the mean chord length value is used. These events have a very high y_T and thus can be relevant for radiobiological effects, especially for y up to 150 keV/μm (overkill effect).

Particles tracked by HDM

We investigated HDM tracking efficiency as well as the characteristics of the tracked events. Tab. 9.2 illustrates for carbon ions and protons the percentage of particles tracked by HDM, their mean track length values, their standard deviations and the average discrepancy between the reconstructed and the real track length. The latter values are reported for the three sensor geometries (34, 71 and 288 strips) described in 9.1.

The results show that, as expected, the 71 strips configuration collects the least amount of events because of the reduced fill factor. Increasing the sensor dimension while keeping the same fill factor increases the number of collected events (288 strips configuration). The mean track length and standard deviation obtained with the tracking algorithm are in good agreement with the real values obtained directly from the simulation. This is confirmed also by the small values of the mean absolute error, defined as the average absolute value of the difference between the real track length and the reconstructed one.

The accuracy of the reconstructed tracks in the three sensor configurations (34, 71

and 288 strips) was further studied in Figs. 9.6 and 9.7 for protons and carbon ions, respectively. We compared the track length distribution obtained directly from Geant4 with that reconstructed with the algorithm. The data are presented as density color plots in panels **A, C** and **E**; the green dotted line marks a perfect prediction of the algorithm, the red and blue colors represent regions of high and low events density, respectively. The distributions have a cone-like shape, implying a better accuracy of the reconstructed tracks of large lengths. This result is further supported by the presence of high density regions around the green line in the large track lengths zones.

To further assess the accuracy of the tracking algorithm, in panels **B, D** and **F** we compared the track distributions of all particles traversing the TEPC with those detected by HDM and either obtained directly from the simulation or estimated with the tracking algorithm. Independently of the primary ion type, the 34 and 71 strips configurations systematically underestimate the distributions for small tracks. On the contrary, the 288 strips configurations provide a more accurate estimation of the whole track distributions, especially for protons.

The track distributions obtained with the three configurations were used to calculate microdosimetric $yd(y)$ and $y_T d(y_T)$ spectra for all particles tracked by HDM. The results are shown in Figs. 9.8 and 9.9 for protons and carbon ions respectively. Results show that the $yd(y)$ spectra differ from the $y_T d(y_T)$ ones, with a peak value shifted to the right in all cases. On the contrary, the $y_T d(y_T)$ distributions obtained with the real track length and with the reconstructed track length are similar mostly in the bell shape regions. It can be nonetheless seen that they differ in the tails due to the higher discrepancy between the real track length and the reconstructed ones for small track lengths. The accuracy between the two increases from the first sensor configuration (panels **A**) to the last one (panels **C**) under both radiation fields.

Particles lost by HDM

As discussed in Section 9.1.3, we can group lost particles into three categories: i) particles with a kinetic energy under the minimum required to traverse all the detectors, ii) particles lost due to MCS and iii) particles that reach all detectors, but cross an inter-strip in at least one LGAD.

The minimum kinetic energies necessary to pass all detectors have been studied and are reported in Tab. 9.3 for all particles of interest. The values calculated with LISE++ are indicated for all particles while those obtained with Geant4 only for selected ions representative of the radiation field. The results obtained with the two methodologies agree very well for protons but have a a higher discrepancy for carbon ions.

Using Geant4 outputs, we characterized the particles lost in terms of kinetic energy when entering the TEPC and track length traversed inside the detector. The results are reported in Figures 9.4 and 9.10 for protons and carbon ions, respectively. In panels **A** and **B** of Figures 9.4 and **A**, **B**, **C** and **D** of Figure 9.10 the kinetic energy spectra of all particle types are plotted with and without the contribution from the primaries. Independently of the fragment type, the energy spectra have the same shape of those reported in Figures 9.3 and 9.5, where all events are considered. These result indicate that the probability for a particle to be lost is independent of the charge and energy (for energies above the minimum threshold reported in Table 9.3). Panels **C** of Figure 9.4 and **E** of Figure 9.10 illustrate the track distributions of lost particles, together with the mean chord length (red dotted line). The left side of the distribution appears to be more populated compared to the distribution of all events (Figures 9.3 and9.5), suggesting that there is a higher chance

	Proton [MeV/A]	**Deuterium** [MeV/A]	**Tritium** [MeV/A]	**Helium-3** [MeV/A]	**Helium-4** [MeV/A]	**Lithium-7** [MeV/A]	**Beryllium-9** [MeV/A]	**Boron-11** [MeV/A]	**Carbon-12** [MeV/A]
LISE++	17	11	8	20	17	20	24	28	34
GEANT4	17	12	9	-	17	-	-	-	37

Table 9.3: Minimum kinetic energies for several isotope types necessary to traverse all the detectors. The values have been calculated with LISE++ toolkit and, for the most representative of the radiation field, also with Geant4.

	y_D	$y_{D_T,TEPC}$	$y_{D_T,TRACKED}$
Carbon	16.53	40.49	40.79
Proton	3.75	23.73	21.08

Table 9.4: y_D and y_{D_T} values evaluated for both carbon and proton ions. y_{D_T} values have been calculated for all the particles that traverse the TEPC ($y_{D_T,TEPC}$) and for particles that HDM is able to track $y_{D_T,TRACKED}$ with the 288 strips configuration.

of loosing a particle if it has a small track length. Such events, in fact, traverse the TEPC edges and geometrically have a larger probability of missing the sensors, considering also MCS effects. In panels **D** for protons and **F** for carbons, the microdosimetric $yd(y)$ and $y_Td(y_T)$ spectra of particles that are not tracked by HDM are shown. Similarly to panels **D** of Figure 9.3 and **F** of Figure 9.5, where all particles are taken into account, the $y_Td(y_T)$ distribution peaks are shifted to the left for both protons and carbon ions radiation fields. Further, the high-y regions are significantly lower than the high-y_T regions; again, this is due to the real track lengths overestimation caused by using the mean chord length value.

Furthermore, the number of particles that reach at least one of the inter-strip passive regions with respect to the total number of events reaching the detectors (i.e. traversing either an active strip or an inter-strip region) has been estimated to be 63% for the 34 strips configuration and 81.5% for the 71 strips configuration. Increasing the number of strips in each sensor results in a significant improvement of the detection efficiency.

Finally, to asses the HDM capability to provide an accurate microdosimetric characterization of the radiation field, the following quantities have been calculated: standard y_D values considering all events traversing the TEPC; $y_{D_T,TEPC}$ values calculated from the $y_Td(y_T)$ spectra for all particles traversing the TEPC; $y_{D_T,TRACKED}$ values calculated from the $y_Td(y_T)$ spectra for all particles tracked by HDM. The results for both ions of interested are reported in Table 9.4.

The large discrepancy between y_D and $y_{D_T,TEPC}$ confirm the results shown in Figs. 9.3, 9.5, 9.8 and 9.9) and proves that the mean chord length approximation applied to the TEPC spectra does not provide an accurate description of the radiation field quality. The $y_{D_T,TRACKED}$, instead, is very close to $y_{D_T,TEPC}$, suggesting that the population of events tracked by HDM is representative of the actual field.

9.1.5 Discussion

An innovative design for a hybrid microdosimeter (HDM: hybrid detector for microdosimetry) is presented in this section. HDM is a two-stage detector composed by a TEPC and four layers of LGAD sensors. The combination of two different types of sensors (gas- and silicon-based) results in detection performances not offered by any existing microdosimeter. In fact, the TEPC gives a direct measurement of energy deposition in tissue while the LGADs provides particle tracking. The latter information has two main advantages: it improves the TEPC lateral spatial resolution to submillimetric precision and offer the real track length traverse by each particle in the TEPC. An improved lateral spatial resolution is helpful especially in a non homogeneous field, as the beam edges (both lateral and near the end-of-range). In these regions, the TEPC is only partially traversed by primary particles, whose path length might substantially deviate from the mean chord, depending on the detector position.

To assess the detector capability, we performed Monte Carlo simulations using Geant4 toolkit. As the primary application of HDM is particle therapy, we investigated its performances exposed to protons and carbon ions at a certain water depth.

The limitations of the mean chord length for our geometry are evident by looking at the track length distributions of all particles traversing the TEPC (Figs. 9.3,9.5 and Tab. 9.1). This approximation is based on the specific assumption that the TEPC is exposed to a uniform isotropic radiation field. In the cases considered here, although the beam generates such type of randomness, the water surrounding the TEPC causes the isotropy assumption to drop, with a direct consequence on the resulting mean track length. To further validate this, simulations without the water phantom has been performed and a mean track length value of 8.56 mm has been obtained for protons and 8.45 for carbon ions, both in accordance to the nominal mean chord length value.

However, even if a mean value of chord length based on more appropriate kind of randomness is used, the data reveal that a mean value is non representative of the whole track length distribution, since the standard deviations are rather large. This behavior is noticeable by the broadness of the track distributions in panels **C** of the Figure 9.3 and **E** of Figure 9.5.

Discrepancies between the mean chord and the real track length translate into difference between the standard $yd(y)$ and the alternative $y_T d(y_T)$ microdosimetry spectra (Figs. 9.3,9.5), the more evident being in the high y_T regions. The majority of particles populating these areas have a track length substantially smaller than the mean chord, and thus their actual lineal energy is is systematically underestimated if using the mean chord approximation.

The detector efficiency is defined by the number of particles that traverse the LGADs active regions, i.e. those that are tracked. This number depends on the LAGD configuration, i.e. the number of detection strips contained in a sensor. As the dead interstrip area is the same independently of the configuration, for a given total area of the sensor, by lowering the number of strips the detection efficiency increases. However, a larger number of strips results in a superior spatial resolution. To optimize the detector design for our application, we investigated HDM performances using three different LGAD configurations: 34, 71 and 288 strips per sensor.

Detection and tracking efficiencies were assessed by studying the composition of the radiation field detected by HDM versus the radiation field incoming on the TEPC. We identified three categories of events: i) particles detected by the entire system (i.e tracked events); ii) particles lost (i.e. only traversing the active volume of some detectors); iii)

particles non-trackable (i.e. those with not enough energy to reach the fourth LGAD).

For each category, we studied the kinetic energy spectra, track length distribution, real track versus track reconstructed with the tracking algorithm and microdosimetric spectra.

Independently of the primary ion and LGAD configuration, the mean track length of the tracked events is always higher that the value of all incoming particles. Events traversing the TEPC with a small track have a higher probability to miss the LGAD detectors. In fact, LGADs with 34 and 71 strips configurations have a total height and width comparable to the TEPC diameter, so if a particle reaches the TEPC with a given angle with respect to the primary beam direction, it is probable that its path will not cross all the LGADs. This hypothesis is confirmed by the fact that the 288 strips configuration collects a significantly higher portion of small-track particles (Figs. 9.6 and 9.7). Furthermore, for this configuration the mean track length of the tracked events is closer to the value of all particles (see Tab. 9.2). The mean tracks obtained when HDM is exposed to protons and carbon ions are similar for the 34 and 71 strips configurations. For the 288 strips configuration, HDM provides a more accurate track distribution for protons than for carbon ions. In fact, secondary fragments produced by protons reach, on average, smaller scattering angles compared to those generated by carbon ions [Rovituso and La Tessa, 2017].

However, those are the chords that suffer most from a high error on the tracking, as panels **A**, **C** and **E** of Figs 9.6 and 9.7 show for all the configurations. Furthermore, panels **A**, **D** and **G**, besides confirming the above mentioned fact that the bigger sensor takes better into account lower track lengths, they demonstrate also that the spatial resolution of the sensors, namely the widths of their strips, has a clear effect on the homogeneity of the track distribution. In fact, it can be noticed that the lower the spatial resolution is, the more the reconstructed tracks will have some preferential track lengths.

Finally, a comparison between panels **F** of Figure9.6 and 9.7 supports the hypothesis that, for protons, the 288 strips configuration is able to collect a track distribution which is very similar to the real one, while for carbon ions the distribution is still slightly underestimated for the small tracks.

Differences in the track length distributions for the LGAD configurations translate into different microdosimetric $y_T d(y_T)$ spectra (Figs. 9.8 and 9.9). A bigger sensor, like the 288 strips configuration, is able to collect more events with smaller TEPC tracks, which are the main contributors of the high y_T region.

The characterization of lost events indicates that the majority is caused to the LGADs fill-factor (interstrip regions). Thus, this issue can be resolved by increasing the measurement time to collect enough statistics.

For events that suffer MCS in the detectors, if the deviation angle is large enough they will be lost. In fact, even trying to enlarge LGADs or place them at a given angle with respect to the beam direction, the reconstructed track would be affected by errors too large to make the data of any value. The probability of loosing a particle because of MCS strongly depends on the HDM position in the radiation field. Depths in the Bragg peak regions as well as distal positions represent the worst cases because of the low kinetic energy of the particles populating these regions. This reflects on the fact that the proposed HDM configuration has limitations in those regions since it cannot operate close to and at distal part of the SOBP.

Finally, particles that do not have enough kinetic energy to reach all detectors are also a limit of HDM detection efficiency. Nonetheless, this issue can be partly solved by exploring the possibility of producing LGADs with thinner active layers or decreasing the

substrate width. For instance, reducing the total LGAD thickness down to 100 μm is considered achievable in the near future.

9.2 HDM read-out

Electronic read-out system

The LGAD sensors are read out through the ABACUS chip [Mazza et al., 2019], designed and produced at the University and INFN of Turin (Italy). The raw signals are fed to a board hosting a *Field Programmable Gate Array* (FPGA) and an Advanced RISC Machines (ARM) processor running Linux. A suitable FPGA program identifies the events according to the time of occurrence with a 1 μs resolution, along with the channel number corresponding to the detector strip hit by a particle. The data are then saved in the on-board Random Access Memory (RAM). The board processor allows to program the FPGA and to read out the data from the RAM. The board can be remotely accessed via Ethernet, and the data transferred as simple text files.

ABACUS read-out

The first step necessary to design the data acquisition (DAQ) has been the characterization of the detector output signal. In particular, the LGADs back-hand read-out system is managed by *ABACUS* board [Fausti et al., 2018], currently available in two versions:

1. an older version, used for testing purposes, that is capable of managing a maximum of 24 LGADs strips , and whose picture is shown in Figure 9.11. Additional features are:
    - the threshold level for the input channels is adjustable, but it is common to all the inputs;
    - the output signal is provided via a *Current Mode Logic* (CML) differential standard.
2. A newer version of the ABACUS board is currently under its final development state at University and INFN of Turin, featuring important improvements with respect to the older one, such as a channel-specific trigger level setting and the possibility to manage up to 96 LGAD strips. In addition, the output standard has been changed from CML to *Low Voltage Differential Signaling* (LVDS), which is a more common signal standard.

The signal read from the two CML output pins of the first version ABACUS board is reported in Figure 9.12. To characterize this output signal, we used the following instrumentation :

- Triple Output Power Supply.
- Function / Arbitrary Waveform Generator.
- Digital Oscilloscope.

The ABACUS board voltage of 3.3 V was provided by the power supply. As input signal, we selected a ramp waveform generated by the arbitrary waveform generator, ranging

from 0 V to 1 V to guarantee the comparator triggering. In order to record the CML output signal, we connected two oscilloscope probes to two output pins, and calculated the difference of the obtained signals using the oscilloscope integrated math functions. As it can be seen in Figure 9.12, the ABACUS output signal is a pulse with $V_{PP} = 1.5$ V and a duration of $\simeq 3$ ns, in accordance with the CML signal characteristics. On the basis of this characterization, the later stage of the acquisition chain can be designed.

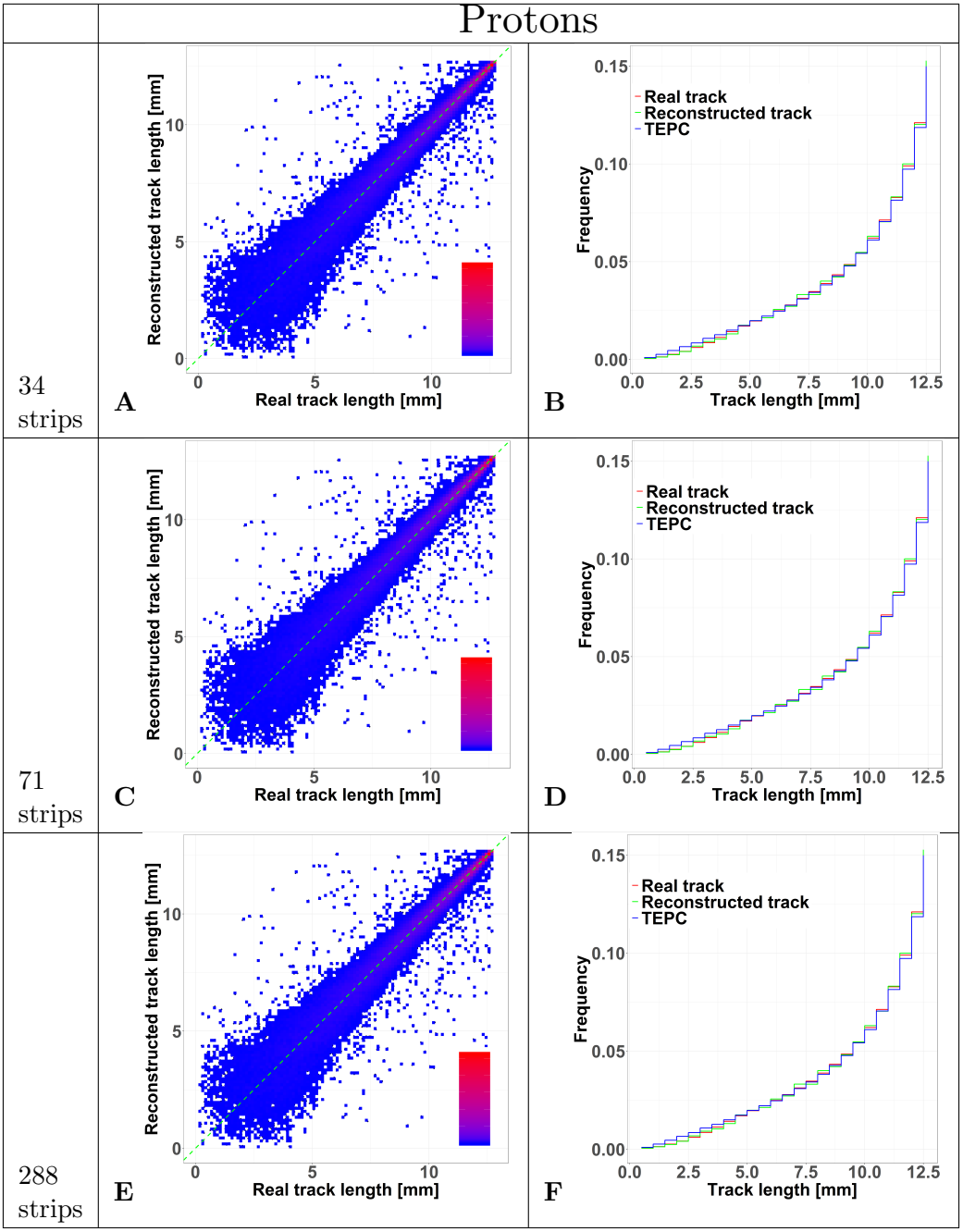

Figure 9.6: HDM performances when exposed to 150 MeV protons at 10.74 cm depth in water. The results are shown for 34, 71 and 288 strips LGAD configurations. Panels **A, C, E** shows 2D color plots of track length obtained with HDM versus real track length calculated directly with Geant4. The green dashed line at 45 degrees indicates the perfect agreement between the two datasets. The colors represent regions with a high (red) or low (blue) density of events. Panels **B, D, F** illustrate the comparison between the track length distributions of particles tracked by HDM considering the real track lengths calculated with Geant4 (blue line) or that reconstructed with the tracking algorithm (green line). The distributions of the real track lengths obtained directly from the simulation is also shown (red line).

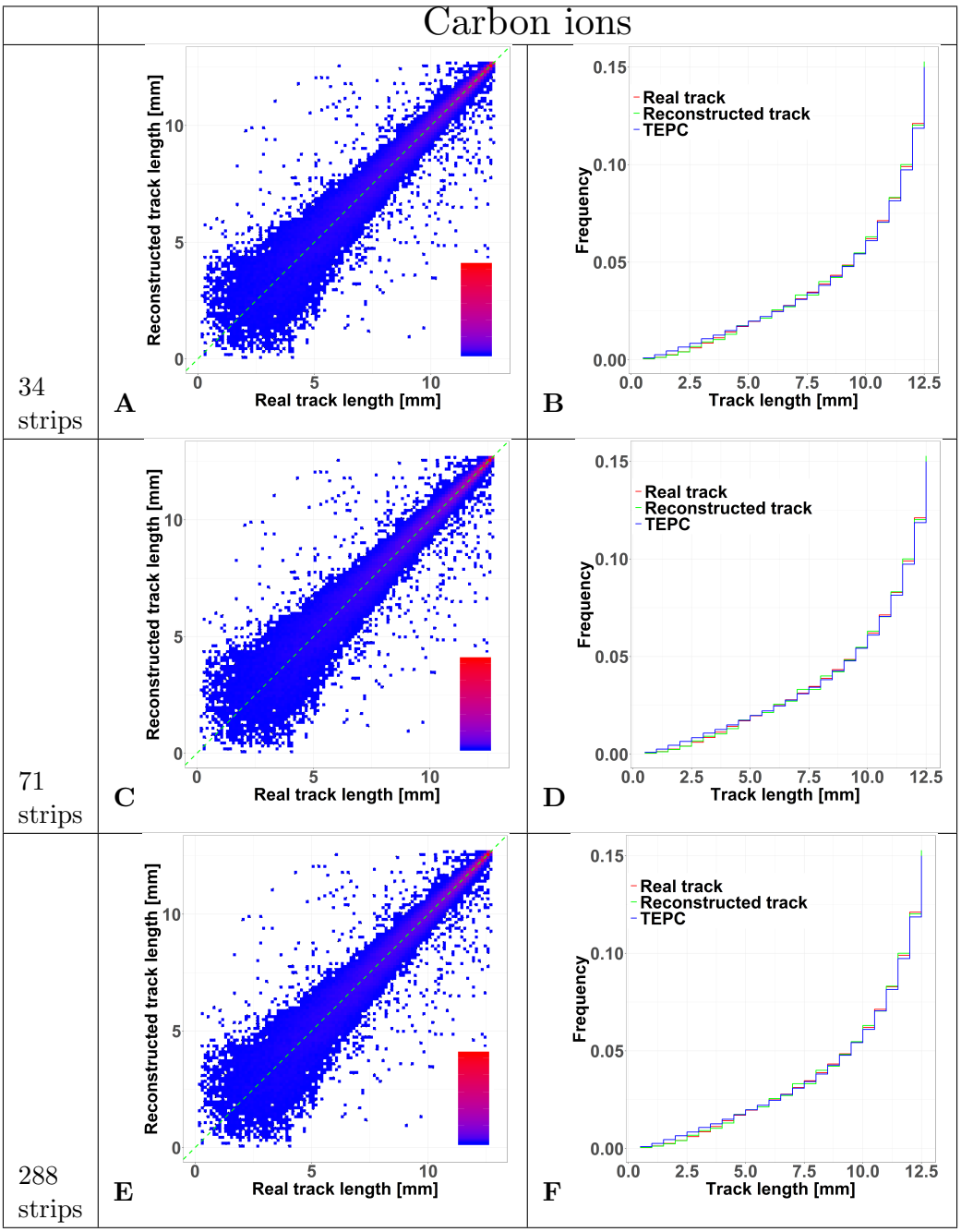

Figure 9.7: HDM performances when exposed to 290 MeV/u carbon ions at 10.74 cm depth in water. The results are shown for 34, 71 and 288 strips LGAD configurations. Panels **A, C, E** shows 2D color plots of track length obtained with HDM versus real track length calculated directly with Geant4. The green dashed line at 45 degrees indicates the perfect agreement between the two datasets. The colors represent regions with a high (red) or low (blue) density of events. Panels **B, D, F** illustrate the comparison between the track length distributions of particles tracked by HDM considering the real track lengths calculated with Geant4 (blue line) or that reconstructed with the tracking algorithm (green line). The distributions of the real track lengths obtained directly from the simulation is also shown (red line).

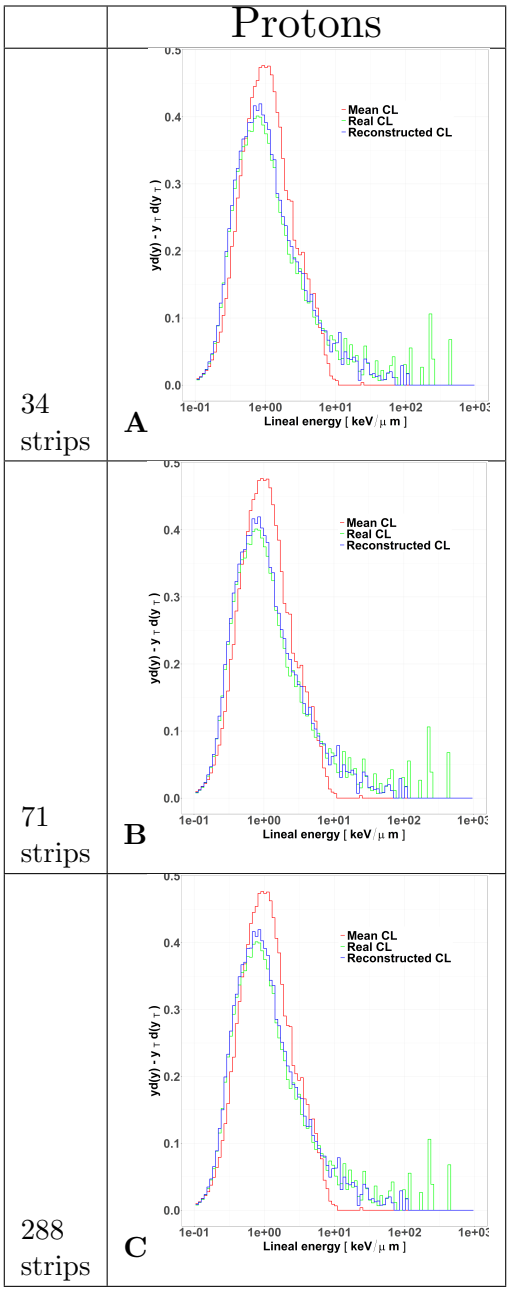

Figure 9.8: Microdosimetric spectra of all particles tracked by HDM when irradiated with 150 MeV protons at a depth or 10.74 cm in water. The distributions include the standard $yd(y)$ spectra calculated with the mean chord length (red line) and the $y_T d(y_T)$ spectra obtained either with the real track length (green line) or with the value estimated with the tracking algorithm (blue line). The distributions are shown for LGAD configurations with 34 (panel **A**), 71 (panel **B**) and 288 (panel **C**) strips..

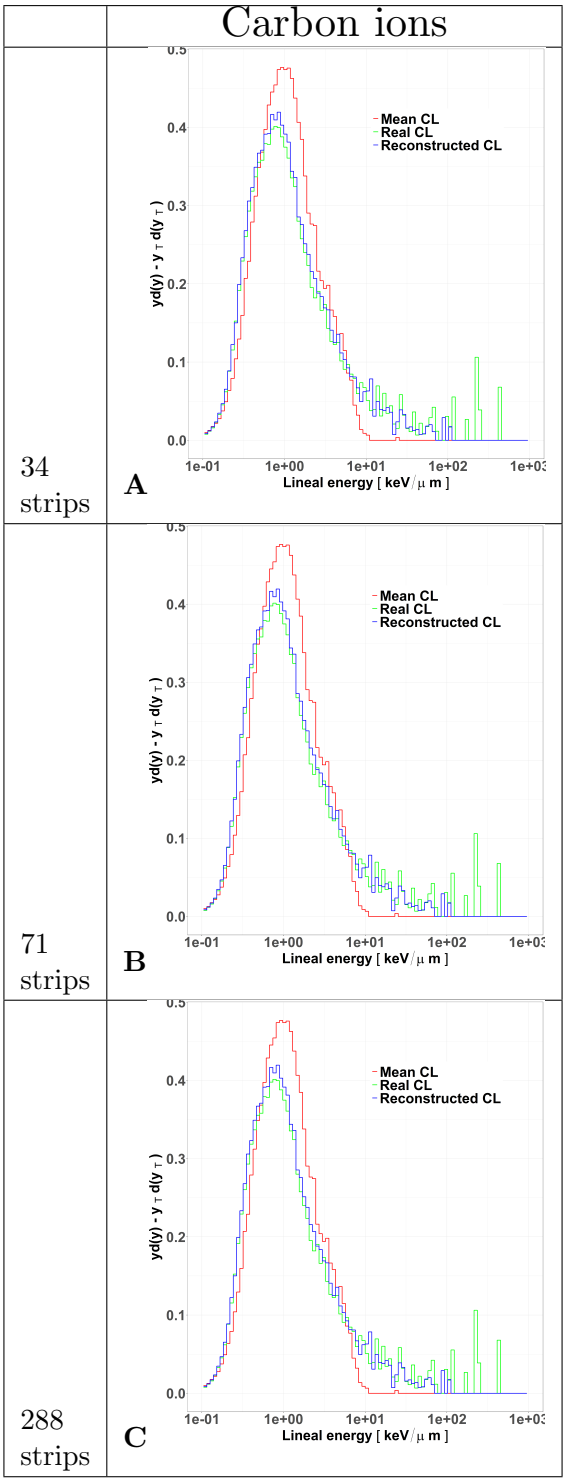

Figure 9.9: Microdosimetric spectra of all particles tracked by HDM when irradiated with 290 MeV/u carbon ions at a depth or 10.74 cm in water. The distributions include the standard $yd(y)$ spectra calculated with the mean chord length (red line) and the $y_T d(y_T)$ spectra obtained either with the real track length (green line) or with the value estimated with the tracking algorithm (blue line). The distributions are shown for LGAD configurations with 34 (panel **A**), 71 (panel **B**) and 288 (panel **C**) strips.

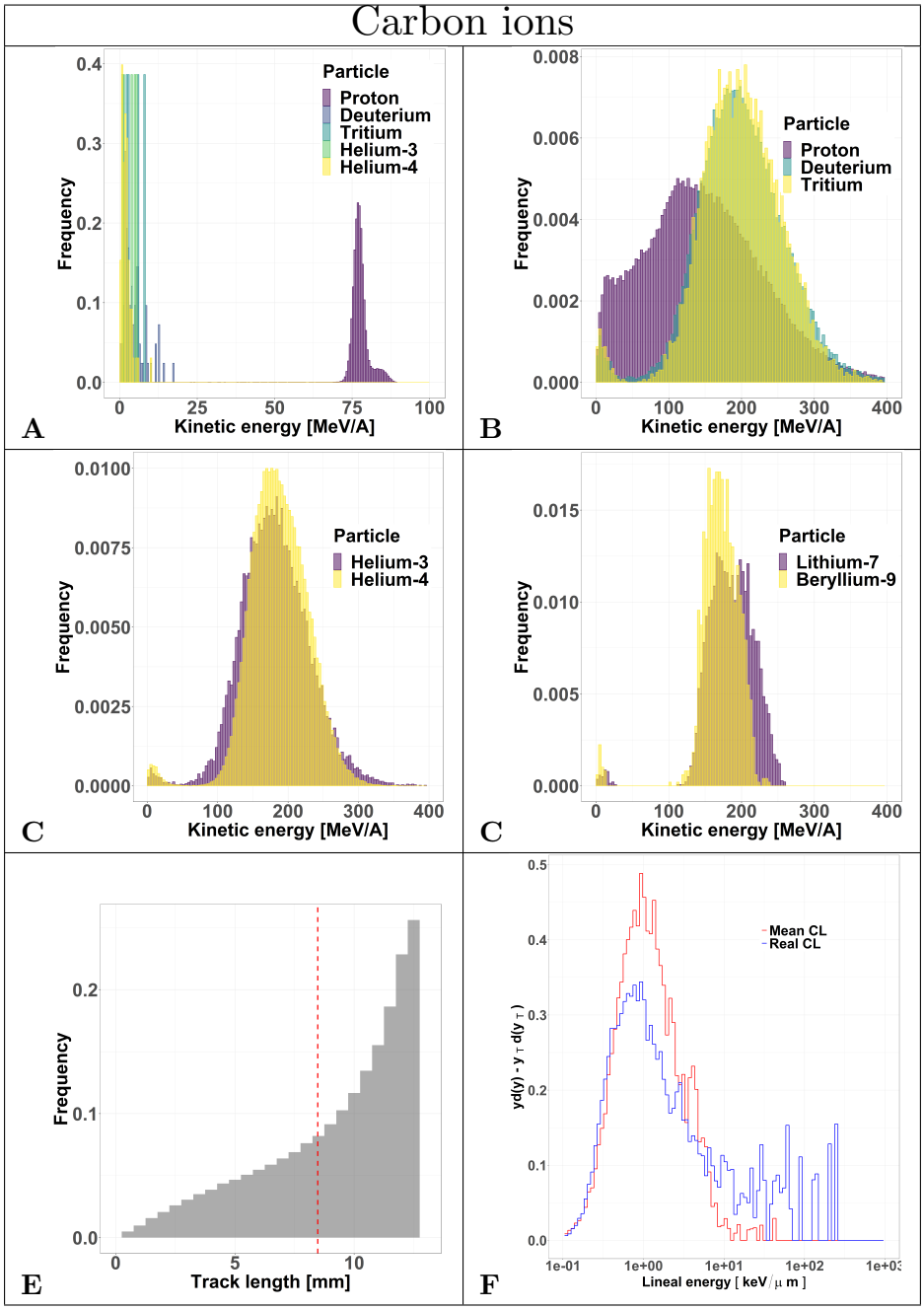

Figure 9.10: Characterization of the particles lost by HDM when irradiated with 290 MeV/u carbon ions at a depth or 10.74 cm in water. Panels **A** to **D**: kinetic energy spectra of the most abundant components of the radiation field including and excluding the primary ions. Panel **E**: track length distribution of all the particles detected by the TEPC. The mean chord length at 8.47 mm is marked with a red dotted line. Panel **F**: microdosimetric yd(y) spectra obtained with the mean chord length approximation (red line) and microdosimetric $y_T d(y_T)$ spectra obtained using the real chord length values (blue line).

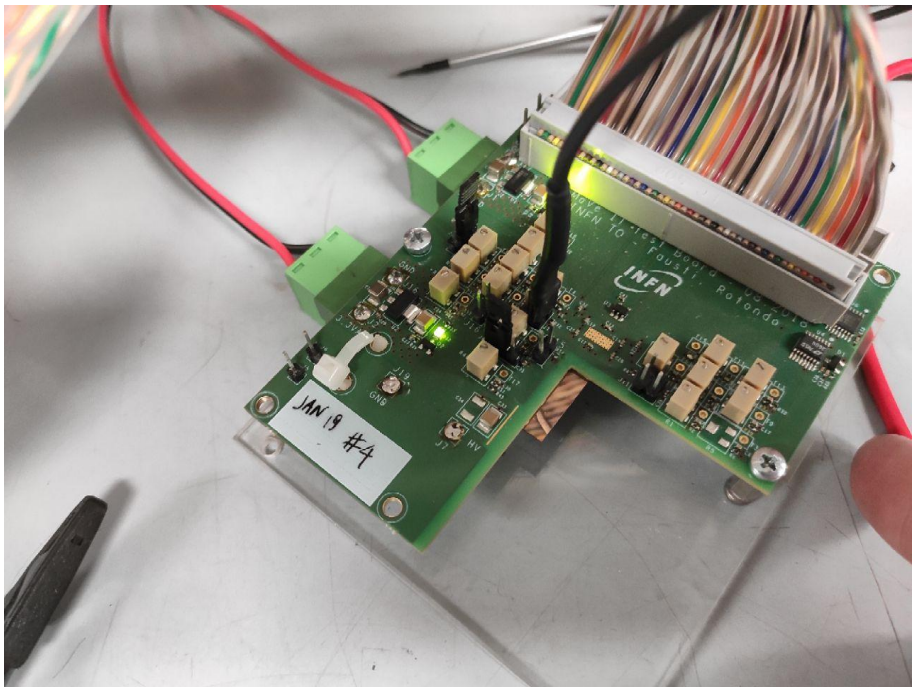

Figure 9.11: A photo of the ABACUS board first version tested.

Figure 9.12: Output signal from the ABACUS board (first version). The cyan and yellow signals are measured on the CML output pins. The purple signal represents the difference between the two CML outputs, obtained with the integrated oscilloscope math functions.

Chapter 10

Machine learning approach to predict the real track length using HDM

The previous Chapter 9 has made clear that the detection efficiency is the most critical issue of HDM. To tackle such issue as well as to improve tracking accuracy, in this Chapter we exploit modern *Machine Learning* techniques, introduced in Chapter 5.3. In the present Chapter we propose a stacked ML model composed by different connected modules suitably structured to solve a specific task.
The first ML module aims at improving the detector efficiency, filling the missing spatial positions on the LGADs. The second one instead, using the augment information provided by previous module, reconstructs the tracks of the particles to calculate the real track length microdosimetric spectra. Both modules are *Random Forest* based, so that several decision trees are trained in parallel and the final output of each module is an ensemble of the trees, called *forest*. Such methodology allows for a new tracking paradigm, so that missing values as well as multivariate physical information can be exploited. Finally, thanks to this novel tracking approach, we were able to achieve a 100% efficiency of HDM and a higher accuracy in the tracking reconstruction algorithm compared to the linear algorithm proposed in the first feasibility study 9.

As regard tracking algorithm, recently ML methods have gained attention due to their accuracy compared with standard methods. In this direction, one of the major strengths of ML consists in the possibility to pass several types of information to the model. This allows a more efficient and accurate track reconstruction compared to more classical methods, [Schulte et al., 2008]. It is further worth stressing that much of the attention of ML algorithms for tracking particle come from high energy physics, [Britton et al., 2020, Duarte and Vlimant, 2020, Amrouche et al., 2020]. Such application has a profound difference in the physical scenario we consider in the current work, where particles at therapeutic energies has a significantly lower energy. This implies that scattering of particles is highly non-negligible and should be carefully taken into account. As it will be seen in details in later discussion, scattering is among the major source of uncertainty in the proposed application with a consequent high number of undetected particles.

10.1 Simulation setup of HDM

In order to develop the advanced *Machine Learning* (ML) tracking model, the 288-strip configuration depicted in 9 has been chosen. This is the LGADs configuration with the sensors largest area, and we chose it to apply the machine learning techniques at the case with the largest amount of tracking information.

In the simulation setup, HDM has been placed in a water phantom with PMMA walls (1.74 cm water equivalent thickness) and irradiated with 150 MeV protons and 290 MeV/u carbon ions, which have the same range in water (∼160 mm). For both beams, we used a circular beam spot of 3 cm radius, and positioned HDM at a depth of 10.74 cm along the beam direction. This setup was designed to reproduce realistic irradiation conditions, and to expose the detector to a mixed radiation field, that can lead to major difficulties in tracking with respect to a pure monoenergetic beam.

The model: Random Forest

A *Random Forest* algorithm ,[Ho, 1995, Friedman et al., 2001], has been used both to predict the real track length of particles traversing the TEPC as well as to fill the missing positions of the LGADs; regarding this latter topic, it is worth stressing that due to dead zone of the detector and scattering, some particles are not seen by some of the LGAD layers. This would imply that in such situations it is not possible to track the particle inside the TEPC. Instead, we developed a *Random Forest* based module that reconstructs and fills the missing information so that the real track length of particles inside the TEPC can be predicted. *Random Forest* is an ensemble ML algorithm that combines weaker models, such as decision trees, to create a more robust final model. Being a bagging algorithm, the ensemble model is created in parallel, and thus the output is the average of all trees outcomes. Compared to decision-trees, the *Random Forest* reduced the overfitting on the train data, and for this reason it improves the predictions accuracy.

More formally, given the dataset $\mathcal{D} := \{\mathcal{X}_i, y_i\}_{i=1}^{N}$ composed by N samples, where

$$\mathcal{X}_i := \left(x_1^i, \ldots, x_n^i\right) , \quad n \in \mathbb{N}, i = 1, \ldots, N < \infty ,$$

are the n features to predict the target variable $y_i \in \mathbb{R}$.

As most ML algorithms, given a function ϕ that describes the link between the features $\mathcal{X}_i$ and the target y_i, the final goal is to look for an approximation $\hat{\phi}$ of ϕ. Such approximation is chosen as the function that minimizes the expected values of a certain suitable loss $\mathcal{L}$. Typically, a regularization term λ is added to the loss function to avoid overfitting.

The *Random Forest* algorithm assumes that $\hat{\phi}$ is the average of weaker learners decision–trees ψ_k, that is

$$\hat{\phi}\left(x_1^i, \ldots, x_n^i\right) = \frac{1}{K} \sum_{k=1}^{K} \psi_k \left(x_1^i, \ldots, x_n^i\right) ,$$

where ψ_k is the outcome of the $k-$th decision tree.

In this study, we will consider two different models, that are both random forest based but the training is done on different sets. The first model, denoted in the following as RF, is a model trained on a subsample of the whole dataset; the second model, denoted by RF_0, on the contrary is a model trained solely on particles that have been seen by all the LGADs. The choice of considering two different training sets is due to the fact that

the tracking algorithm is strongly affected by the ML module dedicated to missing value filling, so that a careful choice on the most proper training dataset must be done in order not to induce some bias into the tracking model.

Bayesian optimization

Hyper–parameters optimization is a key aspects in the development of a robust and efficient ML algorithm that has to improve the final accuracy of the chosen model. An extensive hyper–parameters search is nonetheless extremely expensive. A typical approach consists in a grid search or, in the last years, in random searches [Snoek et al., 2012, Bergstra et al., 2011, Hutter et al., 2011]. In a nutshell, the hyper–parameters optimization aims at finding the model hyper–parameters $\theta = (\theta_1, \ldots, \theta_d)$ that yields the lowest error on the validation set [Frazier, 2018]. More formally, writing explicitly the dependence of the cost $\mathcal{L}$ upon the hyperparameters $\bar{\theta} \in \Theta$, any hyper–parameter search goal is to solves

$$\theta^* := \arg\min_{\theta \in \Theta} \mathcal{L}(y, x|\theta)\,.$$

The Bayesian hyper–parameter search, in contrast to the random or grid searches, first creates a surrogate model typically denoted as $p(y|x)$. Such surrogate model is usually easier to optimize than the original objective function, and is created along with a so–called *prior density distribution*; thus, the hyper–parameters are optimized on this surrogate model rather than on the original cost function. Thus, the *posterior density distribution* is obtained updating the prior via *Bayes theorem*. Given an *acquisition function* α, that depends on the posterior, the next hyper–parameter point is decided according to

$$\theta_n := \arg\min_{\theta \in \Theta} \alpha(\theta|\theta_{1:n-1})\,,$$

where $\theta_{1:n-1}$ are sample previously drawns and are defined as $\{\theta_1, \ldots, \theta_{n-1}\}$. The most popular *acquisition functions* are the *maximum probability of improvement* (MPI), the *expected improvement* (EI) and the *upper confidence bound* (UCB), [Brochu et al., 2010]. For instance, the EI is defined as

$$EI(\theta) = \mathbb{E}\left[\max(\mathcal{L}(\theta) - \mathcal{L}(\theta^+), 0)\right]\,,$$

being θ^+ the best values reached.

Therefore, the new point θ_n is added and the surrogate model posterior is updated, so that previous steps are repeated until a certain fixed threshold on either the iteration or the time is reached.

A Bayesian hyper–parameter optimization algorithm keeps track of the past evaluations, and use the past information to choose future steps.

Recently, the Bayesian optimization has proven to be able to provide more accurate results than the grid or random searches with a lower computational time. In fact, unlike the classical approaches, the Bayesian optimization keeps track of the past results when choosing the future hyper–parameters values.

Error metric assessment

We considered two different metrics: the *weighted Mean Absolute Percentage Error* (MAPE_ω) and the *weighted Mean Absolute Error* (MAE_ω).

MAPE_ω is defined as

$$\text{MAPE}_\omega = \frac{1}{\bar{\omega}} \sum_{n=1}^{N} \omega(n) \frac{|(l(n)+1) - (\hat{l}(n)+1)|}{l(n)+1}\,, \tag{10.1}$$

where N is the number of samples in the test set, ω is a weighting factor to be suitably introduced in a while, $\bar{\omega} := \sum_{n=1}^{N} \omega(n)$, $l(n)$ and $\hat{l}(n)$ are the $n-$th event real and predicted track–length, respectively. It is worth stressing that, since $l(n)$ and $\hat{l}(n)$ can assume values close to 0, in order to avoid singularities and consequent explosion of the error function both the real and the predicted track–length have been shifted by one, namely $l(n)+1$ and $\hat{l}(n)+1$ have been considered in Equation (10.1).

Using the same noation, MAE_ω is defined as

$$\text{MAE}_\omega = \frac{1}{\bar{\omega}} \sum_{n=1}^{N} \omega(n) |l(n) - \hat{l}(n)|\,, \tag{10.2}$$

We tested this metric for different weights ω_i, namely for

$$\begin{aligned}
&\omega_1 = 1\,,\\
&\omega_l = \frac{1}{l}\,,\\
&\omega_y = y := \frac{\epsilon(n)}{l}\,,\\
&\omega_{y^2} = y^2 := \left(\frac{\epsilon}{l}\right)^2\,.
\end{aligned}$$

Besides the trivial choice of no weight used as a benchmark, the rationale behind the choice of ω_l is that a greater weight can be attributed to low track lengths values, and a lower weight to high track length values. In addition, ω_y and ω_{y^2} have been considered among the possible weights since, the ultimate goal of the ML reconstruction algorithm is to reconstruct the $yd(y)$ real-track length simulated microdosimetric spectra, which we recall to be proportional to the y^2 values.

10.2 The data

A different set of data is passed to each module of the model. As regard the missing values module, the *Random Forest* model is trained with 5 variables denoting the energy deposition inside the TECP and the spatial positions detected by the LGADs. A set of 5 different models is trained, and each time the position of a particle that traverses the LGAD but has a missing value is filled with a prediction. At last, the average and standard deviations of the 5 different outcomes is taken. This is done to obtain a more robust missing value prediction, as well to estimate the error for filling a missing value. Intuitively, a higher standard deviation should corresponds to a higher uncertainty in the predicted missing value.

To predict the real track-length, we trained a new model based on 14 variables. The first 5 variables, as in the previous module, are the energy deposition inside the TEPC and the positions of the particles traversing the LGADs, either the real value or the prediction obtained from the missing values module. We further consider the standard deviation in

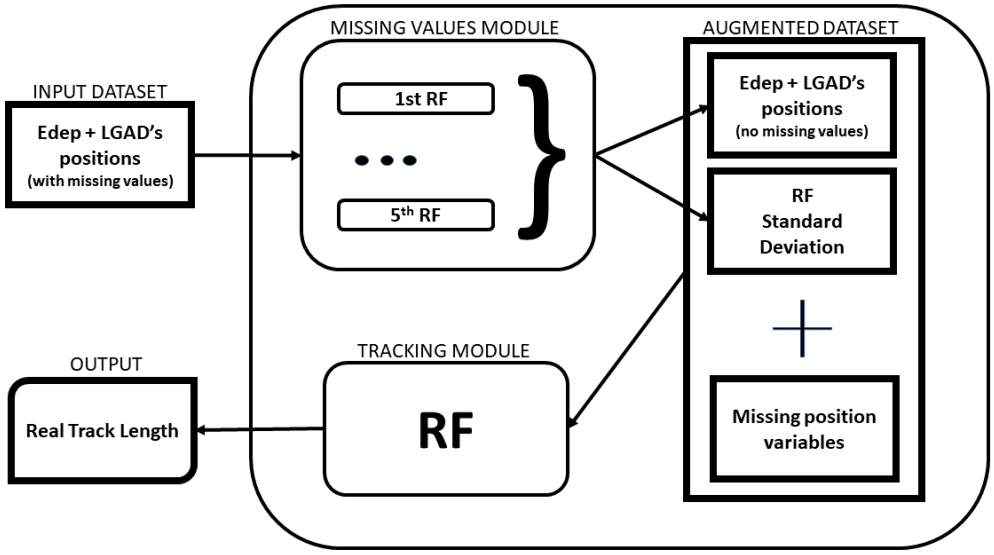

Figure 10.1: Machine learning algorithm workflow: i) the energy deposition together with the 4 LGADs positions (also with missing values) compose the input that is passed to ii) the missing value module, made of 5 *Random Forest* modules that predict the missing values of the input dataset. The mean value of the predictions computed by 5 *Random Forest* models is finally calculated together with the standard deviation, so that iii) the energy depositions and the positions of particle traversing the 4 LGAD's compose the input of the tracking module in order to achieve iv) the real track length prediction.

the predictions of filling undetected particles of the 5 *Random Forest* algorithms developed in the missing value filling module; we then considered 4 binary variables, named LGAD_i with values either 0 or 1 whenever the $i-$th LGADs has seen or not the particle. At last we included a variable, denoted in the following by *Miss* assuming values in $\{0, 1, 2, 3, 4\}$ denoting the number of LGADs that have not seen the particle.

As mentioned above, we consider two different training sets for the module that predicts the *real track length*: a model (RF) trained on the whole dataset, with no distinction on the number of missed LGADs, and another model (RF_0), that has been trained only on particle seen by all the LGADs, and therefore with variable *Miss* equals to 0.

A schematic representation of the model is depicted in Figure 10.1.

10.3 Results

The results of this study first include an exploratory analysis of the missing values, aiming to identify specific patterns among the non detected particles. Then, thanks to the two model modules, the $y_T d(y_T)$ spectra are derived and analyzed. Finally, the results on the metric assessment are presented.

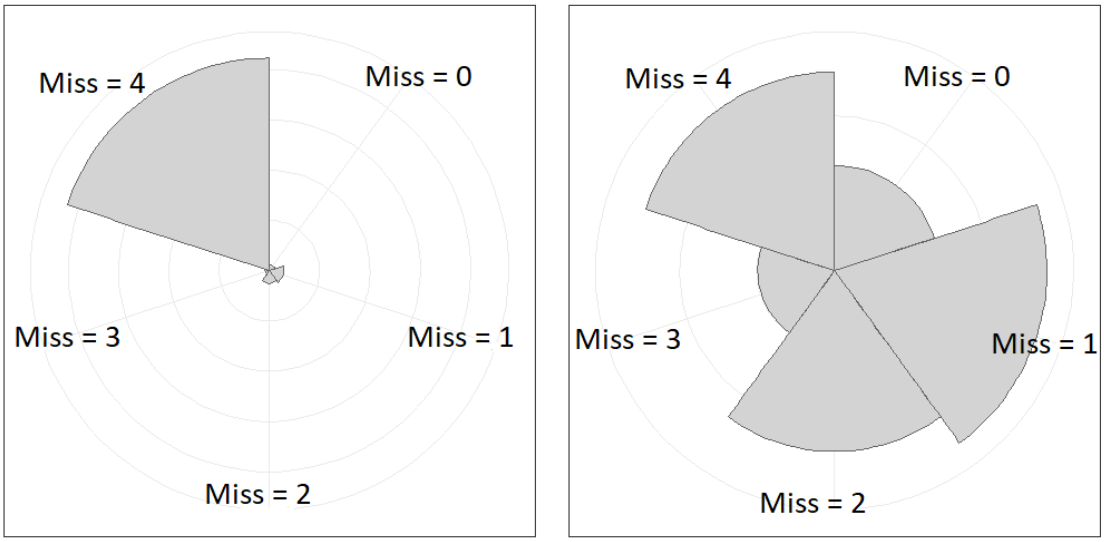

Figure 10.2: Pie charts of missing information on the 4 layers of LGADs for protons (left panel) and carbon ions (right panel). The missing values are identified as *Miss* followed by the number of LGAD layers that did not detect a particle, namely Miss=0 refers to the complete information, Miss=1 refers to one missing value in one layer, Miss=2 to two missing values in two different layers and so on.

Missing values analysis

The missing value analysis focuses on studying the behavior of particles not detected by one or more LGADs layers. The causes of non detection include the passage in the LGADs dead zones, like the interstrip regions, as well as lateral deviations from the inital linear path due to electromagnetic or nuclear interactions occurring within the TEPC, with the LGADs themselves and with water. In Fig 10.2, the pie charts of missing values are shown for protons (left panel) and for carbon ions (right panel). The results clearly indicate that HDM is heavily affected by a low efficiency especially for protons, with 97% of the particles not detected by at least one of the layers, and 85% of the particles missing all the layers. For carbon ions, instead, the efficiency is higher, with 85% of particles that are not detected by at least one layer, and just 25% of them not tracked by any layers. Machine learning is a powerful tool to overcome this big limitation of non detected events affecting HDM, being able to reconstruct the missing values by training on the complete dataset of detected events.

In Figure 10.3 a more detailed analysis on the missing values is plotted for protons (up panel) and carbon ions (bottom panel). This plot confirms that for protons the majority of the events are not detected by any of the LGAD layers, which represents the dominant behavior of the incoming particles. For carbon ions, instead, the data indicate that the majority of non detected particles missed either all four layers, or just one. Furthermore, a relevant proportion of particles has not been detected by the couple of layers with strips oriented in the same direction.

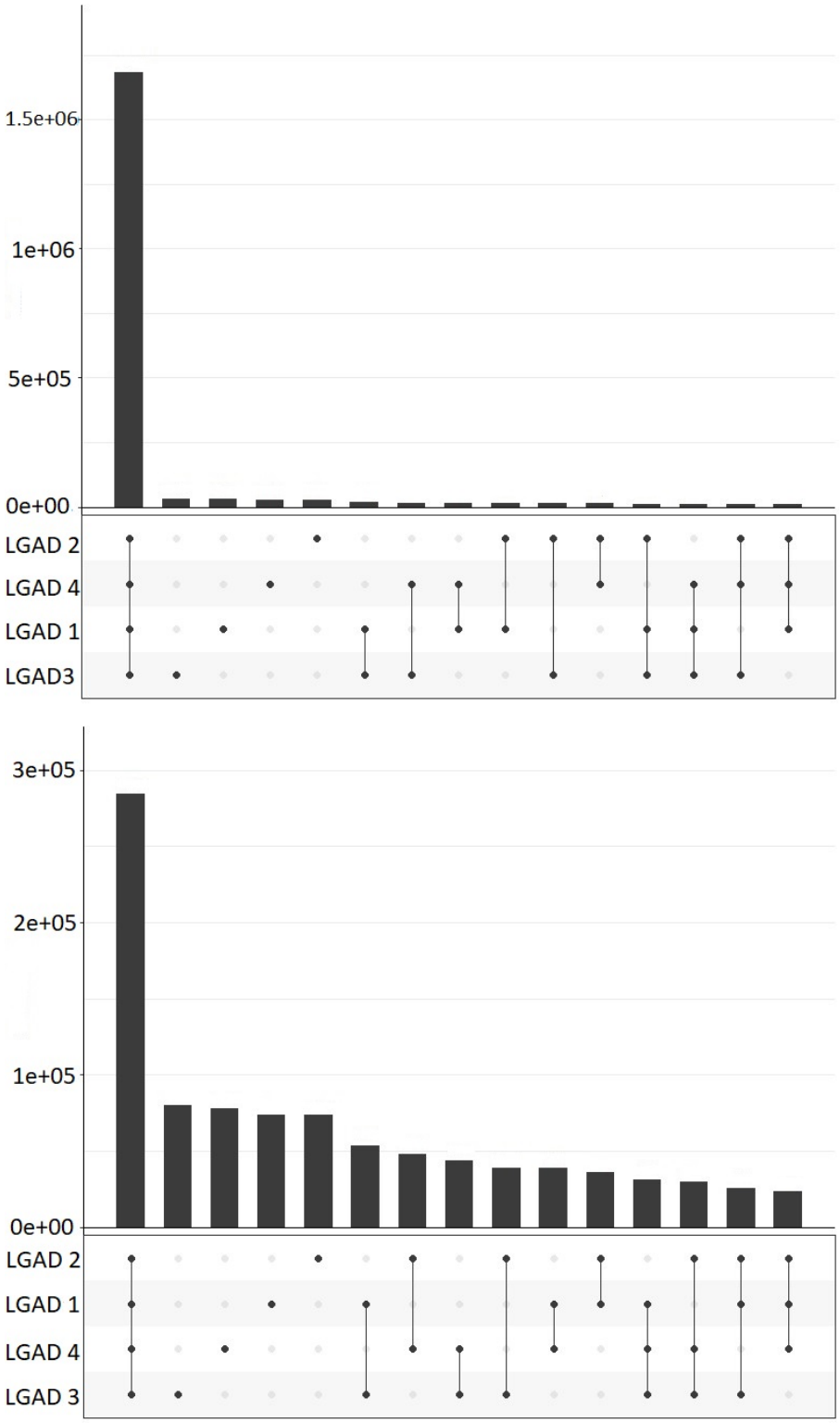

Figure 10.3: Histograms of missing values on the 4 LGAD layers for protons (left panel) and carbon ions (right panel), ordered in descending counting values. The black points represent the position where the value is missing, and are labelled as the LGAD identification number written on the left (LGAD1, LGAD2, LGAD3, LGAD4), with LGAD1 being the nearest to the TEPC.

	MCL	RTL	Linear	RF	RF_0		MCL	RTL	Linear	RF	RF_0
$y_F - y_{F_T}$	0.78	1.02	0.60	0.99	0.61	$y_F - y_{F_T}$	5.18	6.57	5.16	6.01	5.63
$y_D - y_{D_T}$	5.71	36.22	5.08	14.56	5.25	$y_D - y_{D_T}$	15.16	58.77	24.15	22.14	31.62
$y^* - y^*{}_T$	2.79	13.04	2.62	6.57	2.79	$y^* - y^*_T$	13.84	23.96	16.69	17.36	20.55

Table 10.1: y_F, y_{F_T}, y_D, y_{D_T}, y^*, y^*_T and values for protons in left panel and for carbon ions in right panel.

Microdosimetric analysis

The main goal of this investigation is to accurately reconstruct the $y_T d(y_T)$ microdosimetric spectra, predicting the real track lengths of particles traversing the TEPC.
In Figure10.4, the standard $yd(y)$ microdosimetric spectra obtained with the mean chord length approximation are plotted against the $y_T d(y_T)$ spectra calculated using the real track lengths, and with the three $y_T d(y_T)$ spectra obtained exploiting the predicted track lengths from the three model used (linear, RF, RF_0). The results are reported in Figure10.4 for protons (top panel) and carbon ions (bottom panel). For protons, the RF appears to be the most accurate model compared to both the linear and the RF_0. In fact, it reproduces better the distribution left rise as well as the main peak. In the right tail region, despite a clear difference with the RTL spectrum, the RF still better forecast the correct curve than the other models. In this sense, the major issue is an overestimation of the y_T around 10 keV/μm, that results both in a second peak with no physical meaning and in a small tail in the extremely high y_T region above 10 keV/μm.

For carbon ion (bottom panel), the RF_0 is the most accurate method for reconstructing the RTL spectrum. The $yd(y)$ distribution behaviour is similar to protons, with a good prediction of the main peak and a second peak appearing at high y_T. Furthermore, the RF_0 is the only method that can reproduce the fat tail of the RTL spectrum.

To further investigate the cause of the second peak predicted by the best performing methods, that we recall to be RF for protons and RF_0 for carbon ions, in Figure 10.5 we reported the contribution of the track-length to the $y_T d(y_T)$ microdosimetric spectra. Both panels clearly suggest that the reconstructed spectra for particles with a real track length above 6 mm is in a good agreement with the RTL spectrum, while particles with a track length smaller than 2 mm are rarely predicted by any models. This result point to the fact that the main error in the spectrum reconstruction is due to extremely short tracks, that are very few but at the same time very relevant in the $y_T d(y_T)$ spectrum. The second peak in the reconstructed spectrum for both ions is almost completely composed by tracks predicted to be between 2 and 6 mm whose real track length is instead below 2 mm.

In addition, Table 10.1 shows the comparison between the y_F, y_D and $y*$ values obtained from the microdosimetric $yd(y)$ and $y_T d(y_T)$ spectra.

In order to look for patterns as a function of the track length specifically for particles with extremely high reconstruction errors, we separated the particles into two groups: particles whose reconstructed track length has an error above the 99^{th} percentile and particles whose reconstructed track length has an error below the 99^{th} percentile. To understand the main features of the two groups, we plotted the positions of particles traversing the first LGAD pair in the X versus Y spatial planes. Events have been divided into three group based on the real-track length as estimated from *Geant4* simulations. Higher panels denote events with an error above the 99 percentile, while lower panels report the complementary group. The linear tracking algorithm might give a null result (NA

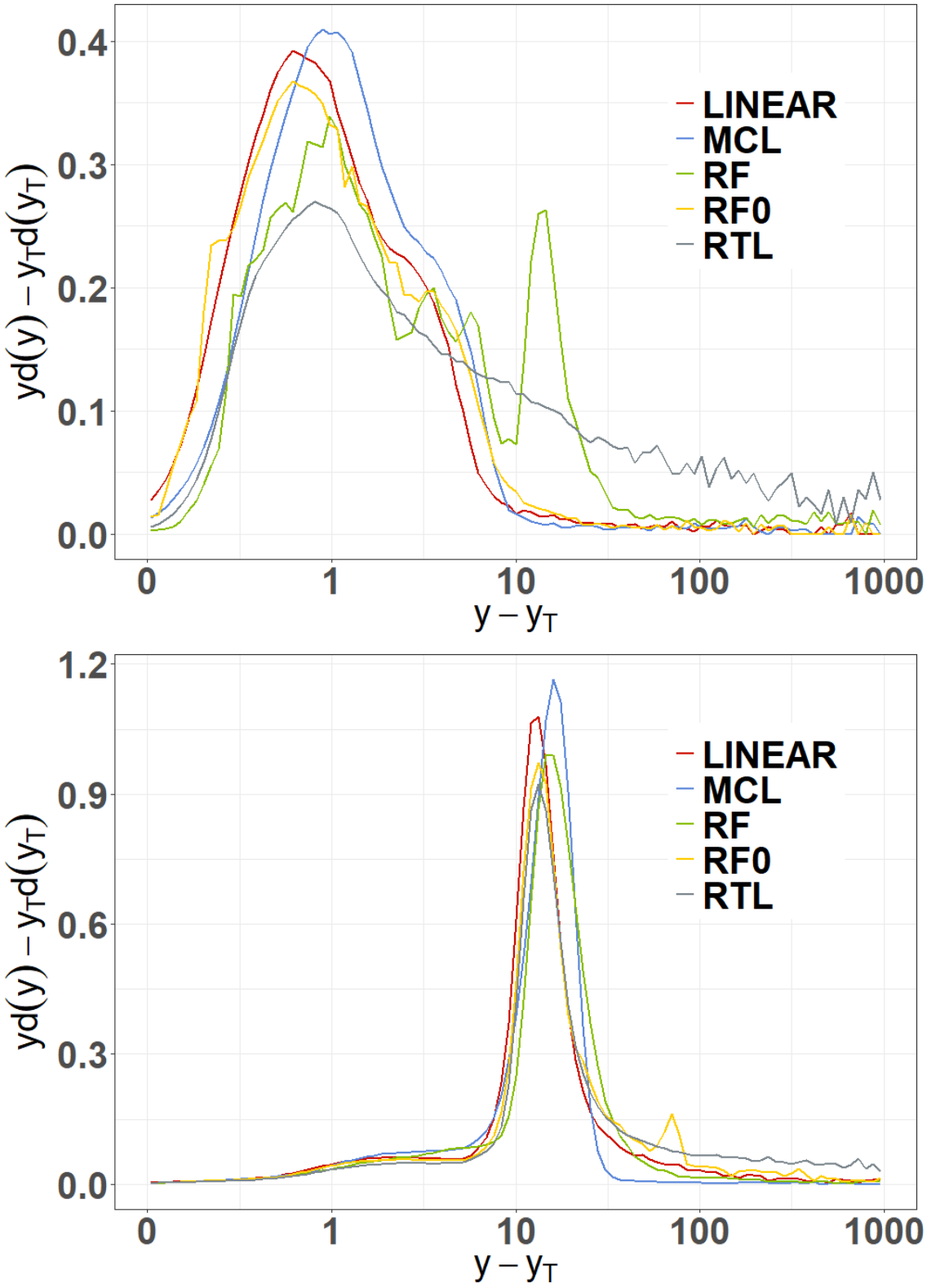

Figure 10.4: Standard $yd(y)$ (MCL) distribution compared to the $y_T d(y_T)$ spectra obtained with the real track length (RTL) and the predicted track length for protons from the three models: linear, RF, RF_0. The upper figure presents the results for protons, while the bottom figure for carbon ions.

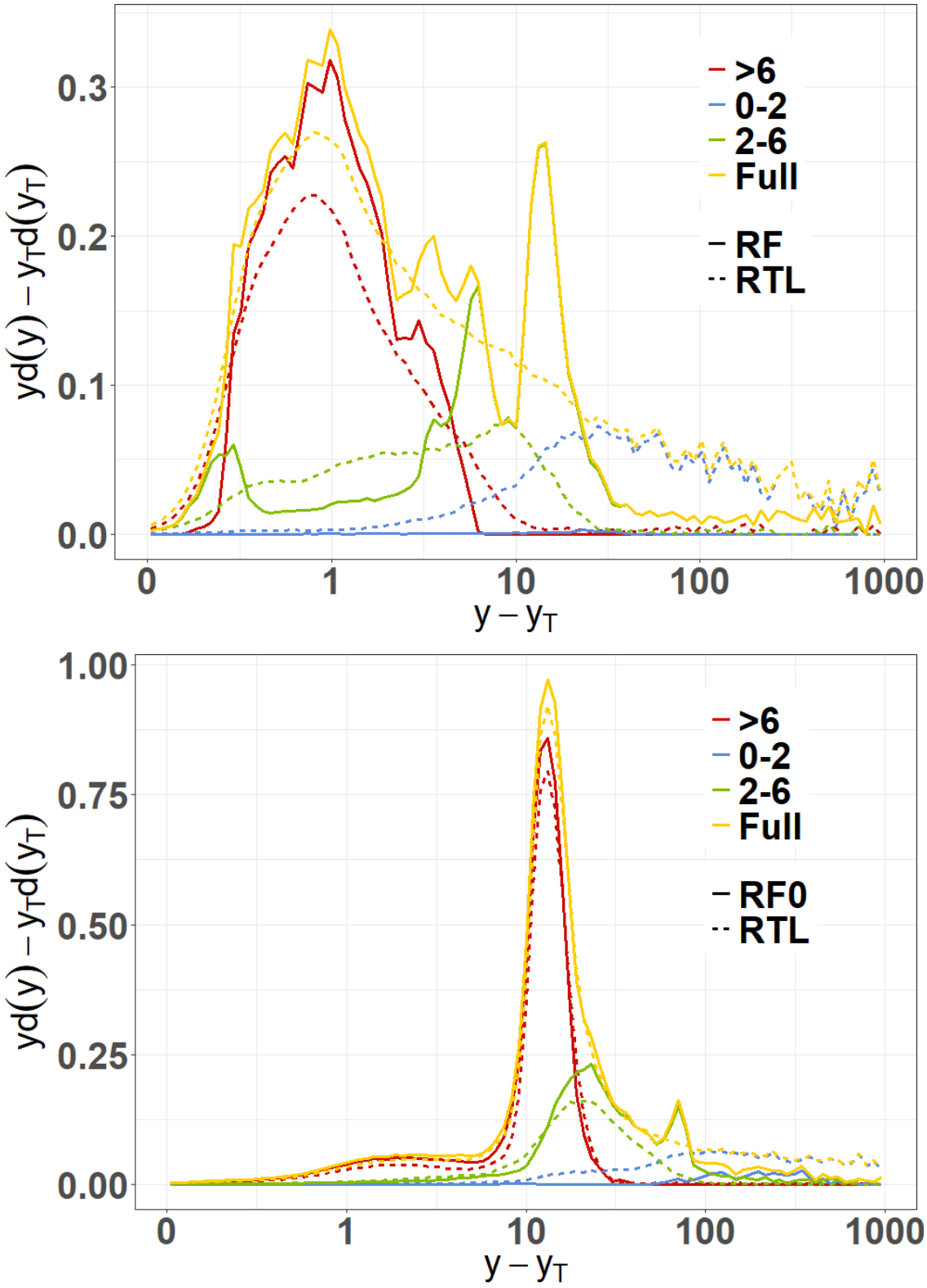

Figure 10.5: Upper figure shows in yellow dotted line $y_T d(y_T)$ (RTL) spectra compared to the reconstructed $y_T d(y_T)$ (RF), in yellow continuous line. The contributions of different track length values to the total $y_T d(y_T)$ spectra are given by different colors according to the legend. Continuous lines reports the reconstructed spectra whereas dotted line plots show the RTL spectra. Bottom figure shows the same comparison for carbon ions, where RF0 model is considered instead of RF model.

Protons				Carbon ions			
	Linear	**RF**	**RF$_0$**		**Linear**	**RF**	**RF$_0$**
MAPE$_{\omega_1}$	0.52	0.30	0.40	**MAPE**$_{\omega_1}$	0.38	0.29	0.35
MAPE$_{\omega_l}$	0.48	0.29	0.37	**MAPE**$_{\omega_l}$	0.34	0.27	0.32
MAPE$_{\omega_y}$	2.40	1.08	2.13	**MAPE**$_{\omega_y}$	1.27	0.86	0.87
MAPE$_{\omega_{y^2}}$	10.81	3.14	9.07	**MAPE**$_{\omega_{y^2}}$	8.73	2.66	0.95
MAE$_{\omega_1}$	3.07	2.49	2.47	**MAE**$_{\omega_1}$	2.10	2.20	2.11
MAE$_{\omega_l}$	2.96	2.50	2.40	**MAE**$_{\omega_l}$	2.01	2.20	2.05
MAE$_{\omega_y}$	5.48	3.23	4.87	**MAE**$_{\omega_y}$	2.95	2.22	2.29
MAE$_{\omega_{y^2}}$	11.42	3.37	9.57	**MAE**$_{\omega_{y^2}}$	9.30	2.68	0.97

Table 10.2: MAPE and MAE values calculated with different weights for protons (left panel) and carbon ions (right panel).

panel), being unable to track the particle path inside the TEPC if the linear interpolation starting from the LGAD positions does not intersect the microdosimeter. The results (Figure 10.6) show that the low tracking errors are associated with high track lengths, especially in the plane center. This region is where we find the particles with the mostly linear path from the TEPC to the LGADs, i.e. that traverse the TEPC active sphere around the diameter, and therefore have a long track. The plot also shows a circle of approximately 6 mm radius of short and medium tracks, populated by particles that are traversing the TEPC at the border regions with a linear path, easily predictable by the model with a low error.

The population of track lengths predicted with a high error is mostly composed by short tracks, clustered in two groups. The first group is a circular pattern corresponding to the TEPC external region but populated with short tracks; this result was expected since lower tracks often yield to high y_T values, and therefore the y_T-weighted MAE might be significantly higher. The other cluster is located in the LGAD central part. Those tracks predicted positions must be incorrect, since it is impossible that a particle walked a short path through the center part of the TEPC.

Metric considerations

The MAPE and MAE errors with different weights have been evaluated for the linear, RF and RF$_0$ models. The values can be found in Tabs.10.2 for protons and for carbon ions.

From these values, ω_{y^2} has been chosen as the best weight because it is the most robust choice, reflecting the discrepancies that emerge in the microdosiemtric spectra reconstruction. Focusing on the weight choice, for both the MAE and the MAPE metric, it is clear that for protons the RF is the best predicting model, while for carbon ions it is the RF$_0$ model. For consistency with the previous study on HDM, we chose to focus on the MAE metric.

To validate the choice on the best prediction model for the two different ion beams, we calculated the density plots of the discrepancies between the model and the linear tracking algorithm, which is used as a benchmark (Figure10.8). In both cases, the area under the positive values of the density functions is bigger than the negative ones. This confirm the fact that the chosen prediction models (RF for protons and RF$_0$ for carbon

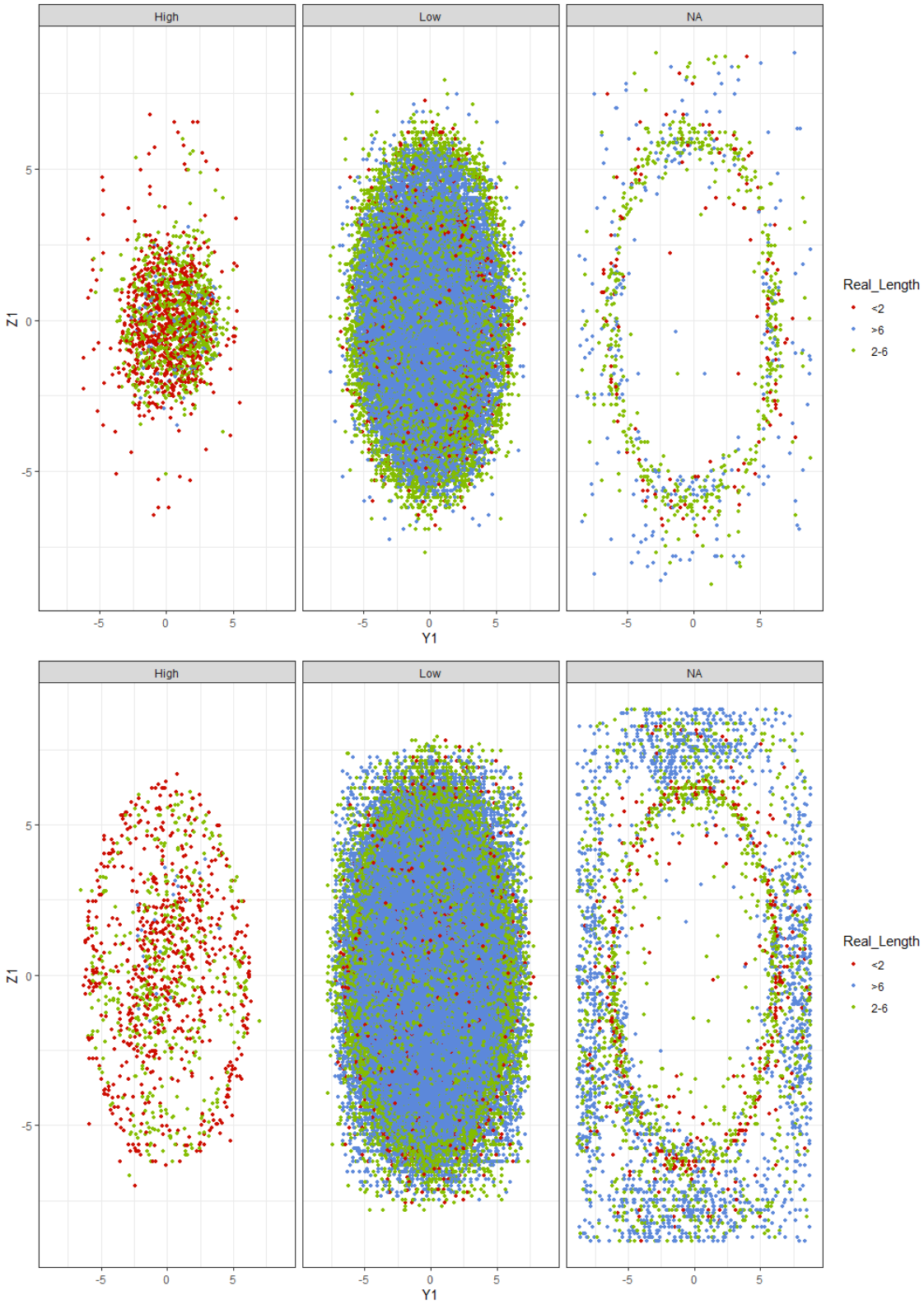

Figure 10.6: Spatial positions of particles passing through the first LGAD couple (either real or predicted by the first ML module). Panels separates high and low $MAE_{\omega_{y^2}}$ of the linear tracking algorithm. The NA panel refers to events with a null linear intersection between the LGAD positions and the TEPC. The colors refer to the real track length values estimated from *Geant4*: pink shows the short tracks (< 2 mm), blue refers to the medium track lengths (from 2 to 6 mm) and green is used for the long tracks (>6 mm). The upper panels show the results for protons, while the bottom panels for carbon ions.

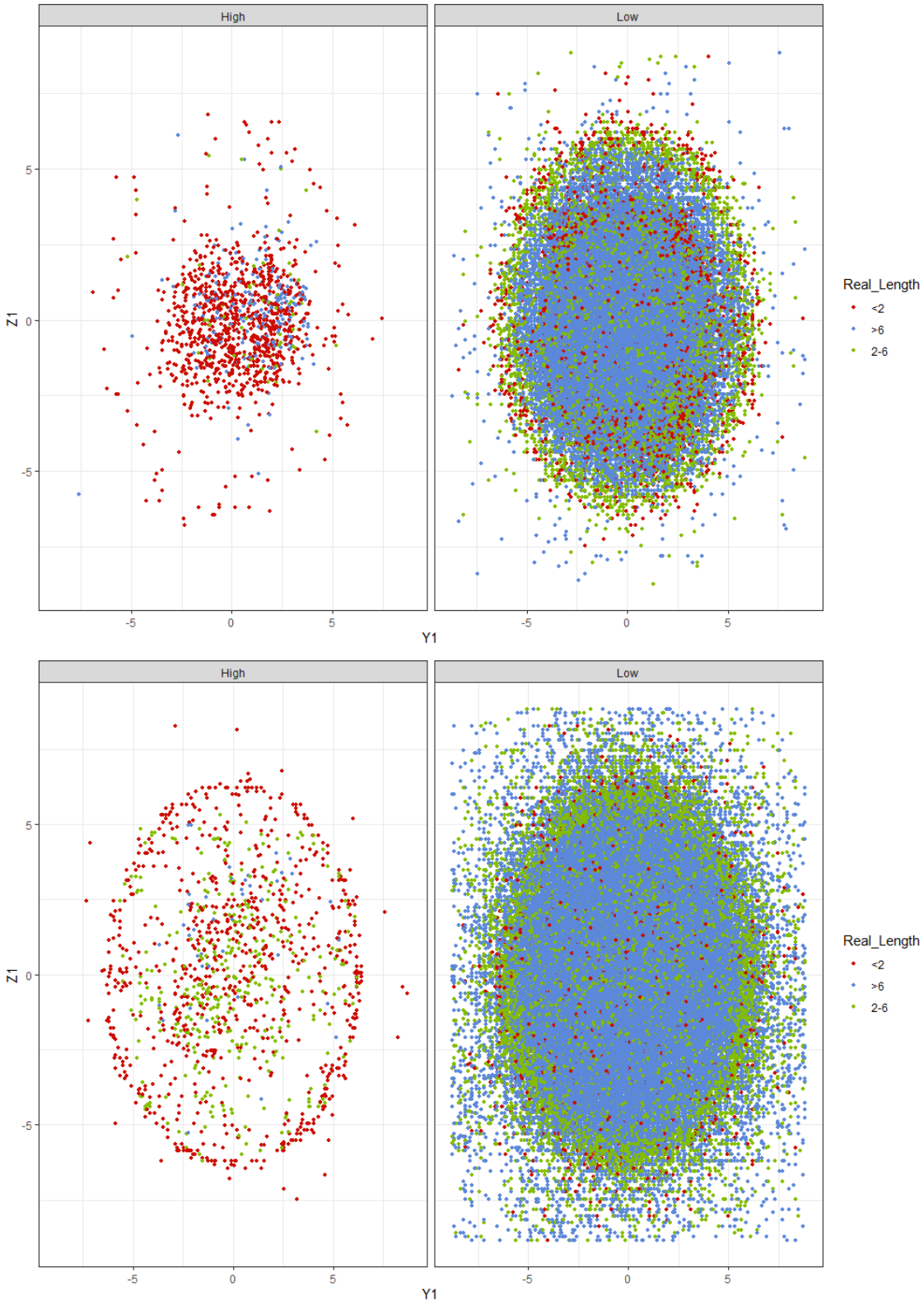

Figure 10.7: Spatial positions of particles passing through the first LGAD couple (either real or predicted by the first ML module). Panels separates high and low $MAE_{\omega_{y2}}$ for the RF tracking algorithm. Colors according to the legend refers to the real track length values estimated from *Geant4*: pink shows the short tracks (< 2 mm), blue refers to the medium track lengths (from 2 to 6 mm) and green is used for the long tracks (>6 mm). Upper panels refer to protons, bottom panels to carbon ions.

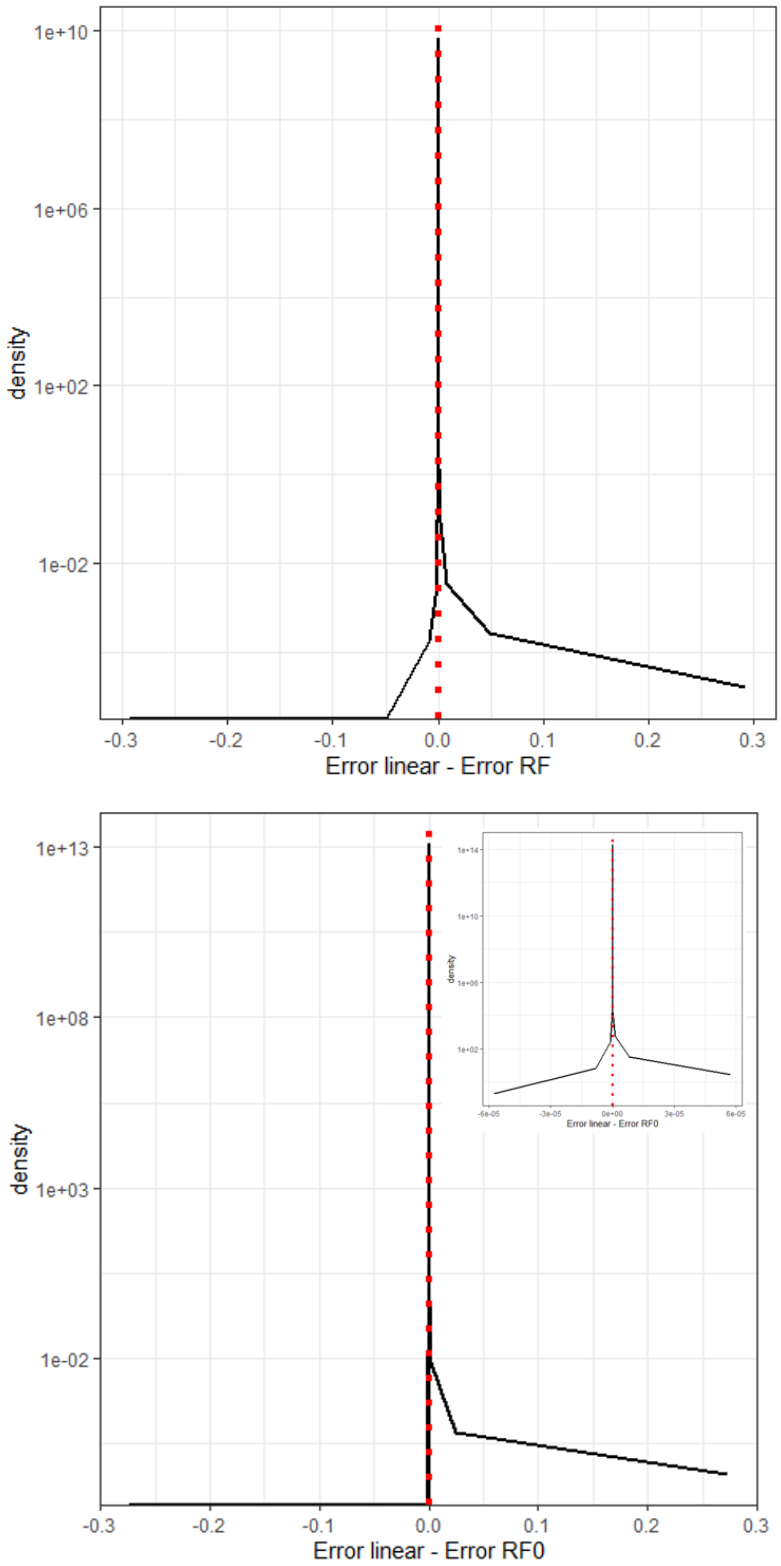

Figure 10.8: Upper panel: density plot of the discrepancies between the MAE error using the linear model and the RF model for the proton beam. Bottom panel: density plot of the discrepancies between the MAE error using the linear model and the RF_0 model for the carbon beam. A zoom of the central part, with errors between $-6\,10^{-5}$ and $6\,10^{5}$ is plotted.

ions) have smaller errors than the linear tracking algorithm in predicting the track lengths. Furthermore, the density plots highlights how the selected model is a better predicting algorithm than the linear model not only on average but also particle by particle.

10.4 Discussion

Missing values analysis

The results on the missing values analysis highlight that two major undetected particles may be due to two main reasons: dead zones and particles scattering in the detector. The former can be improved with the detector optimization, such as reducing the inter-strip widths or changing the LGADs angle with respect to the TECP to increase the active area of the trackers. Instead, decreasing the probability of scattering is less easy, but it would be the most relevant improvement. Together with thinning the silicon layers, a possible solution is to modify the LGADs positions during the irradiation period, in order to collect particles from different angles. Another possible approach is to develop a suitable ML-based module that identifies scattered particles. In doing so, the problem would be split and solved with two different models, the first ML model is trained specifically on scattered events whilst the second one is trained on non scattered ones. Furthermore, it is difficult to distinguish the number of particles lost because of scattering from those hitting the LGAD dead zones. Nonetheless, we can get a hint from Figure10.3: particles that have missed all the layers are unlikely to be traversing only dead zones, while, on the other hand, particles that misses just one layer are most likely not scattered events. In addition, it is worth noticing that whenever an undetected particle traverses just two layers with parallel strips, it is likely that such missing information is due to the dead interstrip regions of the perpendicular LGADs layers rather than a scattering of the particle.

Microdosimetry-based analysis of the tracking ML model

The microdosimetric spectra comparison plotted in Figure 10.4, the mean quantites reported in Tab. 10.1 and the error assessment of Tab. 10.2 all indicate that the RF is the best predictive tracking model for protons, while the RF_0 is the most accurate one for carbon ions. The fact that for protons RF is the best performing model is related to the fact that just 2% of the events are seen by all the LGADs layers. In this case, the training dataset for RF_0 is extremely small and biased by the fact that it is mostly composed by non-scattered events; therefore, a model trained on such a dataset would only seen event that are non representative of the whole population, implying the poor performances exhibited by RF_0 in the case of protons. Further, there is a clear difference between the mean chord length and the real track length spectra, especially for $y_T \geq 10$ keV/μm. The predictions for all the methods are in good agreement with the RTL spectrum in the main peak region, meaning that the primary protons are well reconstructed. This result is expected since these protons are mostly non-scattered particles that follows a straight line direction. This hypothesis is confirmed by the fact that the y_F values are similar for all the models. In the right tail, the method that better reproduce the RTL $y_T d(y_T)$ spectrum is the RF, despite the fact that the reconstructed spectrum exhibits a second peak that can be identified as an artefact. Nonetheless, the RF is clearly the method that predict high y_T particles with the best accuracy, as it is pointed out by the fact that the

y_D value is the closest to the RTL one. In addition, the RF is the model that reproduces better the distribution left rise upstream of the main peak.

For carbon ion, instead, the number of particles seen by all the LGAD is higher. In addition, there is a significant proportion of events that are undetected because they traverse an LGAD dead zone. Both Figs. 10.4 and 10.5 show the same pattern for both ions: the best predicting model overestimates events with tracks of 2-6 mm length, and underestimates those with tracks of 0-2 mm length. Furthermore, 60% of the particles with MAE above the 99 percentile leads to $y_T \geq 10$ keV/μm. In addition, 86% of the samples with $y_T \geq 10$ keV/μm has a reconstruction error above the 99^{th} percentile for protons. These considerations tell us that there is a correspondence between high MAE errors and the extreme right tail of the $y_T d(y_T)$ distribution. On one hand, this result confirms that the chosen metric is a good representative of the error on the reconstructed spectra, and on the other side, it highlights the main issue in reconstructing the microdosimetric distribution. It is clear that it is important to reconstruct the right tail, which is the region where a higher error has been made. In particular, high y_T values cannot be obtained by simply increasing the statistics, and therefore a reconstruction model is necessary to predict such events.

It is worth further stressing that the second peak at high y_T, despite seeming at first to be a huge issue of the model, can be on the contrary exploited to improve the understanding of the ML model. In particular, the second peak in the reconstructed $y_T d(y_T)$ spectrum clearly highlights that the main issue of the currently developed model is the inability to predict 0-2 mm chord lengths. A similar conclusion can be drawn from Figs. 10.6 and 10.7, where the highest tracking errors are associated to extremely small track lengths. These events of short track length are positioned in the central part of the LGADs, which is unrealistic because in the central part of the LGAD track lengths should be comparable to the diameter of the TEPC. Therefore, these events are probably scattered particles that have been wrongly positioned in the LGAD by the first ML module.

In conclusion, we illustrated that advanced ML algorithms can improve both the accuracy and efficiency of particle tracking detectors.

The next step will be to quantitative analyze the effect that the new microdosimetric spectra provided by HDM can have on predicting the radiation biological effects and on the RBE estimate. Today, few radiobiological models can predict a biological endpoint using the entire microdosimetric spectrum, e.g. the *Loncol model*, but they are valid just for specific ion types and biological endpoints. On the contrary, the MKM, which is the most commonly used model in the clinics for predicting RBE, is based on the dose-average value y_D or on its saturation corrected value y^* as described in 2.2. This feature implies that the MKM and its variants are unsuited to consider the RTL spectrum measured by HDM. Recently, we proposed a new model called *Generalized Stochastic Microdosimetric Model* (GSM^2) [Cordoni et al., 2021, Cordoni et al., 2022b, Missiaggia et al., 2022c], that can predict cell survival, and consequently RBE, considering the whole microdosimetric spectrum. For this reason, GSM^2 is particularly suitable to use the microdosimetrc data provided by HDM. GSM^2 mathematical framework and the main characteristics will be presented in the next part.

Part V

The Generalized Stochastic Microdosimetric Model (GSM^2)

Chapter 11

The main setting

11.1 The rationate for a new model

The main goal of the present part is to develop a fully probabilistic model of DNA damage formation and its kinetic evolution based on microdosimetry. The new model, called (*Generalized Stochastic Microdosimetric Model* (GSM^2), will provide a rigorous and general mathematical description of DNA damage time–evolution without using any a priori assumption on the lesion distribution (e.g. a Poisson). GSM^2 will be shown to agree with the MKM predictions in the irradiation conditions where the latter is accurate, but will also provide a more precise and robust description of the cell survival probability where the MKM fails.

To provide a rigorous mathematical formulation of the DNA damage kinetics, we will consider lethal and sublethal lesions inside a single cell nucleus. Potentially lethal lesions can either be repaired or die, in which case they become lethal lesions. A cell in which at least one lethal lesion has been formed is considered inactivated. A potentially lethal damage induced by radiation can undergo three main processes: (i) it can spontaneously repair at a rate r; (ii) it can spontaneously become a lethal damage at a rate a; or (iii) it can combine with another potentially lethal lesion to form a lethal lesion at rate b.

We will introduced an equation, referred to as *Microdosimetric Master Equation* (MME), that governs the time evolution of the joint probability density function for lethal and sub–lethal damages inside the cell nucleus and is based on the parameters a, b and r. It is important to underline in the MME there two aspects in which stochasticity is considered: i) the random lesion formation due to the stochastic nature of energy deposition of ionizing radiation; and ii) the evolution of lethal and lethal damages as well as for the random lesion formation due to the stochastic nature of energy deposition of ionizing radiation. The latter aspect can be precisely characterized by microdosimetric spectra.

The classical approach for mathematically describe a complex physical system, such as the one resulting from the interaction between cells and ionizing radiation that leads to the formation of DNA lesions, is achieved with deterministic models, so that, given an initial condition the system time–evolution can be completely characterized at each given state. Recent studies, [Smith and Grima, 2019], have indicated that this approach fails due to mainly three reasons: i) a precise and accurate estimation of the parameters if often non feasible; ii) it is unrealistic to account for all possible interactions as the system complexity increases; and iii) certain systems can be over-sensitive to some input parameters, typically the initial values. All above reasons have led to include stochasticity in the models via suitable random variables.

For modeling lesions formation following a radiation exposure, the standard method is to consider the *macroscopic* system, which can be fully described by mean values only. On the contrary in a *microscopic* approach, instead, each element of the system is usually modelled using the Brownian dynamics [Van Kampen, 1992, Solov'yov et al., 2009]. Nonetheless, the complexity of the system we want to reproduce makes this strategy not feasible.

To obtain a more general and accurate description of a system than the one provided by a *macroscopic* approach, and yet maintain some reasonable assumptions on the model main aspects, a hybrid methodology, know as *mesoscopic* approach, is typically considered. The latter has proved to be able to take into account the stochastic nature of a system and still being manageable from both the analytical and numerical points of view. The *mesoscopic* method is based on the assumption that the process driving the system evolution is a Markov jump process [Gardiner et al., 1985]. In detail, the equations of motion are described via the so–called *master equation* that contains the probability density function of the whole system [Gardiner et al., 1985, Van Kampen, 1992, Weber and Frey, 2017].

Starting from some probabilistic assumptions on the lesions formation, we will derive a master equation that describes the time evolution for the joint probability density function of DNA lesions, lethal and potentially lethal. The density function solution will be shown to have first moment in agreement with the standard MKM driving equations. The main goal of this study is to overcome the Poissonian assumption on lethal lesions.

In the present work we will further generalize the MME in two main directions. In particular, besides mechanism (i)-(ii)-(iii) in the damage kinetic introduced above, we additionally consider that (iv) either a lethal or sublethal damage can be formed randomly due to the effect of the ionizing radiation at a rate $\dot{d}$ and (v) lethal lesion can move inside the cell nucleus. In fact, (iv) represents DNA damage formation resulting from a continuous irradiation field. In fact, together with standard lesion interactions, we will also take into account random jumps in the number of lethal and sub–lethal lesions caused by the stochastic nature of energy deposition.

Reaction (v) accounts for the the fact that we allow lesions to move inside the domain. Since the model consider pair–wise interaction of potentially lethal lesions. In fact, a domain too big implies that lesions who are created far away from each other can interact to form a lethal lesion. On the contrary, a domain too small results in a very small number of lesions per domain so that the probability of double events is underestimated. The domains size plays a crucial role and to minimize the model dependence on this choice, we will consider interactions between different domains [Smith and Grima, 2019]. An example of this effect is the limit case where the domain size approaches zero and no interaction can occur [Isaacson, 2009, Hellander et al., 2012]. This result follows from the fact that most domains contain a single lesion as their size decrease. Thus, the reaction rates and domain size must be chosen carefully.

Summarizing above discussion, We will introduce a general master equation that models the joint probability distribution of DNA lethal and potentially lethal lesion inside a cell nucleus. The derived master equation will consider, besides potentially lethal lesion repair and death due to either spontaneous dead or pair–wise interaction, also the stochastic effect of energy deposition due to ionizing radiation and lesions movements between adjacent domains, providing a global description of the cell nucleus as a whole.

For above reasons we will name the proposed model *Generalized Stochastic Microdosimetric Model* (GSM^2).

The innovations presented in this work are several. We will develop a fully probabilistic

description of the DNA damage kinetic. In particular, the joint probability distribution of the number of sub-lethal and lethal lesions will be modelled. We will further generalize the model including inter–domain movements and continuous damage formation due to protracted dose. The resulting *master equation* solution will provide the real probability distribution without any a priori assumption on the density function, allowing to compute several biological endpoints. The proposed approach will be able to fully describe the stochastic nature of energy deposition both in time and space, improving the existing models were the energy deposition is averaged over both the whole cell nucleus and cell population. In doing so, we will be able to reproduce several behaviours referred to in literature as *non–Poissonian effects*, that cannot be predicted by the MKM and its variants and are typically included in the models with ad hoc corrections [Kase et al., 2006, Sato and Furusawa, 2012, Hawkins, 2003, Hawkins and Inaniwa, 2014a].

To validate GSM^2, we will consider microdosimetric energy spectra obtained from Geant4 simulations [Agostinelli et al., 2003]. We will show how different assumptions related to the probability distribution of damages number, as well as model parameters, show significant deviation from the Poisson distribution assumed by all existing models, including the MKM. We will further compute the survival probability and compare it to the classical *linear–quadratic* (LQ) model [Bodgi et al., 2016, McMahon, 2018b].

Because of GSM^2 flexibility and generality, analytical solutions both on the probability density function and on the resulting survival curve are not of easy derivation. Therefore, the present study has to be intended as a first step of a systematic investigation of the stochastic nature of energy deposition and how it influences lesions formation. In particular, a further investigation will focus on long–time behaviour of the *master equation* and the resulting survival curve. Furthermore, the principles used in the current approach will be used to develop a fully stochastic model of inter-cellular damage formation optimized to improved radiation field characterization via a novel hybrid detector for microdosimetry, [Missiaggia et al., 2021].

With GSM^2 and its future developments, we try to shade a new light on non–Poissonian effects, to obtain a deeper understanding which will allow us to model them more accurately.

11.2 The Generalized Stochastic Microdosimetric Model

As a part of this study, we investigated how the models described in Section 2.2 could be developed to rely on the whole probability distribution rather than simmply on its mean value. In fact, all proposed generalizations of the MKM always consider deterministic driving equations for predicting the number of lethal and sub-lethal lesions. Non-Poissonian effects are often proposed as corrections terms added to the survival fraction predicted by the MKM with no formal mathematical derivation and mainly based on empirical evaluations.

The MKM formulation is based on the probability distribution of inducing a damage when a specific energy z is deposited. Once the survival for a given z is computed, the specific energy is averaged over the whole cell population to yield the overall expected survival probability. To the best of our knowledge, there is no systematic investigation that aims at capturing the true stochasticity of both energy deposition and lesion formation.

The main goal of the present work is thus to generalize microdosimetric based models in order to describe the full probability distribution of lethal and sub-lethal lesions. We will take advantage of assumptions $1-5$ described in Section 2.2. Regarding assumption 4,

the MKM assumes that the lethal lesions initial distribution, given an energy deposition z, follows a Poisson law. We will generalize this assumption assuming a general initial distribution, allowing to fully describe the stochastcic nature of energy deposition. This point will be treated in detail in Section 11.2.

An additional remark on the importance of the initial distribution is necessary to fully understand the implication of the generalization we will carry out in this study. The stochasticity of energy deposition in a microscopic volume is the basic foundation of microdosimetry, and assuming every probability distribution to be Poissonian is a restrictive assumption that limits the model application.

In order to capture the real stochastic nature of energy deposition and related DNA damage formation we will provide a probabilistic reformulation of Equation (2.16). We denote by $(Y(t), X(t))$ the system state at time t, where X and Y are two $\mathbb{N}_0-$valued random variables representing the number of lethal and sub–lethal lesions, respectively. We will consider a standard complete filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$ that satisfies the usual assumptions of right–continuity and saturation by $\mathbb{P}-$null sets.

Let us consider two different sets $\mathcal{X}$ and $\mathcal{Y}$ denoting the number of type I and type II lesions, respectively. The heuristic interpretation of the coefficients in Equation (2.16) is that a is the rate at which a lesion of type II becomes a lesion of type I, r is the rate at which a lesion of type II recovers and goes to the set $\emptyset$ (i.e. that of the healthy cells), whereas b is the rate at which two lesions interact to become a single type I lesion. These considerations can be mathematically expressed as

$$\begin{aligned} &X \xrightarrow{a} Y\,, \\ &X \xrightarrow{r} \emptyset\,, \\ &X + X \xrightarrow{b} Y\,. \end{aligned} \tag{11.1}$$

Thus, at a given time t, the probability to observe x lesions of type II and y of type I is

$$p(t, y, x) = \mathbb{P}\left((Y(t), X(t)) = (y, x)\right)\,.$$

Also,

$$p_{t_0,y_0,x_0}(t, y, x) := p(t, y, x|t_0, y_0, x_0) = \mathbb{P}\left((Y(t), X(t)) = (y, x)|\,(Y(t_0), X(t_0)) = (y_0, x_0)\right)$$

is the probability conditioned to the fact that at $t = t_0$ there were x_0 and y_0 sub–lethal and lethal lesions, respectively.

To determine the governing master equation for the above probability density $p(t, y, x)$, we need to account for all possible system changes in the infinitesimal time interval dt

Thus, the following scenarios may happen:

(i) at time t we have exactly (y, x) lesions and they remain equal with a rate $(1 - (a + r)x - bx(x - 1))dt$, namely

$$\begin{aligned} &\mathbb{P}\left((Y(t + dt), X(t + dt)) = (y, x)|\,(Y(t), X(t)) = (y, x)\right) = \\ &= 1 - ((a + r)x - bx(x - 1))dt + O(dt^2)\,; \end{aligned}$$

(ii) at time t we have exactly $(y, x+1)$ lesions, and one lesion recovers with rate $(x+1)rdt$, namely

$$\mathbb{P}\left((Y(t+dt),X(t+dt))=(y,x)|\,(Y(t),X(t))=(y,x+1)\right)=(x+1)rdt+O(dt^2)\,;$$

(iii) at time t we have exactly $(y-1,x+1)$ lesions, and one type II lesion becomes of type I with a rate $(x+1)adt$, namely

$$\mathbb{P}\left((Y(t+dt),X(t+dt))=(y,x)|\,(Y(t),X(t))=(y-1,x+1)\right)=(x+1)adt+O(dt^2)\,;$$

(iv) at time t we have exactly $(y-1,x+2)$ lesions, and two type II lesions become one type I with a rate $(x+2)(x+1)bdt$, namely

$$\mathbb{P}\left((Y(t+dt),X(t+dt))=(y,x)|\,(Y(t),X(t))=(y-1,x+2)\right)=(x+2)(x+1)bdt+O(dt^2)\,;$$

Grouping the equations derived in 11.2 we obtain

$$\begin{aligned}p(t+dt,y,x)=&\,p(t,y,x)\left(1-((a+r)x-bx(x-1))dt+O(dt^2)\right)+\\&+p(t,y,x+1)\left((x+1)rdt+O(dt^2)\right)+\\&+p(t,y-1,x+1)\left((x+1)adt+O(dt^2)\right)+\\&+p(t,y-1,x+2)\left((x+2)(x+1)bdt+O(dt^2)\right)\,,\end{aligned}$$

Writing down above relation and taking the limit as $dt\to 0$ we eventually obtain the *microdosimetric master equation* (MME)

$$\begin{aligned}\partial_t p(t,y,x)=&-\left((a+r)x-bx(x-1)\right)p(t,y,x)+(x+1)rp(t,y,x+1)+\\&+(x+1)ap(t,y-1,x+1)+(x+2)(x+1)bp(t,y-1,x+2)\,,\end{aligned}\tag{11.2}$$

where above ∂_t denotes the partial derivative with respect to the first argument of $p(t,y,x)$, that is the time variable. Equation (11.2) must be equipped with suitable initial condition $p(0,y,x)=p_0(y,x)$.

We remark that the above derived MME arises solely from the probabilistic assumptions regarding lesion formation.

The MME (11.2) can be written for short as

$$\begin{aligned}\partial_t p(t,y,x)=&\left(E^{-1,2}-1\right)\left[x(x-1)bp(t,y,x)\right]+\left(E^{-1,1}-1\right)\left[xap(t,y,x)\right]+\left(E^{0,1}-1\right)\left[xrp(t,y,x)\right]=\\&=\mathcal{E}^{-1,2}\left[x(x-1)bp(t,y,x)\right]+\mathcal{E}^{-1,1}\left[xap(t,y,x)\right]+\mathcal{E}^{0,1}\left[xrp(t,y,x)\right]\,,\end{aligned}\tag{11.3}$$

where above we have denoted the creation operators defined as

$$\mathcal{E}^{i,j}\left[f(t,y,x)\right]:=\left(E^{i,j}-1\right)\left[f(t,y,x)\right]:=f(t,y+i,x+j)-f(t,y,x)\,.$$

Connection with the MKM

The present section aims at showing that the mean value of the master equation does satisfy, under certain assumptions, the kinetic Equations (2.16). In what follows, $\mathbb{E}$ denotes the mean value of a random variable defined as

$$\begin{aligned}\bar{x}(t) &:= \mathbb{E}[X(t)] = \sum_{x,y\geq 0} xp(t,y,x)\,,\\ \bar{y}(t) &:= \mathbb{E}[Y(t)] = \sum_{x,y\geq 0} yp(t,y,x)\,.\end{aligned}$$

Note that, for a general function f, the following holds true

$$\begin{aligned}\sum_{x,y\geq 0} x\mathcal{E}^{i,j}\left[f(y,x)p(t,y,x)\right] &= -\mathbb{E}jf(Y,X)\,,\\ \sum_{x,y\geq 0} y\mathcal{E}^{i,j}\left[f(y,x)p(t,y,x)\right] &= -\mathbb{E}if(Y,X)\,.\end{aligned} \tag{11.4}$$

Therefore, multiplying the MME (11.3) by x and y, we obtain using (13.2)

$$\begin{cases}\frac{d}{dt}\mathbb{E}[Y(t)] &= b\mathbb{E}[X(t)(X(t)-1)] + a\mathbb{E}[X(t)]\,,\\ \frac{d}{dt}\mathbb{E}[X(t)] &= -2\mathbb{E}[X(t)(X(t)-1)] - (a+r)\mathbb{E}[X(t)]\,.\end{cases} \tag{11.5}$$

Equations (13.3) are still not of the form of Equations (2.16); in particular they depend on a second order moment $\mathbb{E}[X(t)(X(t)-1)]$. Nonetheless explicit computation will show that, if we try to compute a kinetic equation for the second order moment $\mathbb{E}[X(t)(X(t)-1)]$, we would obtain a dependence on higher moments, and so to obtain an infinite set on coupled ODE. To solve the impasse we shall make what is called a *mean–field* assumption, that is we assume that

$$\mathbb{E}[X(t)(X(t)-1)] \sim \mathbb{E}[X(t)]^2\,.$$

Under the above *mean–field assumption*, Equations (13.3) become

$$\begin{cases}\frac{d}{dt}\bar{y}(t) &= b\bar{x}^2(t) + a\bar{x}(t)\,,\\ \frac{d}{dt}\bar{x}(t) &= -2\bar{x}^2(t) - (a+r)\bar{x}(t)\,,\end{cases} \tag{11.6}$$

and the original kinetic equations are in turn recovered.

A quick remark on the *mean–field assumption* is needed. In the case of x being large enough, we have that the following approximation holds true $\mathbb{E}[X(t)(X(t)-1)] \sim \mathbb{E}[X^2(t)]$; therefore the *mean field assumption* means that $\mathbb{E}[X^2(t)] - \mathbb{E}[X(t)]^2 \sim 0$. Noticing that the last term is nothing but the variance, and recalling that the variance for a random variable is null if and only if the random variable is in fact deterministic, if the *mean field assumption* is realistic than the realized number of lesion does not differ much from the mean value so that everything we need to know is the mean value. On the contrary if there are evidence that the mean value is not a realistic approximation for the realized number of lesion, the *mean–field assumption* must be considered unrealistic so that the knowledge of the full probability distribution is essential to have a complete understanding of the system.

On the initial distribution for the number of lethal and sub-lethal lesions

One of the main advantages of the proposed model is that the distribution of DNA damages induced by an ionizing radiation z does not need to be chosen as Poissonian. In the present section we will show how the number of induced lesions can be evaluated starting from microdosimetric spectra.

Let $f_{1;d}(z)$ be the single–event distribution of energy deposition on a domain d, see [Rossi and Zaider, 1991]. The single–event distribution $f_{1;d}$ can be either computed numerical via Monte Carlo toolkit or by experimental microdosimetric measurements.

The full probability distribution of an energy deposition thus depends on the number of events that deposit energy on the cell nucleus. Given a cell nucleus, composed by N_d domains, the probability that ν events deposit an energy z obeys to a Poissonian distribution of mean $\lambda_n := \frac{z_n}{z_F}$, being z_n the mean energy deposition on the nucleus, i.e.

$$z_n = \int_0^\infty z f(z|z_n) dz \,,$$

and z_F the first moment of the single event distribution $f_{1;d}$ defined as

$$z_F := \int_0^\infty z f_{1;d}(z) dz \,. \tag{11.7}$$

Then, assuming a Poissonian probability that a domain registers ν events, the energy deposition distribution is given by

$$f(z|z_n) := \sum_{\nu=0}^{\infty} \frac{e^{-\frac{z_n}{z_F}}}{\nu!} \left(\frac{z_n}{z_F}\right)^n f_{\nu;d}(z) \,,$$

where $f_{\nu;d}(z)$ is the energy deposition distribution resulting from ν depositions.

In particular, given a domain d suffers ν energy deposition events, the distribution resulting from ν events can be computed convolving ν times the single event distribution, see, [Rossi and Zaider, 1991, Sato and Furusawa, 2012]. Therefore, the imparted energy z has distribution $f_{\nu;d}$, computed iteratively as

$$\begin{aligned} f_{2;d}(z) &:= \int_0^\infty f_{1;d}(\bar{z}) f_{1;d}(z-\bar{z}) d\bar{z} \,, \\ &\dots \,, \\ f_{\nu;d}(z) &:= \int_0^\infty f_{1;d}(\bar{z}) f_{\nu-1;d}(z-\bar{z}) d\bar{z} \,. \end{aligned}$$

For a certain energy deposition z, the induced number of lesions is a random variable. The standard assumption is that the distribution of X given z is a Poisson random variable of mean value κz. Analogous reasoning holds for Y, being the number of induced lesion given z a Poisson random variable of mean λz. Given the high–flexibility of the proposed approach, the number of induced lesions given an energy deposition z can be any random variables. It is worth stressing that the chosen distribution may vary with LET.

In the following general treatment we will denote by $p_z^X(x|\kappa z)$, resp. $p_z^Y(y|\lambda z)$, the initial random distribution for the number of sub–lethal, resp. lethal, lesions given an energy deposition z. We remark again that both $p_z^X(x|\kappa z)$ and $p_z^Y(y|\lambda z)$ can be any probability distribution. Specific relevant examples will be considered in the numerical implementation.

Putting all the above reasoning together, the MME (11.3) reads

$$\begin{cases} \partial_t p(t,y,x) & = \mathcal{E}^{-1,2}\left[x(x-1)bp(t,y,x)\right] + \mathcal{E}^{-1,1}\left[xap(t,y,x)\right] + \mathcal{E}^{0,1}\left[xrp(t,y,x)\right] , \\ p(0,y,x) & = p_0^X(x)p_0^Y(y) , \end{cases} \tag{11.8}$$

where the initial distribution is obtained as

$$\begin{aligned} p_0^X(x) &= \int_0^\infty p_z^X(x|\kappa z) f(z|z_n) dz , \\ p_0^Y(y) &= \int_0^\infty p_z^Y(y|\kappa z) f(z|z_n) dz . \end{aligned} \tag{11.9}$$

Assuming $p_z^X(x|\kappa z)$ to be a Poisson random variable of mean κz, the first moment of the distribution $p_0^X(x)$ can be computed as

$$\begin{aligned} &\sum_{x\geq 0} x p_0^X(x) = \\ &= \sum_{x\geq 0} x \int_0^\infty p_z^X(x|\kappa z) f(z|z_n) dz = \\ &= \int_0^\infty f(z|z_n) \sum_{x\geq 0} x e^{-\kappa z} \frac{(\kappa z)^x}{x!} dz = \\ &= \int_0^\infty \kappa z f(z|z_n) dz = \kappa z_n . \end{aligned}$$

It further follows that

$$\begin{aligned} &\sum_{x\geq 0} x(x-1) p_0^X(x) = \\ &= \sum_{x\geq 0} x(x-1) \int_0^\infty p_z^X(x|\kappa z) f(z|z_n) dz = \\ &= \int_0^\infty f(z|z_n) \sum_{x\geq 0} x(x-1) e^{-\kappa z} \frac{(\kappa z)^x}{x!} dz = \\ &= \int_0^\infty \kappa z^2 \kappa^2 f(z|z_n) dz = \kappa^2 M_2 , \end{aligned}$$

being M_2 the second moment of the distribution $f(z|z_n)$. Using [Rossi and Zaider, 1991, Section II.2], we have

$$M_2 = z_n(z_n + z_D) , \quad z_D = \frac{m_2}{m_1} ,$$

with m_n the $n-$th moment of the single event distribution $f_{1;d}$. Here we have used that $m_1 = z_F$.

Thus, the variance of $p_0^X(x)$ is given by

$$\begin{aligned}
&\sum_{x\geq 0} x^2 p_0^X(x) - \left(\sum_{x\geq 0} x p_0^X(x)\right)^2 = \\
&= \sum_{x\geq 0} x(x-1) p_0^X(x) + \\
&+ \sum_{x\geq 0} x p_0^X(x) - \left(\sum_{x\geq 0} x p_0^X(x)\right)^2 = \\
&= \kappa^2 M_2 + \kappa z_n - \kappa^2 z_n^2 = \\
&= \kappa^2 z_n z_D + \kappa z_n = \vartheta(\kappa, z_n, z_D) + \kappa z_n \,.
\end{aligned} \tag{11.10}$$

The variable $p_0^X(x)$ is Poissonian distributed if the first two moments are equal, i.e. $\kappa z_n = \kappa^2 z_n z_D + \kappa z_n$. If $\kappa^2 z_n z_D << 1$, which is true for low–LET radiation, then $p_0^X(x)$ follows a Poisson distribution, but in general for high–LET radiation the distribution significantly deviates from a Poisson law. Even if $p_0^X(x)$ is Poisson, the long–time lethal lesion distribution might be different from a Poissonian distribution due to non–linear effects, such as the double lesions combination.

The protracted dose case for the Generalized Stochastic Microdosimetric Model

The MME can be further generalized to consider *protracted dose* irradiation. We refer to *protracted dose* as a continuous dose delivery in time. On the contrary a fixed in time, asymptotically short impulse-like dose irradiation is called *acute dose irradiation*, whereas a series of acute irradiations at prescribed timesteps is referred to as *split dose irradiation*. Existing models fail at properly describing protracted dose, being unable to fully capture the stochasticity inherent to energy deposition. Usually, strong assumptions are used to treat protracted dose, [Hawkins, 1996], or a split dose is used to approximate a continuous dose delivery, [Inaniwa et al., 2013]. Nonetheless, models cannot fully predict experimental data, [Inaniwa et al., 2013].

The generalization of the GSM^2 master Equation (11.3) to consider a continuous dose irradiation is not trivial. In fact, at random time t the number of lesions, either lethal or sublethal, exhibits a jump upward of a random quantity that depends on the energy deposition z, that we recall is a random variable.

More formally, the possible interactions now become

$$\begin{aligned}
&X \xrightarrow{a} Y \,, \\
&X \xrightarrow{r} \emptyset \,, \\
&X + X \xrightarrow{b} Y \,. \\
&X \xrightarrow{\dot{d}} X + Z_\kappa \,. \\
&Y \xrightarrow{\dot{d}} Y + Z_\lambda \,,
\end{aligned}$$

being Z_λ and Z_κ two random variables with integer–valued distributions p_0^X and p_0^Y respectively, defined as in Equation (11.8). The parameter $\dot{d}$ represents the dose rate, see,

[Hawkins and Inaniwa, 2014a, Hawkins and Inaniwa, 2014b], and it is given by $\dot{d} := \frac{z_n}{T_{irr} z_F}$, being z_F given in Equation (11.7) and T_{irr} is the total irradiation time.

(i)

$$\begin{aligned}&\mathbb{P}\left((Y(t+dt), X(t+dt)) = (y,x) | (Y(t), X(t)) = (y,x)\right) = \\ &= 1 - ((a+r)x + bx(x-1) + \dot{d}(1 - p_0^X(0))(1 - p_0^Y(0)))dt + O(dt^2)\,;\end{aligned}$$

(ii)

$$\begin{aligned}&\mathbb{P}\left((Y(t+dt), X(t+dt)) = (y,x) | (Y(t), X(t)) = (y - i_y, x - i_x)\right) = \\ &= \dot{d} p_0^X(i_x) p_0^Y(i_y) dt + O(dt^2)\,, \quad i_x = 1, \ldots, x\,, i_y = 1, \ldots, y\,,\end{aligned}$$

Further, reactions (ii), (iii) and (iv) in Section 11.2 remain valid.

Therefore, a similar analysis to the one carried out in Section 11.2 leads to the following MME

$$\begin{aligned}\partial_t p(t,y,x) &= \left(E^{-1,2} - 1\right)\left[x(x-1)bp(t,y,x)\right] + \left(E^{-1,1} - 1\right)\left[xap(t,y,x)\right] + \\ &+ \left(E^{0,1} - 1\right)\left[xrp(t,y,x)\right] + \left(\sum_{i_x=1}^{x}\sum_{i_y=1}^{y} E_{\dot{d}}^{-i_y,-i_x} - (1 - p_0^X(0))(1 - p_0^Y(0))\right)\left[\dot{d}p(t,y,x)\right] = \\ &= \mathcal{E}^{-1,2}\left[x(x-1)bp(t,y,x)\right] + \mathcal{E}^{-1,1}\left[xap(t,y,x)\right] + \\ &+ \mathcal{E}^{0,1}\left[xrp(t,y,x)\right] + \mathcal{E}_{\dot{d}}^{-y,-x}\left[\dot{d}p(t,y,x)\right]\,.\end{aligned} \tag{11.11}$$

The operator in the last line of Equation (11.11) right end side has been defined as

$$\begin{aligned}\mathcal{E}_{\dot{d}}^{-y,-x} f(t,y,x) &:= \left(\sum_{i_x=1}^{x}\sum_{i_y=1}^{y} E_{\dot{d}}^{-i_y,i_x} - (1 - p_0^X(0))(1 - p_0^Y(0))\right) f(t,y,x) = \\ &= \sum_{i_x=1}^{x}\sum_{i_y=1}^{y} p_0^X(i_x) p_0^Y(i_y)\, f(t, y - i_y, x - i_x) - (1 - p_0^X(0))(1 - p_0^Y(0))\, f(t,y,x)\,.\end{aligned}$$

The protracted dose is assumed to be delivered up to a finite time $T_{irr} < \infty$, beyond which no irradiation is considered and the systems evolves according to (11.3).

The diffusive cell nucleus model for GSM^2

In Section 11.2, we investigated the time evolution for lethal and sub–lethal lesions in the cell nucleus.

As we discussed above, one of the major weaknesses of the standard MKM and its extensions is the choice of the cell domains [Smith and Grima, 2019]. In fact, too small domains translate in a null probability of double events, whereas too big domains imply that distant lesions may combine to produce a lethal lesion. To overcome this problem, the cell nucleus is split into several domains so that the time evolution in each domain can be considered independently. Further, following treatment's aim is to encompass above limitations, allowing domains interaction and variability in shape and dimension.

In the current Section, we will show how the MME (11.11) can be extended to include interactions between the domains. In order to keep the treatment as clear as possible, no protracted dose will be consider. The general case of a continuous irradiation can easily included in the following treatment via arguments analogous to the ones used in Section 11.2.

Let us consider N_d domains (referred to also as voxels) that can undergo one of the following possible reactions

$$\begin{aligned} &X_i \xrightarrow{a} Y_i\,, \quad i = 1, \dots, N_d\,, \\ &X_i \xrightarrow{r} \emptyset\,, \quad i = 1, \dots, N_d\,, \\ &X_i + X_i \xrightarrow{b} Y_i\,, \quad i = 1, \dots, N_d\,. \end{aligned} \tag{11.12}$$

A reasoning analogous to the one carried out in Section 11.2 leads to the following MME

$$\partial_t p(t,y,x) = \sum_{i=1}^{N_d} \mathcal{E}_i^{-1,2}\left[x_i(x_i-1)bp(t,y,x)\right] + \sum_{i=1}^{N_d}\left(\mathcal{E}_i^{-1,1}\left[x_i a p(t,y,x)\right] + \mathcal{E}_i^{0,1}\left[x_i r p(t,y,x)\right]\right). \tag{11.13}$$

In Equation (11.13), the variables x and y are $N-$ dimensional vectors with $i-$th component given by x_i and y_i, representing the number of sub–lethal or lethal lesions, respectively, within the $i-$th domain ($i = 1, \dots, N$).

Remark 11.2.1. In order to keep the notation as simple as possible, in Equation (11.12) we chose the rates a, b and r independent of the domain. Similar results would be obtained with voxel-dependent rates a_i, b_i and r_i, $i = 1, \dots, N_d$.

Empirical evidence shows that the lesions, together with interacting within the same voxel, may also move to a different voxel. In fact, lesion spatial movements inside a cell has been demonstrated to be significantly higher than the typical voxel size [Schettino et al., 2011] . To account for this behaviour, we will add an additional term to the MME (11.13).

Besides reactions considered in Equation (11.12), we now assume further the following

$$\begin{aligned} &X_i \xrightarrow{\kappa^X_{i;j}} X_j\,, \quad i,\, j = 1, \dots, N_d\,, \\ &Y_i \xrightarrow{\kappa^Y_{i;j}} Y_j\,, \quad i,\, j = 1, \dots, N_d\,. \end{aligned} \tag{11.14}$$

Remark 11.2.2. We assumed possible interactions also between non adjacent domains. If the reactions described by Equation (11.14) are to be intended as lesions movements inside the cell nucleus, the most reasonable choice for the interaction rates is to set

$$\kappa^X_{i;j} = \kappa^Y_{i;j} = 0\,,$$

for $j \notin \Gamma_i$, being Γ_i the set of adjacent domains to i.

Following the same process described in in Section 11.2, we obtain the MME

$$\begin{aligned} \partial_t p(t,y,x) = &\sum_{i=1}^{N} \mathcal{E}_i^{-1,2}\left[x_i(x_i-1)bp(t,y,x)\right] + \sum_{i=1}^{N}\left(\mathcal{E}_i^{-1,1}\left[x_i a p(t,y,x)\right] + \mathcal{E}_i^{0,1}\left[x_i r p(t,y,x)\right]\right) + \\ &+ \sum_{i,j=1}^{N} {}^X\mathcal{E}_{i,j}^{-1,1}\left[x_i \kappa^X_{i;j} p(t,y,x)\right] + \sum_{i,j=1}^{N} {}^Y\mathcal{E}_{i,j}^{Y;-1,1}\left[y_i \kappa^Y_{i;j} p(t,y,x)\right]\,, \end{aligned} \tag{11.15}$$

where the operators are defined as

$$
\begin{aligned}
{}^{X}\mathcal{E}_{i,j}^{-1,1} f(t,y,x) &= \left(E_i^{0,1} E_j^{0,-1} - 1\right) f(t,y,x)\,, \\
{}^{Y}\mathcal{E}_{i,j}^{-1,1} f(t,y,x) &= \left(E_i^{1,0} E_j^{-1,0} - 1\right) f(t,y,x)\,.
\end{aligned}
$$

The first two lines of Equation (11.15) accounts for reactions within the same voxel, whereas the last line described movements between adjacent domains.

Using the same approach for modeling the initial damage distribution (Section 11.2) the resulting MME reads

$$
\begin{cases}
\partial_t p(t,y,x) & = \sum_{i=1}^{N_d} \mathcal{E}_i^{-1,2}\left[x_i(x_i-1)bp(t,y,x)\right] + \\
& + \sum_{i=1}^{N_d} \left(\mathcal{E}_i^{-1,1}\left[x_i a p(t,y,x)\right] + \mathcal{E}_i^{0,1}\left[x_i r p(t,y,x)\right]\right) + \\
& + \sum_{i,j=1}^{N_d} {}^{X}\mathcal{E}_{i,j}^{-1,1}\left[x_i \kappa_{i;j}^{X} p(t,y,x)\right] + \sum_{i,j=1}^{N_d} {}^{Y}\mathcal{E}_{i,j}^{Y;-1,1}\left[y_i \kappa_{i;j}^{Y} p(t,y,x)\right]\,, \\
p(0,y,x) & = \prod_{i=1}^{N} p_{0;i}^{X}(x_i) p_{0;i}^{Y}(y_i)\,,
\end{cases}
\tag{11.16}
$$

where $p_{0;i}^{X}(x_i)p_{0;i}^{Y}(y_i)$ denotes the initial distribution for the voxel i as computed in Equations (11.8)–(11.9).

11.3 Numerical implementation

To calculate a numerical solution to the MME (11.3), the following steps are performed:

1. We choose the number N_d of domains in which the cell nucleus is divided. As GSM^2 does not rely on any specific assumption for the probability distribution, the domains do not need to be assumed of equal size. For each domain, the *single event* energy deposition distribution $f_{1;d}(z)$ is obtained with Geant4 [Agostinelli et al., 2003] simulations.

2. The number of lethal and sub-lethal lesions are sampled from the distributions $p_0^X(x)$ and $p_0^Y(x)$ as derived in Equation (11.9). The standard assumption is that p_z^X, resp. p_z^Y, is a Poisson distribution of mean κz_d, resp. λz_d. Given the general setting, we will compare the results with an initial Gaussian distribution of different possible variances.

3. Given the initial number of lesions, the evolution paths are simulated via the *stochastic simulation algorithms* (SSA) [Weinan et al., 2019, Chapter 13].

4. Steps 1-3 are repeated to obtain the Monte Carlo empirical distribution of lethal and sublethal lesions over the cell nucleus;

5. The survival probability in the single domain as well as the cell nucleus are calculated from the empirical distribution obtained in step 4;

Previous steps can be computed independently for each domain if no interaction between domains is assumed or the paths for the whole nucleus can be estimated simultaneously, in case of a dependent-voxel model. The computational effort for the latter is substantially higher. It should be noted here that developing an efficient simulation algorithm is beyond the aim of the present work and we refer to [Simoni et al., 2019] for a review of possible simulation algorithms.

The numerical solution

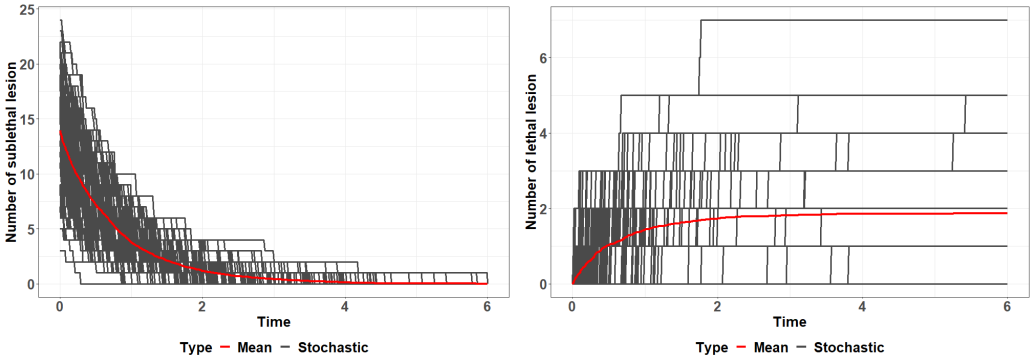

Figure 11.1: Sub-lethal lesions (left panel) and lethal lesions (right panel) evolution. GSM^2 parameters were set to $r = 1$, $a = 0.1$ and $b = 0.01$. The red line represents the average value.

The present Section is devoted to finding and discussing the numerical solution of MME derived in Section 11.2. In particular, the full master Equation (11.3) is solved via the *stochastic simulation algorithms* (SSA) [Weinan et al., 2019, Chapter 13], so that the density is estimated with a Monte Carlo simulation. We simulate 10^6 events and the density function is thus reconstructed empirically.

The goal of this Section is also to highlight how a different setting affects the lesions density distribution. In particular, it will emerge how the density distribution resulting from the corresponding master equation changes for different lesion evolution parameters, initial probabilistic conditions or also irradiation conditions.

To assess the energy deposited on the domain, we used the microdosimetry approach as discussed in Section 11.2. With Geant4, we simulated microdosimetric spectra of a 20 MeV/u carbon ion beam traversing a 1.26 cm diameter sphere filled with pure propane gas with a low density ($1.08 * 10^{-4} g/cm^3$), such that the energy depositions are equivalent to those in 2 μm of tissue. This geometry reproduces a standard Tissue Equivalent Proportional Counter (TEPC) as used for example in [Missiaggia et al., 2020]. Specific energies acquired with the TEPC are then converted to the domain size of interest as reported in [Bellinzona et al., 2021, Section 2]. The choice to simulate a microdosimeter has been made with the aim of remaining as consistent as possible with real experiments. In addition, carbon ions have been chosen since existing model fails at predicting relevant radiobiological endpoints under high-LET regimes.

In the calculations, we consider high doses, so that multi-event distributions as described in Section 11.2 are computed for $z_n >> 1$. This choice is due to the fact that the plotted distributions refer to a single cell nucleus domain and thus, to highlights differences at such a small scale, high dose needs to be considered. At lower doses, differences between the MME solution for a single nucleus domain for different parameters are more difficult to appreciate. Nonetheless, small differences at the domain level can translates into relevant dissimilarities at the macroscopic level.

Figure 11.1 reports different path realizations for the lethal and sub–lethal evolution; the stochastic paths are also compared to the mean value, that evolves according to the MKM kinetic Equations (2.16). The plots indicate that the mean value can not be representative of the whole path realizations distribution.

Figure 11.2 shows the master equation solution at different times. The left panels show

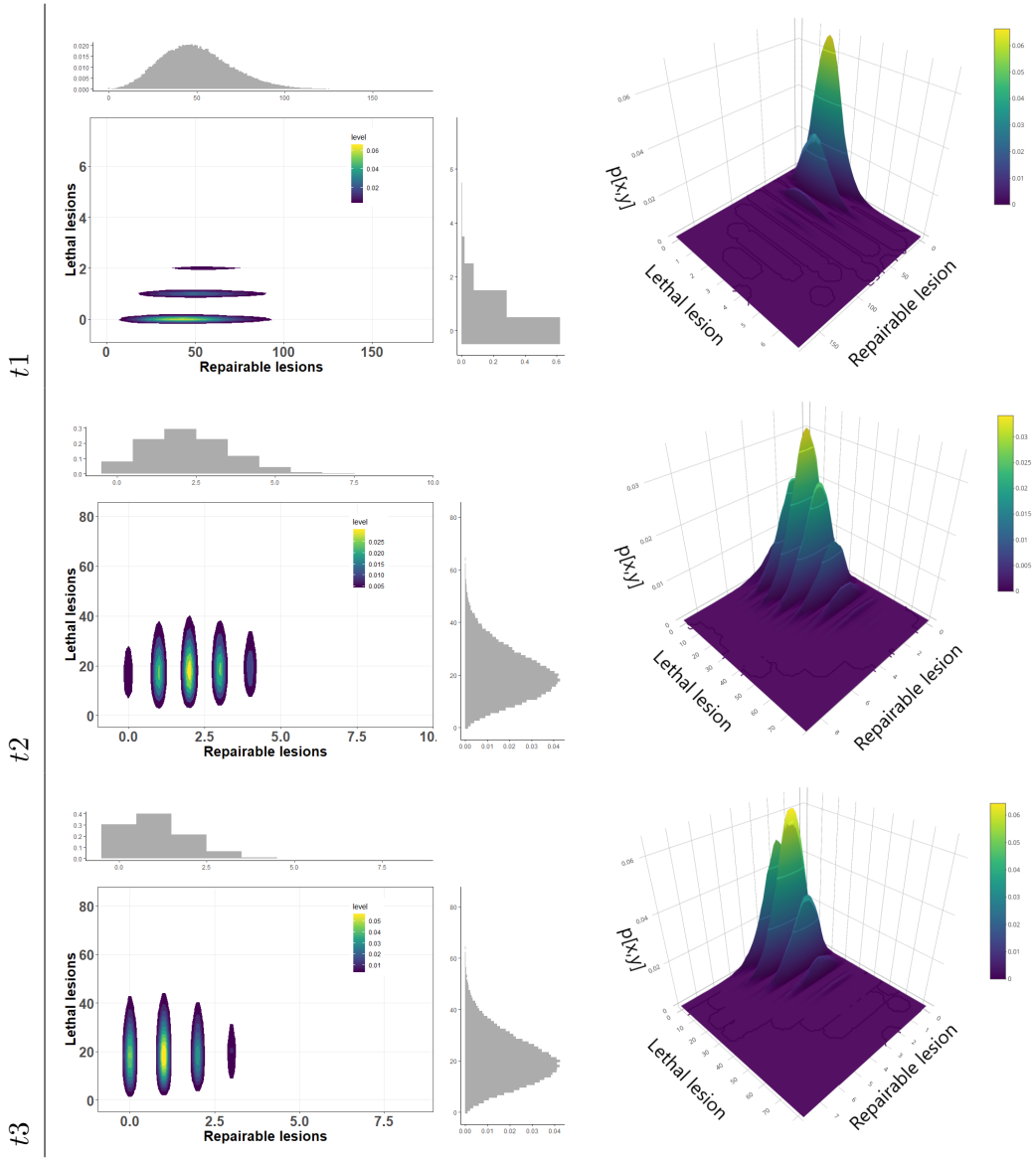

Figure 11.2: Master equation solution at time $t = 1m$ (top panel), $t = 100$ arb. unit (middle panel) and $t = 150$ arb. unit. GSM^2 parameters were set to $r = 1$, $a = 0.2$ and $b = 0.1$.

the contour plots of the joint probability distributions of lethal and sub–lethal damages, together with their marginal distributions. The rights panels are 3D representations of the density function solutions. At a starting time t_1, there is a high variability in the number of reparable lesions while small fluctuations are present in the number of lethal lesions. At a later time t_3, instead, the situation is the exact opposite, with a greater variability in the number of lethal lesions against small fluctuations in the number of sub–lethal lesions.

Figure 11.3 compares lethal and sub-lethal lesion distributions for different types of irradiation conditions, namely acute dose delivery at initial time, split dose at uniform time steps and protracted dose according to Equation (11.11). A split dose at uniform times yields a rather similar lesion distribution as a fully stochastic protracted dose irradiation, while the solution differ significantly for the acute dose case. This result is caused by the

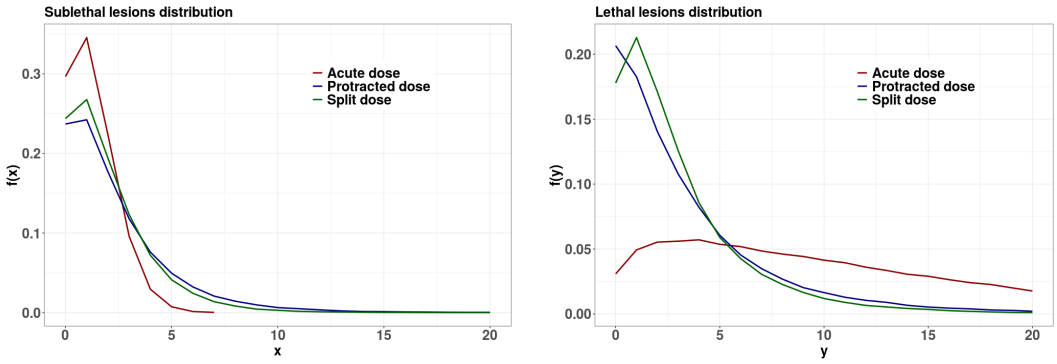

Figure 11.3: Master equation solution for acute, split and protracted doses of 100 Gy. GSM^2 parameters were set to $r = 1$, $a = 0.2$ and $b = 0.1$.

non-linear effect that double events have on the lesions probability distribution.

The long time distribution of lethal lesions is compared with a Poisson distribution for different parameters and doses in Figure 11.4. At lower doses and for b negligible with respect to r, the MME solution is in fact Poissonian (top panel). As the dose increases, the MME solution can be non poissonian even if r dominates b (middle panel). Finally, for higher doses and higher b, the MME solution differs significantly from a Poisson distribution (bottom panel).

Effect of the initial law on the lethal lesions distribution and cell survival

The goal of the present Section is to emphasize the dependence on the initial law of the long–time lethal lesions distribution, showing that the lethal lesions marginal distribution might differ from the Poisson distribution that is typically assumed.

We considered different initial conditions for Equation (11.9). In particular, the following initial distributions were selected for $p_z^X(x|\kappa z)$ and $p_z^Y(y|\lambda z)$: i) a Poisson random variable with mean value μ; and ii) a Gaussian with mean value μ and variance between 0.5μ and 1.5μ. The mean value μ has been set to λz for sub–lethal lesions and κz for lethal lesions. The results are plotted in Figure 11.5 and indicate that a more peaked initial distribution correspond to a more peaked long–time distribution, meaning that the initial value can sharpen or broaden lethal and sub–lethal lesion distributions. This effect has a straightforward consequence on the resulting survival probability shown in Figure 11.6.

we test both the typically used Poisson initial distribution and a Gaussian random variable with different variance.

Figure 11.5 shows the comparison of lethal and sub–lethal lesion distributions for different initial conditions. In particular, initial datum has been taken to be a Poisson random variable with mean value μ. Additionally, the case of an initial distribution to be Gaussian and with mean value μ and variance 0.5μ and 1.5μ has been considered. The mean value μ has been set to λz for sub–lethal lesions and κz for lethal lesions. It can be seen how the initial value can sharpen or broaden lethal and sub–lethal lesion distributions, with a straightforward consequence on the resulting survival probability, see Figure 11.6.

Survival probability is one of the most used and relevant radiobiological observables. Figure 11.6 highlights how a different initial condition affects the resulting survival curve. In particular, it is important to notice that the probability of survival rises or falls in the

high dose region. One of the major flaws in classical models, with particular reference to the linear–quadratic model, is the fact that it significantly underestimates the probability of survival for high doses.

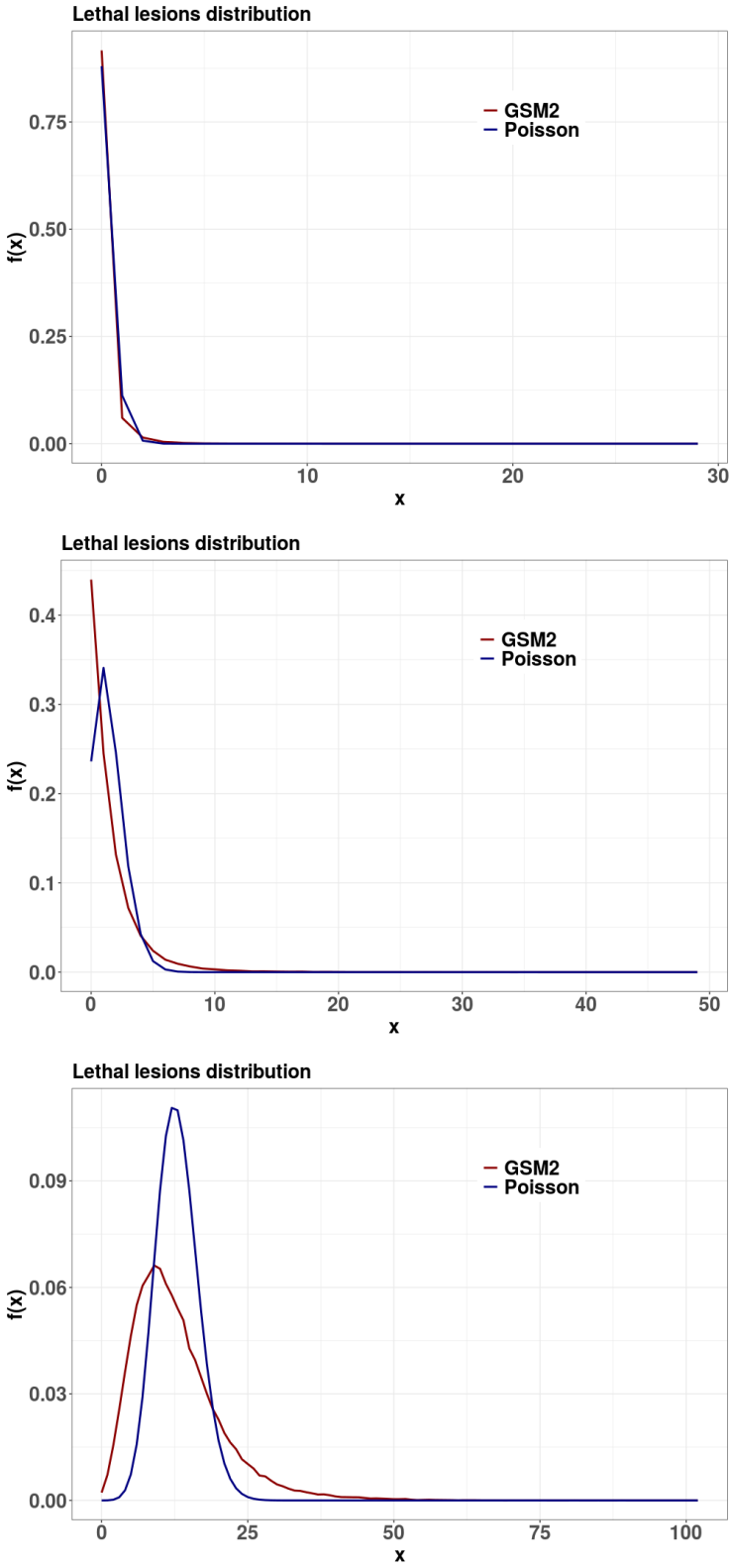

Figure 11.4: Comparison of long–time lethal lesion distributions and Poisson distributions. Top panel: dose=5 Gy, $r = 1$, $a = 0.1$ and $b = 0.01$. Middle panel: dose=100 Gy, $r = 5$, $a = 0.1$ and $b = 0.01$. Bottom panel: dose=150 Gy, $r = 5$, $a = 0.2$ and $b = 0.1$.

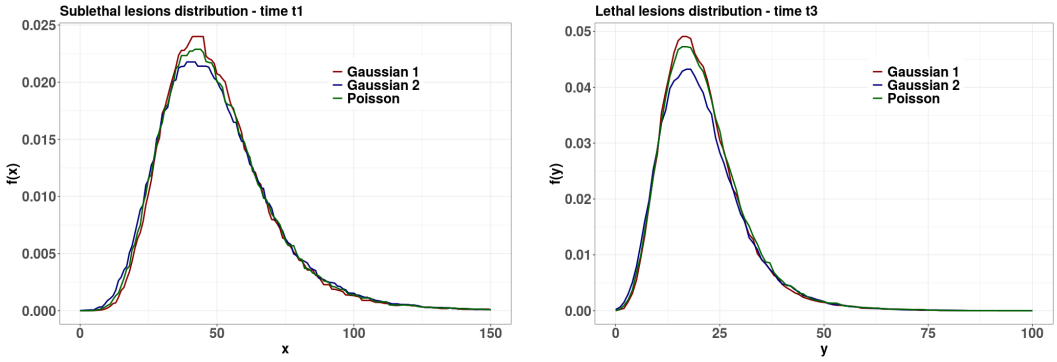

Figure 11.5: Lethal and sub-lethal lesions distribution depending on the chosen initial distribution at time $t1 = 1$ arb. unit and $t3 = 150$ arb. unit. The initial distributions p_z^X and p_z^Y have been chosen as a Poisson distribution of mean $\mu = \lambda, \kappa z$ or as a Gaussian distribution with mean $\mu = \lambda, \kappa z$ and variance $\sigma^2 \in \{0.5\mu, , 1.5\mu\}$. The MME parameters were set to $r = 1$, $a = 0.2$ and $b = 0.1$.

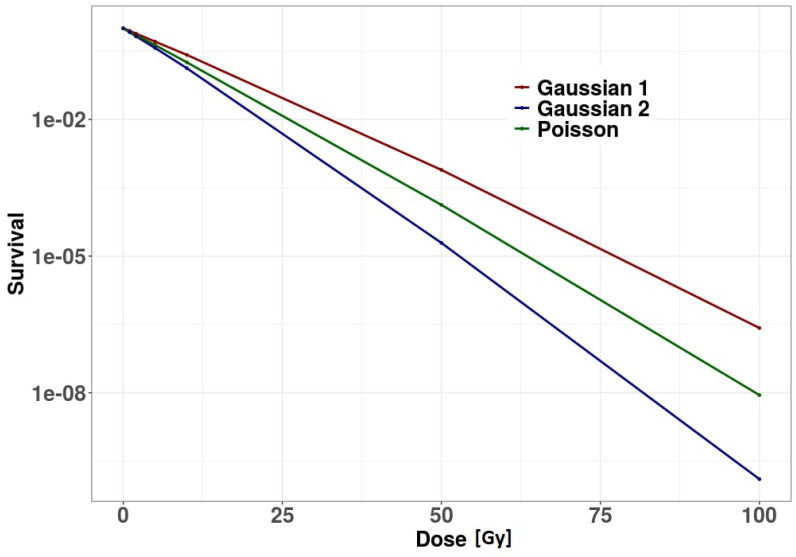

Figure 11.6: Cell survival function calculated for different initial conditions. The initial distributions p_z^X and p_z^Y have been chosen as a Poisson distribution of mean $\mu = \lambda, \kappa z$, or as a Gaussian distribution with mean $\mu = \lambda, \kappa z$ and variance $\sigma^2 \in \{0.5\mu, , 1.5\mu\}$. The MME parameters were set to $r = 1$, $a = 0.2$ and $b = 0.1$.

Chapter 12

Cell survival curve predictions with GSM^2

Exploiting the findings reported in chapter 11, the current section further explores the stochastic effects related to energy deposition and derive an explicit expression for the survival curve.
Simulating energy deposition by different particle beams using TOPAS, [Zhu et al., 2019], it is investigated how the survival curve changes for different ions and different energies. Particular attention is dedicated to higher LET particle for which the existing models fail to accurately predict the cell survival curve. Mixed radiation fields are also considered showing that GSM^2 flexibility allows to model a wide variety of irradiation situations.

12.1 Theory and calculations

Survival curve computation

An estimate of the survival probability following a radiation exposure can be obtained with GSM^2. We will describe the calculation process for a single domain, and then extend it to the entire cell population. The survival probability can be calculated by solving the MME (11.3):

$$S_d(z_n) := \mathbb{P}\left(\lim_{t\to\infty} Y(t) = 0\right) , \tag{12.1}$$

where the dependence of S_d on the energy deposition z_n is explicitly reported.

Assuming that the domains are independent of each other, the survival of a single cell can be obtained as

$$S_n(z_n) := (S_d(z_n))^{N_d} = \left(\mathbb{P}\left(\lim_{t\to\infty} Y(t) = 0\right)\right)^{N_d} . \tag{12.2}$$

In the present Chapter, the domains are considered independent, as it is typically assumed in similar studies. Nonetheless, as we demonstrated in Chapter 11, this assumption can be dropped to allow for diffusive movements of the lesions within the entire cell nucleus.

Thus, a multi-event distribution on the whole cell population can be defined as

$$f_n(z_n|D) := \sum_{\nu=0}^{\infty} p_n(\nu|D) f_{\nu;c}(z_n) ,$$

where $f_{\nu;c}(z_n)$ is the energy deposition distribution resulting from ν depositions in a single nucleus and $p_n(\nu|D)$ is the probability that ν energy depositions occur in a nucleus. The $f_{\nu;c}(z_n)$ distribution can be obtained convolving the single event distribution $f_{1;c}(z_n)$, as described in Section 11.2.

As

$$D = \int_0^\infty z_n f_n(z_n|D) dz_n \,,$$

the *cell survival probability* for the whole cell population can be computed as

$$S(D) = \int_0^\infty S_n(z_n) f_n(z_n|D) dz_n \,. \tag{12.3}$$

Equation (12.3) provides the cell survival probability for a macroscopic dose D.

Therefore, the survival probability for the domain d, in agreement with Equation (12.1), is defined as the quantity

$$S_d(z_n) := \mathbb{P}\left(\lim_{t\to\infty} Y(t) = 0\right) . \tag{12.4}$$

Figure 12.1 reports a schematic representation of GSM^2 cell-survival assesment workflow.

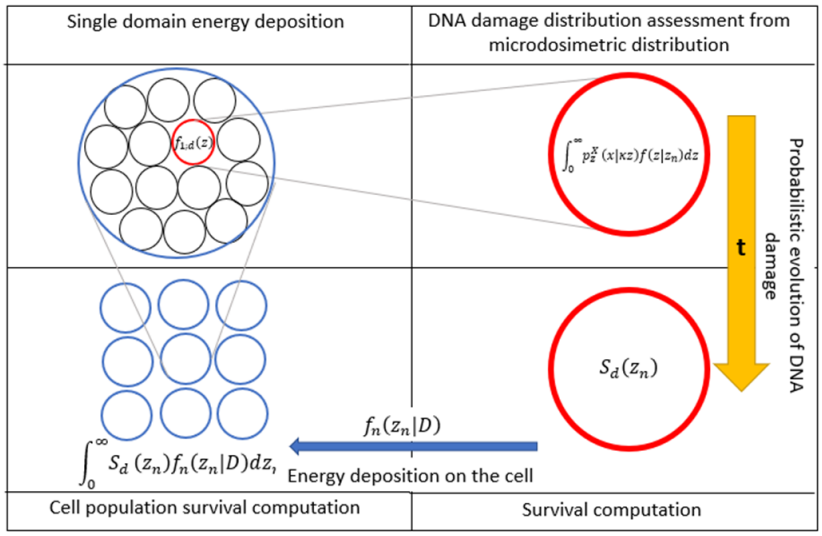

Figure 12.1: Schematic representation of the procedure for cell survival fraction assessment implemented by means of GSM^2

From an heuristic point of view, since the number of sub–lethal lesion can only decrease, the points $\{(y, 0) : y \in \mathbb{N}_0\}$ are absorbing states. Furthermore, the system reaches an absorbing state in finite time with probability 1, most likely converging towards a limiting stationary distribution. In particular, for any initial condition $\mathbf{x}$, it holds that

$$\mathbb{P}\left(\lim_{t\to\infty} X^{\mathbf{x}}(t) = 0\right) = 1 \,,$$

and we can infer that

$$p_\infty(y, x) = \delta(x) * p_\infty(y) =: \bar{p}_\infty(y) \,. \tag{12.5}$$

By *absorbing state*, we mean that once the system reaches the point $(y, 0)$ it cannot leave anymore that state and future evolutions are of no longer of interest. A rigorous proof of this result can be found in [Gardiner et al., 1985, Van Kampen, 1992].

It is worth highlighting that, in explicitly deriving above limiting distribution, the quadratic term in the MME (11.3) implies that the considered process does not fall into the broad category of *birth-death process*, and thus we could not take advantage of the extensive literature available on such type of processes. In this direction, usually approximation of the general model enables to obtain explicit form for the stationary distribution, [Gardiner et al., 1985].

Despite these issues, in the present work we will illustrate a derivation of an explicit form for the survival curve.

For any x, x_0 and $t_0 < t$, we denote

$$\begin{aligned}&\mathbb{P}_{t_0,x_0}\left(Y(t) = y\right) = \\ &= \mathbb{P}\left(Y(t) = y | \, Y(t_0) = y \, , \, X(t_0) = x_0\right) \, ,\end{aligned}$$

as the probability to have exactly x lesion of type II and y of type I at the time t.

Knowing that at time $t_0 < t$ the system was in the state $(0, x_0)$, we define $S(t|x_0, t_0)$ as the survival probability at time t. Thus,

$$\begin{aligned}&S(t|x_0, t_0) := \\ &:= \mathbb{P}\left(Y(t) = 0 | \, Y(t_0) = 0 \, , \, X(t_0) = x_0\right) = \\ &= \sum_{x=0}^{x_0} \mathbb{P}_{t_0,x_0}\left((Y(t), X(t)) = (0, x)\right) \, .\end{aligned} \tag{12.6}$$

Differentiating with respect to time both side of Equation (12.6), we obtain

$$\begin{aligned}&\partial_t S(t|x_0, t_0) = \\ &= \sum_{x=0}^{x_0} \partial_t \mathbb{P}_{t_0,x_0}\left((Y(t), X(t)) = (0, x)\right) = \\ &= \sum_{x=0}^{x_0} \partial_t p_{t_0,x_0}(t, 0, x) \, .\end{aligned} \tag{12.7}$$

Note that $p_{t_0,x_0}(t, 0, x)$ is the solution to Equation (11.3), considering the initial value

$$p_{t_0,x_0}(t_0, x, y) = \delta(y)\delta(x - x_0) \, .$$

As discussed in Section 11.2, the solution for a general probabilistic initial datum $p(t_0, x, y) = p_0^X(x)p_0^Y(y)$ can be obtained using the total law of probability

$$p(t, y, x) = \sum_{x_0,y_0} p_{t_0,y_0,x_0}(t, y, x)p(t_0, y_0, x_0) \, . \tag{12.8}$$

Therefore, merging Equations (12.7) and (12.8), we obtain that the survival probability must satisfy the following differential Equation

$$\partial_t S(t) = \sum_{x=0}^{x_0} \sum_{x_0 \geq 0} \partial_t p_{t_0,x_0}(t, 0, x)p_0^X(x_0)p_0^Y(0) \, . \tag{12.9}$$

To simplify the notation, we will drop the subscript (t_0, x_0), and instead use $p(t, x) = p_{t_0,x_0}(t, 0, x)$. Furthermore, we will describe the calculations for the specific case $t_0 = 0$, being the general case completely analogous.

The solution of Equation (12.7) can be derived from Equation (11.3), and then calculated iteratively. Using Equation (11.3) we have that

$$\begin{aligned} \partial_t p(t,x) = (x+1)rp(t,x+1) + \\ - \left((a+r)x + bx(x-1)\right) p(t,x)\,. \end{aligned} \tag{12.10}$$

As the number of sub–lethal lesion (i.e. the $x-$variable) can only decrease, we have that $p(t, x_0 + 1) = 0$ for $x = x_0$. Therefore, $p(t, x_0)$ satisfies

$$\partial_t p(t,x_0) = -\left((a+r)x_0 + bx_0(x_0-1)\right) p(t,x_0)\,,$$

whose explicit solution is given by

$$\begin{aligned} p(t,x_0) &= e^{-\gamma(x_0)t}\,, \\ \gamma(x_0) &:= \left((a+r)x_0 + bx_0(x_0-1)\right)\,. \end{aligned} \tag{12.11}$$

For $x = x_0 - 1$, we obtain

$$\begin{aligned} &\partial_t p(t, x_0 - 1) = \\ &= -\gamma(x_0-1)p(t,x_0-1) + \rho(x_0)p(t,x_0)\,, \end{aligned}$$

with $\rho(x_0 + 1) := (x_0 + 1)r$. Using $p(t, x_0)$ from Equation (12.11), it follows

$$\begin{aligned} &p(t, x_0 - 1) = \\ &= \int_0^t \rho(x_0) e^{-\gamma(x_0-1)(t-s)} p(s,x_0)ds = \\ &= \frac{\rho(x_0)}{\gamma(x_0) - \gamma(x_0-1)} \left(e^{-\gamma(x_0-1)t} - e^{-\gamma(x_0)t}\right)\,. \end{aligned} \tag{12.12}$$

Iterating above reasoning for any x_0 we infer that the general solution $p(t, x)$ must be a sum of exponential functions

$$p(t,x) = \sum_{k=x}^{x_0} C(k,x,x_0) e^{-\gamma(k)t}\,, \tag{12.13}$$

for a given function $C(k, x, x_0)$.

Inserting Equation (12.13) into Equation (12.12), and comparing exponential functions of the same order, we conclude that the general solution is given by

$$p(t,x) = \sum_{k=x+1}^{x_0} C(k,x,x_0) \left(e^{-\gamma(k)t} - e^{-\gamma(x)t}\right)\,, \tag{12.14}$$

with

$$C(k,x,x_0) := \frac{\rho(x+1)\rho(x+2)\rho(x+3)\ldots\rho(x_0)}{C_1(k,x)C_2(k,x_0)}\,,$$

being

$$\begin{aligned} C_1(k,x) &:= (\gamma(x) - \gamma(k))(\gamma(x+1) - \gamma(k)) \cdots \times \\ &\quad \times \ldots (\gamma(k-1) - \gamma(k))\,, \\ C_2(k,x_0) &:= (\gamma(k+1) - \gamma(k)) \ldots (\gamma(x_0) - \gamma(k))\,. \end{aligned}$$

The calculation described above demonstrates that the repair probability follows a multi-exponential form, and thus deviates from the predictions of the existing theory, which assumes an exponential repair kinetics [Fowler, 1999].

For $x = 0$, we obtain

$$\begin{aligned} p(t,0) &= \int_0^t rp(s,1)ds = \\ & r\sum_{k=2}^{x_0} C(k,1,x_0) \int_0^t \left(e^{-\gamma(k)s} - e^{-(a+r)s} \right) ds = \\ &= \sum_{k=2}^{x_0} C(k,1,x_0) \left(\frac{1-e^{-\gamma(k)t}}{\gamma(k)} - \frac{1-e^{-(a+r)t}}{a+r} \right) = \\ &= \sum_{k=0}^{x_0} C(k,0,x_0) e^{-\gamma(k)t} , \end{aligned} \tag{12.15}$$

which agrees with Equation (12.13).

Integrating Equation (12.7) with respect to time, we obtain from Equation (12.14)–(12.15),

$$\begin{aligned} & S(t|x_0) - 1 = \\ & = \sum_{x=0}^{x_0} (p(t,x) - p(0,x)) = \\ & \sum_{x=0}^{x_0} p(t,x) - 1 = \\ & = \sum_{x=1}^{x_0-1} \sum_{k=x+1}^{x_0} C(k,x,x_0) \left(e^{-\gamma(k)t} - e^{-\gamma(x)t} \right) + \\ & + e^{-\gamma(x_0)t} + \sum_{k=0}^{x_0} C(k,0,x_0) e^{-\gamma(k)t} - 1 . \end{aligned}$$

For $t \to \infty$, we obtain

$$e^{-\gamma(k)t} \to 0 ,$$

so that only the term with $k = 0$ does not converge to 0 and yields

$$S_{\infty;x_0} := \lim_{t\to\infty} S(t|x_0) = C(0,0,x_0) . \tag{12.16}$$

Equation (12.16) has a natural and intuitive meaning. Since the number of sub–lethal lesions can only decrease, it can reach 0 in a finite time with probability 1. Therefore, as $t \to \infty$, the only term that remains in Equation (12.16) is the one coming from $p(t,0)$, that does not converge to 0.

Using the rule of total probability, we can obtain the solution for a probabilistic initial data from Equation (12.8) as

$$\begin{aligned} p(t,0,x) &= \sum_{x_0=x}^{\infty} p_{x_0}(t,0,x) p(0,0,x_0) = \\ &= \sum_{x_0=x}^{\infty} p_{x_0}(t,0,x) p_0^X(x_0) p_0^Y(0) , \end{aligned}$$

where $p_{x_0}(t,0,x)$ can be computed as in Equation (12.14). In this case, the survival probability becomes

$$\begin{aligned} S(t) &= \mathbb{P}\left(Y(t)=0\right) = \\ &= \sum_{x=0}^{\infty} \mathbb{P}\left(Y(t)=0, X(t)=x\right) = \\ &= \sum_{x=0}^{\infty} p(t,0,x) = \\ &= \sum_{x=0}^{\infty} \sum_{x_0=x}^{\infty} p_{x_0}(t,0,x) p_0^X(x_0) p_0^Y(0)\,. \end{aligned} \tag{12.17}$$

The probability $p_{x_0}(t,0,x)$ can be calculated according to Equation (12.14)) only for $x_0 \geq 2$. We can also compute the values for $x_0 \geq 1$, considering that in such scenario no double interactions can occur. Thus, we obtain

$$\begin{cases} p_1(t,0,1) = e^{-(a+r)t}\,, \\ p_1(t,0,0) = r\frac{1-e^{-(a+r)t}}{a+r}\,, \\ p_1(t,0,x) = 0,\ x \geq 2\,, \\ p_0(t,0,0) = 1\,, \\ p_0(t,0,x) = 0\,, x \geq 1\,. \end{cases} \tag{12.18}$$

Combining a similar methodology as the one described above with Equations (12.14)-(12.15)-(12.17)–(12.18), yields the following

$$\begin{aligned} S(t) &= \sum_{x=0}^{\infty} \sum_{x_0=x}^{\infty} p_{x_0}(t,0,x) p_0^X(x_0) p_0^Y(0) = \\ &= \sum_{x_0=0}^{\infty} p_{x_0}(t,0,0) p_0^X(x_0) p_0^Y(0) + \\ &+ \sum_{x=1}^{\infty} \sum_{x_0=x}^{\infty} p_{x_0}(t,0,x) p_0^X(x_0) p_0^Y(0) = \\ &= p_0^X(0) p_0^Y(0) + \\ &+ \frac{r}{a+r} p_0^X(1) p_0^Y(0) \left(1 - e^{-(a+r)t}\right) + \\ &+ \sum_{x_0=2}^{\infty} \sum_{k=0}^{x_0} p_0^X(x_0) p_0^Y(0) C(k,0,x_0) e^{-\gamma(k)t} + \\ &+ \sum_{x=1}^{\infty} e^{-\gamma(x)t} p_0^X(x) p_0^Y(0) \\ &+ \sum_{x=1}^{\infty} \sum_{x_0=x+1}^{\infty} \sum_{k=x+1}^{x_0} p_0^X(x_0) p_0^Y(0) C(k,x,x_0) \times \\ &\times \left(e^{-\gamma(k)t} - e^{-\gamma(x)t}\right)\,. \end{aligned} \tag{12.19}$$

Considering Equation (12.19) in the limit $t \to \infty$, and using that $e^{-\gamma(x)t} \to 0$ as $t \to \infty$, we obtain that only the terms with $k = 0$ do not converge to 0. Thus

$$
\begin{aligned}
&S_d(z_n) := \lim_{t\to\infty} S(t) = \\
&= p_0^X(0|z_n)p_0^Y(0|z_n) + \\
&+ \frac{r}{a+r} p_0^X(1|z_n)p_0^Y(0|z_n) + \\
&+ \sum_{x_0=2}^{\infty} p_0^X(x_0|z_n)p_0^Y(0|z_n)C(x_0)\,,
\end{aligned}
\tag{12.20}
$$

with $C(x_0) := C(0,0,x_0)$. Here we have employed the notation $p_0^X(x_0|z_n)$, $p_0^Y(0|z_n)$ and $S_\infty(z_n)$ to emphasize the dependence of the initial distributions on the average deposited energy on the cell nucleus z_n.

We can give a probabilistic interpretation of the terms appearing in Equation (12.20): i) $p_0^X(x_0|z_n)$ and $p_0^Y(0|z_n)$ represent the probability that the domain suffers x_0 sub–lesions and 0 lethal lesion, respectively; (ii) $C(x_0)$ are weighting terms that represent the probability that x_0 sub–lethal lesions are repaired so that the domain survives.

The *survival probability* of the whole cell, when receiving an average energy deposition of z_n, is therefore evaluated as

$$
\begin{aligned}
&S(z_n) = \\
&= \mathbb{P}\left(\lim_{t\to\infty}\left(\left\{Y^1(t)=0\right\}\cap\cdots\cap\left\{Y^{N_d}(t)=0\right\}\right)\right) = \\
&= \prod_{i=1}^{N_d}\mathbb{P}\left(\lim_{t\to\infty} Y^i(t)=0\right) = \prod_{i=1}^{N_d} S_\infty(z_n) = \\
&= \prod_{i=1}^{N_d} p_0^X(0|z_n)p_0^Y(0|z_n) + \\
&+ \prod_{i=1}^{N_d}\sum_{x_0=1}^{\infty} C(x_0)p_0^X(x_0|z_n)p_0^Y(0|z_n)\,.
\end{aligned}
\tag{12.21}
$$

If we assume that all domains have the same probability distribution, we obtain

$$
\begin{aligned}
S(z_n) = \Big(&p_0^X(0|z_n)p_0^Y(0|z_n) + \\
&+ \sum_{x_0=1}^{\infty} C(x_0)p_0^X(x_0|z_n)p_0^Y(0|z_n)\Big)^{N_d}\,.
\end{aligned}
\tag{12.22}
$$

The survival Equation (12.22) calculated with GSM^2 has a substantial difference from the standard Poisson–based models. As shown in Section 2.2, the survival probability for a cell is generally computed by averaging the survival on a single domain, after it received an energy deposition z. On the contrary, we account for stochasticity deriving from energy deposition at the very beginning. In fact, the single domain survival probability is calculated taking into account all possible stochastic energy depositions z. Therefore, the cell survival probability is not the average over all possible energy depositions, but it is estimated as the probability that none of the domains suffers a lethal lesion.

Explicit formulation of the initial conditions for the survival fraction computation

The survival probability of Equation (12.22) is expressed in terms of the initial damage distribution $p_0^X(x_0|z_n)$. We will derive a more explicit form for the survival Equation by using the microdosimetric distribution $f_{1;d}$, so that the relation of the survival curve on the variable z_n appears explicitly.

For $x > 0$, using Equation (11.9), the following holds

$$\begin{aligned} & p_0^X(x) = \\ & = \int_0^\infty e^{-\kappa z} \frac{(\kappa z)^x}{x!} f(z|z_n) dz = \\ & = \sum_{\nu \geq 0} \int_0^\infty e^{-\kappa z} \frac{(\kappa z)^x}{x!} \frac{z_n^\nu}{z_F^\nu \nu!} e^{-\frac{z_n}{z_F}} f_{\nu;d}(z) dz = \\ & = \sum_{\nu \geq 0} h(z_n, \nu, \kappa, x) \int_0^\infty e^{-\kappa z} z^x f_{\nu;d}(z) dz \,. \end{aligned} \tag{12.23}$$

with

$$h(z_n, \nu, \kappa, x) := \frac{\kappa^x}{x!} \frac{z_n^\nu}{z_F^\nu \nu!} e^{-\frac{z_n}{z_F}} \,.$$

We will denote by $\mathcal{L}[g(z)]$ the *Laplace transform* of the function g, defined as

$$\mathcal{L}[g(z)](\kappa) := \int_0^\infty e^{-\kappa z} g(z) dz \,.$$

Taking advantage of the *Laplace transform* property, we have

$$\begin{aligned} & \mathcal{L}[z^x g(z)](\kappa) = (-1)^x \frac{d^x}{d\kappa^x} \mathcal{L}[g(z)](\kappa) \,, \quad x > 0 \,, \\ & \mathcal{L}[g(z) * g(z)](\kappa) = (\mathcal{L}[g(z)](\kappa))^2 \,. \end{aligned} \tag{12.24}$$

Substituting Equation (12.24) into Equation (12.23), we find

$$\begin{aligned} & p_0^X(x) = \\ & = \sum_{\nu \geq 0} h(z_n, \nu, \kappa, x) \int_0^\infty e^{-\kappa z} z^x f_{\nu;d}(z) dz = \\ & = \sum_{\nu \geq 0} h(z_n, \nu, \kappa, x) (-1)^x \frac{d^x}{d\kappa^x} \left(\mathcal{L}[f_{1;d}(z)]^\nu (\kappa) \right) = \\ & = (-1)^x \frac{\kappa^x}{x!} e^{-\frac{z_n}{z_F}} \times \\ & \quad \times \frac{d^x}{d\kappa^x} \sum_{\nu \geq 0} \frac{z_n^\nu}{z_F^\nu \nu!} \left(\int_0^\infty e^{-\kappa z} f_{1;d}(z) dz \right)^\nu = \\ & = (-1)^x \frac{\kappa^x}{x!} e^{-\frac{z_n}{z_F}} \times \\ & \quad \times \frac{d^x}{d\kappa^x} \exp\left[\frac{z_n}{z_F} \int_0^\infty e^{-\kappa z} f_{1;d}(z) dz \right] \,. \end{aligned} \tag{12.25}$$

$M(\kappa)$ and $M_x(\kappa)$ represent the *moment generating function* and the $x-$translated *moment generating function*, respectively, of the single–event distribution f_1 defined as

$$M(\kappa) := \int_0^\infty e^{-\kappa z} f_1(z) dz = \sum_{n\geq 0} \frac{(-1)^n}{n!} \kappa^n m_n \,,$$
$$M_x(\kappa) := \int_0^\infty z^x e^{-\kappa z} f_1(z) dz = \sum_{n\geq 0} \frac{(-1)^n}{n!} \kappa^n m_{n+x} \,,$$

where m_n is the $n-$th moment of single–event distribution $f_{1;d}(z)$.

Defining for short

$$g(\kappa) = \frac{z_n}{z_F} \int_0^\infty e^{-\kappa z} f_{1;d}(z) dz = \frac{z_n}{z_F} M(\kappa) \,,$$

the following equalities hold true, [Craik, 2005],

$$\begin{aligned} &\frac{d^x}{d\kappa^x} e^{g(\kappa)} = \\ &= e^{g(\kappa)} \sum_{i=1}^{x} \frac{(-1)^i}{i!} \sum_{j=1}^{i} (-1)^j \binom{i}{j} \frac{d^x}{d\kappa^x} g^j(\kappa) g(\kappa)^{i-j} \,, \\ &\frac{d^x}{d\kappa^x} g^j(\kappa) = \\ &= \sum_{l=1}^{j} \left(\frac{z_n}{z_F}\right)^l \frac{j!}{(j-l)!} g^{j-l}(\kappa) B_{x,l}\left(M_1, \ldots, M_{x-l+1}\right) \,, \end{aligned} \tag{12.26}$$

where $B_{x,l}\left(M_1, \ldots, M_{x-l+1}\right)$ is the *Bell's polynomial* defined as

$$\begin{aligned} &B_{x,l}\left(M_1, \ldots, M_{x-l+1}\right) := \\ &= \sum \frac{x!}{j_1! \ldots j_{x-l+1}!} (M_1)^{j_1} \ldots (M_{x-l+1})^{j_{x-l+1}} \,, \end{aligned}$$

and the summation ranges over multi-indexes such that

$$\begin{aligned} &j_1 + j_2 + \cdots + j_{x-l+1} = l \,, \\ &j_1 + 2j_2 + \cdots + (x-l+1) j_{x-l+1} = x \,. \end{aligned}$$

Using Equation (12.26), we thus get

$$\begin{aligned} &\frac{d^x}{d\kappa^x} e^{g(\kappa)} = \\ &= e^{g(\kappa)} \sum_{i=1}^{x} \sum_{j=1}^{i} \sum_{l=1}^{j} \frac{(-1)^i (-1)^j}{(i-j)!(j-l)!} \left(\frac{z_n}{z_F}\right)^j \times \\ &\quad \times M^{i-l} B_{x,l}\left(M_1, \ldots, M_{x-l+1}\right) \,. \end{aligned} \tag{12.27}$$

Inserting Equation (12.27) into Equation (12.25) we finally obtain

$$\begin{aligned} & p_0^X(x) = \\ & = \frac{(-\kappa)^x}{x!} \exp\left[-\frac{z_n}{z_F}\int_0^\infty \left(1-e^{-\kappa z}\right) f_{1;d}(z)dz\right] \times \\ & \times \sum_{i=1}^{x}\sum_{j=1}^{i}\sum_{l=1}^{j} \frac{(-1)^i(-1)^j}{(i-j)!(l-2)!)} \left(\frac{z_n}{z_F}\right)^j \times \\ & \times M^{i-l} B_{x,l}\left(M_1,\ldots,M_{x-l+1}\right) = \\ & = \exp\left[-\frac{z_n}{z_F}\int_0^\infty \left(1-e^{-\kappa z}\right) f_{1;d}(z)dz\right] \times \\ & \times \frac{(-\kappa)^x}{x!} H(z_n,x,M)\,, \end{aligned} \tag{12.28}$$

with

$$\begin{aligned} & H(z_n,x,M) := \\ & = \sum_{i=1}^{x}\sum_{j=1}^{i}\sum_{l=1}^{j} \frac{(-1)^i(-1)^j}{(i-j)!(j-l)!} \left(\frac{z_n}{z_F}\right)^j \times \\ & \times M^{i-l} B_{x,l}\left(M_1,\ldots,M_{x-l+1}\right)\,. \end{aligned} \tag{12.29}$$

For $x = 0$, we find the particular case

$$p_0^X(0) = \exp\left[-\frac{z_n}{z_F}\int_0^\infty \left(1-e^{-\kappa z}\right) f_{1;d}(z)dz\right]\,, \tag{12.30}$$

and analogously for $p_0^Y(0)$

$$p_0^Y(0) = \exp\left[-\frac{z_n}{z_F}\int_0^\infty \left(1-e^{-\lambda z}\right) f_{1;d}(z)dz\right]\,. \tag{12.31}$$

From Equations (12.28)–(12.30)–(12.31), and exploiting the fact that

$$\begin{aligned} & p_0^X(x_0|z_n)p_0^Y(0|z_n) = \\ & = \exp\left[-\frac{z_n}{z_F}\int_0^\infty \left(2-e^{-\kappa z}-e^{-\lambda z}\right) f_{1;d}(z)dz\right] \times \\ & \times \frac{(-\kappa)^x}{x!} H(z_n,x,M)\,, \end{aligned}$$

the survival Equation (12.20) becomes

$$\begin{aligned} & S_d(z_n) = \\ & = \exp\left[-\frac{z_n}{z_F}\int_0^\infty \left(2-e^{-\kappa z}-e^{-\lambda z}\right) f_{1;d}(z)dz\right] \times \\ & \times \left[1+\sum_{x_0=1}^{\infty} \frac{(-\kappa)^{x_0}}{x_0!} H(z_n,x_0,M)C(x_0)\right]\,. \end{aligned} \tag{12.32}$$

Thus, the survival for the whole cell nucleus is calculated as

$$\begin{aligned} &S(z_n) = \\ &= \exp\left[-N_d \frac{z_n}{z_F} \int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz\right] \times \\ &\times \left[1 + \sum_{x_0=1}^{\infty} \frac{(-\kappa)^{x_0}}{x_0!} H(z_n, x_0, M) C(x_0)\right]^{N_d} . \end{aligned} \tag{12.33}$$

To emphasize the dependence of Equation (12.33) on z_n, the terms of $H(z_n, x, M)$ contained in Equation (12.33) can be rearranged to obtain (12.33)

$$\begin{aligned} &S(z_n) = \\ &= \exp\left[-N_d \frac{z_n}{z_F} \int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz\right] \times \\ &\times \left(1 + \sum_{k=1}^{\infty} \left(\frac{z_n}{z_F}\right)^k G_k(M)\right)^{N_d} , \end{aligned} \tag{12.34}$$

with

$$\begin{aligned} &G_k(M) = \\ &:= \sum_{x_0=k}^{\infty} \sum_{i=1}^{x_0} \sum_{j=1}^{i} \frac{(-1)^i (-1)^j}{(i-j)!(j-k)!} \frac{(-\kappa)^{x_0}}{x_0!} C(x_0) \times \\ &\quad \times M^{i-k} B_{x_0,k}\left(M_1, \ldots, M_{x_0-k+1}\right) . \end{aligned} \tag{12.35}$$

Remark 12.1.1. In the very particular case of low–LET radiation, as showed in Section 11.2, since $\kappa^2 z_n z_D << 1$, we infer that p_0^X follows a Poisson distribution.

Therefore, using the fact that the initial damage distribution is Poisson distributed, Equation (12.34) simplifies to

$$\begin{aligned} &S(z_n) = \\ &= e^{-N_d z_n (\kappa + \lambda)} \left(1 + \sum_{x_0=1}^{\infty} z_n^{x_0} \frac{\kappa^{x_0}}{x_0!} C(x_0)\right)^{N_d} . \end{aligned} \tag{12.36}$$

Comparison with existing radiobiological models for predicting survival fraction

Equation (12.34) depends on both the biological parameters a, b, r, κ and λ, as well as physical parameters related to energy deposition, mainly described by the single-event specific energy spectra $f_{1;d}(z)$. In particular, the survival Equation (12.34) is the product of a linear exponential function

$$\exp\left[-N_d \frac{z_n}{z_F} \int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz\right] ,$$

and a polynomial function of general order on z_n

$$\left(1 + \sum_{k=1}^{\infty} \left(\frac{z_n}{z_F}\right)^k G_k(M)\right)^{N_d} .$$

The linear exponential function describes the initial damage formation inside the cell. This term depends on the biological parameters κ and λ, describing the number of initial lethal and sublethal lesions for a given energy event z, and on the radiation quality via the single–event spectra $f_{1;d}(z)$. It is worth stressing that, differently from the vast majority of existing models, the whole microdosimetric distribution is considered rather than solely mean values.

The polynomial function, instead, accounts for the time evolution of the DNA damages. Each term G depends on the full energy spectrum $f_{1;d}(z)$ through the moment generating function M; this is a relevant difference with the majority of existing models, which are mainly based on the first two moments of the microdosimetric distribution. The probabilistic description of the G terms is that the $i-$th term accounts for the damage caused by i events. Each term is weighted by the probability that such event occurs and that the induced DNA damage is repaired.

The survival curve in Equation (12.20), written in compact form as,

$$S_d(z_n) := \sum_{x_0 \geq 0} p_0^X(x_0|z_n) p_0^Y(0|z_n) C(x_0)\,, \tag{12.37}$$

can be seen as a generalization of a *multi–hit* (MH) model. If we further approximate the multi–event microdosimetric distribution as

$$f_{\nu;d}(z) = \delta(z - \nu z_F)\,, \tag{12.38}$$

with $\delta(z - \nu z_F)$ the Dirac delta centered in νz_F, we obtain from Equation (12.23),

$$\begin{aligned} & p_0^X(x) = \\ & = \int_0^\infty e^{-\kappa z} \frac{(\kappa z)^x}{x!} f(z|z_n) dz = \\ & = \sum_{\nu \geq 0} \int_0^\infty e^{-\kappa z} \frac{(\kappa z)^x}{x!} \frac{z_n^\nu}{z_F^\nu \nu!} e^{-\frac{z_n}{z_F}} \delta(z - \nu z_F) dz = \\ & = \sum_{\nu \geq 0} e^{-\kappa \nu z_F} \frac{(\kappa \nu z_F)^x}{x!} \frac{z_n^\nu}{z_F^\nu \nu!} e^{-\frac{z_n}{z_F}} = \\ & = e^{-\frac{z_n}{z_F}\left(1 - e^{-\kappa z_F}\right)} B_x\left(e^{-\frac{z_n}{z_F}} e^{-\kappa z_F}\right)\,, \end{aligned} \tag{12.39}$$

with B_x the Bell polynomial. Equation (12.39) is equal to [Vassiliev, 2012, Equation 5]. Therefore, inserting Equation (12.39) into Equation (12.22) provides a generalization of the multi–hit model proposed in [Vassiliev, 2012].

Although a distribution $f_{\nu;d}(z)$ of the type described by Equation (12.38) is reasonable, it is nonetheless an approximationm because it completely neglects the variance in the microdosimetric energy deposition event. To overcome this limitation, in this work we have exploited the full microdosimetric information coming from the entire spectrum.

Unlike the standard MH model, we do not assume *a priori* any a priori upper limit for the number of hits. In fact, GSM^2 suitably weighs the number of observed damages by the probability that a certain number of damages results from a specific microdosimetric and further consider the probability that damages are repaired.

Another difference between the HM and the proposed model, is that the number of lesions generated by an event is derived exploiting microdosimetry. In particular, the probability that a certain number of damages results from energy deposition is not assumed

to Poissonian but it is derived from a mechanistic dynamical Equation (11.3) allowing for repair and damage interaction. As shown in Chapter 11, the resulting probability can be Poissonian under specific irradiation conditions, but can also differ significantly from a Poisson distribution, e.g. in high dose regimes.

After taking the logarithm, Equation (12.34) becomes

$$\begin{aligned}&\log S(z_n) = \\ &= \left[-N_d \frac{z_n}{z_F} \int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz\right] + \\ &+ N_d \log\left(1 + \sum_{k=1}^{\infty} \left(\frac{z_n}{z_F}\right)^k G_k(M)\right) .\end{aligned} \tag{12.40}$$

For low doses, we have that $z_n << 1$, and expanding up to the second term in z_n in Equation (12.40), we obtain

$$\begin{aligned}&\log S(z_n) \sim_0 \\ &\sim_0 -\frac{z_n}{z_F} N_d \int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz + \\ &+ \frac{z_n}{z_F} N_d G_1(M) + \\ &- N_d \left(\frac{z_n}{z_F}\right)^2 \left(\frac{G_1(M)^2}{2} - G_2(M)\right) = \\ &= -\alpha_0(M) z_n - \beta_0(M) z_n^2 ,\end{aligned} \tag{12.41}$$

with

$$\begin{cases} \alpha_0(M) & = \frac{N_d}{z_F} \int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz + \\ & \quad - \frac{N_d}{z_F} G_1(M) , \\ \beta_0(M) & = \frac{N_d}{z_F^2} \left(\frac{G_1(M)^2}{2} - G_2(M)\right) . \end{cases} \tag{12.42}$$

Therefore, from the fact that

$$\frac{d}{dD} \log S(D) = \frac{1}{S(D)} \frac{d}{dD} S(D) ,$$

using asymptotic arguments it can be shown that

$$\begin{aligned}\frac{d}{dD} \log(S(D)\Big|_{D=0} &= -\frac{N_d}{z_F} \int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz + \\ &+ \frac{N_d}{z_F} \sum_{x_0 \geq 1} C(x_0) \int_0^\infty \frac{e^{-\kappa z}}{x_0!} (\kappa z)^{x_0} f_{1;d}(z) dz .\end{aligned} \tag{12.43}$$

In Equation (12.41), we are neglecting higher order polynomials in z_n coming from Equation (12.34). Additionally, the linear and quadratic terms depend on all the moments of the single–event microdosimetric distribution $f_{1;d}$, rather than just on the first two moments z_F and z_D as in all classical models. By dropping all moments greater than the second, GSM^2 provides the similar results as the existing models [Bellinzona et al., 2021].

For high doses, i.e. $z_n >> 1$, the linear term in Equation (12.40) dominates the logarithm, so that

$$\begin{aligned}
&\log S(z_n) \sim_\infty \\
&\sim_\infty -\frac{z_n}{z_F} N_d \left(\int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz \right) + \\
&+ \frac{z_n}{z_F} N_d \int_0^\infty e^{-\kappa z} \kappa z f_{1;d}(z) dz \,,
\end{aligned} \tag{12.44}$$

so that

$$\begin{aligned}
\frac{d}{dD} \log(S(D)\Big|_{D=\infty} &= -N_d \frac{1}{z_F} \int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz + \\
&+ N_d \frac{1}{2 z_F} \int_0^\infty e^{-\kappa z} \kappa z f_{1;d}(z) dz \,.
\end{aligned} \tag{12.45}$$

Above low and high dose approximation emphasize how the proposed model naturally incorporates the linear–quadratic–linear (LQL) behaviour of the survival curves experimentally observed.

Furthermore, GSM^2 satisfies the *Hugh–Kellerer theorem*, [Rossi and Zaider, 1988, Rossi and Zaider, 1991]. In fact, assuming that no bystandard effects can occur, cells that experience no event must survive. As shown above, the event frequency for a given absorbed dose D is D/z_F. As the events are statistically independents, the Poisson statistics implies that there is a natural lower bound to the survival, that is

$$S_d(D) \geq e^{-\frac{D}{z_F}} \,. \tag{12.46}$$

It can be seen that the standard linear–quadratic model violates the lower bound in the high dose region.

Assuming that each cell has received the same dose, i.e. $z_n = D$, Equation (12.44) shows that in the high dose regime, the derived survival follows a linear exponential function

$$\begin{aligned}
S_d(D) &= e^{-\frac{D}{z_F}\left(\int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz\right)} \geq \\
&\geq e^{-\frac{D}{z_F}} \,,
\end{aligned}$$

where the inequality follows from the fact that

$$\int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz \leq 1 \,.$$

The bound is clearly satisfied under the low dose regimes, proving that GSM^2 does not violate the *Hugh–Kellerer theorem*.

Non–Poissonian inter-cellular corrections

Given a dose D, the energy deposition distribution can be computed as described in Section 11.2 for $f(z|z_n)$. We therefore have that

$$f_n(z_n|D) := \sum_{\nu=0}^{\infty} p_n(\nu|D) f_{\nu;c}(z_n) \,, \tag{12.47}$$

where $f_{\nu;n}(z_n)$ is the distribution resulting from ν energy depositions in a single cell nucleus and $p_n(\nu|D)$ is the probability that a ν energy deposition occurs in a cell nucleus. The distribution $f_{\nu;n}(z_n)$ can be obtained convolving the single event distribution $f_{1;n}(z_n)$.

Therefore, the total *cell survival probability* can be obtained as described in [Rossi and Zaider, 1991, Ch. VI]

$$S(D) = \int_0^\infty S_d(z_n) f_n(z_n|D) dz_n \,. \tag{12.48}$$

Equation (12.48) contains all the over-killing corrections typically added to the survival curves to account for the stochastic nature of energy deposition.

Given the survival curve $S_n(z_n)$ explicitly computed in Equation (12.34), the survival curve $S(D)$ can efficiently calculated using standard numerical integration techniques. Nonetheless, using some physical considerations concerning the stochasticity of energy deposition some useful simplifications can be derived.

It is worth stressing that, as highlighted in Section 2.2, in the original MKM formulation [Hawkins, 1994], it has been assumed that no stochasticity in energy deposition among cell domains happens, so that $S(D)$ can be obtained via the approximation

$$f_n(z_n|D) \approx \delta(z_n - D) \,,$$

yielding using Equation (12.48) the survival curve

$$\begin{aligned} S(D) &= S_d(D) = \\ &= \exp\left[-N_d \frac{D}{z_F} \int_0^\infty \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;d}(z) dz\right] \times \\ &\quad \times \left(1 + \sum_{k=1}^\infty \left(\frac{D}{z_F}\right)^k G_k(M)\right)^{N_d} , \end{aligned}$$

with $G_k(M)$ as in Equation (12.35).

In the following, suitable approximation of the microdosimetric spectrum $f_n(z_n|D)$ will be introduced and considered under different LET and doses regimes. Thus, the resulting survival curve will be computed highlighting its main aspects.

The low dose regime

For low doses, and in particular in the case

$$\alpha_0(M) z_n + \beta_0(M) z_n^2 << 1 \,,$$

with α_0 and β_0 given as in Equation (12.42), Equation (12.53) yields

$$\begin{aligned} S(D) &\approx 1 + \\ &- \int_0^\infty \left(\alpha_0(M) z_n + \left(\beta_0(M) - \frac{\alpha_0^2(M)}{2}\right) z_n^2\right) f_n(z_n|D) dz_n = \\ &= 1 - \alpha_0(M) m_1^c - \left(\beta_0(M) - \frac{\alpha_0^2(M)}{2}\right) m_2^c \,, \end{aligned} \tag{12.49}$$

with m_1^c and m_2^c the first and second moment, respectively, of the multi–event distribution $f_n(z_n|D)$. Using the explicit forms for the moments m_1^c and m_2^c, [Rossi and Zaider, 1991,

Section II.2], we can conclude using Equation (12.49) that

$$\begin{aligned} S(D) \approx 1 - \left(\alpha_0(M) + \left(\beta_0(M) - \frac{\alpha_0^2(M)}{2}\right) z_D\right) D + \\ - \left(\beta_0(M) - \frac{\alpha_0^2(M)}{2}\right) D^2 \approx e^{-\alpha(M)D - \beta(M)D^2} , \end{aligned} \tag{12.50}$$

with

$$\begin{cases} \alpha(M) &= \alpha_0(M) + \left(\beta_0(M) - \frac{\alpha_0^2(M)}{2}\right) z_{D;n} , \\ \beta(M) &= \left(\beta_0(M) - \frac{\alpha_0^2(M)}{2}\right) , \end{cases} \tag{12.51}$$

where $z_{D;n}$ is the dose average of the specific energy in multi events. Equations (12.50)–(12.51) emphasize the connection of the current models with the classical MKM. In fact, as shown in 11, in the low dose regimes the GSM^2 predicts a Poissonian behaviour of lethal damages so that the standing assumption of the MKM is recovered.

The medium/low–dose and high–LET regime

For high–LET regimes and medium/low doses, more precisely in regimes for which $\frac{D}{z_F} << 1$, the probability of more than one event is negligible. In such a situation, either 0 or 1 event is registered, so that the multi–event energy distribution can be approximated as

$$\begin{aligned} f_n(z_n|D) := \sum_{\nu=0}^{\infty} \frac{e^{-\frac{D}{z_F}}}{\nu!} \left(\frac{D}{z_F}\right)^n f_{\nu;c}(z) \approx \\ \approx e^{-\frac{D}{z_F}} \left(\delta(z_n) + \frac{D}{z_F} f_{1;c}(z_n)\right) . \end{aligned} \tag{12.52}$$

Calculating Equation (12.48) with the approximated multi–event distribution (12.52), we obtain

$$\begin{aligned} S(D) = \int_0^{\infty} S_d(z_n) f_n(z_n|D) dz_n \approx \\ \approx e^{-\frac{D}{z_F}} + e^{-\frac{D}{z_F}} \frac{D}{z_F} \int_0^{\infty} S_d(z_n) f_{1;c}(z_n) dz_n . \end{aligned} \tag{12.53}$$

A further approximation can be included into Equation (12.53); in fact, Equation (12.41) demonstrated that the survival at low doses is linear–quadratic, and thus, combining Equation (12.41) with Equation (12.53) we obtain

$$\begin{aligned} S(D) \approx e^{-\frac{D}{z_F}} + \\ + e^{-\frac{D}{z_F}} \frac{D}{z_F} \int_0^{\infty} e^{-\alpha_0 z_n - \beta_0 z_n^2} f_{1;c}(z_n) dz_n , \end{aligned} \tag{12.54}$$

with α_0 and β_0 as in Equation (12.42).

From Equation (12.54), we find further

$$\begin{aligned} S(D) \approx \\ \approx \exp\left[-\frac{D}{z_F} \int_0^{\infty} \left(1 - e^{-\alpha_0 z_n - \beta_0 z_n^2}\right) f_{1;c}(z_n) dz_n\right] . \end{aligned} \tag{12.55}$$

It is worth stressing that Equation (12.55) highlights and insightful connection to the DNA-lesion theory of radiation action, [Rossi and Zaider, 1991]. In fact Equation (12.55) recover the main Equation of DNA-lesion theory of radiation action, [Rossi and Zaider, 1991, equation VI.80].

The high–dose regime

For a high–dose regime, more rigorously in regimes for which it holds that $\frac{D}{z_F} >> 1$, the multi–event distribution becomes Gaussian distributed, [Rossi and Zaider, 1988]. We therefore have that

$$\begin{aligned} f(z|z_n) &:= \sum_{\nu=0}^{\infty} \frac{e^{-\frac{z_n}{z_F}}}{\nu!} \left(\frac{z_n}{z_F}\right)^n f_{\nu;d}(z) \approx \\ &\approx \frac{1}{\sqrt{2\pi z_D D}} e^{-\frac{1}{2}\frac{(z_n-D)^2}{z_D D}} , \end{aligned} \tag{12.56}$$

where z_D is the dose average of the specific energy in single events.

Assessing Equation (12.48) with the approximated multi–event distribution (12.52), we obtain using the purely linear behaviour for the survival curve $S_d(z_n)$ showed in Equation (12.44),

$$\begin{aligned} S(D) &= \int_0^{\infty} S_d(z_n) f_n(z_n|D) dz_n \approx \\ &\approx \frac{1}{\sqrt{2\pi z_D D}} \int_0^{\infty} e^{-\alpha_{HL} z_n} e^{-\frac{1}{2}\frac{(z_n-D)^2}{z_D D}} dz_n , \end{aligned} \tag{12.57}$$

where z_D is the dose average of the specific energy in single events on the cell-nucleuous and

$$\alpha_{HL} := N_d \left(\int_0^{\infty} \left(2 - e^{-\kappa z} - e^{-\lambda z}\right) f_{1;c}(z) dz \right) .$$

Solving Equation (12.57), we find

$$S(D) = e^{-\alpha_{HL}\left(1-\frac{\alpha_{HL} z_D}{2}\right)D} . \tag{12.58}$$

As mentioned above, Equation (12.58) accounts for overkilling effects. In fact, a cell irradiated with a higher LET will have a lower probability of surviving, resulting in a greater value for α_{HL}. According to Equation (12.58), the corresponding survival curve will be corrected and shifted upward by a greater term $\frac{\alpha_{HL}^2 z_D}{2}$, yielding in fact the typical effect of overkilling corrections.

12.2 Numerical results

All numerical Monte Carlo simulations for producing input microdosimetric spectra have been carried out with the microdosimetric extension of TOPAS toolkit [Perl et al., 2012]. In all calculations, we used the geometry of spherical TEPC with the active area equivalent to a 2 μm diameter sphere of tissue, as in experiments of Part III and as explained in II. In Table 12.1, we list all beam types selected for the simulations, including the initial kinetic energy, the residual range in water after traversing the microdosimeter, the y_F and y_D microdosimetric values measured by the detector. The $yd(y)$ spectra obtained from the TEPC for all beams are plotted in Figure 12.2,

Table 12.1: Particle type, incident kinetic energy and residual energy of all particle types used for the calculations. The, y_F and y_D values estimated from the TEPC are also reported.

Particle type	**Proton**			**Helium**			**Carbon**			**Oxygen**		
Residual range [mm]	2	5	10	2	5	10	2	5	10	2	5	10
Incident kinetic energy [MeV/u]	18.6	25.8	80	18.6	25.8	80	34	40.5	149	39.8	55.5	177
y_F **[keV/μm]**	3.67	2.10	0.92	14.32	8.27	2.79	76.48	54.06	14.73	119.39	68.18	22.69
y_D **[keV/μm]**	4.92	4.61	2.15	18.80	13.57	7.65	84.70	62.43	18.43	130.40	81.67	26.27

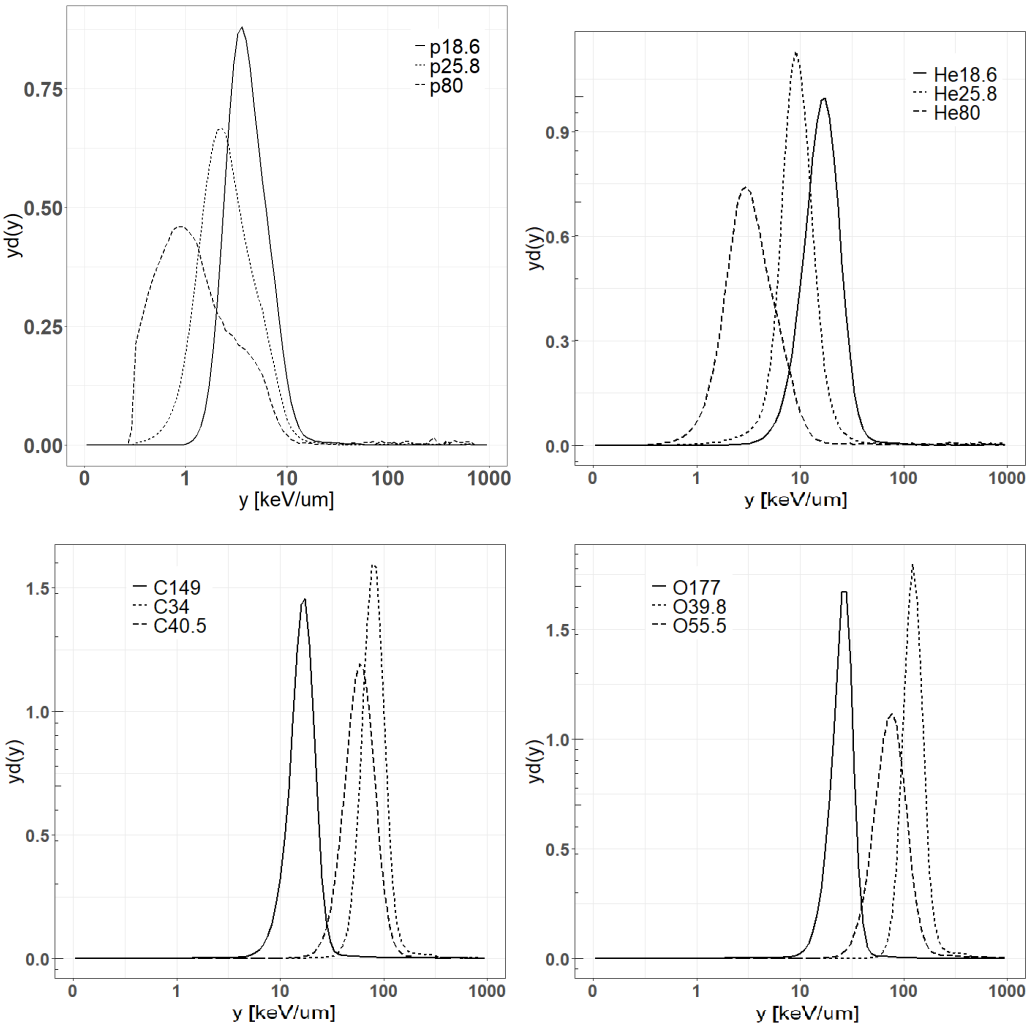

Figure 12.2: yd(y) microdosimetric spectra for the simulated radiation fields in the TECP. Labels in the legends refer to particle type and initial kinetic energy of the particle.

On the initial DNA damage distribution: GSM2 predicted distribution vs classical Poissonian distribution

For all radiation types listed in Table 12.1, we calculated the initial damage distributions using Equation and compared it to the classical Poissonian approach. Here we only show the distributions for sublethal damages X, because the probability for a direct damage Y is typically low, leading to a distribution closer to a Poissonian one. Current analysis further focuses on the combination of high dose and LET regimes to a possibly non Poissonian DNA damage distribution.

First, we considered protons at different residual ranges (i.e. different LET) and at different doses. The results are shown in Figure 12.3.

Figure 12.3 proves that for protons at low LET and low dose, the Poissonian assumption is valid. When the LET or the dose increase, the deviation from the classical assumption grows also, reaching the maximum in panel (d). We also evaluated the initial damage distribution induced by helium, carbon and oxygen ions, to investigate higher LET. Results for helium and oxygen are shown in Figure 12.5. The distribution obtained with carbon

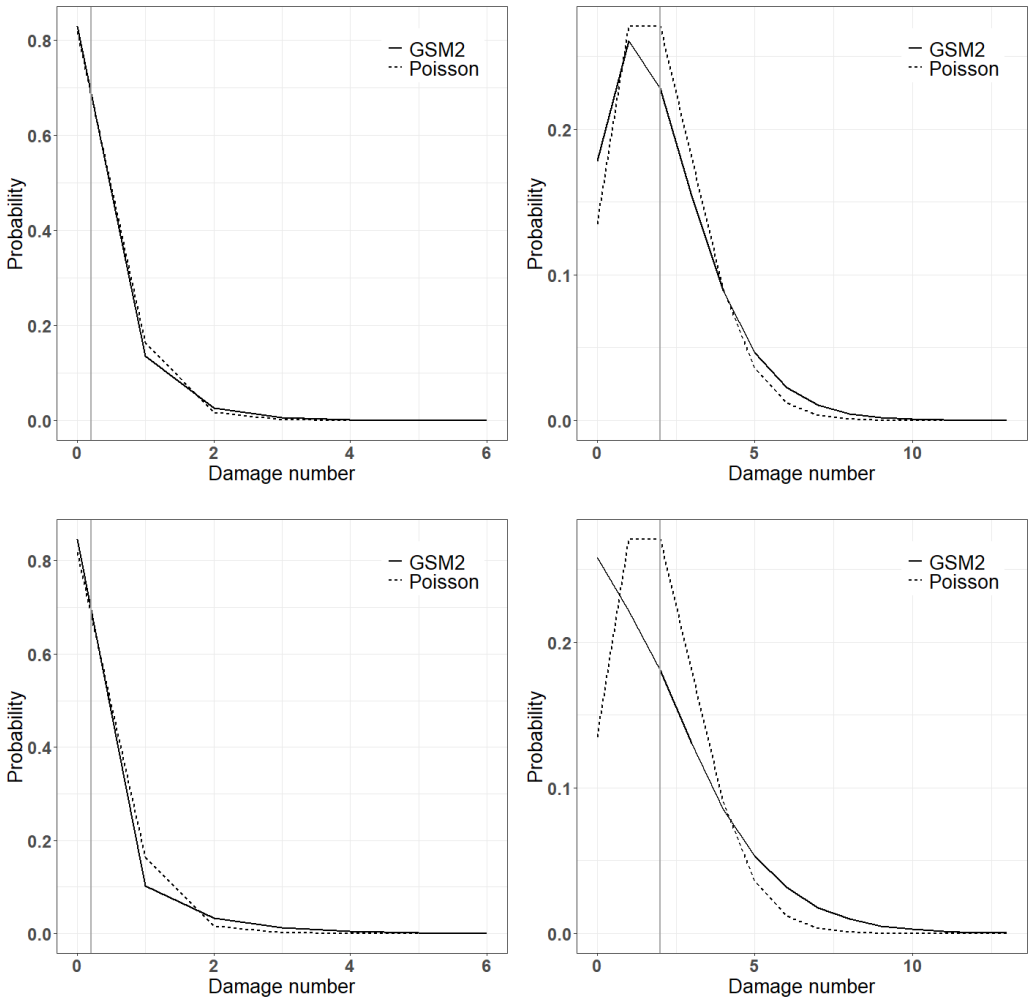

Figure 12.3: Initial sublethal DNA damage distribution for protons at (a) 10 mm residual range, z_n=0.5 Gy, (b) 10 mm residual range, z_n=5 Gy, (c) 2 mm residual range, z_n=0.5 Gy, (d) 2 mm residual range, z_n=5 Gy. In all calculations, we used r_d=0.5 μm and k=0.4.

ions are not plotted, since they are very similar to the oxygen one.

All heavy ions show a pattern similar to protons. At low doses and 10 mm residual range, the distributions are close to a Poisson one, even though in the case of oxygen the difference from the Poisson distribution is still evident also at low doses, indicating a failure of Poisson approximation at all dose ranges for high LET. At high doses, both ions show a bimodal distribution, indicating a significant difference from a Poisson. Also in this case, the damage distribution for oxygen shows a more pronounced bimodal behavior, typical of high LET ions. A summary of the most relevant parameters calculated for the DNA damage distributions of Figures 12.2 and 12.3 are reported in Table 12.2. Together with the physical properties of the considered radiation fields, such as kinetic energy, z_D and z_n, the mean values of the sublethal damage distributions and the non-Poissonian correction $\vartheta(\kappa, z_n, z_D)$ as defined in Equation (11.10) are also reported. We recall that $\vartheta(\kappa, z_n, z_D)$ quantifies the difference between the first and second moment of the damage distribution. Therefore, $\vartheta(\kappa, z_n, z_D) << 1$ implies that the damage distribution is almost Poissonian whereas $\vartheta(\kappa, z_n, z_D) >> 1$ implies significant deviation from the initial distribution to a

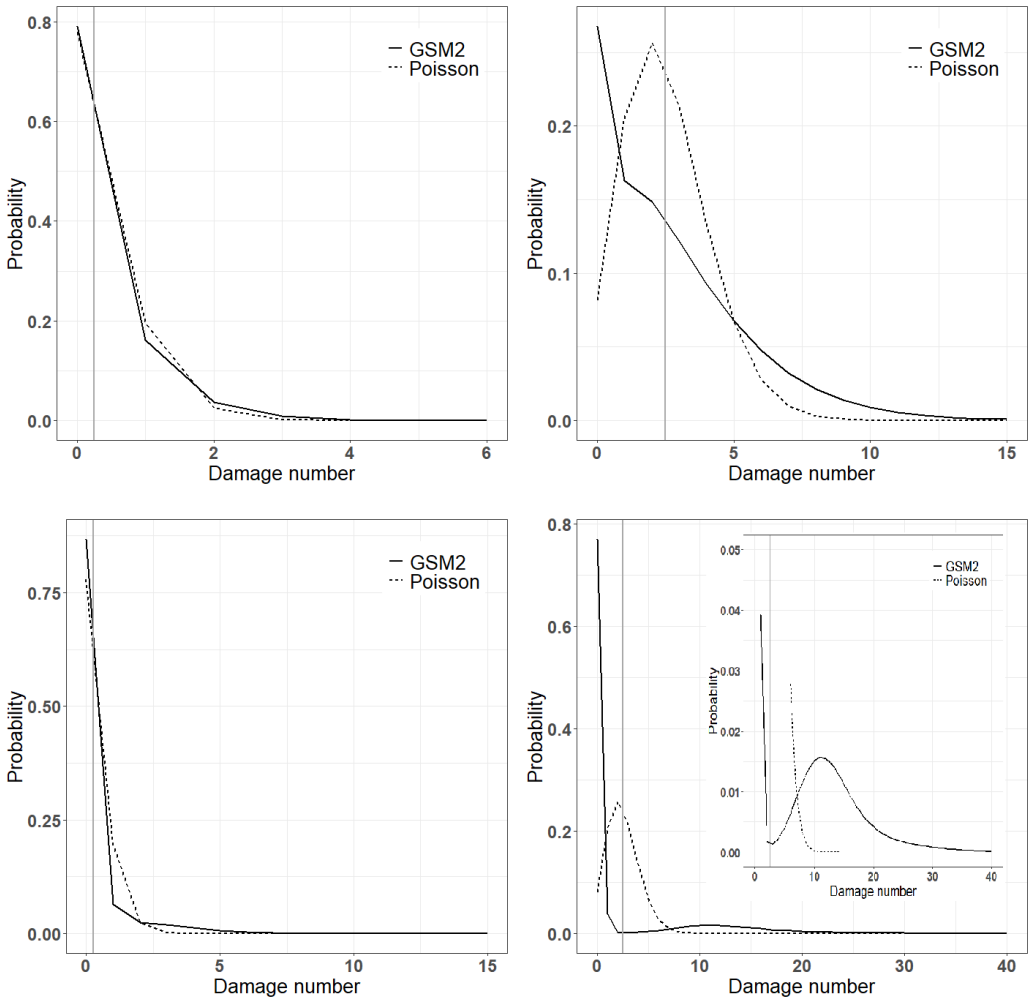

Figure 12.4: Initial sublethal DNA damage distribution for helium at (a) 10 mm residual range, z_n=0.5 Gy, and (b) 10 mm residual range, z_n=5 Gy. Initial sublethal DNA damage distribution for oxygen at (c) 2 mm residual range, z_n=0.5 Gy, and (d) 2 mm residual range, zn=5 Gy. Panel (d) also contains a zoom for probabilities up to 0.05, to better appreciate the distribution bimodality trend. For all calculations, we used r_d=0.5 μm, k=0.5.

Poisson.

We investigated the $f_n(z_n)$ multi-event microdosimetric distributions calculated with Equation (12.47) under different irradiation regimes, and compared them to the proposed approximations given in Equations (12.52)–(12.56).

The analysis is divided in two parts. First, we consider the low LET and high doses regimes approximation, as given in Equation (12.56). In this case, the multi-event distribution can be approximated by a Gaussian density function. Then, we consider the high LET and low doses regimes, where Equation (12.52) holds. The main intuition behind such approximation is that either 0 or 1 event occur, so that the multi-event distribution can be described as a mixture of probability of no hits and the single event microdosimetric distribution $f_{1;c}(z_n)$. Figure 12.5 illustrates the comparison between the multievent distribution $f_n(z_n)$ and the Gaussian approximation (12.56), for low LET ions (protons

Particle type	**Proton**				**Helium**		**Oxygen**	
Kinetic energy [MeV]	**80**		**18.6**		**80**	**18.6**	**177**	**39.8**
z_D	**1.17**		**0.75**		**1.17**	**2.87**	**4**	**20**
z_n	0.5	5	0.5	5	0.5	5	0.5	5
Mean value	0.2	2	0.2	2	0.25	2.5	0.25	2.5
$\vartheta(\kappa, z_n, Z_D)$	0.03	0.3	0.06	0.6	0.14	3.6	0.5	25

Table 12.2: Summary of main parameters for the DNA damage distributions plotted in Figures 12.2 and 12.3. For each particle type, the microdosimetric quantity z_D, the average value of the sub-lethal lesion distribution and the deviation from a Poisson distribution, as described in Equation (11.10) are reported. 3.2 Inter-cellular corrections

and helium) of 2 mm residual range. Ions were considered at same residual range in order to make a more direct comparison of different particles features in a close-to-maximum LET condition for the respective cases.

The results indicate that for protons, the Gaussian approximation is already valid at low doses. For helium, since the LET is higher than protons, the Gaussian approximation holds at higher doses.

Figure 12.6 shows the comparison between the multievent distribution $f_n(z_n)$ and the high LET bimodal approximation (12.52) for higher LET ions (carbon and oxygen) with 2 mm residual range. At 1 Gy, the agreement between the two distributions is relatively good. At the higher dose, however, the contribution coming from the probability of registering more than 1 event is appreciable, given by the second peak.

Survival computation

We computed the cell survival rate using Equation (12.48). In Figure 12.7, the survival curves for all irradiation conditions listed in Table 12.1 are shown. The biological parameters of the MME have been set to $a = 0.05$, $b = 0.001$ and $r = 2.5$, corresponding to a generic tissue type while the cell nucleus radius was $R = 5$ μm The choice of parameters has been done in accord to typical values used in MKM and its variants [Bellinzona et al., 2021]. It is further worth stressing that, despite later in the work we will extract GSM^2 from experimental data, in the current section we decided to use generic parameters not belonging to a specific cell line.

From Figure 12.7 it is clear how the survival behavior is sensitive to the physical characteristics of the radiation field, changing both the linear and the quadratic shapes.

Non Poissonian behavior of the cell survival

A comparison between the survival curves predicted with the Poissonian assumption and with GSM^2 is shown in Figure 12.9. In particular, the Poisson cell survival is computed assuming that the initial DNA damage distribution follows a Poisson distribution, so that the resulting survival curve is given by Equation (12.36). The comparison has been reported only for protons at 2 mm residual range, being the most interesting case since higher LET ions are clearly non Poissonian. Furthermore, cell survival curves have been calculated only for the cell domain, without considering the inter-cellular correction. This choice has been made to investigate whether a non Poissonian behavior already emerges locally at the domain level.

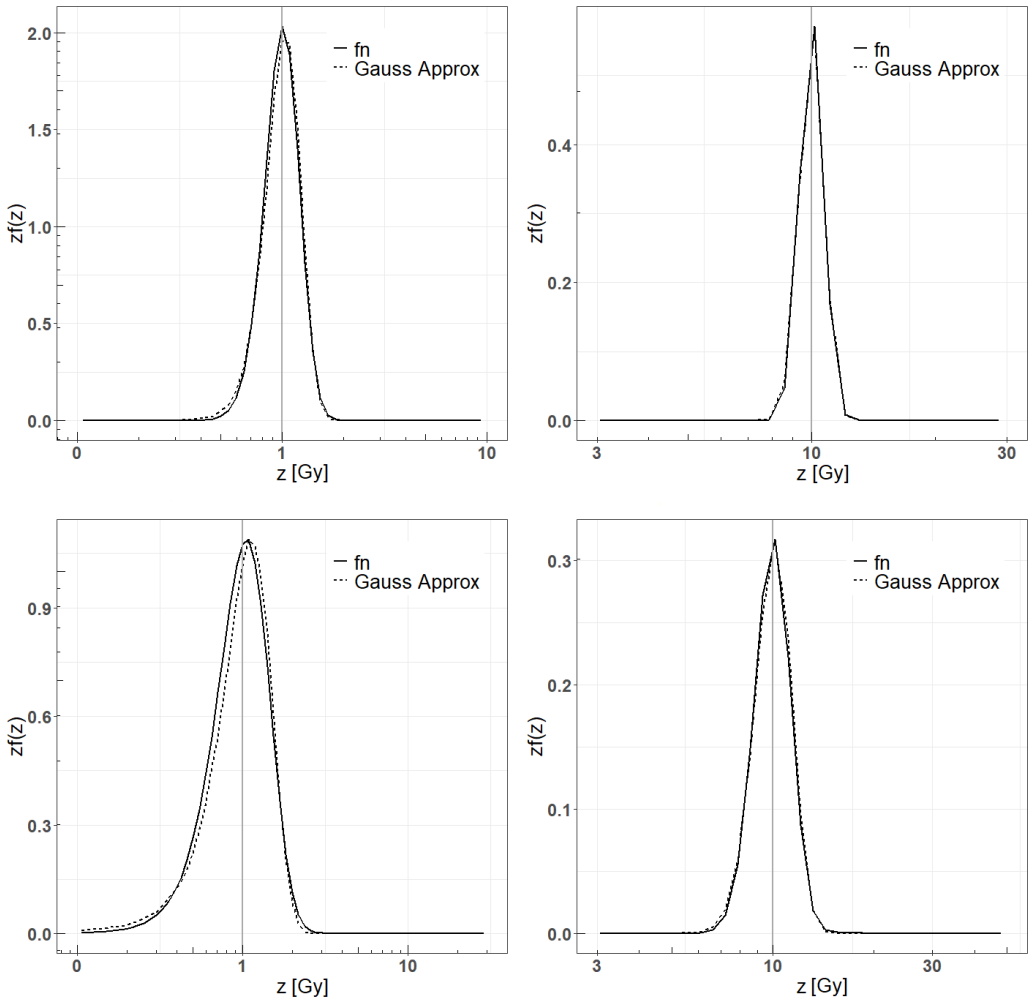

Figure 12.5: Gaussian approximation for protons and helium ions with 2 mm residual ranges at D = 1 Gy and D = 10 Gy. The plots show the full multi-event microdosimetric distribution $f_n(z_n)$ (continuous line) against the Gaussian approximation (dashed line) as defined in Equation (12.56).

The comparison shows that even at the domain level, there is a significant discrepancy between the Poisson approach and GSM^2 even for low LET protons at sufficiently high dose. The analysis on the initial datum, showing that for low dose regimes the initial distribution is almost Poissonian while it deviates at higher doses, is confirmed by the results of Figure 12.9. In fact, the two survival curves coincide at low doses while, at higher doses, a significant difference is appreciable. The Poissonian assumption implies a behavior which typically lead to underestimation of the cell survival fraction, whereas GSM^2 prediction naturally includes overkilling effects in the initial damage distribution.

Linear-quadratic-linear behavior of the cell survival curve

A study on the linear-quadratic-linear (LQL) behavior of the cell survival curve prediction via GSM^2 , as theoretically described above, is reported in the present section. In particular, in Figure 12.9, a LQL cell survival is fitted on the GSM^2 cell survival predic-

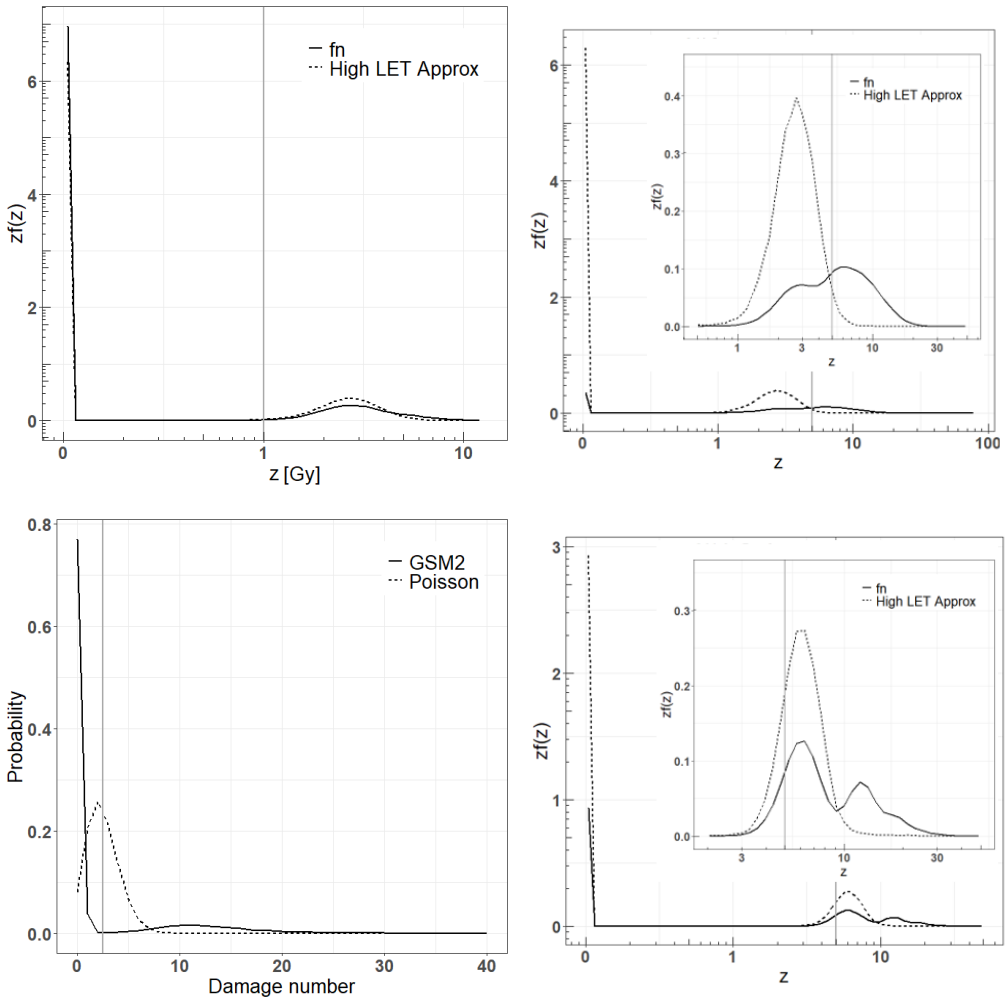

Figure 12.6: Bimodal approximation for carbons and oxygens with 2 mm residual ranges at D = 1 GyD = 5 Gy. The plots show the full multi-event microdosimetric distribution $f_n(z_n)$ (continuous line) against the bimodal approximation (dashed line) as defined in Equation (12.52). The bottom right figure reports a detail of the distribution at high z values.

tion. We remark that LQL fits three parameters, namely α and β for the LQ part and the threshold D_t above which the cell survival is assumed to be purely linear. This is the typical approach used in LEM [Friedrich et al., 2013a]. In particular, the fitted parameters are α=0.18, β=0.09, with a consequent α/β=1.94. The LQL cell survival is also compared to the purely LQ model, in which the α and β parameters previously fitted with a LQ trend are used. As in the previous section, we considered just protons with a residual range of 2 mm since they are expected to exhibit a LQL behavior, in contrast to higher LET ions in which a purely linear trend emerges.

Model fit to experimental data

To validate GSM^2 , the predicted cell survival curve has been fit to experimental data given in the PIDE dataset, [Friedrich et al., 2013b]. Resulting parameters, cell line and

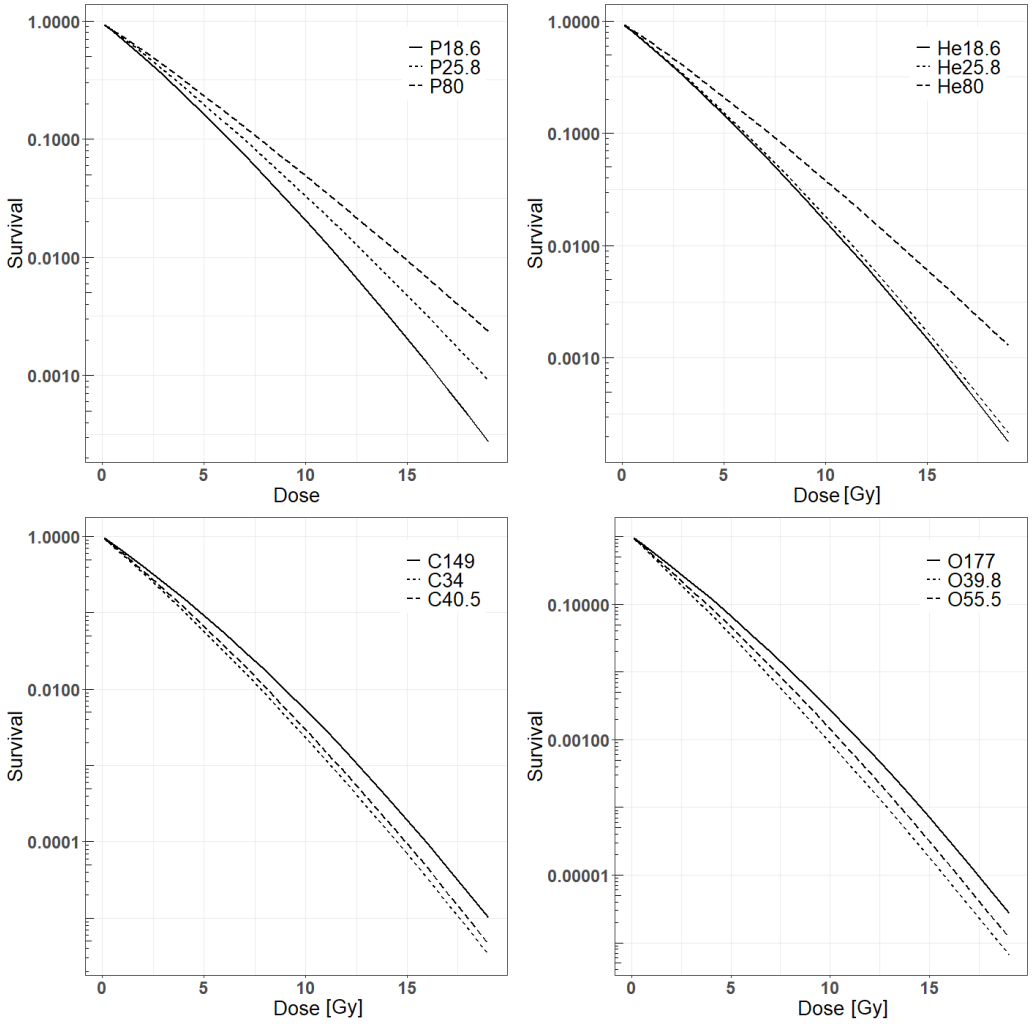

Figure 12.7: Survival curves calculated for the same cell population under different irradiation conditions listed in Table 12.1. Each panel include results for a specific ion type (protons in panel (a), helium in panel (b), carbon in panel (c) and oxygen in panel (d)) at several residual ranges of 2mm, 5mm and 10mm. All curves were calculated using the same biological parameters of the MME, namely $a = 0.05$, $b = 0.001$ and $r = 2.5$.

radiation quality have been gathered in Table 12.3. Further, Figure 12.10 shows the real experimental data plotted against GSM^2 prediction. It is worth stressing further that, differently from what done in other radiobiological models, we did not fit neither domain radius nor cell nucleus radius. On the contrary, we only optimized a, b and r parameters, whereas other parameters have been set equal to $r_d = 1$ μm, R = 5 μm, $\kappa = 1.2$ Gy^{-1}, $\lambda = \kappa\ 10^{-2}$. The choice of a domain of 1 mum is motivated by the TECP size so that energy deposition is first consider in a domain of equal size of the considered detector. In particular, it is believed that the domain radius is not a physical parameter, but it is instead a user hyperparameter used to limit long-range interaction between damages. Therefore, to avoid overfitting of the model we decided not to fit the domain radius.

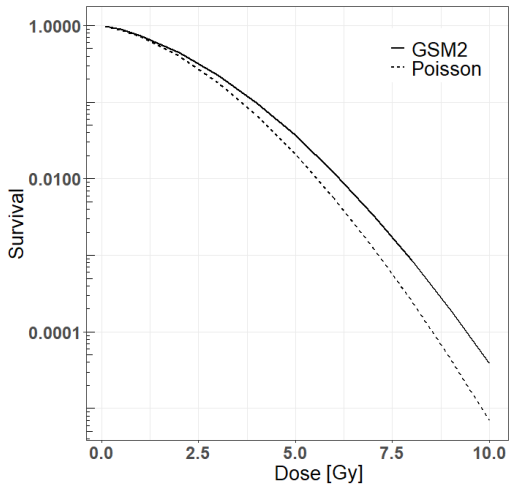

Figure 12.8: Survival curves calculated with the Poissonian assumption and GSM2 for protons at 2 mm residual range. The values have been assessed on the cell domain without taking into account inter-cellular correction.

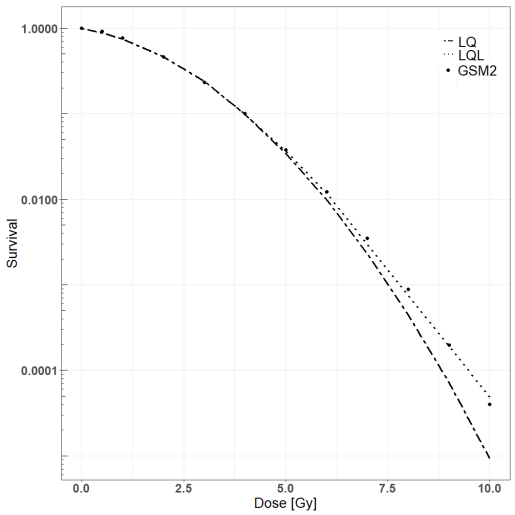

Figure 12.9: Cell survival curves comparison between a LQ and a LQL fits for protons with 2 mm residual range. Survival curves have been calculated on the domain without inter-cellular correction.

12.3 Discussion

In this chapters, we calculated cell survival probability with GSM^2 accounting for different levels of stochasticity. We obtained microdosimetric spectra for different radiation fields (protons, helium, carbon and oxygen) at three different residual ranges (2,5, 10 mm) exploiting the TOPAS microdosimetric extension. Using the energy depositions scored with TOPAS an input, we calculated multi-event microdosimetric distributions with an in-house built-in R code. We then assessed the initial DNA damage distributions, from which we predicted cell survival.

Cell line	Ion	LET [keV/μm]	Energy [MeV/u]	a [h−1]	b [h−1]	r [h−1]
V79	12C	50.3	41.7	0.011	0.009	2.2
V79	1H	2.06	27	0.011	0.01	2.9
HFIB2	1H	2.25	24	0.018	0.027	2.7
T1	4He	4.6	79	0.01	0.007	2.3

Table 12.3: GSM^2 estimated parameter using the PIDE dataset. In particular, cell line, ion type with its LET and initial kinetic energy are reported together with model dynamical parameters value.

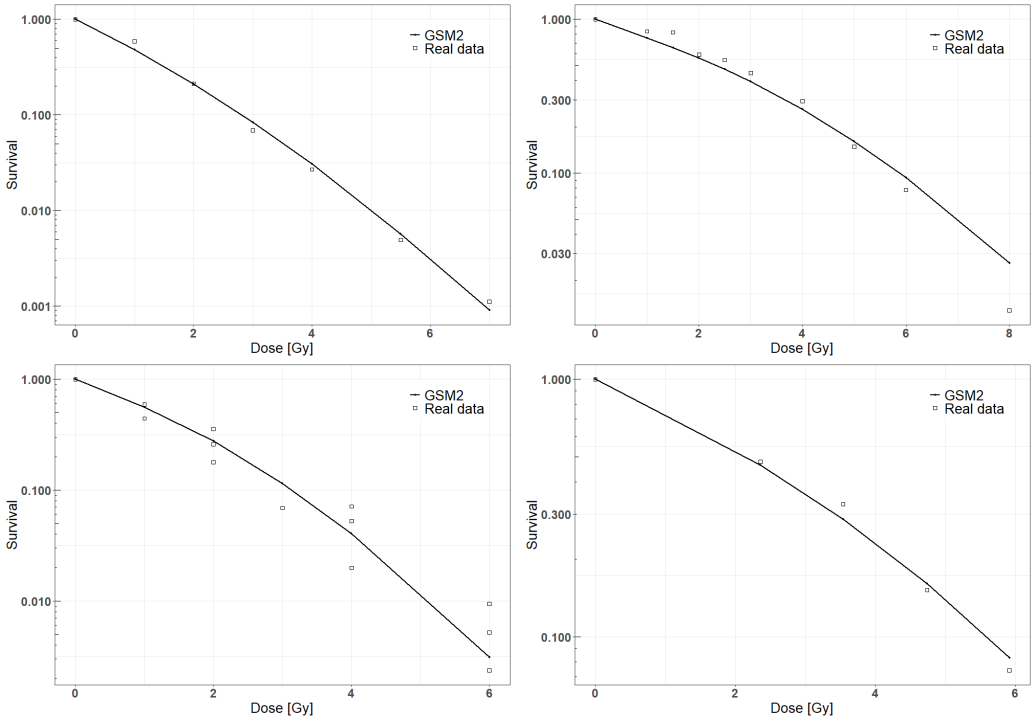

Figure 12.10: Cell survival predicted by GSM^2 (continuous line) fitted against experimental data as contained in the PIDE dataset (dotted). Top left panel shows V79 cell line with 41.7 MeV/u carbon ion, top right shows V79 cell line with 27 MeV protons, bottom left shows HFIB2 cell line with 24 MeV protons and bottom right shows T1 with 79 MeV/u helium

DNA initial damage distribution and microdosimetric inter-cellular correction

Several key aspects emerge from the calculated DNA damage distributions. First, we observe that, together with LET, also the dose can have a significant impact on Poissonianity of the DNA damage distribution. This implies that also low LET radiation can be characterized by a non negligible non Poissonian effects. For example, the discrepancy between the first two moments of the initial damage distributions caused by 5 Gy of protons ranges from 15% for the 10 mm residual range to 33% for the 2 mm residual range. We recall that the Poisson distribution is characterized by equal moments, so that the higher is the discrepancy between the first two moments, the more the distribution differs from a Poisson. Assumption on the Poissonianity of the initial damage distribution

implies an underestimation of the probability for a null DNA damage yield, resulting in a lower survival fraction prediction.

We also observed that a bimodal distribution emerges for higher LET radiation. Such behavior can be guessed already for the 2 mm residual range helium beam at 5 Gy. As the ion LET increases, such as for oxygen, bimodality becomes predominant. This trend is typically added with ad-hoc a posteriori corrections to the model [Bellinzona et al., 2021], while for GSM^2 is naturally taken into account, leading to a straightforward prediction of the overkilling effect.

We also investigated, as another level of stochasticity the inter-cellular correction to take into account the whole cell populations. Two approximations have been numerically studied to establish a range of validity. For low LET irradiations, a Gaussian approximation for the multi-event microdosimetric distribution has been proposed based on Equation (12.56). It is clear that such approximation is valid for protons even at a dose of 1 Gy. As the ion LET increases (e.g. for helium), we noted that a higher dose is need in order for the approximation to be valid. Nonetheless, at 1 Gy the discrepancy between the multi-event distribution for 2 mm residual range helium and the corresponding Gaussian approximation is relatively small. Carbon and oxygen ions have not been investigated since they are not expected to follow the Gaussian approximation at doses of therapeutic interest. Having the possibility to use a Gaussian approximation for the considered microdosimetric distribution could be extremely useful and important. In fact, the Gaussian density function is at the core of many mathematical analytical theories, leading often to straightforward and less numerically demanding computations.

The second proposed approximation is valid for high LET ions at low doses regimes. It reflects the fact that, since the number of hits is inversely proportional to z_F at sufficiently low doses, only 0 or 1 hit may occur. Both oxygen and carbon ions satisfy this approximation at 1 Gy. Nonetheless, as the dose increases, the contribution due to more than 1 single hit becomes non negligible, dropping the approximation validity. It must be said that a further approximation could be introduced by adding the contribution of a non null probability of registering two events. This approach can be useful in deriving an asymptotic behavior at low doses, because it is more easily treatable to fit the model parameters.

Cell survival curve

We computed the cell survival curves for 4 different ions (protons, helium, carbon and oxygen) at 3 different residual ranges each (2, 5, 10 mm), and investigated the difference between GSM^2 , the LQL behavior and Poissonian predictions. The survival curve behavior depends heavily on the ion type: low LET ions (protons and helium) show a more dominant contribution from the quadratic term, whereas high LET radiation (carbon and oxygen) exhibit a more linear trend.

The non Poissonian correction has also been explored. For 2 mm residual range protons, the cell survival is equivalent to a Poissonian one at low doses, while at high doses the discrepancy is non negligible. This is a direct consequence of the initial damage distribution behavior discussed in Chapter 12 GSM^2 naturally incorporates many non Poissonian behaviors: i) the non Poissonianity of the initial DNA damage distribution, and ii) non clustered lethal lesions represented by the term b in the MME. The LQL behavior typically found inthe cell survival curve predicted by GSM^2 is studied and reported in Figure 12.9. The data indicate that a purely linear quadratic curve cannot fit GSM^2 predictions, with a significant underestimation due to a non linear trend. To the best of our knoweledge,

GSM^2 is the first mechanistic model that predicts a LQL surival curve.

Model fit to experimental data

Exploiting the PIDE dataset, [Friedrich et al., 2013b], we validate GSM^2 estimating the relevant biological parameters of the model using real experimental survival data. It is worth stressing that parameters estimation performed is among the main strength of the current work; in fact, using the main formulation of GSM^2 , 11, DNA repair dynamic is needed to estimate the model parameters. Such radiobiological data are nonetheless rare and difficult to collect, making the model difficult to be implemented. Instead parameters can be estimated using cell survival data, of which an extensive database is publicly available, [Friedrich et al., 2013b] therefore making GSM^2 easily usable.

Results plotted in Figure 12.3 show a good agreement between the predicted cell survival and real experimental data. Further, parameters reported in Table 12.3 are coherent with typical parameters estimated using the MKM, [Inaniwa et al., 2013]. It is worth nonetheless stressing that, given the many approximations included into all the different MKM formulations, the parameters are not required to be identical. More important, this consideration is valid in particular for the parameter b. In fact, in the MKM dynamical equation the parameter b is typically dropped assuming it to be small. In addition, differently from MKM, we did not include domain and nucleus radius as fitted parameters. This choice has been made in order to reduce possible overfitting. At last, it is important to notice that, in the V79 cell line case, two different radiation fields lead to similar biological parameters estimation. This implies that GSM^2 potentially proves to be a robust predictor of the cell survival curve.

Chapter 13

Multiple levels of stochasticity included in GSM^2

In Section 11.2, we demonstrated that GSM^2 main equation of the, namely the *Microdosimetric Master Equation* (MME), can be linked to the main equations of the *Repair-misrepair* (RMR) model [Tobias, 1980, Tobias, 1985] under a specific Poissonian assumption. The RMR is a purely Poissonian model that has been developed to interpret radiobiological experiments with heavy ions. The present chapter investigates the connection between GSM^2 and RMR, with a focus on the cell survival curves predictions and on their dependence on the radiation quality and biological parameters.

Together with GSM^2 and the RMR, two other models will be newly introduced and studied. Such models place themselves in between RMR and the GSM^2 in terms of generality from a purely probabilistic point of view. The first model, called in the current work *Poisson GSM*^{2} , is based on the MME introduced in 11 with initial datum being a Poisson law. In this sense, we have already observed in 12 that the GSM^2 initial DNA damage assessment reproduces a Poissonian behaviour for low LET radiation. The other new model considered is the named *Dirac GSM*^{2} , which is a probabilistic model based on GSM^2 with a specific assumption on the energy deposition described by microdosimetry. Such assumption replace the microdosimetric energy spectra by its average value z_F, and in doing so, the Dirac GSM^2 neglects energy deposition fluctuations. This model is computationally less complex than GSM^2 , and therefore it can be regarded as a computationally fast version of the GSM^2 . The microdosimetric approximation for the *Dirac GSM*^{2} , to the best of our knowledge, has been already discussed in a previous study [Albright, 1989]. The DNA damage initial distribution considered in the Dirac GSM^2 is a known probability distribution studied from a mathematical point of view in [Neyman, 1939], and more recently it has been used as main distribution in the *Multi-hit model* [Vassiliev, 2012].

The aim of the current chapter is to study the validity range of different probabilistic foundations of different models, and their effects on the survival curve predicted by the models themselves. This study will also highlight how different stochastic effects are naturally embedded in the GSM^2 original formulation. In particular, such terms are usually referred to in literature as non-Poissonian effects, since they are non-linear terms that contribute to deviates the true underlying probability distribution from a Poisson random variable. We will show how such non-Poissonian behavior of the number of DNA damages may arise either at the DNA damage formation level or also due to cluster effects in the kinetic evolution of lesions.

We analytically studied the survival curve predicted by the four models, i.e. the

RMR, the Poisson GSM^2, the Dirac GSM^2 and GSM^2, characterizing the asymptote of the log-survival at both low and high dose limits. We will analyze and compare the high and low dose limits, studying similarities and differences as well as their dependence on either biological or radiation quality parameters. Then, we will further validate the results using Monte Carlo simulations of microdosimetric spectra for different ions of therapeutic interest, such as protons and carbon ions at different energies.

13.1 Theory and calculations

The Repair-misrepair model

The RMR model considers that the amount of DSBs in the DNA, $U(t)$, is linearly proportional to the radiation dose rate. A number of DSBs evolve in lethal lesions, $L(t)$, while most breaks are successfully repaired with a first-order process. The model includes also the possibility of a misrepair as a second-order process, since it involves two broken DNA strands to form a chromosomal aberration.

These assumptions yield the following kinetic Equations:

$$\begin{cases} \frac{d}{dt}U(t) &= -\rho U - \psi U^2\,, \\ \frac{d}{dt}L(t) &= (1-\phi)\rho U + \sigma\psi U^2\,. \end{cases} \tag{13.1}$$

where ρ is the rate at which the DSBs are repaired, ψ is the rate constant for second-order DSB interactions, and ϕ is the fraction of simple repairs that are successful. The fraction of misrepairs that result in a lethal lesion is called σ.

Using also the MME given in Equation (11.3), in what follows, we will denote by $\mathbb{E}$ the mean value of a random variable defined as

$$\begin{aligned} \bar{x}(t) &:= \mathbb{E}[X(t)] = \sum_{x,y\geq 0} xp(t,y,x)\,, \\ \bar{y}(t) &:= \mathbb{E}[Y(t)] = \sum_{x,y\geq 0} yp(t,y,x)\,. \end{aligned}$$

Note that the following holds true

$$\begin{aligned} \sum_{x,y\geq 0} x\mathcal{E}^{i,j}\left[f(y,x)p(t,y,x)\right] &= -\mathbb{E}jf(Y,X)\,, \\ \sum_{x,y\geq 0} y\mathcal{E}^{i,j}\left[f(y,x)p(t,y,x)\right] &= -\mathbb{E}if(Y,X)\,. \end{aligned} \tag{13.2}$$

Therefore, by multiplying the MME (11.3) by x and y, we obtain using (13.2)

$$\begin{cases} \frac{d}{dt}\mathbb{E}[Y(t)] &= b\mathbb{E}[X(t)(X(t)-1)] + a\mathbb{E}[X(t)]\,, \\ \frac{d}{dt}\mathbb{E}[X(t)] &= -2b\mathbb{E}[X(t)(X(t)-1)] - (a+r)\mathbb{E}[X(t)]\,. \end{cases} \tag{13.3}$$

Setting the parameters in the RMR driving Equations (13.1) as

$$\psi = 2b\,, \quad \sigma = \frac{1}{2}\,, \quad \rho = a+r\,, \quad \phi = \frac{r}{a+r}\,,$$

Equations (13.1) become

$$\begin{cases} \frac{d}{dt}U(t) &= -(a+r)U - 2bU^2\,, \\ \frac{d}{dt}L(t) &= aU + bU^2\,. \end{cases} \tag{13.4}$$

Equations (13.4) and (13.3) have a similar form, but they are not identical. In particular, in Equation (13.4) it appears a second order moment $\mathbb{E}[X(t)(X(t)-1)]$. Assuming a Poissonian distribution for the random variable X, we obtain that

$$\mathbb{E}[X(t)(X(t)-1)] = \mathbb{E}[X(t)]^2$$

so that Equation (13.4) becomes

$$\begin{cases} \frac{d}{dt}\bar{y}(t) &= b\bar{x}^2(t) + a\bar{x}(t)\,, \\ \frac{d}{dt}\bar{x}(t) &= -2b\bar{x}^2(t) - (a+r)\bar{x}(t)\,. \end{cases} \tag{13.5}$$

which are exactly the RMR model main kinetic Equations (13.1).

Remark 13.1.1. Typically to solve an infinite system of interconnected moments equations, the so–called *mean–field* assumption is required, that is

$$\mathbb{E}[X(t)(X(t)-1)] \sim \mathbb{E}[X(t)]^2\,.$$

A comment on the *mean–field assumption* is needed. When x is large enough, we have that $\mathbb{E}[X(t)(X(t)-1)] \sim \mathbb{E}[X(t)]^2$. Therefore, the *mean field assumption* means that $\mathbb{E}[X(t)(X(t)-1)] - \mathbb{E}[X(t)]^2 \sim 0$, whose last term is nothing but the variance. Recalling that the variance for a random variable is null if and only if the random variable is deterministic, if the *mean field assumption* is realistic then the number of lesion does not differ much from the mean value, so that everything we need to know is the mean value. On the contrary, if there are evidence that the mean value is not a realistic approximation for the number of lesion, then the *mean–field assumption* must be considered unrealistic, and the knowledge of the entire probability distribution becomes essential to fully describe the system.

Using the connection between the MME and the RMR, we can explicitly solve Equations (13.1). In particular, we have that

$$\frac{d}{dt}U(t) = -(a+r)U - 2bU^2\,, \quad U(0) = \kappa D\,. \tag{13.6}$$

Equation (13.6) is known in literature as the Bernoulli equation. In order to solve it, we can apply the transformation $u = \frac{1}{U}$, so that an explicit calculation leads to the following differential Equation

$$\dot{u}(t) = (a+r)u(t) + 2b\,.$$

This is a linear equation in u, and the explicit solution is given by

$$u(t) = ce^{(a+r)t} - \frac{2b}{(a+r)}\,.$$

Coming back to the original Equation (13.6) we obtain

$$U(t) = \frac{(a+r)}{ce^{(a+r)t} - 2b}\,, \tag{13.7}$$

with

$$c := \frac{(a+r)}{\kappa D} + 2b\,.$$

We can therefore substitute U in Equation (13.7) into the Equation for L in (13.1) to obtain

$$\begin{aligned}\frac{d}{dt}L(t) &= aU + bU^2 = \\ &= -\frac{1}{2}\left(-(a+r)U(t) - 2bU^2(t)\right) + \frac{a-r}{2}U(t) = \\ &= -\frac{1}{2}\dot{U}(t) + \frac{a-r}{2}U(t)\,,\end{aligned}$$

and after integration we obtain

$$L(t) = \frac{1}{2}\left(\kappa D - U(t)\right) + \frac{a-r}{4b}\log\left(-\frac{2b\kappa D - e^{(a+r)t}(a+r+2b\kappa D)}{e^{(a+r)t}((a+r)}\right)\,. \tag{13.8}$$

We can eventually calculate the long-time convergence toward the stationary solution of Equations (13.7)–(13.8) to be

$$\begin{aligned}\lim_{t\to\infty} U(t) &=: U_\infty = 0\,,\\ \lim_{t\to\infty} L(t) &=: \frac{1}{2}\kappa D + \frac{a-r}{4b}\log\left(1 + \frac{2b\kappa D}{a+r}\right)\,.\end{aligned} \tag{13.9}$$

The survival probability at time t is thus obtained by computing the probability of having 0 lethal or sublethal lesions, under the assumption that they follow a Poissonian law. This implies, using Equations (13.9), that

$$S_{RMR}(D) = \exp\left[-L_\infty\right] = e^{-\frac{1}{2}\kappa D}\left(1 + \kappa D\frac{2b}{a+r}\right)^{\frac{r-a}{4b}}\,. \tag{13.10}$$

Proceeding as above we can calculate the tangent of the cell–survival logarithm at low doses, i.e.

$$\frac{d}{dD}\log(S_{RMR}(D)\Big|_{D=0} = -\kappa(\frac{a}{a+r})\,. \tag{13.11}$$

In the same way, at as the doses increases, the tangent of the log-survival becomes

$$\frac{d}{dD}\log(S_{RMR}(D)\Big|_{D=\infty} = -\frac{1}{2}\kappa\,. \tag{13.12}$$

The GSM^2 in the Poisson initial damage case

As studied in details in Section 11.3, the initial damage–distribution according to the GSM^2 does not necessarily follow a Poisson distribution. Nonetheless, in certain regimes such as low-dose and low-LET, the initial distribution is significantly closer to a Poisson law of average value κD. In this case, the initial–damage distribution described by Equation (11.9) simplifies to

$$p_0^X(x) = e^{-\kappa D}\frac{(\kappa D)^x}{x!}\,. \tag{13.13}$$

We consider the MME (11.3) coupled with a Poisson initial condition (13.13). The resulting Poisson GSM^2 will only exhibits non-Poissonian effects due to the kinetic recombination of sub-lethal lesions.

Therefore, plugging-in the initial damage distribution (13.13) into the survival Equation (12.20), we obtain

$$S_P(D) = e^{-\kappa D}\left(1 + \sum_{x_0=1}^{\infty} \frac{(\kappa D)^x}{x!} C(x_0)\right), \tag{13.14}$$

Although the initial-damage distribution is Possonian, the predicted cell-survival Equation (13.14) includes some non-Poissonian terms coming from clustered terms linked to the coefficient b in the factor $C(x_0)$.

Again, we can study the log-survival tangent at high and low doses and obtain

$$\begin{aligned} \frac{d}{dD}\log(S_P(D))\Big|_{D=0} &= -\kappa(\frac{a}{a+r}), \\ \frac{d}{dD}\log(S_P(D))\Big|_{D=\infty} &= -\frac{1}{2}\kappa. \end{aligned} \tag{13.15}$$

The Dirac GSM^2 in the $f_\nu(z) \approx \delta(z - \nu z_F)$ case

According to the GSM^2, the initial damage distribution at a given dose D is given by

$$p_0^X(x) := \sum_{\nu\geq 0} \int_0^\infty p_z^X(x|\kappa z) \frac{e^{-\frac{D}{z_F}}}{\nu!} \left(\frac{D}{z_F}\right)^\nu f_\nu(z) dz. \tag{13.16}$$

In order to simplify the GSM^2 initial damage distribution, we can notice that the multi-event distribution $f_\nu(z)$ is sharply peaked around its average z_F. Neglecting all the fluctuations around the mean value, we can assume that

$$f_\nu(z) \approx \delta(z - \nu z_F). \tag{13.17}$$

Equation (13.17) implies that the energy deposition by ν events equals νz_f, so that every event always deposits z_F. This assumption is clearly strong and not always valid. In fact, in the case of a low-LET radiation it is true that f_ν is peaked at νz_F for a number of events close to the mean number of events $\frac{D}{z_F}$. On the contrary, if the number of events is significantly smaller than $\frac{D}{z_F}$, then the multi-event distribution is not well represented by the approximation (13.17). In this case, that is whenever $\nu << \frac{D}{z_F}$, we have that the term

$$\frac{e^{-\frac{D}{z_F}}}{\nu!}\left(\frac{D}{z_F}\right)^\nu,$$

in the initial damage distribution (13.16) is small, so that the approximated damage distribution according to Equation (13.17) can still be sufficiently close to the original damage distribution.

Under the assumption (13.17), the initial damage distribution (13.16) becomes

$$p_0^X(x) = \sum_{\nu\geq 0} e^{-\kappa\nu z_F} \frac{(\kappa\nu z_F)^x}{x!} \frac{e^{-\frac{D}{z_F}}}{\nu!} \left(\frac{D}{z_F}\right)^\nu. \tag{13.18}$$

The distribution appearing in Equation (13.21) is known in literature as *Neyman distribution*, and it has been deeply studied and characterized in [Neyman, 1939].

The approximated initial damage distribution (13.21) can be inserted into the survival Equation (12.20), to obtain

$$S_D(D) = p_0^X(0|D) + \frac{r}{a+r} p_0^X(1|D) + \sum_{x_0=2}^{\infty} p_0^X(x_0|D) C(x_0)\,, \tag{13.19}$$

where $p_0^X(x_0|D)$ is given by Equation (13.21).

The the log-survival tangent at low doses is thus given by

$$\begin{aligned}\frac{d}{dD}\log(S(D)\Big|_{D=0} &= -\frac{1}{z_F}\left(1 - e^{-\kappa z_F}\right) + \\ &+ \frac{1}{z_F}\sum_{x\geq 1} C(x_0)\frac{e^{-\kappa z_F}}{x!}\left(\kappa z_F\right)^x\,,\end{aligned}$$

whereas the tangent to the cell-survival curve at high dose is given by

$$\frac{d}{dD}\log(S(D)\Big|_{D=\infty} = -\frac{1}{z_F}\left(1 - e^{-\kappa z_F} - \frac{1}{2}e^{-\kappa z_F}\kappa z_F\right)\,.$$

Connection to the Multi–hit model

As briefly mentioned in 12, Equation (12.20) can be seen as a microdosimetric generalization of the multi–hit model to account for the damage dynamics and repair.

In particular, a non–Poissonian multi–hit model has been derived in [Vassiliev, 2012]. The survival curve is given by

$$S = \sum_{k=0}^{n-1} p(k)\,, \quad p(k) = e^{-a-ae^{-b}}\frac{b^k}{k!}B_k(ae^{-b})\,, \tag{13.20}$$

with B_k the *Bell's polynomial*.

Under assumption (13.17), the initial damage distribution, expressed by (13.16), is given by

$$p_0^X(x) = \sum_{\nu\geq 0} e^{-\kappa\nu z_F}\frac{(\kappa\nu z_F)^x}{x!}\frac{e^{-\frac{D}{z_F}}}{\nu!}\left(\frac{D}{z_F}\right)^{\nu}\,. \tag{13.21}$$

Choosing

$$a = \frac{D}{z_F}\,, \quad b = \kappa z_F\,,$$

we can rewrite $p_0^X(x)$ of Equation (13.20), using the relation

$$\sum_{i=0}^{\infty}\frac{i^k x^i}{i!} = e^x B_k(x)\,,$$

and obtain

$$p_0^X(x) = e^{-\frac{D}{z_F}\left(1-e^{-\kappa z_F}\right)}\frac{\kappa z_F^x}{x!}B_k\left(\frac{D}{z_F}e^{-\kappa z_F}\right)\,. \tag{13.22}$$

The distribution given in Equation (13.22) provides a microdosimetric version of the distribution given in Equation (13.20) and used in [Vassiliev, 2012].

GSM^2 survival curve given in Equation (12.20) provides a generalization to the survival predicted by the *Multi-hit model* in [Vassiliev, 2012] with the following improvements:

	Low-dose tangent	High dose tangent
RMR	$-\kappa(\frac{a}{a+r})$	$-\frac{1}{2}\kappa$
Poisson GSM^2	$-\kappa(\frac{a}{a+r})$	$-\frac{1}{2}\kappa$
Dirac GSM^2	$-\frac{1}{z_F}\left(1-e^{-\kappa z_F}\right)+$ $+\frac{1}{z_F}\sum_{x_0\geq 1} C(x_0)\frac{e^{-\kappa z_F}}{x_0!}\left(\kappa z_F\right)^x$	$-\frac{1}{z_F}\left(1-e^{-\kappa z_F}-\frac{1}{2}e^{-\kappa z_F}\kappa z_F\right)$
GSM^2	$-\frac{1}{z_F}\int_0^\infty\left(1-e^{-\kappa z}\right)f_1(z)dz+$ $+\frac{1}{z_F}\sum_{x_0\geq 1} C(x_0)\int_0^\infty\frac{e^{-\kappa z}}{x_0!}\left(\kappa z\right)^x f_1(z)dz$	$-\frac{1}{z_F}\int_0^\infty\left(1-e^{-\kappa z}\right)f_1(z)dz+$ $+\frac{1}{2z_F}\int_0^\infty e^{-\kappa z}\kappa z f_1(z)dz$

Table 13.1: Low and high dose log-survival tangent limits for the RMR, Poisson GSM^2, Dirac GSM^2 and GSM^2.

- the parameters a and b given in [Vassiliev, 2012] have a natural physical and mechanistic interpretation derived from microdosimetric considerations;
- the maximum number of hits is not a model parameter, but it is derived again by physical arguments;
- each term $p_0^X(x)$ is weighted by a suitable probability that accounts for a possible repair;
- a term accounting for direct killing is added.

Summary

Putting together the results presented previously, we have the following tangents of the log-survival predicted including different levels of stochasticity.

In particular, considering the high and low–dose asymptotes in the GSM^2 with assumption (13.17), we recover the high and low–dose asymptotes of the Dirac GSM^2. Further, if we consider a low–LET radiation, namely $\kappa z_F << 1$ in the Dirac GSM^2, we obtain

$$-\frac{1}{z_F}\left(1-e^{-\kappa z_F}\right)+\frac{1}{z_F}\sum_{x_0\geq 1} C(x_0)\frac{e^{-\kappa z_F}}{x_0!}\left(\kappa z_F\right)^{x_0}\sim_{\kappa z_F<<1}-\kappa(\frac{a}{a+r})$$
$$-\frac{1}{z_F}\left(1-e^{-\kappa z_F}-\frac{1}{2}e^{-\kappa z_F}\kappa z_F\right)\sim_{\kappa z_F<<1}-\frac{1}{2}\kappa$$

so that for a low–LET radiation quality we recover a Poissonian behaviour.

13.2 Numerical results

The numerical results of the theoretical predictions described in Section 13.1 are presented here. The microdosimetric spectra have been simulated using TOPAS Monte Carlo toolkit, 5.2, exploiting its microdosimetric extension. We employed the spherical TEPC geometry available in the code, and simulated four radiation fields: (i) 80 MeV and (ii) 18.6 MeV proton pencil beam, (iii) 34 MeV/u carbon ion pencil beam and (iv) 70 MeV/u carbon ion with 2 cm of water upstream of the TEPC. The first three beams have been already used in Chapter 12, whereas the latter have been chosen to study a realistic mixed

radiation field. Typically assumed approximations used in many existing radiobiological models fail in a heterogeneous field, where GSM^2 might give an accurate prediction. The biological parameters used for the calculations are: $a = 0.01$, $b = 0.01$ and $r = 2.9$. Such parameters have been fitted in Chapter 12 to predict the survival curve for the V79 cell line. We also set $\kappa = 1$, and considered an enhancement of κ depending on the LET according to [Chen et al., 2017]. We will analyze how the selected radiation fields may affect the cell survival prediction.

Figure 13.1 reports the multi-event microdosimetric distribution (2.2) compared to the Dirac approximation as described in Equation (13.17) for a given dose $D = 1$ Gy delivered by the four radiation fields.

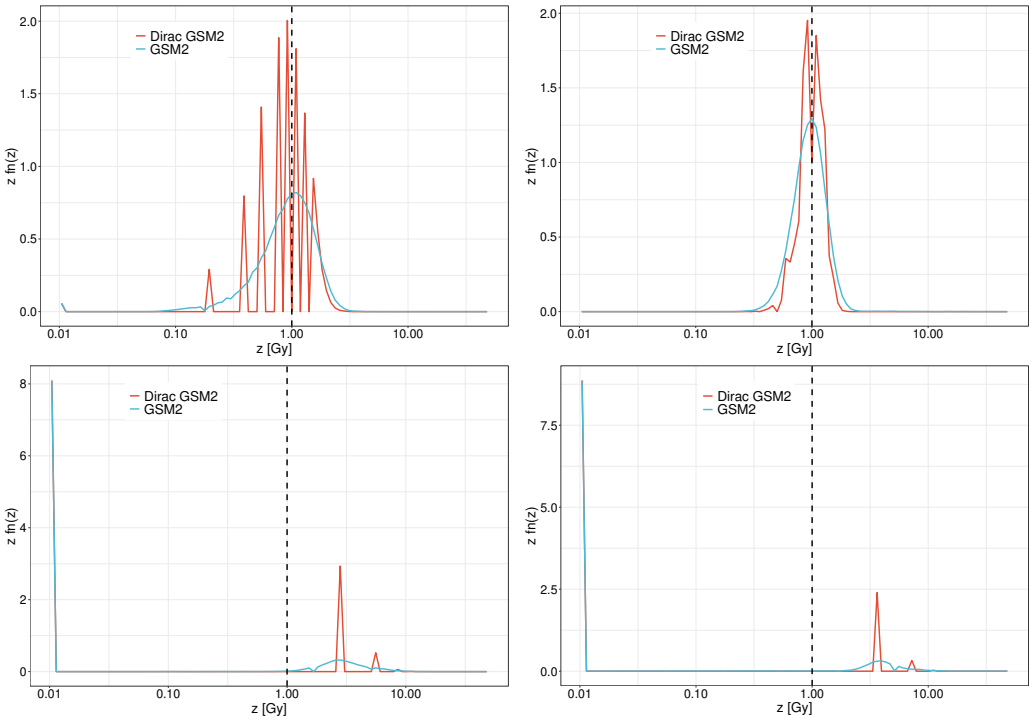

Figure 13.1: Comparison between the multi–event microdosimetric distribution f_n (blue line) of Equation (2.2), and to the Dirac approximation multi-event distribution (red line) at $D = 1$ Gy. The vertical black dotted line marks the average of the distribution, which corresponds to $D = 1$ Gy. Panels refer to (i) protons at 18.6 MeV (top left), (ii) protons at 80 MeV (top right), carbon ions at 24 MeV (bottom left) and (iv) carbon ions after 2 cm of water (bottom right).

The data for 18.6 MeV (panel (a)) and 80 MeV proton beams (panel (b)) suggest that the Dirac approximation, despite having a clear Dirac comb-like behaviour, reproduces well the first two moments of the multi-event distribution. Panel (c) show how a slight underestimation in the Dirac approximation of the variance appears in the 34 MeV/u carbon beam. Finally, the data for the mixed-field radiation beam illustrated in panel (d), we observe a significant deviation from the Dirac approximated distribution.

In Figure 13.2, we report the same study of Figure 13.1 at a higher dose of $D = 10$ Gy. For both proton beams (panel (a) and (b)), the goodness of the Dirac approximation is confirmed. For 34 MeV/u carbon ions (panel (c)), the approximated Dirac distribution is in a better agreement than at $D = 1$ Gy. The mixed field scenario still show clear

discrepancies, especially in the second moment of the multi-event distribution.

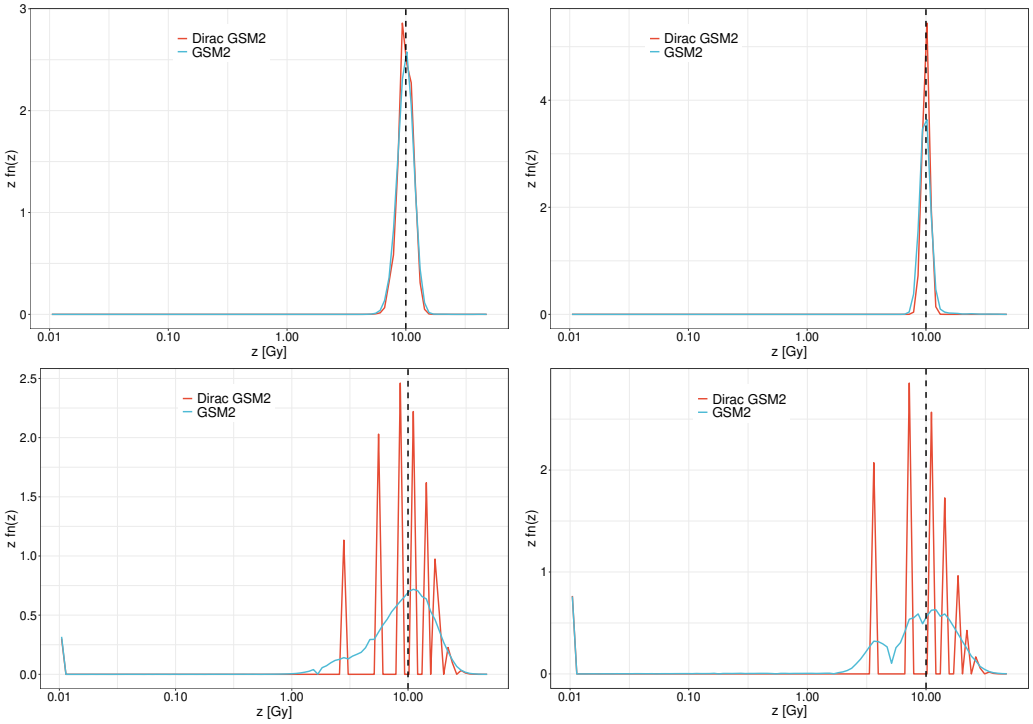

Figure 13.2: Comparison between the multi–event microdosimetric distribution f_n (blue line) of Equation (2.2), and to the Dirac approximation multi-event distribution (red line) at $D = 10$ Gy. The vertical black dotted line marks the distribution average corresponding to $D = 10$ Gy. Panels refer to (i) protons at 18.6 MeV (top left), (ii) protons at 80 MeV (top right), carbon ions at 24 MeV (bottom left) and (iv) carbon ions after 2 cm of water (bottom right).

Figure 13.3 shows the comparison of the sub-lethal damage distribution as described by GSM^2 , by the Dirac GSM^2 and the Poisson distribution at low doses, i.e. $D = 1$ Gy. Panel (a) and (b) illustrate the data for high- and low-LET protons, respectively, showing how the three distributions are similar, with a slight discrepancy in the Poisson GSM^2 case. For carbon ions (panel (c)), the Dirac GSM^2 and the GSM^2 damage distributions are identical, whereas the Poisson GSM^2 is significantly different. As in Figures 13.1–13.2, both the Dirac damage distribution and the Poisson density differ from GSM^2 predictions for the mixed field.

We repeated the calculations plotted in Figure 13.4 for $D = 10$ Gy. The same comments of previous figure hold also in the higher dose scenario, with no significant differences to be reported.

Figure 14.3 illustrates the cell-survival curve predicted by the four different models for four radiation fields. For 18.6 MeV protons (panel (a)), all models agree at low dose. As the dose increases, the models start to differ, with the RMR giving a lower survival probability, followed by the Poisson GSM^2 and then the Dirac GSM^2 and the GSM^2 predicting the higheest survival fraction (panel (b)).

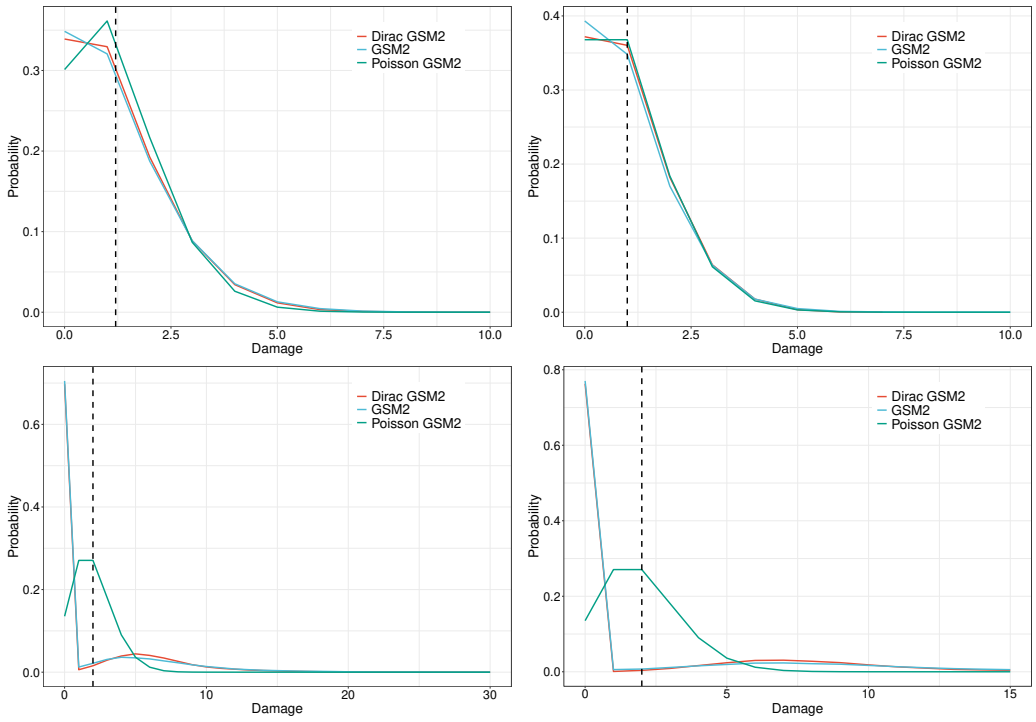

Figure 13.3: Comparison between the sub–lethal damage distribution (blue line), as given in Equation (11.9), the Dirac approximation sub–lethal distribution of Equation (13.21) (red line), and a Poisson distribution (green line) at $D = 1$ Gy. The vertical black dotted line marks the distributions average . Panels refer to (i) protons at 18.6 MeV (top left), (ii) protons at 80 MeV (top right), carbon ions at 24 MeV (bottom left) and (iv) carbon ions after 2 cm of water (bottom right).

13.3 Discussion and Conclusions

In Section 13.1 we have studied four different models, namely the RMR, the Poisson GSM^2 , the Dirac GSM^2 and the GSM^2 , that include different levels of stochasticity. In particular, the RMR assumes a Poissonian initial damage distribution as well as a Poissonian dynamics, and the kinematic equations are purely deterministic, describing the time evolution of the damages average values. Also the Poisson GSM^2 neglects non–Poissonian stochastic fluctuations in the initial damage distribution, but includes Poissonian effects of repair and cluster death using the MME (11.3). The Dirac GSM^2 is instead a non-Poissonian model, in the sense that both the initial damage distribution and the dynamical equation include the non-Poissonian effects of overkilling and clustered death. For this reason, the Dirac GSM^2 can be regarded as a faster implementation of the GSM^2 , since the initial damage distribution calculation is numerically less demanding than GSM^2 original formulation. Nonetheless, the second moment of the microdosimetric distribution z_D is not included into the Dirac GSM^2 , so that energy deposition fluctuations are not fully described. This approximation is valid for monoenergetic pencil beams, where the microdosimetric distribution is in fact sharply peaked at the average value z_F. GSM^2 is the more general model of the four taken into account in terms of stochasticity. This model accounts for non-Poissonian effects both at the initial and dynamical levels. Furthermore, the energy deposition variations at micron scale are included into the model using a mi-

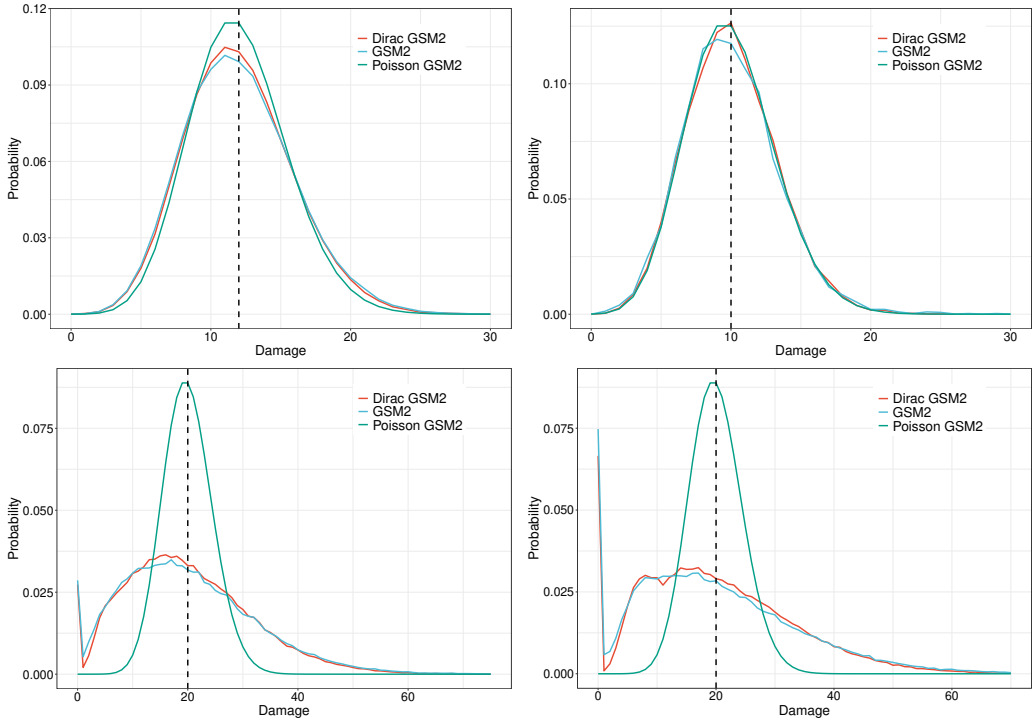

Figure 13.4: Comparison between the sub–lethal damage distribution (blue line), as given in Equation (11.9), the Dirac approximation sub–lethal distribution of Equation (13.21) (red line), and a Poisson distribution (green line) at $D = 10$ Gy. The vertical black dotted line marks the distributions average . Panels refer to (i) protons at 18.6 MeV (top left), (ii) protons at 80 MeV (top right), carbon ions at 24 MeV (bottom left) and (iv) carbon ions after 2 cm of water (bottom right)..

crodosimetry approach. Such generality is believed to be significantly different especially for mixed radiation fields. An overview of stochasticity levels that are included in the four models is given in Table 13.2.

In Section 13.1, we showed how the low and high dose log-survival tangent predicted by the four models may be significantly different and, most important, how different information on the radiation quality is reflected into the predicted cell survival curve. We summarize the results in Table 13.1. The RMR and the Poisson GSM^2 predict cell survivals with the same tangent at both low and high dose. Nonetheless, the two cell survival curves may have different curvatures due to non-Poissonian effects included in the GSM^2 via the MME (11.3). In addition, both asymptotes of the low and high dose do not depend on the radiation quality. Regarding the biological parameters, the low dose limits depend on the linear rates a and r, and the rate of log-survival decrease is proportional to the rate at which a sub-lethal lesion dies normalized to the overall linear rates a and r, namely $\frac{a}{a+r}$. Also the quadratic rate given by b does not appear in the low-dose asymptote, since it gives a term of order 2, which is negligible compared to the linear terms. The high-dose asymptotics depends only on the direct damage parameter κ, since at infinite dose only the direct damage matters and no dynamics is involved.

The Dirac GSM^2 has both high and low–dose log-survival tangent that depend on the radiation quality only via the first microdosimetric moment z_F. This is a direct consequence of Equation (13.17), where all the fluctuations in the energy deposition are assumed

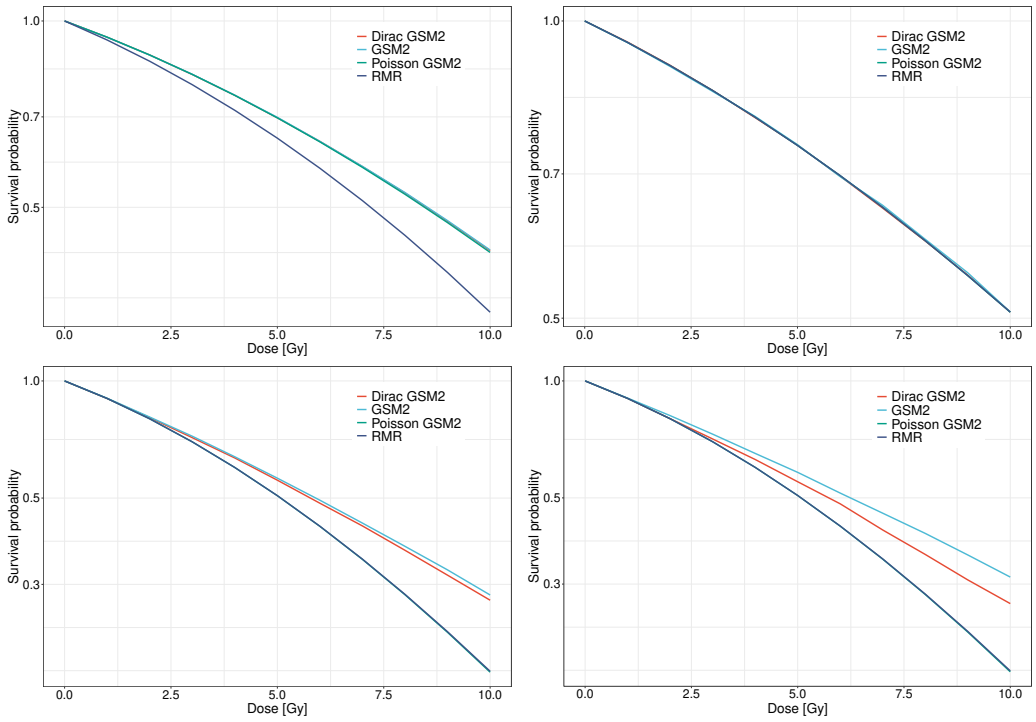

Figure 13.5: Survival curves for the GSM^2 (blue line), the Dirac GSM^2 (red line), Poisson GSM^2 (green line) and the RMR (purple line), at different doses. Panels refer to (i) protons at 18.6 MeV (top left panel), (ii) protons at 80 MeV (top right panel), carbon ions at 24 MeV (bottom left panel) and (iv) carbon ions after 2 cm of water (bottom right panel).

to be negligible. In this case, the low-dose limit is more complex than in the Poisson case, with an explicit dependence on the damage complexity given by the summation over x_0 weighted for the repair weighting function $C(x_0)$. The high–dose limit depends as before on the direct damage parameter κ with no repairs involved. Unlike the Poisson case, the asymptote is more complex and the radiation quality affects the limit via the first moment z_F.

GSM^2 shows low and high–dose limits that depends on the entire microdosimetric spectrum of the radiation field, together with a general description of the DNA–damage complexity. At high-doses only the direct damage is relevant.

As shown in Section 13.1, the models asymptotes are strictly related to each other, in the sense that starting from GSM^2 , we can recover the other models asymptotes by making specific assumptions. In particular, if we assume a Dirac-like microdosimetric initial distribution in the GSM^2 , we obtain the tangents predicted by the Dirac GSM^2 . Furthermore, for a low–LET radiation, meaning for $\kappa z_F << 1$, the Dirac GSM^2 gives the same asymptotes as the Poisson GSM^2 and the RMR.

Figures 13.1–13.2 show that the multi–event distribution computed with the Dirac assumption (13.17) is a Dirac comb-like version of the full multi–event distribution f_n. For the two proton beam energies, the Dirac multi–event distribution reproduces well both the mean and variance of the multi–event distribution f_n, and the distributions of the two ions converge to similar values especially as the dose increases. For the carbon beam, a larger discrepancy than for protons can be seen, with the Dirac distribution still being the best in reproducing the main features of the distribution f_n. In particular, both distribu-

	Non Poissonian effects		Radiation quality dependence	
	Initial damage	Dynamical	First moment	Microdosimetric distribution
RMR	×	×	×	×
Poisson GSM^2	×	√	×	×
Dirac GSM^2	√	√	√	×
GSM^2	√	√	√	√

Table 13.2: Levels of non–Poissonian effects included into the models, both for the initial damage distribution and for the kinetics equations. The radiation quality depends on the high and low-dose log-survival asymptotics, either including the first moment of the microdosimetric distribution or the whole microdosimetric spectra.

tions shares a similar description of the no-hit probability, which is extremely relevant for including the overkilling effects of high–LET radiation. At high dose, however, the two distributions appear to have a slightly different variance, which result in a discrepancy of the predicted survival fraction. The results for mixed field radiation, on the contrary, show clear discrepancies between the distributions for GSM^2 and Dirac GSM^2 , suggesting that a radiation quality description only based on the microdosimetric first moment is inadequate. Figures 13.1–13.2 suggests that although the Dirac assumption 13.17 completely neglects the energy deposition fluctuations, it is still able to provide an accurate description of the energy deposition at the micron-scale. This is reasonable since for a monoenergetic pencil beam the microdosimetric distribution is sharply peaked around its average. As the Dirac–like multi–event distribution neglects the fluctuations in the energy deposition but still include the fluctuations in the number of registered events, a good agreement between the Dirac multi–event and the full microdosimetric multi–event distribution emerges. On the other hand, more complex radiation fields are not very well described by the Dirac multi–event distribution. In this cases, the microdosimetric distributions are not peaked around the mean values since the contributions from different ions at different energies are extremely relevant. The quality of such a mixed radiation field required a more complex description, as the one included into GSM^2 .

The multi–event distribution directly affects the DNA damage distribution. Since the Dirac multi–event distribution provides an accurate description of the multi–event microdosimetric distribution for monoenergetic beams, it follows that the initial damage distributions described by GSM^2 and the Dirac GSM^2 are in a good agreement. Furthermore, the initial damage distributions for protons are also well described by a Poisson law, with the damage distribution in the case of lower LET predicted both by GSM^2 and the Dirac GSM^2 being slightly different from a Poisson distribution. We reported a similar analysis in 12. The Dirac GSM^2 and the GSM^2 also give a similar initial damage distribution for carbon ions, and they are both extremely different from a Poisson distribution. GSM^2 predicts a clear bi-modal distribution, with a significant probability of inducing no damages, as typical for high–LET radiation. At last for the mixed field, differences are clear in the DNA damage distribution.

Above discrepancy and analogies in the damage distributions emerges in the cell–survival curve prediction. Lower–LET protons shows an overall agreement in the survival fraction prediction by all four models. In this case, GSM^2 recovers a Poissonian description of the DNA damage formation and evolution. For higher LET protons, GSM^2 , the

Dirac GSM^2 and the Poisson GSM^2 predict the same cell survival curve, whereas the RMR gives a significantly lower values. Since the three GSM^2 versions share the same dynamical description, the similarity observed in the initial damage distribution is reflected in the cell survival computation. Since the RMR and the Poisson GSM^2 predict different survival curves, it can be deduced that the recombination of lesions leading to the cell death are already relevant for higher LET protons. Even if the RMR and the Poisson GSM^2 have the same high and low–dose asymptotes, the predicted cell–survival curves can differ. The results for carbon pencil beam show instead how a Poissonian initial damage distribution leads to different values compared to the GSM^2 and the Dirac GSM^2 . Both the RME and the Poisson GSM^2 give the same cell–survival curve which is clearly different from the one predicted by the Dirac GSM^2 and GSM^2 . The cell–survival curve given by the Dirac GSM^2 and the GSM^2 has a clear linear trend, which can be also observed in radiobiological experiments with high LET ions. On the contrary, the RMR and the Poisson GSM^2 predict a cell survival curve where a quadratic non-linear term lower the cell survival curve at higher doses. Furthermore, slight differences between the GSM^2 and the Dirac GSM^2 emerge in the cell survival curve prediction at high doses. In accordance with considerations done so far, the four models predict different values for the mixed radiation field, suggesting that in this case various effects may lead to the cell death. The RMR predicts the lower cell survival, followed by the Poisson GSM^2 , then the Dirac GSM^2 and at last the GSM^2 .

Chapter 14

Predicting the survival curves with GSM^2 and the real track length microdosimetric spectra from HDM

This chapter brings together all the pieces developed in this thesis. We used the microdosimetric spectra obtained from HDM, i.e. with the particle real track legth, as input for GSM^2 to describe the radiation field quality. We calculated the initial damage distributions and compared them with the predictions obtained with standard microdosimetric spectra, i.e. using the mean chord length approximation.

Material and Methods

The single event microdosimetric spectrum $f_1(z)$ has been assessed for three different Monte Carlo simulated $f_1(z)$ microdosimetric spectra as input: i) $f_1(z)$ is obtained using the mean chord length, ii) $f_1(z)$ is calculated using the real track length and iii) $f_1(z)$ is obtained considering a realistic detection efficiency and exploiting Machine Learning techniques. We applied the three approaches to two irradiation conditions, namely 150 MeV proton and 290 MeV/u carbon ions after traversing 9 cm of water, as done in Sections 9 and in 10.
Then, based on the single event distributions, we calculated the multi-event $f_n(z)$ spectrum using Equation (2.2). This spectra have been used as input to the GSM^2 to assess the sub–lethal damage distribution and then the cell survival curves using Equations (12.22)–(12.48). The parameters used for the cell survival computation has been set to $a = 0.02$, $b = 0.14$ and $r = 2.9$, and have been fitted in 12.2 and derived for the *Human Salivary Glands* (HSG) cell-line.

Results and Discussion

In Figure 14.1, the comparison of the single event microdosimetric distributions $f_1(z)$ for the three situations (mean chord length (MCL), real track length (RTL) and reconstructed track length via ML (ML)) are reported for protons and carbon ions.
The proton distributions peak at lower specific energies z (around 0.1 Gy), whereas carbon

ions exhibit a main peak around 5 Gy, with an additional second peak at lower z (between 0.1 Gy and 1 Gy), due to the secondary particles produced in water. There is a good agreement between the RTL and the ML predicted spectra, while the MCL distributions show a shift towards higher z values of the main peak both for protons and carbon ions.

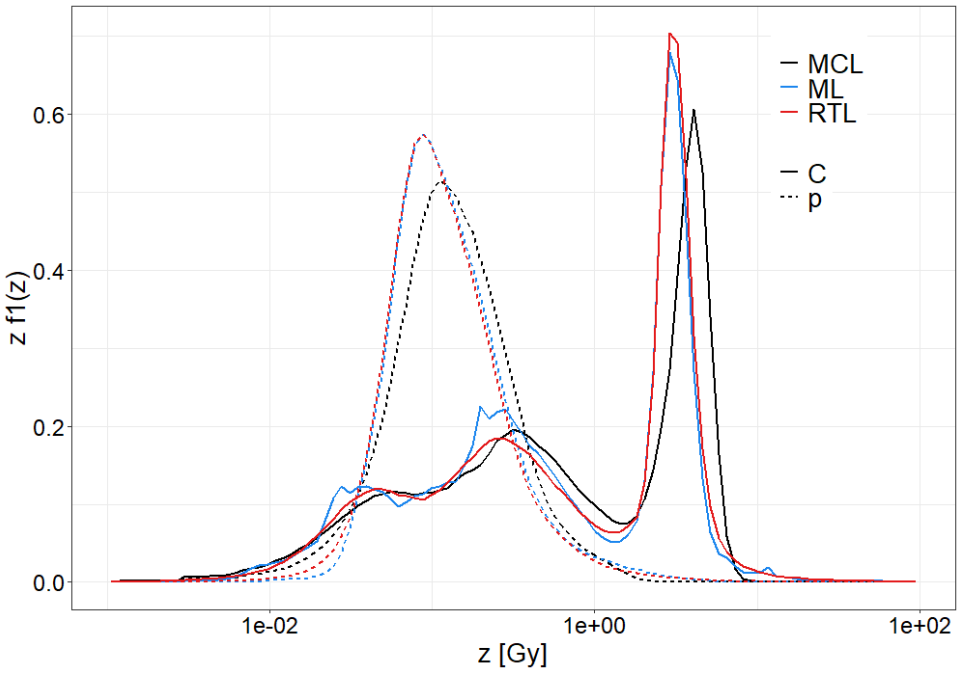

Figure 14.1: Single–event distributions $f_1(z)$ computed using the microdosimetric spectra with the mean chord length (black line), the real track length (red line) and the real track length predicted with the ML module (blue line) as illustrated in Section 10. The microdosimetric spectra were simulated using 150 MeV protons (dotted line) and 290 MeV/u carbon ions (full line)after traversing 9 cm of water.

Figure 14.2 reports the multi-event distributions $f_n(z)$ calculated for the three different approaches (MCL, RTL and ML) when applied to both protons and carbon ions, at two different macroscopic absorbed dose values of 1 Gy and 10 Gy. The general behaviour of the multi-event distributions is the same as for the single-event ones, showing similarities between RTL and ML spectra, while the MCL spectra are peaked at higher z values. It is interesting to notice the presence of the high peak at z=0 Gy of carbon ions at 1 and 10 Gy: since the radiation field has a high LET, just few events are needed in the detector to reach the low dose of 1 Gy, and even at 10 Gy there is still a non negligible probability of registering no energy depositions. This behavior is in agreement with the results shown in Section 12.2.

Figure 14.3 reports the survival probability as computed by GSM^2 for protons and carbon ions. The survival curve for protons exhibits a linear-quadratic trend, whereas for the carbon ions it shows a straight trend in agreement with the available radiobiological experimental data [Furusawa et al., 2000]. The agreement between the RTL and ML microdosimetric spectra is reflected in the cell survival predictions. For both ions, the discrepancy between the cell survival curves emerges only at high doses (above 5 Gy). Instead, the cell survival prediction using the MCL spectra differs at lower doses (around 2 Gy), with a larger discrepancy for protons.

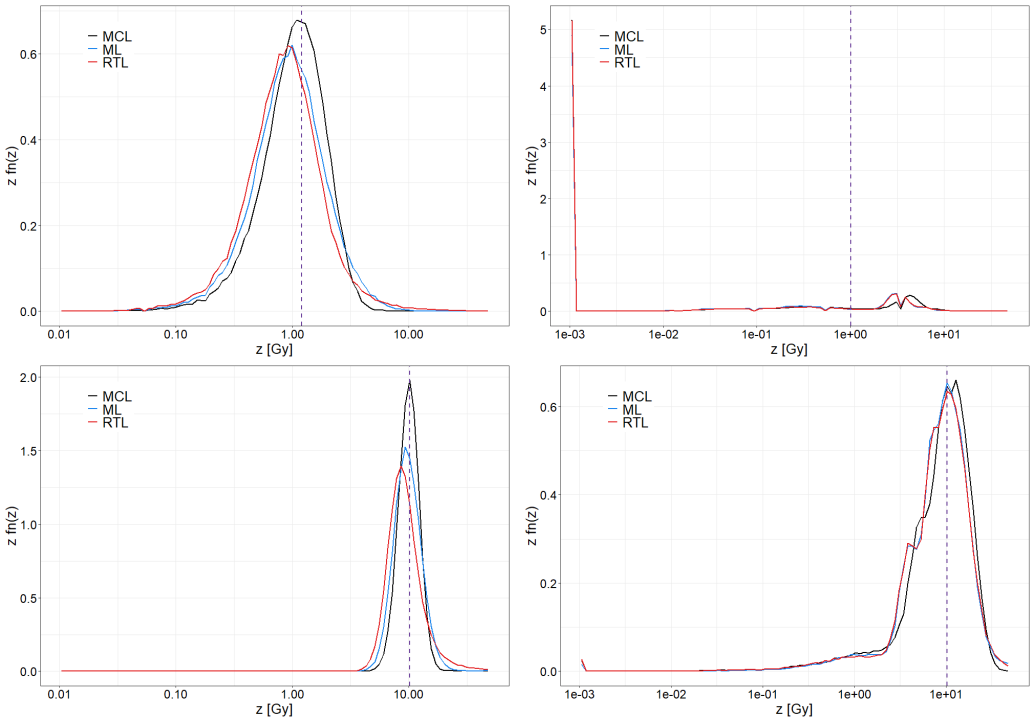

Figure 14.2: Sub–lethal damage distributions (Equation (11.9)) estimated using the microdosimetric spectra computed with the mean chord length (black line), the real track length (red line) and the real track length predicted with the ML module (blue line) as discussed in Section 10. The vertical purple dotted line indicates the distributions average that coincide with the macroscopic imparted dose. The left panels refer to protons at 150 MeV after 9 cm of water with macroscopic imparted dose equal to 1 Gy (top panel) and 10 Gy (bottom panel). The right panels refer to carbon ions at 290 MeV/u after 9 cm of water with macroscopic imparted dose equal to 1 Gy (top panel) and 10 Gy (bottom panel).

However, focusing on the differences between MCL and RTL cell survival curves, it is interesting to notice that there is a lower discrepancy in carbon ions cell survival curve with respect to the proton case. On the contrary, from Figure 14.1 the difference in carbon ions between the MCL $f_1(z)$ and RTL $f_1(z)$ is even higher than this difference in the protons case. This is due to the fact that, as deeply shown both from a theoretical and numerical point of View in Chapter 12, GSM^2 predicts the cell survival curve taking into account the whole microdosimetric spectra. In the particular case of protons and carbon ions the protons primary peaks are in the low z region; instead, carbon ions main peaks are in a high z region, where the overkilling effect reduces their overall radiobiological damage potential. This means that differences in the main peak of the microdosimetric spectra are not seen in the biological response of the tissue. In the low z region of the spectra, that is not affected by the overkilling effect, the MCL and RTL of carbon ions are similar, thus explaining the similarity in the final cell survival curve predicted.

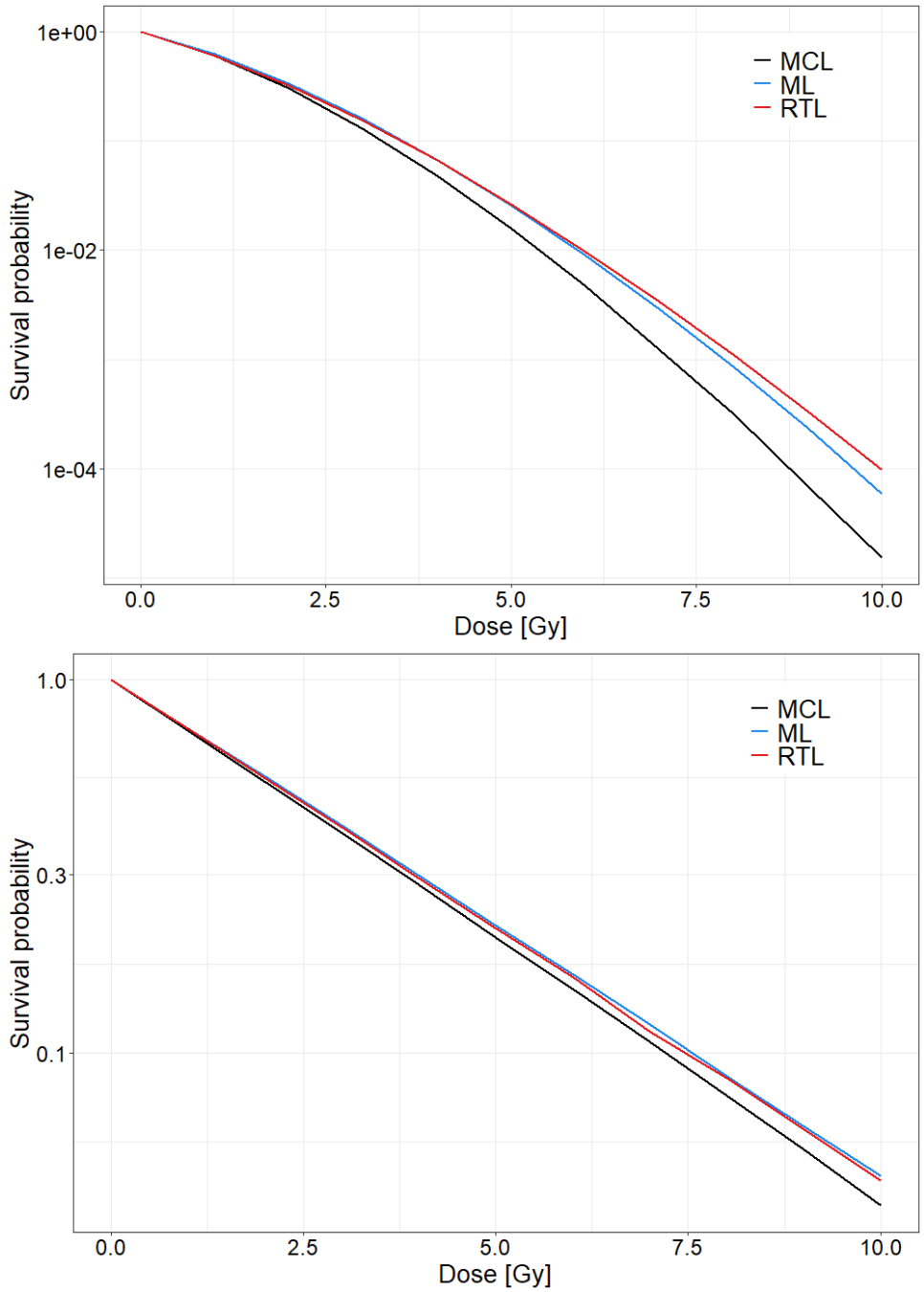

Figure 14.3: Survival curves predicted by GSM^2 using the initial damage distribution estimated with microdosimetry. The microdosimetric spectra were computed with the mean chord length (black line), the real track length (red line) and the real track length predicted with the ML module (blue line) as developed in Section 10. Top panels show the results for 150 MeV protons, while bottom panels for 290 MeV/u carbon ions.

Conclusions and future perspectives

The driving force of the present work has been microdosimetry, with the ultimate purpose of fully exploiting the vast potential of the microdosimetric quantities in describing the radiation field and in providing one of the finest and most meaningful ingredient to model the radiobiological damage.

A detailed knowledge of the field composition and quality is required to estimate the biological effect of radiation. The most widely used approach to assess the radiation quality is the Linear Energy Transfer (LET). By definition, LET is measured in macroscopic volumes, i.e. on a scale much larger than the cellular level where the actual biological damage occurs. This approach, however, neglects the stochastic nature of energy deposition. The limited capability in describing local radiation-induced effects at the cellular level, especially for non-uniform dose deposition patterns delivered by mixed fields, has posed the question whether LET is the most accurate radiation quality predictor for RBE assessment [Grün et al., 2019]. The limitations discussed in 1.1.3 led the radiobiological and medical communities to consider substituting the LET concept with with a more meaningful and experimentally measurable physical quantity.

Microdosimetry allows the investigation of the effects of radiation on cells in a region comparable to the structures of interest, being the energy deposition of a particle scored in a micrometer-sized volume, e.g. the cell nucleus. For this reason, it has been proposed as an alternative methodology to overcome LET limitations [Cortés-Giraldo and Carabe, 2015, Bertolet et al., 2019]. One of the advantages of microdosimetry compared to the dosimetric LET approach is that it is more suited to detect short-range heavy fragments, which can generate hot spots of energy deposition and have been recognized as a potential cause of toxicity in the normal tissue [Haas-Kogan et al., 2018].

Exploiting microdosimetry, we characterized both homogeneous and complex radiation fields, either along the in-beam direction and in different off-beam positions. This study has been performed for monoenergetic and Spread Out Bragg peak protons, for helium and oxygen ion beams and for neutrons produced from a iron beam impinging on a aluminum target, and the outcomes are reported in Part III. Aside from providing a better understanding of the particle effectiveness in tumor killing, data from protons, helium and oxygen characterizations can help predict the risk of undesired effects following the treatment. In addition, the microdosimetric study of neutrons from iron stopping in an aluminum target is of great interest for space radioprotection. Microdosimetry has proven to be an invaluable tool in characterizing radiation fields for both applications, particle therapy and space radioprotection.

The results demonstrate in addition that the microdosimetric characterization is as key ingredient for a final biological damage assessment, that can be derived exploiting different radiobiological models. This investigation uncovered a strong dependence of the biological damage and RBE assessment on the model and indicated that a more robust mathematical model to predict RBE is advisable.

The new tool should be based on microdosimetry, providing a superior description of the energy deposited in the cell nuclei, from which the initial damage distribution is assessed. The accuracy in characterizing the radiation field quality directly reflects on the precision in predicting the cell damage and RBE values. Therefore, expanding the microdosimetric information could benefit the reliability of the radiobiological damage estimation.

A general approximation underlying experimental microdosimetry is the mean chord length concept. The main physical quantity considered in experimental microdosimetry is the *lineal energy* y, [Zaider et al., 1996], which is obtained dividing the energy deposition

ϵ by the mean chord length l traversed by the particle in the detector. While ϵ is directly measured, the value of l is calculated as the average path travelled by a particle inside the detector, and thus depends both on the detector geometry and on specific assumptions on the radiation field (typically considered isotropic and uniform). Nonetheless, as early noted by [Vassiliev, 2017], to achieve a more meaningful connection between the radiation field and the consequent biological outcome, the knowledge of the real track length traversed by the particles that ionize the biological tissue is desirable.

Taking into consideration the strengths and limitations of the the existing detectors, we designed a new hybrid 2-stage microdosimeter (HDM: *Hybrid Detector for Microdosimetry*), 9, composed of a spherical TEPC, [Missiaggia et al., 2020], followed by four *Low Gain Avalanche Detectors* (LGADs), [Pellegrini et al., 2014]. The proposed design benefits from the fact that the TEPC provides the energy deposition ϵ directly in a tissue-equivalent medium and has a good detection limit for low-LET particles (around 0.1 keV/μm). Furthermore, the 4 LGAD planes offer the particle tracking. Using the track information, we can obtain the actual path length travelled by each particle inside the TEPC (and thus avoid the mean chord length approximation). In addition, HDM improves the TEPC spatial resolution.

To assess the detector performances, we performed a feasibility study described in Chapter 9 via Monte Carlo simulations, where different configurations have been tested (changing the distance between detectors, the number of strips in a single LGAD) and the HDM response has been investigated when irradiated with protons and carbon ions at different water depths. The results showed all HDM advantages, but also indicated that the detection efficiency is the detector main weakness. To tackle such issue as well as to improve the tracking accuracy, we exploited modern *Machine Learning* (ML) techniques in Chapter 10.

The next natural step was to quantify the effect that the augmented microdosimetric spectra provided by HDM could have on predicting biological outcomes and estimating the RBE. The vast majority of existing models fail to consider the whole microdosimetric spectrum and are rather based on averages values, such as the MKM and its variants. Biological weighting functions such as the Loncol one [Loncol et al., 1994], consider the whole microdosimetric distribution, but have a limited validity since they have been calculated only for specific ion types and biological endpoints. Therefore, up to date, no radiobiological model is general enough to both i) fully exploit the potential of microdosimetry and of the extended information provided by HDM, and ii) to be valid in many different physical (ion type, energy and dose range) and biological (cell type) scenarios.

For this reason, we have developed the *Generalized Stochastic Microdosimetric Model* (GSM^2) described in Part V. GSM^2 predicts cell survival, and consequently RBE, considering the whole microdosimetric spectrum. The original motivation behind GSM^2 development was to introduce a model that provide a rigorous and general mathematical description of DNA damage time–evolution without using any a priori assumption on the lesion distribution (e.g. a Poisson). This model would include several non-linear and stochastic effects that are typically implemented into existing models by suitable semi-phenomenological corrections. Since GSM^2 aims at providing a true stochastic description of DNA damage formation and kinetic evolution, microdosimetry represents the most suitable physical input to describe the radiation field quality, including HDM augmented microdosimetric information.

Finally, Chapter 14 brings together many of the new concepts presented in the thesis: microdosimetric spectra obtained with HDM and the ML approach provide the radiation

quality description that is used as an input for GSM^2. The outcomes obtained from this novel approach represent an encouraging first step to achieve a superior characterization of the radiation field and the biological damage estimation. Nonetheless, the results showed in this thesis are not the end of the story. HDM can greatly benefits from further technological improvements, such as new LGADs technologies to thin the layers in order to lower the particle scattering within the tracker stage and an optimization of the fill factor to reduce the dead region. In addition, the use of a miniTEPC instead of the TEPC model LET 1/2 will be consider to apply HDM at higher particle rates, in view of a therapeutic range application. Furthermore, the next step in the direction of an augmented microdosimetry includes the particle charge prediction via an additional ML module.

Experimental radiobiological campaigns are already planned to validate the GSM^2. Because the model is based on a full description of the damage initial distribution and evolution, the accuracy and quality of the necessary radiobiological data is not a trivial request, and new studies not available in literature are needed to fully validate the model. The potential of using the real track length additional information in microdosimetry provided by HDM will be better exploited in an improved version of GSM^2, where the concept of the nucleus domains will be substituted with a description of the true spatial resolution of DNA damage.

In conclusion, in this dissertation microdosimetry was explored from different angles, including both the experimental and theoretical sides, and we demonstrated that it can yield impactful advances with wide-reaching implications.

It is my hope that, also thanks to the small steps forward from this work, the potential of microdosimetry can be fully realized.

Bibliography

[Abolfath et al., 2020] Abolfath, R., Helo, Y., Carlson, D. J., Stewart, R., Grosshans, D., and Mohan, R. (2020). A new approach to modeling the microdosimetry of proton therapy beams. *Medical physics*.

[Agosteo et al., 2006] Agosteo, S., Colautti, P., Fazzi, A., Moro, D., and Pola, A. (2006). A solid state microdosimeter based on a monolithic silicon telescope. *Radiation protection dosimetry*, 122(1-4):382–386.

[Agosteo and Pola, 2011] Agosteo, S. and Pola, A. (2011). Silicon microdosimetry. *Radiation protection dosimetry*, 143(2-4):409–415.

[Agostinelli et al., 2003] Agostinelli, S., Allison, J., Amako, K. a., Apostolakis, J., Araujo, H., Arce, P., Asai, M., Axen, D., Banerjee, S., Barrand, G. ., et al. (2003). Geant4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303.

[Ahlen, 1980] Ahlen, S. P. (1980). Theoretical and experimental aspects of the energy loss of relativistic heavily ionizing particles. *Reviews of Modern Physics*, 52(1):121.

[Albright, 1989] Albright, N. (1989). A markov formulation of the repair-misrepair model of cell survival. *Radiation research*, 118(1):1–20.

[Amrouche et al., 2020] Amrouche, S., Basara, L., Calafiura, P., Estrade, V., Farrell, S., Ferreira, D. R., Finnie, L., Finnie, N., Germain, C., Gligorov, V. V., et al. (2020). The tracking machine learning challenge: Accuracy phase. In *The NeurIPS'18 Competition*, pages 231–264. Springer.

[Arce et al., 2021] Arce, P., Bolst, D., Bordage, M.-C., Brown, J., Cirrone, P., Cortés-Giraldo, M. A., Cutajar, D., Cuttone, G., Desorgher, L., Dondero, P., et al. (2021). Report on g4-med, a geant4 benchmarking system for medical physics applications developed by the geant4 medical simulation benchmarking group. *Medical physics*, 48(1):19–56.

[Asaithamby et al., 2011] Asaithamby, A., Hu, B., and Chen, D. J. (2011). Unrepaired clustered DNA lesions induce chromosome breakage in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, 108(20):8293–8298.

[Asaithamby et al., 2008] Asaithamby, A., Uematsu, N., Chatterjee, A., Story, M. D., Burma, S., and Chen, D. J. (2008). Repair of HZE-Particle-Induced DNA Double-Strand Breaks in Normal Human Fibroblasts. *Radiation Research*, 169(4):437–446.

[Barkas, 1973] Barkas, W. H. (1973). Nuclear research emulsions. ii. particle behavior and emulsion applications.

[Bellinzona et al., 2021] Bellinzona, V. E., Cordoni, F., Missiaggia, M., Tommasino, F., Scifoni, E., La Tessa, C., and Attili, A. (2021). Linking microdosimetric measurements to biological effectiveness in ion beam therapy: a review of theoretical aspects of mkm and other models. *Frontiers in Physics*, 8:623.

[Bergstra et al., 2011] Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 24.

[Bertolet et al., 2019] Bertolet, A., Baratto-Roldán, A., Barbieri, S., Baiocco, G., Carabe, A., and Cortés-Giraldo, M. (2019). Dose-averaged LET calculation for proton track segments using microdosimetric Monte Carlo simulations. *Med. Phys.*, 46(9).

[Bertolet et al., 2021] Bertolet, A., Cortés-Giraldo, M., and Carabe-Fernandez, A. (2021). Implementation of the microdosimetric kinetic model using analytical microdosimetry in a treatment planning system for proton therapy. *Physica Medica*, 81:69–76.

[Bethe, 1953] Bethe, H. A. (1953). Molière's Theory of Multiple Scattering. *Phys. Rev.*, 89.

[Bianchi et al., 2020] Bianchi, A., Selva, A., Colautti, P., Bortot, D., Mazzucconi, D., Pola, A., Agosteo, S., Petringa, G., Cirrone, G. P., Reniers, B., et al. (2020). Microdosimetry with a sealed mini-tepc and a silicon telescope at a clinical proton sobp of catana. *Radiation Physics and Chemistry*, 171:108730.

[Bishop and Nasrabadi, 2006] Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.

[Bodgi et al., 2016] Bodgi, L., Canet, A., Pujo-Menjouet, L., Lesne, A., Victor, J.-M., and Foray, N. (2016). Mathematical models of radiation action on living cells: From the target theory to the modern approaches. a historical and critical review. *Journal of theoretical biology*, 394:93–101.

[Bolst et al., 2020] Bolst, D., Guatelli, S., Tran, L. T., Chartier, L., Davis, J., Biasi, G., Prokopovich, D. A., Pogossov, A., Reinhard, M. I., Petasecca, M., et al. (2020). Validation of geant4 for silicon microdosimetry in heavy ion therapy. *Physics in Medicine & Biology*, 65(4):045014.

[Bolst et al., 2017] Bolst, D., Guatelli, S., Tran, L. T., Chartier, L., Lerch, M. L., Matsufuji, N., and Rosenfeld, A. B. (2017). Correction factors to convert microdosimetry measurements in silicon to tissue in 12C ion therapy. *Physics in Medicine & Biology*, 62(6):2055.

[Booz et al., 1983] Booz, J., Braby, L., Coyne, J., Kliauga, P., Lindborg, L., Menzel, H., and Parmentier, N. (1983). Report 36. *Journal of the International Commission on Radiation Units and Measurements*, (1):NP–NP.

[Braby, 2015] Braby, L. (2015). Experimental microdosimetry: history, applications and recent technical advances. *Radiation protection dosimetry*, 166(1-4):3–9.

[Bragg and Kleeman, 1905] Bragg, W. H. and Kleeman, R. (1905). Xxxix. on the α particles of radium, and their loss of range in passing through various atoms and molecules. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(57):318–340.

[Britton et al., 2020] Britton, T., Lawrence, D., and Gavalian, G. (2020). Ml track fitting in nuclear physics. In *EPJ Web of Conferences*, volume 245, page 06015. EDP Sciences.

[Brochu et al., 2010] Brochu, E., Cora, V. M., and De Freitas, N. (2010). A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*.

[Burigo et al., 2015] Burigo, L., Pshenichnov, I., Mishustin, I., and Bleicher, M. (2015). Comparative study of dose distributions and cell survival fractions for ^{1}H, ^{4}He, ^{12}C and ^{16}O beams using Geant4 and Microdosimetric Kinetic model. *Phys. Med. Bio.*, 60:3313–3331.

[Carabe et al., 2012] Carabe, A., Moteabbed, M., Depauw, N., Schuemann, J., and Paganetti, H. (2012). Range uncertainty in proton therapy due to variable biological effectiveness. *Physics in Medicine and Biology*, 57(5):1159–72.

[Carabe-Fernandez et al., 2011] Carabe-Fernandez, A., Dale, R. G., and Paganetti, H. (2011). Repair kinetic considerations in particle beam radiotherapy. *The British journal of radiology*, 84(1002):546–55.

[Carlson et al., 2008] Carlson, D. J., Stewart, R. D., Semenenko, V. A., and Sandison, G. A. (2008). Combined use of monte carlo dna damage simulations and deterministic repair models to examine putative mechanisms of cell killing. *Radiation research*, 169(4):447–459.

[Cekanaviciute et al., 2018] Cekanaviciute, E., Rosi, S., and Costes, S. V. (2018). Central nervous system responses to simulated galactic cosmic rays. *Int. J. Mol. Sci.*, 19:3669–83.

[Chartier et al., 2018] Chartier, L., Tran, L. T., Bolst, D., Guatelli, S., Pogossov, A., Prokopovich, D. A., Reinhard, M. I., Perevertaylo, V., Anderson, S., Beltran, C., et al. (2018). Microdosimetric applications in proton and heavy ion therapy using silicon microdosimeters. *Radiation Protection Dosimetry*, 180(1-4):365–371.

[Chen et al., 2017] Chen, Y., Li, J., Li, C., Qiu, R., and Wu, Z. (2017). A modified microdosimetric kinetic model for relative biological effectiveness calculation. *Physics in Medicine & Biology*, 63(1):015008.

[Chiriotti et al., 2015a] Chiriotti, S., Moro, D., Colautti, P., Conte, V., and Grosswendt, B. (2015a). Equivalence of pure propane and propane TE gases for microdosimetric measurements. *Radiation protection dosimetry*, 166(1-4):242–246.

[Chiriotti et al., 2015b] Chiriotti, S., Moro, D., Conte, V., Colautti, P., Grosswendt, B., Sterpin, E., and Vynckier, S. (2015b). Influence of the physical data to calibrate tepcs. *Radiation protection dosimetry*, 166(1-4):238–241.

[Colautti et al., 2020] Colautti, P., Magrin, G., Palmans, H., Cortés-Giraldo, M. A., and Conte, V. (2020). Characterizing radiation effectiveness in ion-beam therapy part ii: Microdosimetric detectors. *Frontiers in Physics*, 8:451.

[Comfort et al., 1966] Comfort, J. R., Decker, J. F., Lynk, E. T., Scully, M. O., and Quinton, A. R. (1966). Energy Loss and Straggling of Alpha Particles in Metal Foils. *Phys. Rev.*, 150.

[Conte et al., 2020] Conte, V., Agosteo, S., Bianchi, A., Bolst, D., Bortot, D., Catalano, R., Cirrone, G., Colautti, P., Cuttone, G., Guatelli, S., et al. (2020). Microdosimetry of a therapeutic proton beam with a mini-tepc and a microplus-bridge detector for rbe assessment. *Physics in Medicine & Biology*, 65(24):245018.

[Conte et al., 2019] Conte, V., Bianchi, A., Selva, A., Petringa, G., Cirrone, G., Parisi, A., Vanhavere, F., and Colautti, P. (2019). Microdosimetry at the CATANA 62 MeV proton beam with a sealed miniaturized TEPC. *Phys. Med.*, 64:114–122.

[Conte et al., 2017] Conte, V., Selva, A., Colautti, P., Hilgers, G., and Rabus, H. (2017). Track structure characterization and its link to radiobiology. *Radiat. Meas.*, 106:506–511.

[Cordoni et al., 2021] Cordoni, F., Missiaggia, M., Attili, A., Welford, S., Scifoni, E., and La Tessa, C. (2021). Generalized stochastic microdosimetric model: The main formulation. *Physical Review E*, 103(1):012412.

[Cordoni et al., 2022a] Cordoni, F. G., Missiaggia, M., La Tessa, C., and Scifoni, E. (2022a). Multiple levels of stochasticity accounted in different radiation biophysical models: from physics to biology. *Submitted to International Journal of Radiation Biology.*

[Cordoni et al., 2022b] Cordoni, F. G., Missiaggia, M., Scifoni, E., and La Tessa, C. (2022b). Cell survival computation via the generalized stochastic microdosimetric model (gsm2); part i: The theoretical framework. *Radiation research*, 197(3):218–232.

[Cortés-Giraldo and Carabe, 2015] Cortés-Giraldo, M. A. and Carabe, A. (2015). A critical study of different monte carlo scoring methods of dose average linear-energy-transfer maps calculated in voxelized geometries irradiated with clinical proton beams. *Phys. Med. Bio.*, 60:2645–2669.

[Coutrakon et al., 1997] Coutrakon, G., Cortese, J., Ghebremedhin, A., Hubbard, J., Johanning, J., Koss, P., Maudsley, G., Slater, C., Zuccarelli, C., and Robertson, J. (1997). Microdosimetry spectra of the loma linda proton beam and relative biological effectiveness comparisons. *Medical physics*, 24(9):1499–1506.

[Craik, 2005] Craik, A. D. (2005). Prehistory of faà di bruno's formula. *The American Mathematical Monthly*, 112(2):119–130.

[Cruz et al., 2001] Cruz, G. S., Palmer, M., Matatagui, E., and Zamenhof, R. (2001). A theoretical model for event statistics in microdosimetry. i: Uniform distribution of heavy ion tracks. *Medical physics*, 28(6):988–996.

[Curtis, 1986] Curtis, S. B. (1986). Lethal and potentially lethal lesions induced by radiation—a unified repair model. *Radiation research*, 106(2):252–270.

[Curtis, 1988] Curtis, S. B. (1988). The lethal and potentially lethal model—a review and recent development. In *Quantitative mathematical models in radiation biology*, pages 137–146. Springer.

[Dale et al., 1999] Dale, R. G., Fowler, J. F., and Jones, B. (1999). A new incomplete-repair model based on a 'reciprocal-time' pattern of sublethal damage repair. *Acta Oncologica*, 38(7):919–929.

[De Nardo et al., 2004] De Nardo, L., Cesari, V., Donà, G., Magrin, G., Colautti, P., Conte, V., and Tornielli, G. (2004). Mini-tepcs for radiation therapy. *Radiation protection dosimetry*, 108(4):345–352.

[Delaney et al., 2005] Delaney, G., Jacob, S., Featherstone, C., and Barton, M. (2005). The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 104(6):1129–1137.

[Dewey et al., 2017] Dewey, S., Burigo, L., Pshenichnov, I., Mishustin, I., and Bleicher, M. (2017). Lateral variations of radiobiological properties of therapeutic fields of ^{1}H, ^{4}He, ^{12}C and ^{16}O ions studiede with Geant4 and microdosimetric kinetic model. *Phys. Med. Bio.*, 62:5884–5907.

[Dicello et al., 1980] Dicello, J., Amols, H., Zaider, M., and Tripard, G. (1980). A comparison of microdosimetric measurements with spherical proportional counters and solid-state detectors. *Radiation Research*, 82(3):441–453.

[Dieter Schardt, 2010] Dieter Schardt, D. S.-E. (2010). In *Heavy-ion tumor therapy: physical and radiobiological benefits*. Rev.Mod.Physics 82.

[Dikomey and Franzke, 1986] Dikomey, E. and Franzke, J. (1986). DNA repair kinetics after exposure to X-irradiation and to internal β-rays in CHO cells. *Radiation and Environmental Biophysics*, 25(3):189–194.

[Duarte and Vlimant, 2020] Duarte, J. and Vlimant, J.-R. (2020). Graph neural networks for particle tracking and reconstruction. *arXiv preprint arXiv:2012.01249*.

[Durante and Loeffler, 2010] Durante, M. and Loeffler, J. S. (2010). Charged particles in radiation oncology. *Nature reviews Clinical oncology*, 7(1):37.

[Durante and Paganetti, 2016] Durante, M. and Paganetti, H. (2016). Nuclear physics in particle therapy: a review. *Rep. Prog. Phys.*, 79:096702.

[Durante et al., 2019] Durante, M., Paganetti, H., Pompos, A., Kry, S., Wu, X., and Grosshans, D. (2019). Report of a National Cancer Institute special panel: Characterization of the physical parameters of particle beams for biological research. *Med. Phys.*, In press.

[Elsässer et al., 2008] Elsässer, T., Krämer, M., and Scholz, M. (2008). Accuracy of the local effect model for the prediction of biologic effects of carbon ion beams in vitro and in vivo. *International Journal of Radiation Oncology* Biology* Physics*, 71(3):866–872.

[Elsässer and Scholz, 2007] Elsässer, T. and Scholz, M. (2007). Cluster effects within the local effect model. *Radiation research*, 167(3):319–329.

[Elsässer et al., 2010] Elsässer, T., Weyrather, W. K., Friedrich, T., Durante, M., Iancu, G., Krämer, M., Kragl, G., Brons, S., Winter, M., Weber, K.-J., et al. (2010). Quantification of the relative biological effectiveness for ion beam radiotherapy: direct experimental comparison of proton and carbon ion beams and a novel approach for treatment planning. *International Journal of Radiation Oncology* Biology* Physics*, 78(4):1177–1183.

[Endo, 2018] Endo, M. (2018). Robert r. wilson (1914–2000): the first scientist to propose particle therapy—use of particle beam for cancer treatment.

[Fano, 1963] Fano, U. (1963). Penetration of protons, alpha particles, and mesons. *Annual Review of Nuclear Science*, 13(1):1–66.

[Farahmand et al., 2004] Farahmand, M., Bos, A., De Nardo, L., and Van Eijk, C. (2004). First microdosimetric measurements with a tepc based on a gem. *Radiation protection dosimetry*, 110(1-4):839–843.

[Fausti et al., 2018] Fausti, F., Mazza, G., Cartiglia, N., Ferrero, M., Giordanengo, S., Ali, O., Mandurrino, M., Monaco, V., Olave, J., Sacchi, R., et al. (2018). Characterization of a silicon detector and front-end electronics prototype for single ion discrimination in hadrontherapy. In *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, pages 1–3. IEEE.

[Fowler, 1989] Fowler, J. F. (1989). The linear-quadratic formula and progress in fractionated radiotherapy. *The British journal of radiology*, 62(740):679–694.

[Fowler, 1999] Fowler, J. F. (1999). Is repair of DNA strand break damage from ionizing radiation second-order rather than first-order? A simpler explanation of apparently multiexponential repair. *Radiation research*, 152(2):124–36.

[Frazier, 2018] Frazier, P. I. (2018). A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*.

[Frese et al., 2012] Frese, M. C., Victor, K. Y., Stewart, R. D., and Carlson, D. J. (2012). A mechanism-based approach to predict the relative biological effectiveness of protons and carbon ions in radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 83(1):442–450.

[Friedman et al., 2001] Friedman, J., Hastie, T., Tibshirani, R., et al. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

[Friedrich et al., 2013a] Friedrich, T., Durante, M., and Scholz, M. (2013a). The local effect model-principles and applications. *The Health Risks of Extraterrestrial Environments*.

[Friedrich et al., 2012] Friedrich, T., Scholz, U., Elsässer, T., Durante, M., and Scholz, M. (2012). Calculation of the biological effects of ion beams based on the microscopic spatial damage distribution pattern. *International journal of radiation biology*, 88(1-2):103–107.

[Friedrich et al., 2013b] Friedrich, T., Scholz, U., ElsäSser, T., Durante, M., and Scholz, M. (2013b). Systematic analysis of rbe and related quantities using a database of cell survival experiments with ion beam irradiation. *Journal of radiation research*, 54(3):494–514.

[Furusawa et al., 2000] Furusawa, Y., Fukutsu, K., Aoki, M., Itsukaichi, H., Eguchi-Kasai, K., Ohara, H., Yatagai, F., Kanai, T., and Ando, K. (2000). Inactivation of aerobic and hypoxic cells from three different cell lines by accelerated 3he-, 12c-and 20ne-ion beams. *Radiation research*, 154(5):485–496.

[Gardiner et al., 1985] Gardiner, C. W. et al. (1985). *Handbook of stochastic methods*, volume 3. springer Berlin.

[Goldhaber and Heckman, 1978] Goldhaber, A. S. and Heckman, H. H. (1978). High energy interactions of nuclei. *Annual Review of Nuclear and Particle Science*, 28(1):161–205.

[Goodhead, 2013] Goodhead, D. (2013). https://melodi-online.eu/wp-content/uploads/2021/04/02-What-is-radiation-quality-Goodhead.pdf.

[Granville and Sawakuchi, 2015] Granville, D. A. and Sawakuchi, G. O. (2015). Comparison of linear energy transfer scoring techniques in monte carlo simulations of proton beams. *Phys. Med. Bio.*, 60:N283–N291.

[Grün et al., 2019] Grün, R., Friedrich, T., Traneus, E., and Scholz, M. (2019). Is the dose-averaged let a reliable predictor for the relative biological effectiveness? *Med. Phys.*, 46:1064–1074.

[Guardiola et al., 2015] Guardiola, C., Fleta, C., Rodríguez, J., Lozano, M., and Gómez, F. (2015). Preliminary microdosimetric measurements with ultra-thin 3d silicon detectors of a 62 mev proton beam. *Journal of Instrumentation*, 10(01):P01008.

[Gueulette et al., 2007] Gueulette, J., Wambersie, A., Octave-Prignot, M., De Coster, B., and Grégoire, V. (2007). Radiobiological characterisation of clinical beams: Importance for the quality assurance (qa) programme in ion beam therapy. *IAEA-TECDOC-1560*, page 29.

[Gunzert-Marx et al., 2008] Gunzert-Marx, K., Iwase, H., Schardt, D., and Simon, R. (2008). Secondary beam fragments produced by 200 mev u- 1 12c ions in water and their dose contributions in carbon ion radiotherapy. *New journal of physics*, 10(7):075003.

[Haas-Kogan et al., 2018] Haas-Kogan, D., Indelicato, D., Paganetti, H., Esiashvili, N., Mahajan, A., Yock, T., Flampouri, S., MacDonald, S., Fouladi, M., Stephen, K., Kalapurakal, J., Terezakis, S., Kooy, H., Grosshans, D., Makrigiorgos, M., Mishra, K., Poussaint, T., Cohen, K., Fitzgerald, T., Gondi, V., Liu, A., Michalski, J., Mirkovic, D., Mohan, R., Perkins, S., Wong, K., Vikram, B., Buchsbaum, J., and Kun, L. (2018). National cancer institute workshop on proton therapy for children: Considerations regarding brainstem injury. *Int. J. Radiat. Oncol. Biol. Phys.*, 101:152–168.

[Hall and Giaccia, 2006] Hall, E. J. and Giaccia, A. J. (2006). *Radiobiology for the Radiologist*, volume 6. Lippincott Williams & Wilkins.

[Hall et al., 1978] Hall, E. J., Kellerer, A. M., Rossi, H. H., and Lam, Y.-M. P. (1978). The relative biological effectiveness of 160 mev protons—ii biological data and their interpretation in terms of microdosimetry. *International Journal of Radiation Oncology* Biology* Physics*, 4(11-12):1009–1013.

[Hawkins, 1996] Hawkins, R. (1996). A microdosimetric-kinetic model of cell death from exposure to ionizing radiation of any let, with experimental and clinical applications. *International journal of radiation biology*, 69(6):739–755.

[Hawkins, 1994] Hawkins, R. B. (1994). A statistical theory of cell killing by radiation of varying linear energy transfer. *Radiation research*, 140(3):366–374.

[Hawkins, 1998] Hawkins, R. B. (1998). A microdosimetric-kinetic theory of the dependence of the rbe for cell death on let. *Medical physics*, 25(7):1157–1170.

[Hawkins, 2003] Hawkins, R. B. (2003). A microdosimetric-kinetic model for the effect of non-poisson distribution of lethal lesions on the variation of rbe with let. *Radiation research*, 160(1):61–69.

[Hawkins and Inaniwa, 2014a] Hawkins, R. B. and Inaniwa, T. (2014a). A Microdosimetric-Kinetic Model for Cell Killing by Protracted Continuous Irradiation II: Brachytherapy and Biologic Effective Dose. *Radiation Research*, 182(1):72–82.

[Hawkins and Inaniwa, 2014b] Hawkins, R. B. and Inaniwa, T. (2014b). A microdosimetric-kinetic model for cell killing by protracted continuous irradiation ii: brachytherapy and biologic effective dose. *Radiation research*, 182(1):72–82.

[Hellander et al., 2012] Hellander, S., Hellander, A., and Petzold, L. (2012). Reaction-diffusion master equation in the microscopic limit. *Physical Review E*, 85(4):042901.

[Ho, 1995] Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE.

[Horst et al., 2022] Horst, F., Boscolo, D., Cartechini, G., Durante, M., Hartel, C., Kozlova, E., La Tessa, C., Missiaggia, M., Pierobon, E., Radon, T., et al. (2022). A multi-detector experimental setup for the study of space radiation shielding materials: Measurement of secondary radiation behind thick shielding and assessment of its radiobiological effect. In *EPJ Web of Conferences*, volume 261, page 03002. EDP Sciences.

[Hutter et al., 2011] Hutter, F., Hoos, H. H., and Leyton-Brown, K. (2011). Sequential model-based optimization for general algorithm configuration. In *International conference on learning and intelligent optimization*, pages 507–523. Springer.

[Hüfner et al., 1975] Hüfner, J., Schäfer, K., and Schürmann, B. (1975). Abrasion-ablation in reactions between relativistic heavy ions. *Phys. Rev. C*, 12:1888.

[ICRU, 1964] ICRU (1964). Physical Aspects of Irradiation Report 10b. *International Commission on Radiological Units and Measurements*.

[Inaniwa et al., 2010] Inaniwa, T., Furukawa, T., Kase, Y., Matsufuji, N., Toshito, T., Matsumoto, Y., Furusawa, Y., and Noda, K. (2010). Treatment planning for a scanned carbon beam with a modified microdosimetric kinetic model. *Physics in Medicine & Biology*, 55(22):6721.

[Inaniwa et al., 2013] Inaniwa, T., Suzuki, M., Furukawa, T., Kase, Y., Kanematsu, N., Shirai, T., and Hawkins, R. B. (2013). Effects of dose-delivery time structure on biological effectiveness for therapeutic carbon-ion beams evaluated with microdosimetric kinetic model. *Radiation research*, 180(1):44–59.

[International Commission on Radiation Units and Measurements, 1983] International Commission on Radiation Units and Measurements (1983). Microdosimetry. *ICRU*, 36.

[Isaacson, 2009] Isaacson, S. A. (2009). The reaction-diffusion master equation as an asymptotic approximation of diffusion to a small target. *SIAM Journal on Applied Mathematics*, 70(1):77–111.

[James et al., 2021] James, B., Tran, L. T., Bolst, D., Prokopovich, D. A., Lerch, M., Petasecca, M., Guatelli, S., Povoli, M., Kok, A., Petringa, G., et al. (2021). In-field and out-of-field microdosimetric characterisation of a 62 mev proton beam at catana. *Medical Physics*, 48(8):4532–4541.

[Jarlskog and Paganetti, 2008] Jarlskog, C. Z. and Paganetti, H. (2008). Physics settings for using the geant4 toolkit in proton therapy. *IEEE Transactions on nuclear science*, 55(3):1018–1025.

[Jones, 2015] Jones, B. (2015). Towards Achieving the Full Clinical Potential of Proton Therapy by Inclusion of LET and RBE Models. *Cancers*, 7(1):460–480.

[Kalholm et al., 2021] Kalholm, F., Grzanka, L., Traneus, E., and Bassler, N. (2021). A systematic review on the usage of averaged let in radiation biology for particle therapy. *Radiotherapy and Oncology*, 161:211–221.

[Kanai et al., 1999] Kanai, T., Endo, M., Minohara, S., Miyahara, N., Koyama-Ito, H., Tomura, H., Matsufuji, N., Futami, Y., Fukumura, A., Hiraoka, T., et al. (1999). Biophysical characteristics of himac clinical irradiation system for heavy-ion radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 44(1):201–210.

[Kanai et al., 1997] Kanai, T., Furusawa, Y., Fukutsu, K., Itsukaichi, H., Eguchi-Kasai, K., and Ohara, H. (1997). Irradiation of mixed beam and design of spread-out bragg peak for heavy-ion radiotherapy. *Radiation research*, 147(1):78–85.

[Kase et al., 2007] Kase, Y., Kanai, T., Matsufuji, N., Furusawa, Y., Elsässer, T., and Scholz, M. (2007). Biophysical calculation of cell survival probabilities using amorphous track structure models for heavy-ion irradiation. *Physics in Medicine & Biology*, 53(1):37.

[Kase et al., 2006] Kase, Y., Kanai, T., Matsumoto, Y., Furusawa, Y., Okamoto, H., Asaba, T., Sakama, M., and Shinoda, H. (2006). Microdosimetric measurements and estimation of human cell survival for heavy-ion beams. *Radiation research*, 166(4):629–638.

[Kase et al., 2011] Kase, Y., Kanai, T., Sakama, M., Tameshige, Y., Himukai, T., Nose, H., and Matsufuji, N. (2011). Microdosimetric approach to nirs-defined biological dose measurement for carbon-ion treatment beam. *Journal of radiation research*, 52(1):59–68.

[Kase et al., 2013] Kase, Y., Yamashita, W., Matsufuji, N., Takada, K., Sakae, T., Furusawa, Y., Yamashita, H., and Murayama, S. (2013). Microdosimetric calculation of relative biological effectiveness for design of therapeutic proton beams. *J. Radiat. Res.*, 54:485–493.

[Kellerer et al., 1985] Kellerer, A. M. et al. (1985). Fundamentals of microdosimetry. *The dosimetry of ionizing radiation*, 1:77–162.

[Kellerer and Rossi, 1972] Kellerer, A. M. and Rossi, H. H. (1972). The theory of dual radiation action. *Current topics in radiation research*, 8:85–158.

[Kellerer and Rossi, 1978] Kellerer, A. M. and Rossi, H. H. (1978). A generalized formulation of dual radiation action. *Radiation research*, 75(3):471–488.

[Kiffer et al., 2019] Kiffer, F., Boerma, M., and Allen, A. (2019). Behavioral effects of space radiation: A comprehensive review of animal studies. *Life Sci. Space Res.*, 21:1–21.

[Kliauga et al., 1978] Kliauga, P. J., Colvett, R. D., Lam, Y.-M. P., and Rossi, H. H. (1978). The relative biological effectiveness of 160 mev protons i. microdosimetry. *International Journal of Radiation Oncology• Biology• Physics*, 4(11):1001–1008.

[Kuang et al., 2016] Kuang, Y., Nagy, J. D., and Eikenberry, S. E. (2016). *Introduction to mathematical oncology*, volume 59. CRC Press.

[La Tessa et al., 2016] La Tessa, C., Sivertz, M., Chiang, I.-H., Lowenstein, D., and Rusek, A. (2016). Overview of the nasa space radiation laboratory. *Life sciences in space research*, 11:18–23.

[Lea and Catcheside, 1942] Lea, D. E. and Catcheside, D. G. (1942). The mechanism of the induction by radiation of chromosome aberrations in Tradescantia. *Journal of Genetics*.

[Lindborg and Waker, 2017] Lindborg, L. and Waker, A. (2017). *Microdosimetry: experimental methods and applications*. CRC Press.

[Loncol et al., 1994] Loncol, T., Cosgrove, V., Denis, J., Gueulette, J., Mazal, A., Menzel, H., Pihet, P., and Sabattier, R. (1994). Radiobiological effectiveness of radiation beams with broad let spectra: microdosimetric analysis using biological weighting functions. *Radiation Protection Dosimetry*, 52(1-4):347–352.

[M. Kellerer, 2002] M. Kellerer, A. (2002). Microdosimetry: reflections on harald rossi. *Radiation protection dosimetry*, 99(1-4):17–22.

[Manganaro, 2018] Manganaro, L. (2018). *Dose delivery time structure effects in particle therapy: development of a time-resolved microdosimetric-kinetic model and implementation of spatiotemporal treatment plan optimization*. PhD thesis, University of Turin, Italy.

[Manganaro et al., 2017] Manganaro, L., Russo, G., Cirio, R., Dalmasso, F., Giordanengo, S., Monaco, V., Muraro, S., Sacchi, R., Vignati, A., and Attili, A. (2017). A Monte Carlo approach to the microdosimetric kinetic model to account for dose rate time structure effects in ion beam therapy with application in treatment planning simulations. *Medical physics*, 44(4):1577–1589.

[Mariotti et al., 2013] Mariotti, L. G., Pirovano, G., Savage, K. I., Ghita, M., Ottolenghi, A., Prise, K. M., and Schettino, G. (2013). Use of the γ-H2AX assay to investigate DNA repair dynamics following multiple radiation exposures. *PLoS ONE*, 8(11):1–12.

[Martino et al., 2010] Martino, G., Durante, M., and Schardt, D. (2010). Microdosimetry measurements characterizing the radiation fields of 300 MeV/u 12C and 185 MeV/u 7Li pencil beams stopping in water. *Physics in Medicine & Biology*, 55(12):3441.

[Matsuya et al., 2014] Matsuya, Y., Ohtsubo, Y., Tsutsumi, K., Sasaki, K., Yamazaki, R., and Date, H. (2014). Quantitative estimation of DNA damage by photon irradiation based on the microdosimetric-kinetic model. *Journal of Radiation Research*, 55(3):484–493.

[Mazza et al., 2019] Mazza, G., Cirio, R., Olave, J., Fausti, F., Sacchi, R., and Monaco, V. (2019). ABACUS: Two fast amplifiers for the readout of LGAD detectors. *PoS*, page 071.

[Mazzucconi et al., 2019] Mazzucconi, D., Bortot, D., Agosteo, S., Pola, A., Pasquato, S., Fazzi, A., Colautti, P., Conte, V., Petringa, G., Amico, A., et al. (2019). Microdosimetry at nanometric scale with an avalanche-confinement tepc: response against a helium ion beam. *Radiation protection dosimetry*, 183(1-2):177–181.

[McMahon, 2018a] McMahon, S. J. (2018a). The linear quadratic model: usage, interpretation and challenges. *Physics in Medicine & Biology*, 64(1):01TR01.

[McMahon, 2018b] McMahon, S. J. (2018b). The linear quadratic model: usage, interpretation and challenges. *Physics in Medicine & Biology*, 64(1):01TR01.

[McNamara et al., 2015] McNamara et al., A. L. (2015). A phenomenological relative biological effectiveness (rbe) model for proton therapy based on all published in vitro cell survival data. *Physics in Medicine & Biology*, 60(21):8399.

[Menzel et al., 1990] Menzel, H., Pihet, P., and Wambersie, A. (1990). Microdosimetric specification of radiation quality in neutron radiation therapy. *International journal of radiation biology*, 57(4):865–883.

[Missiaggia et al., 2020] Missiaggia, M., Cartechini, G., Scifoni, E., Rovituso, M., Tommasino, F., Verroi, E., Durante, M., and La Tessa, C. (2020). Microdosimetric measurements as a tool to assess potential in-and out-of-field toxicity regions in proton therapy. *Physics in Medicine & Biology*.

[Missiaggia et al., 2022a] Missiaggia, M., Cartechini, G., Tommasino, F., Rovituso, M., Scifoni, E., and La Tessa, C. (2022a). Helium and oxygen beams microdosimetric in- and out-of- field characterization. *in preparation*.

[Missiaggia et al., 2022b] Missiaggia, M., Cartechini, G., Tommasino, F., Scifoni, E., and La Tessa, C. (2022b). Investigation of in- and out-of-field radiation quality and its impact on rbe in proton therapy: microdosimetry experimental assessment and monte carlo validation. *Submitted to International Journal of Radiation Oncology, Biology, Physics*.

[Missiaggia et al., 2022c] Missiaggia, M., Cordoni, F., Scifoni, E., and La Tessa, C. (2022c). Cell survival computation via the generalized stochastic microdosimetric model (gsm^2) - part ii: numerical results via monte carlo simulated spectra. *Submitted to Radiation Research*.

[Missiaggia et al., 2021] Missiaggia, M., Pierobon, E., Castelluzzo, M., Perinelli, A., Cordoni, F., Centis Vignali, M., Borghi, G., Bellinzona, E., Scifoni, E., Tommasino, F., et al. (2021). A novel hybrid microdosimeter for radiation field characterization based on the tissue equivalent proportional counter detector and low gain avalanche detectors tracker: A feasibility study. *Front. Phys. 8: 578444. doi: 10.3389/fphy*.

[Missiaggia et al., 2022d] Missiaggia, M., Pierobon, E., La Tessa, C., and Cordoni, F. G. (2022d). Machine learning techniques applied to therapeutic energies particle tracking in microdosimetry using the novel hybrid detector for microdosimetry (hdm). *Submitted to Physics in Medicine & Biology*.

[Mohamad et al., 2018] Mohamad, O., Yamada, S., and Durante, M. (2018). Clinical indications for carbon ion radiotherapy. *Clinical Oncology*, 30(5):317–329.

[Moliere, 1948] Moliere, G. (1948). Theory of scattering of fast charged particles ii multiple and multiple scattering. *Journal for Natural Research A*, 3(2):78–97.

[Moro et al., 2015] Moro, D., Chiriotti, S., Conte, V., Colautti, P., and Grosswendt, B. (2015). Lineal energy calibration of a spherical tepc. *Radiation protection dosimetry*, 166(1-4):233–237.

[Morstin et al., 1989] Morstin, K., Bond, V., and Baum, J. (1989). Probabilistic approach to obtain hit-size effectiveness functions which relate microdosimetry and radiobiology. *Radiation research*, 120(3):383–402.

[Muraro et al., 2020] Muraro, S., Battistoni, G., and Kraan, A. (2020). Challenges in monte carlo simulations as clinical and research tool in particle therapy: a review. *Frontiers in Physics*, page 391.

[Neyman, 1939] Neyman, J. (1939). On a new class of" contagious" distributions applicable in entomology and bacteriology. *Ann. Math. Statistics*, 10:35–57.

[Nikjoo, 2003] Nikjoo, H. (2003). Radiation track and dna damage. *Iranian Journal of Radiation Research*, 1(1):14–17.

[Norbury et al., 2012] Norbury, J. W., Miller, J., Adamczyk, A. M., Heilbronn, L. H., Townsend, L. W., Blattnig, S. R., Norman, R. B., Guetersloh, S. B., and Zeitlin, C. J. (2012). Nuclear data for space radiation. *Radiation measurements*, 47(5):315–363.

[O'Bryan et al., 2000] O'Bryan, M. V., LaBel, K. A., Reed, R. A., Howard, J., Ladbury, R. L., Barth, J. L., Kniffin, S. D., Seidleck, C. M., Marshall, P. W., Marshall, C. J., et al. (2000). Radiation damage and single event effect results for candidate spacecraft electronics. In *2000 IEEE Radiation Effects Data Workshop. Workshop Record. Held in conjunction with IEEE Nuclear and Space Radiation Effects Conference (Cat. No. 00TH8527)*, pages 106–122. IEEE.

[Olko, 1989] Olko, P. (1989). *Fluctuations of energy deposited in biological targets by ionising radiation*. PhD thesis, PhD thesis, Institute of Medicine, KFA Jülich.

[on Radiological Protection. Committee 4 and on Radiological Protection, 1984] on Radiological Protection. Committee 4, I. C. and on Radiological Protection, I. C. (1984). *Protection of the Public in the Event of Major Radiation Accidents: Principles for Planning: a Report*, volume 12. Pergamon.

[Orlic et al., 1989] Orlic, M., Lazarevic, V., and Boreli, F. (1989). Microdosimetric counters based on semiconductor detectors. *Radiation Protection Dosimetry*, 29(1-2):21–22.

[ORTEC, 2019] ORTEC (2019). Maestro, https://www.ortec-online.com/products/application-software/maestro-mca.

[Otten, 1989] Otten, E. (1989). Treatise on heavy-ion science, vol. 8. *DA Bromley Plenum, New York*.

[Paganetti et al., 1997] Paganetti, H., Olko, P., Kobus, H., Becker, R., Schmitz, T., Waligorski, M. P., Filges, D., and Müller-Gärtner, H.-W. (1997). Calculation of relative biological effectiveness for proton beams using biological weighting functions. *International Journal of Radiation Oncology* Biology* Physics*, 37(3):719–729.

[Pellegrini et al., 2014] Pellegrini, G., Fernández-Martínez, P., Baselga, M., Fleta, C., Flores, D., Greco, V., Hidalgo, S., Mandić, I., Kramberger, G., Quirion, D., et al. (2014). Technology developments and first measurements of Low Gain Avalanche Detectors (LGAD) for high energy physics applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 765:12–16.

[Perl et al., 2012] Perl, J., Shin, J., Schümann, J., Faddegon, B., and Paganetti, H. (2012). Topas: an innovative proton monte carlo platform for research and clinical applications. *Medical physics*, 39(11):6818–6837.

[Pihet et al., 1990] Pihet, P., Menzel, H., Schmidt, R., Beauduin, M., and Wambersie, A. (1990). Biological weighting function for rbe specification of neutron therapy beams. intercomparison of 9 european centres. *Radiation Protection Dosimetry*, 31(1-4):437–442.

[Powers et al., 1968] Powers, E. L., Lyman, J. T., and Tobias, C. A. (1968). Some effects of accelerated charged particles on bacterial spores. *International Journal of Radiation Biology*, 14(4):313–330.

[PTC, 2022] PTC, T. (2022). Trento ptc website. https://protonterapia.provincia.tn.it/eng.

[PTCOG, 2020] PTCOG (2020). Ptcog website. https://www.ptcog.ch/index.php/ptcog-patient-statistics.

[Richter, 2012] Richter, D. (2012). *Treatment planning for tumors with residual motion in scanned ion beam therapy*. PhD Thesis, Technischen Universität Darmstadt.

[Robertson et al., 1994] Robertson, J., Eaddy, J., Archambeau, J., Coutrakon, G., Miller, D., Moyers, M., Siebers, J., Slater, J., and Dicello, J. F. (1994). Relative biological effectiveness and microdosimetry of a mixed energy field of protons up to 200 mev. *Advances in Space Research*, 14(10):271–275.

[Rosenfeld, 2016] Rosenfeld, A. B. (2016). Novel detectors for silicon based microdosimetry, their concepts and applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 809:156–170.

[Rosenfeld et al., 2000] Rosenfeld, A. B., Bradley, P., Cornelius, I., Kaplan, G., Allen, B., Flanz, J., Goitein, M., Van Meerbeeck, A., Schubert, J., Bailey, J., et al. (2000). A new silicon detector for microdosimetry applications in proton therapy. *IEEE Transactions on nuclear science*, 47(4):1386–1394.

[Rossi and Zaider, 1988] Rossi, H. and Zaider, M. (1988). Saturation in dual radiation action. In *Quantitative mathematical models in radiation biology*, pages 111–118. Springer.

[Rossi, 1959] Rossi, H. H. (1959). Specification of radiation quality. *Radiation Research*, 10(5):522–531.

[Rossi and Rosenzweig, 1955] Rossi, H. H. and Rosenzweig, W. (1955). Measurements of neutron dose as a function of linear energy transfer. *Radiation research*, 2(5):417–425.

[Rossi and Zaider, 1991] Rossi, H. H. and Zaider, M. (1991). Elements of microdosimetry. *Medical physics*, 18(6):1085–1092.

[Rovituso and La Tessa, 2017] Rovituso, M. and La Tessa, C. (2017). Nuclear interactions of new ions in cancer therapy: impact on dosimetry. *Transl Cancer Res*, 6:S914–33.

[Santa Cruz et al., 2001] Santa Cruz, G., Palmer, M., Matatagui, E., and Zamenhof, R. (2001). A theoretical model for event statistics in microdosimetry. ii: Nonuniform distribution of heavy ion tracks. *Medical physics*, 28(6):997–1005.

[Santa Cruz, 2016] Santa Cruz, G. A. (2016). Microdosimetry: Principles and applications. *Reports of Practical Oncology and Radiotherapy*, 21(2):135–139.

[Sato and Furusawa, 2012] Sato, T. and Furusawa, Y. (2012). Cell survival fraction estimation based on the probability densities of domain and cell nucleus specific energies using improved microdosimetric kinetic models. *Radiation research*, 178(4):341–356.

[Schardt, 2016] Schardt, D. (2016). Hadrontherapy. In *Basic Concepts in Nuclear Physics: Theory, Experiments and Applications*, pages 55–86. Springer.

[Schardt et al., 2010] Schardt, D., Elsässer, T., and Schulz-Ertner, D. (2010). Heavy-ion tumor therapy: Physical and radiobiological benefits. *Reviews of modern physics*, 82(1):383.

[Schardt et al., 2008] Schardt, D., Steidl, P., Krämer, M., Weber, U., Parodi, K., and Brons, S. (2008). Gsi scientific report 2007. *GSI Report*, 1.

[Schettino et al., 2011] Schettino, G., Ghita, M., Richard, D. J., and Prise, K. M. (2011). Spatiotemporal investigations of DNA damage repair using microbeams. *Radiation Protection Dosimetry*, 143(2-4):340–343.

[Scholz, 2003] Scholz, M. (2003). Effects of ion radiation on cells and tissues. *Radiation effects on polymers for biological use*, pages 95–155.

[Scholz, 2006] Scholz, M. (2006). Dose response of biological systems to low-and high-let radiation. In *Microdosimetric response of physical and biological systems to low-and high-LET radiations*, pages 1–73. Elsevier.

[Scholz et al., 1997] Scholz, M., Kellerer, A., Kraft-Weyrather, W., and Kraft, G. (1997). Computation of cell survival in heavy ion beams for therapy. *Radiation and environmental biophysics*, 36(1):59–66.

[Scholz and Kraft, 1992] Scholz, M. and Kraft, G. (1992). A parameter-free track structure model for heavy ion action cross sections. In *Biophysical modelling of radiation effects*.

[Scholz and Kraft, 1996] Scholz, M. and Kraft, G. (1996). Track structure and the calculation of biological effects of heavy charged particles. *Advances in Space Research*, 18(1-2):5–14.

[Schulte et al., 2008] Schulte, R., Penfold, S., Tafas, J., and Schubert, K. (2008). A maximum likelihood proton path formalism for application in proton computed tomography. *Medical physics*, 35(11):4849–4856.

[Schürmann et al., 2018] Schürmann, R., Vogel, S., Ebel, K., and Bald, I. (2018). The physico-chemical basis of dna radiosensitization: Implications for cancer radiation therapy. *Chemistry–A European Journal*, 24(41):10271–10279.

[Serber, 1947] Serber, R. (1947). Nuclear reactions at high energies. *Physical Review*, 72(11):1114.

[Simoni et al., 2019] Simoni, G., Reali, F., Priami, C., and Marchetti, L. (2019). Stochastic simulation algorithms for computational systems biology: Exact, approximate, and hybrid methods. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 11(6):e1459.

[Smith and Grima, 2019] Smith, S. and Grima, R. (2019). Spatial stochastic intracellular kinetics: A review of modelling approaches. *Bulletin of mathematical biology*, 81(8):2960–3009.

[Snoek et al., 2012] Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25.

[Sokol et al., 2017] Sokol, O., Scifoni, E., Tinganelli, W., Kraft-Weyrather, W., Wiedemann, J., Maier, A., Boscolo, D., Friedrich, T., Brons, S., Durante, M., et al. (2017). Oxygen beams for therapy: advanced biological treatment planning and experimental verification. *Physics in Medicine & Biology*, 62(19):7798.

[Sola et al., 2019] Sola, V., Arcidiacono, R., Boscardin, M., Cartiglia, N., Dalla Betta, G.-F., Ficorella, F., Ferrero, M., Mandurrino, M., Pancheri, L., Paternoster, G., et al. (2019). First FBK production of 50 μm ultra-fast silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 924:360–368.

[Solov'yov et al., 2009] Solov'yov, A. V., Surdutovich, E., Scifoni, E., Mishustin, I., and Greiner, W. (2009). Physics of ion beam cancer therapy: a multiscale approach. *Physical Review E*, 79(1):011909.

[Steidl et al., 2008] Steidl, P., Schardt, D., Iancu, G., Kraemer, M., and Weber, U. (2008). Precision measurements of bragg curves of light-ion beams in water. *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 43.

[Stewart et al., 2018] Stewart, R. D., Carlson, D. J., Butkus, M. P., Hawkins, R., Friedrich, T., and Scholz, M. (2018). A comparison of mechanism-inspired models for particle relative biological effectiveness (rbe). *Medical physics*, 45(11):e925–e952.

[Stineman, 1980] Stineman, R. W. (1980). A consistently well-behaved method of interpolation. *Creative Computing*.

[Tarasov and Bazin, 2008] Tarasov, O. and Bazin, D. (2008). LISE++: Radioactive beam production with in-flight separators. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 266(19-20):4657–4664.

[Team, 2000] Team, R. C. (2000). R language definition. *Vienna, Austria: R foundation for statistical computing*.

[Tobias, 1980] Tobias, C. A. (1980). The repair-misrepair model of cell survival.

[Tobias, 1985] Tobias, C. A. (1985). The repair-misrepair model in radiobiology: comparison to other models. *Radiation Research*, 104(2s):S77–S95.

[Tommasino and Durante, 2015] Tommasino, F. and Durante, M. (2015). Proton radiobiology. *Cancers*, 7(1):353–381.

[Tommasino et al., 2017] Tommasino, F., Rovituso, M., Fabiano, S., Piffer, S., Manea, C., Lorentini, S., Lanzone, S., Wang, Z., Pasini, M., Burger, W. J., La Tessa, C., Scifoni, E., Schwarz, M., and Durante, M. (2017). Proton beam characterization in the experimental room of the Trento Proton Therapy facility. *NIMA*, 869:15–20.

[Tommasino et al., 2015] Tommasino, F., Scifoni, E., and Durante, M. (2015). New ions for therapy. *International journal of particle therapy*, 2(3):428–438.

[Tommasino et al., 2019] Tommasino et al., F. (2019). A new facility for proton radiobiology at the trento proton therapy centre: Design and implementation. *Physica Medica*, 58:99–106.

[Tran et al., 2017] Tran, L. T., Chartier, L., Bolst, D., Pogossov, A., Guatelli, S., Petasecca, M., Lerch, M. L., Prokopovich, D. A., Reinhard, M. I., Clasie, B., et al. (2017). Characterization of proton pencil beam scanning and passive beam using a high spatial resolution solid-state microdosimeter. *Medical physics*, 44(11):6085–6095.

[Tsujii et al., 2004] Tsujii, H., Mizoe, J.-e., Kamada, T., Baba, M., Kato, S., Kato, H., Tsuji, H., Yamada, S., Yasuda, S., Ohno, T., et al. (2004). Overview of clinical experiences on carbon ion radiotherapy at nirs. *Radiotherapy and Oncology*, 73:S41–S49.

[Van Houten et al., 2018] Van Houten, B., Santa-Gonzalez, G. A., and Camargo, M. (2018). Dna repair after oxidative stress: current challenges. *Current opinion in toxicology*, 7:9–16.

[Van Kampen, 1992] Van Kampen, N. G. (1992). *Stochastic processes in physics and chemistry*, volume 1. Elsevier.

[Vassiliev, 2012] Vassiliev, O. N. (2012). Formulation of the multi-hit model with a non-poisson distribution of hits. *International Journal of Radiation Oncology* Biology* Physics*, 83(4):1311–1316.

[Vassiliev, 2017] Vassiliev, O. N. (2017). Microdosimetry. elements of stochastic transport theory. In *Monte Carlo Methods for Radiation Transport*, pages 195–223. Springer.

[Vignati et al., 2017] Vignati, A., Monaco, V., Attili, A., Cartiglia, N., Donetti, M., Mazinani, M. F., Fausti, F., Ferrero, M., Giordanengo, S., Ali, O. H., et al. (2017). Innovative thin silicon detectors for monitoring of therapeutic proton beams: preliminary beam tests. *Journal of Instrumentation*, 12(12):C12056.

[Wambersie, 1994] Wambersie, A. (1994). Contribution of microdosimetry to the specification of neutron beam quality for the choice of the clinical rbe'in fast neutron therapy. *Radiation Protection Dosimetry*, 52(1-4):453–460.

[Wambersie et al., 1990] Wambersie, A., Pihet, P., and Menzel, H. (1990). The role of microdosimetry in radiotherapy. *Radiation Protection Dosimetry*, 31(1-4):421–432.

[Weber and Frey, 2017] Weber, M. F. and Frey, E. (2017). Master equations and the theory of stochastic path integrals. *Reports on Progress in Physics*, 80(4):046601.

[Wedenberg and Toma-Dasu, 2014] Wedenberg, M. and Toma-Dasu, I. (2014). Disregarding RBE variation in treatment plan comparison may lead to bias in favor of proton plans. *Medical Physics*.

[Weinan et al., 2019] Weinan, E., Li, T., and Vanden-Eijnden, E. (2019). *Applied stochastic analysis*, volume 199. American Mathematical Soc.

[Wilson, 1946] Wilson, R. R. (1946). Radiological use of fast protons. *Radiology*, 47(5):487–491.

[Wroe et al., 2009] Wroe, A., Clasie, B., Kooy, H., Flanz, J., Schulte, R., and Rosenfeld, A. (2009). Out-of-field dose equivalents delivered by passively scattered therapeutic proton beams for clinically relevant field configurations. *Int. J. Radiat. Oncol. Biol. Phys.*, 73(1):306–313.

[Wroe et al., 2007] Wroe, A., Rosenfeld, A., and Schulte, R. (2007). Out-of-field dose equivalents delivered by proton therapy of prostate cancer. *Medical physics*, 34(9):3449–3456.

[Zaider and Brenner, 1985] Zaider, M. and Brenner, D. J. (1985). On the microdosimetric definition of quality factors. *Radiation research*, 103(3):302–316.

[Zaider et al., 1996] Zaider, M., Rossi, B. H. H., and Zaider, M. (1996). *Microdosimetry and its Applications*. Springer.

[Zeitlin and La Tessa, 2016] Zeitlin, C. and La Tessa, C. (2016). The role of nuclear fragmentation in particle therapy and space radiation protection. *Front. Oncol.*, 6:1–13.

[Zhu et al., 2019] Zhu, H., Chen, Y., Sung, W., McNamara, A. L., Tran, L. T., Burigo, L. N., Rosenfeld, A. B., Li, J., Faddegon, B., Schuemann, J., et al. (2019). The microdosimetric extension in topas: development and comparison with published data. *Physics in Medicine & Biology*, 64(14):145004.

[Zirkle et al., 1951] Zirkle, R. E., Marchbank, D. F., and Kuck, K. D. (1951). Exponential and sigmoid survival curves resulting from alpha and x irradiation of aspergillus spores. Technical report, Institute of Radiobiology and Biophysics, Univ. of Chicago; Argonne National.