

**DEVELOPMENT OF MULTISCALE
RADIOBIOLOGY SIMULATION AND
DAQ SYSTEM FOR RADIOBIOLOGY
BEAMLINE**

TAN HONG QI

NATIONAL UNIVERSITY OF SINGAPORE

2018

**DEVELOPMENT OF MULTISCALE
RADIOBIOLOGY SIMULATION AND
DAQ SYSTEM FOR RADIOBIOLOGY
BEAMLINE**

Tan Hong Qi

(B.Sc (Hons), NUS)

Supervisor: Associate Professor Andrew Anthony Bettoli

Examiner: Associate Professor Thomas Osipowicz

Examiner: Professor Mark Breese

Examiner: Professor Roger Webb, University of Surrey

**A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF
PHILOSOPHY
NUS GRADUATE SCHOOL FOR INTEGRATIVE
SCIENCES AND ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE**

2018

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety.

I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

hongqi

Tan Hong Qi

10th November 2018

Acknowledgment

This PhD turns out to be tougher than I originally conceived, but nonetheless it was a rewarding experience and I am grateful to have gone through it. I chose to work on a project in which there was no predecessor in Singapore and learned a lot in the process. There are several people I would like to thank during my PhD journey. Firstly, I would like to thank my supervisors A/P Andrew Bettoli and Prof. Frank Watt for encouraging me to pursue a work that I am genuinely interested in, and giving me the freedom to do so. Secondly, I would like to thank Khong Wei from NCCS for widening my perspective on the utility of GEANT4 simulation in clinical settings. Thirdly, I would like to thank my fellow colleagues such as Mike, Prashant, Tao Ye and HueiMing for many interesting technical discussion and the casual small talks which adds a bit of spice to my PhD life. Fourth, I would like to thank my family for their support, and my Father in particular, who has served as a role model to me. He was an excellent RF engineer and programmer during his time, and was still intrigued and eager to learn about new technologies till this age. His passion for learning has an influence on me at a very young age. Lastly, I would like to thank my wife Rongxuan, whom I married during the third year of my PhD. Her motivation and constant pushing helps me to get through moments of self-doubt.

Contents

Abstract	viii
List of Figures	ix
1 Introduction	1
1.1 Backgrounds and Motivations	1
1.2 Objectives	3
1.3 Thesis Organisations	4
2 Introduction to the Radiobiology Beamline	6
2.1 Introduction	6
2.2 CIBA Radiobiology Beamline	9
2.2.1 Singletron Accelerator	9
2.2.2 Magnetic Focusing	10
2.2.3 Beam Scanning and Blanking	15
2.2.4 End Station	16
2.3 Typical Experiment Protocol	20
2.4 Summary	22
3 Beamline Control and DAQ system	23
3.1 Introduction	23
3.2 Hardware	24
3.2.1 Overview of the set-up	24
3.2.2 Detector	25

3.2.3	Signal Conditioning Unit and ADC	26
3.2.4	FPGA	30
3.3	Software	32
3.3.1	Planning	32
3.3.2	STIM Imaging	35
3.3.3	Scanning-imaging calibration	42
3.3.4	Manual Cellular Targeting Module	48
3.3.5	Single Proton Detection Algorithm and Results	50
3.4	Future Development	55
3.5	Summary	56
4	Overview of Radiobiology Simulation Software	58
4.1	Motivation and Aim	58
4.2	Review of existing simulation software	60
4.3	Features and Novelties in IRSGS	62
4.4	Architecture and Blueprints of IRSGS	63
5	Initializing the Simulation	65
5.1	Chromosome and cellular target geometry construction	65
5.1.1	Implementation details of Nucleosome construction . .	66
5.1.2	Implementation details of Chromatin construction . .	67
5.1.3	Implementation details of Chromosome construction .	70
5.1.4	Implementation details of Nucleus and Cell construc- tion	72
5.2	Introduction to GEANT4-DNA	74
5.2.1	Introduction to GEANT4 Monte Carlo	74
5.2.2	Overview of Physical Models in GEANT4DNA	76
5.2.3	Overview of Radiation Chemistry Module in GEANT4- DNA	78
5.3	Summary	81

6 Quantifying Radiation Damage with IRSGS	83
6.1 Introduction	83
6.2 Determination of DSB	84
6.2.1 Prerequisites	84
6.2.2 Calculating Direct and Indirect damage to DNA . . .	85
6.2.3 Algorithm for classification of SSB and DSB	87
6.2.4 Results of DSB for Protons	88
6.2.5 Results of DSB for X-rays	90
6.3 DNA Repair Model	93
6.3.1 Biological Introduction to DNA repairs and Chromo- some Aberration	93
6.3.2 Non-Homologous End Joining Model	94
6.3.3 Chromosome Aberration Model	104
6.3.4 Results and Discussions	107
6.4 Summary	114
7 Conclusion and Future Outlook	116
7.1 Conclusion	116
7.2 Future Outlook	117
Bibliography	120
Appendix A Derivation of $P_{spatial}(synapse)$	140

Abstract

Proton Therapy will be available in Singapore in 2021 and it offers a greater conformal dose distribution compared to conventional radiotherapy. However, despite this advantage, proton radiobiology is still not well understood and this work seeks to tackle this problem via two approaches. The first approach is the development of the DAQ system of the radiobiology beam line which will subsequently be used to perform deterministic targeted irradiation of cell for fundamental study. The second involves the development of a mechanistic radiobiology simulation software to quantify the initial double strand break (DSB) damage in a chromosome model and the subsequent DSB repair rate. The simulation shows that Proton and X-rays result in both a very different initial DSB spectrum and repair kinetics. The proton beam is shown in the simulation to be more biologically lethal compared to X-ray.

List of Figures

1.1	A figure showing the comparison of the dose deposition pattern between different forms of radiation including proton, carbon, electron and photon from Ref. [11].	2
1.2	A figure showing the number of particle therapy centres around the world up to early 2020s from Ref. [11].	3
2.1	A figure showing the layout of the accelerator facility and all the beamlines and their research purpose in CIBA, including the new radiobiology beamline which is shown in red.	9
2.2	A figure the 4 magnetic quadrupoles for focusing and the magnetic scanner for deflecting the beam	11
2.3	A figure of the schematics of the Quadruplets (left) and the Triplets (right).	11
2.4	A figure of the WINTRAX simulation of the proton rays focused at a working distance of 20 cm and a scan size of 1 mm in both directions for both pre-scan and post-scan configurations. The scale is shown in the figure with 500 μm per division for top two figures and 200 μm for bottom two figures.	13
2.5	A figure of the WINTRAX simulation result of pre-scan compared to post-scan for various working distance for a scan size of 1 mm in both directions.	14

2.6	A figure of the scan controller OM-1015E and the constant current amplifier for both horizontal and vertical direction.	16
2.7	An image of the entire end station of the radiobiology beam-line.	17
2.8	An image of the vacuum chamber and the nozzle with a 1 mm opening.	18
2.9	An image of the XYZ stage used to position the cell holder relative to the nozzle.	19
2.10	An image of the custom optical microscope used in our beam-line.	20
3.1	An overview of the DAQ system design. It shows the relationship between various hardwares and the software. The arrow refers to the flow of input signal.	24
3.2	An image of the hardware components of the DAQ system of the radiobiology beam line.	25
3.3	An image of the PIN diode for measuring energy of the proton in the radiobiology beam line.	26
3.4	A circuit diagram of a typical charge sensitive preamplifier .	28
3.5	The programme organization of the DAQ software consisting of the FPGA level code and the GUI level code	34
3.6	A screenshot of the STIM imaging interface. The words in red refers to the description of the various windows and buttons. The setting windows set the parameters for acquiring STIM image.	36
3.7	The comparison of the Kepco current output and FPGA output trace for short dwell time of 1 microsecond (top) and long dwell time of 10 microseconds (bottom). The delay between the amplifier and FPGA output is non-negligible for short dwell time.	38

3.8 Energy spectrum of the proton beam measured by the Silicon barrier detector through the gold grid in air. The rightmost high energy peak originates from proton travelling through empty space of the grid and the left most low energy peak originates from proton travelling through the bar of the grid.	40
3.9 The STIM image of the gold grid by extracting the median energy in each pixel. Yellow color refers to high median energy and green refers to low median energy. The grid pattern is evident through STIM imaging.	41
3.10 The extraction of the edge profile in both the X and Y direction together with the best fit error function to determine the resolution in the horizontal and vertical direction respectively.	42
3.11 The fluorescence image of the cell nucleus together with the damage regions in both line (top) and array scans (bottom). The blue fluorescence image correspond to DAPI label for the nucleus and the red image correspond to γ -H2AX label for the DNA damage sites. The damage pattern can be clearly seen in the nucleus.	43
3.12 A screenshot of the calibration interface for mapping the optical image pixel coordinate to the magnetic scanner scanning voltage.	45
3.13 A photo of the experimental set-up to test the calibration and targeting functionality of the software.	46
3.14 A screenshot of the actual calibration process at the 14th and 33th iterations. The white circles in the right graph are the (X,Y) coordinates of the detected beam spots from the optical image in the left figure. The red and green line are the linear fit to the (X,Y) coordinates.	47
3.15 A screenshot of the manual targeting interface	48

3.16 Testing of manual targeting functionality and accuracy with the software by targeting random positions (shown in red solid circle) and determining the actual beamspot position (shown in white spot). The scale is shown in red in the upper right corner of the figure.	50
3.17 Images of using the targeting software to write "CIBA" in a 300 microns (top) and 100 microns (bottom) scale	51
3.18 The colour scale represents probability of only 1 proton arriving within a specified <i>integration time</i> period (Y-axis) under a specific proton count rate (X-axis).	52
3.19 The schematics of the experiment for measuring the fidelity of the SPD system	54
3.20 The result of the SPD fidelity measurement for different integration time from 0 to 0.5 milliseconds.	55
3.21 The result of the SPD fidelity measurement for two different count rates: 200 and 800 s^{-1}	55
4.1 A figure showing the multi-scale, multi-disciplinary and multi-physics nature of the understanding of biological effect of radiation. Reprinted from [1], with the permission of AIP Publishing.	60
4.2 A table of the existing attempts to simulate biological effect of radiation	61
4.3 Figure of the blueprint of IRSIGS	63
5.1 3D rendering of the nucleosome structure used in IRSIGS using POV-ray	67
5.2 3D rendering of the repeating unit of the zigzag chromatin structure used in IRSIGS using POV-ray	68

5.3	3D rendering of a section of the zigzag chromatin structure used in IRSGS using POV-ray	69
5.4	3D rendering of the repeating unit of the solenoidal chromatin structure used in IRSGS using POV-ray	70
5.5	3D rendering of a section of the solenoidal chromatin structure used in IRSGS using POV-ray	71
5.6	3D rendering of a chromatin fibre loop structure used in IRSGS using POV-ray. The fibre loop is constructed for the solenoidal chromatin.	72
5.7	3D rendering of a chromosome used in IRSGS using POV-ray. The chromosome is modelled as a cube with length 1.5 microns containing 1575 chromatin fibre loops	73
5.8	3D rendering of the nucleus model used in IRSGS using POV-ray. The nucleus shown as a yellow sphere has a diameter of 10 microns and contains 23 pairs (46 blue cubes in total) of chromosomes with limited overlap as supported by experiment [4]	74
5.9	Table showing the details of the physical models available in GEANT4-DNA and used in IRSGS for an electron	77
5.10	Table showing the details of the physical models available in GEANT4-DNA and used in IRSGS for a proton	77
5.11	Table showing the decay branching ratio of the excited or ionized water molecule at 1 ps after the physico-chemical stage in GEANT4-DNA from [5]	79
5.12	Table showing the diffusion constants of the free radicals used in GEANT4-DNA from [5]	80
5.13	Table showing reaction and reaction rates between free radicals in GEANT4-DNA from [5]	80
6.1	Schematics of Direct and Indirect damage processing	87

6.2	Comparison of SSB and DSB of simulation and experimental data for 1500 eV electron	88
6.3	A comparison of the DSB from IRSIGS and other simulation software and experimental data	89
6.4	The comparison of the direct and indirect strand breaks	91
6.5	The comparison of the SSB yield from direct and indirect effect, and the DSB yield. The simulation result from IRSIGS is compared to that from [6]	93
6.6	A review of the existing NHEJ repair models	95
6.7	This figure shows the biological network of DSB repair mechanisms with the rate constant definition for each reaction. The repair mechanism differs for simple (shown in blue) and complex DSBs (shown in red).	97
6.8	Function of $P_{spatial}(synapse)$ against time for different parameters of Δ and D for a DSB ends of separation 300 nm .	101
6.9	Function of $P_{spatial}(synapse)$ against time for different distance between a pair of DSB ends for $D = 100 \text{ nm}^2/\text{s}$ and $\Delta = 20 \text{ nm}$	102
6.10	This figure shows the possible chromosome aberration patterns that can arise from DSBs in one or two chromosomes. This figure is reprinted from [7] with the permission of John Wiley and Sons.	104
6.11	A) Comparison of the rate of repair of DSBs for different diffusion constants in the NHEJ repair model. B) Comparison of the rate of repair of DSBs for different Δ constants in the NHEJ repair model.	108

6.12 A) Comparison of the rate of repair of DSBs for different rate constants c_5/s_3 (last reaction) in the NHEJ repair model. B) Comparison of the rate of repair of DSBs for different rate constants c_1/s_1 (first reaction) in the NHEJ repair model. . .	109
6.13 This figure shows the experimental data of fraction of DSB within a nucleus after 40 Gy of Ce-137 irradiation and the simulation data fitting result.	112
6.14 This figure shows the histogram of the probability of correct rejoining for DSBs formed under 10 Gy of Ce-137 irradiation.	113

Chapter 1

Introduction

1.1 Backgrounds and Motivations

Cancer accounted for 30% of the deaths in Singapore from 2014 to 2017 [8] and it is expected that there will be a 3 to 5 % rise per annum in new cancer patients up to year 2030 partly due to ageing population. This motivates an increased efforts in Singapore to improve Cancer treatment. In particular, a new radiation therapy technology using proton beams instead of conventional X-rays is being constructed in National Cancer Centre of Singapore (NCCS) [9]. Radiation therapy uses energetic radiation to kill cancer cells and is an important component of the initial and subsequent treatments of 55-60% of the cancer patients. The preference of Proton therapy over conventional X-rays is due to the physics of energy deposition [10]. A proton beam (or charged particle beam in general) deposits most of its energy at the end of range in a localized region known as the Bragg peak, whereas X-rays undergoes exponential attenuation as it travels through matter. A comparison of these two forms of radiation is shown in Fig. 1.1 [11]. The Bragg peak is highly desirable as it concentrates dose on the tumour and reduces dose to normal healthy tissues. This is important not only to prevent unwanted collateral damage to healthy tissue, but also

to reduce secondary cancer occurrences [12, 13] which could arise due to mutation induced by the radiation effect on normal cells.

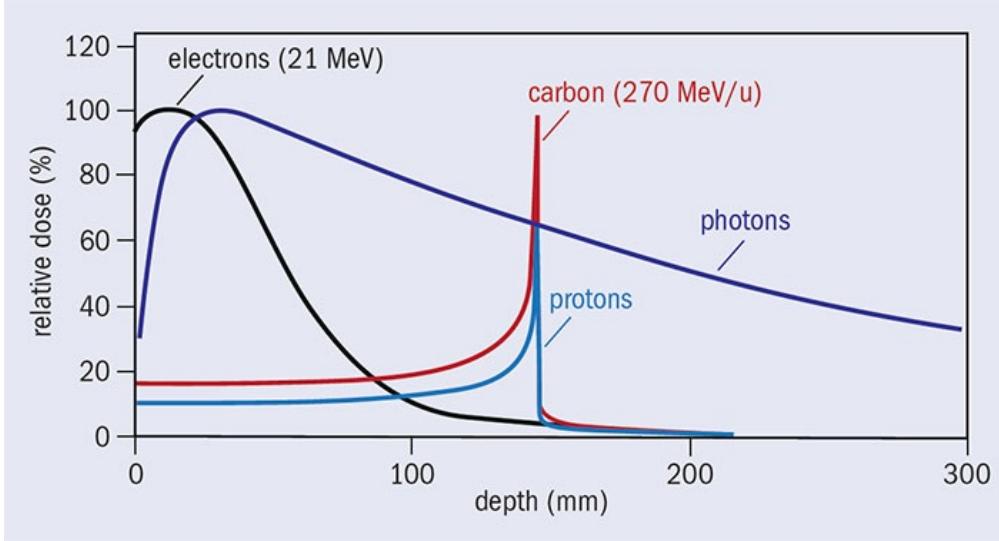


Fig. 1.1. A figure showing the comparison of the dose deposition pattern between different forms of radiation including proton, carbon, electron and photon from Ref. [11].

Due to the superior dose deposition patterns of particle beams, there are increasing number of particle therapy centres being constructed around the world. The statistics of current and anticipated numbers of particle therapy centres to be constructed is shown in Fig. 1.2. Hence, it is evident that particle therapy is indeed gaining popularity around the world with a two-fold increase of particle therapy centres from 2016 to the early 2020s.

However, despite the growing interest in proton therapy, the radiobiological effect of protons is still not well-elucidated. There is evidence that proton triggers a different biological effect compared to X-rays [14] and the Relative Biological Effectiveness (RBE) [15] value is still subjected to a large uncertainty [16]. More research is definitely required in proton radiotherapy to fully utilize the benefits of Proton Therapy. Thus, coupled with the building of a new proton therapy centre in NCCS projected for completion in 2021, the Centre of Ion Beam Applications (CIBA) based in the National University of Singapore is currently building a proton microbeam facility which is dedicated to perform targeted single cell irradiation in a

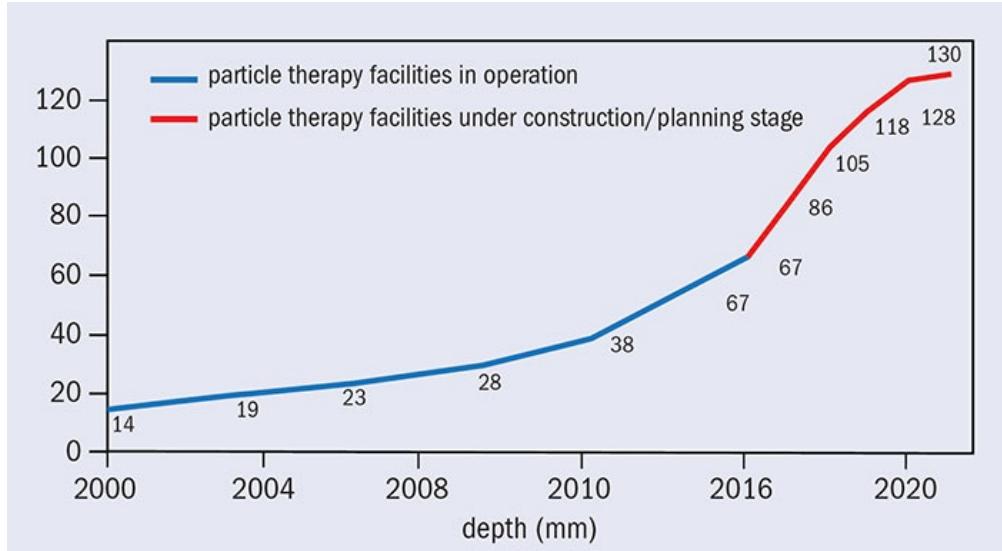


Fig. 1.2. A figure showing the number of particle therapy centres around the world up to early 2020s from Ref. [11].

deterministic manner. This will allow a fundamental study of damage interactions and subsequent cellular response between a MeV proton beam and cells or tissue, which could potentially give insights on the uniqueness of proton radiobiology.

1.2 Objectives

In CIBA, there are two main modes of inquiries towards a better understanding of proton radiobiology. The first main mode is mentioned briefly in Section 1.1, and concerns the use of a proton microbeam to induce localized damage in the cell with an accurate dose of protons, and observe the subsequent biological response. The second mode is the development of simulation software to understand and predict the biological damage (Double Strand Breaks in particular). This not only allows the testing of various working hypotheses, but also gives us the capacity to explain the experimental results on cellular damage. The contributions and objectives of my thesis in these two channels are as follows:

1. Develop the Data Acquisition (DAQ) system and dose targeting for

the radiobiology beamline. This is integral for automatic control of the operation of the beamline and includes targeting of selected cells with a known proton dose. This involves both hardware and software development. The hardware interface unit is responsible for controlling the beam scanning and beam blanking. The software interface consists of developing a Graphical User Interface (GUI) to allow untrained users to operate the beamline easily and efficiently.

2. Develop a simulation software to predict firstly, the DSB distributions of both protons and photons in a cellular geometry, including direct and indirect contributions. Secondly, model and simulate the temporal structure of the DNA repair after the initial radiation insult. The software is then used to generate DSB yields for different energies of protons and photons to better understand the mechanistic difference between them in generating DSBs.

1.3 Thesis Organisations

The thesis is divided into two main parts. The first part, contained in chapters 2 to 3, details the contribution made towards beamline development, especially in the DAQ and cell targeting aspect. The second part (chapters 4 to 6), details the contribution towards development of simulation software for cell damage and response, and the scientific results that ensue. Chapter 2 presents the overview of the radiobiology beamline including the components and the designs, and chapter 3 gives a detailed presentation of the development of the DAQ system. Chapter 4 gives an introduction to the radiobiology software including its architecture and current capability, as well as a review of the existing radiobiology softwares and their strengths and weaknesses. Chapter 5 presents the construction of the chromosome geometry that is used in IRSIGS and an introduction to GEANT4-DNA

including the physics models that are used in IRSGS. Chapter 6 presents a two core module in IRSGS. The first is the damage-processing module that is responsible for calculating DSB and SSB and the second is the development of the novel modelling method of the NHEJ process which is important for determine Chromosome Aberration in cells. Lastly, chapter 7 is the conclusion of this thesis and presents a future outlook.

Chapter 2

Introduction to the Radiobiology Beamline

2.1 Introduction

The development of automated charged particle microbeam technology represents a significant step towards understanding radiation effects at the cellular level [17, 18]. Such technology is unique for two reasons. First, it allows precise control of the number of particles hitting the cell which translates to an accurate determination of dose in the cell. This is in contrast to the use of a radioactive source such as alpha or gamma rays or broad beam irradiation which leads to a Poisson distribution in the number of particle transversals in the cell. Second, the microbeam facility employs either a collimated or focused beam, which results in a beamspot size down to around one micrometre. This allows an accurate spatial position of the beam at the cell. Hence, such a system allows *deterministic* and *targeted* cellular irradiation which is unique to charged particle microbeam radiobiology experiments.

The first rudimentary charged particle microbeam irradiation of cells experiment took place in 1950 [19]. Subsequently with technological ad-

vances, the first fully automated system was developed in Gray laboratory in Northwood, UK [20] and RARAF in Columbia University, USA [21]. These facilities used collimated beam produced from Van de Graaf Accelerator and achieved beamspot size in air of several micrometres. Currently, there are many such facilities in the world which use magnetic or electrostatic focusing to achieve better lateral resolutions, and coupled with other technologies can achieve excellent measurements of the damage response in the cell. There are currently 11 active microbeam facilities for cellular irradiation [22]. The more noteworthy facilities are 1) SNAKE in Munich, Germany [23], 2) RARAF in Columbia University, USA [21], 3) CENBG in Bordeaux, France [24], 4) RIKEN in Japan [25] and 5) IMP in Fudan, China [26]. The SNAKE facility in Munich has made tremendous progress in using a high energy ion microbeam to probe radiation damage in cells. For example, they did time-lapsed fluorescence imaging of the formation of 53BP1 foci in the nucleus directly after carbon ion irradiation [27] and recently they discovered that mitochondria loses membrane potential rapidly after irradiation [28]. The leading facilities mainly differ in the choice of the particles (proton or carbon ion etc) and their energy as well as the imaging capability at the end station [22].

In CIBA, we have constructed a new radiobiology beamline to add to the list of *live cell micro-irradiation* facilities above. The layout of all the beamlines in CIBA is shown in Fig. 2.1 including the new radiobiology beamline which is labelled by a red box. CIBA is an interdisciplinary research group that specializes in producing highly focused proton beams for various applications, for example:

1. Proton beam writing [29]
2. Material analysis using techniques such a PIXE, RBS, ERDA or IBIC [30, 31]

3. High resolution imaging of cells and tissues using Scanning Transmission Ion Microscopy (STIM) or Ionoluminescence techniques [32, 33]
4. Irradiation of live cell in air for radiobiological research (new)

Each of this work is carried out using separate beamlines as shown in Fig. 2.1. CIBA has produced the best resolution proton beam in proton beam writing and in STIM imaging applications with less than 50 nanometres in vacuum [29, 34] being routinely achieved. We hope to use the same focusing technology to achieve high resolution in the new radiobiology beamline, since this will enable us to produce the most spatially precise proton beam on the cell, thereby pushing the targeting ability to an unprecedented scale.

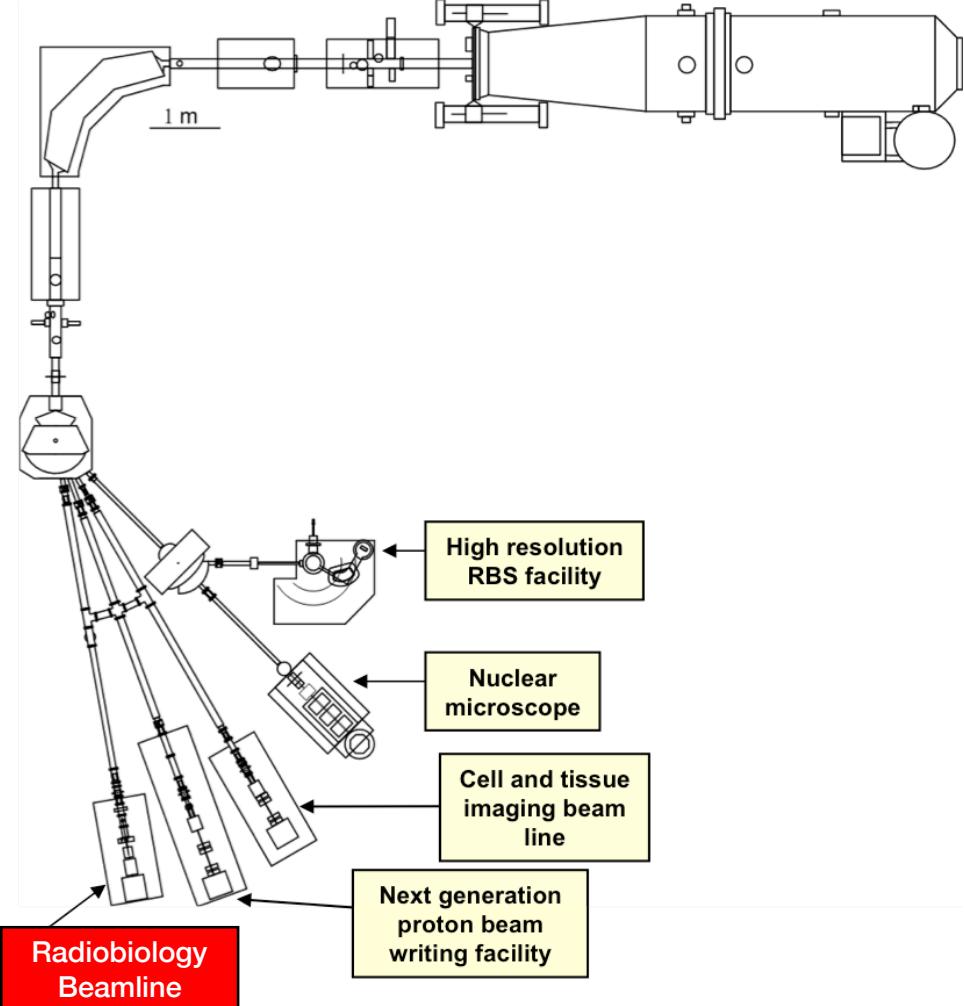


Fig. 2.1. A figure showing the layout of the accelerator facility and all the beamlines and their research purpose in CIBA, including the new radiobiology beamline which is shown in red.

2.2 CIBA Radiobiology Beamline

The schematics of the entire accelerator facility in CIBA is shown in Fig. 2.1 and further descriptions of these facilities are covered in the following subsections. The radiobiology beamline is shown in Fig. 2.2 and most development in the radiobiology beamline occurs at the End Station.

2.2.1 Singletron Accelerator

The proton beam is provided by a 3.5 MV high brightness High Voltage Engineering Europa Singletron ion accelerator [35, 36]. This particle accelerator has a high energy stability which results in a stable beam in terms of

brightness. This is beneficial for accurate dose determination in radiobiology experiments. The accelerator generates its high voltages electronically by following the Cockcroft-Walton principle. It uses a radio frequency (RF) ion source to produce a variety of gaseous ion species such as proton, hydrogen (H_2^+), helium and oxygen. For the radiobiology experiment, we typically run at proton energies around 2 MeV.

2.2.2 Magnetic Focusing

The focusing of MeV proton beam is achieved using magnetic quadrupoles lenses from Oxford Microbeams ltd. A single magnetic quadupole is only capable of line focus, thus more than 2 quadrupoles are needed to focus the beam into a spot. In the radiobiology beamline, there are a total of 4 magnetic quadrupoles which can be linked into a variety of different configurations. In our case we have concentrated on the Oxford un-spaced triplet and Russian Quaduplet [37] configurations. The four quadrupoles are shown in Fig. 2.2 and the schematics of the two different configurations utilised are shown in Fig. 2.3. The working distance is defined as the distance from the last quadupole lens to the plane containing the cells sample and it is 20.0 cm in the case of our radiobiology beamline. **C** refers to *converging* lens which converges the beam into a horizontal line and **D** refers to *diverging* which focuses the beam into a vertical line.

Apart from the choice of the two quadrupole configurations, there are considerations on the placement of the magnetic scanner, which can either be *before*(*pre-scan*) or *after*(*post-scan*) the quadrupoles lenses. The decision relies heavily on the final beamspot size of the 2 arrangements at the cell plane in Fig. 2.3. The final beamspot size at the cell plane depends on three factors - demagnification factor from the focusing quadrupoles, aberrations effect and lateral straggling from proton transversing the air into the cell. These are encompassed in

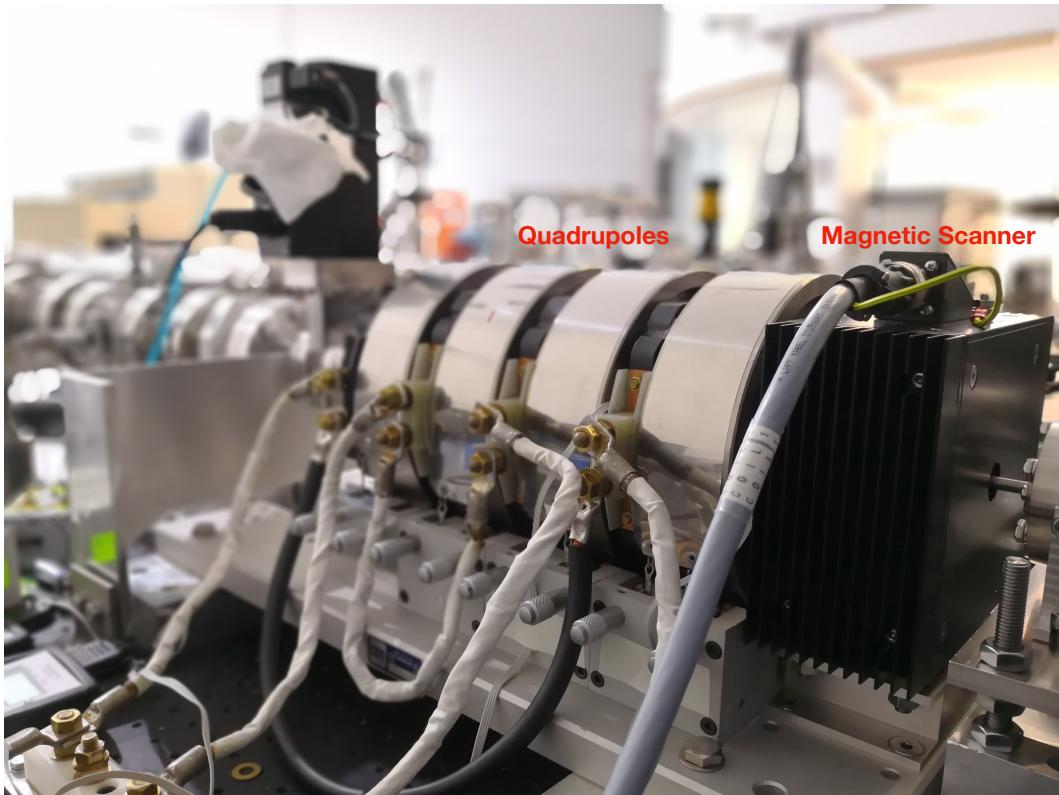


Fig. 2.2. A figure the 4 magnetic quadrupoles for focusing and the magnetic scanner for deflecting the beam

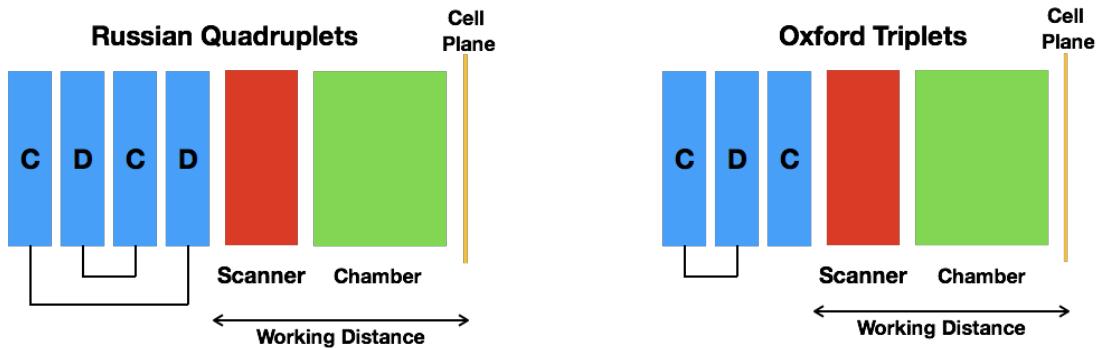


Fig. 2.3. A figure of the schematics of the Quadruplets (left) and the Triplets (right).

$$\Delta_{final} = \Delta_{demag} + \Delta_{aberration} + \Delta_{straggling}, \quad (2.1)$$

where Δ_{final} is the final beamspot size which is a sum of contributions from the three factors (assuming they are independent from each other).

A pre-scan system causes the beam to enter the quadrupoles with a larger entrance angles and larger deviations from the central axis, which results

in larger aberration effects especially the third order spherical aberration. However, such arrangement can also result in a smaller working distance of 10 cm which in turn leads to a greater demagnification factor and smaller geometric image size. On the other hand, a post-scan system gives a smaller demagnification factor (due to longer working distance of 20 cm) and a smaller aberration effect. Hence, a closer examination of the trade-off between aberration and demagnification is required to arrive at an optimised position for the scanner. In addition, the use of a pre-scan system in general leads to higher aberrations at the extremities of the scan, whereas the post scan system allows good resolutions across the scan profile. This is important for targeting accuracy of cells throughout the scan area.

The study of the trade-off is done using WINTRAX simulation software [38] and an object size of 2 x 2 microns and a proton beam of 2 MeV (with zero energy dispersion) through an oxford triplet configuration are assumed. The aim is to calculate the beamspot size at various working distance by varying the magnetic field in the quadrupole to optimize the beamspot size. An example of the proton rays for pre-scan and post-scan at a maximal scan size of 1 mm is shown in Fig. 2.4. The final result of the simulation is shown in Fig. 2.5. Pre-scan system results in a larger beam spot size than post-scan system which decreases with increasing working distance whereas the post-scan system shows a gradual increase in beamspot size with working distance (not visible in Fig. 2.5). This shows that for a 1 mm scan size, the aberration effect dominates the demagnification effect and the post-scan configuration is clearly a better choice. Thus, for large area irradiation of cells where considerable scanning is required, post-scan configuration will maintain the beamspot integrity at all scanning distance with little degradation compared to pre-scan.

The second design consideration involves the choice between Oxford triplet and Russian quadruplet. Oxford triplet is a high excitation lens

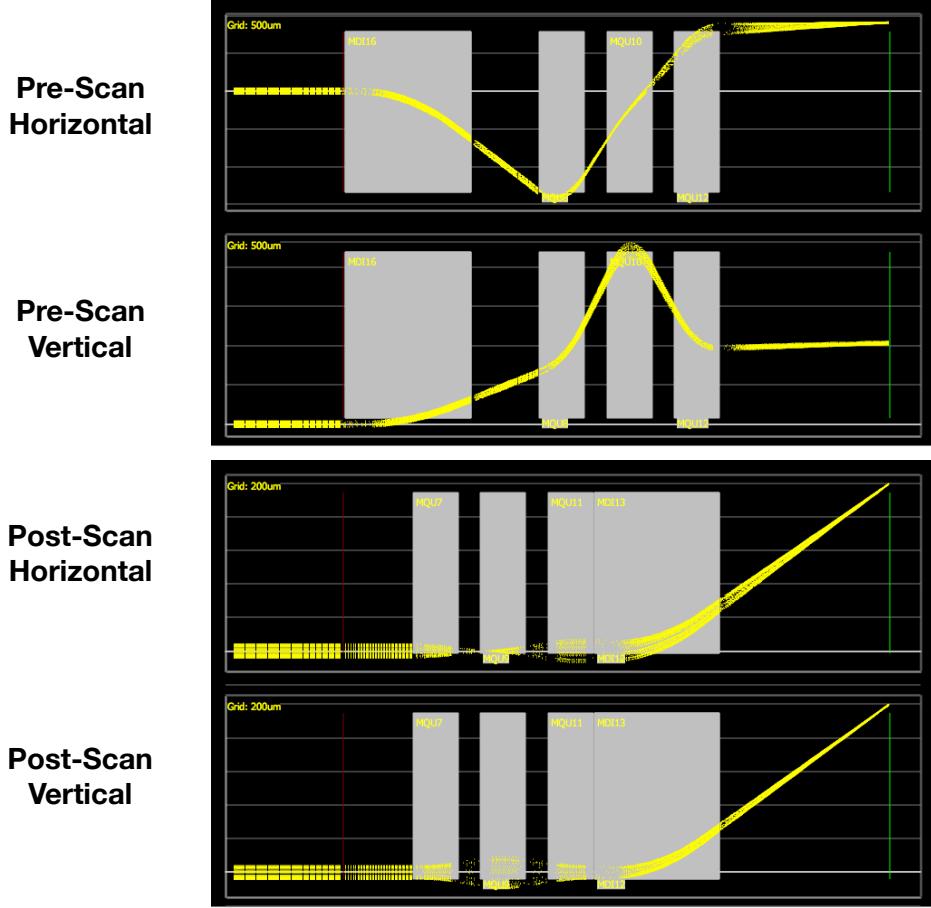


Fig. 2.4. A figure of the WINTRAX simulation of the proton rays focused at a working distance of 20 cm and a scan size of 1 mm in both directions for both pre-scan and post-scan configurations. The scale is shown in the figure with $500 \mu m$ per division for top two figures and $200 \mu m$ for bottom two figures.

system with unequal demagnification in both directions. It uses high lens currents which results in the beam envelop crossing the beam axis and having a short focal length, yielding high demagnification factors. For a working distance of 20 cm, the Oxford triplet gives a horizontal and vertical demagnification of 87 and 29 respectively, whereas the Russian quadruplet gives a demagnification of 20 in both directions. Thus, from the point of view of optimizing beam resolution, The Oxford triplet is clearly a better configuration. Furthermore, the Oxford triplet also benefits from the de-coupled characteristic between the orthogonal directions. This means the vertical direction focusing can be done almost independently of the hori-

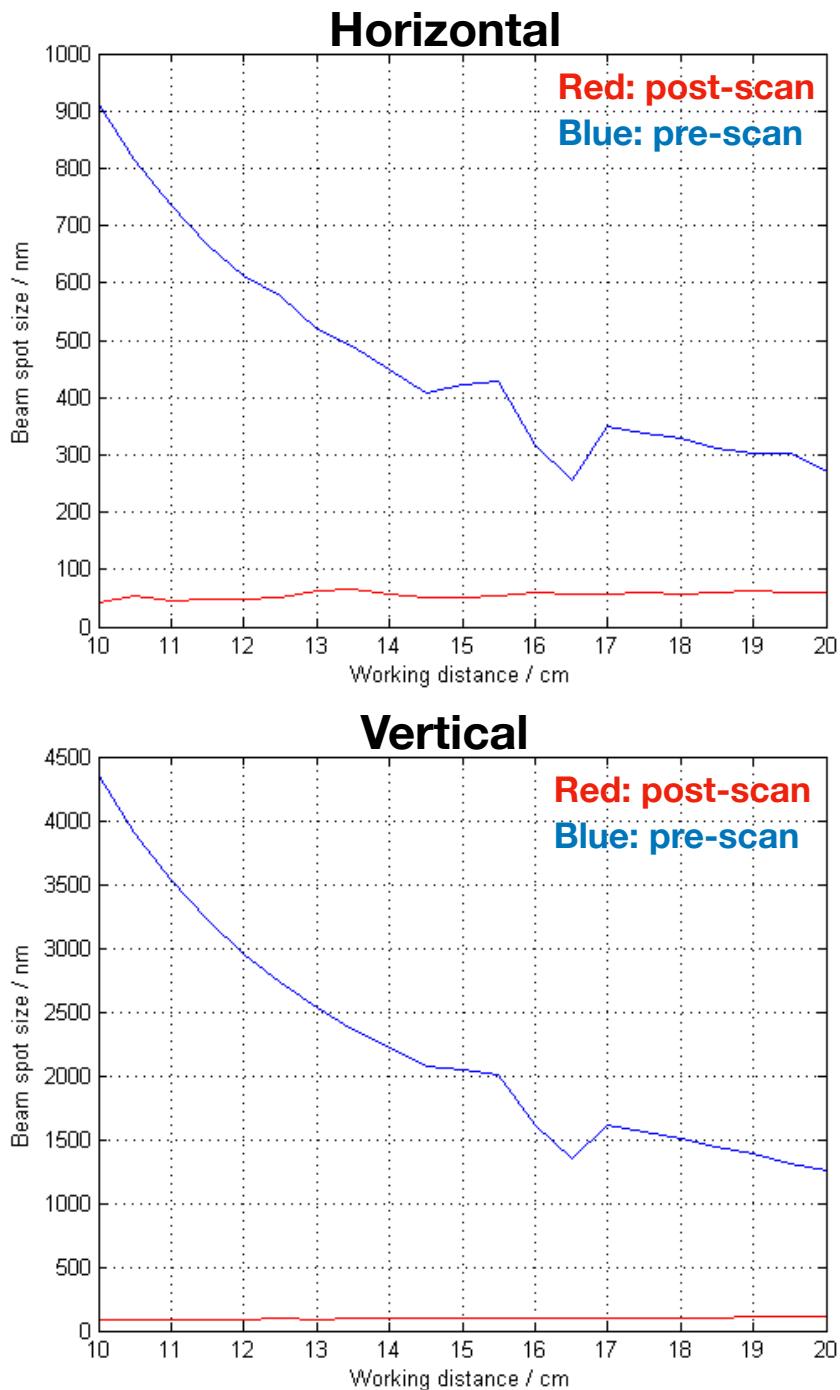


Fig. 2.5. A figure of the WINTRAX simulation result of pre-scan compared to post-scan for various working distance for a scan size of 1 mm in both directions.

zontal direction focusing and this is highly desirable during operation when focusing is performed manually. One possible disadvantage of such a high

demagnification factor is the *small depth of field*. In practice, the cell plane may not be consistent in terms of positioning in the direction of the beam, and so a small depth of field may result in very degraded beamspot size between different target locations. Overall, the Oxford triplet is preferred mainly for the superb resolution it offers, which is important for precise cell targeting.

2.2.3 Beam Scanning and Blanking

Beam scanning is an extremely important element of the beamline as it manipulates the proton to move in predefined trajectory. There are two main scanning methods - magnetic or electrostatic. Electrostatic scanning uses a voltage supplied to parallel plates to generate an electric force to deflect the beam whereas magnetic scanning uses current supplied to a coil to generate magnetic deflection. The radiobiology beamline uses the magnetic scanner and is shown in Fig. 2.2. The scanner has a *0.1 tap* option which allows the use of a smaller coil to achieve smaller scan size. The radiobiology beamline uses the OM-1015E scan controller shown in Fig. 2.6 and it accepts two inputs voltage of -10 V to 10 V for both the horizontal and vertical direction scanning. These voltages are then converted to a signal input to the *Kepco* bipolar amplifiers to produce constant current for the coils. The conversion is linear with a maximal current output of -4 A to 4 A. However, there are limitations to this scan design under certain conditions. The first limitation is magnetic hysteresis, which leads to different magnetic flux density profiles produced by the magnet under different scanning directions as powered by an AC supply [39]. This effect will lead to different beam deflections for different scan direction during STIM imaging. The second limitation arises due to bandwidth and slew constraints of the amplifier. This effect introduces time delay and non-linearity in the output current of the amplifier which can lead to a distorted STIM image.

However, this limitation can be reduced significantly by using a slower scan speed. An in-depth discussion of the scan speed effect on STIM image will be covered in Chapter 3.



Fig. 2.6. A figure of the scan controller OM-1015E and the constant current amplifier for both horizontal and vertical direction.

Beam blanking is a convenient and important feature and is required to "turn off" the beam remotely for better dose control to the cells. In CIBA, electrostatic blanking system work with a 5 V input signal into a 'fast switching amplifier'. A strong electric field is then created between two plates close to the switching magnet in Fig. 2.1 to deflect the proton beam away from the end station.

2.2.4 End Station

The end station is the main platform for performing cell-irradiation experiment and is shown in Fig. 2.7. It consists of several components which will be discussed in turn.

Vacuum chamber and Nozzle

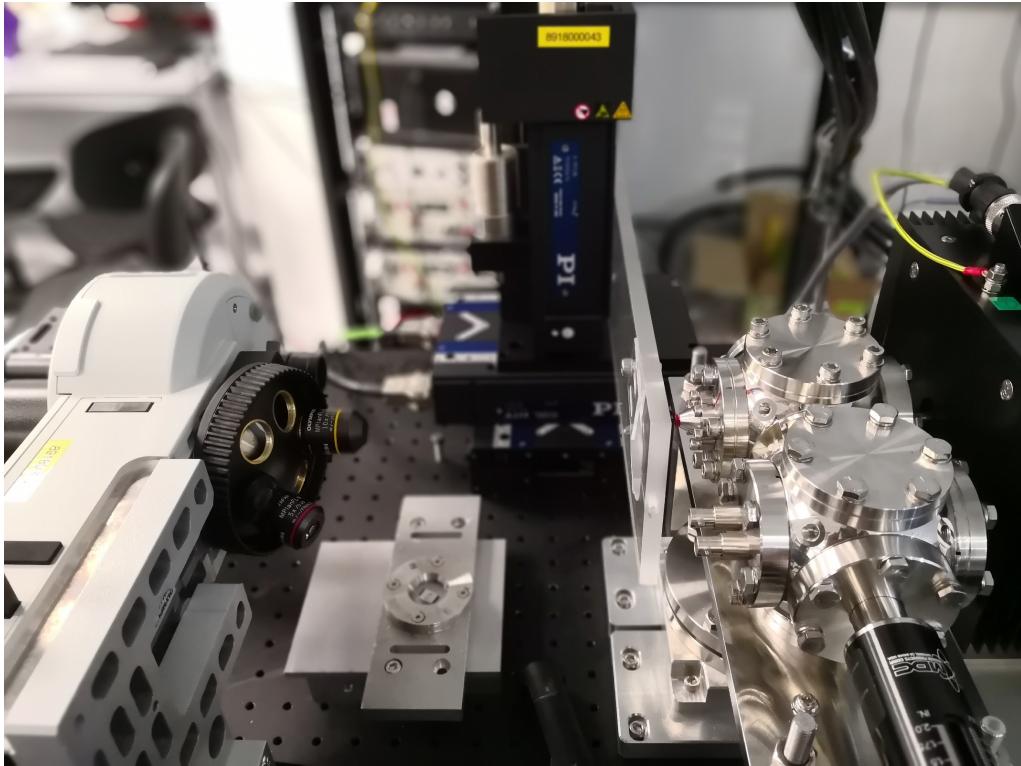


Fig. 2.7. An image of the entire end station of the radiobiology beamline.

The image of the vacuum chamber and the nozzle, which allows the beam to pass into air, is shown in Fig. 2.8. The chamber contains a Faraday cup to measure the large beam current, a PIN diode to measure small current (or number of particles) accurately and Quartz to visualize the position of the beam through the fluorescence produced from the molecular excitation by the proton beam. These three detectors are positioned using a linear and manual manipulator at the side of the chamber. The nozzle design proves to be challenging for the setting up of the radiobiology beamline. This is due to the need for a thick enough window to withstand the pressure difference between the air and vacuum, and yet maintain minimal energy loss of the proton through the window to ensure the protons will reach the cell target. A Silicon Nitride window of 100 nm thickness is used for this purpose. The nozzle has an opening size of 1 mm as shown in Fig. 2.8. A larger nozzle opening can give a larger scan size or target areas at the cells but will increase the degree of bending in the silicon nitride window

which could compromise its structural integrity. This may lead to increased probability of vacuum failure and hence is not recommended.

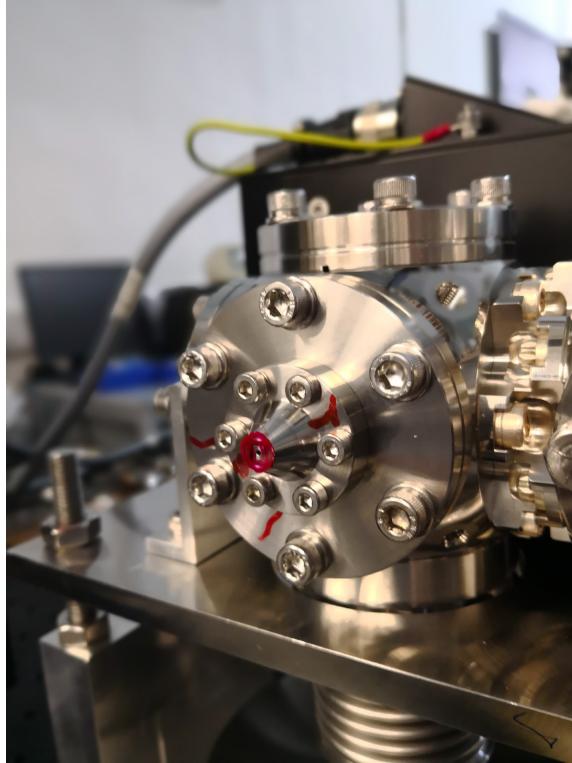


Fig. 2.8. An image of the vacuum chamber and the nozzle with a 1 mm opening.

XYZ stage

The image of the XYZ stage is shown in Fig. 2.9 and it is used to positioned the cells or cell holder precisely relative to the nozzle. It comprises of three PITM V-551 close-loop linear stages stacked on top of each other. The two stages below have a travel range of 2 inches and the topmost stage has a range of 1 inches. The bidirectional repeatability is 50 nanometers and the encoder resolution is 1 nanometre. This set-up can support a 5 kg load mounted on the topmost stage, which allows room for further development such as including an environmental chamber etc.

Optical Microscope

The optical microscope is from Olympus and is specially designed to have the objective facing the direction of the proton beam. This is shown in Fig. 2.10. It has both an LED light source which produced less heat and

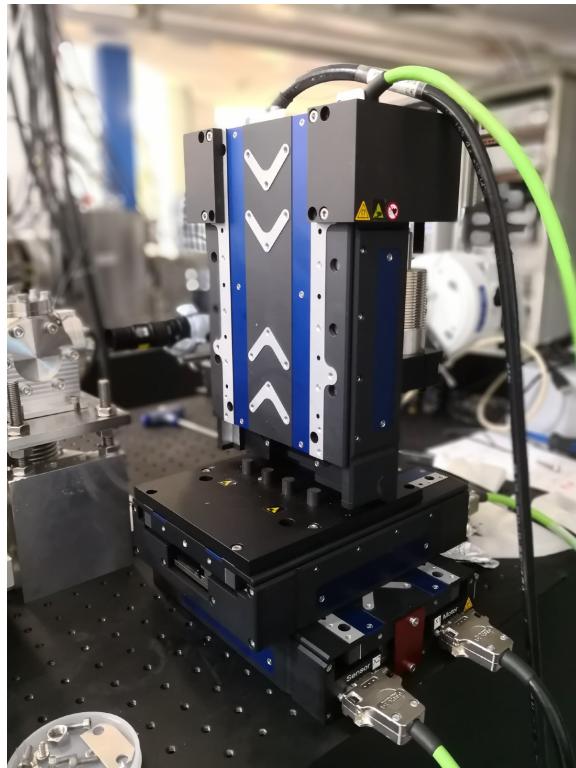


Fig. 2.9. An image of the XYZ stage used to position the cell holder relative to the nozzle.

less damaging to cells, and a metal halide lamp for fluorescence imaging. The objective lenses have two degrees of freedom - one in the focusing direction and the other in the vertical direction due to the stage beneath. The control of these two degrees of freedom can only be done manually.

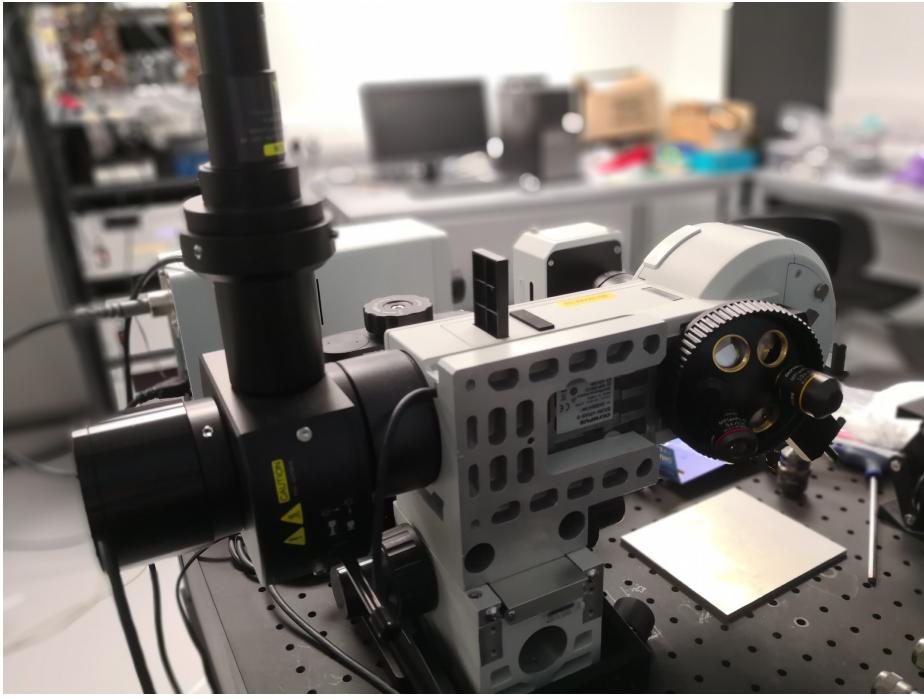


Fig. 2.10. An image of the custom optical microscope used in our beamline.

2.3 Typical Experiment Protocol

We now describe a typical protocol for performing radiobiology experiments on the new beamline. This protocol is not finalized and could be changed in the future with development of new detectors or instruments (such as using having EMCCD or diamond transmission detector). After setting up for the hardware and software aspects of the radiobiology beamline, the next step is to develop clear experimental procedures for carrying out safe and robust experiments. This procedure is also important for appreciating how our initial development efforts actually plays a part in the daily operation of the irradiation experiment. The procedure is as follows:

1. The energy and type of particles are first chosen and the beam current is optimized from the accelerator to the vacuum chamber in the end station.
2. The aim of this step is to measure and optimize the beam spot size in *vacuum*. The Quartz slide and grid are inserted into the beam's

path in turn for this step. The particle beam is focused via a two step process. The first step is *coarse focusing* where the current through the quadrupoles are varied to obtain the smallest fluorescence spot on the quartz. The second step is the *fine focusing* where the beam is focused onto a grid. With the use of STIM imaging technique, the current is varied until the sharp edge of the grid is obtained either through visual inspection or through algorithmic evaluation of the FWHM of the beamspot by fitting with an error function. Further details on the STIM imaging of grid will be discussed in chapter 3.

3. This step involves changing the focal plane of the beam from the middle of the chamber in the previous step, to the Quartz nozzle. At this stage the beam is observable as a bright fluorescent spot on the Quartz nozzle. The position of the beam is visualized through the microscope and the camera, and the position is then recorded using the image capturing software. It is important that after this step, the microscope and camera should not be moved to ensure that the recorded position in the software is accurate. After this step, the beam is *ready for use*.
4. The Quartz nozzle is then replaced with the thin Silicon Nitride window to allow the beam to reach the air. A PIN diode is placed directly outside the nozzle to ensure the beam exits the nozzle and into the air. After which, the cell sample is mounted carefully on the XYZ stage and moved slowly and precisely toward the SiN window using software control. The beam is ready for use when both the desired location at the cell sample and the SiN window are in focus under the microscope (which implies they are very close together) and possibly touching (The positioning depends on the depth of focus of the microscope objective lens).

5. The cell is then irradiated using the software developed (to be discussed in Chap.3) to control the irradiation pattern and dose.
6. The XYZ stage is moved slowly away from the nozzle and the cell sample is dismounted for further analysis. The set-up is now ready for another round of irradiation.

2.4 Summary

1. This chapter introduces the new radiobiology beamline in CIBA. The components of the radiobiology beamline are covered in details especially the magnetic quadrupole of scanning and the end station of the beamline.
2. The simulation with WINTRAX in section 2.2.2 shows that Oxford triplet will produce a smaller beamspot size (higher resolution) compared to the Russian quadruplet and therefore the triplet configuration is used to achieve greater targeting accuracy.
3. The hardwares of the end station of the new beamline has been set up and described in section 2.2.4.
4. The preliminary proposed experiment protocol is given in section 2.3.

Chapter 3

Beamline Control and DAQ system

3.1 Introduction

Data Acquisition (DAQ) and control systems are integral components of any sophisticated experimental or industrial equipments. DAQ converts physical signals in the real world into digital signals that can be processed by the computer. This allows a greatly increased level of automation and synchronization in the system. Thus, in the radiobiology beamline, a good DAQ and control system design is as equally important as the experimental design of the beamline. The DAQ system consists of both hardware and software components. The hardware includes the sensor (or detector) which converts physical energy into an analogue signal, the signal conditioning units and the ADC to convert analogue to digital signal, which can then be processed by the computer. The software consists of codes to automate the acquisition and process the signals in real-time. A well-designed GUI is important for cases of multiple functions for acquiring and processing the data and for easy operation by inexperienced users.

This chapter of the thesis is focused on the development of the hard-

ware and software of the DAQ system in the radiobiology beamline. The overview of the DAQ system is shown in Fig. 3.1. It shows the flows of signals between various hardware components and it also lists down the required functions of the software used for controlling the DAQ. This chapter will first introduce the DAQ hardware including the detectors, signal conditioning unit and most importantly the NI-FPGA card which is the main controller. Then, the software aspects will be covered in details including the algorithms and GUI for various functions. This chapter also includes experimental measurements and results to prove that the DAQ hardware and software are functional.

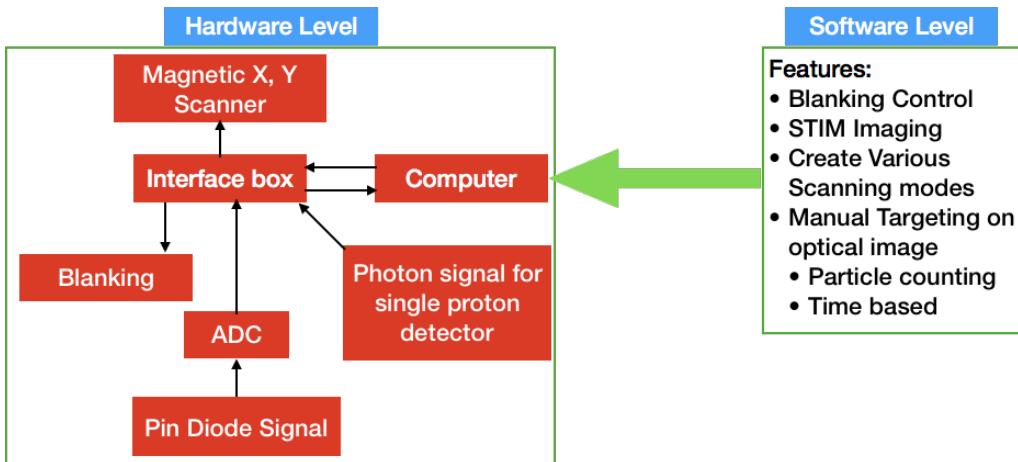


Fig. 3.1. An overview of the DAQ system design. It shows the relationship between various hardwares and the software. The arrow refers to the flow of input signal.

3.2 Hardware

3.2.1 Overview of the set-up

The image of the hardware of the DAQ system is shown in Fig. 3.2. The signal conditioning units consisting of a spectroscopy amplifier and the ADC are mounted on the NIM (Nuclear Instrument Module) rack, which is the oldest and simplest nuclear electronics standard. The CPU contains the FPGA card for acquiring data and producing analogue output voltage

to the scanner and blanking amplifiers. The interface box contains break-out board for the FPGA card which is connected to the BNC connectors which is in turn connected to the X and Y scanner, and the blanking amplifiers. It also interfaced with the 25-pin D-Sub-Connector from the ADC to accept direct input data from it. The magnetic scanner controller and the Kepco amplifiers have been discussed in Section 2.2.3.

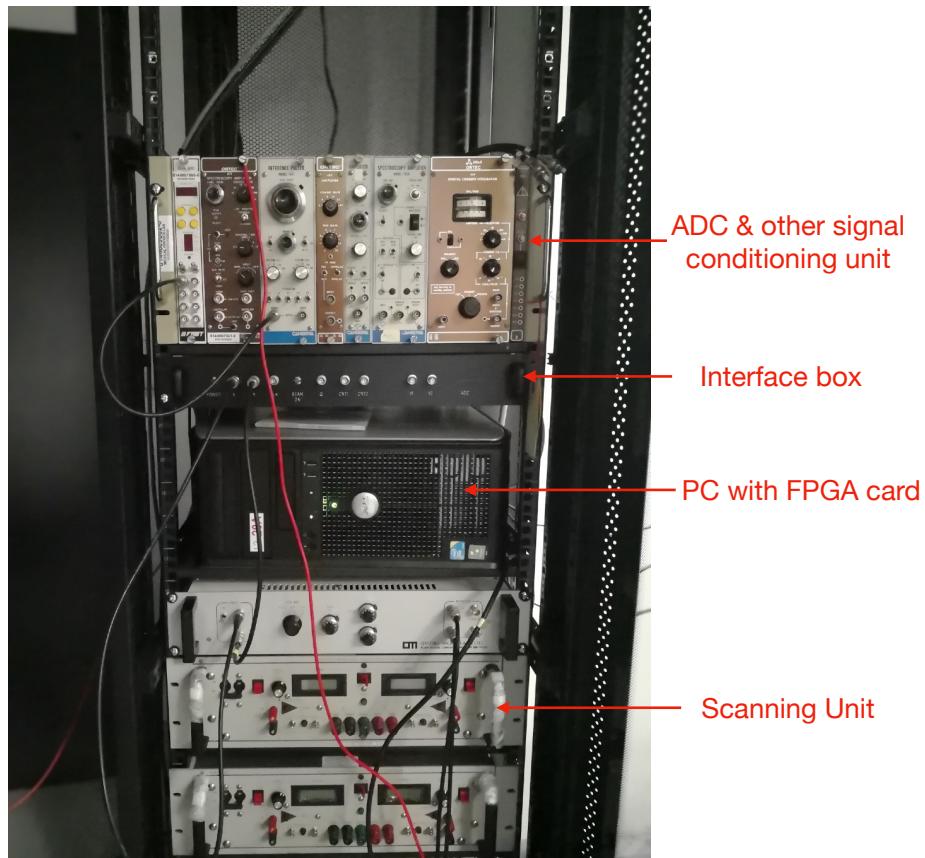


Fig. 3.2. An image of the hardware components of the DAQ system of the radiobiology beam line.

3.2.2 Detector

Detectors are devices that directly converts physical energy into analogue signals. In our case, there are two types of detectors that are interfaced with the DAQ system - PIN diodes and photo-multiplier tubes (PMT). The PIN

diode is a semiconductor device which can be used for *measuring the energy* of the proton and is shown in Fig. 3.3. The proton beam deposits energy in the detector to create electron-hole pairs [40, 41] which will drift in opposite directions under the electric field produced by an applied bias voltage. This in turn results in an induced current by the Shockley-Ramo's theorem [42], thus producing the desired analogue signal for further processing. This induced current is then integrated using a charge sensitive preamplifier [41] to yield the total induced charge which is proportional to the energy deposited by the proton beam. The preamplifier will be discussed further in Chapter 3.2.3.

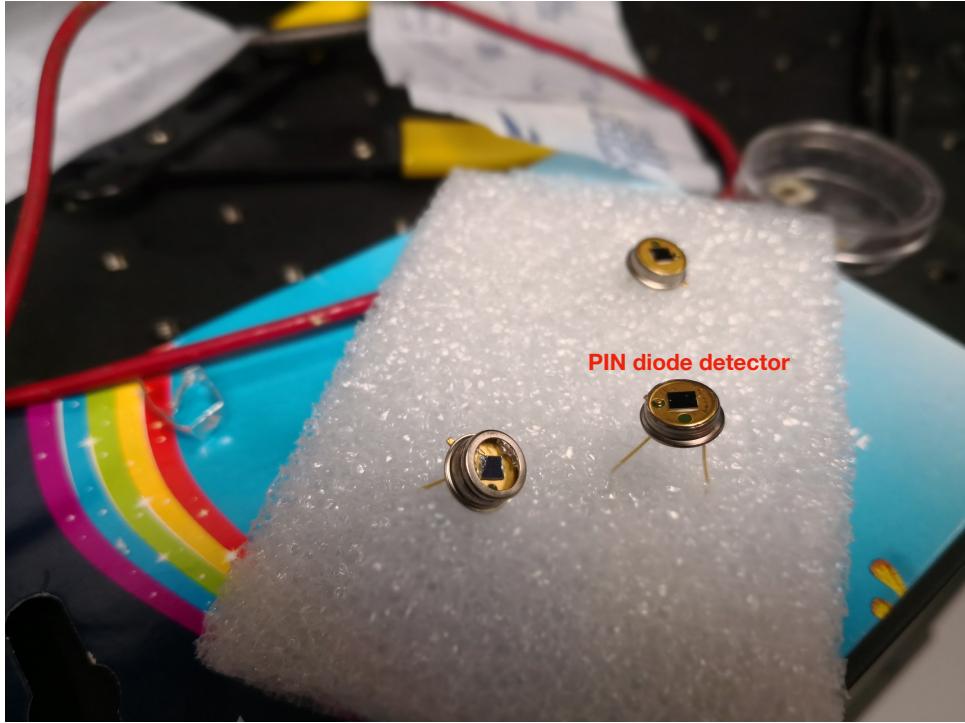


Fig. 3.3. An image of the PIN diode for measuring energy of the proton in the radiobiology beam line.

3.2.3 Signal Conditioning Unit and ADC

The direct output signal from the PIN diode needs to be conditioned in two stages before it can be fed into the ADC. The first stage involves the Charge Sensitive Preamplifier (ORTEC 142A model) and the second stage

involves a spectroscopy amplifier (ORTEC 672 model). The ADC used in this DAQ system is a dual ADC 7072 by fast COMTEC.

Charge Sensitive Preamplifier

The Charge Sensitive Preamplifier (CSP) is essentially an integrator circuit with a capacitor, C_f connected in parallel to the op-amp [43] shown in Fig. 3.4, where C_d is the detector's capacitance and A is the open-loop gain. The voltage output is proportional to the total charge collected and the output voltage (v_o) per unit input charge (Q_i) is

$$\frac{v_o}{Q_i} = \frac{Av_i}{C_f v_f} = \frac{Av_i}{C_f(A+1)v_i} \approx \frac{1}{C_f}. \quad (3.1)$$

Thus, the feedback capacitance of the preamplifier controls the *voltage gain* of the detector. The above result follows from an ideal op-amp and considering realistic op-amp with finite Gain Bandwidth Product (GBWP), the charge $Q(t)$ measured by the preamplifier given an initial signal charge of Q_s is

$$Q(t) \approx \frac{Q_s}{1 + \frac{C_d}{C_f(A+1)}} \left(1 - \exp\left(-\frac{t}{\tau}\right) \right), \quad (3.2)$$

where τ is defined by

$$\tau = \frac{1}{\omega_0 C_f} C_d, \quad (3.3)$$

and ω_0 is the GBWP. The result in Eq. (3.2) arises from the fact that for an inverting amplifier, the input impedance of a CSP is capacitative for low frequency and resistive at high frequency [41, 43]. Two interesting observations can be made from Eq. (3.2) to (3.3). First, the total measurable charge from the CSP is determined by $C_d/(A+1)C_f$ factor. A high open-loop gain, A or C_f/C_d will result in a charge measurement equal to Q_s . Second, the rise time of the output signal of CSP is determined by τ . A higher C_f/C_d factor and GBWP will lead to a smaller rise time which is

a measure of the response time of the CSP. However, a higher C_f will lead to a smaller output voltage per input charge as in Eq. (3.1) which affects the sensitivity. There is thus, a delicate balance in the choice of C_f .

The ORTEC 142A preamplifier is suitable for our detector with C_d between 0 to 100 pF and has a charge sensitivity of 45 mV/MeV in Silicon detector. The rise time varies from 5 ns (for 0 pF detector) to 12 ns (for 100 pF detector) and the decay time is 500 μ s. The open loop gain is more than 40000.

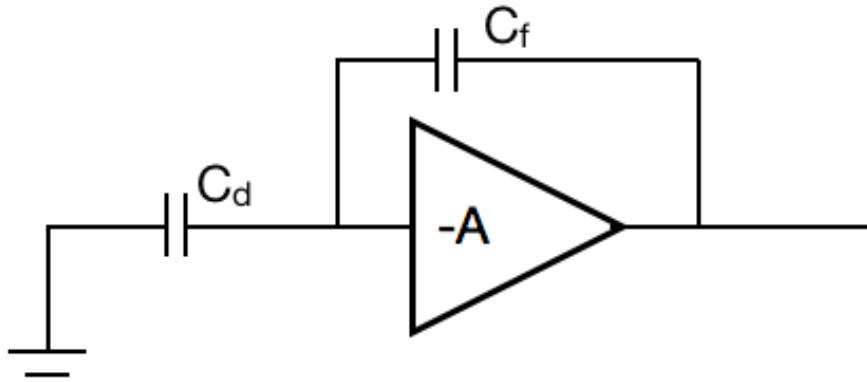


Fig. 3.4. A circuit diagram of a typical charge sensitive preamplifier

Spectroscopy Amplifier

The spectroscopy amplifier accepts the output from the CSP to achieve the followings [41]:

1. To amplify the signal from the preamplifier from the millivolt range to between 0.1 to 10 V range. This is to ensure the signal is large enough to be digitized by the ADC to give sufficient dynamic range for further analysis.
2. To shape the pulse to optimize energy resolution in order to achieve better pulse height approximation and to lower the probability of overlap between consecutive pulses (the output from preamplifier has a decay time of 500 μ s) to accommodate higher count rates.
3. There is also a *baseline restorer* in the spectroscopy amplifier to en-

sure the baseline between pulses are kept firmly at ground potential during changes in count rates or temperature. This step is crucial for high count rates application where the baseline maybe altered due to an AC-coupled amplification effects. It is important to minimise any baseline shift as it will subsequently affect the pulse height of the signals being processed.

The ORTEC 672 spectroscopy amplifier offers a voltage gain from 2.5 to 1500 and output either a unipolar triangular, unipolar gaussian or a bipolar pulse. The shaping time can be changed from $0.5 \mu s$ to $10 \mu s$. Unipolar gaussian pulse shaping is usually recommended for energy spectroscopy application. A bipolar pulse is useful for application involving measuring zero-crossing or minimizing baseline shift if there is any ac-coupled circuitry after this amplifier. However, bipolar output suffers from lower SNR and longer pulse duration issues, thus making it non-ideal for energy spectroscopy especially in a high count rate setting.

Analogue to Digital Conversion (ADC)

After processing by the spectroscopy amplifier, the signal pulse height is proportional to the energy of the proton particle, and the pulse will then be processed by the ADC. The ADC in Pulse Height Analyzer mode will digitize the pulse heights according to the resolution of the ADC (from 256 to 8192 in binary steps) and output through a 13 bit binary data line to the controller. These pulse height data of the protons are collected to yield the *energy spectrum* information. Our ADC has a fixed conversion time of 500 ns and a bandwidth of 3.3 MHz. The conversion time is related to the latency of the ADC and the bandwidth affects the degree of distortion of the incoming signal. The ADC trigger can be adjusted using the Lower Limit Discriminator (LLD) which in our case ranges from 78 mV to 1 V. Two other important specifications of the ADC are the *differential nonlinearity* and *integral nonlinearity*. The first specification measures

the non-uniformity of the widths of the voltage channels and the second measures the non-uniformity of the position of the conversion point within a channel. The 7072 model has a differential nonlinearity of 0.7% and an integral nonlinearity of 0.16% at 8192 resolution.

3.2.4 FPGA

There are several approaches that can be used in developing the control system for the DAQ in nuclear instrumentation. These include using microcontroller[44], DSP[45], CPU[46, 47, 48], GPU or FPGA[49, 50]. Our radiobiology beamline uses the PCI-7831R FPGA card from NI as the choice of controller for the DAQ system. A FPGA card is an array of *configurable logic blocks* which can be configured to perform combinatorial or sequential functions [51, 52] using software programming tools. The traditional programming tools such as *verilog* or VHDL [52] are highly complex to use for applications and are comparable to assembly language. However, the FPGA from NI breaks this tradition by allowing programming in Labview which is a graphical programming language (or "G" language). This allows faster development time and a milder learning curve for using FPGA.

The motivations for using a FPGA are the followings:

1. Precise timing control down to the hardware level governed directly by the clock rates up to tens of MHz. This is important for precise hardware trigger and maintaining synchronicity between the scanning (X,Y) voltage and the data acquisition from the detector. Also, the high clock rate in MHz implies faster processing capability not achievable in microcontroller or DAQmx.
2. It has the flexibility of a software due to the *ease* of reprogramming the logic blocks, but yet possesses the low latency characteristics con-

sistent with hardware systems. This is important for future upgrades in the functionality of the DAQ system.

3. FPGA is a truly robust and reliable system. Unlike processor-based solution which have many abstraction layers to help in tasking scheduling and resource management (such as driver layer for hardware control and OS for memory and processor control), FPGA is capable of completing tasks *deterministically* and achieving *high levels of parallelism* [52].

There are other benefits of using FPGA involving security and backward compatibility but are more applicable to commercial application or deployment. The PCI-7831R card uses a Virtex-II V1000 FPGA chip consisting of 10240 flip-flops and 720 kbits embedded block RAM. There are 8 analog inputs with independent sampling rates up to 200 kHz (5 μ s conversion time) and 16-bit resolutions from -10 V to 10V. There are 8 analog outputs with independent update rates up to 1 MHz (1 μ s conversion time) and 16-bit resolution from -10 V to 10 V. Lastly, there are 96 digital I/O lines at a rate up to 40 MHz. The ADC in the FPGA is much slower than the one from fast COMTEC, thus the decision to perform the ADC using fast COMTEC module and transfer the 13-bit digital signals into the *fast* digital input lines of the FPGA. This will minimize data loss from the spectroscopy amplifier to the FPGA. The (X,Y) scanning voltage are outputed using 2 analog output channels into OM-1015E. The 5V blanking voltage, necessary to switch the beam off and on, is outputed through the third analog output channel. This connects to the switching amplifier feeding the blanking plates, which are situated upstream in the beam line.

The PCI-7831R FPGA card can only be programmed using 32 bit Labview program [53]. In addition, it does not support any floating point data structure which precludes operation such as division. Thus, algorithms need to be designed to fit in with these constraints.

3.3 Software

3.3.1 Planning

After choosing and finalizing the DAQ hardware, the next step is to develop the software for controlling the hardware. The design strategy of the software follows the well-known saying: "Measure twice, Cut once". Prior to programming the software, a well thought-out design strategy will go a long way in reducing the re-coding iterations and therefore save a significant amount of time. There are three considerations [54] to be tackled before embarking on the construction phase - Requirements, Architecture and Programming Language.

Requirements

The crafting of the requirements is the most important part of software development. It defines the functions and architecture of the software.

The software requirements at the user and system level is as follows:

1. The software must provide an *integrated platform* with user friendly GUI to perform various tasks involving the operation of the beamline.
2. The software must perform basic functions such as measuring the ion *count rate* and *energy* through the ADC, and also control the *blanking* of the beam.
3. The first key function of the software is to provide an interface and system capability for executing *STIM imaging* and its visualization under different scan settings. This is important not only for *measuring beam resolutions* but also for implementing accurate damage patterns to the biological cells under investigation.
4. The second key function is the ability to deflect the proton beam to a designated position (with cursor control) using an image captured

by an optical microscope. To achieve this, there must also be a *calibration* phase to map the position on the image to the (X,Y) voltage. This is the most vital requirement as it allows the targeting of cells with the proton beam.

5. There is also a diagnostic function to test the fidelity of single proton detection based on single proton fluorescence yields. The idea is to test the hypothesis that a single proton can produce either a single or multiple photons that can be detected using a suitably positioned photomultiplier tube (PMT). The PMT signal is inputted through the Digital Channel in the FPGA, which can process much faster than the Analog Input channel.

Architecture

The software architecture contains more details and design principles regarding the software. The program organization is shown in Fig. 3.5. It shows that the software is composed of two layers - one at the FPGA level and one at the higher level for creating the GUI and calling the FPGA reference. The key functions in each layer are also shown in Fig. 3.5 together with their relations to each other. The transfer of data from the FPGA to the upper layer is done through DMA (Direct Memory Access) using a FIFO buffer and the transfer only happens when the *pushData* input is True. The data which is pushed to the upper layer include the (X,Y) scan coordinates and the *energy* from the ADC which is crucial for constructing the STIM image. There are three main event loops in the upper layer running in parallel and independently. As the FPGA reference can only be called once at any one time, all the sub-VI which involves calling FPGA reference are placed in a single event structure to prevent calling the reference twice concurrently. The rest of the sub-VI including *dialogs* are placed in a different event structure.

With regards to *Error processing*, we decided to employ a *detective* and *active* approach. *Detective* means the software will detect an error and continue running or quit instead of correcting it. *Active* means the software will check for validity of user inputs instead of passively responding to the error arising from invalid inputs.

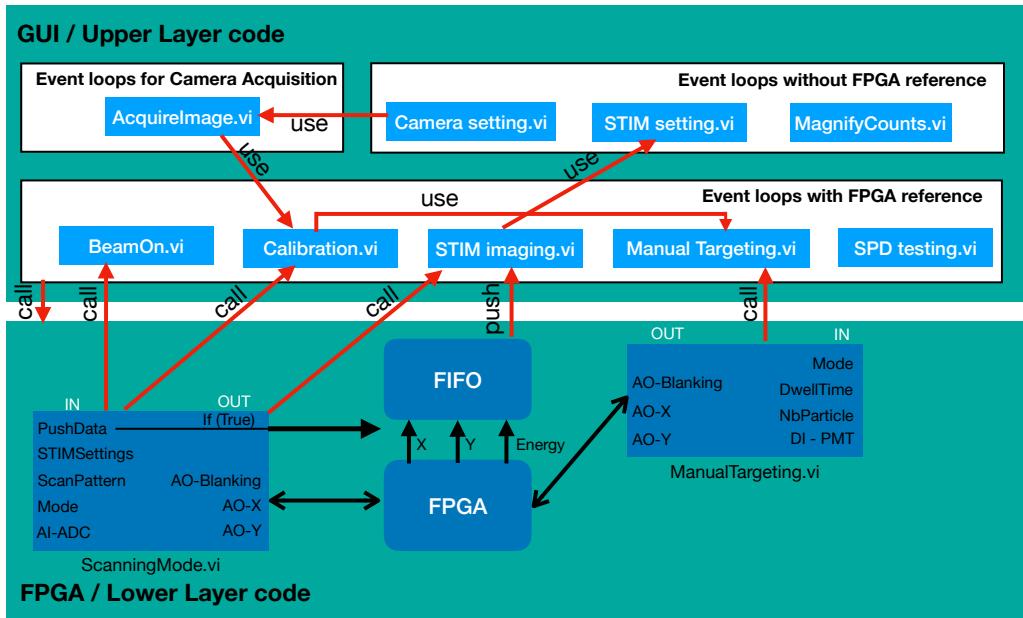


Fig. 3.5. The programme organization of the DAQ software consisting of the FPGA level code and the GUI level code

Programming Language

The programming of the FPGA is done in 32 bit Labview language and the code is then compiled in the FPGA into *bitfiles*. This constitutes the coding at the lower level which directly controls the FPGA. The choice of the programming language of the higher level code which controls the GUI and run instances of the FPGA code is the main subject of this section. The fact that the FPGA codes can be exported as *.dll* files imply that common frameworks such as Qt or .Net can be used for creating the GUI and interfacing with the FPGA. As the two frameworks require regular

maintenance to ensure backward compatibility either due to OS updates or obsolete class libraries, we decided to develop the GUI with 64 bit Labview. A Labview-based software offers *convenience* of editing the codes (without compilation) due to changing operational requirements and it offers straightforward tools for implementing standard design patterns such as *producer-consumer*, *master-slave*, *event-based user interface* and *state machine*. The labview code can also be compiled into single executable machine code that can be run with the help of the labview run-time engine (similar to MATLAB). This makes the labview code portable across platforms. The two disadvantages of using a labview-based software are the slow speed and the aesthetically displeasing interface. However, the first disadvantage can be overcome by making use of appropriate software design patterns.

3.3.2 STIM Imaging

Interface

The interface of STIM imaging is shown in Fig. 3.6. In STIM imaging the proton beam is scanned in a predefined manner, and the energies and number of protons measured in each pixel position (X,Y) is processed. If the energy of the proton falls within the user defined *energy gate*, the count in that pixel will be accumulated and subsequently visualized using the interface. Due to DMA, the reading of data from FIFO is fast and could handle count rates up to 20000 s^{-1} . However, the updating of 2D image arrays is slow and can only keep up for count rate up to 3000 s^{-1} . Thus, updating the image sequentially after *popping* the data from the FPGA FIFO buffer is not very efficient. To circumvent this problem, the *producer-consumer* architecture is used to handle two processes with different speed. Each process is placed in each *parallel* loop and the data from the buffer (or producer) is placed in a *queue*, which is subsequently dequeued in the

other loop (or consumer) for an image update.

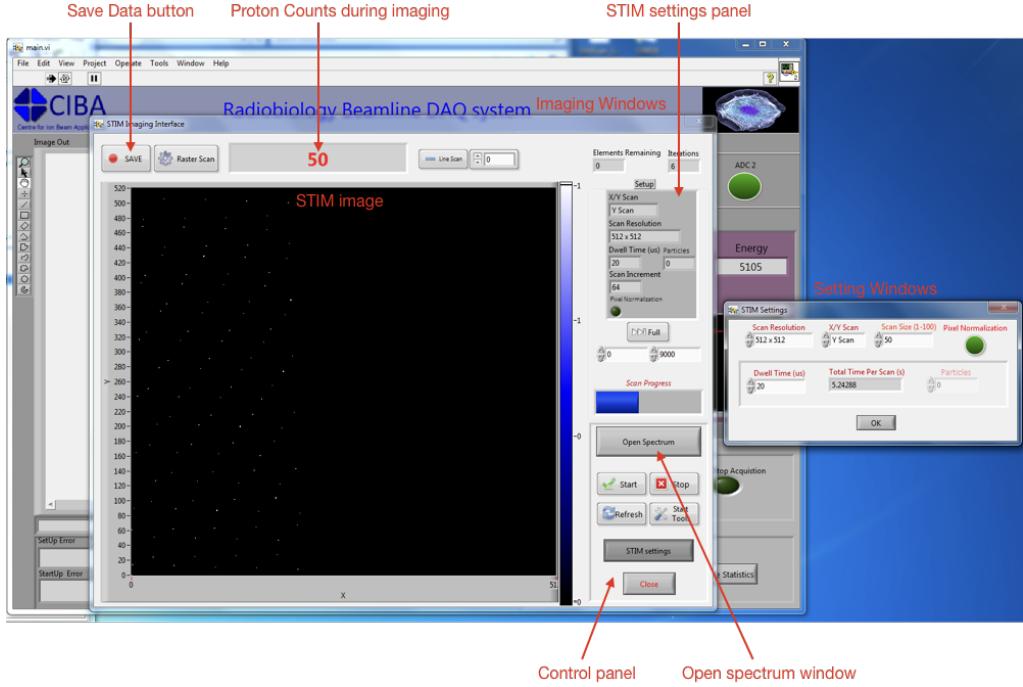


Fig. 3.6. A screenshot of the STIM imaging interface. The words in red refers to the description of the various windows and buttons. The setting windows set the parameters for acquiring STIM image.

The settings of STIM imaging interface consists of:

1. **Scan Resolution:** This defines the number of pixels in the image in both the X and Y direction (equal in both directions). The software offers resolutions from 128 to 2048 in binary steps.
2. **Scan Size:** This defines the scan size of the imaging and assumes values between 0 to 100. A value of 100 refers to maximal scan size with output voltage from the FPGA from -10 V to 10 V, and any value in between decreases the output voltage *linearly*. For instance, a value of 50 refers to a scanning voltage between -5 V and 5V. In principle, the scan size can be set to be arbitrarily small, but in practice, this is limited by the DAC resolution of the FPGA. The smallest scan size will satisfy this equation:

$$2 \times 32768 \times \frac{\text{ScanSize}}{100} \times \frac{1}{\text{ScanResolution}} = 1. \quad (3.4)$$

This equation arises from the constraint that the minimal increment in scan voltage is $(1/2^{15} \times 10)$ V.

3. **Pixel Normalization:** There are two methods of determining the time spent on a single pixel. The first method is based on fixed dwell time and the second method is based on fixed number of particles collected at each pixel. The second method is known as *pixel normalization*.
4. **Dwell Time:** If the Pixel Normalization option is *off*, the STIM imaging will spend a fixed amount of time on each pixel. This fixed amount of time is user-defined and is specified by the *Dwell Time* settings in microseconds. It is important to note that the exact time spent on each pixel is in practice larger than the *Dwell Time* due to the conversion time of the DAC mentioned in Section 3.2.4. This amounts to an actual dwell time of 1 microsecond in addition to the user-defined *Dwell Time*. Also, a small dwell should not be used in practice due to a range of artefacts it introduced during the imaging process. Fig. 3.7 shows the comparison of the trace from the FPGA output and the current monitor of the amplifier (in the *long scan* direction) for short dwell time of 1 μs and long dwell time of 10 μs (every other settings are kept constant). It became evident that with short dwell times, it introduced non-linearity and non-negligible delay (about 300 microseconds) in the output current. These effects diminished at longer dwell time such that the current monitor graph coincides completely with the FPGA output voltage in Fig. 3.7. The non-linearity will result in a wrong scaling between the user defined *Scan Size* and the actual scan area, whereas the time delay results in pixel off-set in images for different scan directions. The pixel off-set is

$$Offset = 2 \times \frac{DwellTime}{TimeDelay}. \quad (3.5)$$

Thus, the parasitic effect from time delay can be eliminated by using a longer dwell time. These effects arise due to the finite slew rate and bandwidth of the amplifier. Thus, these settings must be optimised by the user so as to operate well beyond the constraints of the hardware.

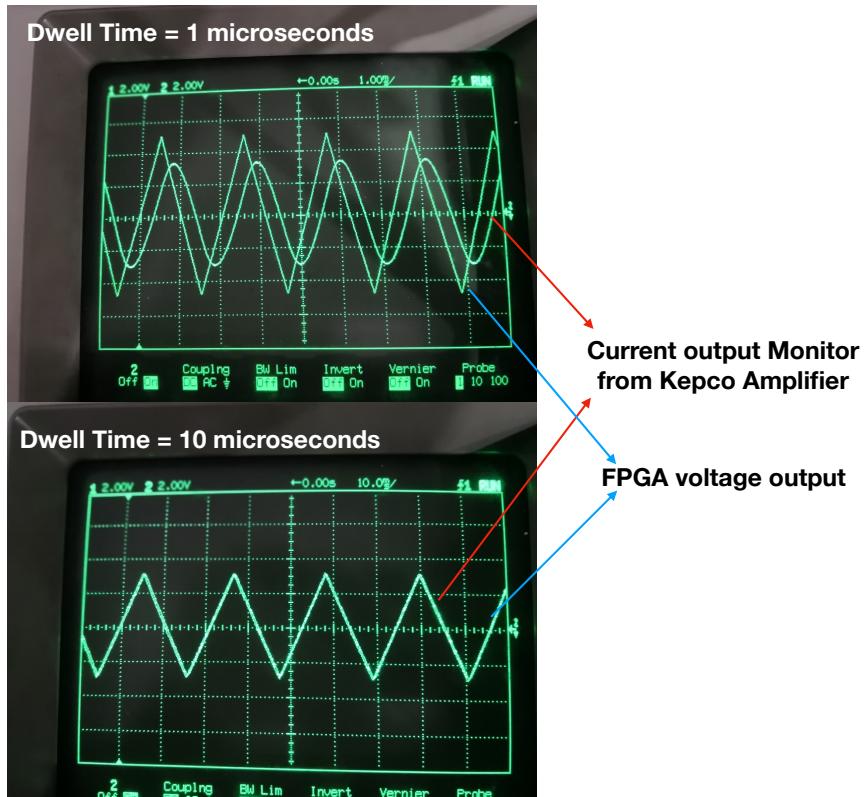


Fig. 3.7. The comparison of the Kepco current output and FPGA output trace for short dwell time of 1 microsecond (top) and long dwell time of 10 microseconds (bottom). The delay between the amplifier and FPGA output is non-negligible for short dwell time.

5. Number of Particles : If the Pixel Normalization option is *on*, the STIM imaging scanning system will spend the required time on each pixel until a fixed number of particle is detected by the ADC. This fixed number of particles is user-defined and must be an integer.

6. Line Scan: This is a special mode of scanning the beam with non-

equal X and Y resolution using user defined spacing between the *slow scan* direction. This function is mainly for implementing line irradiation of the cells.

7. Raster Scan: This affects the scan direction of STIM imaging. If the option is *on*, the scanning will only be done in a *single* direction. Otherwise, the scan direction will alternate in a *triangular wave-like* manner. The Raster Scan option is for circumventing possible hysteresis effects of the magnetic scanning coils at high scan rates.

The settings of STIM imaging can be set via the interface shown in Fig. 3.6.

Saving Data

Saving the STIM image data is important for subsequent post-processing. The software is able to save *all* the information of the particle that comes in, into a binary format known as the *list mode* [55]. Each of the (X,Y) and energy information of the particle is converted into a *unique* U64 binary data format using the following formula:

$$data = X + R \times Y + R^2 \times Energy, \quad (3.6)$$

where R is the scanning resolution. The *data* variable is then converted to U64 binary data in the little endian byte ordering format [52] and saved (or appended) to a file in the hard disk. Eq. (3.6) together with the data format and byte ordering information are all that is required to back-determine the unique (X,Y) and energy value of each proton.

STIM Imaging of Gold Grid

In this section, in order to measure the beam resolution, the STIM image of the gold grid in *air* were collected using the software protocols described above. The image in Fig. 3.9 is collected using the pixel normalization op-

tion with 9 particles in each pixel and a 512 x 512 image resolution. Each proton is detected by a silicon surface barrier detector, and the resultant energy profile for all protons is shown in Fig. 3.8. The energy is measured using a 8192 resolution ADC. The rightmost peak represents the high energy protons which have passed through the hole in the grid, whereas the leftmost lower energy peak represents the protons that have lost energy by passing through the grid bar. Such a spectrum, containing two independent peaks, is typical for a grid target.

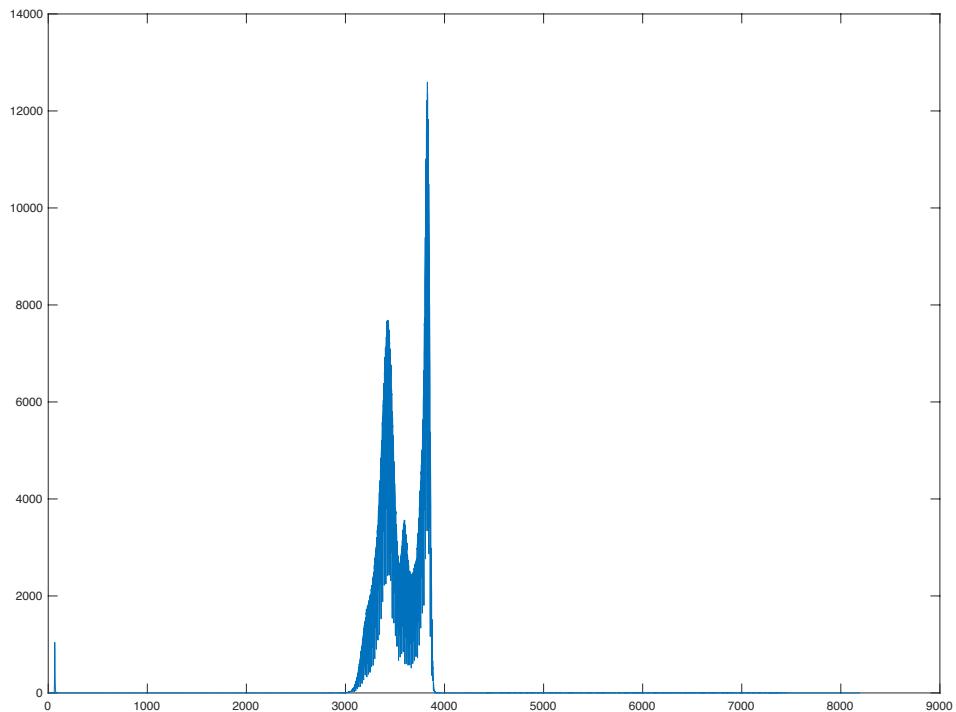


Fig. 3.8. Energy spectrum of the proton beam measured by the Silicon barrier detector through the gold grid in air. The rightmost high energy peak originates from proton travelling through empty space of the grid and the left most low energy peak originates from proton travelling through the bar of the grid.

There are various ways to visualize the STIM image. In this case, the STIM image is constructed by calculating the median energy in each pixel and is shown in Fig. 3.9. The image of the grid is clearly shown in the figure. The beam resolution in the X and Y directions can be determined by fitting an *error function* [56] to the vertical and horizontal edge profiles, and

calculation of the FWHM can be subsequently achieved. This procedure is shown in Fig. 3.10. In this particular case, the calculated FWHM in the X and Y directions from the error functions are $1.02 \mu m$ and $840 nm$ respectively. This is an example of how beam resolutions can be estimated from the STIM imaging data using the relevant software.

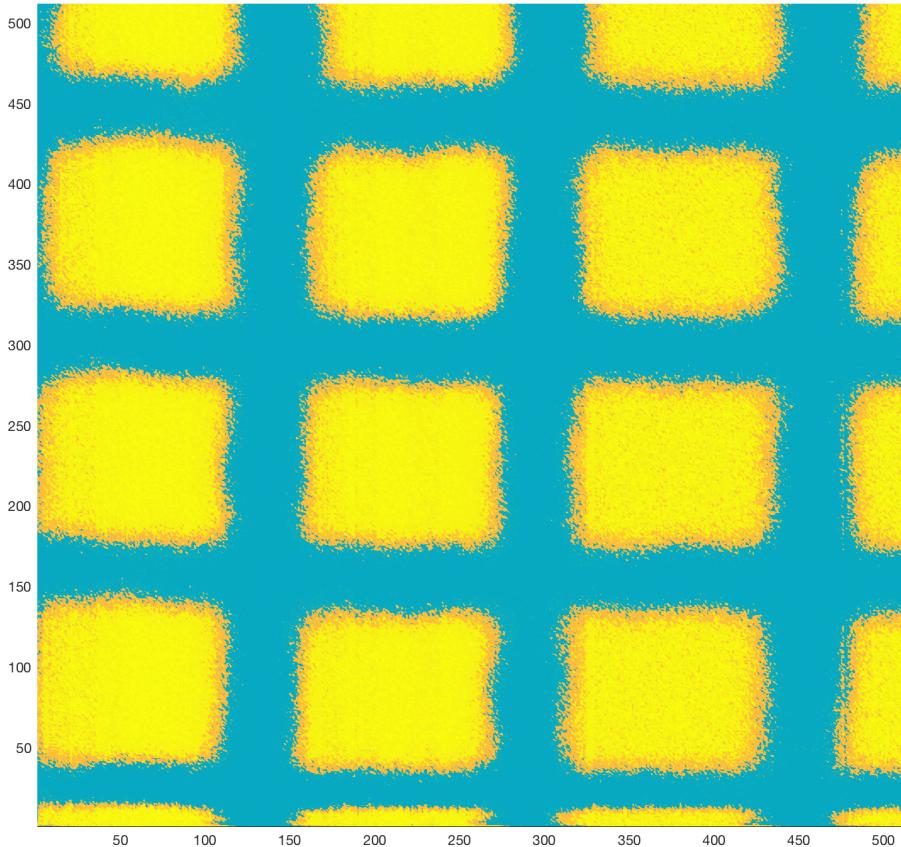


Fig. 3.9. The STIM image of the gold grid by extracting the median energy in each pixel. Yellow color refers to high median energy and green refers to low median energy. The grid pattern is evident through STIM imaging.

STIM Imaging for Cellular Targeting

As mentioned previously, as well as STIM imaging, this interface can be used for inducing periodic damage to the cells. In this experiment, we used our control system to irradiate HepG2 cells with 2 MeV proton beam (in vacuum). The cells' nucleus were stained with DAPI (in blue) in order to visualise the cell nuclei [57] and also stained using γ -H2AX (in red) [58] for visualizing DNA damage sites. The γ -H2AX indicates DNA double strand

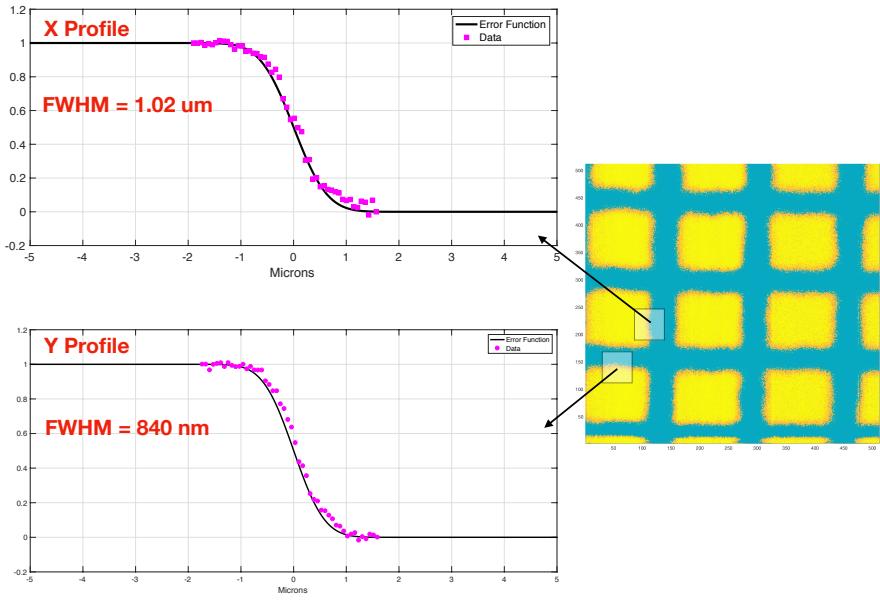


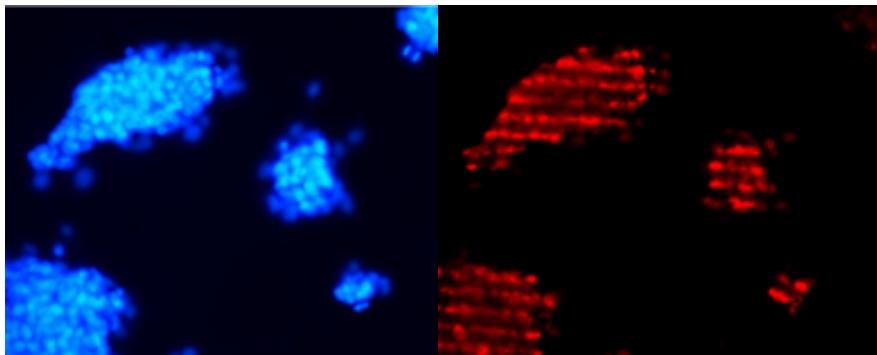
Fig. 3.10. The extraction of the edge profile in both the X and Y direction together with the best fit error function to determine the resolution in the horizontal and vertical direction respectively.

breaks (DSBs) caused by proton damage, although false positives can still arise from various sources [59, 60, 61]. The irradiation dose is such that each pixel has on average *100 protons* passing through it, and the *Scan Size* was chosen such that a $700 \mu m$ area is irradiated. The *Line Scan* option was used to irradiate a cluster of cells shown in the top figure of Fig. 3.11 with a pixel scan resolution of 512×32 . To induce a point-like damage foci as shown in the bottom figure of Fig. 3.11, we used conventional scanning with a resolution of 128×128 so that the damage foci are spaced far enough apart and no overlapping occurs. Overall, this shows the success and utility of the interface in inducing a periodic damage to cells in terms of a line or an array scan.

3.3.3 Scanning-imaging calibration

Carrying out periodic damage to cells using line or array scanning is however limited. Ideally, the software should also be able to carry out targeted

Line Scan on clusters of HepG2 cells



Array Scan on a monolayer of HepG2 cells

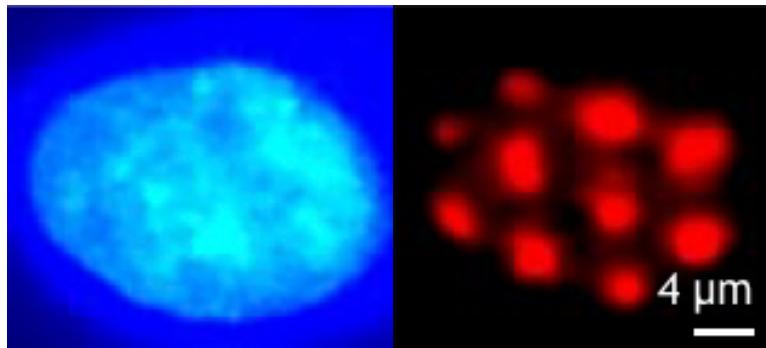


Fig. 3.11. The fluorescence image of the cell nucleus together with the damage regions in both line (top) and array scans (bottom). The blue fluorescence image correspond to DAPI label for the nucleus and the red image correspond to γ -H2AX label for the DNA damage sites. The damage pattern can be clearly seen in the nucleus.

irradiation to each cell. This means that the proton beam should be deflected to a position indicated by the user from the optical image of the cell target. Such an operation requires a *mapping* from the (X,Y) pixel position on the optical image to the (X,Y) voltage input to the magnetic scanner. This section is dedicated to explain the algorithm for establishing the map and calibration *automatically*.

The calibration process involves observing the proton induced fluorescent beamspot on a Quartz located at the (or close to) plane of the cell target. The beamspot, which tends to have a circular form, can be detected automatically using Hough Transform [62]. The (X,Y) voltage of the magnetic scanner is then varied and the new (X,Y) position of the beamspot in the image is recorded. This process is repeated several times to establish

the relations between the (X,Y) voltage and the beam positions either using linear regression (if one assumes or observes the relation to be linear) [63] or interpolation routines [64]. The pseudocode for the algorithm of calibration is shown below.

```

Require: N = Number of Iterations
Require: dV = Voltage Step
//Initialization of zeroes array
xPosition, yPosition <- zeroes(N)
xVoltage, yVoltage <- zeroes(N)
//Main Algorithm
for k from 1 to N do
    OutputVoltage(xVoltage[k], yVoltage[k])
    wait
    xPosition[k], yPosition[k] = HoughTransform(Image)
    xVoltage[k+1] <- xVoltage[k-1] + dV
    yVoltage[k+1] <- yVoltage[k-1] + dV
end for
xMap = Regression(xVoltage, xPosition)
yMap = Regression(yVoltage, yPosition)

```

The interface for the calibration process is shown in Fig. 3.12. The only user-input in this interface is the *voltage increment* indicated by dV in the pseudocode above. The left panel shows the *live* camera view of the beamspot and the circle detection via Hough Transform. The right panel shows the plot of the pixel coordinates of the beamspot against the applied voltage to the scanner. These two windows are important for monitoring the progress of the calibration. Currently, we assume the relation between the position and voltage to be linear, which is one reason why we opted for post lens scanning (ie to reduce off axis aberrations). Thus, the algorithm

above amounts to finding the a and b parameters in this equation $Voltage = a \times Position + b$. There are a total of 4 parameters (2 for each direction) to be determined and it is important to note that this process needs to be repeated if the microscope or cells are moved relative to each other. With each voltage increment, the parameters values will be updated and once the regression line and parameters stabilize, the user can *stop* the calibration process and proceed to targeting.

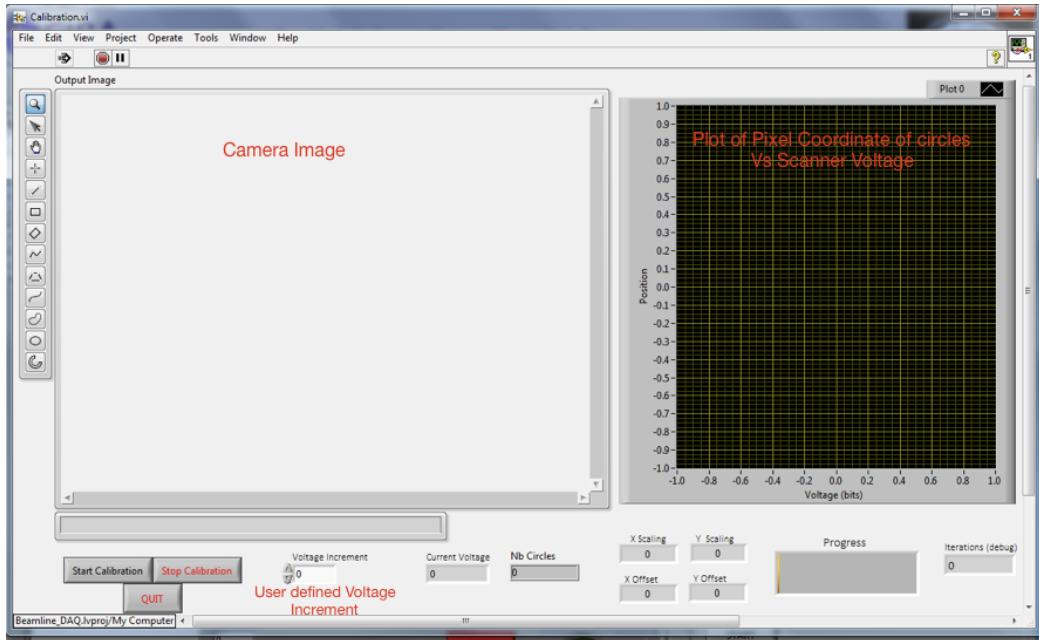


Fig. 3.12. A screenshot of the calibration interface for mapping the optical image pixel coordinate to the magnetic scanner scanning voltage.

An experiment is conducted to test the calibration and the manual targeting function (to be covered in the next section) in the software. The experimental setup is shown in Fig. 3.13 where a 2 MeV proton beam is incident on a Quartz target mounted on the external beam nozzle, to produce a fluorescent beamspot. This beamspot is visualized through a 5X objective microscope fitted with a Qimaging Retiga R6 CCD camera. The CCD image is displayed on the left panel in Fig. 3.12. The calibration algorithm is then run using this experimental setup and the result is shown in Fig. 3.14. The top figure in Fig. 3.14 shows the interface at the 14th iterations and the bottom figure shows that at the 33th iterations. The

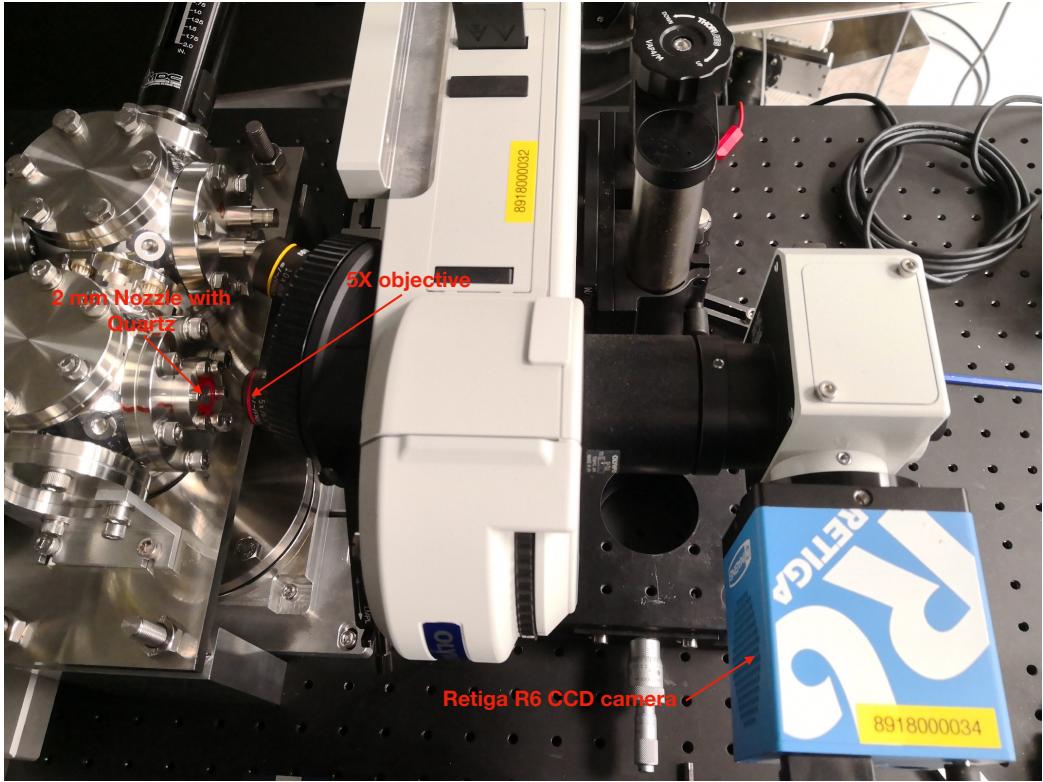


Fig. 3.13. A photo of the experimental set-up to test the calibration and targeting functionality of the software.

(X,Y) voltage is increased from 0 V to 3.5 V at a step size of 0.1 V. The left panel in Fig. 3.14 shows the automatic detection of the beamspot by the software and it is clear that the beamspot has moved between the 14th to 33th iterations due to the increasing X and Y voltage. The software has also successfully estimated the (X,Y) positions from the detected circles and these positions are indicated by the white (X) and red (Y) circles in the right panel of Fig. 3.14. A linear regression is performed on the (X,Y) position data and the estimated regression parameters are updated at each iteration. The live update of the 4 parameters are indicated in the *scaling* and *offset values* in Fig. 3.14. The calibration is *stopped* at 35th iterations when it was observed that the data follows a linear trend very well (at least for the voltage range considered) and the estimated regression parameters have stabilised. Now, the 4 parameters can be used for the manual targeting process which will be covered in the next section. In overall, this result shows that the automated calibration function in the

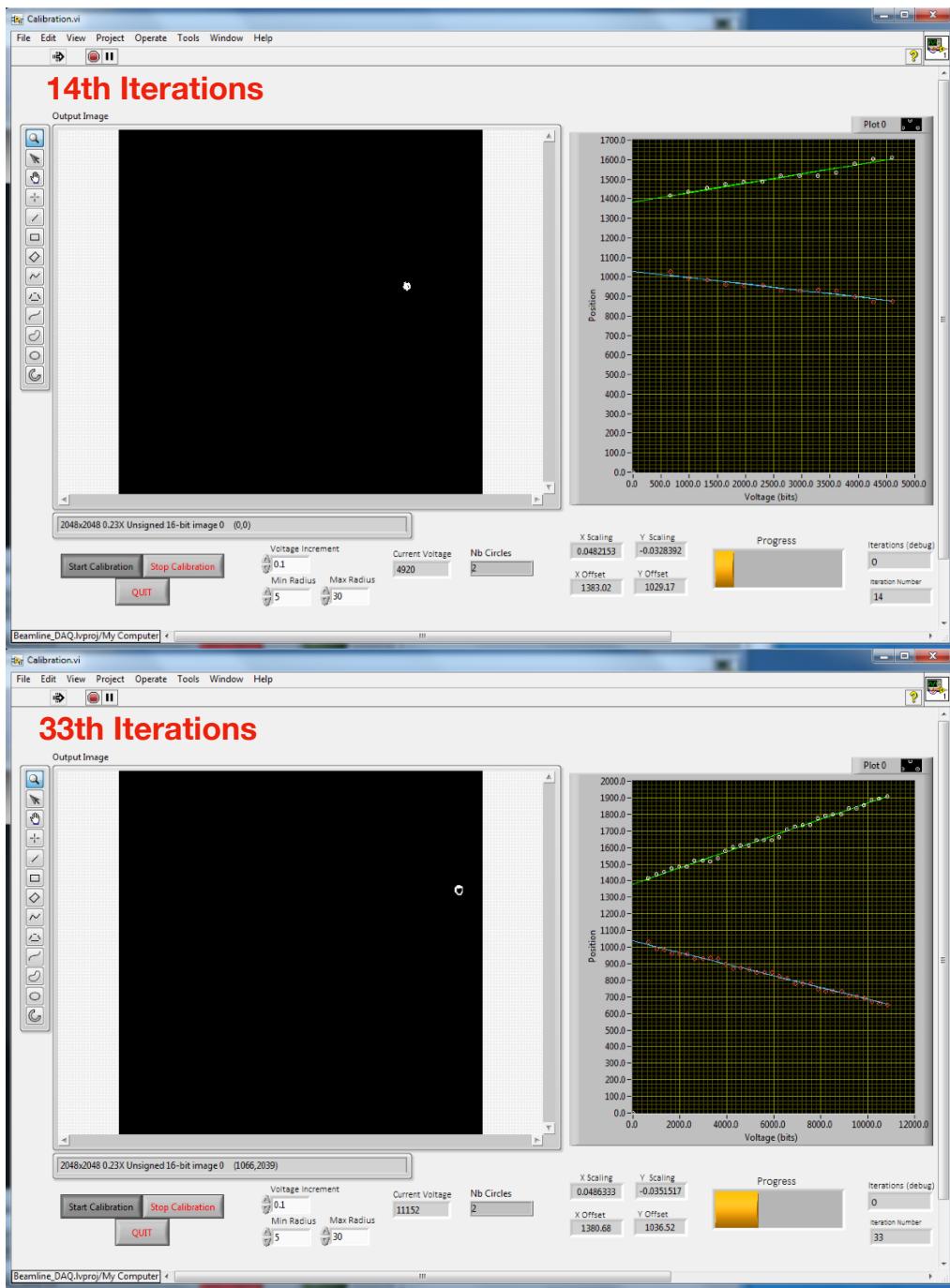


Fig. 3.14. A screenshot of the actual calibration process at the 14th and 33th iterations. The white circles in the right graph are the (X,Y) coordinates of the detected beam spots from the optical image in the left figure. The red and green line are the linear fit to the (X,Y) coordinates.

software works very well during operation.

3.3.4 Manual Cellular Targeting Module

After the calibration between the scanning voltage and pixel position is accomplished in Section 3.3.3, the software is ready to carry out targeting on the cells. The targeting can be classified into *manual* or *automatic*. Manual targeting refers to irradiation at a specific point upon the clicking of cursor on the image, and the second refers to automatic irradiation of all identifiable cell nuclei. This section focuses on the development of the manual targeting interface.

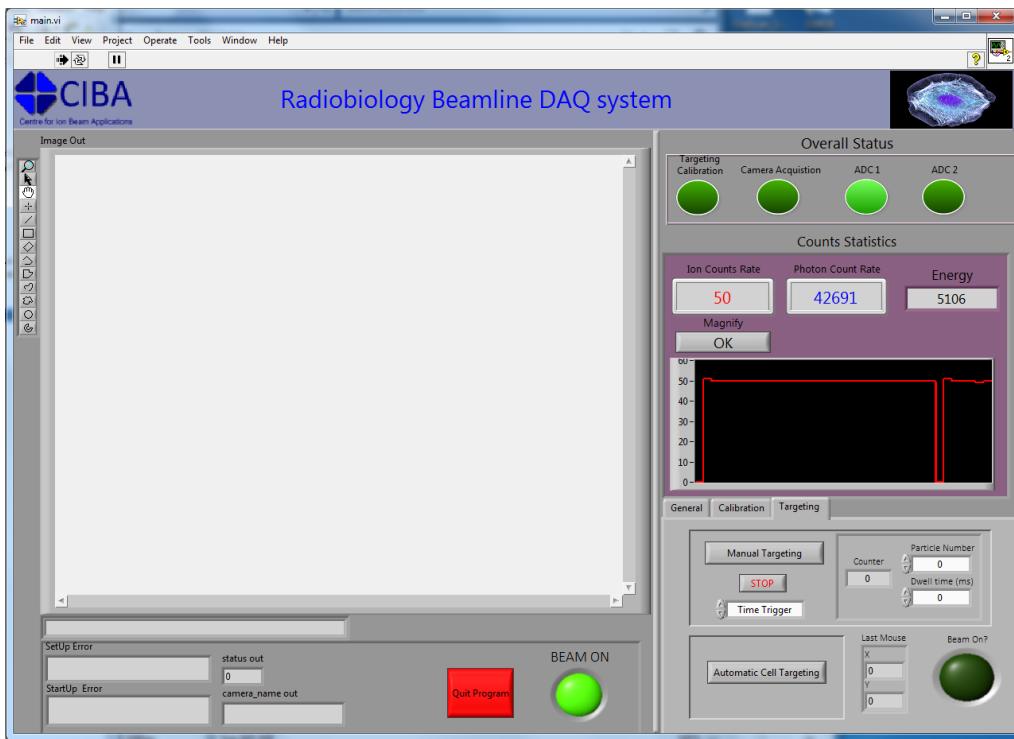


Fig. 3.15. A screenshot of the manual targeting interface

There are two modes of manual targeting - *particle trigger* or *time trigger*. *Particle Trigger* refers to delivering a fixed number of protons to a particle pixel position indicated by the cursor, and *Time trigger* refers to delivering the protons for a fixed amount of time to the similar position. The manual targeting interface is shown in the lower right hand corner of Fig. 3.15. The only required user-inputs are either the number of particles required per pixel or dwell time (in milliseconds) which depends on the

choice of *trigger*. There is a counter to keep track of the number of irradiated targets in *manual targeting* mode. Apart from the *targeting* panel, Fig. 3.15 is also the *home* interface of the software. Once *BEAM ON* is pressed, the beam will be unblanked and the protons or photons count rate will be constantly updated in the *Counts Statistics* panel. The screen in the left panel is the image output from the camera. Since this software is used mainly for identifying and irradiating cells, a constant *live* image is highly desirable during operation.

An experimental setup in Fig. 3.13 is used to test the manual targeting function in the software. The details and calibration process is performed and covered in Section 3.3.3. The result is shown in Fig. 3.16 where random points (shown in red dots) are clicked on the image panel in Fig. 3.15 and the actual beamspot (shown as bright white spot) are recorded using the CCD camera. The measured beamspot lies close to the selected points in the image (red dot). The best targeting accuracy is $10 \mu\text{m}$ and the average targeting accuracy is $18 \mu\text{m}$. One can observe that better accuracy is achieved at the bottom right hand corner of Fig. 3.16 as the calibration is carried out in this region. The deterioration of the accuracy towards the left side of Fig. 3.16 could be attributed to either 1) the *non-parallel* optical axis and the beam axis or 2) the quartz is not flat and perfectly perpendicular to the beam axis. Furthermore, it is observed during experiment that the calibration is lost after about 30 mins and this could be due to temperature variation of *minute movement* in the set-up. This misalignment is of the order of several microns and is hard to avoid with the current equipment: This needs to be improved in the future to achieve multiple sub-nuclear targeting

Apart from the *click-to-shoot* manual targeting function, it is also possible to input a list of predefined pixel coordinates in which the beam will be deflected to into the software. This feature is used to irradiate patterns

such as the word "CIBA" in Fig. 3.17 at a scale of $300\ \mu m$ and $100\mu m$.

To sum, the manual targeting function of the software is working perfectly well and the current set-up can easily achieve *sub-cellular* targeting.

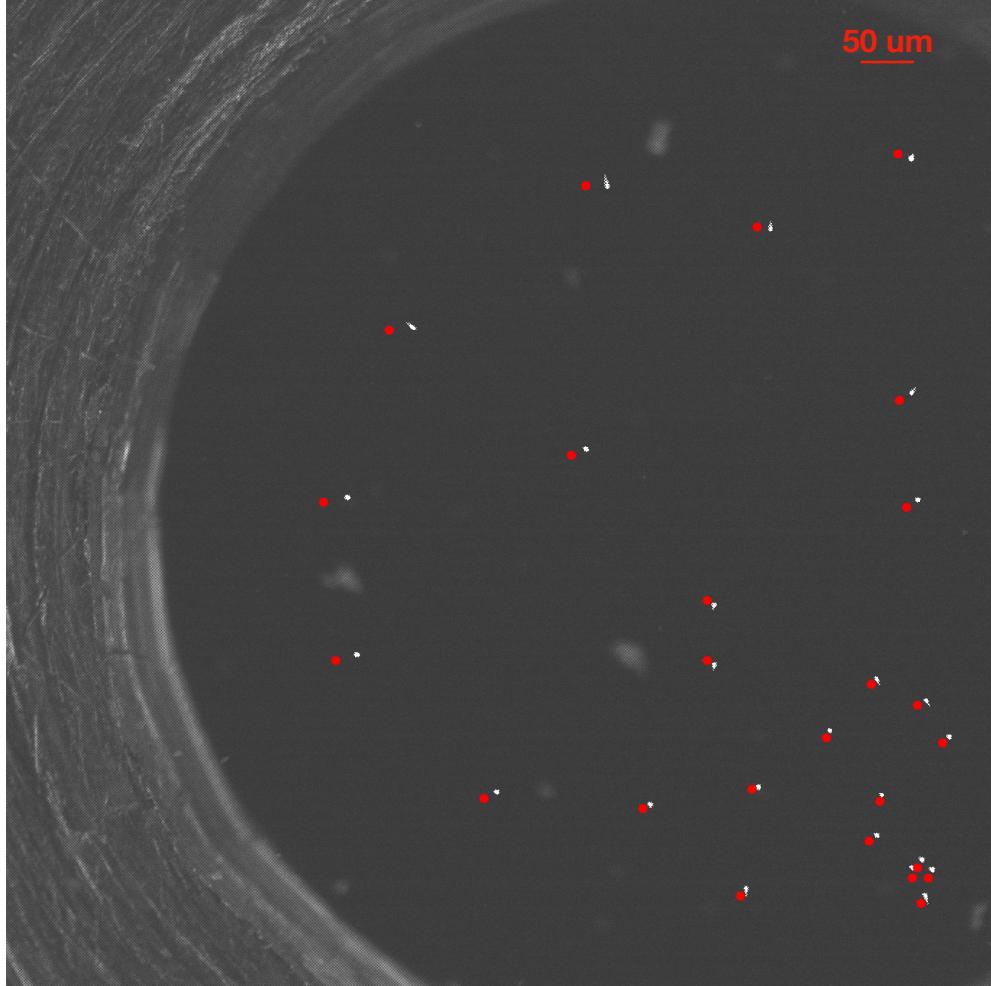


Fig. 3.16. Testing of manual targeting functionality and accuracy with the software by targeting random positions (shown in red solid circle) and determining the actual beamspot position (shown in white spot). The scale is shown in red in the upper right corner of the figure.

3.3.5 Single Proton Detection Algorithm and Results

The fidelity of Single proton detection (SPD) fidelity will affect the performance of the *particle trigger* mode as described in Section 3.3.4. This thesis will focus more on the algorithm and software aspect of SPD, since details of the single proton measurement design and subsequent experi-

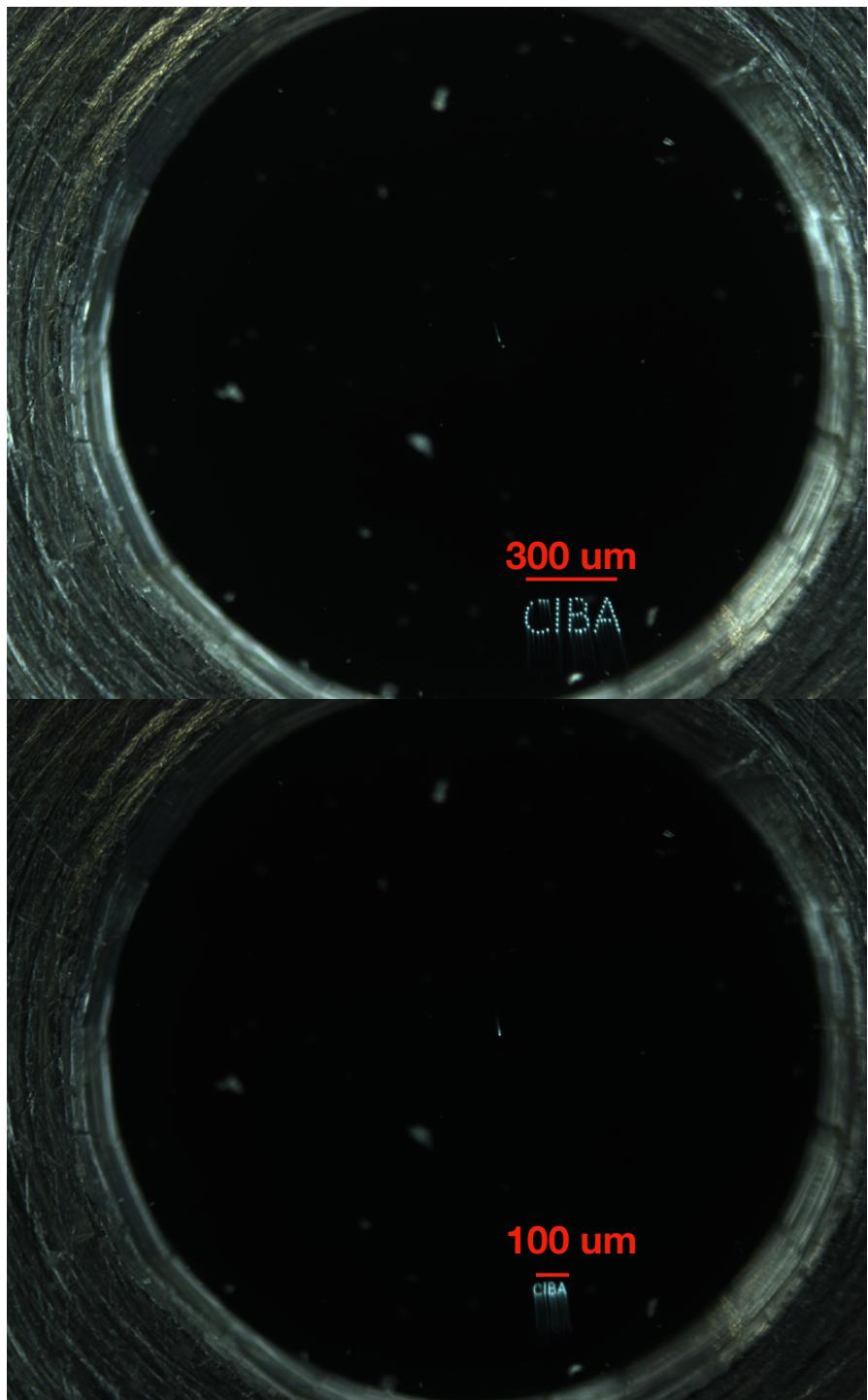


Fig. 3.17. Images of using the targeting software to write "CIBA" in a 300 microns (top) and 100 microns (bottom) scale

ments pertaining to the radiobiology beam line will be covered elsewhere. Here we discuss the design of the single proton detection algorithm for the radiobiology beamline and the *fidelity* measurement carried out in the CIBA proton microscopy beam line to show that the SPD function in the

software is functional. The SPD design is based on the scintillation signal from BC490 material [65]. Fig. 3.18 shows the upper bound of the SPD fidelity given a specified integration time where 2 protons cannot be distinguished due to hardware and software latency (or programmed by users). This result assumes the ideal conditions with no noise pulse (or false positive) and each proton results in exactly one photon pulse into the FPGA. It is calculated using

$$P(|t_1 - t_2| < t_{int}) = 1 - \exp(-C \times t_{int}), \quad (3.7)$$

where t_1 and t_2 are the arrival time of the two photon pulses, t_{int} is the integration time and C is the count rate of the protons. This equation arises from the memory-less property of the *exponential distribution*. The result in Fig. 3.18 shows that ideally a lower count rate is preferred to lower the probability of two protons reaching within integration time. However, in practice a lower count rate will lead to an increase in false positive from the intrinsic detector noise which is detrimental to the SPD performance.

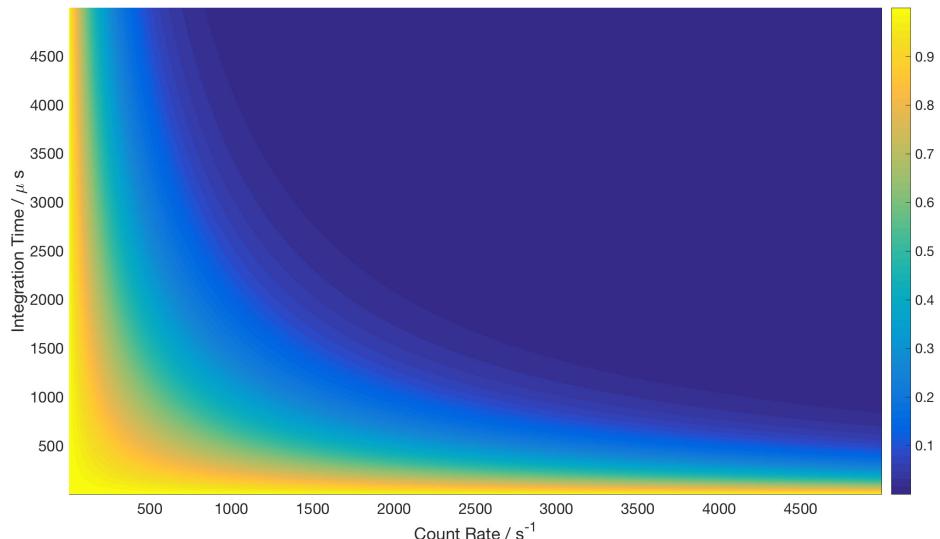


Fig. 3.18. The colour scale represents probability of only 1 proton arriving within a specified *integration time* period (Y-axis) under a specific proton count rate (X-axis).

The experimental schematics to measure the fidelity of scintillator-based SPD system is shown in Fig. 3.19. As the proton traverses through a *thin* scintillator, photons will be emitted. A PMT and PIN diode is placed at the rear of the scintillator to detect both the photon and proton respectively. *If* each proton emits strictly one photon pulse from the counting unit, then the proton can be counted *deterministically* by monitoring the number of photon pulses. The signal through the hardwares from the PIN diode to FPGA has been explained in Section 3.2. The PMT used here is a Hamamatsu R7400P model, which is suitable for photon counting and has a very fast rise time of 0.78 ns. This signal is then directed to a *fast* photon counting unit which consists of an amplifier and a *discriminator* circuit [43]. A TTL pulse of width 10 ns will be produced when a photon signal with a good SNR is inputted into the photon counting unit. The pulse-pair resolution is 25 ns. There are *two main problems* with counting the number of protons using the photon pulse produced by the counting unit.

1. The PMT will produce noise pulse due to ambient light not originating from the scintillator. This noise pulse will be detected as protons and contribute to false positives.
2. Each proton passing through the scintillator may result in more than one photon pulse output from the counting unit depending on the bias applied to the PMT. This will also contribute to the false positive if the time difference between the two pulses are substantially higher than 25 ns, which can happen in many fluorescent materials.

Due to these problems, counting protons directly from photon pulse can be inaccurate. To overcome these problems, the concept of user-defined *integration time* is introduced in the software. This time window defines the sampling time for which regardless of the number of photon pulses

detected by the software, it will be regarded as *one proton*. If well-chosen, the *integration time* can help to reduce false positives in the SPD system arising from multiple photon pulses per proton. The false positive rate from detector noise can be reduced by increasing the proton count rates.

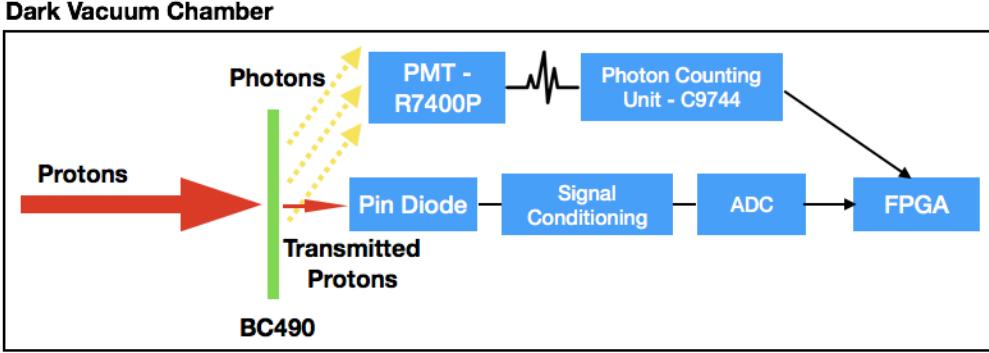


Fig. 3.19. The schematics of the experiment for measuring the fidelity of the SPD system

In this experiment, carried out using the CIBA proton microscopy line, the software is set to detect a *single* proton based on the photon pulse and integration time concept, and the *actual* proton numbers detected by the PIN diode during this process is recorded. This procedure is repeated 1000 times and the distribution of the actual number of proton detected are shown in Fig. 3.20 and Fig. 3.21. Fig. 3.20 shows the result of the measurement by varying the integration time while keeping the count rate constant at 800 s^{-1} . The fidelity of the SPD system improves with smaller integration time and reaches 90% with zero integration time. This result suggests that each proton is producing approximately one photon pulse measured by the PMT, and any further integration time will just increase the probability of additional protons arriving within this time. Fig. 3.21 shows the result of the measurement with different proton count rates. When a lower count rates of 200 is used, the performance is poorer with an increased false positive rate. This is due to lower proton count relative to the noise which results in a higher false positives.

This experiment shows that the software is capable of performing single

proton detection using a scintillation signal and achieve up to 90% fidelity with the CIBA proton microscopy beamline set-up.

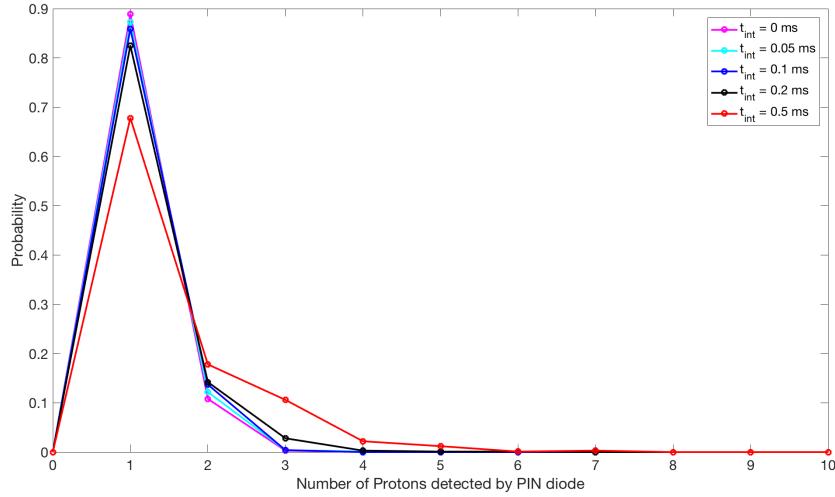


Fig. 3.20. The result of the SPD fidelity measurement for different integration time from 0 to 0.5 milliseconds.

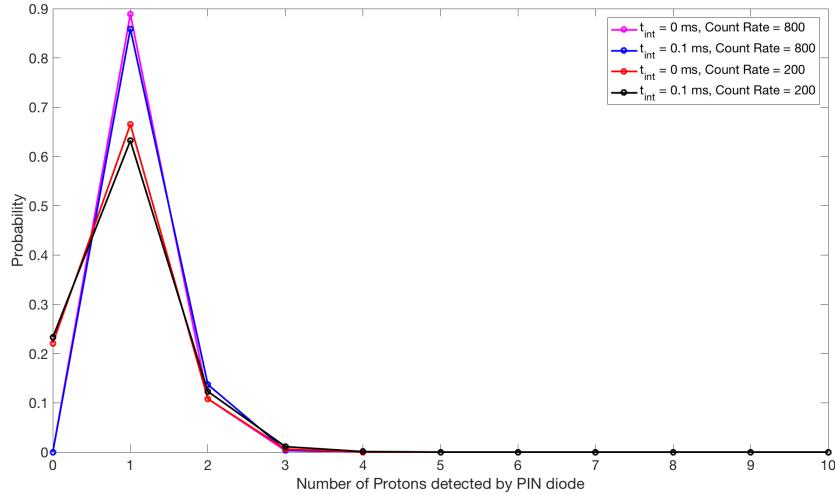


Fig. 3.21. The result of the SPD fidelity measurement for two different count rates: 200 and 800 s^{-1} .

3.4 Future Development

In general, the planned programming requirements for the new radiobiology beam line has been completed and have been shown to work well during

test experiments. However, there are always future developments to be carried out in order to improve the efficiency and performance of the beam line, for example:

1. Development of an automatic cell/nuclei targeting interface for high throughout irradiation of cell nuclei. This function is important for the measurement of biological cells that involve large sample sizes.
2. Develop a diamond detector-based SPD system and interface. This system will utilise a thin film diamond as the exit window, and carry out the functions of separating the vacuum from the ambient environment as well as allowing single proton measurements for accurate dose studies.

3.5 Summary

1. The aim of this chapter is to present the work on the DAQ and control system of the CIBA radiobiology beamline: This work includes both the hardware and software aspects.
2. The hardware aspect of the DAQ has been designed and set up. This includes the choice of the detectors, signal conditioning units and the FPGA, assembled and interfaced together using the interface unit. The hardware has also been tested and works well during the running of the beamline.
3. The software for carrying out specific tasks with the radiobiology beamline has been programmed using LABVIEW, and extensive programming has been carried out for each task.
4. The software can perform STIM imaging, which is required for measuring the beamspot resolution. This feature has been tested experimentally by imaging the proton STIM image of a gold grid. This

module can also be used to induce periodic damage patterns in cells, and has proven to work well from cellular irradiation experiments, where bright lines or spots corresponding to γ -H2AX or DNA damage are observed.

5. The software has been experimentally tested and has been found to be capable of performing automatic targeting calibration to map (X,Y) scanner voltage to (X,Y) pixel position in an optical image. This is achieved by automatically detecting the movement of the fluorescent beamspot on a Quartz or scintillator window.
6. The software has been experimentally tested to be capable of performing time-based manual targeting with average accuracy less than 20 μm . Thus, sub-cellular targeted irradiation is easily achievable with the current radiobiology beamline.
7. The software has the function to carry out single proton detection using scintillation signal through a PMT. The fidelity of the SPD is measured to be 90% using the CIBA proton microscopy beam line.

Chapter 4

Overview of Radiobiology Simulation Software

4.1 Motivation and Aim

Radiobiology has been predominantly viewed as a biological field involving medical clinicians and biologists who have worked to develop experiments to investigate how cells, tissues or animals respond to various types of radiation. However, another mode of inquiry is possible in radiobiology, which depends on modelling and computational techniques. These approaches are important as it complements experiments and seeks to provide a quantitative understanding in radiobiology by quantifying relevant biological endpoints. In the rest of this thesis, the focus will be on the development of a radiobiology simulation program. In addition, the results from the application of this program to answer some scientific and clinical questions relevant to Proton Therapy will be presented. The development of this type of simulation program is still in its infancy, and commercial radiobiology simulation software is essentially non-existent at the point of writing the thesis. The lack of operational software in this important field is a primary motivation for this work. The second motivation is to develop the software

from different foundations to cross-validate with other developing simulation data, and this is important for understanding the impact of modelling on the observables.

The eventual aims of our program is as follows:

1. Ability to explain experimental measurements quantitatively and in an *ab initio* manner.
2. Allows testing of hypothesis *in silico* on the physical, chemical and possibly biological phase of the radiation-cell interaction.
3. Ability to eventually predict the radiation effect on tumors realistically. This can be helpful in clinical setting where better irradiation strategies can be devised to optimize therapeutic ratio [15].
4. Integrate the radiobiology simulation program with the Treatment Planning System (TPS) to achieve a fully radiobiology-oriented TPS which will better predict the effect of radiation.

There are strong motivations and benefits for developing such a simulation program but a full realization of the benefits stated above will take many years to materialize. This is because to fully model the biological effect of radiation is very challenging. Fig. 4.1 shows the multi-scale, multi-physics and multi-disciplinary nature of this type of investigation [1, 2, 3]. The need to integrate all the different spatiotemporal scales, fields and modelling techniques together makes a complete quantitative understanding very challenging. This is further hindered by the fact that a complete experimental understanding is still lacking in many areas including the biological processes of damage and repair. However the situation is improving with synergistic efforts between modellers and experimentalists. Hence, in this thesis, I will present in details the initial steps that I took towards these goals.

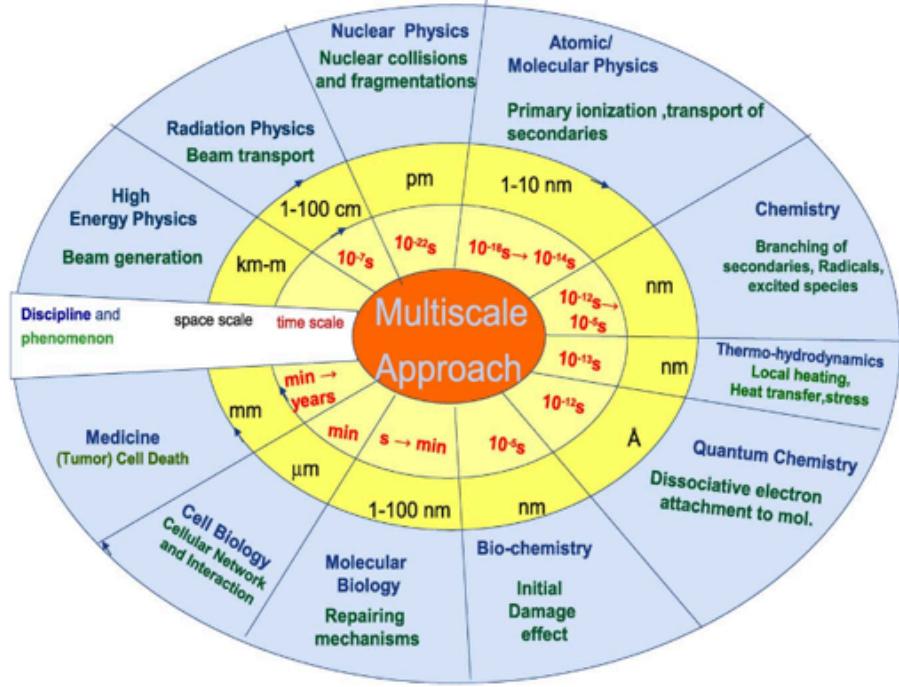


Fig. 4.1. A figure showing the multi-scale, multi-disciplinary and multi-physics nature of the understanding of biological effect of radiation. Reprinted from [1], with the permission of AIP Publishing.

4.2 Review of existing simulation software

Fig. 4.2 shows a table containing a review of the existing programs and endeavours to simulate biological effect of radiation. These include named programs such as PARTRAC by Dr. W. Friedland [66, 67], BIANCA II by Dr F. Ballarini [68, 69, 70], RITRACK by NASA [71] and KURBUC by Dr. H. Nikjoo [72, 73, 74, 75]. There are also un-named model and simulation such as Ref. [76] by Dr. M. Douglas during his PhD, Refs. [2, 3] by Dr. E. Surdutovich et. al., Ref. [6] by Dr. Y. Liang et. al. and the most recent Ref. [77] by Dr. S. Meylan and the GEANT4-DNA collaborators. Each program is unique in its approach and has different degrees of complexity and realism. Some programs are meant to be task-specific (for quantifying a single biological endpoint) while others are ambitious in trying to be multi-purpose. In designing the program, the author has to decide on the amount of details to include in the simulation ranging from

how many spatio-temporal processes to simulate (as shown in Fig. 4.1) and the method of simulation. The method can be a coarse-grained modelling approach [2, 3] to a very detailed *ab initio* simulation [66, 74]. The benefit of a simpler coarse-grained model is that it often involves fewer free parameters and could yield an analytical expression that shows clear connection between several quantities. However, this comes at a cost of lack of flexibility to include further details when further experimental observations are made. Hence, in radiobiology (especially with charged particles) where new experimental findings and data are still being produced regularly, a program with considerable details and flexibility is highly desirable. An example of such a program is PARTRAC [66].

Author & Affiliation	Program Name	Features	Disadvantages
Dr Werner FriedLand, Helmholtz Zentrum Hospital (1998 - present)	PARTRAC	<ul style="list-style-type: none"> Currently the state of the art radiobiology simulation (closed to 20 years of work). Direct and indirect effect are simulated with MC and realistic chromosome geometry up to atomic resolution are used. DNA repairs simulations are included and have been well-verified with various particles and photon radiation. 	<ul style="list-style-type: none"> Physical and physico-chemical part are well simulated. The immediate biological endpoint such as DNA fragmentation and chromosome aberrations are well treated. However, this simulation code is not open-source and cannot be accessed by other research groups or medical centres.
Francesca Ballarini , Pavia University, Italy	BIANCA II	<ul style="list-style-type: none"> Simulate chromosome aberrations and developed realistic biophysical model to predict Cell Survival Curve (dicentric formation etc). 	<ul style="list-style-type: none"> Limited in its use to other biological endpoints.
NASA	RITRACK	<ul style="list-style-type: none"> Good treatment of direct and indirect effect. Simulation of radical reaction with DNA are taken into account. Nice and user-friendly GUI 	<ul style="list-style-type: none"> Little or no biological endpoints modelling and simulation. Works purely in the physical and physico-chemical stage. Distribution regulated by NASA
Hooshang Nikjoo, Karolinska Institute (2001 - present)	KURBUC	<ul style="list-style-type: none"> Use KURBUC for track structure simulation Use realistic DNA geometry up to nucleotide resolution for calculating initial DSB yields Predict DSB, SSB yields for proton, alpha and photon Modelled NHEJ repair using differential equation method 	<ul style="list-style-type: none"> NHEJ repair modelling does not include diffusion of DSB ends thus not able to predict chromosome aberration.
Dr Michael Douglas, Adelaide Hospital (2012-2015)	PhD Thesis	<ul style="list-style-type: none"> Use Geant4 for simulation of direct effect for multicellular case modelled as ellipsoids. Use clustering algorithm to identify strand breaks Use TLK model to predict CSC 	<ul style="list-style-type: none"> No chromosome geometry information No indirect effect simulation
Eugene Surdutovich and Andrey V. Sol'yoz, MBN Research Center, Germany (2008-2014)	Analytical expression	<ul style="list-style-type: none"> Purely analytical and phenomenological in the approach to calculate RBE and Cell survival curve No MC simulation 	<ul style="list-style-type: none"> DNA repair is not taken into account Chromosome geometry not included Indirect effect not realistically tackled
Ying Liang, Gen Yang et al. , State Key Lab of nuclear physics and tech., Center for quantitative biology (2016)	-	<ul style="list-style-type: none"> Use Geant4 for track simulation Use chromosome geometry up to nucleotide resolution Predict DSB and SSB yields for photon, proton and alpha 	<ul style="list-style-type: none"> Indirect effect is modelled instead of simulated Direct effect not adequately dealt with Currently, might be limited in its application to other endpoints
Sylvain Meylan, Sebastian Incerti et al. , IRSN and CENBG (2017)	DNAFabric + Geant4DNA	<ul style="list-style-type: none"> Use Geant4-DNA for track simulation Chromosome geometry is constructed with DNAFabric with nucleotide resolution and can be configured. Able to predict DSB and SSB for proton radiation in fibroblast nucleus in G0/G1 phase with realistic chromatin domain 	<ul style="list-style-type: none"> Unable to predict the subsequent DNA repair and chromosome aberration yields.

Fig. 4.2. A table of the existing attempts to simulate biological effect of radiation

4.3 Features and Novelties in IRSGS

This section will cover the core features of our work so far on the simulation program known as Integrated Radiobiology Simulation with Geant4 and System biology (IRSGS; pronounced as eye-ris). The features will guide the architecture and modelling used in the simulation. The features and novelties that are implemented in IRSGS are as follows:

1. The modelling approach used in this program should be of sufficient complexity and details so that the program is versatile enough to test a wide range of hypotheses and can be easily extended to include new experimental observations.
2. However, the software should also be application driven and not include excessive details which will not affect the quantification of the biological endpoints. Thus, appropriate coarse-graining is allowed in certain part of the modelling.
3. Use GEANT4-DNA for track structure and radiation chemistry simulation. GEANT4-DNA has excellent support from the international community. This software will not involve developing the track structure code as in PARTRAC [78] or KURBUC [79].
4. Increased focus in tools developed in system biology to model the biology phase of the radiation-cell interaction.
5. The code is developed using an object oriented approach and be modular so that the code can be easily extended or adapted. Python is used for developing the software due to its rapid development. Cython and Multi-processing libraries are used on portions of code where performance is critical.

4.4 Architecture and Blueprints of IRSYS

Fig. 4.3 shows the blueprint of the IRSYS. IRSYS is divided into several modules as shown in the boxes in Fig. 4.3. The arrows show that the output of one module will be used as the input for another module. The blueprint is the architecture of the current state of IRSYS and is currently focused at the cellular level. All the modules in Fig. 4.3 are *designed, developed and coded by the author for this thesis* except the GEANT4-DNA module which is developed by the GEANT4 collaboration [80, 5]. The description of the modules are as follows:

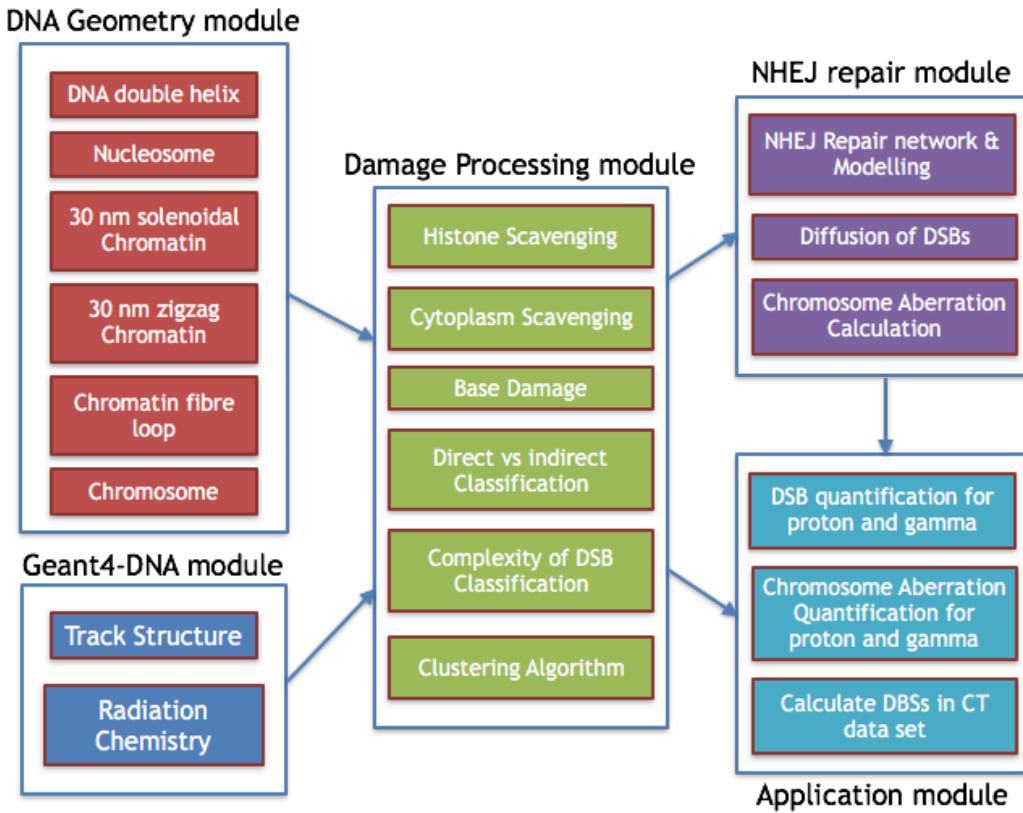


Fig. 4.3. Figure of the blueprint of IRSYS

1. **DNA Geometry Module:** This module creates a data file containing information on the chromosome and chromatin geometry to be irradiated.
2. **GEANT4-DNA Module :** This module creates two separate data

files for the quantification of DNA damage due to direct and indirect effect of radiation. The track structure simulation produces spatial energy deposition data for direct effect processing. The radiation chemistry produces spatiotemporal information of the free radicals for indirect effect processing.

3. **Damage Processing Module:** This module accepts outputs from the previous two modules to identify Single Strand Breaks (SSBs) or Double Strand Breaks (DSBs), and the origin of the damage (direct or indirect). The output of this module will be a data file containing information on the DSBs including spatial position, complexity, genomic index and chromosome number.
4. **NHEJ Repair Module:** This module accepts output from the Damage Processing Module to initialize the DSB repair algorithm. The repair module contains the temporal recruitment of repair proteins and diffusion of DSB ends. The repair module will determine which DSB will be repaired and which will be mis-rejoined to cause Chromosome Aberration.
5. **Application Module:** This module accepts inputs from Damage Processing and NHEJ Repair Modules to answer scientific questions involving biological effects of radiation. The scientific application of the program in this thesis is shown in the cyan boxes. There could in principle be a wide range of applications but we only focused on a subset in this thesis as preliminary work.

Chapter 5

Initializing the Simulation

5.1 Chromosome and cellular target geometry construction

This section covers the methodology and specification of the construction of DNA geometry at different hierarchical levels for the DNA geometry module in Fig. 4.3. There are several programs produced in recent years that are capable of producing DNA geometry within the GEANT4 application [72, 81, 82]. However, most of these programs do not have atomic resolution and by using them, we are limited by the geometry choice which exists which the program has to offer. Thus, to allow for greater flexibility and to test a greater range of hypotheses involving various DNA geometries, we decided to develop our own DNA geometry code. The aim of this code is to produce a data file that contains the following information:

1. 3 float variables representing spatial coordinate information, (x,y,z), of each atom.
2. Boolean variable of whether each atom belongs to the base or backbone of DNA.
3. Integer variable representing the genomic index of each atom. The

genomic index is an unique number representing each nucleotide base in the chromatin.

4. Integer variable to represent the element of each atom.
5. Integer variable to represent which DNA strand does the atom belongs to. There are only two possible values for this variable.

for each hierarchical structure up to a single chromosome level. The construction of the geometry follows closely to Ref. [78], but there are still differences in the final geometry such as the number of nucleotides in each structure.

5.1.1 Implementation details of Nucleosome construction

The DNA construction begins with the information of the atoms within 10 basepairs of pure Adenine in a B-DNA configuration [83, 84], which is available in Ref. [85]. B-DNA configuration is used as it is the most common form found in cell [86]. Pure Adenine is chosen for simplicity and could be replaced with other nucleotides sequence in future development to increase realism or to test hypotheses. After which, a series of transformations are used to wrap the DNA around the histone to form the nucleosome [87]. The 3D rendered images of the nucleosome are shown in Fig. 5.1 using POV-ray software [88]. The magenta cylinder is the histone molecule with height 6 nm and diameter 6.4 nm, and the DNA chain is wrapped 1.8 times around the cylinder. The radius of the DNA chain axis around the cylinder is 4.4 nm and the pitch is 2.7 nm per turn. The red sphere represents Phosphorus atom, white is Oxygen, blue is Carbon and Yellow is Nitrogen. The van der Waals radius of Phosphorus is 0.19 nm, Carbon is 0.17 nm, Nitrogen is 0.15 nm and Oxygen is 0.14 nm. The hydration shell around the DNA is modelled by increasing the van der Waals radius of all atoms by a factor of

2. These data are obtained from [89, 78]. There is a total of 148 nucleotides in the nucleosome in our construction.

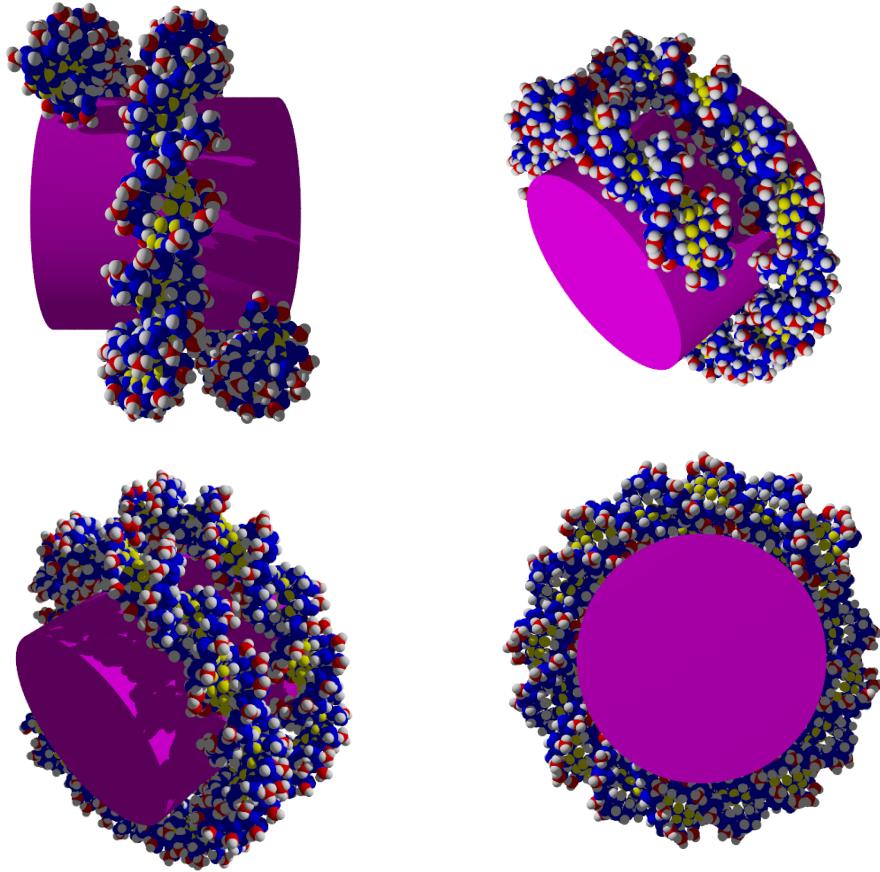


Fig. 5.1. 3D rendering of the nucleosome structure used in IRSQS using POV-ray

5.1.2 Implementation details of Chromatin construction

After constructing the nucleosome, the next step is the chromatin structure, which is formed by connecting multiple nucleosome modules together with linker molecules in between. There are several chromatin structures that were known to exist in the nucleus [83, 78]. We have constructed both the zigzag and solenoidal chromatin structures and they were known to exist concurrently within cell nucleus [90]. The zigzag chromatin [91] is shown in Fig. 5.3 and the repeating unit is shown in Fig. 5.2. The width of

the chromatin is 30 nm and the distance between each repeating unit is 20 nm. Each repeating unit comprises of 2 nucleosomes rotated 180° from each other with a 38 basepairs linker in between. There is a total of 372 nucleotides in each repeating unit. The zigzag geometry is constructed but is not adopted for damage processing module.

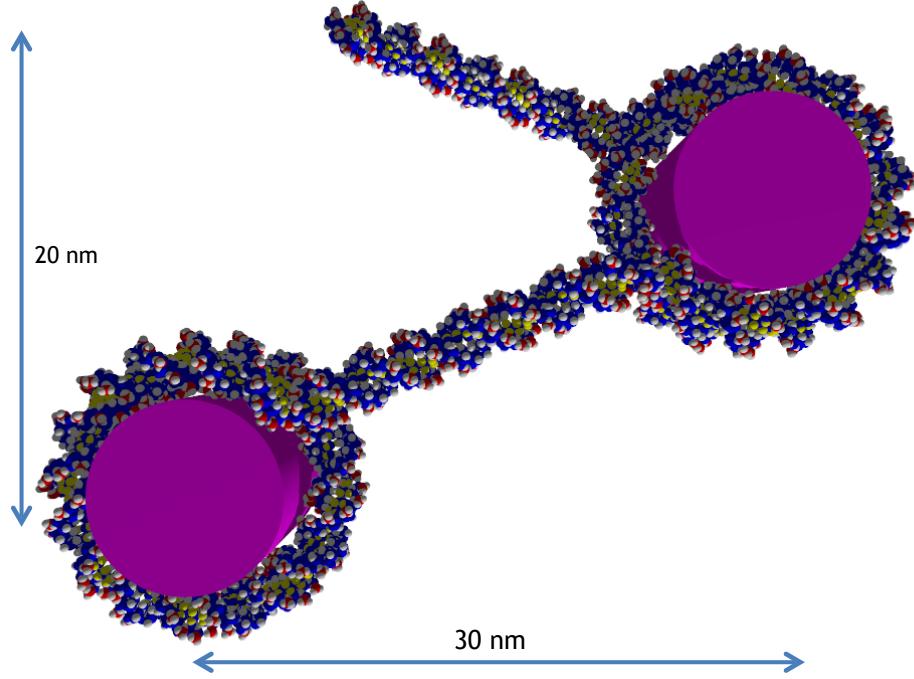


Fig. 5.2. 3D rendering of the repeating unit of the zigzag chromatin structure used in IRSGS using POV-ray

The second chromatin structure that is constructed is the solenoidal form [92] which is shown in Fig. 5.4 and Fig. 5.5. The solenoidal chromatin has a diameter of 30 nm which is similar to the zigzag chromatin. Each repeating unit is made up of 6 nucleosomes as shown in Fig. 5.4, and the angle between each neighbouring nucleosome is 60°. The linker molecule has a length of 36 basepairs. There is a total number of 1092 nucleotides in each repeating unit. A section of the solenoidal chromatin is shown in Fig. 5.5. The solenoidal chromatin will be used for the damage processing module.

The last step in the chromatin construction is to piece several solenoidal

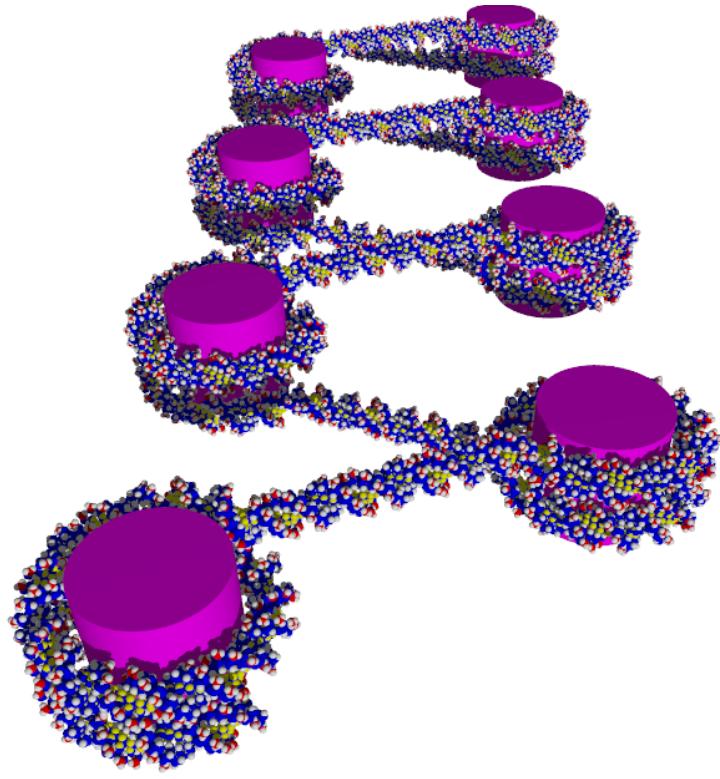


Fig. 5.3. 3D rendering of a section of the zigzag chromatin structure used in IRSGS using POV-ray

chromatin together to form a fibre loop [93]. The 3D rendered image of the basic element of the fibre loop is shown in Fig. 5.6. The dimension of the fibre loop is 279 nm by 192 nm by 30 nm. This is consistent with Ref. [93, 94] where the reported length of the fibre loop is 300 nm . The path length along the loop is about 814 nm and there is a total of 80808 nucleotides in each fibre loop element in this construction. The fibre loop will be the foundational building block for the construction of chromosome. There are more realistic approaches to model chromatin within the chromosome by using worm-like chain [6]. However, for simplicity and reduced statistical fluctuation in later results, we followed the approach in Ref. [78] and constructed the chromosome from the fibre loop in Fig. 5.6.

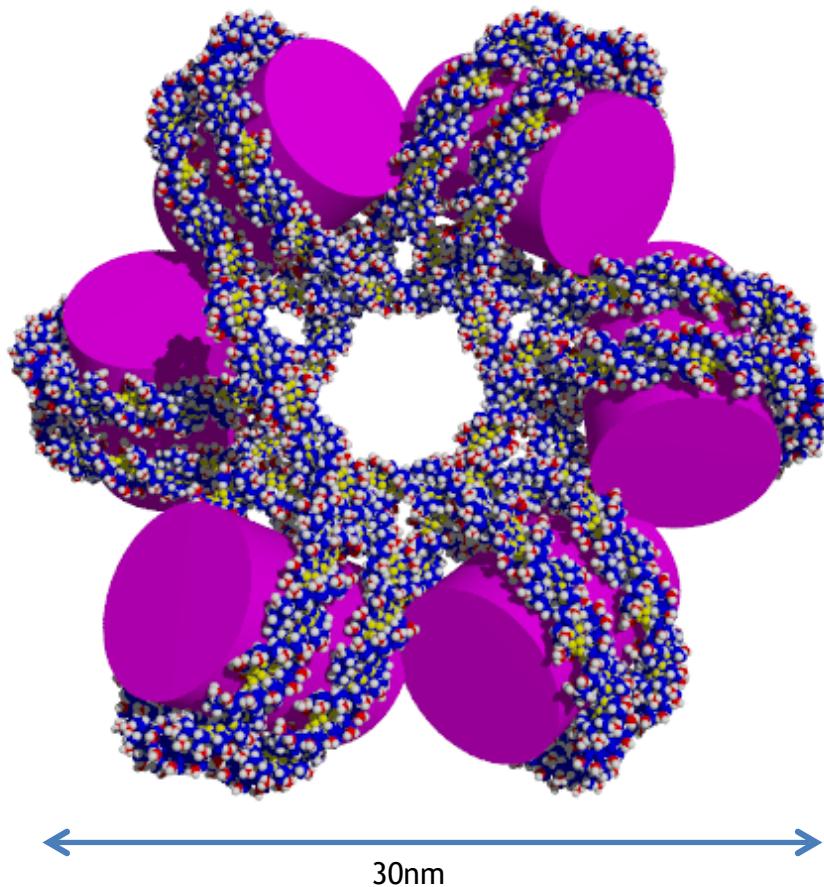


Fig. 5.4. 3D rendering of the repeating unit of the solenoidal chromatin structure used in IRSGS using POV-ray

5.1.3 Implementation details of Chromosome construction

The chromosome geometry in the nucleus varies considerably depending on the stage of the cell cycle [95]. In IRSGS, it is modelled as a distinct cubic domain as shown in Fig. 5.7 to represent the chromosome in the G1 phase (during interphase) [96]. The G1 phase is chosen because the *DNA content* remains relatively constant and is not duplicated during this phase. The chromatin fibre loop is placed periodically within the cube of 1.5 micron. Randomness is introduced by having the orientation of the fibre loop alternating between upward or downward. A total of 1575

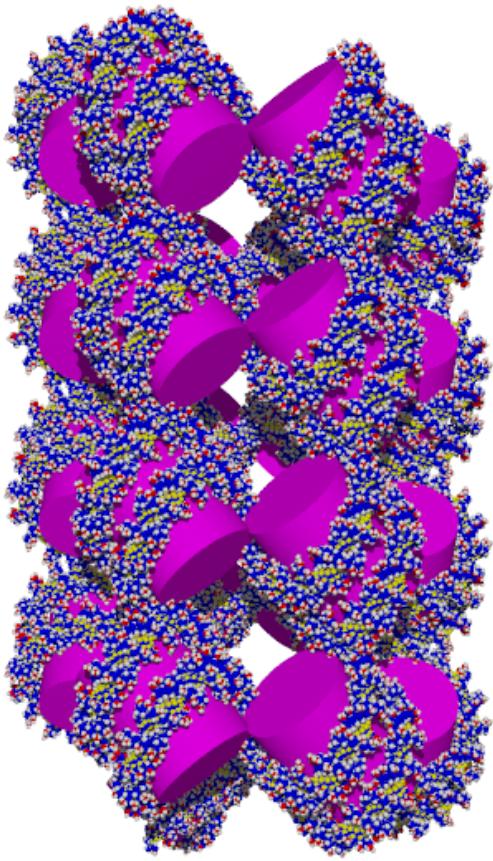


Fig. 5.5. 3D rendering of a section of the solenoidal chromatin structure used in IRSGS using POV-ray

chromatin fibre loops are placed within the chromosome cube in Fig. 5.7. This number is chosen as it gives a total number of 5.85 Mega-basepairs in a nucleus (diploid) which agrees with the number in the human chromosome [97]. The realism of the chromosome geometry only matters when the mean free path of the electron coincides with the characteristic length scale of the chromosome which is about 200 nm. This is reasonable, since most of the electrons produced during proton (of medically relevant energy) transversal in water are low energy electrons averaged at 50 eV or less [1] and have an Inelastic Mean Free Path (IMFP) of less than 10 nm [98]. Thus, this justifies the use of a cubic chromosome.

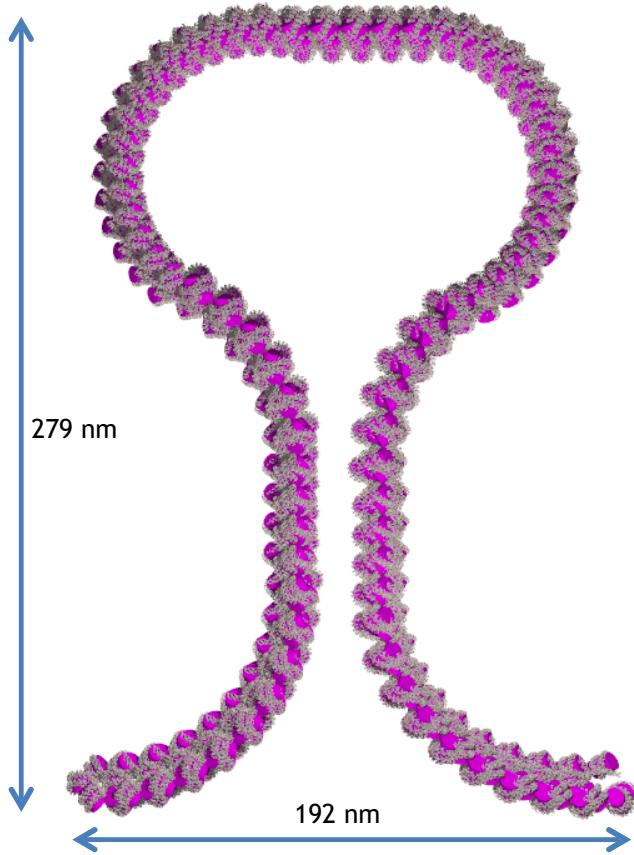


Fig. 5.6. 3D rendering of a chromatin fibre loop structure used in IRSGS using POV-ray. The fibre loop is constructed for the solenoidal chromatin.

5.1.4 Implementation details of Nucleus and Cell construction

The chromosome geometry in Fig. 5.7 will suffice for calculating double strand Breaks or direct and indirect strand break yields. However, the calculation of the amount of chromosome aberration will involve determining the degree of DNA misrejoining amongst the DSBs within the entire nucleus. Thus, a nucleus geometry containing 23 pairs of chromosome needs to be devised. These chromosomes occupy specific regions in the nucleus known as chromosome territories [99]. The 3D rendering of the nucleus geometry used in IRSGS is shown in Fig. 5.8. The yellow sphere represents the boundary of the nucleus of diameter 10 microns and the blue cubes are chromosome domains. The chromosome cubes are placed randomly (but

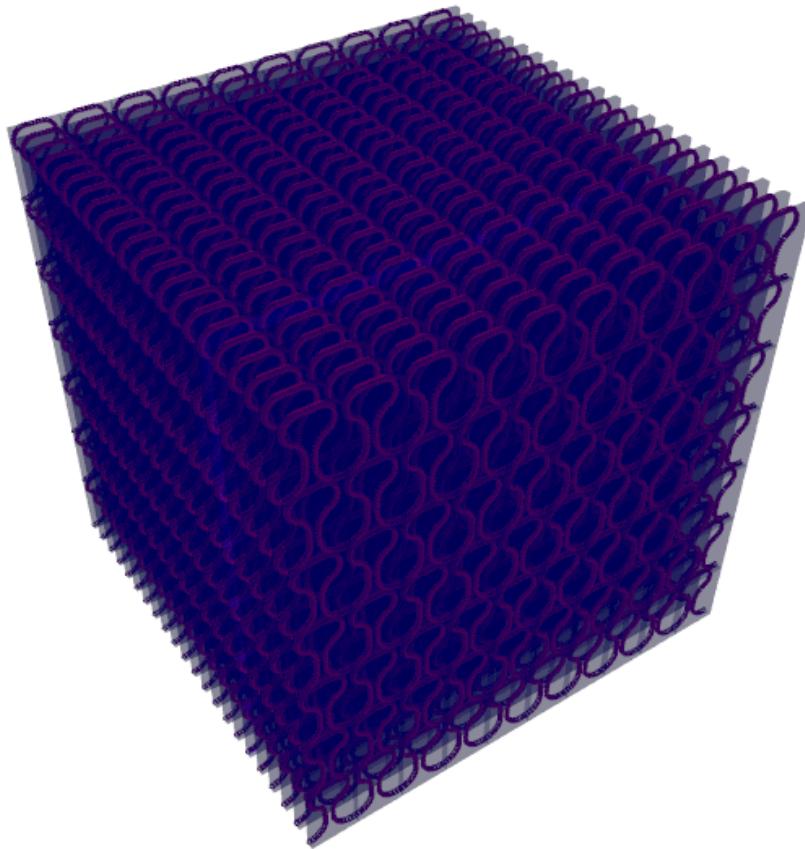


Fig. 5.7. 3D rendering of a chromosome used in IRSGS using POV-ray. The chromosome is modelled as a cube with length 1.5 microns containing 1575 chromatin fibre loops

not rotated relative to each other) subjected to two constraints. First, there must be no overlap between the cube and the spherical boundary and second, the overlap between cubes is kept at a minimal of at most 20 nm. The overlap stems from the Interchromatin Network Model [100, 4] where there are intermingling decondensed chromatin loops between neighbouring chromosome territories.

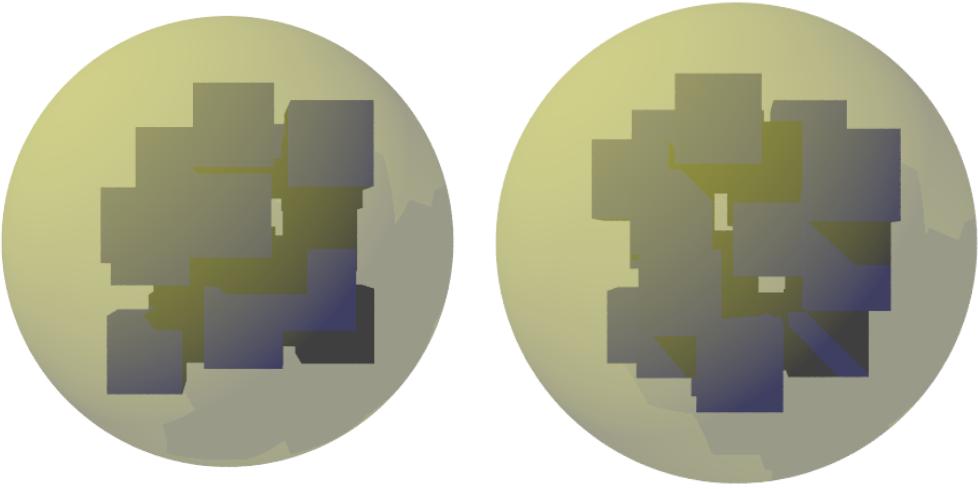


Fig. 5.8. 3D rendering of the nucleus model used in IRSGS using POV-ray. The nucleus shown as a yellow sphere has a diameter of 10 microns and contains 23 pairs (46 blue cubes in total) of chromosomes with limited overlap as supported by experiment [4]

5.2 Introduction to GEANT4-DNA

5.2.1 Introduction to GEANT4 Monte Carlo

GEANT4 framework [80] is highly important for the simulation as seen in Fig. 4.3 where it forms one of the two inputs into the Damage Processing module. Geant4 is a radiation transport code which has a huge library of physics models and can perform radiation transport in any material and geometry. It is widely used for medical, radiation protection and space application. IRSGS uses a specific extension of Geant4 known as GEANT4-DNA which is designed for radiation transport at the nanometre scale [5, 101, 102].

There are two main types of Monte Carlo simulation - condensed history algorithm and event-by-event (or track structure) algorithm [103]. The condensed history algorithm speeds up the simulation of charged particle

transport (especially for ion tracks such as proton or alpha particle) by using Continuous Slowing Down Approximation (CSDA) and not explicitly producing every single electron if it's energy is less than the cut-off energy. Whereas in track structure algorithm, every secondary electron is explicitly simulated and tracked, thus resulting in a more accurate result but at the expense of increased computation time. GEANT4-DNA belongs to the track structure algorithm class and GEANT4 belong to the condensed history. Hence, GEANT4-DNA offers the highest accuracy in simulation at the nanometre scale. However, currently GEANT4-DNA is only applicable to water molecules which forms the bulk of biological materials. A short overview of the Monte Carlo algorithm for track structure method will be covered next before giving an overview of the physical models and radiation chemistry module used in GEANT4-DNA. This overview is important as a proper understanding of the underlying models and assumptions in a simulation will allow us to better understand the implications of the results.

A general track structure algorithm for charge particle transport is as follows [103]:

1. Sample the initial phase space of an electron, either from the source distribution if it is a primary particle or from double differential cross section (DDCS) if it is a secondary particle.
2. Calculate the total interaction cross section per atom including excitation, ionization, elastic scattering and possibly bremsstrahlung process [104, 105]. The mean free path or interaction length, λ of the electron of energy E is given by

$$\lambda(E) = (\Sigma_i[n_i\sigma(Z_i, E)])^{-1}. \quad (5.1)$$

$\sigma(Z_i, E)$ is the total cross section of atom Z_i and the Σ_i runs through

all elements in the medium following Bragg’s rule of addition [105]. n_i is the number density of material i in the mixture. λ will be used to sample the next point of interaction from an exponential distribution. The final distance to the interaction point is the minimum of *step size* (which is defined by the user) and the sampled distance. A smaller step size will give a more accurate result but a longer computation time.

3. The probability of physical process k happening at the point of interaction is given by the ratio of the cross section of k and the total cross section. This can be determined by sampling from a uniform distribution. The energy and direction of the primary particle is updated according to the physical process that takes place. If ionization takes place, secondary electron will be produced and the energy and angular position is determined from the DDCS.
4. This process is then repeated until the energy of the primary particle approaches zero.

This concludes the overview of the track structure algorithm. The main bulk of the computation cost arises from the need to repeat the process for a large number primary particles to obtain sufficient statistic and this is made worse by the tracking of secondary electrons produced explicitly in this algorithm.

5.2.2 Overview of Physical Models in GEANT4DNA

From the overview of the algorithm of Monte Carlo simulation in the previous section, it is evident that the cross sections are the most important inputs in the simulation. It determines the mean free path and the relative occurrences of the various physical processes, which in turn affect the energy deposition patterns. This section will review the physical models

available in GEANT4-DNA [5]. The information on the physical models and their condition of validity for electron and proton are shown in Fig. 5.9 and Fig. 5.10 respectively.

Interactions	Process Class	Model Class	Min Energy	Max Energy	Tracking cut for electron	Type	Used in IRSGS?
Elastic Scattering	G4DNAElastic	G4DNACHampionElasticModel	7.4 eV	1 MeV	7.4 eV	Interpolated	Yes
Elastic Scattering	G4DNAElastic	G4DNAScreenedRutherfordElasticModel	0 eV	1 MeV	9 eV	Analytical	No
Elastic Scattering	G4DNAElastic	G4DNAUeharaScreenedRutherfordElasticModel	9 eV	10 keV	10 eV	Analytical	No
Elastic Scattering	G4DNAElastic	G4DNACPA100ElasticModel	11 eV	255 keV	11 eV	Interpolated	No
Electronic Excitation	G4DNAExcitation	G4DNABornExcitationModel	9 eV	1 MeV	-	Interpolated	Yes
Electronic Excitation	G4DNAExcitation	G4DNAEmfietzoglouExcitationModel	8 eV	10 keV	-	Interpolated	No
Electronic Excitation	G4DNAExcitation	G4DNACPA100ExcitationModel	11 eV	255 keV	-	Interpolated	No
Ionisation	G4DNAIonisation	G4DNABornIonisationModel	11 eV	1 MeV	-	Interpolated	Yes
Ionisation	G4DNAIonisation	G4DNAEmfietzoglouIonisationModel	10 eV	10 keV	-	Interpolated	No
Ionisation	G4DNAIonisation	G4DNACPA100IonisationModel	11 eV	255 keV	-	Interpolated	No
Vibrational Excitation	G4DNAVibExcitation	G4DNASancheExcitationModel	2 eV	100 eV	-	Interpolated	Yes
Attachment	G4DNAAttachment	G4DNAMeltonAttachmentModel	4 eV	13 eV	-	Interpolated	Yes

Fig. 5.9. Table showing the details of the physical models available in GEANT4-DNA and used in IRSGS for an electron

Interactions	Process Class	Model Class	Min Energy	Max Energy	Tracking cut for proton	Type	Used in IRSGS?
Nuclear Scattering	G4DNAElastic	G4DNAIonElasticModel	100 eV	1 MeV	100 eV	Interpolated	Yes
Electronic Excitation	G4DNAExcitation	G4DNAMillerGreenExcitationModel	10 eV	500 keV	-	Analytical	Yes
Electronic Excitation	G4DNAExcitation	G4DNABornExcitationModel	500 keV	100 MeV	-	Interpolated	Yes
Ionisation	G4DNAIonisation	G4DNARuddionisationModel	0 eV	500 keV	100 eV	Interpolated	Yes
Ionisation	G4DNAIonisation	G4DNABornIonisationModel	500 keV	100 MeV	-	Interpolated	Yes
Electron Capture	G4DNAChargeDecrease	G4DNADingfelderChargeDecreaseModel	100 eV	100 MeV	-	Analytical	Yes

Fig. 5.10. Table showing the details of the physical models available in GEANT4-DNA and used in IRSGS for a proton

The *Min Energy* and *Max Energy* columns refer to the minimum and maximum energy of the particle where the model will be used by GEANT4. The *Tracking Cut* column refers to the energy where the particle is killed and no longer tracked and lastly, the *Type* column refers to the method

of cross section calculation. In addition to elastic scattering, excitation and ionisation, the electron interactions include vibrational excitation and electron attachment which are significant for low energy electron, thus enabling GEANT4-DNA to simulate nanometre scale with great accuracy. There are 5 main ionization and excitation channels for water molecules [106, 107]. The ionization channels include 1b_1 , 3a_1 , 1b_2 , 2a_1 and oxygen K-shell. The excitation channels include A^1B^1 , B^1A^1 , two Rydberg series Ryd A+B and Ryd C+D and diffuse bands. These channels are important for determining the dissociation products of water molecules during the pre-chemical stage in GEANT4-DNA. There are also a greater selection of models for each type of interaction for electron and they differ in energy range of validity and the cross section values. There is currently no experimental support for the preference of one model over another, thus we employed the default physics constructor for IRSIGS. The models used in IRSIGS are shown in the last column of Fig. 5.9 and Fig. 5.10. From a physical point of view, the physical processes ends at around 100 femtoseconds after irradiation leaving behind ionized and excited atoms or molecules, and thermalized electrons [108]. The thermalization of electrons occurs through vibrational excitation of water molecule and eventually becomes solvated after 250 fs. The sub-excitation electron then interact via dissociative attachment, vibrational excitation and elastic scattering until they reach 25 meV which they will be regarded as solvated and diffuse through Brownian motion. This solvated electron will enter the radiation chemistry module as e_{aq}^- chemical species.

5.2.3 Overview of Radiation Chemistry Module in GEANT4-DNA

The aim of the radiation chemistry module in GEANT4-DNA is to determine the spatial and temporal distribution of free radicals produced after

irradiation. This section gives an overview of the radiation chemistry module including the methodology, reactions and parameters used [5, 101].

Prior to the *chemical stage* of the simulation, there is a *prechemical stage* which determines the initial free radicals placement in the medium. The initial free radicals are produced from the dissociations of ionized or excited water molecules, and the decay channels and branching ratios are shown in Fig. 5.11. These decay channels can be classified into three main pathways - relaxation to ground state, dissociative decay or auto-ionization. The excited water molecule can undergo all three pathways whereas ionized water molecule is assumed to undergo solely dissociation process. It is important to note that there is little information on the branching ratios of the formation of radiolytic species in liquid water [109], thus the parameters in Fig. 5.11 might be updated in the future with greater experimental inputs. After a pathway is chosen, the dissociation fragment will undergo thermalization to remove the excess kinetic energy from the decay process. This causes the placement of the free radicals at the pre-chemical stage to be different from the original position of the excited or ionized water molecule. The thermalization distance model is taken care of by the GEANT4-DNA.

Electronic State	Decay Channel	Fraction (%)
All ionisation states	$\text{H}_3\text{O}^+ + \text{OH}$	100
Excitation states: A1B1: (1b1) \rightarrow (4a1/3s)	$\text{OH} + \text{H}$ $\text{H}_2\text{O} + \text{Energy}$	65 35
Excitation state: B1A1: (3a1) \rightarrow (4a1/3s)	$\text{H}_3\text{O}^+ + \text{OH} + \text{e}_{\text{aq}}^-$ $\text{OH} + \text{OH} + \text{H}_2$ $\text{H}_2\text{O} + \text{Energy}$	55 15 30
Excitation states: Rydberg, diffusion bands	$\text{H}_3\text{O}^+ + \text{OH} + \text{e}_{\text{aq}}^-$ $\text{H}_2\text{O} + \text{Energy}$	50 50

Fig. 5.11. Table showing the decay branching ratio of the excited or ionized water molecule at 1 ps after the physico-chemical stage in GEANT4-DNA from [5]

After the initial placement of all the free radical products from water,

Species	Diffusion constants ($10^{-9} \text{ m}^2 \text{ s}^{-1}$)
e_{aq}^-	4.9
OH	2.8
H	7.0
H_3O^+	9.0
H_2	4.8
OH^-	5.0
H_2O_2	2.3

Fig. 5.12. Table showing the diffusion constants of the free radicals used in GEANT4-DNA from [5]

Reaction	Reaction rate ($10^{10} \text{ M}^{-1} \text{ s}^{-1}$)
$\text{H} + e_{\text{eq}}^- + \text{H}_2\text{O} \rightarrow \text{OH}^- + \text{H}_2$	2.65
$\text{H} + \text{OH} \rightarrow \text{H}_2\text{O}$	1.44
$\text{H} + \text{H} \rightarrow \text{H}_2$	1.20
$\text{H}_2 + \text{OH} \rightarrow \text{H} + \text{H}_2\text{O}$	4.77×10^{-3}
$\text{H}_2\text{O}_2 + e_{\text{eq}}^- \rightarrow \text{OH}^- + \text{OH}$	1.41
$\text{H}_3\text{O}^+ + e_{\text{eq}}^- \rightarrow \text{H} + \text{H}_2\text{O}$	2.11
$\text{H}_3\text{O}^+ + \text{OH}^- \rightarrow 2\text{H}_2\text{O}$	14.3
$\text{OH} + e_{\text{eq}}^- \rightarrow \text{OH}^-$	2.95
$\text{OH} + \text{OH} \rightarrow \text{H}_2\text{O}_2$	0.44
$e_{\text{eq}}^- + e_{\text{eq}}^- + 2\text{H}_2\text{O} \rightarrow 2\text{OH}^- + \text{H}_2$	0.50

Fig. 5.13. Table showing reaction and reaction rates between free radicals in GEANT4-DNA from [5]

the *chemical stage* will begin. In GEANT4-DNA, the chemical stage is simulated using particle-continuum approach where each radical is simulated

explicitly and the solvent is neglected. This is regarded as a stochastic approach in contrast to a deterministic alternative. During this stage, all the free radicals will time and space stepping via Wiener process (or Brownian process) [110] using diffusion constants from Fig. 5.12. A chemical reaction will occur according to Fig. 5.13 if the physical distance between the two molecules are closer than a calculated reaction radius using Smoluchowski reaction theory [101],

$$k = 4\pi N_A D R_0. \quad (5.2)$$

R_0 is the reaction radius, N_A is Avogadro's constant, k is the rate of reaction is Fig. 5.13 and $D = D_1 + D_2$ is the relative diffusion constant between the two participating species. This process of stepping and reacting is repeated until the specified stopping time, and the spatio-temporal data of the free radical yields can be scored. One limitation of the radiation chemistry module is that currently it can only be run with water medium, but the density can be varied.

5.3 Summary

1. The aim of this chapter is to present the development work and the background knowledge on the steps involved in initialising the IRSGS simulations.
2. Chapter 5.1 presents the first step of initialising the simulation which is to generate the chromosome geometry for irradiation. The chromosome geometry was created with atomistic resolution from the B-DNA structure to nucleosome to solenoidal chromatin and finally to chromatin fibre loop. The chromosome box is then placed randomly within a spherical nucleus to represent the chromosome territories which is consistent with experimental observation.

3. Chapter 5.2 presents the second step of initialising the simulation which is to generate output data files from Geant4-DNA (to be covered in more details in the next chapter). This chapter gives an introduction to the capability of Geant4-DNA focusing on the physical model and radiation chemistry module.

Chapter 6

Quantifying Radiation Damage with IRSGS

6.1 Introduction

This chapter features the fusion of the physical modelling approach with radiobiology to yield a *quantitative prediction* of radiobiological damage. The modelling approach in Biology is very different from in Physics where most models in Physics can be constructed from first principles. For instance, the modelling of material properties is fundamentally rooted in Schroedinger's equation [111] and then applying it to *many atoms* scenarios together with approximations to ease the calculations [112]. However, such methods do not exist for modelling in Biology [113] in which there is often more than one way to model the system.

This chapter will focus on the *models* and *results* for the quantification of two specific endpoints - Double Strand Breaks (DSBs) and Chromosome Aberrations (CAs) due to DSB misrepairs. DSBs represent the *initial* biological damage upon irradiation and there are many experimental measurements using different types of radiation such as Alpha particles, X-rays and Protons. However, DSB is not indicative of *cell-killing potential* as

DSB can be repaired depending on the damage complexity. CA is a better endpoint in determining the *cell-killing potential*, thus is a highly relevant endpoint for the purpose of therapy. The Single Strand Breaks (SSBs) are also produced in large quantities compared to DSBs during radiation damage, but they are easily repairable and does not impact the cell viability [114]. This chapter will first discuss the development of DSB models and the results from Proton and X-ray irradiation. After which, the development of the DNA repair and CA quantification models will be examined, followed by the result with X-ray irradiation.

6.2 Determination of DSB

6.2.1 Prerequisites

There are two main important prerequisites for running the damage processing module as shown in Fig. 4.3. These are the outputs from the DNA geometry module and the Geant4-DNA module. The former module output datas containing atomic information of the DNA in a chromosome geometry as detailed in Chapter 5 whereas the output of the latter module will be discussed now.

Prior to running the *Damage Processing Module*, the *Geant4-DNA Module* needs to be run to generate two important data files. The first file contains the information on the *energy deposition patterns* which includes the energy deposited in the medium at each 3D spatial position. The second file contains information on the 3D position of all *OH radicals* at each successive time step up to 10 ns. The default physics model settings are used in the simulation and the medium is a unit density water cube of length $1.5 \mu\text{m}$ (similar to the chromosome dimension). The energy deposition data is important as this is related to the amount of *direct DNA damage* where the secondary electrons or primary ions directly attack the

DNA via Ionization. The OH radical data is related to the *indirect DNA damage* where the DNA is attacked by reactive OH radicals. There are other harmful radicals produced during the radiation chemistry reaction, but OH radicals are known to be the most lethal [114, 115, 116]. Further details on the calculations of direct and indirect damage will be covered in the next section. One important assumption used in the Damage Process Module is that the spatial *sites* of the DNA damage can be determined by *overlaying* the chromosome geometry data file and the GEANT4 output files. This essentially implies that the DNA atoms do not affect the trajectory of the electrons and free radicals. This assumption is valid for energetic electrons but is questionable for low energy electrons and free radicals. However, almost all radiobiology software uses the *overlay* method and there has been no report on any discrepancies originating specifically from this assumption. Hence, the *overlay* method will be used for the current version of the software and later developments can focus on improving on this assumption.

6.2.2 Calculating Direct and Indirect damage to DNA

It is well-known in the conventional radiobiology paradigm that Direct and Indirect effects are the two main modalities of DNA damage [117, 114, 115]. The definitions of these two forms of damage were introduced in Section 6.2.1. Simulation studies [118] have shown that mechanical damage to DNA through shockwave induction at the Bragg Peak is a possible new modality of damage, however this effect will not be considered due to lack of experimental evidence. The quantification of direct and indirect effect will be discussed in turn.

Direct Effect

The *first step* in calculating the direct strand breaks is to determine the *genomic index* of the strand breaks. A successful direct damage is scored

if the spatial position of the energy deposition lies within twice the van der waal radius of the atom (to simulate the hydration shells surrounding the DNA) and the energy deposited is above a certain *threshold energy*, E_{th} . After which, the genomic index and Boolean flag (belonging to base or backbone) corresponding to the atom will be registered in the memory. The value of the threshold energy is determined using electron data described in a later section.

Indirect Effect

The calculation of indirect effect is more involved than the direct effect. At each time step, every OH radical is checked for a reaction with the DNA. The OH radical will only interact with the DNA backbone if it lies within 0.85 angstroms of the atom and will interact with the bases if it is within 2.9 angstroms from the atom. In addition, the probability of interaction with the backbone is 0.65 and that with base is 0.50. These parameters are obtained from [78] and are determined either from experiments or intuitive reasonings. Apart from reacting with the DNA, the scavenging effect of histone and the medium are modelled explicitly. If an OH radical falls within the histone domain, it will be scavenged immediately and the medium scavenging rate is given by 0.4 ns^{-1} [78, 119]. If any of the above processes happen, the specific OH radical will be *eliminated* and not tracked at the next time instance.

The schematics for the calculation of both direct and indirect DNA damage is shown in Fig. 6.1. Since DSB is the entity of interest, only damage to the DNA backbone is considered in the schematics and base damage will not be considered for the time being. The main bottleneck limiting the speed of the program is the search for the nearest atom for each energy deposition and OH radical position. To circumvent this issue, the nearest neighbour search algorithm is done using KD-Tree algorithms [120]. This algorithm offers better scalability with the size of the search

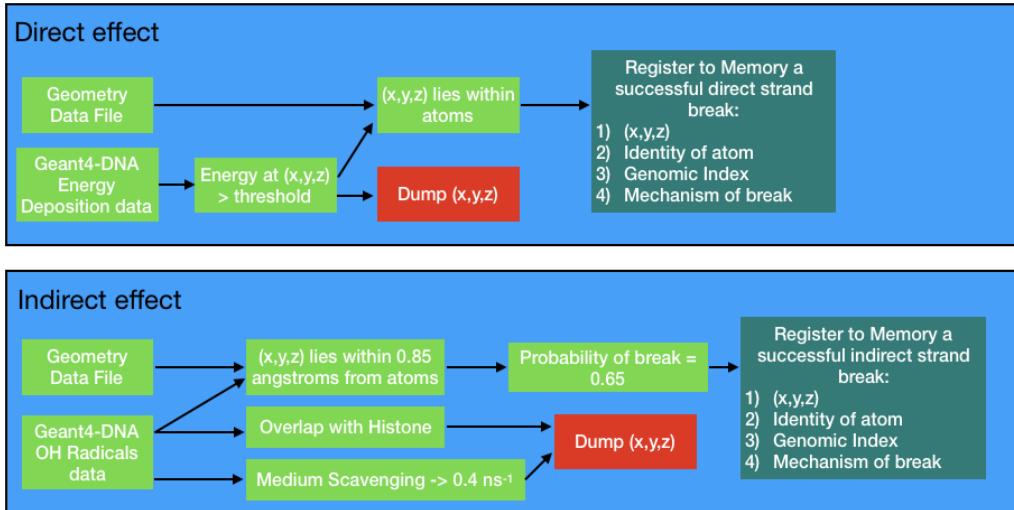


Fig. 6.1. Schematics of Direct and Indirect damage processing

space (logarithmic time complexity) at the expense of an initial time spent for sorting the atomic positions into the tree.

6.2.3 Algorithm for classification of SSB and DSB

After the description of the calculation of direct and indirect effects, the results can be combined to determine the initial DNA damage - SSB and DSB yields. The yields are measured in $Gy^{-1}Gbp^{-1}$ which means per Gray per Giga-basepairs. The implementation details of the SSB and DSB calculations are often inadequate in the literature, thus we devise our own implementation procedures. The genomic index of each break in the backbone is sent through a clustering algorithm known as Density Based Spectral Clustering and Noise (DBSCAN) [121] to determine the cluster. The distance and member threshold for cluster formation is 10 basepairs and 2 breaks respectively. Hence, from the clustering results, all isolated damage or *noise* identified by the algorithm is SSB, and DSB occurs in a cluster when the break happens on opposite strand between 10 basepairs apart [122, 123, 124]. Furthermore, if multiple breaks occurs within a cluster, it will still be regarded as *one* DSB (for experimental comparison purposes, techniques such as Pulse Field Gel Electrophoresis (PFGE) [125, 126] can-

not resolve DSBs within several basepairs apart).

With the SSB and DSB algorithm in place, the rest of this section will describe the determination of the energy threshold, E_{th} for direct effects using experimental SSB and DSB data using 1500 eV electron [127, 128]. The simulation is run using different E_{th} , and the calculated SSB and DSB are compared with the experimental data. This result is shown in Fig. 6.2 and $E_{th} = 16.5\text{eV}$ gives the best agreement with experimental data. Interestingly, this value is very similar to the threshold energy of 17.5eV used in other simulation [73, 74] studies despite entirely different transport codes and method of simulation being used.

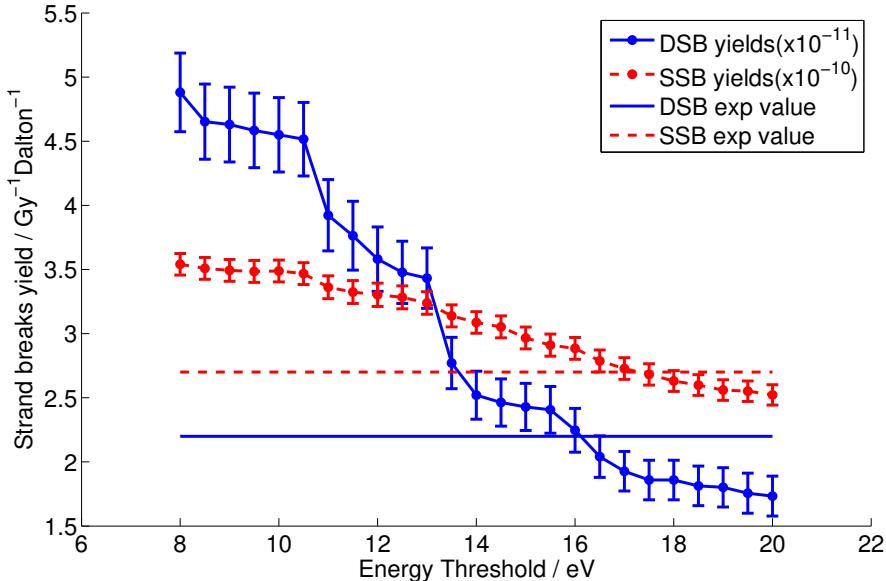


Fig. 6.2. Comparison of SSB and DSB of simulation and experimental data for 1500 eV electron

6.2.4 Results of DSB for Protons

To better understand the biological damage induced at the cellular level during Proton Therapy towards the *Bragg Peak*, the IRSGS program is run for proton beam of energies from 400 keV to 20 MeV. 1000 protons are sent through the chromosome cube for each energy and the strand breaks are scored separately for each incident proton. This is to prevent *cross-fire* in

proton interaction with the DNA which could result in inaccuracies in SSB and DSB calculations. The Linear Energy Transfer (LET) of the proton is calculated from the ratio of the energy deposited within the cube to the dimension of the cube. LET is defined as the energy transferred to the medium per unit penetration depth. The result of the DSB against LET of protons are shown in Fig. 6.3. The simulation results are compared with other simulations [6, 124] and experimental result [129]. In general, the DSB prediction for proton from IRSAGS agrees very well with other simulation despite a wholly different DNA geometry, particle transport and damage processing methodology. This implies that the DSB calculation is *not sensitive* to the details of the mentioned input information. Furthermore, as expected, the DSB increases with increasing LET due to an increase in magnitude and density of energy deposition. The DSB-LET relation is also *linear* for the range of LET considered in this simulation.

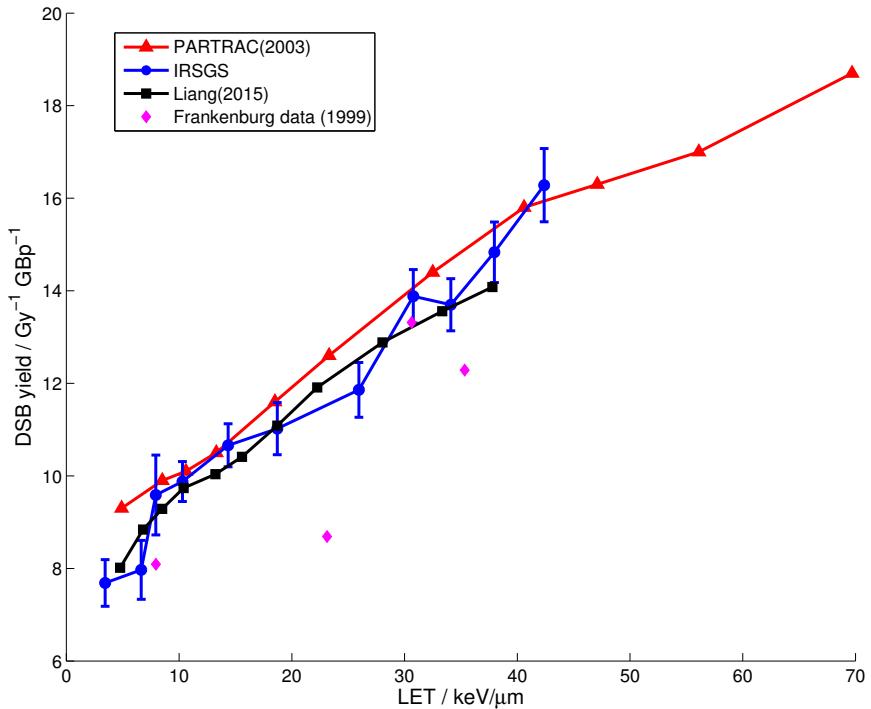


Fig. 6.3. A comparison of the DSB from IRSAGS and other simulation software and experimental data

The results of the SSBs (including breaks within the DSB) due to direct and indirect strand breaks respectively are shown in Fig. 6.4. The error bars arise from the statistical variation in running the simulation multiple times. The indirect SSB remain consistently higher than direct SSB by 1.5 to 2 times for the LET range considered, thus showing that the indirect damage is always the predominant damage process in radiation damage to DNA. The direct effect remains relatively constant for LET between 0 $keV/\mu m$ to 30 $keV/\mu m$. After which, there is a small dip followed by a constant SSB value again. The dip arises due to the mean free path of the proton approaching 1 nm [130] which coincides with the resolution of the energy deposition scoring grid. In contrast, the indirect SSB decreases by up to 1.5 times with increasing LET of the proton. This is due to higher quantities and thus concentrations of free radicals produced by high LET proton which will eventually *self-react*, thus leading to smaller number of free radicals to attack the DNA. Thus, the indirect damage to the DNA decreases as one approaches the Bragg peak region of the proton beam. It is interesting to note that despite a decrease in total SSB with increasing LET, the DSB still increases which is due to the increase in *density* of the strand breaks. Overall, the present IRSIGS program is able to predict the DSB and SSB with individual direct and indirect contributions for proton beam. This represents the initial damage to the DNA from the physical proton beam.

6.2.5 Results of DSB for X-rays

This section will focus on the simulation of the initial strand break distribution using photons. The simulation for photons is challenging due to the low interaction cross section with water within the 1.5 μm chromosome cube resulting in an unrealistically large number of simulations needed to obtain sufficient statistics. There are *biasing* techniques in Monte Carlo

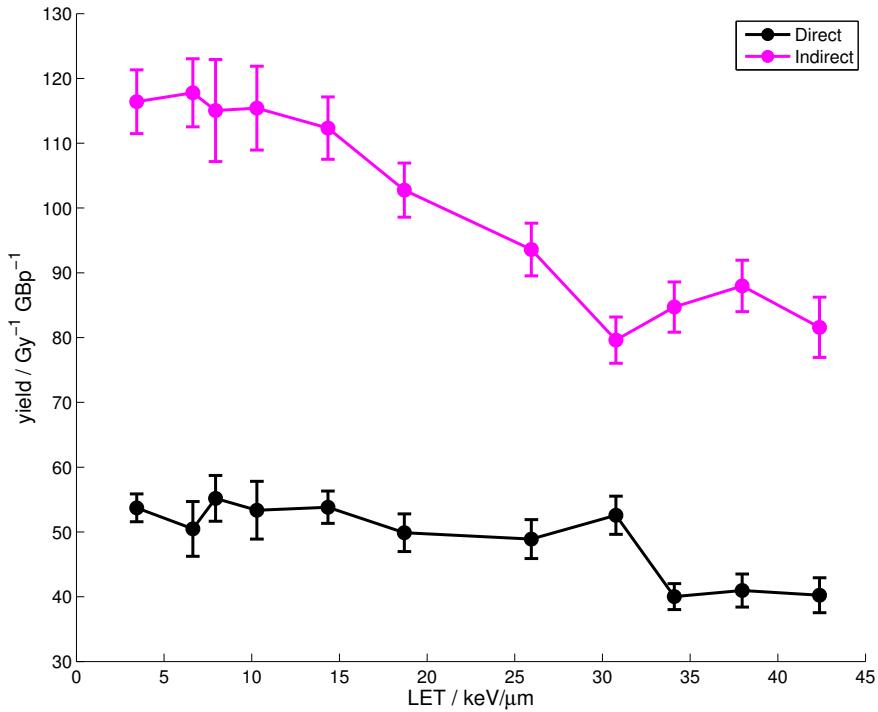


Fig. 6.4. The comparison of the direct and indirect strand breaks

simulations designed to overcome this problem [103]. One such (and perhaps most suitable) technique involves increasing the cross-section of all the interactions, however this could lead to inaccurate spatial patterns in energy depositions. Hence, we adopt an entirely different approach towards this problem by noting two observations. First, the main mode of photon-atom interaction in this energy range [131] is Compton scattering [132] and it is the Compton-scattered electron which is responsible for DNA damage. Second, an electron will definitely interact within the chromosome cube due to the large inelastic cross section with water. Therefore, making use of these observations, we first sample the energy of the Compton-scattered electron using *Klein-Nishina cross-section* [133, 134],

$$\frac{d\sigma}{d\Omega} = \alpha^2 r_c^2 P(E_\gamma, \theta)^2 [P(E_\gamma, \theta) + P(E_\gamma, \theta)^{-1} - \sin(\theta)]/2, \quad (6.1)$$

where α is the *fine structure constant*, r_c is the reduced Compton wave-

length of the electron, E_γ is the energy of incident photon, θ is the scattering angle and $P(E_\gamma, \theta)$ is the ratio of the energy of the photon before and after the interaction. It is given analytically by

$$P(E_\gamma, \theta) = \frac{1}{1 + (E_\gamma/m_e c^2)(1 - \cos \theta)}. \quad (6.2)$$

The *sampling* of the Compton electron energy begins with sampling the scattered angle, *theta* from the cross section in Eq. (6.1). After which, the electron energy can be determined from $[1 - P(E_\gamma, \theta)]E_\gamma$. The IRSIGS simulation is then initialized with the following energy of the electron to determine the SSB and DSB yield spectrum. It must be emphasized that the SSB and DSB yield calculated from IRSIGS with photon represents the initial biological damage at the *vicinity of the photon transversal*. This is important as Compton scattered electrons tends to be energetic and could deposit most of its energy far away from the incident photon (maybe several cells away). This means our simulation will determine the damage directly at the nucleus where the photon traverses.

The simulation was done for high energy X-Ray with energy 80 *keV*, 250 *keV* and 1.25 *MeV*. These energies were chosen for comparison with the simulation data by [6]. The results are shown in Fig. 6.5 where the SSB yields from direct and indirect origins are plotted together with the DSB yield. Overall, IRSIGS simulation results agree well with the results from [6] even though IRSIGS predicts a slightly larger direct SSB and smaller indirect SSB yields. The predicted DSB yields of photons decreases with increasing photon energy. This can be understood from the higher average energy of Compton-scattered electrons leading to smaller LET and lesser biological damage along the photon's path. In contrast to protons, the indirect SSB contributions remain relatively constant throughout the clinically relevant photon energy and is larger than the proton values especially for low energy protons. The direct SSB contributions are comparable for both

photon and proton. Finally, low energy protons (towards the Bragg peak) have higher DSB yields compared to photons, thus confirming the fact that end of range protons results in greater biological damage compared to photons.

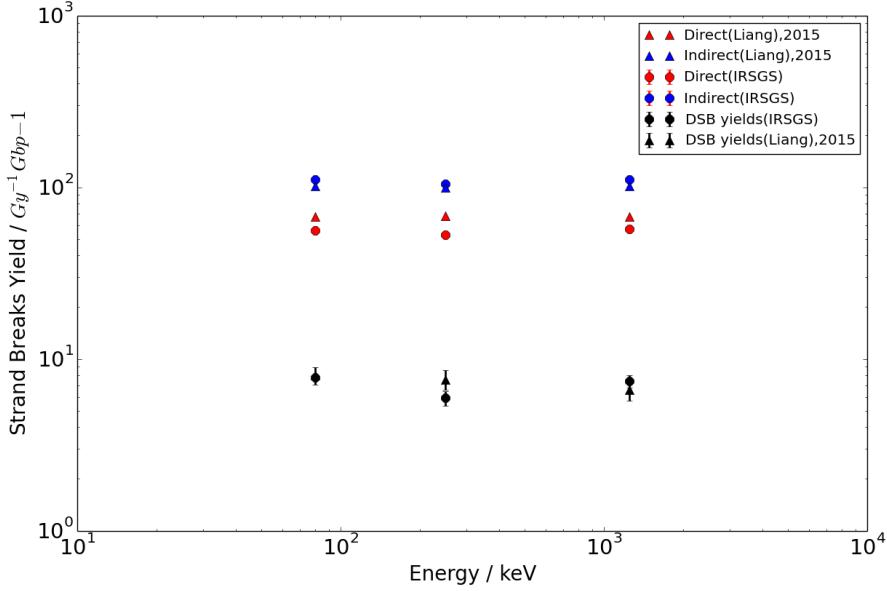


Fig. 6.5. The comparison of the SSB yield from direct and indirect effect, and the DSB yield. The simulation result from IRSIGS is compared to that from [6]

6.3 DNA Repair Model

6.3.1 Biological Introduction to DNA repairs and Chromosome Aberration

After the initial radiation damage on the DNA which results in a spectrum of damage (ranging from single base damage to single and double strand breaks as covered in Section 6.2), the DNA repair mechanism will commence within seconds [135, 136, 137] to repair the damage. For eukaryote cells, there are several possible DNA repair mechanisms [136, 138] namely, Homologous Recombination (HR), Non-Homologous End Joining

(NHEJ). The first repair mechanism tends to be error-free whereas the latter tends to be error prone [136]. The error prone NHEJ repair mechanism is of concern here for two main reasons. First, due to its lower repair fidelity, it is related to the efficacy of cell-killing. Second, contrary to HR which is present only at G2 and S phase of cell cycle [136], NHEJ repair mechanism is present throughout the entire cell cycle. There are two main forms of NHEJ - *Canonical* NHEJ (C-NHEJ) and *Backup* NHEJ (B-NHEJ) [138, 139]. The B-NHEJ uses different repair proteins from C-NHEJ and as a backup pathway, it is more error-prone and will only take precedence when cells are deficient in C-NHEJ. Thus, in this work, the focus will be on C-NHEJ (NHEJ will refer unambiguously to C-NHEJ in the thesis).

Unlike in HR where there is a replica copy of the DNA for repair, the NHEJ mechanism results in mis-repairs due to the diffusion of DSB ends away from its original damage site. Thus, the DSB end could rejoin incorrectly with other free DSB end within or between chromosomes. When such mis-repairs occur, it results in *Chromosome Aberration* (CA) such as the formation of dicentric ring, acentric fragment, translocation and pericentric [138]. Most of such aberrations are highly lethal to cell and will be more closely correlated to cell viability compared to DSBs. Hence, in this chapter, we will first outline in detail the development of the novel mathematical framework and algorithm for modelling NHEJ repair and CA formation. Then, we will present the result showing excellent agreement between our simulation and other measurements from X-ray radiation.

6.3.2 Non-Homologous End Joining Model

There have been several attempts by various groups [140, 141, 142, 143, 144, 145, 146, 147, 148] to model the NHEJ process with different degrees of complexity. The summary of such endeavor is shown in Fig. 6.6. No two NHEJ models in Fig. 6.6 are the same as they differ either in the

Author and Affiliations	Method of Modelling	Parameters	Radiation Type	Details	Refs.
F. A. Cucinotta et. al., NASA (2008)	Differential Equation	9	Low LET radiation (X-rays)	<ul style="list-style-type: none"> Predict Gamma H2AX fluorescence intensity with time. Predict number of Foci against initial number of DSB No classification of simple and complex DSB First NHEJ modelling paper Diffusion of DSB ends not included DSB spatial positions are not included. 	[146]
S. J. McMahon et. al., Massachusetts General Hospital (2016)	Analytical Expression	11	Low LET radiation (X-rays)	<ul style="list-style-type: none"> Predict Cell Survival Curve and Chromosome Aberrations with dose Predict DSB repair and misrepairs with time. Diffusion of DSB ends not included Included MMEJ, HR and NHEJ repair mechanism with each mechanism being parametrised by 1 or 2 constants Lack of details for NHEJ mechanisms No classification of simple and complex DSB 	[145]
O. V. Belov et. al., Joint Institute of Nuclear Research, Russia (2015)	Differential Equation	54	Low LET radiation (X-rays)	<ul style="list-style-type: none"> Predict H2AX Foci and other repair protein recruitment with time Diffusion of DSB ends not included Included network of NHEJ, HR and SSA repair mechanism No classification of simple and complex DSB 	[144]
R. Taleei et. al., Karolinska Institute (2013)	Differential Equation	9	Low LET radiation (X-rays and low energy electrons)	<ul style="list-style-type: none"> Predict H2AX Foci and other repair protein recruitment with time Diffusion of DSB ends not included Modelled NHEJ repair network Classification of simple and complex DSB 	[143], [142]
Y. Li et. al., NASA (2014)	Differential Equation	10	High LET radiation (Ultrasoft X-rays)	<ul style="list-style-type: none"> Predict Ku70/80 and DNA-PKcs repair protein recruitment with time Diffusion of DSB ends not included Modelled NHEJ repair network Classification of simple and complex DSB 	[140]
W. Friedland et. al., Helmholtz Zentrum Hospital (2012)	Stochastic	18	Low and High LET radiation (proton, alpha, X-rays)	<ul style="list-style-type: none"> Predict Chromosome Aberrations and Foci with dose. Predict repair protein recruitment with time. Modelled NHEJ repair network Classification of simple and complex DSB DSB location and complexity determined from Monte Carlo code, PARTRAC Diffusion of DSB ends are included 	[139], [138]

Fig. 6.6. A review of the existing NHEJ repair models

repair network, mathematical tools employed, or the initialisation of the DSBs (as input for the NHEJ model). It is important to note that Refs [146, 147] modelled the NHEJ repair mechanism together with other repair network such as HR or SSA to predict experimental results. However, due to the reasons that we are focusing only on the G1 phase of the cell cycle and that NHEJ repair mechanism contributes most to the DSB misrepair and subsequent chromosome aberration and cell death, NHEJ is the only repair mechanism that is of interest. This also justifies the greater interest in NHEJ repair mechanism in Fig. 6.6.

The NHEJ repair mechanism (or network) used in the IRSGS program is shown in Fig. 6.7. This network is biologically motivated and is inspired by Ref. [142]. The novel features of this DSB repair model compared to other models are the following:

1. The DSB will be classified into simple and complex DSBs. Simple DSB does not need to recruit DNA-PKcs prior to ligation process

whereas complex DSB require recruitment of DNA-PKcs for completion of DSB repair. This is motivated by biological experiment in Ref. [149, 150, 151] and most DNA repair models in Fig. 6.6 does not include this biological observation.

2. Synapse formation only occurs when two DSB free ends (which have recruited the correct repair proteins) diffuses close to one another. Thus, the inclusion of the diffusion models naturally allows for DSB misrepairs and chromosome aberration formation. Only a small number of repair models include a diffusion model for DSB ends [140, 141].
3. Most importantly, this model uses a novel probabilistic approach in contrast to stochastic or differential equation approach. The differential equation approach do not model the DSB individually but instead model the total DSB as a single collective variable and subject it to various reaction kinetics equations [152, 153]. The stochastic approach models each DSB individually including its state conversion and diffusion using Gillespie Algorithm [154]. The DSB in this approach can only exist in each of the state which corresponds to a particular chemical entity in the reaction network. In the probabilistic approach, the DSB is modelled individually as a state vector with each element representing the probability of existing in one of the states in the reaction network. This approach has the advantage of yielding information on the statistics of various quantities and yet avoids the problem of stochastic simulation where huge number of simulation runs need to be made to sample rare events such as certain specific chromosome aberration patterns.

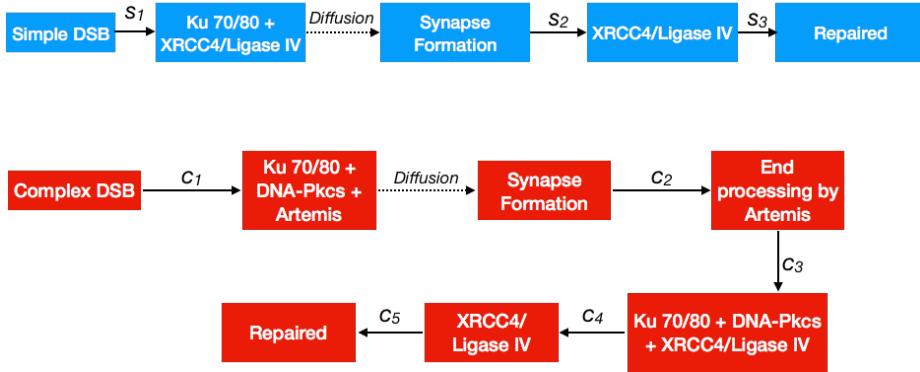


Fig. 6.7. This figure shows the biological network of DSB repair mechanisms with the rate constant definition for each reaction. The repair mechanism differs for simple (shown in blue) and complex DSBs (shown in red).

Basic Mathematical Framework

There are two main inputs to the NHEJ repair model in IRSGS. The first is the DSB information in each chromosome obtained from Section 6.2 which contains the 3D spatial position (relative to the centre of the chromosome) of the DSB, the complexity of DSB, genomic index and finally chromosome index (ranged from 1 to 46). The second is the chromosome position information within the nucleus which is important for assigning the chromosome index to each DSB. The cubic chromosome is placed randomly within a spherical cell nucleus of radius 6 microns. The chromosome cannot touch the nuclear membrane and there should be minimal overlap between the chromosome which is consistent with biological observations [99, 100, 4]. The maximal intersection between two chromosome domains is set arbitrarily to be 20 nm. The 3D image of the Nucleus with the 46 chromosome domains are shown in Fig. 5.8. It is assumed that the proton energy does not change appreciably when traversing through the entire nucleus thus allowing the DSBs damage information for one chromosome (from Section 6.2) to be assigned randomly to each of the 46 chromosomes with equal probability. This forms the basis of assigning chromosome index to each DSB. A DSB is considered *complex* if there is another DNA damage (base or backbone) within 20 basepairs from this DSB [67, 140, 141]. This

is the *only* condition used to determine whether a DSB is simple or complex. Some software models have considered further classification of DSB into varying degrees of complexity or clusters, however currently there is no definite biological support for this classification. Furthermore, DNA fragments of less than 25 basepairs are assumed to be *not joinable* as they are too short for assembly of repair proteins [141, 143].

In the probabilistic approach, each of the DSB free end is characterised by a vector with each element representing the probability that it exists in different biological states as shown in Eq. (6.1).

$$DSB_{simple}(t) = [P_1 \ P_2 \ P_3 \ P_4 \ P_5], \quad DSB_{complex}(t) = [P_1 \ P_2 \ P_3 \ P_4 \ P_5 \ P_6 \ P_7] \quad (6.3)$$

P_i is the probability that a DSB (simple or complex) exists in the i -th state shown in the i -th box (in the same order) in Fig. 6.1 at time t . Also, the condition that $\sum_i P_i(t) = 1$ must hold for every DSB free end at all time t . The general form of the time evolution of each of the probability of the DSB state is determined by Eq. (6.4). Δt is the time step used for the NHEJ repair simulation and should be small by Eq. (6.4) to be valid, and k_i is the rate constant governing the effective transformation from P_i to P_{i+1} . The various k_i rate constants that are used in this model is shown in Fig. 6.7 and $k_i = s_i$ in the case of simple DSB and $k_i = c_i$ in the case of complex DSB.

$$P_i(t + \Delta t) = P_i(t) + k_{i-1}\Delta t P_{i-1}(t) - k_i\Delta t P_i(t) \quad (6.4)$$

Upon close observation of the repair network in Fig. 6.1, one notices that the *diffusion* process governs the rate of conversion from P_2 to P_3 for both complex and simple DSBs. The diffusion process is not controlled by a single rate constant, but an explicit modeling of the diffusion of the

DSB free end to account for the possibility of misrepairs. In contrast to a stochastic approach towards diffusion modelling as used in Refs. [67, 66], we will adopt a probabilistic approach. In this approach, the main question we are interested in is: *What is the probability of any DSB free end rejoining at time t given that the DSB free ends were created in \vec{r}_1 and \vec{r}_2 respectively?* To answer this question, we first assume that the probability of rejoin between DSB end i and j at time t is of this form,

$$P(\text{synapse}; i, j, t) = K(\text{synapse}; i, j, t)\Delta t \quad (6.5)$$

where $K(\text{synapse}; i, j, t)$ is a time dependent rate constant governing the rate of rejoin between DSB end i and j . This time-dependent rate constant can be decomposed into 4 different terms as shown in Eq. (6.6) below. $P_{\text{spatial}}(\text{synapse}; i, j, t)$ represents the probability of two DSB ends overlap with each other spatially at time t . This term arises from the requirement for DSB ends to be located close to each other before synapse can be formed. In addition to being in close proximity, the DSB free ends must also have recruited the relevant proteins for synapse formation. In particular both ends must be in state 2 from Fig. 8.1 for both complex and simple DSB end and the probability of DSB end i being in state 2 at time t is given by $P_2(i, t)$ in Eq. (6.6). Lastly, k_{synapse} represents a constant parameter which control the rate of DSB ends rejoining once they have recruited the relevant repair proteins and are located close to each other.

$$K(\text{synapse}; i, j, t) = k_{\text{synapse}} P_{\text{spatial}}(\text{synapse}; i, j, t) P_2(i, t) P_2(j, t) \quad (6.6)$$

To derive an expression for $P_{\text{spatial}}(\text{synapse}; i, j, t)$, we assume that the DSB free end will undergo Brownian motion where the Probability Density Function (PDF) $f(\vec{r})$ of finding the end at arbitrary spatial position \vec{r} is

given by Eq. (6.7). \vec{r}_i is the initial position of the i -th DSB's free end and D is the diffusion constant. The diffusion constant is taken to be $100\text{\AA}^2/\text{s}$ [140, 141].

$$f(\vec{r}; \vec{r}_i, t) = \frac{1}{(4\pi Dt)^{3/2}} \exp\left(\frac{-|\vec{r} - \vec{r}_i|^2}{4Dt}\right) \quad (6.7)$$

2 DSB free ends will rejoin and form a synapse [140, 141] if they are within a certain threshold distance, Δ , from each other. For the ease of calculation, the threshold distance is applied to the x , y and z distance between the free ends individually instead of the radial distance. Thus, the probability of spatial overlap between 2 DSB free ends within Δ distance are given in Eq. (6.8). The first integral in Eq. (6.8) is over the \mathbb{R}^3 . There will be no further modelling of diffusion once the synapse is formed for two main reasons. First, after a formation of synapse, the diffusion constant decreases [155] and it remains almost spatially fixed, and second, there is no other experimental observation that requires the diffusion of a synapse.

$$P_{spatial}(synapse; i, j, t) = \int f(\vec{r}; \vec{r}_i, t) \int_{\vec{r}-\Delta}^{\vec{r}+\Delta} f(\vec{r}'; \vec{r}_j, t) d^3\vec{r}' d^3\vec{r} \quad (6.8)$$

The integral in Eq. (6.8) can be written in analytical form using error function (erf). This is shown in

$$P_{spatial}(synapse; i, j, t) = \prod_{k=x,y,z} \frac{1}{2} \left[\operatorname{erf}\left(\frac{\gamma_{1,k}}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{\gamma_{2,k}}{\sqrt{2}}\right) \right], \quad (6.9)$$

where $\gamma_{1,x} = \frac{1}{\sqrt{2}\sigma}(x_i - x_j + \Delta)$, $\gamma_{2,x} = \frac{1}{\sqrt{2}\sigma}(x_i - x_j - \Delta)$ and $\sigma = \sqrt{2Dt}$. The multiplication arises from the consideration of overlap between DSB ends in all 3 dimensions. The detailed derivation of Eq. (6.9) is shown in Appendix A. In the limiting scenario of $t = 0$ and $\vec{r}_i = \vec{r}_j$, $P_{spatial}(synapse; i, j, t) =$

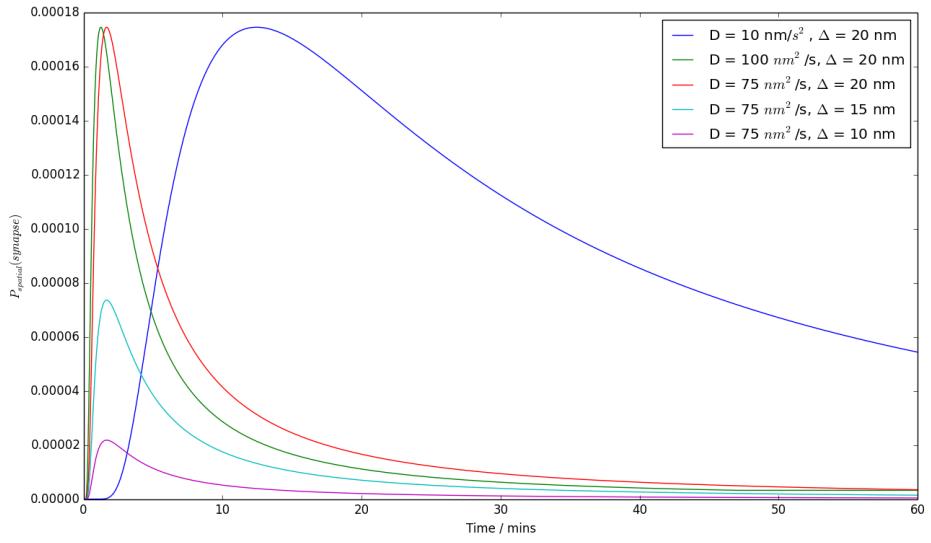


Fig. 6.8. Function of $P_{spatial}(synapse)$ against time for different parameters of Δ and D for a DSB ends of separation 300 nm

1 which implies that if two DSB ends lie at the same position, then the probability of overlap is one during the very first moment. This result is sensible as at $t = 0$ seconds, the DSB ends have not diffused away from their point of formation. The graph of the $P_{spatial}(synapse; i, j, t)$ function for different initial distance between DSB ends and parameters are shown in Fig. 6.8 and Fig. 6.9. Fig. 6.8 shows the variation of $P_{spatial}(synapse; t)$ with different parameters of D and Δ . As the diffusion constant D increases, the maximal overlap probability between DSB ends occurs at a shorter time and as Δ threshold distance increases, the overall overlap probability decreases. Fig. 6.9 shows the variation of $P_{spatial}(synapse; t)$ (omitting i and j for convenience) with different distance between the DSB ends using a fixed parameters of $D = 100 \text{ nm}^2/\text{s}$ and $\Delta = 20 \text{ nm}$. As the distance between the DSB ends decreases, the magnitude of overlap probability decreases and the maximal overlap probability occurs at a longer time. In general, one can observe that $P_{spatial}(synapse; t)$ increases rapidly to a maximal overlap probability and decay slowly to zero.

Eq. (6.9) is important in determining the misrepair rates, because two

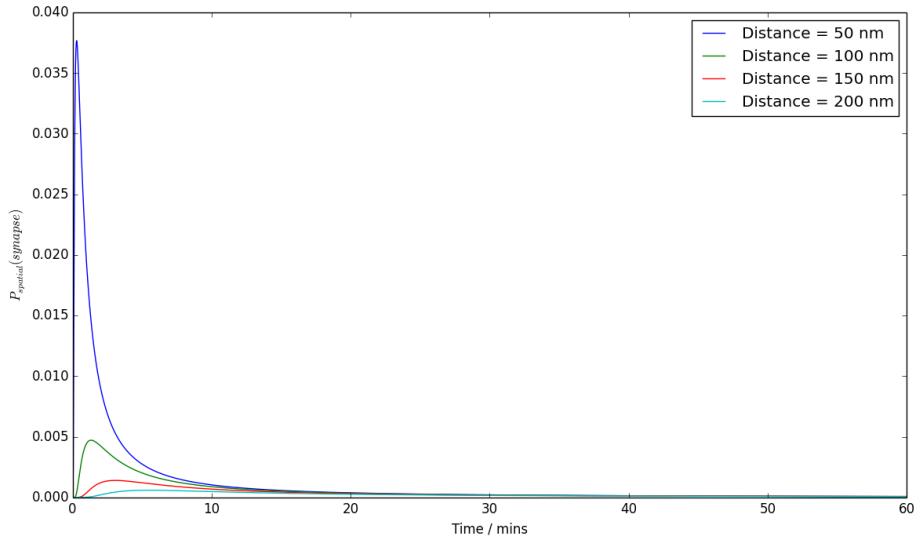


Fig. 6.9. Function of $P_{spatial}(\text{synapse})$ against time for different distance between a pair of DSB ends for $D = 100 \text{ nm}^2/\text{s}$ and $\Delta = 20 \text{ nm}$.

different DSBs free end of different origin can diffuse far away from their original position to rejoin with each other which will constitute a misrepair. Next, to complete the algorithm for the time evolution of DSB state, we need to define the rate constant that is associated with *diffusion* in Fig. 6.1. To do so, we need to make an assumption about the spatial rejoicing process as mentioned above. It is assumed that the probability that 3 or more DSB free ends existing within Δ distance from each other at a common time t is negligible which is reasonable for low LET radiation. Furthermore, such a scenario is more plausible during the initial time when DSB free ends are close together, but in this scenario they would not have recruited the relevant repair proteins to form the synapse. This allows us to consider only *pairwise interaction* between free ends. Thus, the probability that free end i will form a synapse at time t is given by Eq. (6.9) which uses the expression from Eq. (6.8).

$$P(\text{synapse}; i, t) = \sum_{j \neq i}^{\infty} P(\text{synapse}; i, j, t) \quad (6.10)$$

Hence, using Eq. (6.10), the transformation from $P_3(i, t)$ to $P_3(i, \Delta t)$

for both simple or complex DSB free ends in Fig. 6.1 are given in Eq. (6.11) and (6.12)

$$P_3(i, t + \Delta t) = P(\text{synapse}; i, t + \Delta t) \quad (6.11)$$

$$= \left[1 - \sum_{j \neq i}^{\infty} P_{\text{spatial}}(\text{synapse}; i, j, t) P_2(j, t) \right] P_2(i, t) \quad (6.12)$$

This completes the mathematical formulation of the time evolution of the simple and complex DSB states. There is a total of 11 free parameters in this model - 8 rate constants ($3 s_i$ and $5 c_i$) and Δ , D and k_{synapse} constants in diffusion modelling. This is a lot lesser than other models in Fig. 6.6, but still maintain the complexity and universality of the NHEJ repair modelling. This forms the foundational mathematical framework of the NHEJ repair model. The refinements to this model for greater biological accuracy will be covered in the next section.

Model Refinements

The first refinement stems from the existence of chromosomal territories [99] where each chromosome occupies a certain volume of the nucleus and does not fill the entire nucleus. Thus, the DSB ends within a chromosome should not diffuse far beyond the boundary of the chromosome or more precisely, $f(r, r_i, t)$ decreases rapidly to zero beyond the boundary of each chromosome. To implement this constraint into the current NHEJ framework, we restrict the standard deviation, σ of the DSB end diffusion to be less than the distance between the point of formation of the i -th DSB and the chromosome boundary, d_i . This is implemented as

$$\sigma = \min(\sqrt{2Dt}, d_i). \quad (6.13)$$

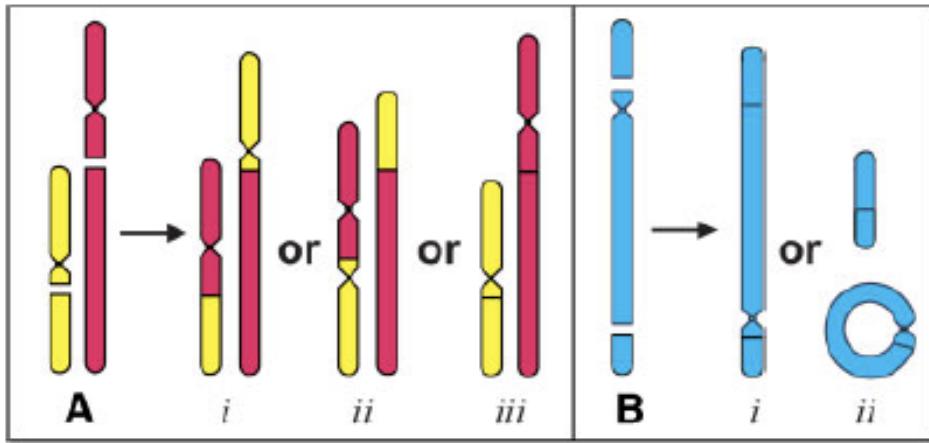


Fig. 6.10. This figure shows the possible chromosome aberration patterns that can arise from DSBs in one or two chromosomes. This figure is reprinted from [7] with the permission of John Wiley and Sons.

6.3.3 Chromosome Aberration Model

As mentioned in Section 6.3.2, the chromosome aberration model is closely linked to the diffusion section of the NHEJ model. Chromosome aberration arises due to mis-rejoining of DSB in the local region [138, 156]. This mis-joining depending on whether it happens *within* or *between* chromosomes, can give rise to several different configurations of chromosome as shown in Fig. 6.10 (adapted from Ref. [7]). The possible chromosome aberration patterns for two chromosomes (left figure of Fig. 6.10) are i) simple translocation, ii) Dicentric and acentric (highly lethal configuration) and iii) double restitution (point mutation at worst), whereas for a single chromosome (right figure of Fig. 6.10) one can have i) pericentric inversion and ii) a centric ring and acentric fragment (high lethal configuration). These chromosome aberrations are possible to *identify* and *quantify* during experiment using the Fluorescence In-Situ Hybridization (FISH) techniques [157, 158, 159, 156]. Due to the lethality of certain chromosome aberrations, it is more directly linked to cell viability after irradiation compared to the initial DSB breaks spectrum.

There has been some works on the modelling of chromosome aberra-

tion, albeit not as extensive as the DNA repair modelling described in Section 6.3.2. The most prominent modelling work were done by Dr. R. K. Sachs [160, 161, 162, 7] in the early 2000s and Dr. F. Ballarini [68, 69, 70, 163]. Dr. R. K. Sachs tested various hypothesis of DSB repair and radiation quality with his *discrete time Markov process based* model and Dr. F. Ballarini uses the concept of *clustered lesions* and realistic chromosome geometry within the nucleus to quantify the various chromosome aberration patterns and the cell survival subsequently for hadron and X-ray therapy. There are also more recent modelling efforts in Refs. [164, 165] which uses less realistic modelling environment compared to the previous two authors but still possess predictive value. Despite the realism involved in the chromosome aberration modelling, all the simulations mentioned above did not use a *mechanistically-derived* initial DSB breaks spectrum. The only simulation program which uses the *mechanistically-derived* DSB breaks for subsequent chromosome aberrations calculation is PARTRAC [67], and now IRSAGS. The rest of this section will be devoted to outlining the methodology of chromosome aberration modelling in IRSAGS.

Misrepair rates Before stratifying the misrepair into different types of aberrations, we first defined the average misrepair and correct repair probability for DSB end i after radiation damage as

$$P(\text{Misrepairs}; i, t) = \int_0^t \sum_{j \neq i'} P(\text{synapse}; i, j, t) dt, \quad (6.14)$$

and

$$P(\text{Repairs}; i, t) = \int_0^t \sum_{j=i'} P(\text{synapse}; i, j, t) dt, \quad (6.15)$$

where i and i' are conjugate DSB ends. If they are rejoined together it will constitute correct repair, otherwise it will be a mis-repair. In practice, the misrepairs and correct repair probability are defined when $t \rightarrow \infty$.

Chromosome Aberrations The mis-repaired DSB ends determined from Eq. (6.14) will be further analyzed to determine the resulting spectrum (and distribution) of chromosome aberrations. There are two assumptions required to quantify the aberration distributions:

1. The *centromere* is assumed to be at the centre of the genomic index of each chromosome, dividing the chromosome into two arms. Thus, if two DSB occur on one side of the arm and misrepaired, it will constitute an *intra-arm exchange*. Alternatively, if two DSB occur on opposite sides of the arms and mis-repaired, it will be an *inter-arm exchange*.
2. Except for DSB fragments that were too small to be repaired and were removed from NHEJ model (discussed in Section 6.3.1), all DSB ends were assumed to be rejoined either correctly or incorrectly. It is important to note that there will always be an *even* number of DSB ends.

To determine the aberration distribution, the only information needed is $P(\text{synapse}; i, j, t)$ for $t \rightarrow \infty$. If there are N number of DSB ends, there will be $N(N - 1)/2$ unique $P(\text{synapse}; i, j, t)$ quantities. These quantities can be used to determine the probability of resulting in any particular repaired configuration $\{C_i\}$, where i refers to the i -th configuration. The configuration $\{C_i\}$ consists of $N/2$ pairs of DSB ends such as $\{(1, 3), (2, 5), (9, 4)\dots\}$, where $(1, 3)$ implies DSB end 1 is rejoined to 3. In IRSGS, conjugate DSB ends occur consecutively. For instance, $(1, 2)$ and $(3, 4)$ are both conjugate DSB ends. Thus, the configuration $\{C_i\}$ provides all the required information to determine the mis-repair rates as well as the chromosome aberration pattern in that configuration. The probability of forming that particular configuration is given by multiplying all the relevant $P(\text{synapse}; i, j, t)$ terms corresponding to each rejoined pairs in $\{C_i\}$. For instance,

$$\begin{aligned}
P(\{C_i\}) &= P(\{(1, 2), (3, 6), (4, 5)\}) \\
&= P(synapse; 1, 2, \infty) \times P(synapse; 3, 6, \infty) \\
&\quad \times P(synapse; 4, 5, \infty)
\end{aligned}$$

The total number of configurations is given by

$$\text{Total Configurations} = \frac{N!}{2^{N/2}}. \quad (6.16)$$

The time complexity of any algorithm attempting to sample each configuration is $O(N!)$ which is definitely not scalable for large number of DSB breaks. However, the saving grace is that most of $P(synapse; i, j, \infty)$ is very small especially for *non-conjugate* DSB ends and can be neglected, thus the number of configurations to be considered is much smaller than given in Eqn. (6.16). The chromosome aberration patterns can then be determined by analysing $\{C_i\}$ and the probability is determined from $P(\{C_i\})$.

6.3.4 Results and Discussions

The NHEJ model requires 11 constants. Some constants are fixed by using information from literature while others are determined from fitting with experimental data published in literature. Before presenting on how we arrived at the values of the parameters, we vary some parameters in the model to understand its impact on the DSB repair rate. These results are shown in Fig. 6.11 and Fig. 6.12, where the vertical axis represents the amount of un-repaired DSB ends given by:

$$P(DSB; t) = \frac{\left(\sum_{i=1}^{N_{simple}} P_5(t) + \sum_{i=1}^{N_{complex}} P_7(t) \right)}{N_{simple} + N_{complex}}. \quad (6.17)$$

Figure A of Fig. 6.11 shows the effect of varying the diffusion constant of

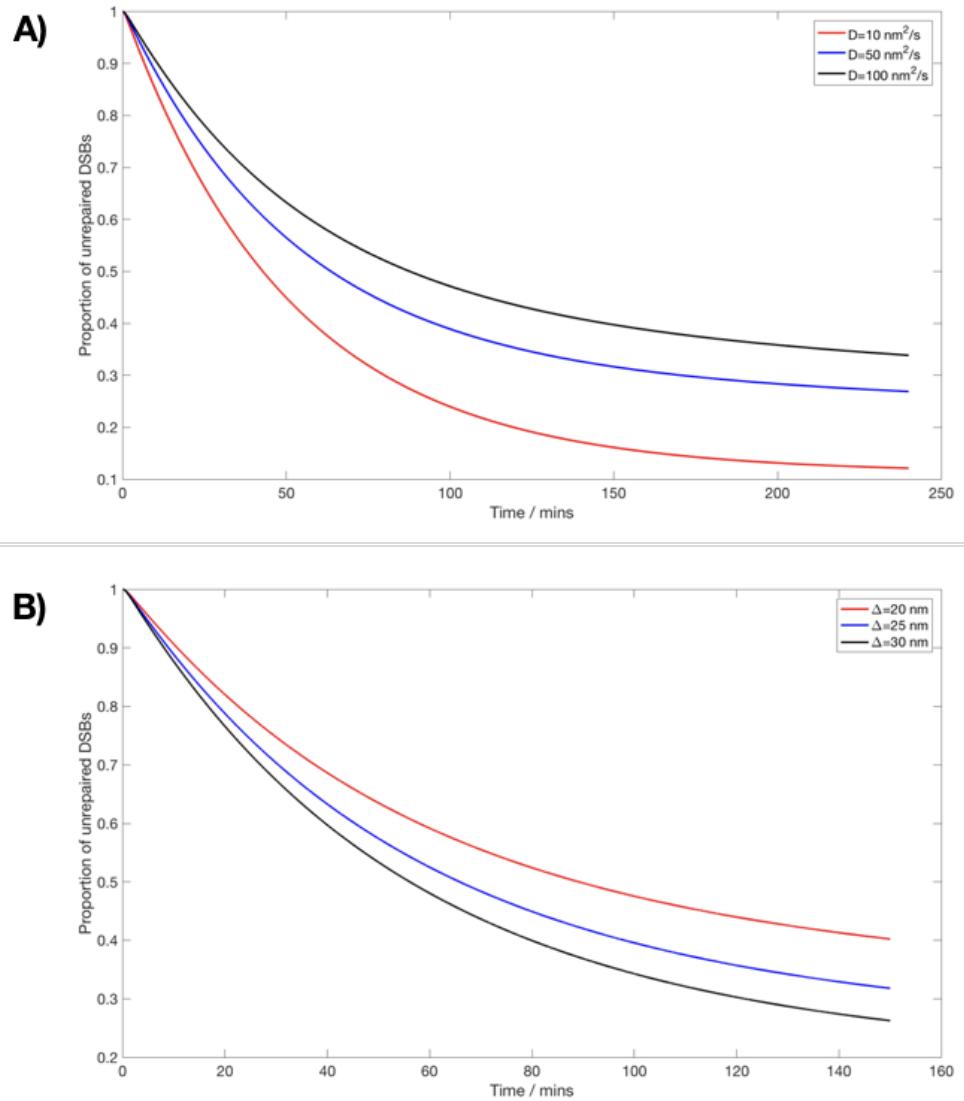


Fig. 6.11. A) Comparison of the rate of repair of DSBs for different diffusion constants in the NHEJ repair model. B) Comparison of the rate of repair of DSBs for different Δ constants in the NHEJ repair model.

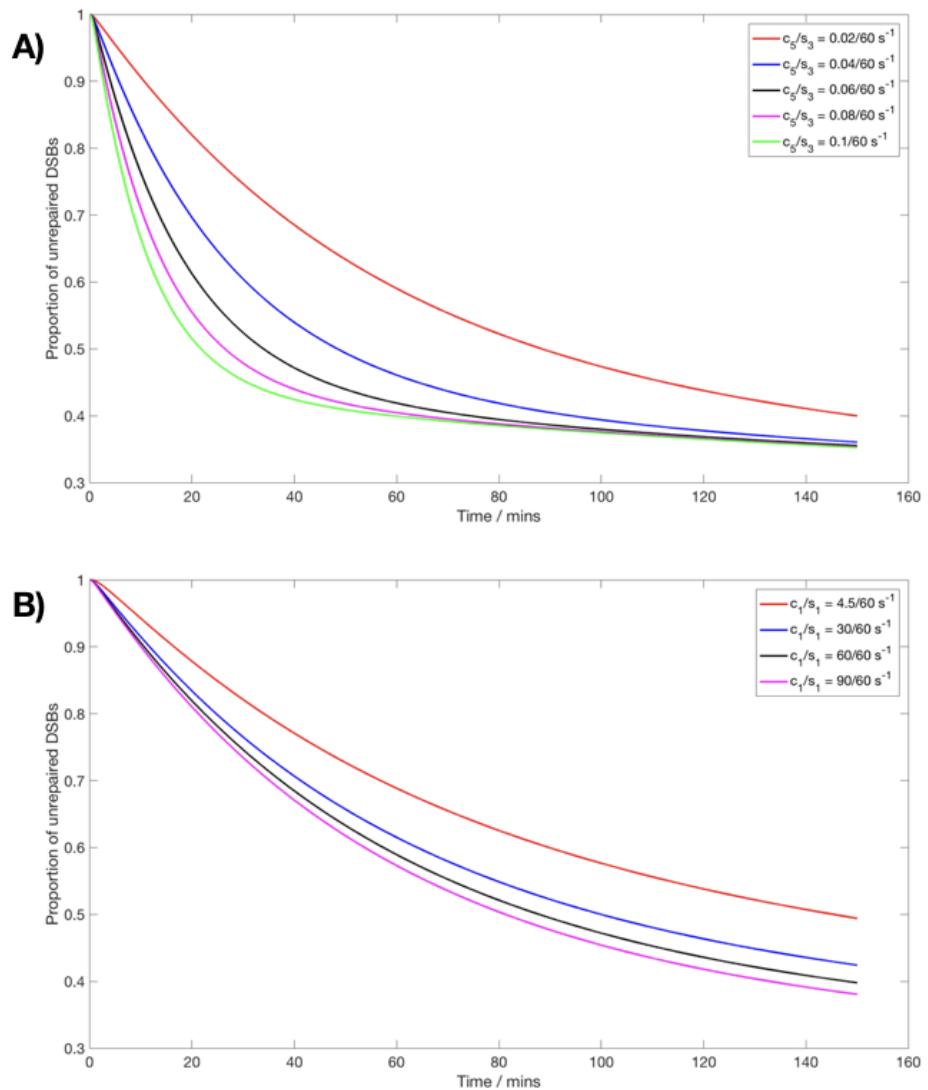


Fig. 6.12. A) Comparison of the rate of repair of DSBs for different rate constants c_5/s_3 (last reaction) in the NHEJ repair model. B) Comparison of the rate of repair of DSBs for different rate constants c_1/s_1 (first reaction) in the NHEJ repair model.

the DSB free end while figure B shows the effect of varying the DSB overlap distance Δ . Both parameters have an impact on the proportion of *residual DSBs* indicated by the vertical axis value at $t \rightarrow \infty$. This represents the proportion of DSBs which can never be repaired. It can be seen that the proportion of residual DSBs decreases with decreasing diffusion constant and increasing Δ . This can be explained by noting that with lower diffusion constants, the DSB end will remain localised for a longer period of time thus allowing higher probability of repair with conjugate DSB end. The increased Δ value results in higher probability of repair as the DSB end do not need to approach very closely for repair to occur. Figure A of Fig. 6.12 shows the impact of varying the reaction constant c_5 and s_3 of the NHEJ model. These constants govern the last step of the repair reactions of simple and complex DSB in Fig. 6.7 where Ligase IV and XRCC4 detached to complete the repair. Figure B of Fig. 6.12 shows how the c_1 and s_1 constants affect the DSB repair rate and they represent the first step of the repair reactions. The c_5 and s_3 constants *does not* affect the proportion of residual DSBs, but affect the rate of reaching this proportion. A higher c_5 and s_3 constants values imply that the DSB will be repaired more quickly and reach the residual DSBs value faster. Finally, varying c_1 and s_1 constants *does* affect the proportion of residual DSBs, as a higher rate constants imply that the DSB end will quickly recruit the required proteins and be *eligible* for synapse formation with its conjugate end, thus resulting in lower proportion of residual DSBs. These 4 parameters are chosen as a representative case, and the other parameters behave very similarly. For example, $k_{synapse}$ affect the DSB repair rate in the same way as Δ and σ .

The determination of the 11 parameters in the NHEJ model is plagued with two problems. The first is due to *degeneracy* issue where more than one set of parameters can be fitted to the data when there are not enough independent data sets to uniquely constraint the 11 parameters. The sec-

ond is the long amount of time required to determine the ideal set of parameters that can be fitted to a data set as this involves optimisation in 11 dimensional space. To circumvent these issues, we do not determine the parameters from scratch but use the parameters that were published in the literature as a guide. We fix parameters c_2 , c_3 , c_4 , s_2 according to values advised in Ref. [142]. Thus, the values of the 4 parameters are as follows:

$$\begin{aligned} s_2 &= \frac{2.5155}{60} s^{-1} \\ c_2 &= \frac{4.2257}{60} s^{-1} \\ c_3 &= \frac{4.5}{60} s^{-1} \\ c_4 &= \frac{2.7559}{60} s^{-1} \end{aligned}$$

The rest of the 7 parameters are determined from fitting the simulation data for DSB repair rate to the data obtained from 40 Gy of Ce-137 irradiation data from Ref. [166]. It is further assumed that the initial repair protein attachment process and the final protein detachment process are governed by the same rate constants for both simple and complex DSBs. Hence, $c_1 = s_1$ and $c_5 = s_3$. This assumption is also used in Ref. [142]. Furthermore, it has been reported in literature that the diffusion constant of the DSB free end σ , is between $100 \text{ nm}^2/\text{s}$ to $200 \text{ nm}^2/\text{s}$ [140, 141], and this range was used to constrain the diffusion constant parameter. The IRSGS program is first run with 40Gy of Ce-137 X-ray (using the method outlined in Section 6.2.5) to generate the distribution of DSBs for input into the NHEJ model. The IRSGS program predict a total of 2421 number of DSBs (4842 DSB ends) which is comparable to that predicted by PARTAC which is 2228 DSBs [141]. The result of the fitting of the NHEJ model to the experimental result is shown in Fig. 6.13. From this fit, the remaining parameters are currently estimated to be:

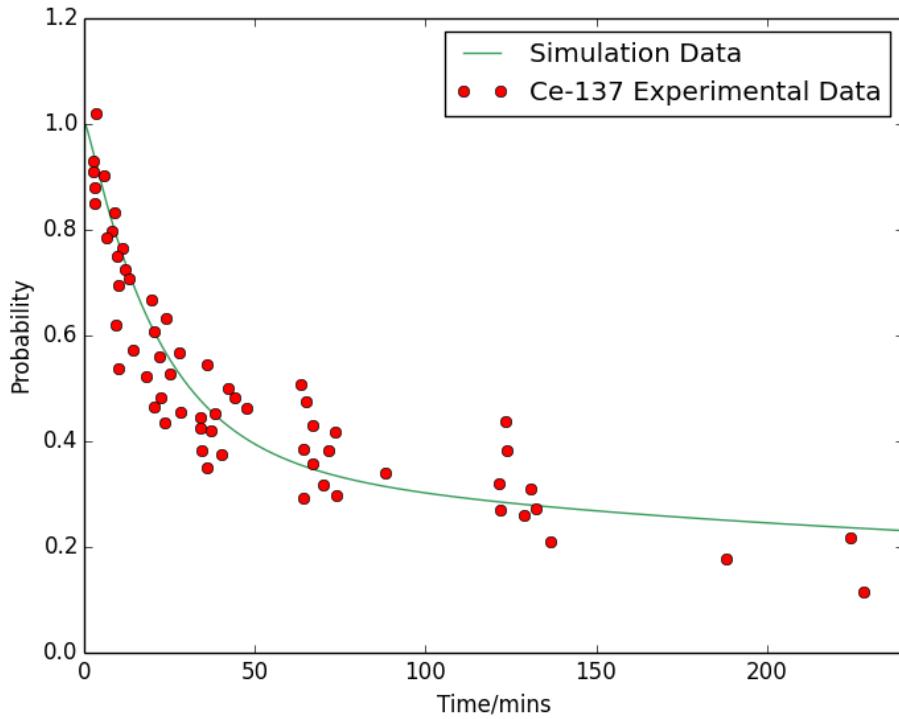


Fig. 6.13. This figure shows the experimental data of fraction of DSB within a nucleus after 40 Gy of Ce-137 irradiation and the simulation data fitting result.

$$\sigma = 100 \text{ nm}^2/\text{s}$$

$$\Delta = 200 \text{ nm}$$

$$k_{synapse} = 2.0 \text{ s}^{-1}$$

$$c_1 = s_1 = 0.5 \text{ s}^{-1}$$

$$c_5 = s_3 = \frac{0.05}{60} \text{ s}^{-1}$$

These parameters will be updated with future testing and fitting with other independent data sets containing various biological endpoints.

With the parameters of NHEJ model determined, we will now examine the number of correctly repaired and mis-repaired DSBs. This part of the work is carried out using 10 Gy of Ce-137 irradiation to compare with data from Ref. [167]. The calculation follows from Eq. (6.14) and (6.15), and

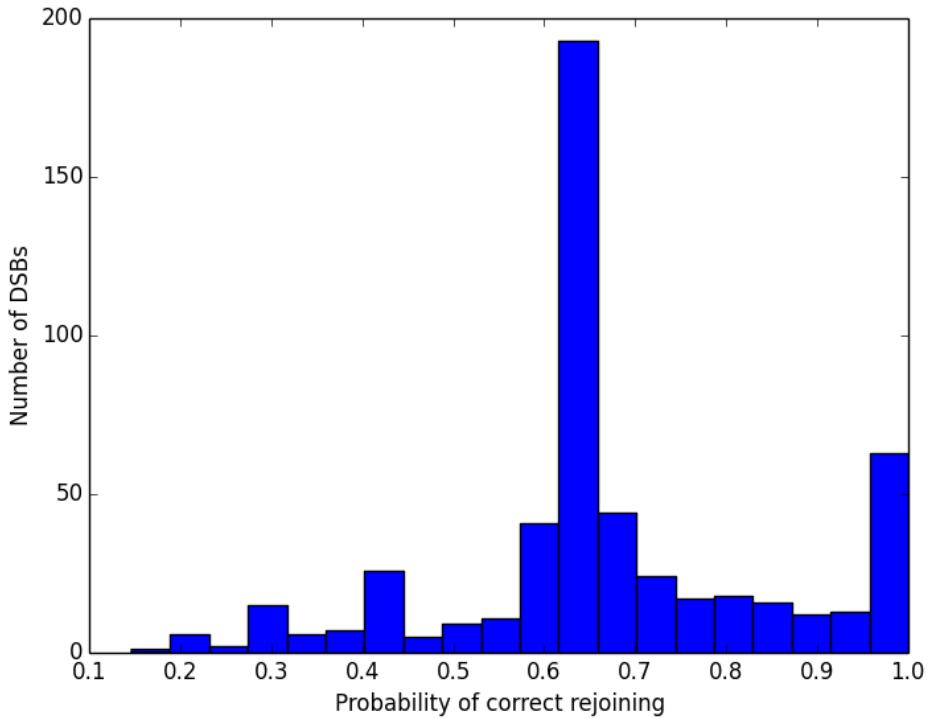


Fig. 6.14. This figure shows the histogram of the probability of correct rejoining for DSBs formed under 10 Gy of Ce-137 irradiation.

the result of the proportion of DSBs that were correctly repaired is shown in Fig. 6.14. As can be seen from the figure, most DSBs undergo correct rejoining with an average probability of 69.8%. Furthermore, Ref. [167] reports 0.02 misrejoined DSB / 3.2 Mbp, and using our model, we obtain 0.0301 misrejoined DSB / 3.2 Mbp, which agrees well with the experimental data.

In conclusion, we have presented the mathematical frame of the NHEJ model and the chromosome aberration model and determined the parameters of the NHEJ model from a combination of values published in the literature and fitting to experimental data. The NHEJ model result is then analysed to examine the proportion of mis-repair of DSB which agrees with the experimental result from Ref. [167]. Hence in this chapter, we have presented the implementation of the entire simulation pipeline from the initial energy deposition to initial formation of DSB spectrum to the NHEJ re-

pair and finally to the chromosome aberration. The simulation results are compared to experimental results at various stages and show good predictive ability. Further work and effort is definitely required in the future to fine-tune the models and parameters to enhance its robustness and utility.

6.4 Summary

1. The aim of this chapter is to present the development of IRSGS to simulate the DSB production and repair after proton and X-ray irradiation. The development and conception of the framework and algorithm is novel in two ways. First, no other author has implemented the entire geometry, transport, DSB calculation and NHEJ repair process in the same way as IRSGS. Second, apart from PARTRAC, no other radiobiology simulation offers the same level of details in the mechanistic simulation as IRSGS. This detailed simulation could allow the testing of a wider range of hypotheses involving the physical to the biological stages of radiation damage.
2. The algorithm for determining initial DSB break spectrum together with direct and indirect contributions are discussed in Section 6.2. This algorithm is then used with X-rays and protons to determine the DSB spectrum arising from different radiation source. Proton irradiation has a higher DSB yield compared to X-rays and the total yield increases with decreasing proton energy. The indirect contribution to strand breaks us always higher than the direct contribution for both protons and photons.
3. This thesis develops a novel algorithm for modelling the NHEJ repair process. It uses the probabilistic approach where each DSB end is defined by a state vector containing probability of being in a particular biological state.

4. The model for the quantification of the chromosome aberration is presented using information on the DSB mis-repairs from the NHEJ repair model.
5. Both the NHEJ and chromosome aberration model is tested on X-ray data.

Chapter 7

Conclusion and Future Outlook

7.1 Conclusion

This thesis consists of two parts. The first covers the work involved in the developing and setting up of the DAQ system for the new radiobiology beamline, and the second covers the radiobiology simulation to understand quantitatively the biological damage induced by protons and photons. In this work, we have shown that the DAQ system for the radiobiology beamline has been successfully implemented on the beamline at CIBA. The software is able to carry out STIM imaging, automatic calibration between Scanning voltage and Pixel positions, and manual targeting. These functions have been proven in experiments to work very well. The second part involves the radiobiology simulation for modeling and quantifying the DSB spectrum from X-rays and protons. This simulation has been designed around a framework incorporating a physical to a biological process in a way which is different from other existing software. This approach gives the benefits of increased realism (but using atomic DNA geometry and implementing a diffusion model into the NHEJ repair model) as well as increased

versatility (using GEANT4 for the transport code to allow the simulations to have access to the latest and most accurate cross section data). The IRSGS simulation codes ranging from the DNA construction, DSB scoring and NHEJ repair modeling have been developed as part of this PhD thesis. IRSGS which uses different transport models and geometry from other simulation, is found to agree well other program and shows that X-rays and protons have different direct and indirect contributions. The DSB output is then inputted into the novel NHEJ model in this work to determine the repair and misrepair rates for all the DSB. This is preliminary work but shows promising initial results with X-rays.

7.2 Future Outlook

This thesis shows the development and testing of the DAQ software and hardware with considerable success, but as mentioned in Section 3.4, further work is required to expand the DAQ software to its maximum capability in experiment. A more detailed view of the further work is:

1. The GUI is an important part of any software development. Depending on the workflows required by the experiment, the GUI needs to be improved to ensure the user can operate the software smoothly.
2. To achieve high throughput cell targeting, two important features need to be included in the DAQ software. The first is the automatic cell recognition ability and the second is the integration of the 3-axis stage control in the software to control the movement of the stage to different region of the cell targets for irradiation.
3. The current time-triggered manual targeting allows us to achieve *targeting* ability but the incident protons are still subjected to the Poisson statistics. To achieve *deterministic targeting*, the entire particle-triggered feature must be developed. Currently, the software aspect

of the *particle-triggered* feature is completed, but the detector hardware (scintillator or diamond detector-based) needs to developed in order to test this feature.

With regard to the IRSGS development, there is plenty of work to be done either in 1) increasing the biological realism of the simulation and extending beyond the cellular level to the tissue level, or 2) using the IRSGS to calculate quantities that are useful in a clinical setting. Several possible directions of future work are:

1. Employ more realistic chromosome geometry in the nucleus instead of a chromosome cube model.
2. There are several parameters in the NHEJ and chromosome aberration models that needs to be determined from *fitting* with experimental data. Thus, further effort is needed to optimize the parameters with existing literature data or newly published data.
3. The NHEJ and chromosome aberration models need to be executed on proton irradiation data to compare the efficacy of cell-killing between proton and X-rays.
4. There are various works in modelling the tumor tissue evolution using Cellular Automata approach [168] and this can be integrated with IRSGS to yield a mechanistic prediction on the radiation effect on tumor at the tissue level.
5. Use the chromosome aberration model in IRSGS to predict cell survival curve which will yield the α/β ratio [169] that is used clinically to determine the Relative Biological Effectiveness (RBE). IRSGS has been used to predict the RBE of proton based on DSB endpoints in a 1 dimension water volume and agree well with the current accepted clinical value of 1.1 [170].

6. Hypoxia [15] is an important consideration affecting radiotherapy efficacy and it usually occurs at the centre of the tumor due to less developed blood vessels system [117]. A hypoxia model needs to be included in IRSGS in the future to better understand and quantify the biological effects of radiation in realistic tissue environment.

Bibliography

- [1] E. Surdutovich. Multiscale approach to the physics of radiation damage with ions. *AIP Conference Proceeding*, 1525:672, 2013.
- [2] A. V. Solov'yov, E. Surdutovich, E. Scifoni, I. Mishustin, and W. Greiner. Physics of ion beam cancer therapy: A multiscale approach. *Phys. Rev. E*, 79:011909, 2009.
- [3] E. Surdutovich and A. V. Solov'yov. Multiscale approach to the physics of radiation damage with ions. *Eur. Phys. J. D*, 68:353, 2014.
- [4] S. Maharana, K. V. Iyer, N. Jain, M. Nagarajan, Y. Wang, and G. V. Shivashankar. Chromosome intermingling-the physical basis of chromosome organization in differentiated cells. *Nucleic Acids Res.*, 44(11):5148–60, 2016.
- [5] M. Karamitros, A. Mantero, S. Incerti, W. Friedland, et al. Modeling radiation chemistry in the geant4 toolkit. *Progress in NUCLEAR SCIENCE and TECHNOLOGY*, 2:503–508, 2011.
- [6] Y. Liang, G. Yang, F. Liu, and Y. Wang. Monte carlo simulation of ionizing radiation induced dna strand breaks utilizing coarse grained high-order chromatin structures. *Phys. Med. Biol.*, 61:445–460, 2016.

- [7] L. Hlatky, R. K. Sachs, M. Vazquez, and M. N. Cornworth. Radiation-induced chromosome aberrations: insights gained from biophysical modeling. *BioEssays*, 24:714–723, 2002.
- [8] https://www.moh.gov.sg/content/moh_web/home/statistics/Health_Facts_Singapore/Principal_Causes_of_Death.html, 2018.
- [9] <https://www.nccs.com.sg/ABOUTUS/OURFACILITIES/Pages/NewNCCSBUILDINGPROTONTHERAPYCENTRE.aspx1>, 2018.
- [10] W. D. Newhauser and R. Zhang. The physics of proton therapy. *Phys. Med. Biol.*, 60:R155, 2015.
- [11] <http://medicalphysicsweb.org/cws/article/opinion/71033>, 2018.
- [12] R. V. Sethi, H. A. Shih, and B. Y. Yeap. Second nonocular tumors among survivors of retinoblastoma treated with contemporary photon and proton radiotherapy. *Cancer*, 120:126–133, 2014.
- [13] C. S. Chung, T. L. Yock, K. Nelson, Y. Xu, N. L. Keating, and N. J. Tarbell. Incidence of second malignancies among patients treated with proton versus photon radiation. *Int J Radiat Oncol Biol Phys*, 87:46–52, 2013.
- [14] S. Girdhani, R. Sachs, and L. Hlatky. Biological effects of proton radiation: What we know and don't know. *Radiat. Res.*, 179(3):252–272, 2013.
- [15] M. Joiner and A. van der Kogel. *Basic Clinical Radiobiology*. CRC Press, 2009.

- [16] H. Paganetti. Relative biological effectiveness (rbe) values for proton beam therapy. variations as a function of biological endpoint, dose, and linear energy transfer. *Phys. Med. Biol.*, 59(22):R419–472, 2014.
- [17] M. Durante and A. A. Friedl. New challenges in radiobiology research with microbeams. *Radiat. Environ. Biophys.*, 50:335–338, 2011.
- [18] K. Prise and G. Schettino. Microbeams in radiation biology: Review and critical comparison. *RRadiat. Prot. Dosim.*, 143:335–339, 2011.
- [19] R. E. Zirkle and W. Bloom. Irradiation of parts of individual cells. *Science*, 117(80):487–493, 1953.
- [20] M. Folkard, B. Vojnovic, K. M. Prise, A. G. Bowey, et al. A charged-particle microbeam: I. development of an experimental system for targeting cells individually with counted particles. *Int. J. Radiat. Biol.*, 72:375–385, 1997.
- [21] G. Randers-Pehrson, C. Geard, G. Johnson, C. Elliston, et al. The columbia university single-ion microbeam. *Radiat. Res.*, 156:210–214, 2001.
- [22] P. Barbaret and H. Seznec. Advances in microbeam technologies and applications to radiation biology. *Radiation Protection Dosimetry*, 166:182–187, 2015.
- [23] V. Hable, C. Greubel, A. Bergmaier, P. Reichart, et al. The live cell irradiation and observation setup at snake. *Nucl. Instrum. Methods Phys. Res. Sect. B*, 267:2090?2097, 2009.
- [24] S. Bourret, F. Vianna, G. Deves, V. Atallah, et al. Fluorescence time-lapse imaging of single cells targeted with a focused scanning charged-particle microbeam. *Nucl. Instrum. Methods Phys. Res. Sect. B*, 325:27–34, 2014.

- [25] Y. Iwai, T. Ikeda, T. M. Kojima, Y. Yamazaki, et al. Ion irradiation in liquid of um3 region for cell surgery. *Appl. Phys. Lett.*, 92:023509, 2008.
- [26] X. F. Wang, J. Q. Li, J. Z. Wang, J. X. Zhang, et al. Current progress of the biological single-ion microbeam at fudan. *Radiat. Environ. Biophys.*, 50:353?364, 2011.
- [27] G. A. Drexler, C. Siebenwirth, S. E. Drexler, S. Girst, et al. Live cell imaging at the munich ion microbeam snake ? a status report. *Radiat. Oncol.*, 10:42, 2015.
- [28] D. W. M. Walsh, C. Siebenwirth, C. Greubel, K. Ilicic, et al. Live cell imaging of mitochondria following targeted irradiation in situ reveals rapid and highly localized loss of membrane potential. *Sci. Rep.*, 7:46684, 2017.
- [29] J. A. van Kan, P. Malar, and A. B. de Vera. The second generation singapore high resolution proton beam writing facility. *Rev. Sci. Instrum.*, 83:02B902, 2012.
- [30] T. K. Chan, F. Fang, A. Markwitz, and T. Osipowicz. Solid phase epitaxy of ultra-shallow sn implanted si observed using high-resolution rutherford backscattering spectrometry. *Appl. Phys. Lett.*, 101(8):081602, 2012.
- [31] M. D. Ynsa, M. Q. Ren, R. Rajendran, J. N. Sidhapuriwala, et al. Zinc mapping and density imaging of rabbit pancreas endocrine tissue sections using nuclear microscopy. *Microsc. Microanal.*, 15:345–352, 2009.
- [32] X. Chen, C. Chen, C. N. B. Udalagama, M. Q. Ren, et al. High-resolution 3d imaging and quantification of gold nanoparticles in a

- whole cell using scanning transmission ion microscopy. *Biophys. J.*, 104 (7):1419–1425, 2013.
- [33] Z. Mi, Y. Zhang, S. K. Vanga, C. Chen, H. Q. Tan, et al. Sub-wavelength imaging through ion-beam-induced upconversion. *Nat. Comm.*, 6:8832, 2015.
- [34] F. Watt, X. Chen, C. Chen, C. N. B. Udalagama, et al. Whole cell structural imaging at 20nm resolutions using mev ions. *Nucl. Instrum. Methods Phys. Res. Sect. B*, 306:6–11, 2013.
- [35] D. J. W. Mous, R. G. Haitsma, T. Butz, R.-H. Flagmeyer, D. Lehmann, and J. Vogt. The novel ultrastable hvee 3.5 mv singletron accelerator for nanoprobe applications. *Nucl. Instr. and Meth. B*, 130:31, 1997.
- [36] F. Watt, J. A. van Kan, I. Rajta, A. A. Bettoli, et al. The national university of singapore high energy ion nano-probe facility: Performance tests. *Nucl. Instr. and Meth. B*, 210:14–20, 2003.
- [37] G. W. Grime and F. Watt. *Beam Optics of Quadrupole Probe-Froming Systems*. Adam Hilger Ltd., 1984.
- [38] G. W. Grime. Wintrax: A raytracing software package for the design of multipole focusing systems. *Nucl. Instr. and Meth. B*, 306:76–80, 2013.
- [39] A. A. Bettoli, J. A. van Kan, T. C. Sum, and F. Watt. A labview-based scanning and control system for proton beam micromachining. *Nucl. Instrum. and Meth. B*, 181:49–53, 2001.
- [40] G. Lutz. *Semiconductor radiation detectors: Device Physics*. Springer, Berlin, 2007.

- [41] H. Spieler. *Semiconductor Detector Systems*. Oxford University Press, 2005.
- [42] Z. He. Review of the shockley-ramo theorem and its application in semiconductor gamma-ray detectors. *Nuclear Instrument and Method A*, 463:250–267, 2001.
- [43] P. Horowitz and W. Hill. *The Art of Electronics*. Cambridge University Press, 1989.
- [44] N. A. Sayko, A. G. Ponomarev, and A. A. Drozdenko. Beam scanning control and data acquisition on the sumy nuclear microprobe. *Nucl. Instrum. and Meth. B*, 89:223–228, 1994.
- [45] S. Aiello, A. Anzalone, M. Bartolucci, G. Cardella, et al. Data acquisition and real-time computing by a dsp-based system. *Nucl. Instrum. and Meth. B*, 136-138:1172–1176, 1998.
- [46] T. Sakai, T. Hamano, T. Hirao, T. Kamiya, K. Murozono, et al. Development of a fast multi-parameter data acquisition system for microbeam analyses. *Nucl. Instrum. and Meth. B*, 136-138:390–394, 1998.
- [47] G. W. Grime and M. Dawson. A pc-based data acquisition package for nuclear microprobe systems. *Nucl. Instrum. and Meth. B*, 260:101–104, 2007.
- [48] D. Zou, C. Ren, J. Tang, and F. Yang. An integrated computer system for fudan nuclear microprobe. *Nucl. Instrum. and Meth. B*, 104:119–123, 1995.
- [49] A. A. Bettoli, C. Udalagama, and F. Watt. A new data acquisition and imaging system for nuclear microscopy based on a field programmable gate array card. *Nucl. Instrum. and Meth. B*, 267:2069–2072, 2009.

- [50] C. G. Ryan, D. P. Siddons, G. Moorhead, R. Kirkham, et al. Large detector array and real-time processing and elemental image projection of x-ray and proton microprobe fluorescence data. *Nucl. Instrum. and Meth. B*, 260:1–7, 2007.
- [51] N. Wirth. *Digital Circuit Design An Introduction Textbook*. Springer, 1995.
- [52] D. M. Harris and S. L. Harris. *Digital Design and Computer Architecture*. Morgan Kaufmann, 2013.
- [53] T. J. Bress. *Effective LabVIEW Programming*. NTS Press, 2013.
- [54] S. McConnell. *Code Complete*. Microsoft Press, 2004.
- [55] K. Perajarvi, J. Keightley, J. Paepan, O. Tengblad, and H. Toivonen. *List-mode data acquisition based on digital electronics - State-of-the-art report*. Publications Office of the European Union, 2014.
- [56] Larry C. Andrews. *Special Functions of Mathematics for Engineers*. Oxford University Press, 1998.
- [57] J. Kapuscinski. Dapi: a dna-specific fluorescent probe. *Biotech. Histochem.*, 70(5):220–233, 1995.
- [58] L. J. Kuo and L. X. Yang. Gamma-h2ax - a novel biomarker for dna double-strand breaks. *In Vivo*, 22(3):305–309, 2008.
- [59] W. Z. Tu, B. Li, B. Huang, Y. Wang, X. Liu, et al. H2ax foci formation in the absence of dna damage: Mitotic h2ax phosphorylation is mediated by the dna-pkcs/chk2 pathway. *FEBS Letters*, 587:3437–3443, 2013.
- [60] J. E. Cleaver, L. Feeney, and I. Revet. Phosphorylated h2ax is not an unambiguous marker for dna double strand breaks. *Cell Cycle*, 10:3223–3224, 2011.

- [61] P. Rybak, A. Hoang, L. Bujnowicz, T. Bernas, K. Berniak, M. Zar?bski, et al. Low level phosphorylation of histone h2ax on serine 139 (h2ax) is not associated with dna double-strand breaks. *Oncotarget*, 7:49574–49587, 2016.
- [62] R. O. Duda and P. E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Comm. ACM*, 15:11–15, 1972.
- [63] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley-Interscience, 1998.
- [64] G. M. Phillips. *Interpolation and Approximation by Polynomials*. Springer-Verlag New York, 2003.
- [65] G. F. Knoll. *Radiation Detection and Measurement*. John Wiley and Sons Ltd, 2010.
- [66] W. Friedland, P. Jacob, and P. Kundrat. Mechanistic simulation of radiation damage to dna and its repair: On the track towards systems radiation biology modelling. *Radiat. Prot. Dosim.*, 143(2-4):542–548, 2011.
- [67] W. Friedland and P. Kundrat. Track structure based modelling of chromosome aberrations after photon and alpha-particle irradiation. *Mutat. Res.*, 756:213–223, 2013.
- [68] M. P. Carante, C. Aime, J. J. T. Cajiao, and F. Ballarini. Bianca, a biophysical model of cell survival and chromosome damage by protons, c-ions and he-ions at energies and doses used in hadrontherapy. *Phys. Med. Biol.*, 63(7):075007, 2018.
- [69] F. Ballarini and M. P. Carante. Chromosome aberrations and cell death by ionizing radiation: Evolution of a biophysical model. *Radiat. Phys. Chem.*, 128:18–25, 2016.

- [70] M. P. Carante, S. Alteri, S. Bortolussi, I. Postuma, N. Protti, and F. Ballarini. Modeling radiation-induced cell death: role of different levels of dna damage cluster. *Radiat. Environ. Biophys.*, 54:305–316, 2015.
- [71] <https://software.nasa.gov/software/MSC-25937-1>, 2018.
- [72] H. Nikjoo and P. Girard. A model of the cell nucleus for dna damage calculations. *Int. J. Radiat. Biol.*, 88(1-2):87–97, 2012.
- [73] H. Nikjoo, P. O’Nelli, W.E. Wilson, and D.T. Goodhead. Computational approach for determining the spectrum of dna damage induced by ionizing radiation. *Radiat. Res.*, 156:577–583, 2001.
- [74] H. Nikjoo, P. O’Nelli, M. Terrissol, and D.T. Goodhead. Quantitative modelling of dna damage using monte carlo track structure method. *Radiat. Environ. Biophys.*, 38:31–38, 1999.
- [75] R. Taleei, P. Girard, K. Sankaranarayanan, and H. Nikjoo. The non-homologous end-joining (nhej) mathematical model for the repair of double-strand breaks: II. application to damage induced by ultrasoft x rays and low-energy electrons. *Radiat. Res.*, 179(5):540–8, 2013.
- [76] M. Douglas, E. Bezak, and S. Penfold. Development of a radiation track structure clustering algorithm for the prediction of dna dsb yields and radiation induced cell death in eukaryotic cells. *Phys. Med. Biol.*, 60:3217–3236, 2015.
- [77] S. Meylan, S. Incerti, M. Karamitros, N. Tang, M. Bueno, I. Clairand, and C. Villagrasa. Simulation of early dna damage after the irradiation of a fibroblast cell nucleus using geant4-dna. *Sci. Rep.*, 7:11923, 2017.

- [78] W. Friedland, P. Jacob, H.G. Partzke, and T. Stork. Monte carlo simulation of the production of short dna fragments by low-linear energy transfer radiation using higher-order dna models. *Radiat. Res.*, 150:170–182, 1998.
- [79] D. Emfietzoglou, F. Cucinotta, and H. Nikjoo. A complete dielectric response model for liquid water: a solution of the bethe ridge problem. *Radiat. Res.*, 164(2):202–211, 2005.
- [80] J. Allison et al. Recent developments in geant4. *Nucl. Inst. Meth. A*, 835:186–225, 2016.
- [81] M. A. Bernal, D. Sikansi, F. Cavalcante, S. Incerti, C. Champion, V. Ivanchenko, and Z. Francis. An atomistic geometrical model of the b-dna configuration for dna?radiation interaction simulations. *Comput. Phys. Commun.*, 184:2840–2847, 2013.
- [82] S. Meylan, U. Vimont, S. Incerti, I. Clairand, and C. Villagrassa. Geant4-dna simulations using complex dna geometries generated by the dnafabric tool. *Comput. Phys. Commun.*, 204:159–169, 2016.
- [83] R. R. Sinden. *DNA structure and function (1st ed.)*. Academic Press, 1994.
- [84] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. Dickerson. Crystal structure analysis of a complete turn of b-dna. *Nature*, 287:755–8, 1980.
- [85] <http://w3dna.rutgers.edu/index.php/rebuild>. 2008.
- [86] T. J. Richmond and C. A. Davey. The structure of dna in the nucleosome core. *Nature*, 423:145–150, 2003.

- [87] K. Luger, A.W. Mader, R.K. Richmond, D.F. Sargent, and T.J. Richmond. Crystal structure of the nucleosome core particle at 2.8 angstrom resolution. *Nature*, 389:251–60, 1997.
- [88] <http://www.povray.org>, 2018.
- [89] W. Saenger. *Principles of Nucleic Acid Structure*. Springer-Verlag, 1984.
- [90] S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock, and T. Schlick. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc. Natl. Acad. Sci.*, 106(32):13317–13322, 2009.
- [91] J. Widom. Solenoidal model for superstructure in chromatin. *Annu. Rev. Biophys. Biophys. Chem.*, 18:365–395, 1989.
- [92] J. Finch and A. Klug. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci.*, 73:1897–901, 1976.
- [93] A. T. Annunziato. Dna packaging: Nucleosomes and chromatin. *Nature Education*, 1(1):26, 2008.
- [94] M. Tajik, A. S.H. Rozatian, and F. Semsarha. Calculation of direct effects of 60co gamma rays on the different dna structural levels: A simulation study using the geant4-dna toolkit. *Nucl. Instr. Meth. B*, 346:53 – 60, 2015.
- [95] J. Wilson. *Molecular biology of the cell: a problems approach*. New York: Garland Science, 2002.
- [96] D. O. Morgan. *The Cell Cycle: Principles of Control*. Oxford University Press., 2007.
- [97] http://vega.archive.ensembl.org/Homo_sapiens/Info/Index, 2018.

- [98] D. Emfietzoglou, I. Kyriakou, I. Abril, R. Garcia-Molina, I. D. Pet-salakis, H. Nikjoo, and A. Pathak. Electron inelastic mean free paths in biological matter based on dielectric theory and local-field corrections. *Nucl. Instr. Meth. Phys. B*, 267:45–52, 2009.
- [99] M. T. Meaburn. Chromosome territories. *Nature*, 445:379–781, 2007.
- [100] M. R. Branco and A. Pombo. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.*, 4(5):e138, 2006.
- [101] M. Karamitros, S. Luan, M. A. Bernal, J. Allison, et al. Diffusion-controlled reactions modeling in geant4-dna. *J. Comput. Phys.*, 274:841–882, 2014.
- [102] M. A. Bernal, M. C. Bordage, J. M. C. Brown, M. Davidkova, et al. Track structure modeling in liquid water: A review of the geant4-dna very low energy extension of the geant4 monte carlo simulation toolkit. *Phys. Med.*, 31:861–874, 2015.
- [103] O. N. Vassiliev. *Monte Carlo Methods for Radiation Transport*. Springer, 2017.
- [104] P. Sigmund. *Particle Penetration and Radiation Effects*. Springer, 2005.
- [105] E. B. Podgorsak. *Radiation Physics for Medical Physicists*. Springer, 2015.
- [106] M. Dingfelder, D. Hantke, M. Inokuti, and H. G. Paretzke. Electron inelastic-scattering cross sections in liquid water. *Radiat. Phys. Chem.*, 53:1, 1998.

- [107] M. Dingfelder, M. Inokuti, and H. G. Paretzke. Inelastic-collision cross sections of liquid water for interactions of energetic protons. *Radiat. Phys. Chem.*, 59:255, 2000.
- [108] M. S. Kreipl, W. Friedland, and H. G. Paretzke. Time- and space-resolved monte carlo study of water radiolysis for photon, electron and ion irradiation. *Radiat. Environ. Biophys.*, 48:11–20, 2009.
- [109] F. Ballarini, M. Biaggi, M. Merzagora, A. ottolenghi, et al. Stochastic aspects and uncertainties in the prechemical and chemical stages of electron tracks in liquid water: a quantitative analysis based on monte carlo simulations. *Radiat. Environ. Biophys.*, 39:179–188, 2000.
- [110] C. Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer, 2009.
- [111] P. A. M. Dirac. *The Principles of Quantum Mechanics*. Oxford University Press, 1958.
- [112] E. B. Tadmor and R. E. Miller. *Modeling Materials - Continuum, Atomistic and Multiscale Techniques*. Cambridge University Press, 2011.
- [113] N. Boccaro. *Modeling Complex Systems*. Springer, 2010.
- [114] E.L.Alpen. *Radiation Biophysics*. Academic Press, 1997.
- [115] A. Mozumder. *Fundamentals of Radiation Chemistry*. Academic Press, 1999.
- [116] R. Roots and S. Okada. Estimation of life times and diffusion distances of radicals involved in x-ray-induced dna strand breaks or killing of mammalian cells. *Radiat. Res.*, 64:306–20, 1975.

- [117] M. Joiner and A. Van der Kogel. *Basic Clinical Radiobiology*. CRC Press, 2009.
- [118] E. Surdutovich, A. V. Yakubovich, and A. V. Solov'yov. Biodamage via shock waves initiated by irradiation with ions. *Sci. Rep.*, 3:1289, 2013.
- [119] D. T. Goodhead, P. O'Neill, and H. G. Menzel. *Microdosimetry: an interdisciplinary approach*. The Royal Society of Chemistry, Cambridge, UK, 1997.
- [120] J.L. Bentley. Multidimensional binary search trees used for associative searching. *Comm. ACM*, 18(9):509, 1975.
- [121] E. Martin, K. Hans-Peter, S. Jorg, and X.W. Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. proceedings of the second international conference on knowledge discovery and data mining (kdd-96). *AAAI Press*, pages 226–231, 1996.
- [122] Ph. Bernhardt, W. Friedland, P. Jacob, and H. G. Paretzke. Modeling of ultrasoft x-ray induced dna damage using structured higher order dna targets. *Int. J. of Mass Spec.*, 223-224:579–597, 2003.
- [123] W. Friedland, P. Jacob, H.G. Paretzke, et al. Simulation of dna fragment distributions after irradiation with photons. *Radiat. Environ. Biophys*, 38:539–47, 1999.
- [124] W. Friedland, P. Jacob, Ph. Bernhardt, H.G. Paretzke, and M. Dingfelder. Simulation of dna damage after proton irradiation. *Radiat. Res.*, 159:401–10, 2003.
- [125] I. Gradzka and T. Iwanenko. A non-radioactive, pfge-based assay for low levels of dna double-strand breaks in mammalian cells. *DNA Rep.*, 4(10):1129, 2005.

- [126] J. A. Longo, B. Nevaldine, S. L. Longo, J. A. Winfield, and P. J. Hahn. An assay for quantifying dna double-strand break repair that is suitable for small numbers of unlabeled cells. *Radiat. Res.*, 147(1):35, 1997.
- [127] C. M. deLara, M. A. Hill, T. J. Jenner, D. Papworth, and P. O'Neill. Dependence of the yield of dna double strand breaks in chinese hamster v79-4 cells on photon energy of characteristic ultrasoft x rays. *Radiat. Res.*, 155:440–448, 2001.
- [128] S. W. Botchway, D. L. Stevens, M. A. Hill, T. J. Jenner, and P. O'Neill. Induction and rejoicing of dna double-strand breaks in chinese hamster v79-4 cells irradiated with characteristic aluminium k and copper l ultrasoft x rays. *Radiat. Res.*, 148:317–324, 1997.
- [129] D. Frankenberg, H. Brede, U. Schrewe, C. Steinmetz, M. Frankenberg-Schwager, G. Kasten, and E. Pralle. Induction of dna double-strand breaks by 1 h and 4 he ions in primary human skin fibroblasts in the let range of 8 to 124 kev ?m?1. *Radiat. Res.*, 151:539–47, 1999.
- [130] Z. Y. Tan, Y. Y. Xia, M. W. Zhao, and X. D. Liu. Proton inelastic mean free path in a group of bioorganic compounds and water in 0.05?10 mev range ? including higher-order corrections. *Nucl. Instr. Meth. Phys. B*, 268(14):2337–2342, 2010.
- [131] M. De Felici, R. Felici, M. S. del Rio, C. Ferrero, T. Bacarian, and F. A. Dilmanian. Dose distribution from x-ray microbeam arrays applied to radiation therapy: An egs4 monte carlo study. *Med. Phys.*, 32(8):2455, 2005.
- [132] Compton and H. Arthur. A quantum theory of the scattering of x-rays by light elements. *Physical Review*, 21(5):483–502, 1923.

- [133] O. Klein and Y. Nishina. Über die streuung von strahlung durch freie elektronen nach der neuen relativistischen quantendynamik von dirac. *Z. Phys.*, 52(11-12):853, 1929.
- [134] D. Griffith. *Introduction to Elementary Particles*. Wiley, 1987.
- [135] L. H. Thompson. Recognition, signaling, and repair of dna double-strand breaks produced by ionizing radiation in mammalian cells: The molecular choreography. *Mutat. Res.*, 751:158–246, 2012.
- [136] G. Giglia-Mari, A. Zottler, and D. L. Spector. Dna damage response. *Cold Spring Harb. Perspect. Biol.*, 3:a000745, 2011.
- [137] M. E. Lomax, L. K. Folkes, and P. O'Neill. Biological consequences of radiation-induced dna damage: Relevance to radiotherapy. *Clinical Oncol.*, 25:578–585, 2013.
- [138] M. Durante, J. S. Bedford, D. J. Chen, S. Conrad, M. N. Cornworth, A. T. Natarajan, D. C. van Gent, and G. Obe. From dna damage to chromosome aberrations: Joining the break. *Mutat. Res.*, 756:5–13, 2013.
- [139] I. Brandsma and D. C. Gent. Pathway choice in dna double strand break repair: observations of a balancing act. *Genome Integr.*, 3:9, 2012.
- [140] W. Friedland, P. Kundrat, and P. Jacob. Stochastic modelling of dsb repair after photon and ion irradiation. *Int. J. Radiat. Biol.*, 88(1-2):129–136, 2012.
- [141] W. Friedland, P. Jacob, and P. Kundrat. Stochastic simulation of dna double-strand break repair by nonhomologous end joining based on track structure calculation. *Radiat. Res.*, 173:677–688, 2010.

- [142] Y. Li, P. Reynolds, P. O'Neill, and F. A. Cucinotta. Modeling damage complexity-dependent non-homologous end-joining repair pathway. *PloS one*, 9(2), 2014.
- [143] Y. Li, H. Qian, Y. Wang, and F. A. Cucinotta. A stochastic model of dna fragments rejoicing. *PloS one*, 7(9), 2012.
- [144] R. Taleei, P. M. Girard, K. Sankaranarayanan, and H. Nikjoo. The non-homologous end-joining (nhej) mathematical model for the repair of double-strand breaks: II. application to damage induced by ultrasoft x rays and low-energy electrons. *Radiat. Res.*, 179(5):540–8, 2013.
- [145] R. Taleei and H. Nikjoo. The non-homologous end-joining (nhej) mathematical model for the repair of double-strand breaks: I. a mathematical model. *Radiat. Res.*, 179(5):530–9, 2013.
- [146] O. V. Belov, E. A. Krasavin, M. S. Lyashko, M. Batmunkh, and N. H. Sweilam. The non-homologous end-joining (nhej) mathematical model for the repair of double-strand breaks: I. a mathematical model. *J. Theor. Biol.*, 366:115–30, 2015.
- [147] S. J. McMahon, J. Schuemann, H. Paganetti, and K. M. Prise. Mechanistic modelling of dna repair and cellular survival following radiation-induced dna damage. *Sci. Rep.*, 6:33690, 2016.
- [148] F. A. Cucinotta, J. M. Pluth, J. Anderson, J. V. Harber, and P. O'Neill. Biochemical kinetics model of dsb repair and induction of gamma-h2ax foci by non-homologous end joining. *Radiat. Res.*, 169(2):214–222, 2008.
- [149] P. Reynolds, J. A. Anderson, J. V. Harper, M. A. Hill, S. W. Botchway, et al. The dynamics of ku70/80 and dna-pkcs at dsbs induced by

- ionizing radiation is dependent on the complexity of damage. *Nucleic Acids Res.*, 40(21):10821–10831, 2012.
- [150] P. O. Mari, B. I. Florea, S. P. Persengiev, N. S. Verkaik, H. T. Bruggenwirth, et al. Dynamic assembly of end-joining complexes requires interaction between ku70/80 and xrcc4. *Proceedings of the Nat. Acad. Sci. USA*, 103(49):18597–18602, 2006.
- [151] K. Yano, K. Morotomi, S. Y. Wang, N. Uematsu, K. J. Lee, et al. Ku recruits xlf to dna double-strand breaks. *EMBO Rep.*, 9(1):91–96, 2008.
- [152] R. Schwartz. *Biological Modeling and Simulation*. MIT Press, 2008.
- [153] B. P. Ingalls. *Mathematical Modeling in Systems Biology: An Introduction*. MIT Press, 2013.
- [154] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.
- [155] G. Du, G. A. Drexler, W. Friedland, C. Greubel, V. Hable, R. Krucken, et al. Spatial dynamics of dna damage response protein foci along the ion trajectory of high-let particles. *Radiat. Res.*, 176:706–715, 2011.
- [156] R. M. Anderson, D. L. Stevens, and D. T. Goodhead. M-fish analysis shows that complex chromosome aberrations induced by alpha particle tracks are cumulative products of localized rearrangements. *PNAS*, 99(19):12167–12172, 2002.
- [157] A.T. Natarajan and J. J. W. A. Boei. Formation of chromosome aberrations: insights from fish. *Mutat. Res.*, 544(2-3):299–304, 2003.

- [158] P. R. Langer-Safer, M. Levine, and D. C. Ward. Immunological method for mapping genes on drosophila polytene chromosomes. *Proc. Natl. Acad. Sci. U S A.*, 79(14):4381–4385, 1982.
- [159] B. D. Loucas and M. N. Cornworth. Complex chromosome exchanges induced by gamma rays in human lymphocytes: an mfish study. *Radiat. Res.*, 155:660–671, 2001.
- [160] R. K. Sachs, D. Levy, P. Hahnfeldt, and L. Hlatky. Quantitative analysis of radiation-induced chromosome aberrations. *Cytogenet. Genome Res.*, 104:142–148, 2004.
- [161] R. K. Sachs, D. Levy, A. M. Chen, P. J. Simpson, M. N. Cornworth, et al. Random breakage and reunion chromosome aberration formation model; an interaction?distance version based on chromatin geometry. *Int. J. Radiat. Biol.*, 76(12):1579–1588, 2000.
- [162] R. K. Sachs, D. Levy, P. Hahnfeldt, and L. Hlatky. Clustering of radiation-produced breaks along chromosomes: modelling the effects on chromosome aberrations. *Int. J. Radiat. Biol.*, 75(6):657–672, 1999.
- [163] M. P. Carante, C. Aime, J. J. T. Cajiao, and F. Ballarini. Bianca, a biophysical model of cell survival and chromosome damage by protons, c-ions and he-ions at energies and doses used in hadrontherapy. *Phys. Med. Biol.*, 63(7):075007, 2018.
- [164] A. L. Ponomarev, K. George, and F. A. Cucinotta. Generalized time-dependent model of radiation-induced chromosomal aberrations in normal and repair-deficient human cells. *Radiat. Res.*, 181:284–292, 2014.

- [165] Y. A. Eidelman and S. G. Andreev. Computational model of dose response for low-let-induced complex chromosomal aberrations. *Radiat. Prot. Dosi.*, 166(1-4):80–85, 2015.
- [166] K. H. Karlsson, I. Radulescu, B. Rydberg, and B. Stenerlow. Repair of radiation-induced heat-labile sites is independent of dna-pkcs, xrcc1 and parp. *Radiat. Res.*, 169:506–512, 2008.
- [167] B. Rydberg, B. Cooper, P. K. Cooper, W. R. Holley, and A. Chatterjee. Dose-dependent misrejoining of radiation-induced dna double-strand breaks in human fibroblasts: experimental and theoretical study for high- and low-let radiation. *Radiat. Res.*, 163:526–534, 2005.
- [168] A. Zuleyha, M. Ziya, Y. Selcuk, O. M. Kemal, and T. Mesut. Simulation of glioblastoma multiforme (gbm) tumor cells using ising model on the creutz cellular automaton. *Physica A*, 486:901–907, 2017.
- [169] H. Paganetti. Relative biological effectiveness (rbe) values for proton beam therapy. variations as a function of biological endpoint, dose, and linear energy transfer. *Phys. Med. Biol.*, 59(22):R419–72, 2014.
- [170] H. Q. Tan, K. W. Ang, and A. A. Bettoli. 3d quantification of biological damage for a 160 mev proton beam. *IFMBE proceedings*, 2019.

Appendix A

Derivation of $P_{spatial}(synapse)$

This appendix outlined the derivation of Eq. (6.9). The starting point of the derivation is Eq. (6.8) where

$$P_{spatial}(synapse; i, j, t) = \int f(\vec{r}; \vec{r}_i, t) \int_{\vec{r}-\Delta}^{\vec{r}+\Delta} f(\vec{r}'; \vec{r}_j, t) d^3\vec{r}' d^3\vec{r}. \quad (\text{A.1})$$

The definition of the variables in Eq. (A.1) is stated in Section 6.3.2. Assuming a Gaussian diffusion model for the DSB free end, we have

$$f(\vec{r}; \vec{r}_i, t) = \frac{1}{(4\pi Dt)^{3/2}} \exp\left(\frac{-|\vec{r} - \vec{r}_i|^2}{4Dt}\right). \quad (\text{A.2})$$

Assuming further that the X, Y and Z motion of the DSB free ends are decoupled from each other, the 3D integral in Eq. (A.1) can be simplified to a product of 1D integral as in:

$$\begin{aligned}
P_{spatial}(synapse; i, j, t) &= \int f(x; x_i, t) \int_{x-\Delta}^{x+\Delta} f(x'; x_j, t) dx' dx \\
&\times \int f(y; y_i, t) \int_{y-\Delta}^{y+\Delta} f(y'; y_j, t) dy' dy \\
&\times \int f(z; z_i, t) \int_{z-\Delta}^{z+\Delta} f(z'; z_j, t) dz' dz \\
&= \left[\int f(x; x_i, t) \int_{x-\Delta}^{x+\Delta} f(x'; x_j, t) dx' dx \right]^3.
\end{aligned}$$

The last equality stems from the fact that each of the integral in each 3 dimensions are equivalent. The problem is now reduced to evaluating the 1D integral as follows:

$$\int f(x; x_i, t) \int_{x-\Delta}^{x+\Delta} f(x'; x_j, t) dx' dx \quad (\text{A.3})$$

$$= \frac{1}{(4\pi Dt)^{1/2}} \int f(x; x_i, t) dx \int_{x-\Delta}^{x+\Delta} \exp\left(-\frac{(x' - x_j)^2}{4Dt}\right) dx' \quad (\text{A.4})$$

$$= \frac{1}{2} \int f(x; x_i, t) dx \left[\operatorname{erf}\left(\frac{x + \Delta - x_j}{\sqrt{4Dt}}\right) - \operatorname{erf}\left(\frac{x - \Delta - x_j}{\sqrt{4Dt}}\right) \right] \quad (\text{A.5})$$

$$= \frac{1}{2} \frac{1}{(4\pi Dt)^{1/2}} \int \exp\left(-\frac{(x - x_i)^2}{4Dt}\right) \operatorname{erf}\left(\frac{x + \Delta - x_j}{\sqrt{4Dt}}\right) dx \quad (\text{A.6})$$

$$+ \frac{1}{2} \frac{1}{(4\pi Dt)^{1/2}} \int \exp\left(-\frac{(x - x_i)^2}{4Dt}\right) \operatorname{erf}\left(\frac{x - \Delta - x_j}{\sqrt{4Dt}}\right) dx \quad (\text{A.7})$$

Each of the integral in Eq. (A.6) and (A.7) are carried out over $(-\infty, \infty)$. The integral of a product of error function and gaussian function can be simplified using the following identity

$$\int_{-\infty}^{\infty} \exp(-\alpha x^2) \operatorname{erf}(\beta x + \gamma) dx = \sqrt{\frac{\pi}{\alpha}} \operatorname{erf}\left(\gamma \sqrt{\frac{\alpha}{\alpha + \beta^2}}\right). \quad (\text{A.8})$$

By inserting Eq. (A.8) into Eq. (A.6) and (A.7), we obtain Eq. (A.10) which is equivalent to the term within the product in Eq. (6.9).

$$\int f(x; x_i, t) \int_{x-\Delta}^{x+\Delta} f(x'; x_j, t) dx' dx \quad (\text{A.9})$$

$$= \frac{1}{2} \left[\operatorname{erf}\left(\frac{x_i - x_j + \Delta}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x_i - x_j - \Delta}{\sqrt{2}}\right) \right]. \quad (\text{A.10})$$

This concludes the derivation of Eq. (6.9).