Monte Carlo modelling of Raman scattering in heterogeneous breast tissue

Laura Jane Moran

Submitted by Laura Jane Moran to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Physics, June 2022.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Signed: .....

Laura Jane Moran

Date: .....

09/06/2022

1st Supervisor: Tim Harries 2nd Supervisor: Nick Stone

Monte Carlo modelling of Raman scattering in heterogeneous breast tissue

This work was funded by EPSRC [EP/M506527/1].

Abstract

Breast cancer is the most common cancer for a woman to develop in her lifetime. By detecting breast cancer at an early stage, the symptoms can be easier to manage and the patient should have the best chance of survival. The current gold standard for breast cancer detection is a mammogram, followed by a biopsy and histopathology. This is effective but can also be expensive and invasive. A promising addition to the diagnostic pathway uses vibrational spectroscopy which utilises non-elastic interactions between light and tissue. Raman spectroscopy has been used widely in industry and research: it is a non-invasive and chemically specific technique. This spectroscopic technique has been proven to be applicable to the detection of microcalcifications in breast tissue to aid in diagnosing breast cancer and potentially reducing the number of biopsies required.

This thesis involves the development of algorithms to model Raman scattering in biological tissues to aid in the improvement of breast cancer detection. The technique used is the numerical modelling method Monte Carlo Radiative Transport (MCRT) to effectively simulate the transport of light through turbid media. There is a need for a fast and flexible code capable of modelling a variety of Raman source materials, tissue types and shapes, input laser beams and detectors. This rapid simulation of light transport through breast tissue can provide more information and insight to complement the practical measurements and analysis of experimental work, which can be used to improve future experiments and probes. By implementing physically correct Raman scattering into a fast and powerful code, and utilising work from the field to estimate the optical properties of tissues, simulations to supplement experimental work and predict potential clinical results are performed and analysed.

This thesis begins by verifying our Raman scattering algorithm against theoretical

and experimental results. Firstly, Chapter 3 shows the work undertaken to demonstrate that our tool is capable of producing the same results as a previously published Monte Carlo code, and the lessons learned from this. Chapter 4 is focused on comparing the simulation output to experimental output, in an investigation using tissue phantoms and a "semi-infinite" layer of Raman source material. Satisfied that `arctk` is performing as expected in returning the same results as another code and laboratory work, the code is then used for more theoretical work into how light propagates throughout a turbid volume. Chapter 5 presents a variety of parameters and how they affect the detected Raman signal: laser beam width, detector type and size, material(s) within the bulk volume. Finally, a link back to laboratory work allows us to show an estimated detection zone, where it should be feasible to identify microcalcifications in breast tissue. A conclusion is then given with comments on the future of this work and how the computational tool `arctk` could be used to continue the work in applying Raman spectroscopy to cancer research.

The results shown in this thesis are intended to improve understanding on how Raman photons behave in biological tissue, and how the geometry of the optical set-up can influence the detected signal. By increasing our knowledge of light transport in the tissue phantom and breast environment, the work here can begin to help design Raman probes to target biomarker signals. This thesis shows that there is real potential for clinical use of Raman spectroscopy for non-invasive breast cancer diagnosis for a more manageable and cost-effective treatment pathway.

Acknowledgements

I would like to begin by thanking my supervisors: Tim Harries and Nick Stone. Their unending support and belief in me made all of this possible. Tim, without your kindness and help and sleepless nights thinking about Raman photons I would not have got to the end. Nick, your enthusiasm, willingness to let me into the lab, and patience in teaching me all of the chemistry and biology I needed have made this work come together. It was a daunting task to take on a project based so deeply in fields I had never heard of and without the two of you always being there and having my back this would have never worked. A special mention to Pavel Matousek as well, for providing interesting discussions and helpful explanations.

The University of Exeter has given me the opportunity to meet some of the most incredible people. I have been lucky enough to spend my time in two research groups, and thus have had double the support, knowledge, and Christmas parties of an average PhD student. To Jenni, Ben, Ryan, Michelle, Doriana and the rest of BioSpec, thank you for taking in this odd astrophysicist who was afraid of the lab and teaching me so much. The patience of this group and their interest in my simulations was invaluable. To Exeter Astrophysics group: too numerous to name, but really everyone who came to pub, loved a crossword, endured a debate at coffee, played board games and watched movies, survived kayaking trips, indulged my love for a boogie... My experience would not have been the same without each of you, thank you.

My friends in Exeter and further afield are the most unbelievably supportive, hilarious, and loving people. Mich and Han, what would this have been without you both and your love? Our chats have got me through the roughest storms and you never fail to make me laugh. Simon: you're truly a rock in my life, I don't know what I would do without

you. Freddy, there's no one better I could have developed code with, and destroyed so many billions of photon packets with: your friendship is one of a kind. Angie and Anna, office 408 and the last four years wouldn't have been the same without you. Amy Rankine, thank you for always being a phone call away and forever my back up woah girl. Sophie, thank you for all of the cups of tea, trash TV, and always listening. Mairi, Hannah H, Amy A, Kirsty - thank you for being your wonderful selves.

To my loving family, moving to Exeter was not an easy decision, and thank goodness for video calls, because being so far away was more difficult than I could have imagined. Mum, Dad, and Kieran, you have always been my biggest supporters no matter what and there aren't words to express my gratitude for the safety you always provide me. To my Nan and Grandad, and Grandad and late Granny, thank you for calling me to check in when I have lost track of time, and for your love and conversations.

It's been an unbelievable chapter, and I'm so proud to have made it to the end.

Laura Jane Moran

Exeter, U.K.

9th June 2022

Contents

Abstract ii

Acknowledgements iii

Acronyms xi

List of Figures xv

List of Tables xxv

1 Introduction 1

1.1 Breast cancer . . . 1

1.1.1 Statistics . . . 1

1.1.2 Breast anatomy . . . 4

1.1.3 Cancer pathology . . . 5

1.1.4 Microcalcifications . . . 7

1.1.5 Diagnosis . . . 9

1.1.5.1 Early diagnosis . . . 9

1.1.5.2 Current diagnostic methods . . . 11

1.1.5.2.1 Clinical examination . . . 11

1.1.5.2.2 Mammography . . . 12

1.1.5.2.3 Ultrasound imaging . . . 13

1.1.5.2.4 Magnetic Resonance Imaging . . . 14

1.1.5.2.5 Biopsy and histopathology . . . 14

1.1.5.3 Treatment methods . . . 15

1.1.5.3.1 Surgery . . . 16
1.1.5.3.2 Adjuvant treatments . . . 16
1.1.6 Discussion . . . 17
1.2 Raman spectroscopy . . . 18
1.2.1 Theoretical background . . . 19
1.2.1.1 Light-matter interactions . . . 19
1.2.1.2 Raman scattering . . . 19
1.2.2 Instrumentation . . . 24
1.2.2.1 Light sources . . . 24
1.2.2.2 Optics . . . 25
1.2.2.3 Detectors . . . 26
1.2.3 Spectroscopic techniques . . . 27
1.2.3.1 Raman microscopy . . . 27
1.2.3.2 Deep Raman spectroscopy . . . 27
1.2.3.2.1 Temporal gating . . . 27
1.2.3.2.2 Spatial offset . . . 28
1.2.4 Raman spectroscopy in cancer diagnosis . . . 31
1.2.5 Summary . . . 32

2 Monte Carlo Modelling **34**
2.1 History . . . 35
2.1.1 Statistical sampling . . . 35
2.1.2 Neutron bombs . . . 39
2.1.3 Modern Monte Carlo in biological systems . . . 41
2.2 Fundamentals and mathematical framework . . . 43
2.2.1 Radiative transfer . . . 43
2.2.1.1 Beer-Lambert law . . . 47
2.2.1.2 Diffusion approximation . . . 48
2.2.2 Monte Carlo radiative transfer method . . . 48
2.2.2.1 The journey of a photon . . . 48
2.2.2.2 Light-matter interactions . . . 49
2.2.2.2.1 Scattering event . . . 52

2.2.2.2.2 Absorption event 53
2.2.2.3 Interface phenomena 53
2.2.2.4 Variance reduction 57
2.2.2.5 Measuring spatial variables 59
2.3 Our code 60
2.3.1 `arctk` 61
2.3.2 Code improvements 65
2.3.3 Implementing Raman scattering 65
2.4 Summary 66

3 Code Benchmarking 68
3.1 Simulation summary 69
3.2 The importance of definitions 73
3.3 Isotropic scattering 76
3.4 Conclusions 80

4 Monte Carlo modelling of Raman scattering in tissue phantoms 81
4.1 Introduction 82
4.2 Experimental setup and results 83
4.2.1 Materials and methods 83
4.2.2 Experimental outcomes 88
4.3 Monte Carlo setup and results 90
4.3.1 Monte Carlo method 90
4.3.2 Optical properties 93
4.4 Comparison 97
4.4.1 Results and discussion 97
4.5 Conclusions 100

5 Simulations of Raman scattering from microcalcifications in breast tissue 102
5.1 Introduction 102
5.2 Methods 103
5.3 Results 109
5.3.1 Homogeneous tissue 109

5.3.1.1 Pork tissue volume 109
5.3.1.1.1 Transmission Raman spectroscopy 110
5.3.1.1.2 Spatially offset Raman spectroscopy 111
5.3.1.2 Breast tissue results 113
5.3.1.2.1 Comparison to pork tissue 113
5.3.1.2.2 Transmission Raman spectroscopy 116
5.3.1.2.3 Spatially offset Raman spectroscopy 117
5.3.2 Heterogeneous tissue 119
5.3.2.1 Skin addition 120
5.3.2.2 Gland addition 124
5.4 Detection limit 126
5.4.1 Laboratory work 127
5.4.2 Simulation work 128
5.4.3 Comparison 130
5.5 Conclusion 134

6 The Conclusion **137**
6.1 Concluding remarks 137
6.2 Future work 140

Appendix: Output from PhD **145**

Bibliography **147**

Acronyms

BRCA1/BRCA2 BReast CAncer gene 1/BReast CAncer gene 2

DCIS Ductal Carcinoma In Situ

LCIS Lobular Carcinoma In Situ

TNM Tumour, Node, Metastasis

CO Calcium Oxalate

HAP HydroxyAPatite

NICE National Institute for health and Care Excellence

BSE Breast Self Examination

MRI Magnetic Resonance Imaging

FNA Fine Needle Aspiration

ER (O)Estrogen Receptors

HER2 Human Epidermal growth factor Receptor 2

IR InfraRed

NIR Near InfraRed

UV UltraViolet

CCD Charge-Coupled Device

SORS Spatially Offset Raman Spectroscopy

i-SORS inverse Spatially Offset Raman Spectroscopy

TRS Transmission Raman Spectroscopy

MC Monte Carlo

MCRT Monte Carlo Radiative Transport

PDF Probability Distribution Function

CDF Cumulative Distribution Function

ENIAC Electronic Numerical Integrator And Computer

MANIAC Mathematical Analyser Numerical Integrator and Automatic Computer

MCML Monte Carlo Multi Layered

RTE Radiative Transfer Equation

JSON5 JavaScript Object Notation 5

IL IntraLipid

PTFE PolyTetraFluoroEthylene

AR Anti-Reflective

VPH Volume Phase Holographic

BMI Body Mass Index

CSV Comma-Separated Values

SNR Signal to Noise Ratio

SERS Surface Enhanced Raman Spectroscopy

SESORS Surface Enhanced and Spatially Offset Raman Spectroscopy

List of Figures

1.1 Pie chart showing incidence of different cancer types in 2020 in both sexes. Four (lung, breast, bowel and prostate) make up more than 4 in 10 new diagnoses. Taken from Global Cancer Statistics.[1] 2
1.2 Mortality rate per 100,000 people in the UK from 1971-2018. Taken from Cancer Research UK.[2] 3
1.3 Diagram of female breast anatomy after puberty. Image and labels taken from Encyclopedia Britannica.[3] 4
1.4 Lymphatic drainage system and lymph nodes found in and around the breast. Image taken from Cancer Research UK[4] 5
1.5 Image of type I microcalcification from a sclerocystic mastopathy (benign). Taken from Frappart et al.[5] 8
1.6 Image of type II microcalcification from an infiltrating carcinoma. Taken from Frappart et al.[5] 8
1.7 Plot showing the age specific five year relative survival rate of breast cancer in adults diagnosed in 2000–2007. Taken from De Angelis et al.[6] 10
1.8 Mammogram illustration from Blausen Medical.[7] 13
1.9 Graphics displaying the procedure for (a) fine needle aspiration biopsy[8] and (b) core needle biopsy.[9] 15
1.10 Jablonski diagram of vibrational energy states of a molecule. The transitions shown are infrared absorption, Rayleigh, Raman Stokes, and Raman anti-Stokes scattering. $h\nu_0$ is incident laser energy, $h\nu_{\text{vib}}$ is vibrational energy, $\Delta\nu$ is frequency shift and ν_{vib} is vibrational frequency. Taken from Baker et al.[10] 21

1.11 A diagram showing the stretching and bending vibrational modes. Symmetric motions tend to be Raman-active and asymmetric movements are usually IR-active. Taken from Mendes and Duarte.[11] 22

1.12 Three pathways for an excited state to return to the ground state. (1) is non-radiative vibrational relaxation, (2) is fluorescence: short-lived photon emission and (3) is phosphorescence: long-lived photon emission. Taken from Cox.[12] 23

1.13 Simplified graphic to show the main components of a Raman spectroscopy setup. Taken from Sandeman.[13] 25

1.14 Schematic of Kerr gate experiment to probe subsurface layer Raman signals. Taken from Matousek et al.[14] 28

1.15 Diagram of (a) SORS and (b) i-SORS geometries. The illumination and collection zones are reversed from traditional SORS to perform inverse SORS. Δs is the spatial offset. 29

1.16 Simplified schematic of (a) SORS and (b) TRS set-ups. From Matousek and Stone.[15] 30

1.17 Properties of biological tissues and photodetectors in the broad optical window. (a) the attenuation of biological entities such as both oxygenated and deoxygenated blood, skin and fat. (b) the quantum efficiency of different photodetectors. Combining the information from these justifies the use of the first NIR window. From Duan.[16] 31

2.1 Sketch to visualise Buffon's needle problem. The black line is the nearest parallel line to the needle. The needle is represented by the blue line. The terms defined in order to solve the question of the probability of the needle crossing the line are shown: x, the distance from the needle centre to the closest line, and θ, the orientation of the needle. 36

2.2 An example of rejection method. A complicated function can be sampled by uniformly sampling x values, and for each generating a corresponding uniform random y. Where $y \leq f(x)$ is satisfied, the x value is kept, otherwise it is rejected. Here, we see that the point (x_1, y_1) is acceptable, but (x_2, y_2) is not and will be rejected. By repeating enough times, the original function can be reproduced. 38

2.3 Replica of the FERMIAC machine, built in Bologna in 2015. There are three main components, from left to right: the lucite platform to select the direction of the neutron, the rear drum measures the elapsed time from the neutron velocity and the front drum measures the distance the neutron travels based on the velocity of the neutron and the material it is travelling through.[17] 40

2.4 Diagram showing the terms involved with defining the energy flow dE through an area dA in direction $\hat{s}$. In equation 2.4, the term θ refers to the angle between the direction of interest $\hat{s}$ and the unit normal vector $\hat{n}$. Adapted from Chandrasekhar.[18] 44

2.5 Diagram illustrating the terms involved with defining the fluence dE through a cylindrical volume dA ds with solid angle dΩ in a direction $\hat{s}$ and a different solid angle dΩ' in another direction $\hat{s}'$, in order to derive the radiative transfer equation. Adapted from Wang and Wu.[19] 45

2.6 Schematic of the light attenuation in a highly absorbing medium. The transmitted light is dependent on both the absorption coefficient and thickness of the medium, and decays exponentially. Diagram adapted from Yoon *et al.*[20] 47

2.7 Flowchart describing the lifetime of a photon packet in a Monte Carlo simulation. 50

2.8 Diagram showing the impact of the anisotropy factor, g, on the Henyey-Greenstein phase function. The initial direction of travel is across the centre of the circle towards 0°. 53

2.9 Schematic diagram showing the geometry of light reflecting at a boundary between two media with different refractive indices. Here, the direction vectors are indicated by $\vec{I}$ for incident light, $\vec{R}$ for reflected and $\vec{N}$ for a normal to the surface. $\vec{A}$ is a coplanar vector with $\vec{I}$ and $\vec{R}$ parallel to the interface. $\vec{B}$ is another vector coplanar with $\vec{I}$ and $\vec{T}$ which is in the opposite direction to $\vec{N}$ at the point of the reflection on the interface. 55

2.10 Diagram showing the geometry of light refracting at a boundary between two media with different refractive indices. Here, the direction vectors are indicated by $\vec{I}$ for incident light and $\vec{T}$ for transmitted. The circle is a unit circle. Vector $\vec{M}$ is a unit vector coplanar with $\vec{I}$ and $\vec{T}$ along the interface; $\vec{N}$ is a unit vector also coplanar with $\vec{I}$ and $\vec{T}$ perpendicular to the interface. Vector $\vec{A}$ is along $\vec{M}$, vector $\vec{B}$ is perpendicular to $\vec{A}$, and the sum of $\vec{A} + \vec{B} = \vec{T}$. $\vec{C}$ is opposite in direction to, but runs along $\vec{N}$ to the point of refraction on the interface. 56

2.11 Graphic to depict the peel-off method. The yellow central circle is the light source and the green oval is the detector. The orange arrows show an example path for a photon packet out of the medium and the green arrows represent a small weight at each scattering event being peeled off towards the detector. 59

3.1 Graphic depicting the simulation setup with a source-detector separation s and a top layer with depth d. The colour change represents the laser photon being re-tagged as a Raman photon, and the location where the change occurred (the grey spiked shapes) is recorded. Figure adapted from the Matousek et al.[21] study used for benchmarking. 69

3.2 Definition of the collection annulus as used in this validation exercise with source detector separation s, and annulus width Δs. Figure adapted from the Matousek et al.[21] study used for benchmarking. 72

3.3 Digitised plot showing the number of Raman photons detected from the top and bottom layers, as a function of the radius of the annulus detector. Figure digitised from Matousek et al.[21] 72

3.4 Plot showing the Raman photon signal from top and bottom layers, as a function of spatial offset of the detector. The purple lines are the results from the original paper, and the blue lines are the results from our first pass attempt to benchmark with our code. 73

3.5 Plot showing the Raman photon signal from top and bottom layers, as a function of spatial offset of the detector. The purple lines are the results from the original paper, and the blue lines are the results from our second pass attempt to benchmark with our code, after using the correct definition of optical density. 75

3.6 Diagram illustrating how the differential surface area element of a sphere is derived. 77

3.7 Plot showing the Raman photon signal from top and bottom layers, as a function of spatial offset of the detector. The purple lines are the results from the original code, with isotropic scattering implemented, and the blue lines are the final results from our code. 79

4.1 Liquid tissue phantom schematic with inset photo of empty quartz tank with PTFE slab inside. 84

4.2 Diagram of the deep Raman setup used, in transmission mode. Based on a figure from Vardaki et al.[22] 85

4.3 Raman spectra of PTFE for all positions in the tank with 0.25% Intralipid as the scattering agent. The dominant Raman peak at 734 cm^{-1} is indicated by the dashed line. 86

4.4 Raman peak at 734 cm^{-1} for three of the positions of the PTFE in the tank with 0.25% IL concentration: PTFE closest to the laser, in the centre, and closest to the detector. The dashed lines show the best fit Python Gaussian for each of these, and the legend contains the R^2 value for each of these fits. 87

4.5 Intensity (measured as area under the peak) of dominant Raman peak in the PTFE spectrum for each of the tested concentrations of Intralipid. The laser is incident on the tank at 0 mm, and the collection is at 25 mm. The error bars represent the shot noise from the camera. 89

4.6 Upper: the fractional comparison of the experimental average difference (purple circles) and the Monte Carlo simulation average difference (blue crosses), both performed relative to 1% Intralipid values. Lower: plot is of our corresponding derived scattering coefficients against concentration. The variation in the scattering coefficient does not scale with the concentration directly as previously thought. The best fit line is the dashed line and has the formula $\mu_s = 5.16 \times \text{concentration}(\%) + 22.1$, and the R^2 value indicated in the legend. The black triangle indicates the estimated scattering coefficient for breast tissue based on our results. 95

4.7 A comparison of the experimental output at each concentration of Intralipid (purple) against the results of the Monte Carlo simulations (dashed teal). The scattering coefficients for the Monte Carlo simulations were calculated from the relative difference ratio and are detailed in Table 4.1. The laser is incident at 0 mm and the collection point is at 25 mm. Monte Carlo simulations were run with 10^9 photon packets and repeated 10 times for each PTFE slab position in order to calculate the variance, shown by the error bars. 98

5.1 Visualisation of the energy density in a gridded volume for a 1 cm diameter, 1 W power laser beam. This is an example output saved from the first step in a two-step Raman scattering simulation. 105

5.2 Visualisation of the energy density in a gridded volume for two different groups of 10 randomly generated microcalcifications. The power of each calcification Raman source is the energy density from the laser calculation at the location of the calcification. These are examples of light sources in the second stage of a two-step Raman simulation. 106

5.3 Graphic showing the post processing detector for TRS: a spectrometer in the simulation on the opposite face to the input laser collects the weights and locations of every shifted packet that exits that face. The image on the left is the output on the face opposite the laser at the end of a simulation; the image on the right shows a detector being added in post processing. All of the collected packet weights inside a circular area representing the detector are summed - thus one simulation can be used for a variety of detector radii. After the end of the simulation, the spectrometer data is processed in a Python script to sum the weights within an arbitrary circle. 108

5.4 Graphic showing the post processing detector for SORS: a CCD in the simulation on the same face as the input laser collects the weights and locations of every shifted packet that exits there. The image on the left is the output on the same face as the input laser at the end of a simulation; the image on the right shows a detector being added in post processing. All of the collected packet weights inside an annulus representing the detector are summed - thus one simulation can be used for a variety of source-detector separations. After the end of the simulation, the spectrometer data is processed in a Python script to sum the weights within an annulus. 108

5.5 Raman signal strength in pork tissue for 10 microcalcifications randomly distributed in a 1 mm sphere, for transmission Raman spectroscopy geometry. Each rectangle in the 2D plot represents the signal detected when the microcalcification distribution is centred at that location. The laser width is shown by the red lines on the left of each subplot and the detector by the black lines on the opposite side. 111

5.6 Raman signal strength in pork tissue for 10 microcalcifications randomly distributed in a 1 mm sphere, for spatially offset Raman spectroscopy geometry. Each rectangle in the 2D plot represents the signal detected when the microcalcification distribution is centred at that location. The laser width is shown by the red lines on the left of each subplot and the detector by the black lines on the same side. 112

5.7 Raman signal strength in pork and breast tissue for 10 microcalcifications randomly distributed in a 1 mm sphere, for transmission Raman spectroscopy geometry. Each rectangle in the 2D plot represents the signal detected when the microcalcification distribution is centred at that location. The laser width is shown by the red lines on the left of each subplot and the detector by the black lines on the opposite side. The laser diameter is 1 mm which is not typical for real world TRS but is done here for comparison purposes and curiosity. 114

5.8 Raman signal strength in breast tissue for 10 microcalcifications randomly distributed in different sphere volumes, for transmission Raman spectroscopy geometry. Each rectangle in the 2D plot represents the signal detected when the microcalcification distribution is centred at that location. The laser width is shown by the red lines on the left of each subplot and the detector by the black lines on the opposite side. 117

5.9 Raman signal strength in breast tissue for 10 microcalcifications randomly distributed in different sphere volumes, for spatially offset Raman spectroscopy geometry. Each rectangle in the 2D plot represents the signal detected when the microcalcification distribution is centred at that location. The laser width is shown by the red lines on the left of each subplot and the detector by the black lines on the sane side. 118

5.10 Graphic to demonstrate the simulation domain when skin layers were added to the top and bottom surface of the bulk average breast tissue volume. For transmission Raman geometry, the laser would be incident at the top, and the detector underneath. 121

5.11 Plot of the detected Raman signal for a bulk breast volume with skin at the top and bottom (in front of the laser and detector). The laser and detector both had a diameter of 1 cm, there were 10 microcalcifications inside a 1 mm spherical volume and these were moved 1 mm at a time along the optical axis. 122

5.12 Plot of the detected Raman signal for a bulk breast volume with skin at the top and bottom (purple), compared to the shape of the same simulation but without skin (pink). The model without skin has been scaled down by a factor of 10 to be plotted on the same y-axis. 123

5.13 Graphic to demonstrate the simulation domain when skin layers were added to the top and bottom surface of a fatty tissue volume with a sphere of glandular material containing the microcalcification point sources. For transmission Raman geometry, the laser would be incident at the top, and the detector underneath. 124

5.14 Plot showing the detected Raman signal for a fatty tissue volume with skin at the top and bottom (in front of the laser and detector), and the calcifications contained within a spherical gland. The scattering and absorption coefficients have been lowered by a factor of 10 to make them more similar to the bulk breast literature results from earlier. The laser and detector both had a diameter of 1 cm, there were 10 microcalcifications inside a 1 mm spherical volume and these were moved 1 mm at a time along the optical axis. 126

5.15 Plot comparing the results of a 3-layer skin model around bulk average breast tissue compared to the simulation with a 3-layer skin, fatty tissue and glandular tissue model. 127

5.16 The result of moving a 4 mm sphere of 50 mg of HAP powder through a volume of bacon on the optical axis. The laser and detector were in a transmission Raman spectroscopy set-up and the acquisition 10× 10 seconds for each point. 128

5.17 The result of simulating a 4 mm sphere of 500 microcalcifications through a volume of bacon on the optical axis. The laser and detector were in a transmission Raman spectroscopy set-up as in the experiment. 129

5.18 Plot made by Jennifer Haskell from her laboratory results of the Raman HAP counts against the signal to noise ratio, with a red line fitted through the data to allow extrapolation to lower SNR than the experiment afforded. 131

5.19 Map of Raman signal for a 1 mm sphere with 10 microcalcifications in it as it is moved through a volume of pork tissue. The input beam was a 1 W laser, 1 cm in diameter, with a circular detector opposite with a 4 cm diameter. The contour shows the region where the calcifications need to be in order to be detected. 132

5.20 Map of Raman signal for a 1 mm sphere with 10 microcalcifications in it as it is moved through a volume of pork tissue. The input beam was a 1 W laser, 1 cm in diameter, with a circular detector opposite with a 4 cm diameter. The smaller contour shows the region where the group of 10 calcifications need to be in order to be detected; the larger contour demonstrates the region where a group of 500 calcifications would need to be in order to be detected. 133

List of Tables

4.1 Absorption (μ_a), scattering (μ_s), anisotropy (g), and reduced scattering coefficient (μ_s') used in the simulations at 830 nm, the laser wavelength. The values for breast tissue[23,24] are included in order to compare with the tissue phantoms. The absorption coefficient values for the diluted IL are that of water, as this is the dominant absorbing material in the tissue phantoms. . 96

5.1 Absorption (μ_a) and reduced scattering coefficient (μ_s') used in the simulations at 808 nm (the laser wavelength) and 876 nm (the Raman wavelength). Values from Mosca et al.[25] for pork muscle. 110

5.2 Absorption (μ_a) and reduced scattering coefficient (μ_s') used in the simulations at 808 nm (the laser wavelength) and 876 nm (the Raman shifted wavelength). Values from Spinelli et al.[26] 113

5.3 Absorption (μ_a), anisotropy (g) and scattering coefficient (μ_s) used in the simulations at 808 nm (the laser wavelength) and 876 nm (the Raman wavelength). Values from Bhandari et al.[27] 121

5.4 Absorption (μ_a), anisotropy (g) and scattering coefficient (μ_s) used as the basis for the simulations at 808 nm (the laser wavelength) and 876 nm (the Raman shifted wavelength). Values from Peters et al.[28] 125

Chapter 1

Introduction

1.1 Breast cancer

Cancer is one of the most prevalent and devastating diseases in modern society; globally in 2020 there were an estimated 19.3 million new cases and 10.0 million deaths.[1] There are more than 200 types of cancer, but four of them account for over 4 in every 10 diagnosed: lung, breast, bowel and prostate. A breakdown of the incidences of different cancer types can be seen in figure 1.1.

Breast cancer is the most commonly diagnosed cancer type in the UK. In this section, a statistical background on the prevalence of breast cancer is presented, as well as the physiology of breast tissue and the cancer pathology.

1.1.1 Statistics

Breast cancer mainly affects women, with one in seven women born after 1960 and living in the UK having a lifetime risk of developing it. Men can also develop breast cancer, but this occurs at a much lower rate of approximately one in 870.[29]

Mortality from cancers, especially in developed countries, has decreased in recent years. This is testament to scientific developments in treatments and in diagnosis. The earlier cancer is detected, the better the survival rates and the less severe the treatment can potentially be. There are increasing numbers of cases, however, with the UK female

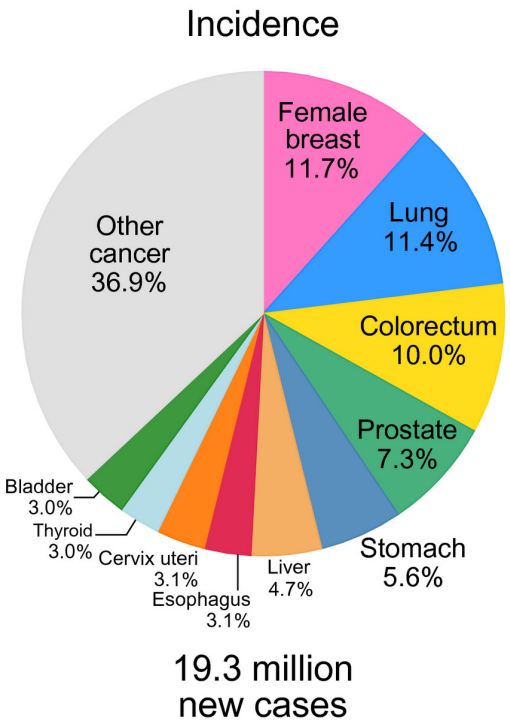


Figure 1.1: Pie chart showing incidence of different cancer types in 2020 in both sexes. Four (lung, breast, bowel and prostate) make up more than 4 in 10 new diagnoses. Taken from Global Cancer Statistics.[1]

incidence increasing by 23% between 1993–1995 and 2015–2017.[30] Some of this increase can be explained by improved detection techniques and data recording. Risk factors and how they change over time also have an impact on the incidence rates.

Developed nations, including Western Europe and North America, see four-fold incidence of breast cancer when compared to developing countries. This is believed to be linked to old age, and the Western lifestyle increasing risk factors - smoking, lack of physical activity, unhealthy diet, obesity and alcohol abuse.[31] Up to 23% of breast cancers are estimated to be avoidable through different lifestyle choices.[32]

Familial history is an unavoidable risk factor, and it is known that the number of relatives with breast cancer increases an individual's incidence likelihood. Hereditary factors are responsible for around one-quarter of breast cancer risk.[33] This risk is about twice as high in women with one first-degree family member with breast cancer, and this risk increases with a larger number of first-degree relatives diagnosed. However, over 85% of women with a first-degree relative with breast cancer will not develop breast cancer themselves.[34]

Two specific genetic mutations responsible for an increased risk in breast cancer have been identified: BRCA1 and BRCA2 mutations. Women with either mutation have a 45-65% chance of developing breast cancer by the age of 70.[35] Early diagnosis or identification of the faulty gene is crucial for these women to maximise their survival rates. Patients at sufficiently high risk frequently opt for a bilateral mastectomy to remove the chance of incidence.

Breast cancer accounted for 7% of all cancer deaths in the UK in 2018, and in women was the second most common cause of cancer death (after lung cancer). Breast cancer European age-standardised (AS) mortality rates in the UK for females and males combined decreased by 39% between 1971–1973 and 2016–2018, as shown in figure 1.2.

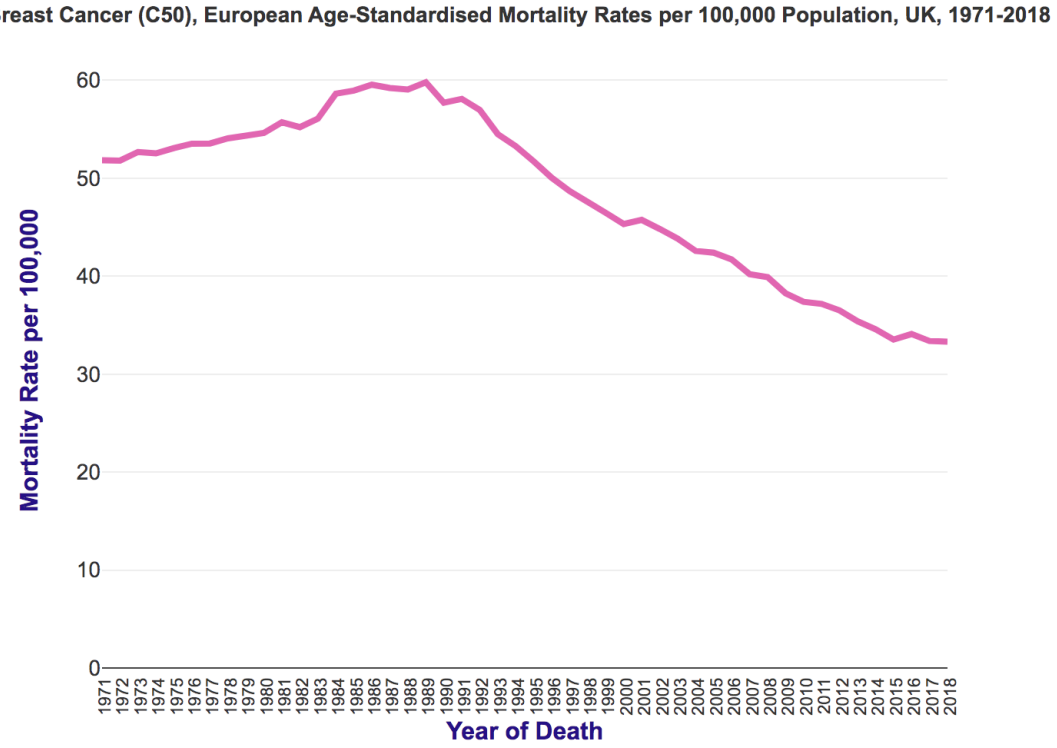


Figure 1.2: Mortality rate per 100,000 people in the UK from 1971-2018. Taken from Cancer Research UK.[2]

Although survival has almost doubled in the last 40 years in the UK, the mortality rates are still high. Around 11,500 women died of breast cancer in 2018 - this is approximately 31 women every day.

In order to improve the survival rates of breast cancer, an understanding of the physiology of the human female breast is important. This knowledge gives greater un-

derstanding into how mammograms and other methods of detection could be utilised to aid early diagnosis.

1.1.2 Breast anatomy

Breasts are present on the upper chest wall in both males and females; the breast bud begins developing in utero. Before puberty, both sexes have breasts made up of a basic ductal tree leading to the nipple in a stromal fat pad.[36] The female breast develops into the mammary gland (as seen in figure 1.3) during puberty in response to hormonal stimulation. The developed breast is made up of mammary glands surrounded by a connective tissue stroma and its primary purpose is to produce milk.

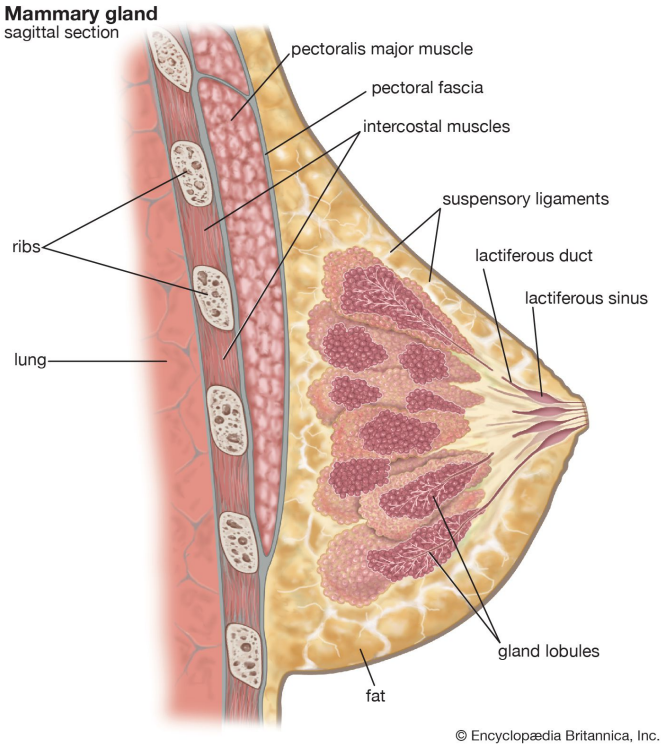


Figure 1.3: Diagram of female breast anatomy after puberty. Image and labels taken from Encyclopedia Britannica.[3]

The gross descriptions of the human breast are mostly unchanged since Astley Paston Cooper dissected cadavers of women who were lactating at time of death.[37] The adult female breast is composed of glandular (secretory) and adipose (fatty) tissue. These are supported by a loose framework of connective tissue called Cooper's ligaments. The glandular anatomy is composed of lobes, which are divided into multiple lobules. Each lobule contains clusters of alveoli which are made of mammary secretory epithelial cells.

Each lobule contains anywhere from 10–100 alveoli, and traditionally it was thought that a breast contained 15–20 lobes. Recently, ultrasound work has suggested the number of lobes may actually be 5–9.[38]

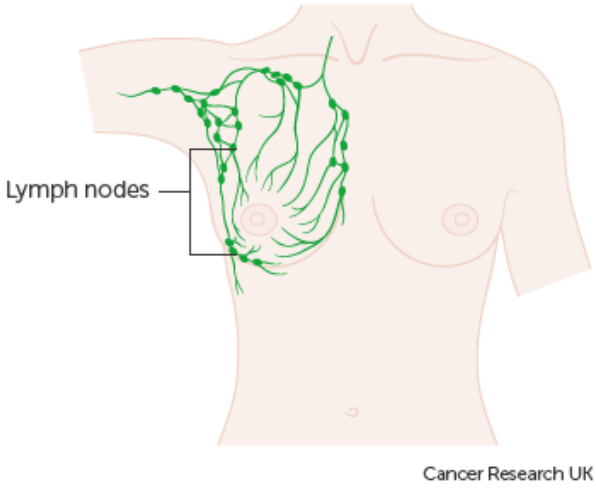


Figure 1.4: Lymphatic drainage system and lymph nodes found in and around the breast. Image taken from Cancer Research UK[4]

The breast also contains a lymphatic network to handle immune responses to pathogens, shown in figure 1.4. The lymph vessels circulate lymphatic fluid and drain into the lymph nodes, small masses of lymphatic tissue in the area around the breast. Due to the proximity of this network, breast cancer cells (once they have invaded local tissues) are able to spread through the body forming metastases, most commonly in bone, lung and liver.[39]

1.1.3 Cancer pathology

Most breast cancers begin in the cells that line the ducts, while others begin in the lobules. A small number begin in the other tissues of the breast. Non-invasive breast cancer cells are confined to the ducts or lobes and have not spread into the surrounding fatty or connective tissues. Ductal carcinoma in situ (DCIS) is the most common form of non-invasive breast cancer: 1 in 5 cancers detected at screenings in the UK are DCIS. Lobular carcinoma in situ (LCIS) is rarer. Invasive breast cancer is the most common type of breast cancer. These cancer cells have invaded through the duct or lobular wall and are in the surrounding tissues. These invasive cancers are not necessarily metastatic. Ductal invasive breast cancer is the most common type of breast cancer, with 70–80% of diagnoses being this type.

Breast cancers are classified by staging and grades. The staging describes how large the cancer is and whether it has spread; the grading describes how abnormal the cancer cells appear. These two metrics are used to determine which treatments are appropriate.

There are two main staging metrics: TNM and number stages. The TNM staging system stands for Tumour, Node, Metastasis, and are:

- T describes the size of the tumour.
- N describes whether or not the tumour has spread to the lymph nodes.
- M describes whether the cancer has spread beyond the lymph nodes to another part of the body.

The number staging system goes from 0 to 4:

- Stage 0 is ductal carcinoma in situ (DCIS). This is a pre-invasive breast cancer that has not begun to spread into surrounding breast tissue.
- Stage 1 means the cancer is small and contained in the breast tissue or in lymph nodes close to the breast. This is an early stage breast cancer.
- Stage 2 breast cancer means the cancer cells are in the breast, nearby lymph nodes, or both. This is also an early stage breast cancer.
- Stage 3 means the cancer has spread from the breast to lymph nodes, the skin of the breast, or the chest wall. This is called locally advanced breast cancer.
- Stage 4 breast cancer means the cancer has spread to other parts of the body.

The grade describes how cancer cells look under a microscope and how similar they are to normal breast tissue cells. The features that can help to predict the behaviour of the cancer include the arrangement of the cells, whether they form tubules, how similar they look to normal breast cells and how many of the cells are dividing. The more different the cells look and the faster they are growing, the poorer the prognosis.

In addition to the cell grading system, there are other biomarkers present in breast

tissue that can be used to identify or rule out malignancies. The primary biomarker used in this thesis is microcalcifications, which are explained in more detail in the next section.

1.1.4 Microcalcifications

Microcalcifications are small calcium deposits that can be identified on a mammogram as tiny white specks. They are a common occurrence, especially with age and post-menopause. They can appear for a variety of reasons, both malignant and benign. They can be an important disease identifier in the case of breast cancer with up to 50% of non-palpable breast cancers detected through microcalcifications present in a mammogram.[40] Moreover, up to 93% of DCIS cases have associated microcalcifications.[41]

These microcalcifications have been characterised on a molecular level and can be identified by their physical and chemical properties. They fall into two categories: Type I and Type II. Type I calcifications are composed of calcium oxalate (CO): amber in colour, partially transparent, and form in pyramid-type structures. Type II calcifications are made of calcium phosphate, primarily hydroxyapatite (HAP): grey-white in colour, opaque, and formed in ovoid shapes with irregular surfaces.[5]

The type of calcification present appears to be related to the pathology. The study by Frappart et al.[5] looked at both types of microcalcification using light microscopy and X-ray diffraction. Their results indicate that type I calcifications are present in benign lesions (or in situ lobular carcinoma at most[42]), whereas type II calcifications are present in both malignant and benign lesions. The shapes can be seen in figures 1.5 and 1.6.

Additionally, it has been shown that the carbonate substitution in type II calcifications can give an indication of disease progression.[43,44] Benign proliferative lesions have higher carbonate levels in the type II calcifications present than those of malignant lesions. There are two main different methods of substitution available for increasing carbonate in the calcification. A-type substitutions have the carbonate group (CO_3^{2-}) replace the hydroxyl groups (OH^-). The charges are balanced by one carbonate group replacing two hydroxyl groups. In B-type substitutions the CO_3^{2-} replaces the phosphate groups (PO_4^{3-}). The charge balance here is more complicated and requires co-substitution of sodium ions.[45] In breast malignancies, HAP undergoes B-type substitutions. Another

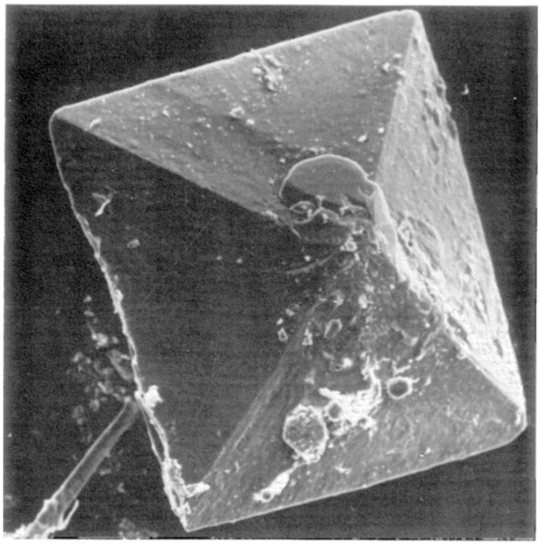


Figure 1.5: Image of type I microcalcification from a sclerocystic mastopathy (benign). Taken from Frappart et al.[5]

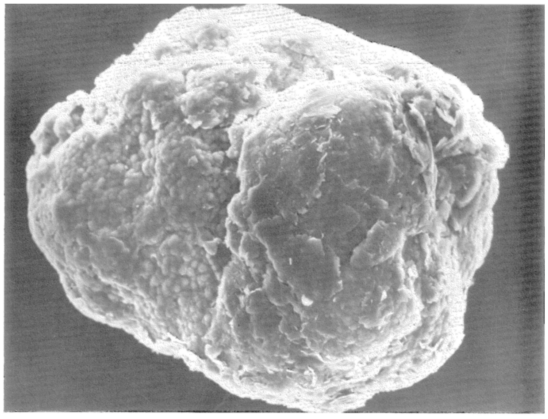


Figure 1.6: Image of type II microcalcification from an infiltrating carcinoma. Taken from Frappart et al.[5]

study by Haka et al.[46] shows that type II calcifications with a larger calcium carbonate level are found in benign ducts, and those in malignant ducts have lower carbonate levels.

Understanding the anatomy of the breast and the structures therein give us the context required to look at how diagnosis of breast cancer is currently performed, and how it might be improved in future. The next section looks at the current diagnostic and treatment pathways, in order to highlight later where spectroscopy could fit into and improve the current methods.

1.1.5 Diagnosis

A number of techniques are currently used in diagnosing breast cancer. The aim is to diagnose patients as early as possible in order to treat them more easily and maximise the chances of a full recovery. The common methods of diagnosis are discussed in this section, followed by a brief overview of the current treatment methods. These discussions will help to frame where the work later in the thesis can fit into the current diagnostic pathway.

1.1.5.1 Early diagnosis

An early diagnosis of cancer can be vital in improving the outcome for the patient, especially when asymptomatic.[47] The benefits are in reducing mortality and morbidity, relieving symptoms and possible well-being improvements for the patient. The World Health Organisation outlines three steps to early diagnosis of cancer:[48]

- Awareness and accessing care - awareness of symptoms, seeking and accessing care.
- Clinical evaluation, diagnosis and staging - accurate clinical diagnosis, diagnostic testing and staging, referral for treatment.
- Access to treatment - accessible, high-quality treatment.

Cancer survival rates are generally poorer in the UK and Ireland than in Europe as a whole.[49–51] The European mean for 5-year breast cancer survival for adults diagnosed in 2000–2007 was 81.8%, whereas the UK and Ireland mean was 79.2%.[6] As can been seen in figure 1.7, the survival of women aged 75 years and over was especially low, and contributed most of the difference in survival rates.

The stage of the cancer at diagnosis is highly predictive of mortality. A higher proportion of cancer patients are diagnosed at a later stage in England than in Europe.[52,53] The records of staging have improved since those studies, exceeding 80% for breast cancer (amongst other cancers). The remaining staging data can be missing for a variety of reasons: certain tumour types have no formal system, it was deemed clinically inappropriate, or there was patient death before staging. A National Awareness and Early Diagnosis Initiative was launched in 2008[54] for England, between Cancer Research UK and the UK

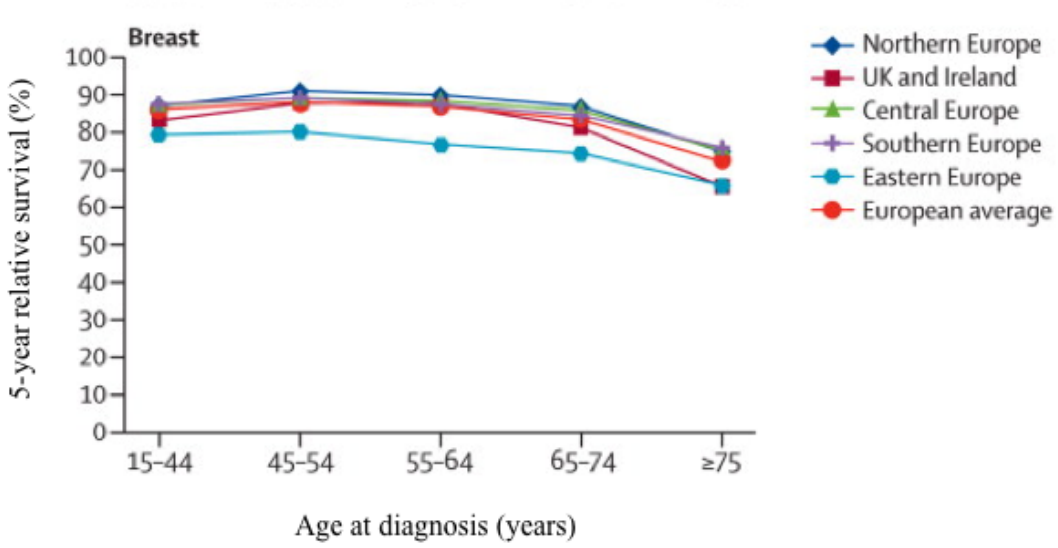


Figure 1.7: Plot showing the age specific five year relative survival rate of breast cancer in adults diagnosed in 2000–2007. Taken from De Angelis et al.[6]

Government to promote awareness of early symptoms of cancer. Richards et al.[55] performed a systematic review of worldwide observational studies looking at symptoms and survival rates. This review showed strong evidence that longer delays are associated with poorer survival and more advanced disease. Specifically, women who have a symptom duration of 3–6 months have an estimated 7% lower 5 year survival rate than those with a shorter duration of symptoms. It concludes that between 5000 and 10,000 cancer deaths within 5 years of diagnosis could be avoided per year in England with successful earlier diagnosis and if appropriate primary surgical treatment was implemented.

A study from 2015[47] looked at which cancers could have the most improved mortality and/or morbidity rates from expedited diagnosis by use of a modified Delphi exercise. The conclusion shows that the greatest mortality benefits could be found in breast cancer, followed by uterine cancer and melanoma. The least benefit was in brain and pancreatic cancers, mostly due to screening being unviable and having limited treatment options. This justifies the work to improve breast cancer detection, as patients will benefit most from earlier diagnosis. In contrast, research efforts for cancers with little to gain could be focused more on prevention, treatment or symptom relief.

Work by Laudicella et al.[56] investigated how the cost of treating breast cancer varies

depending on the age of the patient and the stage at diagnosis. An earlier stage diagnosis results in lower treatment costs, however costs in the first year of diagnosis are higher in patients under 65 compared to those over. Costs in the year of diagnosis of breast cancer are around £11,000 for patients aged 18-64 and £7700 for those 65 and over. In breast cancer, lower stage breast cancer (stages 1–2) saves 13–19% in treatment costs compared to later stage (stages 3–4).

The works highlighted here show that it is very important to diagnose and treat breast cancer as early as possible in patients. In the following section the current methods used for diagnosing breast cancer will be described.

1.1.5.2 Current diagnostic methods

There are several methods for diagnosing breast cancer: the current "gold standard" is biopsy followed by histopathological analysis. The National Institute for Health and Care Excellence (NICE) quality standards require that people with suspected breast cancer who are referred to specialist services are offered a triple diagnostic assessment in a single hospital visit: clinical assessment, mammography/ultrasound, and fine needle aspiration or core biopsy.[57]

1.1.5.2.1 Clinical examination

Regular breast self examination (BSE) is touted as a low-cost and low-risk method to detect early signs of breast cancer and can be performed without technical equipment by the women themselves.[58] Symptoms women are advised to look out for include skin dimpling, nipple inversion (if not already inverted) and a new lump or thickening. There is little evidence to confirm that performing BSE provides tangible benefit, and it has been suggested that these actually increase the number of further tests required by two-fold.[59] Advantages are that it is cheap and can make a difference in low and middle income countries where widespread screening is not commonplace.[60]

Clinical breast examinations are performed by a trained healthcare professional. In these examinations, both breasts and the lymph nodes from armpit to clavicle are investigated. The signs to look for are similar to those in the BSE, and includes checking the

lymph nodes for inflammation or pain. These clinical examinations have low sensitivity rates[61] and so are always followed by imaging diagnostics.

1.1.5.2.2 Mammography

Mammography is used for both screening and diagnostics. It uses low energy X-rays to obtain an image of the inside of the breast, showing early signs of breast cancer such as masses or microcalcifications if present. X-rays are a form of ionising radiation and thus come with an inherent risk: assuming a woman attends 7 mammograms between the ages of 50 to 71, the risk of radiation induced breast cancer is between 1 in 7000 and 1 in 14,000.[62]

In the UK, a mass breast screening programme was launched in 1988 following recommendations made by Professor Sir Patrick Forrest.[63] The programme invites women aged 50–70 years old every three years for a mammogram, however there is an age extension trial currently running that means the most up to date data is for the age range 45–74. Between 2019–2020, 2.12 million women were screened, and 17,771 women had cancers detected: 8.4 cases per 1000 women screened. 78.9% of these were found to be invasive.[64]

During the procedure, the breast tissue is compressed to improve the quality of the image and reduce the dose of X–rays required. This compression is achieved by the mammogram unit, as seen in figure 1.8. Mammograms used for diagnostic purposes have additional views taken. High density breast tissue (usually seen in younger women) is less easily imaged by mammogram and ultrasounds are generally used for these.

Despite mammography being standard practice in many countries, there are several questions around its suitability due to the low specificity and the high rate of false positives diagnosed.[65] Mammograms are also not chemically specific so cannot inform clinicians on the type of microcalcifications present. Therefore, further tests are always required if something suspicious is seen in the results. A false positive occurs when a patient is told they have the disease when they do not: an estimated cumulative risk of 20% for women aged 50–51 who participate in biennial screenings. The risk of having a further invasive procedure is 3%.[66] This is both upsetting for patients and expensive for the NHS.

Moreover, younger breasts tend to return higher rates of false negatives: the disease

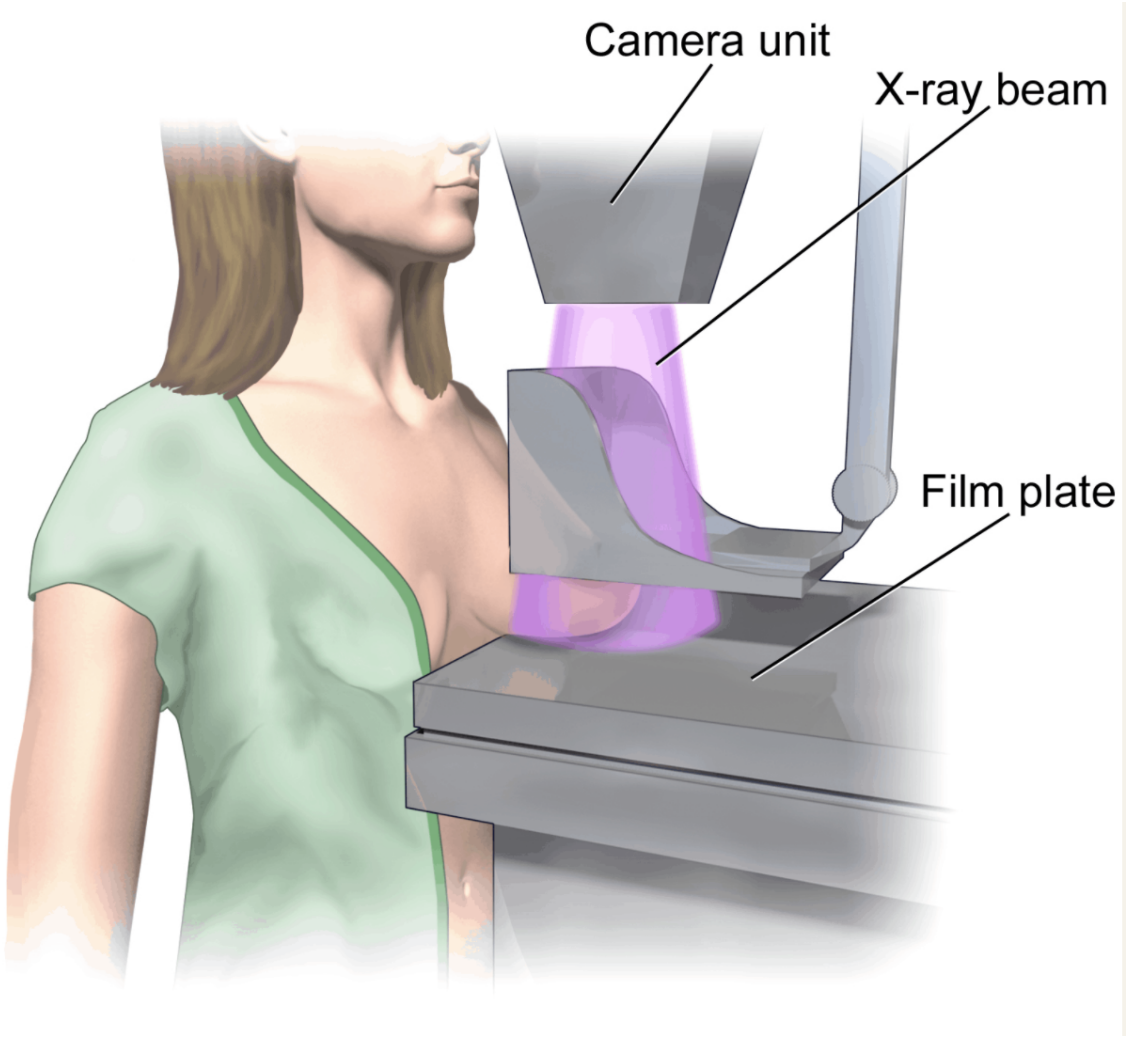


Figure 1.8: Mammogram illustration from Blausen Medical.[7]

is present but the mammogram does not allow for it to be identified. These can include cancers that are not visible, cancers that are only visible in retrospect and cancers that are interpreted as a benign result. Younger women or people with breast implants are considered unsuitable for mammograms due to the density of the breast tissue appearing bright on the mammogram and thus obscuring any microcalcifications or tumours that might be present.[67] These cases, and any patients with significant fibroglandular tissue in the breast, would be better suited to an ultrasound.

1.1.5.2.3 Ultrasound imaging

An ultrasound scan uses high-frequency sound waves to generate an image of the inside of the breast. It is non-ionising, inexpensive and effective in women where mammography

is unsuitable such as those with denser breasts. Ultrasound imaging also has the potential to detect node negative cancers (where there has not been spread to the lymph nodes)[68] and differentiate between solid and liquid filled masses in the breast.[69,70]

There are several disadvantages to ultrasound imaging, as a stand-alone diagnostic tool or for screening. Despite improvements in equipment, the sensitivity of the technique is very dependant on the operator and examining the entirety of both breasts is time consuming. The low sensitivity and specificity of ultrasound imaging, in comparison to mammograms, means it is not suitable as a mammogram replacement.[71] It does remain needed as a complementary tool, or for imaging patients with dense breast tissue.

1.1.5.2.4 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is another non-ionising imaging technique that utilises strong magnetic fields and radio waves. MRI has become a means of screening women with high lifetime risk of breast cancer, such as those carrying the BRCA gene mutation. Using MRI for screening a younger cohort is useful since it is not affected by the denser breast parenchyma.[72] MRI has a much greater sensitivity than mammograms but generates a lot of false positive results.[73] MRI can be expensive, but it is used annually for surveillance of young and high-risk women from the age of 30 onward.[74]

1.1.5.2.5 Biopsy and histopathology

There are two main types of needle biopsy (as seen in figure 1.9) that are used in breast cancer diagnosis: fine needle aspiration (FNA) and core biopsy. Both of these are used to remove tissue from the area of concern in order for it to be examined ex-vivo by a pathologist. These are used in conjunction with mammograms to make up the current gold-standard diagnostic procedure.

FNA uses a very thin needle to remove a small number of cells from the suspicious area, whereas core needle biopsy uses a much larger needle to remove a column of tissue. Fine needle aspiration is quicker than core needle biopsy, as well as not usually requiring a local anaesthetic and having a lower risk of complications. On the other hand, core needle biopsy samples a far greater volume of the tissue and is less likely to lead to false negative

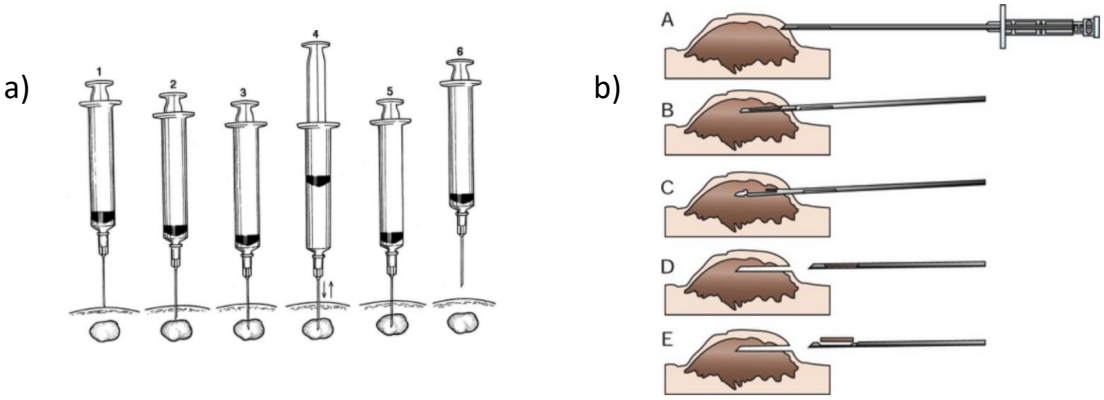


Figure 1.9: Graphics displaying the procedure for (a) fine needle aspiration biopsy[8] and (b) core needle biopsy.[9]

results.[75] Core needle biopsies are most commonly used in the UK currently.

Samples from the biopsy are placed in fixative solution immediately and sent to the laboratory as soon as possible.[76] The biopsy samples are sectioned and stained before being examined under a microscope to be classified by a specialist pathologist. The NICE guidelines specifically state that if microcalcifications are seen in the initial images (mammogram), then the biopsy needs to contain these in order for them to be classified.[77] This sets a priority in identifying microcalcifications and including them in the diagnostic process. The main source of issue in using biopsy samples is the lack of consensus between histopathologists. A study from 2015 shows that, amongst 115 pathologists interpreting 60 cases, there was 75.3% concordance with the consensus-derived reference diagnosis, and 50.5% of them reported that breast pathology is challenging.[78] In order for biopsy to be as precise as possible, analysis by several histopathologists and the use of complementary imaging techniques are key.

1.1.5.3 Treatment methods

Breast cancer patients usually begin their treatment with surgery, followed by adjuvant therapy to maximise the effectiveness. Occasionally, these second therapies will be done before surgery to shrink a tumour and improve outcomes. The specific combination of treatments depends on the stage and grade of the cancer in each individual case. The different therapies are detailed below, and the information is from the NHS website[79] , Cancer Research UK website[80] , and NICE guidelines.[81]

1.1.5.3.1 Surgery

A lumpectomy is a breast conserving surgery that removes the cancer and a small surrounding area, leaving as much normal breast tissue as possible. The margin is important for ensuring that all of the cancer has been removed, so this area is checked for cancer cells. Lymph nodes are occasionally also removed in order to be checked for metastasis. If the tumour is too large, or the cancer is in a higher-risk patient, then a mastectomy may be performed which is where the entire breast is removed. This can also include the surrounding lymph nodes.

1.1.5.3.2 Adjuvant treatments

Radiotherapy is commonly used after surgery to ensure that all of the cancer cells have been killed and to lower the risk of the cancer coming back. This treatment uses high-energy X-rays to disrupt the DNA of the cells and prevent proliferation. The biggest side effects come from the damage done to healthy cells, but these do not usually last long as the healthy cells can repair themselves more easily.

Chemotherapy is a combination of anti-cancer drugs that target cells which multiply quickly. This is good for targeting cancer cells, but unfortunately can lead to unpleasant side-effects such as hair and nail loss, nausea and lack of appetite.

Hormone therapy relies on the identification of hormone receptors on the tumour. Ovaries produce oestrogen and progesterone in women before the menopause, and oestrogen is made in body fat and muscle after the menopause. In some cancers, these hormones can stimulate the growth of breast cancer cells. Hormone therapy works by lowering the levels of these hormones or blocking this proliferative effect. This treatment will only work on cancers that have oestrogen receptors (ER). Around 70% of breast cancers have them, and this is checked when the breast cancer is diagnosed.

Targeted therapy is a variety of drugs that target the differences in cancer cells in order to disrupt them. The drugs work on different proteins, for example some breast cancers have a large amount of HER2 receptor which can be treated with a drug that attaches to this protein and then stops the cancer cells from growing and dividing. This

approach is less destructive than chemotherapy, but is only appropriate for cancers with specific receptors. Around 15 out of 100 women will have triple negative breast cancer: no oestrogen receptors, progesterone receptors, or HER2 receptors. This is more common in younger women, and means that the options are limited to radiotherapy, chemotherapy, and surgery.

1.1.6 Discussion

The number of diagnosed breast cancer cases are increasing globally. There have been many campaigns of awareness and population screening, the results of which have improved survival rates in the UK and other developed countries. Earlier diagnosis is known to be linked to less drastic and invasive treatments and longer life expectancy.

The current imaging techniques discussed in this section have their benefits and have undoubtedly helped to save lives. Unfortunately, they are not optimal for detecting cancer in the earliest stages, when the treatment required could be minimally destructive. Mammograms of diagnostic quality require compression and therefore pain in some patients which can reduce rates of return for asymptomatic screening. Ultrasound imaging can be used for dense breasts which are not suitable for mammography but have reduced sensitivity and specificity. MRI is far more sensitive, however has high associated costs which limits its use on the screening population as a whole.

Needle biopsy followed by histopathology is the current "gold standard" for breast cancer diagnosis, and it is capable of detecting molecular structure changes in disease progression. Biopsies are an invasive procedure, and are linked to high false negative rates. This high rate is caused by the small tissue volume removed in the biopsy compared to the whole tissue volume. If the tumour is small, it could be missed entirely. Moreover, the assessment of a biopsy by a histopathologist is prone to a lack of reproducibility, particularly in intermediate or unclear disease stages. On the other hand, false positive diagnoses lead to over-treatment and patient distress.

Increasing the suite of diagnostic techniques, to include further and unique information would be beneficial to earlier identification. The techniques discussed previously are broadly image-based, and the classification is morphological. Adding a technique

which gives biochemical information could improve the diagnostic pathway. Raman spectroscopy is a promising novel optical approach which uses light to identify molecules. This technique uses inelastic scattering to analyse the vibrational modes of molecules and could be used to differentiate malignant and benign molecular biomarkers. The theoretical background of Raman spectroscopy, instrumentation and current work in applying this to cancer diagnoses are discussed in the next section.

1.2 Raman spectroscopy

Vibrational spectroscopy is a novel diagnostic technique, with two main branches: Raman and infrared (IR). These techniques are chemically specific and label-free, which makes them ideal for identifying the chemical makeup of the microcalcifications described in the previous section.

IR spectroscopy is used in a variety of situations: quality control,[82,83] forensic analysis,[84] and the food industry[85] to name a few. It is extremely sensitive to water, however, and this deems it entirely unsuitable for *in vivo* measurements of the human body. The large water signal overwhelms any small signal from other biomolecules and microcalcifications, for instance.

This leaves us with Raman spectroscopy, which is largely unaffected by water content when used at near infrared wavelengths. It is currently implemented in a number of applications due to its high sensitivity and specificity: characterising materials,[86] geology,[87] art history,[88] security[89] and pharmaceuticals.[90]

Raman spectroscopy can be used in the breast cancer diagnostic pathway by offering a chemically specific method of differentiating between microcalcification types. Determining the type of microcalcification present could reduce the number of biopsies that need to be performed by ruling out malignant lesions in the presence of type I microcalcifications. In this section, we will focus on the theoretical background behind this technique.

1.2.1 Theoretical background

1.2.1.1 Light-matter interactions

When light propagates through a material, such as a biological tissue, there are two primary interactions: scattering and absorption. This results in a "random walk" effect, which varies with the optical properties of the material. In scattering, the direction a photon is travelling in is altered (either with or without losing energy). A highly scattering medium can result in a photon becoming very deviated from its original trajectory. In absorption, the incident photon has all of its energy absorbed due to a match between energy levels (vibrational, rotational and/or electronic) and the energy of the photon. Often this energy is then radiated away thermally.

These interactions can be characterised for different materials by a scattering coefficient (μ_s) and an absorption coefficient (μ_a). The coefficients describe the chance per unit length of a photon being scattered or absorbed, respectively. It is also important to consider that not every material will scatter isotropically; biological tissue is very forward scattering.[91] This effect can be described by the anisotropy factor (g) which is a measure of the cosine of the mean scattering angle. These terms are discussed more mathematically in the context of Monte Carlo simulations in chapter 2. The main physical factors shaping the optical properties of different tissues are the presence of chromophores (such as melanin and haemoglobin) and cell presence and morphology on the same scale as the wavelength of the light. These features increase the absorption coefficient and alter the scattering coefficient,[92] respectively.

1.2.1.2 Raman scattering

In most instances of a scattering event after a photon interacts with a molecule, the photon leaves with the same amount of energy as it entered. This is elastic scattering. In very few scattering interactions (approximately 1 in 10^6),[93] a Raman scattering event occurs: the photon leaves with a different amount of energy, thus changing its wavelength. The difference in the wavelength is characteristic of the scattering molecule. Heavier atoms and stronger bonds create a larger energy shift than smaller atoms and weaker bonds.

The inelastic scattering of light was first predicted by A. Smekal[94] in 1923, however was not observed until 1928. The Raman effect, specifically, is named after one of its experimental discoverers, C.V. Raman.[95] C.V. Raman won the Nobel Prize in Physics in 1930 for this work and became the first Asian to win a Nobel Prize in any science. He was so confident that he would win, that he had booked his travel to the ceremony in advance of being announced.[96] There have been controversies around his discovery though: K. S. Krishnan (his research associate) performed the experiments and co-authored the papers but was not nominated alongside Raman for the Nobel prize. Moreover, Russian scientists Mandelstam and Landsberg also independently observed the Raman effect[97] just a week before Raman and Krishnan[98] ; they were nominated for the Nobel prize alongside Raman but as they cited Raman's paper in their later article they were not awarded it. The discovery, nonetheless, remains important to this day and in the first seven years after its discovery, more than 700 papers in the scientific literature on the topic were published.

Due to the wave-particle duality of light, there are two ways of considering the Raman effect: wave interpretation or quantum mechanically. In the wave description, light has an oscillating electric field which interacts with the molecule via the negatively charged electron cloud. This distortion induces a dipole moment which is related to the strength of the electric field of the radiation and the polarisability (how easily the electron cloud can be distorted) of the molecule. The polarisability of a molecule depends on the number of electrons, their mean distance from the nucleus and the strength of the electron-nucleus attraction.

The quantum mechanical interpretation frames light as a photon interacting with a molecule and can be described in terms of energy levels, as seen in figure 1.10. The molecule gets excited by the incident photon to a virtual state, and then de-excites by three different routes. In most circumstances, the molecule relaxes back to the ground state and this is Rayleigh scattering. Occasionally, the molecule will relax to an excited vibrational state resulting in Raman Stokes scattering, where the emitted photon has a frequency shift ($\Delta\nu$ in figure 1.10) according to the difference between the ground and the vibrational state. Even more rarely, anti-Stokes Raman scattering happens, where the molecule was originally in an excited vibrational state, then after the photon interaction it relaxes to the

ground state. This gives the same wavelength shift as Stokes, but in the opposite direction. Anti-Stokes scattering is rarer than Stokes because at room temperature most molecules are in the ground state, according to Boltzmann's distribution.

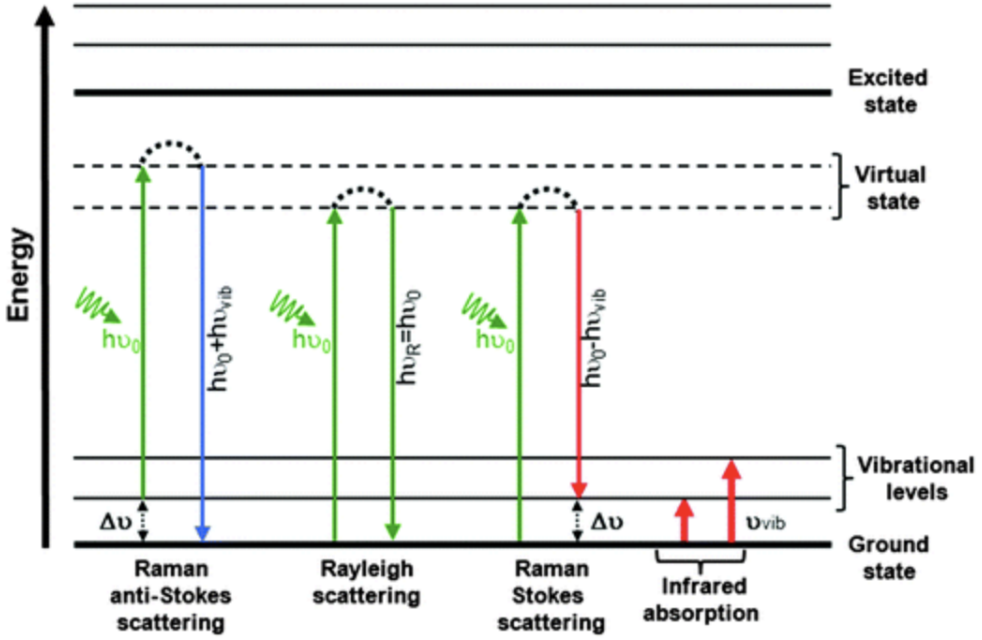


Figure 1.10: Jablonski diagram of vibrational energy states of a molecule. The transitions shown are infrared absorption, Rayleigh, Raman Stokes, and Raman anti-Stokes scattering. $h\nu_0$ is incident laser energy, $h\nu_{\text{vib}}$ is vibrational energy, $\Delta\nu$ is frequency shift and ν_{vib} is vibrational frequency. Taken from Baker et al.[10]

Infrared absorption is also highlighted in figure 1.10. This occurs when an infrared-active molecule absorbs a photon with a frequency that matches a vibrational energy level gap. By analysing a spectrum, a molecular fingerprint can be obtained, similarly to Raman spectroscopy. These techniques can be complementary: different bonds have different vibrational activity. For a molecule to be IR-active, there needs to be a change in the dipole moment, however in Raman there needs to be a change in the polarisability as discussed earlier. The symmetry also plays a role: IR is usually applicable in polar bonds and asymmetric vibrations, whereas Raman active bonds tend to be symmetric vibrations. In centro-symmetric molecules, IR active and Raman active vibrations are mutually exclusive and the two techniques are truly complementary.

The bond vibrations fall into two categories: stretching and bending, which can be seen in figure 1.11. Stretching vibrations involve the atoms moving in and out along the axis of the bond. Bending vibrations are motions which change the angles between the

bonds.

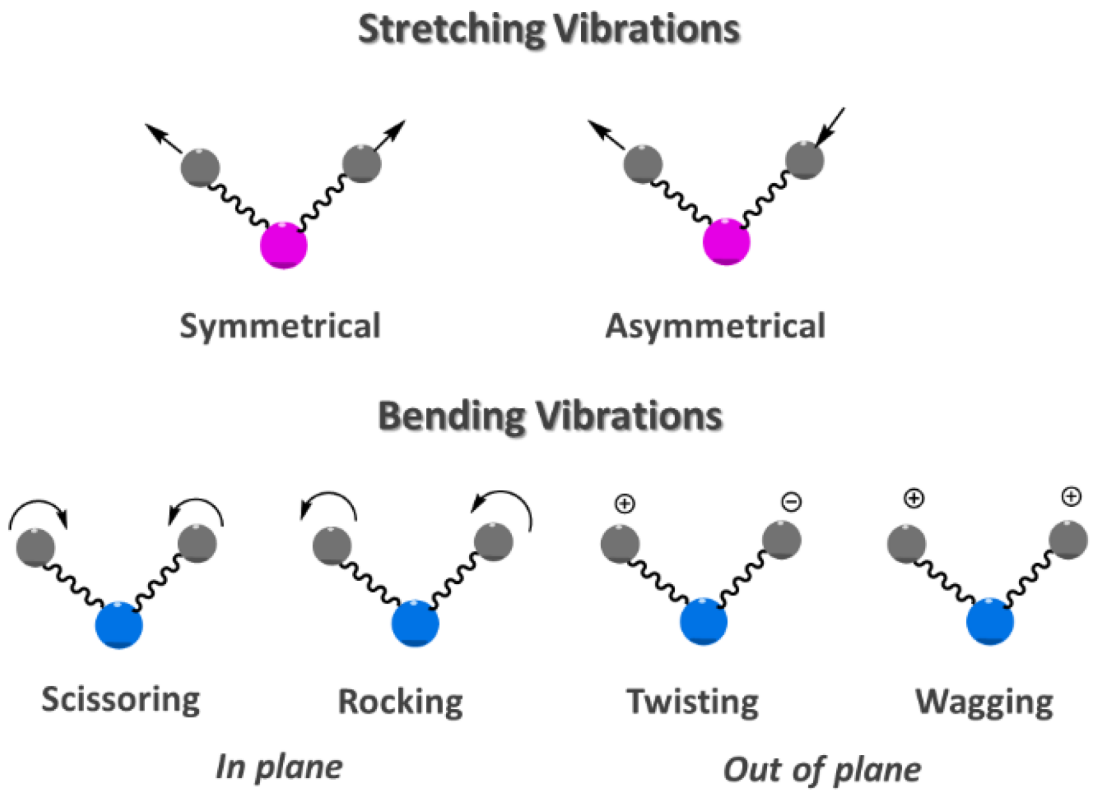


Figure 1.11: A diagram showing the stretching and bending vibrational modes. Symmetric motions tend to be Raman-active and asymmetric movements are usually IR-active. Taken from Mendes and Duarte.[11]

The intensity of a Raman signal is dependent on the Raman cross-section (σ_{Raman}, units of cm^2) of the molecule: this describes the effective interaction area for a molecule to generate a Raman shifted photon. This cross-section is a factor in determining the Raman intensity, given an input laser intensity:

$$I_{\text{Raman}} = I_0 \cdot \sigma_{\text{Raman}} \cdot D \cdot \mathrm{d}z. \tag{1.1}$$

In equation 1.1[99] , I_0 is the laser intensity, D is the number density of scatterers and $\mathrm{d}z$ is the path length of the laser in the sample. The power of the Raman signal is proportional to the intensity of the laser beam, and the inverse fourth power of the photon wavelength:

$$P_{\text{Raman}} \propto I_0/\lambda^4. \tag{1.2}$$

This indicates that reducing the laser wavelength would increase the Raman power,

however it also creates a problem with fluorescence excitation.

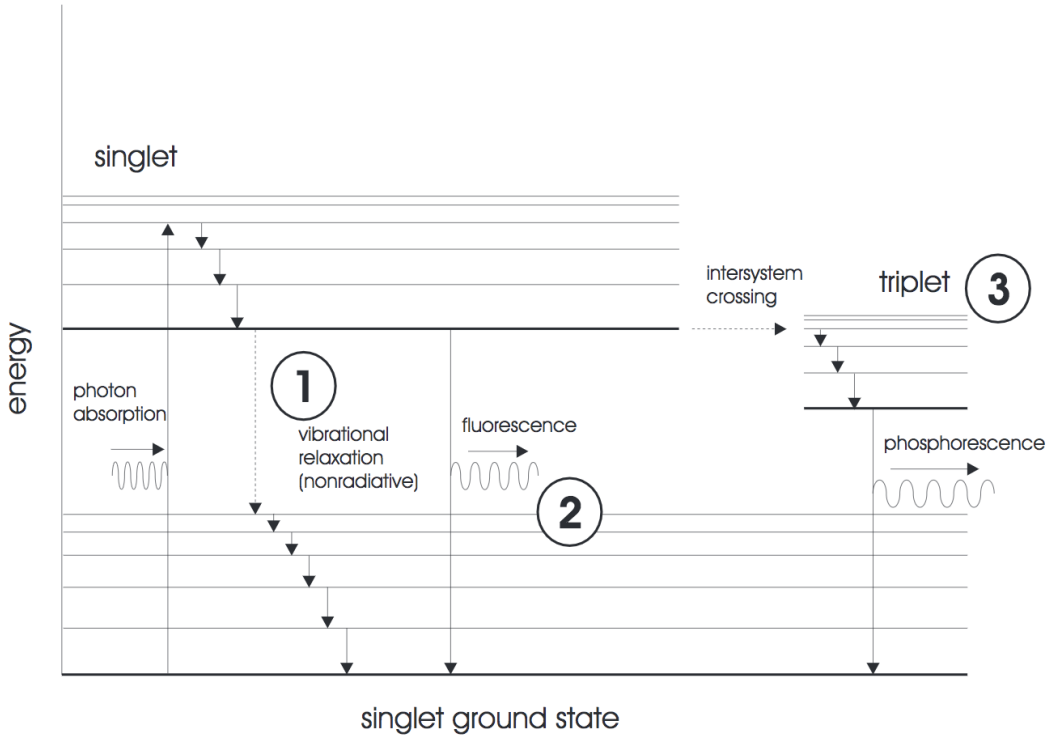


Figure 1.12: Three pathways for an excited state to return to the ground state. (1) is non-radiative vibrational relaxation, (2) is fluorescence: short-lived photon emission and (3) is phosphorescence: long-lived photon emission. Taken from Cox.[12]

Figure 1.12 shows three possible paths for an electron in the first excited electronic level to de-excite back to ground level. The first (labelled 1 in figure 1.12) is vibrational relaxation which is a non-radiative process where the electronic energy is converted into vibrational or rotational energy for the molecule. Through collisions, this energy is transferred to kinetic energy in other molecules: this is defined as a temperature rise. This occurs quickly, within picoseconds (10^{-12} s).

The second deactivation path (2 in figure 1.12) is fluorescence, which can swamp the Raman signal if the laser wavelength is too close to the excitation energy required by the molecule. The choice of a near IR (NIR) laser can prevent this by being too low in energy or a UV laser by separating the emission from the Raman signal in energy. Fluorescence occurs when there is some vibrational relaxation down to the first electronic energy level, and then one sudden jump down to ground level. This releases a photon with a slightly different energy to the incident photon. This process occurs within around 1–100 nanosec-

onds after absorption. Fluorescence can be identified from Raman by changing the input laser energy: Raman peak positions are unaffected by the incident photon wavelength, fluorescence is affected and peak positions will shift.

The third path (3 in figure 1.12) is phosphorescence and this takes the longest to occur, on the order of milliseconds to seconds. This is due to the intersystem crossing required. In this case, the excited electron transitions from a singlet excited state to a triplet excited state: the spins change from being opposite to being parallel. This can happen in molecules that have singlet and triplet states with very similar energies. This transition is "forbidden" according to the selection rules, and thus is unlikely to occur (molecules with a heavy atom are more likely to violate spin selection rules and phosphoresce).

1.2.2 Instrumentation

After the discovery of the Raman effect in 1928, there were still difficulties in widespread and commercial use for it. This was primarily due to the equipment that was (and was not) available at the time. During World War II, infrared spectroscopic techniques became hugely improved due to developments in equipment; Raman still required darkrooms and specially trained operators. Once the laser was invented in the 1960s,[100] Raman spectroscopy began to take off in earnest.

The instrumentation required to perform Raman spectroscopy has remained unchanged in principle; the basic set-up can be seen in figure 1.13. C.V. Raman used focused light from the Sun for his original signal; early spectrometers used mercury arc lamps; modern spectrometers use lasers due to their high intensity and monochromatic photons. Figure 1.13 shows several components which are key to a dispersive Raman spectroscopy set up: laser light source, sample to be analysed, filter to remove the more intense Rayleigh scattering, diffraction grating to disperse the light and a CCD (or other detector) to detect the light and record a spectrum.

1.2.2.1 Light sources

Any type of electromagnetic radiation can produce Raman shifted light by interaction with suitable sample molecules; currently lasers are used because they are powerful enough

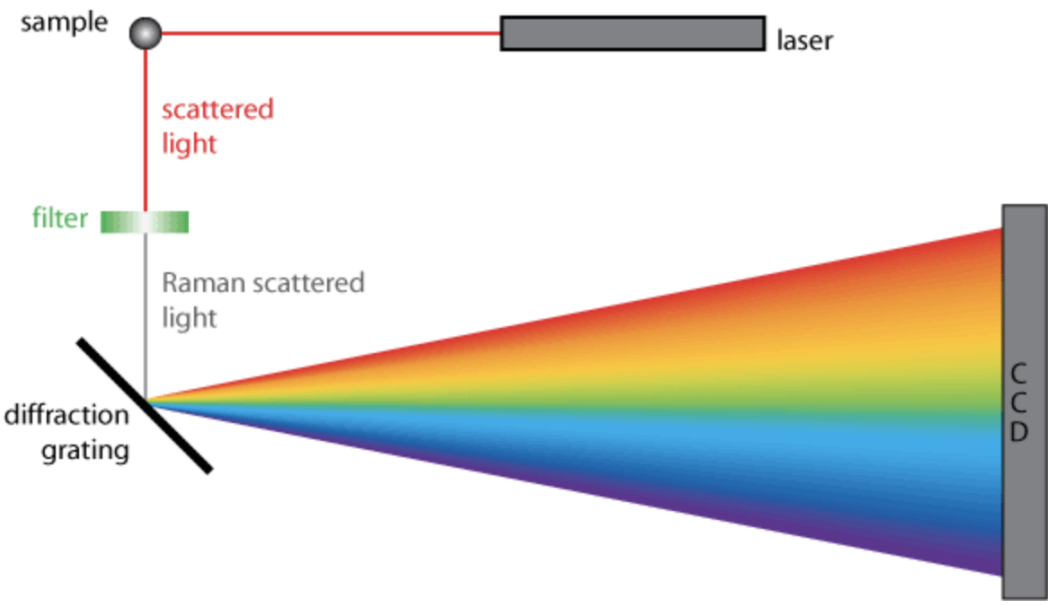


Figure 1.13: Simplified graphic to show the main components of a Raman spectroscopy setup. Taken from Sandeman.[13]

to generate a detectable level of Raman photons. Different laser wavelengths have their own advantages and disadvantages, and the material of the sample dictates a lot about the choice. Raman spectroscopy systems use lasers with wavelengths from UV through to near infrared: < 300 nm through to 1700 nm. Shorter wavelengths are a popular choice since the Raman intensity is related to inverse fourth power of the laser wavelength, as discussed earlier. NIR is also good for fluorescence suppression in biological samples and can penetrate deeper.

1.2.2.2 Optics

Once light has left the chosen source, lenses (and occasionally also filters) are used to guide illumination onto the sample. A bandpass filter can ensure that minimal laser spectral wings or ambient light are incident onto the sample. Laser spectral wings are a physical phenomenon which broaden the emission profile of the laser due to quantum effects: the light emitted from a laser is not perfectly monochromatic. Additionally, collection optics are required to collect the scattered light from the sample and direct it to a detector.

Notch or edge filters are usually used to remove the overwhelming Rayleigh scattered signal and allow the weak Raman light into the detector. An edge filter blocks all wavelengths up to a certain point and then transmits all light above this limit with high

efficiency. These will block Rayleigh and anti-Stokes light. A notch filter has a sharp absorption which can be a few nanometres wide, and is designed to absorb the laser light only and hence transmit both the Stokes and anti-Stokes light to the detector.

A dispersion grating is used to disperse the light onto the detector. Dispersion gratings are defined by the periodicity of the grooves, in units of grooves per mm (g/mm). The higher the groove density, the greater the spectral resolution but the narrower the range. A trade-off is required to balance these two experimental aspects.

1.2.2.3 Detectors

The detectors used in a dispersive Raman spectroscopy system are CCDs - charge coupled devices. These are 2D arrays of pixels, where each pixel can be considered as an individual detector. They are usually made from silicon or another photosensitive semiconductor. The Raman signal is registered via electron-hole pairs; the number of photons incident on a pixel is directly proportional to the charge acquired by the pairs, and this is called the gain. An important characteristic of a CCD is the quantum efficiency curve, which shows the probability of generating a measurable photoelectron as the incident energy changes. These usually peak around 600 nm and zero above 1000 nm and below 250 nm.[101] This has an impact on the laser wavelengths that can be used. Deep depletion CCDs are generally used in order to maximise quantum efficiency in the NIR region.

CCD performance can be determined by the generated noise. Noise typically falls into three types: dark current, readout noise and shot noise. Dark current is the spontaneous formation of electron-hole pairs by thermal effect (instead of incident radiation). This effect has a large dependency on temperature: CCDs are usually operated with active cooling to reduce this effect. The dark current approximately doubles for every 5°C increase in temperature.[102] Readout noise is the error associated with digitising the electrons. It is generally independent of the number of electrons being counted and only becomes an issue when the magnitude of the signal is comparable to the magnitude of the noise (on the order of a few electrons). Shot noise is the term for the statistical process that is the CCD detecting photons. Poisson statistics describe the intervals between photon arrivals and so shot noise is equivalent to the square root of the signal.

1.2.3 Spectroscopic techniques

There are various different techniques available to improve the Raman signal collection. The specific optical set-up that is chosen varies with the material being probed and the depth at which the Raman signal originates from. Moreover, the clinical application plays a role since different body parts and organs have varying levels of accessibility!

1.2.3.1 Raman microscopy

Conventional Raman microscopy uses a confocal microscope to illuminate and collect the Raman light. This means the same objective that focuses the laser light onto the sample is used to collect the scattered light: this is called backscattering geometry. This method has a very shallow penetration depth (on the order of 1–500 microns) and can suffer from strong fluorescence swamping the Raman signal. The depth of Raman photons detected in backscattering geometry is approximately 50 times the scattering mean free path[103] ; this dependence means the penetration depth in biological tissue (a turbid medium) is very low. For the purpose of microcalcification identification at depth in breast tissue, methods of deeper penetration are required.

1.2.3.2 Deep Raman spectroscopy

Deep Raman spectroscopic techniques have been developed to increase the penetration depth and to expand the applicability, especially for non-invasive testing. There have been two main paths to this depth increase: temporal gating and spatial offsets.

1.2.3.2.1 Temporal gating

Temporal gating exploits the different flight times of photons from near the surface escaping compared to those photons which penetrate deeper into the medium. A Kerr gate is a common choice for temporal gating as it works as an optical “switch” on an ultrafast time scale (on the order of picoseconds). Kerr gates have two crossed polarisers and a Kerr medium. When the gate is closed, light is blocked by the cross polarisers. When Raman light reaches the gate, it can be turned on to bypass the polarisers, and the Kerr medium rotates the light from the sample by 90°, allowing it to be transmitted through the

cross polariser for the duration of the pulse. Since Raman scattering is an instantaneous effect and fluorescence takes longer, this can be used to dampen the fluorescence effect. An example experimental set-up schematic can be seen in figure 1.14.

Through employment of Kerr gating, depth penetration of ~1mm has been achieved.[14] This method of temporal gating requires pulsed lasers, on the order of picoseconds. It also requires an expensive and high-powered set-up which can result in sample damage due to the laser powers used and is thus often unsuitable for biomedical applications.

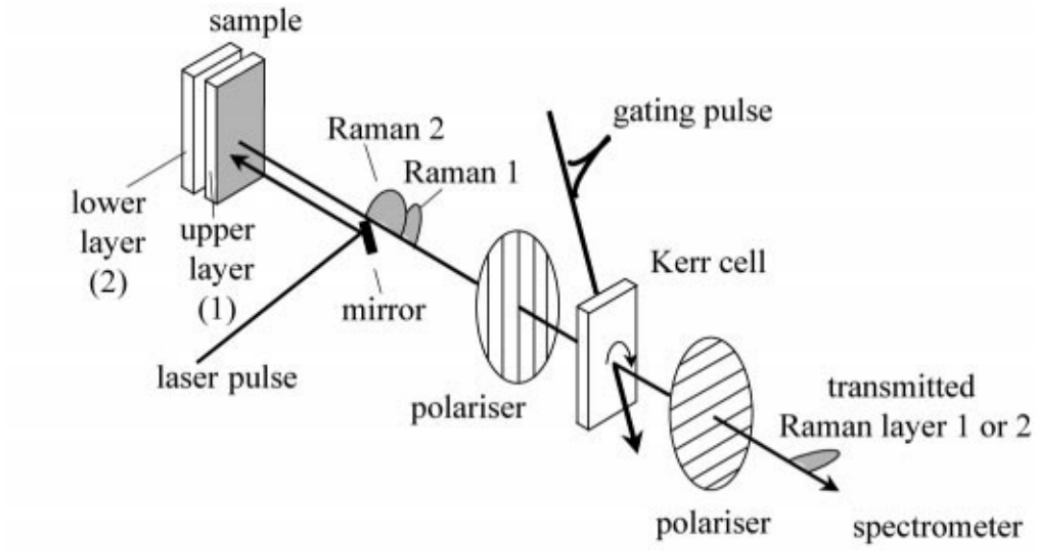


Figure 1.14: Schematic of Kerr gate experiment to probe subsurface layer Raman signals. Taken from Matousek et al.[14]

1.2.3.2.2 Spatial offset

In 2005, Matousek et al. developed a new approach to improve the penetration depth[104] - spatially offset Raman spectroscopy (SORS). This method uses continuous wave laser excitation and laterally offset detection. SORS works on the principle that the deeper a Raman photon is generated, the greater the range of locations it will exit from in the backscattering direction. Therefore, the collection furthest from the incident beam will detect the deepest generated Raman photons. SORS is traditionally performed with the laser light brought onto the sample in a small spot, and collection is via an annulus of variable radius, as shown in figure 1.15 (a). A variation on this technique is inverse SORS (i-SORS) where the laser light is illuminating in the ring shape and the collection is the

small point in the centre area. This is displayed in figure 1.15 (b). i-SORS is useful because a higher powered laser can be used due to the beam being spread over a greater area; this reduces the damage that would ordinarily be done by the same laser in a small area.

SORS (and i-SORS) have limited usefulness in the case of material with high absorption coefficients as the total photon flight path is reduced and thus the detectable depths penetrated will be shallower. Moreover, if fluorescence is high in the sample material this can pose a problem by drowning out the Raman signal. However, if the strong fluorescence is coming from the surface layer (and not the deeper layer of interest), SORS can effectively suppress it.[105]

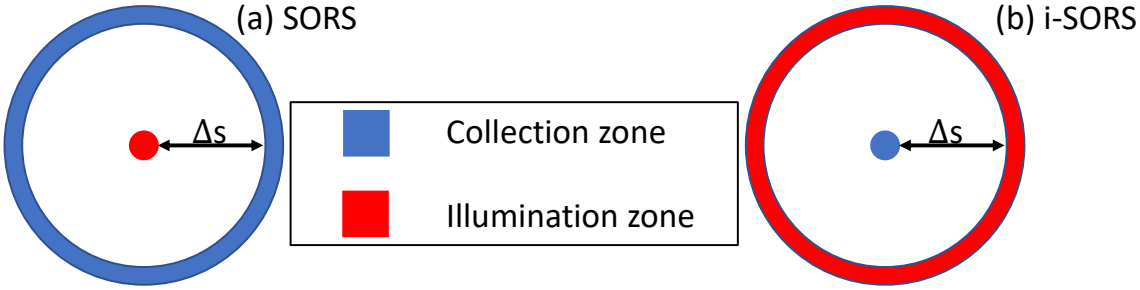


Figure 1.15: Diagram of (a) SORS and (b) i-SORS geometries. The illumination and collection zones are reversed from traditional SORS to perform inverse SORS. Δs is the spatial offset.

The final form of deep Raman spectroscopy that shall be discussed can be considered to be the "most extreme" form of SORS: transmission Raman spectroscopy, TRS. This set-up has the illumination on one side of the sample, and collection on the opposite side. The difference between a SORS set-up and a TRS set-up is shown in figure 1.16. During TRS, the laser light propagates through the entire sample, and Raman photons can be generated anywhere in the volume. The Raman signal that is measured on the opposite side could have been generated at any point, and thus this technique is representative of the Raman properties of the bulk sample volume, rather than discriminating between layers as SORS is effective for. Transmission Raman is effective at suppressing surface Raman signals and excess fluorescence.[106]

The choice of laser wavelength for performing SORS or TRS is a delicate one. Typically, the excitation wavelength for biological samples is chosen so both the illumination and collection wavelengths are in the NIR regime: 650–900 nm. This range is called the "first NIR window" and is chosen due to the absorption response of biological compo-

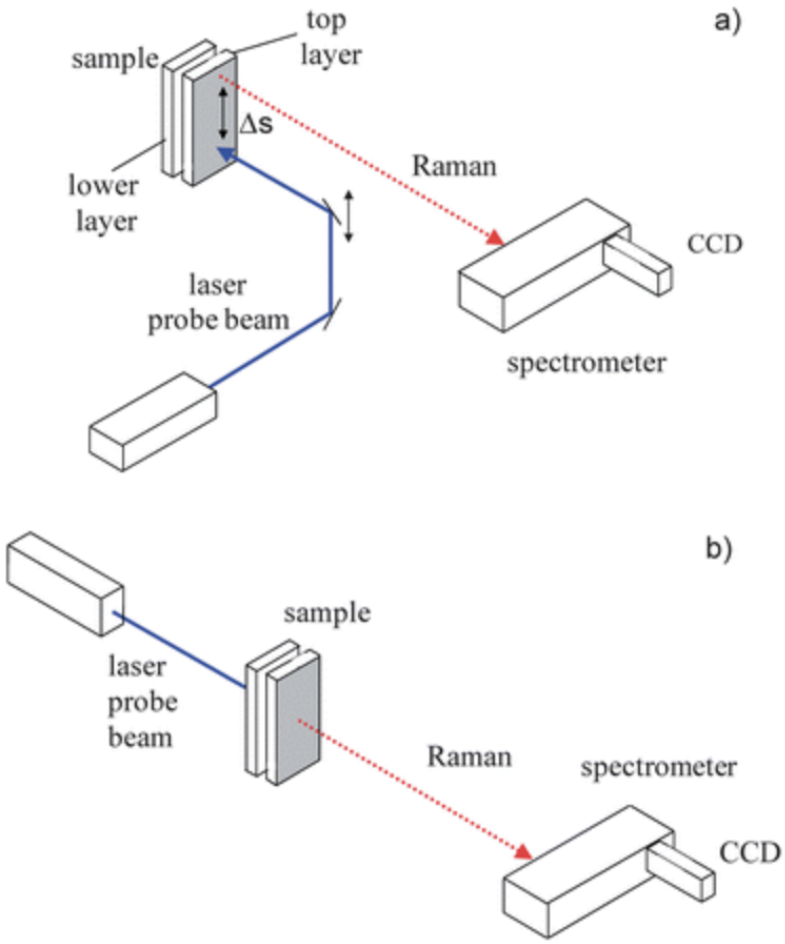


Figure 1.16: Simplified schematic of (a) SORS and (b) TRS set-ups. From Matousek and Stone.[15]

nents at these wavelengths, as can be seen in figure 1.17. Figure 1.17 (a) shows how the attenuation of various biological components varies with wavelength: oxygenated and deoxygenated blood, skin, and fat all have some of their lowest attenuation in the first and second windows. Additionally, figure 1.17 (b) shows how the quantum efficiency of different photodetectors changes with wavelength. As discussed in section 1.2.2, Raman spectroscopy uses CCDs for detection, and these are silicon based devices. Thus, it is clear that the first NIR window is the most appropriate region to target for illumination and detection. This minimising of absorption by controlling the wavelengths used maximises the potential for SORS to be useful in Raman spectroscopy applied to biological tissues.

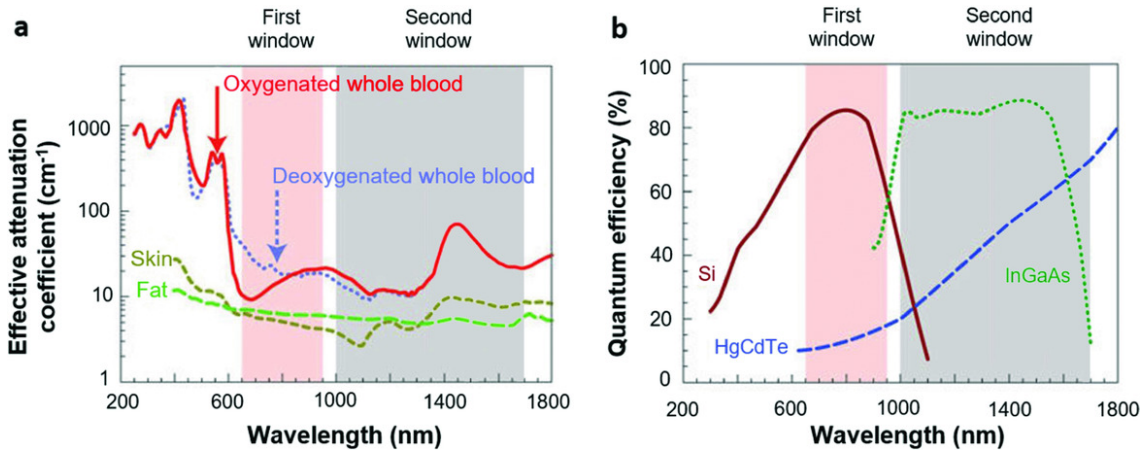


Figure 1.17: Properties of biological tissues and photodetectors in the broad optical window. (a) the attenuation of biological entities such as both oxygenated and deoxygenated blood, skin and fat. (b) the quantum efficiency of different photodetectors. Combining the information from these justifies the use of the first NIR window. From Duan.[16]

1.2.4 Raman spectroscopy in cancer diagnosis

Raman spectroscopy is suitable for both qualitative and quantitative analysis of tissues due to its ability to identify a "fingerprint" of Raman active bonds present and their relative quantities (the strength of the Raman signal is directly related to the concentration of the molecules). Therefore, this can be a useful method for real-time, non-invasive measurements of potential disease, given a suitable biomarker.

The variety of optical geometries allowing the development of deeper sample probing has led to different set-ups being more suited to certain types of samples. Conventional Raman microscopy is primarily used for looking at excised tissue sections or biological fluids. Examples include brain tumour assessment,[107,108] identifying biomarkers in ovarian cancer tissues,[109] prostate cancer cell lines[110] and plasma,[111] and pancreatic cancer detection.[112]

The primary use for Raman spectroscopy that this thesis is interested in is for improving the diagnosis of breast cancer by using microcalcifications as a biomarker. Raman spectroscopy has been demonstrated to be useful for identifying molecules (particularly inorganic compounds) and thus could be used to differentiate between type I and type II microcalcifications in breast tissue. These calcifications are buried in the tissue and so, to be identified *in vivo*, a deep Raman set-up is required. The current deepest measurement of HAP using TRS is 40 mm from Ghita et al.[113] A study[114] showed the feasibility

of differentiating between the calcification types at depth using SORS, although not at clinically relevant sensitivities. SORS and TRS being employed in this way are fairly novel techniques and are the focus of this thesis.

Often, these feasibility studies are done on tissue phantoms, or homogeneous materials such as chicken breasts or porcine tissues. These are useful for demonstrating potential penetration depth, however being able to investigate how heterogeneity and irregular shapes could impact the availability of Raman signal would be a useful tool. Additionally, it would be beneficial to explore how to optimally design a system to maximise sensitivity for small differences in signal at depth. In the next chapter, a statistical modelling technique called the Monte Carlo method in its theory and applicability to these situations will be discussed.

1.2.5 Summary

In this chapter, the impact of breast cancer rates globally and the physiology of the breast environment were discussed. This was important for giving the background material required to understand the current methods of detecting breast cancer. Successfully treating cancer relies on detection of the disease as early as possible, and current imaging diagnostic techniques rely on changes that arise from later stages of the disease.

Needle biopsy is currently the only method of diagnosis which can show changes on a molecular level that can be linked to the progression of the disease. However, biopsies are invasive and can have large rates of false negatives, as well as being expensive and most results being benign. Moreover, there are issues with consensus between different pathologists assessing the biopsies.

There is, therefore, a need for another diagnostic technique which could highlight earlier signs of breast cancer and determine whether microcalcifications present in breast tissue are associated with malignant or benign lesions. A promising method of delivering this is through Raman spectroscopy: an optical technique which uses inelastic light-matter interactions to provide a molecular fingerprint in a label-free, chemically specific manner.

Raman spectroscopy has the potential to provide real-time, *in vivo* measurements

of large tissue volumes, however understanding how the light sources and detectors are set up is vital to optimising a probe design. Additionally, investigating how the microcalcifications are distributed and the impact of location in relation to the input laser beam are crucial to inform future work in bringing Raman spectroscopy to the clinic. In order to explore all of these parameters efficiently, Monte Carlo modelling can be a useful tool to simulate light-tissue interactions accurately and quickly.

The work in this thesis is intended to provide insight into the issue of how to optimise a probe capable of differentiating between microcalcification types at depth in breast tissue. The aim of the thesis is to first develop a Monte Carlo code capable of replicating Raman scattering. The method to this is described in Chapter 2, and the code is benchmarked against a known and published code in Chapter 3. Following this, the code is utilised to model Raman scattering in tissue phantoms, detailed in Chapter 4 and provides useful insight into the optical properties of Intralipid. Finally, the objective to simulate calcification distributions and a variety of potential illumination and detection geometries is achieved and discussed in Chapter 5. The outcomes of the work presented here can be used to inform probe design and move towards a real-time, non-invasive cancer diagnosis.

Chapter 2

Monte Carlo Modelling

This chapter contains a background to Monte Carlo Radiative Transport (MCRT) and how it can be applied to the problem of Raman signals at depth for cancer diagnosis, as outlined in chapter 1. The chapter then goes on to describe the specific code (`arctk` used for the modelling and simulations in subsequent chapters. Monte Carlo (or MC) modelling is able to accurately simulate light-matter interactions within turbid media, making it a powerful tool for predicting how a Raman signal could propagate through breast tissue.

The work in this thesis uses the `arctk` toolkit originally developed by my colleague Dr. Freddy Wordingham*. Wordingham developed this library during his PhD and used it to investigate drug diffusion and photodynamic therapy for skin cancers; he was responsible for the base library and the framework for enabling high performance simulations. I was responsible for the design, implementation, and development of the scientific aspects of the code, ultimately making it possible to model Raman scattering in turbid media.

The design for the MCRT application came from the work by Tim Harries and Sid Visser to implement triangular meshes into the astrophysics code TORUS, which works on an adaptive-mesh-refinement grid. The `arctk` codebase is freely available at https://github.com/FreddyWordingham/arctk and a full description of the development and the skin cancer work can be found in the thesis of F. Wordingham, University of Exeter. The code has been used in a previous publication by Jeynes et al.,[115] in which I am an

*and dream-team science best friend

author and contributed to code development and editing of the manuscript. The study by Jeynes et al. was investigating nanoparticles and heat deposition for photothermal skin cancer therapy, so coupled a Monte Carlo routine to a heat diffusion calculation to estimate cell death. The code is written in Rust, a modern language which is strongly-typed and memory safe. The next section looks at the history of the Monte Carlo method.

2.1 History

2.1.1 Statistical sampling

Monte Carlo method is a numerical technique that is based on random numbers, used to estimate outputs of problems. One of the earliest documented uses for the method is that of Buffon's needle in the 18$^{\text{th}}$ century.[116,117] This is a useful example for instructional purposes. The needle problem that was posed and solved by Buffon was as follows:

"Let a needle of length L be thrown at random onto some horizontal plane ruled with parallel straight lines a distance d apart, where $d > L$. What is the probability that the needle will intersect with one of the lines?"

Assuming that "at random" means the needle centre can be at any position and all orientations of the needle are equally probable. Defining the distance from the centre of the needle to the nearest line as x and the orientation defined by the angle θ, we need only consider a single line as each needle is only compared to its closest. These definitions are based on the diagram shown in figure 2.1.

From examining figure 2.1, it can be seen that for the needle to cross the nearest line, $x < \frac{L}{2} \sin\theta$. x is uniformly distributed from $[0, \frac{d}{2}]$ and θ in $[0, \frac{\pi}{2}]$. Using a probability distribution function for each variable, a combined probability can be found.

A probability distribution function (PDF) is a function of a variable that gives the probability of that variable taking a given value. The PDF is normalised over the range of its variable. For this problem, the variables are x and θ, and as these are independent of each other we can multiply them in order to find the total probability of a needle crossing a line. The PDFs are $p(x) = \frac{2}{d}$ and $p(\theta) = \frac{2}{\pi}$; thus the joint probability density is given by

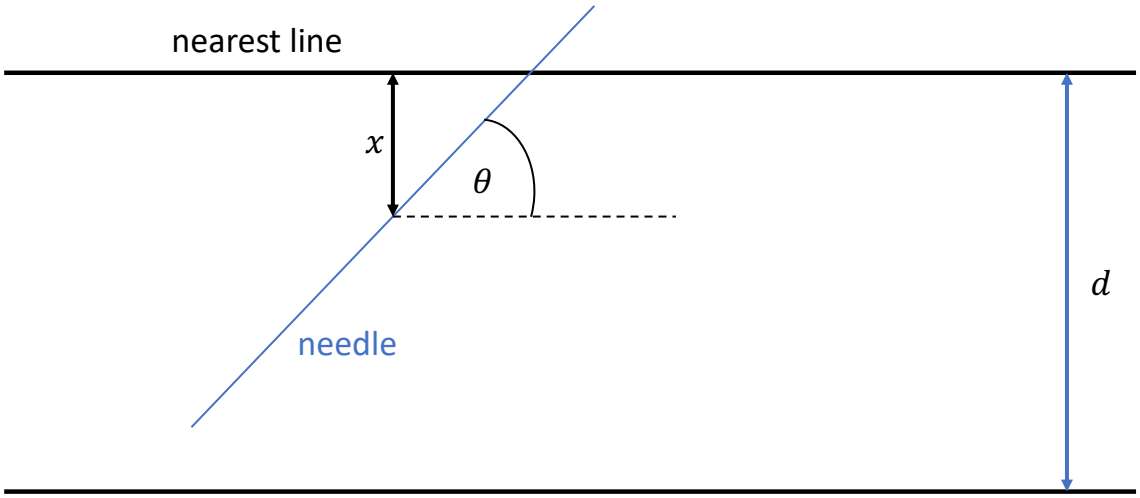


Figure 2.1: Sketch to visualise Buffon's needle problem. The black line is the nearest parallel line to the needle. The needle is represented by the blue line. The terms defined in order to solve the question of the probability of the needle crossing the line are shown: x, the distance from the needle centre to the closest line, and θ, the orientation of the needle.

$p(x, \theta) = \frac{4}{\pi d}$. Finally, the probability is found by integrating the joint probability density over the range of the variables:

$$P = \int_0^{\frac{\pi}{2}} \int_0^{\frac{L}{2}\sin\theta} \frac{4}{\pi d}\, \mathrm{d}x\, \mathrm{d}\theta = \frac{2\,L}{\pi\, d}. \tag{2.1}$$

Equation 2.1 can be rearranged to perform a Monte Carlo method estimation of π. The result yields:

$$\pi = \frac{2\,L}{d\,P} \tag{2.2}$$

where P is the ratio of needles crossing a line to total number of needles dropped. Laplace was the first to recognise that Buffon's result could be used as a π estimator.[117] A very large number of needles must be dropped in order to get an accurate estimation of π. It became a pastime of some people to throw thousands of needles to get the first decimal places, including a Captain Fox while he was recovering from wounds he obtained in the American Civil War.[118] This method is, however, incredibly inefficient: the probability of obtaining π correct to 5 decimal places in 3400 needle throws is less than 1.5%.[117] Post-electronic computing, this idea of randomising variables and observing the experimental

outcome became incredibly powerful.

There are several different ways of using a Monte Carlo method to determine a value. The inverted sampling method is an analytical process to generate variables from an arbitrary distribution by drawing uniform random variables. The inversion method is performed as follows:

1. Compute the cumulative distribution function (CDF) by integrating the PDF giving $P(x)$.

2. Generate a uniformly distributed random number ξ, in the range $[0, 1)$.

3. Calculate the inverse: $P^{-1}(x)$.

4. Calculate the sample: $X_i = P^{-1}(\xi)$.

For example, if we want to sample $\cos(x)$ then we must link $P(x)$ to uniform random number ξ and then invert to get a $\cos(x)$ distribution from uniform random numbers.

$$\xi_i = \int_0^{x_i} P(x)\,\mathrm{d}x = \sin(x_i) \rightarrow x_i = \sin^{-1}\,\xi_i. \tag{2.3}$$

From equation 2.3, it can be seen how drawing uniform random numbers can be transformed into sampling from any arbitrary distribution: in this case, building up $\cos(x)$.

There are occasions where it will not be possible to use the inverse sampling method, such as not being able to get an analytical representation of a PDF or to invert the CDF. In these cases, we can use the rejection method, which does what it says on the tin. Points are randomly chosen and compared to the desired function: if the point is under the function it is accepted, otherwise it is rejected. A simple example of this is evaluating the area of a circle (radius 1 unit):

1. Generate a random point within a 2 unit length square.

2. If point lies within 1 unit of the square centre, add to a count.

3. Repeat sufficient times to sample the area.

4. Area of circle = (fraction of points in circle/total points) multiplied by the area of the square.

5. (bonus: this method can also be used to estimate pi: $\pi = \frac{4\,N_{\text{circle}}}{N_{\text{total}}}$.)

Rejection method is effective for sampling complicated functions, if time consuming. The rejection method can be used for any complicated shape and does not require anything to be "solved", just to know whether a point lies inside or outside a defined area or function. An example of this can be seen in figure 2.2. An arbitrarily complicated function $f(x)$ can be sampled by generating a uniform random number x in $[0, x_{\text{max}}]$ and another uniform random number, y, in $[0, 1]$, provided $f(x)$ is normalised. If $y \leq f(x)$, then we accept the point, otherwise we reject it. In the example of figure 2.2, the point (x_1, y_1) is acceptable, whereas (x_2, y_2) has no such luck and is rejected.

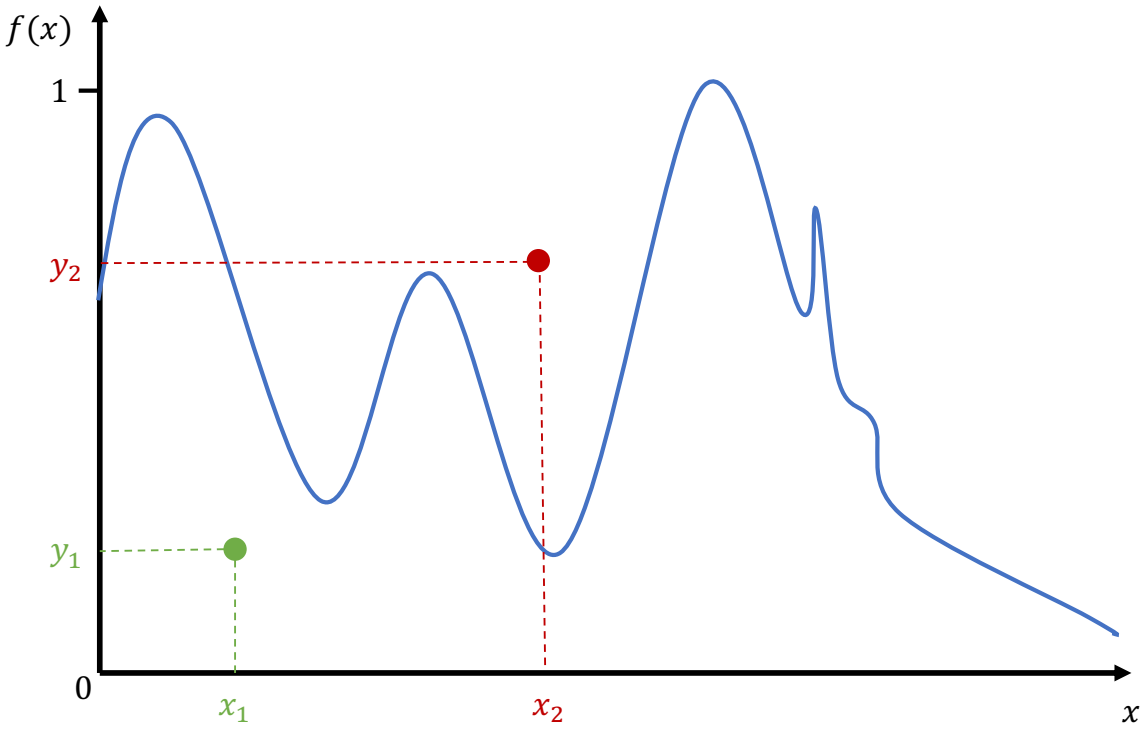


Figure 2.2: An example of rejection method. A complicated function can be sampled by uniformly sampling x values, and for each generating a corresponding uniform random y. Where $y \leq f(x)$ is satisfied, the x value is kept, otherwise it is rejected. Here, we see that the point (x_1, y_1) is acceptable, but (x_2, y_2) is not and will be rejected. By repeating enough times, the original function can be reproduced.

These sampling techniques are examples of Monte Carlo methods: relying on drawing random numbers to replicate a distribution and inform on a problem that cannot be easily solved analytically. Monte Carlo methods have been used in a variety of disciplines:

in the financial sector to model sources of uncertainty in portfolios and investments affecting their value,[119,120] in the applied statistics branch of mathematics[121] and in weather forecasting,[122] to name a few. For physical sciences, a specific type of Monte Carlo simulation is used: Monte Carlo Radiative Transfer (MCRT). MCRT is used widely in astronomy[123,124] and models how radiation propagates through turbid media and is discussed in more detail later in this chapter. First, the history of Monte Carlo simulations will be covered.

2.1.2 Neutron bombs

The modern applications of Monte Carlo modelling to solve radiative transport problems began in the 1930s with work by Italian physicist Enrico Fermi who developed statistical sampling to predict experimental results in nuclear physics experiments. Later, he invented the FERMIAC to aid in the solving of neutron transport. The FERMIAC is described as an analogue computer, a computer that performs analogous behaviour to that which it is simulating. A replica of it can be seen in figure 2.3. It has three main components to allow it to simulate neutron transport. On the left is the lucite platform which selects the direction of the neutron; in the centre is the rear drum which measures how much time has passed based on the velocity of the neutron; and on the right is the front drum which measures the distance travelled by the neutron between collisions from the velocity of the neutron and the properties of the material it is travelling in.[17] Electronic computers were in their infancy, but the transport of neutrons and the development of the neutron bomb was rapidly taking off. The FERMIAC worked by using a random number table to determine the path of the machine, replicating the path of neutrons through dense material.

Fermi was employed by the Manhattan Project to further his group's work into nuclear fission during World War II due to the perceived threat of Germany beating them to the construction of a nuclear weapon. In 1943, theoretical physicist Robert Oppenheimer was appointed director of the Los Alamos Laboratory in New Mexico, where the first bombs were built and tested. The nuclear weapons "Little Boy" and "Fat Man" were among the bombs built there. There were several key figures in the research that made the construction of these bombs possible with Monte Carlo techniques: Nick Metropolis,

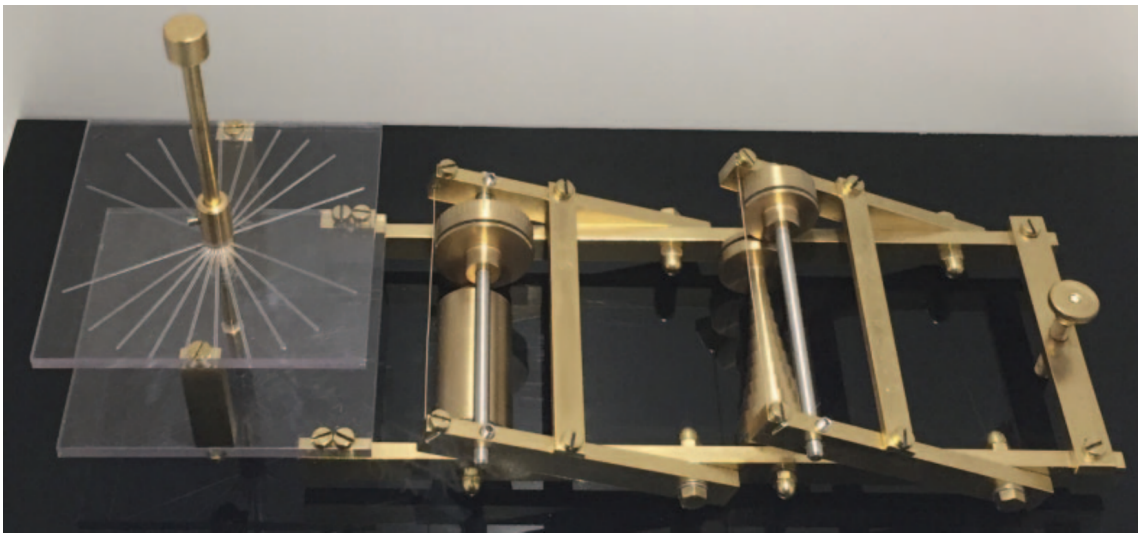


Figure 2.3: Replica of the FERMIAC machine, built in Bologna in 2015. There are three main components, from left to right: the lucite platform to select the direction of the neutron, the rear drum measures the elapsed time from the neutron velocity and the front drum measures the distance the neutron travels based on the velocity of the neutron and the material it is travelling through.[17]

Edward Teller, Stanislaw Ulam and Jon von Neumann. The name "Monte Carlo" comes from when Ulam was ill and resigned to playing solitaire in bed. He wondered if he could calculate his chances of success in a game, by playing many games and counting the number of successful plays. He then extrapolated this method of statistical sampling to the neutron transport work at Los Alamos. The work needed a code name, and Metropolis suggested Monte Carlo after the famous casino in Monaco where Ulam's uncle was known to gamble (with his relatives' money)[125]!

The work carried out by this team was performed on the ENIAC (Electronic Numerical Integrator And Computer), the first electronic computer. The time between its creation and the move to its permanent home was when Fermi invented his FERMIAC. Once the ENIAC was in place, it was capable of performing Monte Carlo calculations at great speed. After the ENIAC came the MANIAC (Mathematical Analyser Numerical Integrator and Automatic Computer model), so named by Metropolis in an attempt to end silly namings...[126] The first paper from this period which established the Monte Carlo method was by Metropolis and Ulam, in 1949.[127]

2.1.3 Modern Monte Carlo in biological systems

Monte Carlo simulations became more viable for complex problems as computers advanced in power. One field which appreciated and advanced MC algorithms was astrophysics: limited to only observing electromagnetic radiation from their subjects, modelling how that radiation interacts with the environment between Earth and the body of interest was incredibly useful. Rather than modelling (relatively) large particles such as neutrons, this work involved modelling the behaviour of photons, the particles of light. One of the earliest and thorough reports of Monte Carlo being combined with radiative transfer in astronomy is in House and Avery (1968)[128] when investigating resonance-line scattering.

Applying MCRT to biological systems began with Adam and Wilson (1983)[129] who modelled the simplest case: isotropic scattering in homogeneous tissue. In 1987, anisotropic scattering and cylindrical co-ordinates were implemented by Keijzer et al.[130] Work by Prahl et al. (1989)[131] reconfigured the program to have the photon propagation in Cartesian co-ordinates, which improved the written form to be easier to understand. Wang and Jacques,[132] in 1992, adapted these works to develop the Monte Carlo Multi Layered (MCML) code which allows simulation of several planar layers, each with different optical properties. The MCML simulation has been widely used and adapted by research groups, continuing to be popular due to its flexibility allowing a wide variety of scenarios to be modelled.

There have been many MC codes for diffuse reflectance and fluorescence in layered biological systems,[132,133] however there are fewer for Raman scattering, and for microcalcifications buried in breast tissue specifically. Matousek et al.[21] developed an MC model that includes Raman scattering in layered turbid media for pharmaceuticals. Pharmaceuticals are similar to biological tissues in their turbidity, but differ in their constitution (fewer layers), scale size and sources of Raman signal. Schulmerich et al.[134] completed a study of Raman tomography, both experimental and simulation for imaging canine bone tissue. In the human body, Enejder et al.[135] developed a Raman MC simulation to predict how system geometry will affect the Raman signal collected from blood. Another

study, by Shih et al.[136] used Monte Carlo simulations to validate their work on correcting turbidity-induced Raman spectral distortions.

There have been investigations for a variety of Raman signals arising in breast tissue. Work from Keller et al.[137] developed MC simulations to replicate SORS optical geometry for investigating breast tumour margins. This was one of the first studies to model Raman scattering in a layered biological medium using SORS. Wang et al.[138] simulated a multi-layer skin model with Raman MC included to investigate how the different layers contribute to the overall detected Raman signal. Depth sensing is crucial for surgical guidance, and an MC package by Akbarzadeh et al.[139] provides analysis of Raman sampling at depth and replicates experimental results. This work on layered systems has led to studies on inclusions within layers: two works by Periyasamy et al.[140,141] show the feasibility of uncovering "tumours" at depth in breast tissue. These were computationally expensive and had the tumours buried at a depth of only 3 mm which is not similar to the depths of realistic breast tumours. Work from Mosca et al.[142] used MC simulations to find relationships between photon path length, depth and spatial offset for nanoparticles in a 3.6 mm bilayer system with external calibration in a SORS geometry. This study showed the potential for SORS to localise abnormalities at depth in tissue. Another use for Mosca et al.'s MC work was in depth determination of a thin layer using SORS.[143] A computationally efficient code by Dumont et al.[144] has been developed to speed up the simulation of Raman scattering and fluorescence emission in bulk Raman spectroscopy set-ups. It has been proven to replicate experimental results for a homogeneous volume with an inclusion.

The codes discussed here have shown the potential for applying MCRT to biological systems, and astrophysical codes like TORUS[145] have implemented techniques to allow simulation on a broad range of length scales. The previous work in the field is largely limited to parallel layers and basic shapes, and works entirely on a grid. The MCRT tool used in this thesis relies on triangle-meshes which allow for all manner of geometries to be simulated. This means, for example, that the simulation volumes are not limited to plane parallel layers. This allows for greater flexibility and for a developer to run a fully customised numerical simulation of their chosen scenario. The software is well documented

and freely available. The next section will discuss the fundamentals and mathematical framework behind a MCRT code, before going into the specifics of the code used for the work in this thesis.

2.2 Fundamentals and mathematical framework

An MCRT code is a numerical simulation of the radiative transfer equation, which can be solved analytically. Radiative transfer codes replicate the interactions between light and matter, so firstly radiative transfer will be discussed, before exploring how to model these numerically, rather than analytically.

2.2.1 Radiative transfer

The radiative transfer equation (RTE) models the gains and losses of a beam of radiation as it propagates through a medium: loss from absorption, loss or gain via scattering, and gain from emission. To derive the RTE, some terms must first be defined.

Spectral irradiance, L_ν, is defined as the energy flow in a direction $\hat{\mathrm{s}}$ within a solid angle $\mathrm{d}\Omega$ per unit time per unit frequency. Radiance is then the spectral irradiance over a small frequency range $[\nu, \Delta\nu]$ and is expressed as $L(\vec{r}, \hat{s}, t)$. Here, $\vec{r}$ is the position at which the radiance is being measured, $\hat{s}$ is the direction the radiance is being measured in and t is the time it is being measured at. L has the units [W m^{-2} sr^{-1}]. The radiance can be used to determine the energy flow (see figure 2.4) $\mathrm{d}E$ through an area $\mathrm{d}A$ in a solid angle $\mathrm{d}\Omega$ in a time $\mathrm{d}t$ by:

$$\mathrm{d}E = L(\vec{r}, \hat{s}, t) \cdot \cos\theta \cdot \mathrm{d}A \, \mathrm{d}\Omega \, \mathrm{d}t \tag{2.4}$$

where $\hat{n}$ is the unit normal vector to $\mathrm{d}A$ and $\cos\theta$ is the angle between $\hat{n}$ and $\hat{s}$.

The radiance can also be used to calculate the fluence rate, ϕ, which is the energy flow per unit time, independent of direction and measured in units of [W m^{-2}]:

$$\phi(\vec{r}, t) = \int_{4\pi} L(\vec{r}, \hat{s}, t) \, \mathrm{d}\Omega. \tag{2.5}$$

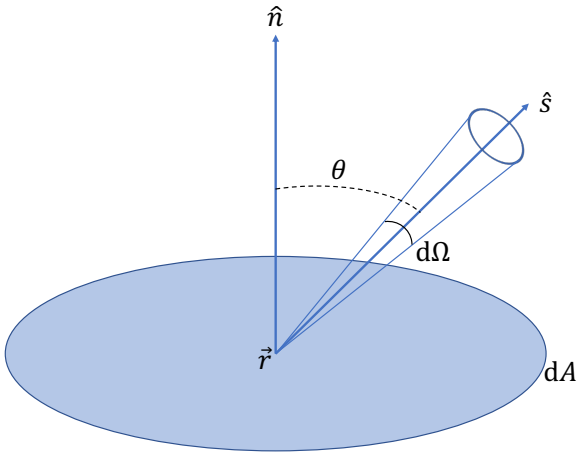


Figure 2.4: Diagram showing the terms involved with defining the energy flow $\mathrm{d}E$ through an area $\mathrm{d}A$ in direction $\hat{s}$. In equation 2.4, the term θ refers to the angle between the direction of interest $\hat{s}$ and the unit normal vector $\hat{n}$. Adapted from Chandrasekhar.[18]

Solving the RTE gives the radiance, hence the RTE can now be derived. Firstly, consider the conservation of energy:

$$\mathrm{d}P = -\mathrm{d}P_{\mathrm{div}} - \mathrm{d}P_{\mathrm{ext}} + \mathrm{d}P_{\mathrm{scatt}} + \mathrm{d}P_{\mathrm{src}}. \tag{2.6}$$

In equation 2.6, $\mathrm{d}P$ is the total change in energy in a volume $\mathrm{d}A\,\mathrm{d}s$ within solid angle $\mathrm{d}\Omega$ per unit time and can be seen in figure 2.5. $\mathrm{d}P_{\mathrm{div}}$ is the the energy loss due to the radiation beam divergence (per unit time) and $\mathrm{d}P_{\mathrm{ext}}$ is the energy loss from scattering and absorption within the solid angle $\mathrm{d}\Omega$ inside the volume element $\mathrm{d}A\,\mathrm{d}s$. $\mathrm{d}P_{\mathrm{scatt}}$ is the energy gain due to scattering from any other direction $\hat{s}'$ into solid angle $\mathrm{d}\Omega$ and $\mathrm{d}P_{\mathrm{src}}$ is gain in energy due to any emission sources within the medium, per unit time. Each of these terms will be discussed and then combined.

The left-hand side of equation 2.6 (the change in energy, per unit time in the volume element) can be expressed as a partial differential of the radiance:

$$\mathrm{d}P = \frac{1}{c}\frac{\partial L(\vec{r},\hat{s},t)}{\partial t}\,\mathrm{d}V\,\mathrm{d}\Omega \tag{2.7}$$

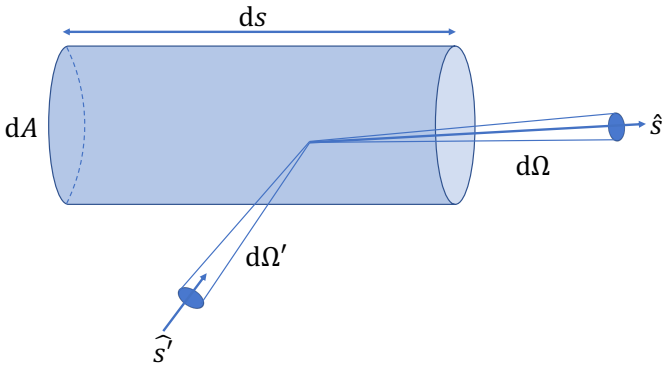


Figure 2.5: Diagram illustrating the terms involved with defining the fluence dE through a cylindrical volume dA ds with solid angle dΩ in a direction $\hat{s}$ and a different solid angle dΩ′ in another direction $\hat{s}'$, in order to derive the radiative transfer equation. Adapted from Wang and Wu.[19]

where c is defined as the speed of light.

Firstly, the energy diverging from the radiation beam, per unit time, can be described as the energy flow out of volume element dV and solid angle dΩ:

$$\mathrm{d}P_{\mathrm{div}} = \frac{\partial L(\vec{r},\hat{s},t)}{\partial s}\,\mathrm{d}s\,\mathrm{d}A\,\mathrm{d}\Omega = \frac{\partial L(\vec{r},\hat{s},t)}{\partial s}\,\mathrm{d}V\,\mathrm{d}\Omega. \tag{2.8}$$

The other loss term is the extinction, energy removed via absorption and scattering combined:

$$\mathrm{d}P_{\mathrm{ext}} = (\mu_{\mathrm{t}}\mathrm{d}s)\,L(\vec{r},\hat{s},t)\,\mathrm{d}A\,\mathrm{d}\Omega = \mu_{\mathrm{t}}\,L(\vec{r},\hat{s},t)\,\mathrm{d}V\,\mathrm{d}\Omega \tag{2.9}$$

where μ_{t} is the extinction or attenuation coefficient, defined as the probability of the photon being scattered or absorbed per unit distance and has units [m^{-1}]. The attenuation coefficient is described more thoroughly later.

Light that has been scattered from any other direction $\hat{s}'$ into the volume element

$\mathrm{d}V$ and the solid angle $\mathrm{d}\Omega$ will contribute to $\mathrm{d}P$ by:

$$\begin{aligned}\mathrm{d}P_{\mathrm{sca}} &= (N_{\mathrm{s}}\mathrm{d}V)\left[\int_{4\pi} L(\vec{r},\hat{s}',t)\,P(\hat{s}',\hat{s})\sigma_{\mathrm{s}}\,\mathrm{d}\Omega'\right]\mathrm{d}\Omega \\ &= \mu_{\mathrm{s}}\mathrm{d}V\left[\int_{4\pi} L(\vec{r},\hat{s}',t)\,P(\hat{s}',\hat{s})\sigma_{\mathrm{s}}\,\mathrm{d}\Omega'\right]\mathrm{d}\Omega. \end{aligned} \tag{2.10}$$

In equation 2.10, $P(\hat{s}',\hat{s})$ is the scattering phase function which is discussed in more detail later. Additionally, μ_{s} is the scattering coefficient and is equal to the number density of scatterers N_{s} multiplied by the cross section of the scatterers σ_{s}.

The final term from equation 2.6 that needs to be defined is the energy gain term due to an emission source in the volume element within the solid angle $\mathrm{d}\Omega$:

$$\mathrm{d}P_{\mathrm{src}} = S(\vec{r},\hat{s},t)\,\mathrm{d}V\,\mathrm{d}\Omega. \tag{2.11}$$

The term $S(\vec{r},\hat{s},t)$ in equation 2.11 is the description of the emission source contributing to the gain.

Finally, substituting all of the expressions 2.8, 2.9, 2.10 and 2.11 into equation 2.6 gives us the RTE:

$$\begin{aligned}\frac{1}{c}\frac{\partial L(\vec{r},\hat{s},t)}{\partial t}\,\mathrm{d}V\,\mathrm{d}\Omega = &-\frac{\partial L(\vec{r},\hat{s},t)}{\partial s}\,\mathrm{d}V\,\mathrm{d}\Omega - \mu_{\mathrm{t}}\,L(\vec{r},\hat{s},t)\,\mathrm{d}V\,\mathrm{d}\Omega \\ &+ \mu_{\mathrm{s}}\mathrm{d}V\left[\int_{4\pi} L(\vec{r},\hat{s}',t)\,P(\hat{s}',\hat{s})\sigma_{\mathrm{s}}\,\mathrm{d}\Omega'\right]\mathrm{d}\Omega + S(\vec{r},\hat{s},t)\,\mathrm{d}V\,\mathrm{d}\Omega. \end{aligned} \tag{2.12}$$

The RTE in its final form, equation 2.12, is not a pleasant sight, and is even uglier to solve especially for arbitrary 3D geometries. Approximations and numerical methods are the weapons of choice for such a beast! The two most common approximations are to assume either scattering or absorption are dominating the light-matter interactions, which will be covered briefly here.

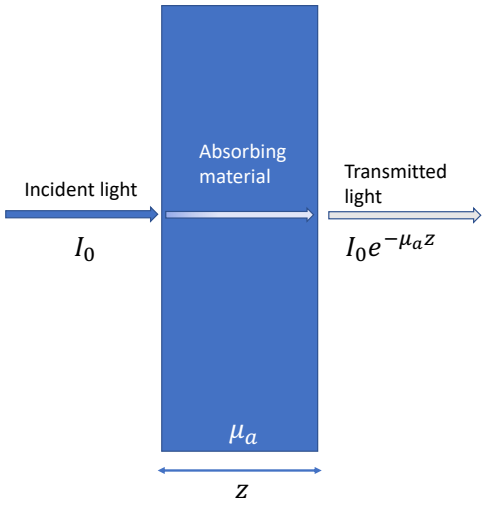


Figure 2.6: Schematic of the light attenuation in a highly absorbing medium. The transmitted light is dependent on both the absorption coefficient and thickness of the medium, and decays exponentially. Diagram adapted from Yoon *et al.*[20]

2.2.1.1 Beer-Lambert law

Taking a medium where the absorption is significantly greater than the scattering, the intensity of a beam of light is found to decrease exponentially with increasing distance through the medium. This can be seen in figure 2.6. This relationship is called the Beer-Lambert law and can be expressed mathematically as:

$$I(z) = I_0 e^{-\mu_a z}. \tag{2.13}$$

Key terms in equation 2.13 are the irradiance ($I(z)$), the incident irradiance (I_0) and the absorption coefficient of the medium (μ_a). Importantly, the absorption coefficient μ_a is dependent on wavelength, and thus the irradiance also depends on the wavelength of the light used. The other factor in how much light is transmitted is the thickness of the slab, z, which when multiplied by the absorption coefficient gives a term called the optical depth, the exponent in equation 2.13. The Beer-Lambert law is applicable for highly absorbing media, where scattering can be assumed to be negligible. This makes it unsuitable for biological systems, as tissue is highly scattering.

2.2.1.2 Diffusion approximation

Looking now to the other simplification: a medium where scattering is the dominant mechanism in light-matter interactions, the radiance can be expressed as spherical harmonics. This is the diffusion approximation, which is valid for isotropic media and requires uniformity of tissue type. This can make it unsuitable for biomedical applications as tissue is highly forward scattering, which breaks the isotropic scattering requirement.[146] However, using similarity relations can improve the accuracy of diffusion approximation in tissue.[147] Additionally, tissue types tend to have overlaps (e.g. a tumour buried in tissue), which makes the diffusion approximation difficult to implement due to the lack of uniformity. It can be used to speed up MCRT in media where scattering dominates as the diffusion approximation is not computationally expensive.[148]

2.2.2 Monte Carlo radiative transfer method

The approximations outlined above are suitable for specific scenarios, unfortunately modelling light in many biological systems is more complicated than the simplifications allow for. Instead of thinking of the beam radiance as a whole, considering the paths of individual photons that make it up can lead to the numerically equivalent result. Applying Monte Carlo methods to radiative transport lets the lifetime of photons in the beam be modelled, where their interactions and motions are chosen by random sampling of probability distributions. These probability distributions are physically motivated by the properties of the environment the photon is travelling through.

MCRT is capable of including arbitrary 3D geometries, regimes with comparable rates of scattering and absorption and various microphysics, albeit at a computationally expensive cost. The following sections will cover the terms required for an MCRT code, the basics of how one functions, and the specifics of the codebase used in this thesis.

2.2.2.1 The journey of a photon

In an MCRT simulation, each "photon" is an energy packet, representing many real photons. There are on the order of 10^{17} photons emitted per second from a 1W light source.

The packet is related to the real number, N_{real}, by:

$$N_{\text{real}} = \frac{E_i}{h\nu_i} \tag{2.14}$$

where E_i is the energy per MCRT packet, h is Planck's constant and ν_i is the frequency of the i^{th} MCRT packet. The energy per packet depends on the total amount of energy from the illuminating source in a time Δt and the total number of packets simulated, N:

$$\frac{E_i}{\Delta t} = \frac{L}{N}. \tag{2.15}$$

In equation 2.15, L is the power of the illuminating light source, equal to the irradiance (in W cm^{-2}) multiplied by the area (in cm^2) of the light source. Each packet has a specific wavelength in order to determine the correct optical properties of the medium, as these vary with wavelength. The initial packet wavelength is found by random sampling of the emission spectrum of the light source. The total energy of each packet is the same, so packets with different wavelengths correspond to different numbers of "real" photons. The initial location of each packet is a random position on the surface of the emission source, represented by a triangle mesh in our code. The initial direction is normal to the emission surface at the emission location. The photon then travels through the medium until it leaves the domain or an interaction occurs. This interaction is either a step within a voxel or reaching the edge of a voxel. At the edge of a voxel, the packet is moved a tiny step over into the next voxel; this is called the bump distance. At the end of a step, the packet is either scattered or absorbed, and the whole process repeats. A simplified flowchart for the progression of a packet, and thus a Monte Carlo simulation can be seen in figure 2.7.

2.2.2.2 Light-matter interactions

As mentioned previously, the propagation of photons through a medium is determined by the optical properties of the medium. The term that covers all interactions that can occur is μ, the interaction coefficient with units of inverse length.

The probability of a packet being scattered or absorbed over unit length ds is equal to

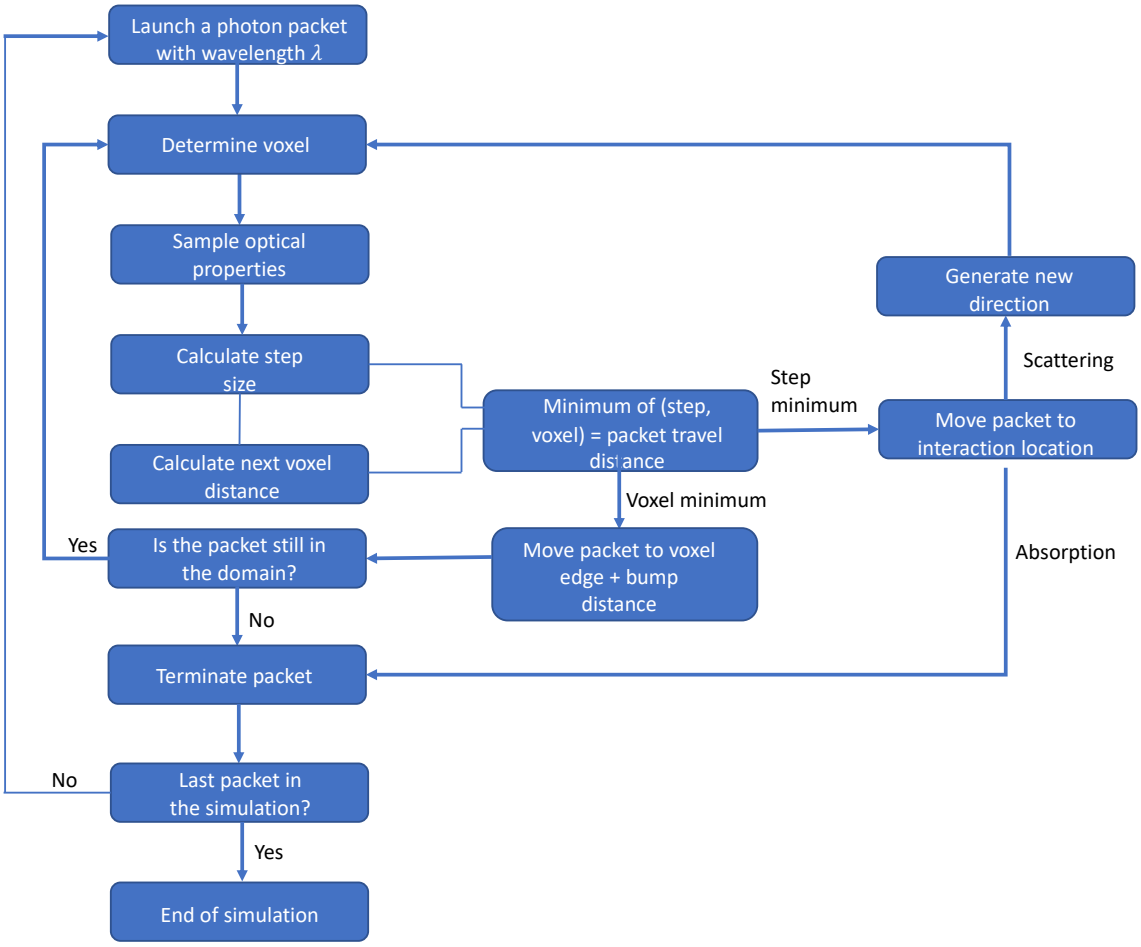


Figure 2.7: Flowchart describing the lifetime of a photon packet in a Monte Carlo simulation.

the interaction coefficient multiplied by the length. Therefore, the probability of a packet travelling a distance $\mathrm{d}s$ without an interaction is $(1 - \mu\,\mathrm{d}s)$. If the whole length S is split into A segments, such that $\mathrm{d}s = \frac{S}{A}$, then the probability of a packet travelling the entire length without interacting can be written as:

$$P(S) = \lim_{A\to\infty}\left(1 - \mu\frac{S}{A}\right)^{A} = e^{-\mu S} = e^{-\tau}$$
$$\implies P(\tau) = e^{-\tau}. \tag{2.16}$$

Here, τ is the optical depth, defined to be the number of mean free paths in the length. The mean free path is the average distance a photon travels between interactions and can be found by the inverse of the interaction coefficient: $\frac{1}{\mu}$. Equation 2.16 is saying that as the length is split into infinitely tiny pieces, i.e. as $\mathrm{d}s$ becomes infinitely small, the probability that a packet will traverse the length without interacting tends to a decaying

exponential that depends on the interaction coefficient and the length itself. This is similar to the result seen in the Beer-Lambert approximation (equation 2.13).

Using what was discussed earlier in section 2.1.1 regarding statistical sampling of a probability distribution function, it is possible to select optical depths in a way that is proportional to the probabilities a packet would go those depths. Choosing a uniform random number ξ in the range [0,1), an optical depth can be selected by:

$$\begin{aligned} \xi &= \int_0^{\tau} e^{-\tau'} d\tau' = 1 - e^{-\tau} \\ \implies \tau &= -\ln(1-\xi). \end{aligned} \tag{2.17}$$

Since ξ is a random number in the range [0,1) it follows that $(1-\xi)$ is an equally random number in the same range, so it is possible to rewrite the result from equation 2.17 slightly more neatly:

$$\tau = -\ln(\xi). \tag{2.18}$$

From the derivation in equation 2.16, it is known that $\tau = \mu S$, and so it is possible to convert the result from equation 2.18 into a physical length step size, s:

$$s = \frac{-\ln(\xi)}{\mu} \tag{2.19}$$

where ξ is a uniform random number in [0,1) and μ is the interaction coefficient, as discussed earlier.

Each interaction (scattering or absorption) has its own coefficient for each medium, and so the overall interaction coefficient is defined as:

$$\mu = \mu_s + \mu_a \tag{2.20}$$

where μ_s is the scattering coefficient, μ_a is the absorption coefficient and both of these have units of inverse length, usually cm^{-1} for biological applications. The decision

of which interaction happens after each step is made by random number generation. The albedo is defined as $a = \frac{\mu_s}{\mu}$ and so a uniform random number ξ in the range [0,1) is generated. If ξ is less than or equal to the albedo, a scattering event occurs and the packet takes another step through the domain. Otherwise, an absorption event happens.

2.2.2.2.1 Scattering event

The scattering coefficient μ_s is a measure of how far a photon (or packet) will travel before a scattering event occurs. When a photon scatters, it will change direction, by some angle. The azimuthal scattering angle ($0 \leq \phi \leq 2\pi$) is assumed to be isotropic and is expressed by:

$$\phi = 2\pi\xi. \tag{2.21}$$

The phase function describes the probability of a photon being deflected by the polar angle ($0 \leq \theta \leq \pi$). The Henyey-Greenstein phase function was introduced to describe the scattering of diffuse light in our galaxy[149] but has since been used for modelling propagation in biological tissue.[150] The Henyey-Greenstein phase function is defined as:

$$P(\theta) = \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 - 2g\cos(\theta))^{\frac{3}{2}}} \tag{2.22}$$

where the parameter g is the anisotropy factor. This is a dimensionless coefficient between -1 and 1 which describes the intensity of forward or backward scattering in the material. A g of -1 corresponds to fully back-scattering, 1 is for fully forward-scattering and $g = 0$ is for isotropic scattering, as shown in figure 2.8.

By normalising, integrating, and rearranging equation 2.22, a function for sampling to derive the scattering distribution is obtained, for $g \neq 0$:

$$\cos(\theta) = \frac{1}{2g}\left[1 + g^2 - \left(\frac{1 - g^2}{1 + g(2\xi - 1)}\right)\right]. \tag{2.23}$$

If $g = 0$ (isotropic scattering) then equation 2.23 becomes $\cos(\theta) = 2\xi - 1$.

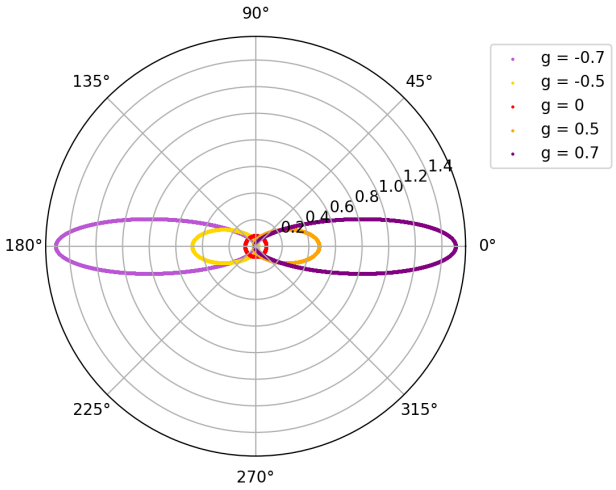


Figure 2.8: Diagram showing the impact of the anisotropy factor, g, on the Henyey-Greenstein phase function. The initial direction of travel is across the centre of the circle towards 0°.

2.2.2.2.2 Absorption event

The absorption coefficient μ_a describes how likely a photon is to be absorbed in an interaction. Being absorbed in the simulation traditionally means the end of the journey for that photon packet.

2.2.2.3 Interface phenomena

Thus far, the description of the code components and calculations could be applied to a generic light transport problem, over large scales and/or homogeneous materials. In biological systems, optical properties can change over very short length scales, which requires handling of interface phenomena: reflection and refraction.

The three important angles to consider are the angles of incidence, reflection and refraction: θ_i, θ_r and θ_t, respectively. Fresnel coefficients (denoted by R are required to be able to model these interface phenomena accurately. For perpendicular (subscript s) and parallel (subscript p) polarised light, and the unpolarised light, the coefficients can be calculated from:

$$R_s = \left| \frac{n_0 \cos\theta_i - n_1 \cos\theta_t}{n_0 \cos\theta_i + n_1 \cos\theta_t} \right|^2, \tag{2.24}$$

$$R_p = \left| \frac{n_0 \cos\theta_t - n_1 \cos\theta_i}{n_0 \cos\theta_t + n_1 \cos\theta_i} \right|^2, \tag{2.25}$$

and

$$R_{\text{eff}} = \frac{1}{2}(R_{\text{s}} + R_{\text{p}}). \tag{2.26}$$

In the above equations, the terms have the following meanings:

- θ_{i} and θ_{t} are the angles of incidence and transmission, respectively;
- n_0 and n_1 are the refractive indices of the current medium and the next medium;
- R_{s} and R_{p} are the reflection coefficients for the s and p polarised lights;
- R_{eff} is the effective reflection coefficient for unpolarised light, defined as the average of the two polarised light components.

The angles of incidence and transmission can be calculated from

$$\sin\theta_{\text{t}} = \frac{n_0}{n_1}\sin\theta_{\text{i}}, \tag{2.27}$$

$$\cos\theta_{\text{t}} = \sqrt{1 - \sin^2\theta_{\text{t}}}\,. \tag{2.28}$$

Using equation 2.27 (Snell's Law), and figure 2.9, the reflection vector can be calculated. Both the incident and reflected rays are in the same medium with the same refractive index n_0, so equation 2.27 simplifies to $\theta_{\text{t}} = \theta_{\text{i}}$. Using the geometry in figure 2.9, the following expressions can be found:

$$\begin{aligned} \vec{I} &= \vec{A} + \vec{B} \\ \vec{R} &= \vec{A} - \vec{B} \\ \vec{B} &= \cos\theta_{\text{i}} \cdot \vec{N}. \end{aligned} \tag{2.29}$$

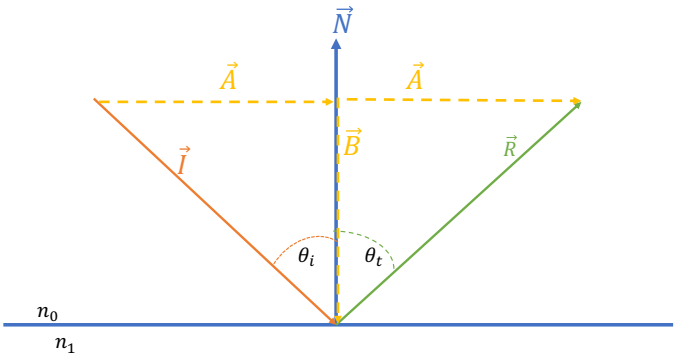


Figure 2.9: Schematic diagram showing the geometry of light reflecting at a boundary between two media with different refractive indices. Here, the direction vectors are indicated by $\vec{I}$ for incident light, $\vec{R}$ for reflected and $\vec{N}$ for a normal to the surface. $\vec{A}$ is a coplanar vector with $\vec{I}$ and $\vec{R}$ parallel to the interface. $\vec{B}$ is another vector coplanar with $\vec{I}$ and $\vec{T}$ which is in the opposite direction to $\vec{N}$ at the point of the reflection on the interface.

By substitution and rearranging, the reflection vector is derived:

$$\begin{aligned} \vec{I} &= \vec{A} + \cos\theta_{\mathrm{i}} \cdot \vec{N} \\ \vec{R} &= \vec{A} - \cos\theta_{\mathrm{i}} \cdot \vec{N} \\ \therefore \vec{R} &= \vec{I} - 2(\vec{N} \cdot \vec{I})\vec{N} \end{aligned} \tag{2.30}$$

Now let us consider the transmission vector, and how the refraction of the light affects it. The geometry of the problem can be seen in figure 2.10, contained within a unit circle.

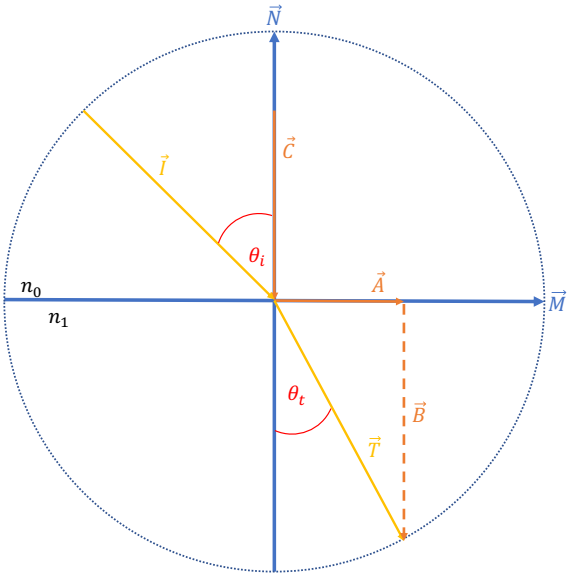


Figure 2.10: Diagram showing the geometry of light refracting at a boundary between two media with different refractive indices. Here, the direction vectors are indicated by $\vec{I}$ for incident light and $\vec{T}$ for transmitted. The circle is a unit circle. Vector $\vec{M}$ is a unit vector coplanar with $\vec{I}$ and $\vec{T}$ along the interface; $\vec{N}$ is a unit vector also coplanar with $\vec{I}$ and $\vec{T}$ perpendicular to the interface. Vector $\vec{A}$ is along $\vec{M}$, vector $\vec{B}$ is perpendicular to $\vec{A}$, and the sum of $\vec{A} + \vec{B} = \vec{T}$. $\vec{C}$ is opposite in direction to, but runs along $\vec{N}$ to the point of refraction on the interface.

From figure 2.10, the following relations can be found:

$$\begin{aligned}
\vec{T} &= \vec{A} + \vec{B} \\
\vec{A} &= \sin\theta_t \vec{M} \\
\vec{B} &= \cos\theta_t(-\vec{N}) \\
\vec{C} &= \cos\theta_i \vec{N} \\
\vec{M} &= \frac{\vec{I} + \vec{C}}{\sin\theta_i}
\end{aligned} \tag{2.31}$$

The expressions for $\vec{A}$ and $\vec{B}$ come from the geometry seen in figure 2.10, and $\vec{N}$ is the normal to the surface. To get the expression for vector $\vec{M}$, it is noted that $\vec{I} + \vec{C}$ gives a vector perpendicular to $\vec{N}$ and tangent to the surface, and then it gets normalised by $\sin\theta_i$.

Substituting the expressions for $\vec{A}$ and $\vec{B}$ into the expression for $\vec{T}$, and using Snell's Law from earlier, equation 2.27, an expression for the transmission vector in terms of the

refractive indices, the incident ray and the normal $\vec{N}$ is obtained:

$$\begin{aligned}
\vec{T} &= \vec{A} + \vec{B} \\
&= \vec{M}\sin\theta_t - \vec{N}\cos\theta_t \\
&= \frac{\vec{I} + \vec{C}}{\sin\theta_i}\sin\theta_t - \vec{N}\cos\theta_t \\
&= \frac{\vec{I} + \cos\theta_i \vec{N}}{\sin\theta_i}\sin\theta_t - \vec{N}\cos\theta_t \\
\therefore \vec{T} &= \frac{n_0}{n_1}(\vec{I} + \cos\theta_i \vec{N}) - \vec{N}\cos\theta_t \\
\vec{T} &= n + (n\, c_1 - c_2)\vec{N}.
\end{aligned} \tag{2.32}$$

It is possible to get to the final line in deriving equation 2.32 by defining some new expressions:

$$\begin{aligned}
c_1 &= \vec{N}\cdot\vec{I} \\
c_2 &= \sqrt{1 - n^2(1 - c_1^2)} \\
n &= \frac{n_0}{n_1}.
\end{aligned} \tag{2.33}$$

As an individual packet propagates a medium, upon colliding with a surface a uniform random number in [0,1) is generated. If it is less than the Fresnel coefficient for reflection, defined in equation 2.26 then the packet is reflected, otherwise it is transmitted along the transmission vector with appropriate refraction.

2.2.2.4 Variance reduction

To improve the signal-to-noise ratio of MCRT simulations, there are several techniques that can be employed to maximise the impact of the photon packets.

Weighting

Upon emission, each photon packet is given a weight w of unity. Instead of giving every packet a probability of being fully absorbed at each interaction location, all of the packets are kept but reduce their weight accordingly. This allows for deeper sampling of a medium by not absorbing as many entire packets, but retaining the same accuracy in depicting the

energy loss to the medium. Recall that the albedo is defined as:

$$a = \frac{\mu_a}{\mu}. \tag{2.34}$$

The albedo describes the fraction of light absorbed in a single scattering event and is expressed by the ratio of the absorption coefficient to the interaction coefficient. Each interaction event becomes a scattering event, and the weight of the packet is reduced by:

$$w' = w\,(1 - a). \tag{2.35}$$

Russian Roulette

Following this weight reduction technique to keep packets in the simulation for longer, it cna be seen that after n scattering events, the weight will be reduced to A^n. Since tissue is so scattering, even a high albedo will quickly reduce the statistical weight of a packet to a small fraction. This can lead to the issue of devoting lots of computing power for less and less statistically valuable packets. Rather than spend energy and time on this, a quick game of Russian roulette can put things right.

An arbitrary weight threshold is defined, and if a packet falls below this weight then a uniform random number, ξ, in [0,1) is generated. If ξ is below a fraction ($\frac{1}{p}, p > 0$) then the packet lives and has its weight multiplied by p. Conversely, the packet is killed if ξ is greater than the fraction. Multiplying the weight of the surviving packet by p ensures energy is conserved.

Peel-off

Another method of improving signal to noise ratio for photon packets is to implement a peel-off method. A graphic depicting this idea can be seen in figure 2.11, where there is a spherical domain with a central light source and a detector represented by the green oval. The orange arrows show an example path that a packet might take from the centre to leaving the domain. The green arrows represent a small amount of weight calculated at each scattering event, proportional to the probability that the packet would have been scattered towards the detector. By doing this, every packet contributes to the detected

image and the signal to noise ratio can be greatly improved.

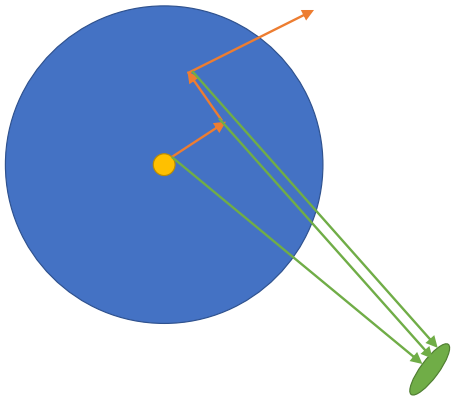


Figure 2.11: Graphic to depict the peel-off method. The yellow central circle is the light source and the green oval is the detector. The orange arrows show an example path for a photon packet out of the medium and the green arrows represent a small weight at each scattering event being peeled off towards the detector.

The probability of a packet being scattered to a point in space can be found by multiplying the probability that the packet is scattered in the given direction by the exponential decay of the optical depth between the scattering location and the point in space:

$$W_{\text{peeloff}} = \frac{1}{2} \frac{1 - g^2}{(1 + g^2 - 2g\cos\theta)^{\frac{3}{2}}} e^{-\tau} \tag{2.36}$$

This method assumes that the packet travels in the same medium from the scattering point to the detector, but becomes invalid when crossing boundaries due to the refraction changing the angles.

2.2.2.5 Measuring spatial variables

There are many variables in a Monte Carlo simulation that could be recorded: the path of every packet, the average number of scattering events before escape, absorbed energy density. However, it can quickly become extremely computationally expensive to calculate all of these values, and impractical to store all of the data. For an individual problem, only the crucial and relevant information is tracked and kept.

There are several variables that vary spatially, and in order to measure these, the domain is split into a uniform grid on the Cartesian axes. This method implements a 3D grid of voxels, and each voxel can contain information about spatial variables, for example the emission density and the absorbed energy density. A photon packet will always be contained in one voxel at any time in the simulation. If the packet reaches the boundary of the domain and the edge of the voxel grid then it is ordinarily removed from the simulation.

As a packet travels, it contributes energy ϵ_0 to the voxel containing it at a given instant. By defining a variable ℓ to be the path length between events (including crossing into a new voxel as an event), the trajectory of the packet contributes $\epsilon_0 \delta t / \Delta t$ to the energy of a voxel, with $\delta t = \ell / c$ and Δt being the duration of the simulation. The average energy density in a given voxel volume V is then:

$$U = \frac{1}{V} \frac{\epsilon_0}{c\,\Delta t} \sum \ell n_v. \tag{2.37}$$

The contained energy density is a measure of the flux (J) through the surface: $U_c = 4\pi J \frac{n}{c}$.

The absorption rate in a given voxel is then described by[151] :

$$\dot{A} = 4\pi\, J\, \mu_a \quad = \frac{1}{V} \frac{\epsilon_0}{\Delta t} \sum \ell \mu_{\mathrm{a}}. \tag{2.38}$$

2.3 Our code

The previous section explored a standard Monte Carlo code, and now the code used for the work in this thesis is presented, along with how it is an improvement on previous works. Implementing Raman scattering into the code is then briefly discussed.

As mentioned in the beginning of this chapter, the `arctk` toolkit used for the simulations in this thesis was started by Dr Freddy Wordingham, with the design inspired from previous works in the astrophysics code TORUS. The `arctk` code is written in Rust,

a modern language (at the time of writing!).

This code was chosen primarily for the way it was designed and developed. As detailed in the following sections, `arctk` allows for complete flexibility when setting up simulations. Objects of any shape can be created from triangle-meshes and imported. Since it is written in Rust, it runs in comparable times to C-like languages, and it is simple to create a binary to then be run on any machine. This avoids the difficulties often experienced with Make files in other languages. The next section discusses the specific and relevant details of the `arctk` code.

2.3.1 `arctk`

The `arctk` library has several data structures alongside binary applications that utilise them to produce a result for a given input simulation. This thesis does not use all of the binaries and possibilities afforded by `arctk` and so only the relevant sections will be highlighted here.

To run an instance of `arctk`, the basics required are the binary path (which binary application is required), an input directory (where the input information can be found), an output directory (where the output will be dumped) and a path to a parameters file (where the details of the simulation are stored). The parameters file is in a standard JSON5 format and either contains or points towards the details of the simulation such as the materials, the objects, the attributes of the objects, and the illumination source. An example can be seen in Listing 2.1. The majority of the work in this thesis uses the MCRT binary application.

The surfaces are set up by naming and placing an object in the domain. Therefore, the optical properties of any point inside the domain are determined by the surfaces, not by a Cartesian grid so they do not vary on a voxel-by-voxel basis. This creates a lot of flexibility in the simulation possibilities but at the expense of performing hit-scan operations on triangle meshes. To reduce the computational cost, a hit-scan tree is created wherein the resolution increases with the number of triangles in the region. Each triangle has information about the surface it represents stored, so for an MCRT simulation it is known how a packet should interact with it. The tree generation is dictated in the parameters file.

The attributes of the objects define how a photon will interact with a surface. An object can therefore represent biological entities, where a packet will be reflected or refracted as appropriate. An object can also be a detector, where the packet will be killed off (removed from the simulation) and the requisite information stored (for example location, wavelength, weight). In this way, it is possible to have embedded objects of any geometry as long as it is a closed surface and the material it is made of is defined, or a detector of any shape is also possible. The light source is also defined in a separate section, but can also be of any shape and be given an emission spectrum (monochromatic or otherwise). The light source file also needs to specify the material the photons will be emitted into.

The following is an example parameters JSON5 for shining a laser into a cube of pork tissue material. More surfaces of different materials can be added by simply listing them in a similar fashion to how the pork tissue cube has been set up. Surfaces with special properties such as a CCD detector or a spectrometer can be inserted: all of the information is easily found in the code documentation.

Listing 2.1: Example json5 parameters file for shining a laser into a volume of pork material.

```
1   grid:   {There  :   'grid.json5'},      \\grid file
2
3   tree:   {Here:  {
4           tar_tris: 5,
5           max_depth: 8,
6           padding: 0.01
7       }
8   },                                      \\tree instructions
9
10  sett: {
11      Here: {
12          block_size          : 1000,     \\block size
13          num_phot            : 1e8,      \\10 million photon
                packets
14          bump_dist           : 1e-12,    \\bump distance (m)
```

```
15          loop_limit          : 1e6,      \\loop limit
16          min_weight          : 1e-3,     \\minimum simulation
                weight before
17                                          triggering Russian
                                               roulette
18          roulette_barrels    : 8,        \\roulette barrels
19          init_mat            : 'pork',   \\initial material
               packets are emitted into
20      }
21  },
22
23  surfs: {
24      Here: {
25          small_cube: {
26              attr: "cube",
27              mesh: [
28                  ["objs/raman/small_cube.obj"], {
29                      scale: 0.081, rot:[0.01,0.02,0.03]
30                  }
31              ]
32          }
33      }               \\surface in simulation is a cube of
           pork material
34  },
35
36  attrs: {
37      Here: {
38
39          cube: {Interface: ["pork", "air"]},
40
41      }               \\material interface with pork inside
```

```
        and air outside
42  },
43
44  mats: {
45      Here: {
46              air:    { There: 'mats/misc/air.json5'   },
                \\air material
47              pork:   { There: 'mats/raman/pork.json5'}
                \\pork material
48          }
49  },
50
51  light: {
52      Here: {
53          power       : 1.0,
54          emit        : { Surface:
55                          [['objs/raman/wide_laser.obj'],
56                          {scale: 1.0e-2, trans:[0.0,0.0,
                              -22.4999e-3]}] },
57          spec        : { Here: {Point: 8.08e-7}},
58          mat         : 'pork'
59      }                                   \\1 cm diameter
        laser, photon packets initialised with a wavelength
        of 808 nm
60  },
61
62  engine      : "Standard"    \\instruction to run a standard
      MC simulation
```

2.3.2 Code improvements

As discussed earlier, simulation of biological systems requires specific work on boundaries between materials and small scale light propagation. This code handles these situations by implementing triangle-meshes and packet tagging in the "classic" Monte Carlo loop.

Modern modelling software, such as Blender, allows models of surfaces and objects to be created and then stored as triangular mesh .obj files. When determining the distance to an interaction point, if the packet will cross from one material into another, the code determines the optical properties by way of a table of stored values. The triangular mesh interface is tagged with an inside and an outside material, with corresponding optical properties. If the photon packet is reflected, the material tag remains the same, and if the packet is refracted then it goes into the new material and the optical properties are updated accordingly.

Another important improvement over the standard Monte Carlo loop is the use of parallelisation. Monte Carlo simulations can be described as "embarrassingly parallelisable".[152] This is due to the fact that each photon packet is independent: the life and path of each packet have no impact on any of the other packets. This means that it is easy to divide up a total number of packets N between a number of threads T and sum up the results at the end for whichever variables the specific simulation is tracking. Parallelising the code allows simulations to be run much faster than before, despite the added complexities of the arbitrary geometries `arctk` is capable of handling.

2.3.3 Implementing Raman scattering

There are two main methods for including Raman scattering into this Monte Carlo loop: a one-step process or a two-step process. The one-step process means that a Monte Carlo simulation is run once, and at each interaction, the definition of the scattering coefficient is altered to include a probability of a Raman scattering event:

$$\mu_\mathrm{s} = \mu_\mathrm{e} + \mu_\mathrm{R} \tag{2.39}$$

where μ_e represents the elastic scattering coefficient and μ_R is the Raman scattering coefficient. Therefore, each photon packet continues as before, but at each scattering event there is a probability that its wavelength could be shifted to a new wavelength. It is also necessary to have the optical properties not only at the illumination wavelength, but also at the new shifted wavelength.

The two-step process utilises another binary (named "babbage") of the `arctk` toolkit that Wordingham developed. For the two-step process, the MCRT tool is run for photons from the illumination source (at the incident wavelength only) through the domain, with all of the correct objects and materials, and store the energy density throughout the Cartesian grid described earlier. Then, run a second MCRT routine but this time the illumination points are the Raman scatterers, the wavelength for the packets is the shifted wavelength and the power is derived from the laser energy density at the location of each of these scatterers in the grid. These Raman scatterers now behave as emitters of photons at the new wavelength. This uses the babbage tool in `arctk` which does post-processing on datacubes. The post-processing required here is to take the locations of the Raman emitters, place them in the Cartesian grid and return the energy density at each location.

These processes have both been used in the work in this thesis and will be explained more fully with their specifications at each application in the chapters that follow. The one-step process can be quicker, requires less intense set-up and is suitable for set-ups with a singular large Raman source, but can have worse signal to noise. The two-step process is good for separated small Raman sources (such as microcalcifications) but does require more elaborate set-up. A bash script and assorted Python scripts were developed to automate the two-step process.

2.4 Summary

Monte Carlo radiative transfer is a popular technique for simulating light propagation through turbid media. This chapter has explored the history of the technique, the development of MCRT simulations and the code used for the work in this thesis. The aim of this thesis is to use the MCRT technique to investigate the Raman signals from microcalcifications in breast tissue, in order to further the use of Raman spectroscopy in breast

cancer detection. The advantages of using a code are the cost-effectiveness, being able to model many more scenarios than physically possible and the wealth of information to be extracted from a simulation. The design of our code means that arbitrary geometries can be used, in 3D domains. This is an improvement on previous codes which have been entirely based on grids, usually 1D or 2D layer systems. The flexibility in light sources, detectors, and materials between the two allow for application to other problems in future.

Chapter 3 will look at validating the code against another theoretical result in the field of Monte Carlo modelling of Raman spectroscopy in a simple 2-layer system with SORS geometry for illumination and detection. Then, chapter 4 will describe the capabilities of the code in replicating experimental work, and combining results from an experiment with MC results to constrain optical properties of tissue phantoms. Finally chapter 5 will present a range of simulations be run with microcalcifications in both homogeneous and heterogeneous breast tissue domains. These simulations show the effects of different beam widths, detector geometries and including skin in the breast model.

Chapter 3

Code Benchmarking

As described in the previous chapter, Monte Carlo simulations are stochastic by nature, relying on random number generation at several points. It is prudent, as with any computational physics code, to check that the results returned by the simulations are correct. The thesis of F. Wordingham has verification of the base Monte Carlo binary, returning expected behaviour in: a refracting prism simulation, a focusing lens, a fibre optic, an absorbing cube, the Beer-Lambert law and a scattering sphere. These tests make it apparent that the Monte Carlo algorithms and variable tracking techniques are working as expected.

The work in this thesis involved implementing Raman scattering into the Monte Carlo binary, and thus this is the area that needs testing against known results. The verification is to set-up and run the same simulation using the codebase as has been run in a published work; the code needs to be checked against results from another Monte Carlo code that was simulating Raman scattering. The code chosen was written by Pavel Matousek, and is detailed (with results) in Matousek et al. (2005).[21] A summary of the original paper, our implementation of the simulation and the following comparison and findings are explained in this chapter.

3.1 Simulation summary

The selected paper "Numerical Simulations of Subsurface Probing in Diffusely Scattering Media Using Spatially Offset Raman Spectroscopy" originally appeared in Applied Spectroscopy in 2005.[21] It presents the first model predicting how Raman intensities vary with a range of experimental variables in a SORS set-up. This technique was initially proposed in another Matousek paper[104] also from 2005, and the chosen study was written as a useful theoretical counterpart in the beginning of the development of SORS. This made it an ideal candidate for benchmarking our code against: simple plane-parallel geometry, annuli detectors and basic materials.

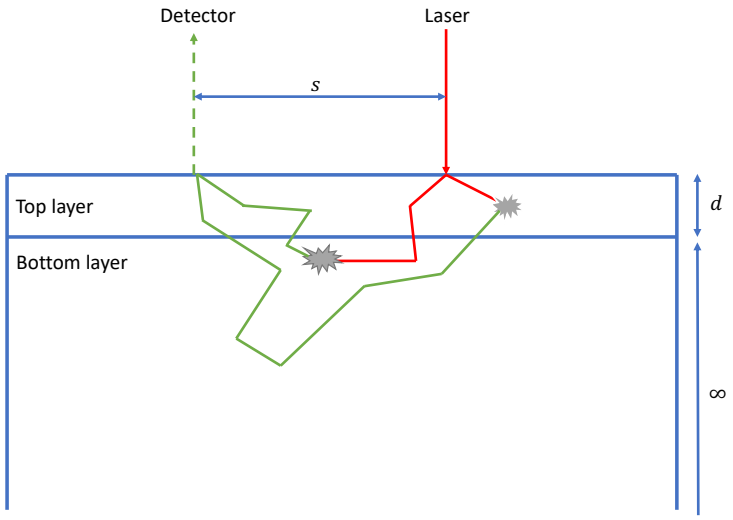


Figure 3.1: Graphic depicting the simulation setup with a source-detector separation s and a top layer with depth d. The colour change represents the laser photon being re-tagged as a Raman photon, and the location where the change occurred (the grey spiked shapes) is recorded. Figure adapted from the Matousek et al.[21] study used for benchmarking.

Spatially offset Raman spectroscopy (SORS) (as discussed more thoroughly in chapter 1) is based on collecting Raman spectra from areas of the sample surface at significant distances from the incident beam. The spectra show variance in the relative intensities from surface and sub-surface layers based on the distance away from the input light source. These can be processed and thus help eliminate surface Raman scattering when probing Raman at depth.

The numerical simulations in this work followed both laser and Raman photons as they propagated through a three-dimensional medium. A simplification was used in assuming that each step a photon packet makes is a straight line over a distance t and at the end of each step is scattered in a fully randomised direction. In reality, a photon would have to undergo several scattering events before its direction is considered random, due to the forward-scattering bias in tissue. The scattering angles in tissue are dependent on the wavelength of the incident light relative to the sizes of the structures that the light interacts with. Scattering in tissue is dominated by Mie scattering, where the structures in the tissue and the wavelength of the light are of comparable size; this results in forward scattering. For large distances, these multiple scattering events and step sizes can be approximated as one event occurring over the step size t. In SORS, with highly scattering media, the total propagation length of a photon is much larger than an individual path length. Therefore, this simplification is justified; in the paper, this is tested and validated.

The model is a semi-infinite slab made of some turbid media, with the air-medium interface at $z = 0$, where z is a Cartesian co-ordinate normal to the interface plane. The sample consists of two layers: the top layer has thickness d and is located from $0 > z > -d$ while the bottom layer is at $z < -d$. The geometry of this domain can be seen in fig 3.1. All of the probe photons are emitted at a depth equal to one transport length into the top layer and distributed around the origin. In reality, the photons are brought onto a sample at near-normal incidence and will travel a transport length before being fully randomised so this is a reasonable approximation. The simulation beam radius r has a uniform intensity across it - a photon has an equal probability of being generated anywhere across it. This is an approximation of real world lasers which often have a Gaussian beam profile.

In a traditional Monte Carlo algorithm, a laser photon packet traverses the medium by moving the packet one step size at a time and then randomly generating a direction. A one-step process can be implemented in order to include Raman scattering. This means that the scattering coefficient of the material is defined by: $\mu_s = \mu_e + \mu_R$, where μ_e is the elastic scattering coefficient and μ_R is the Raman scattering coefficient. In words, at each step there is a finite probability that the photon will be inelastically scattered instead, and thus converted to a Raman photon.

Any laser photons leaving the domain without conversion are lost forever (removed) from the simulation. Any Raman photon leaving the domain is recorded, along with its distance from the origin. Once counted, these Raman photons are also removed from the simulation. In this model, only the Raman photons leaving via the air-sample interface are counted in order to simulate the SORS detection geometry.

In this code, rather than generating a step size using the methods outlined in chapter 2, a fixed step size of 0.2 mm was used as the transport length. The original simulation propagated 1000 photons over an overall path length of 25 cm to reflect typical migration distances in Raman spectroscopy without absorption. This process was repeated 1000 times, meaning an overall number of propagated photons of 10^6. A typical run would take around 3 hours. These experimental parameters were chosen to resemble closely the previous experimental work.

The basic conditions were as follows: the probe beam radius was $r = 0$ mm and the Raman scattering coefficient was 0.005 mm^{-1}. This value was chosen to ensure the majority of the probe photons were converted by the end of the overall propagation distance. This value is higher than the real value (around 1 in 10^6), but that fact only affects the absolute number of Raman photons generated and not the spatial dependencies; the original study tested and discussed this assumption. The absorption optical density was assumed to be the same for the laser and Raman photons, and was set to zero for simplicity. The depth of the top layer was $d = 1$ mm. The thickness of the annulus used for detection was $\Delta s = 0.5$ mm and the inner radius was defined to be the spatial offset s which is plotted in the results graphs. A diagram to demonstrate a collection annulus can be seen in figure 3.2. In this study, the Raman and elastic scattering coefficients are the same in each layer, but not the same as each other. When a Raman scattering event occurs, the layer where this event happens is recorded and then the Raman photon is propagated in the same manner as a probe photon was previously.

This paper runs several simulations to examine different parameters: dependence of signal from layers, effect of changing the thickness of the top layer, altering the transport length, and altering the absorption rate. In order to validate our code the investigation into how the detected Raman signal from the top layer compares to that from the bottom

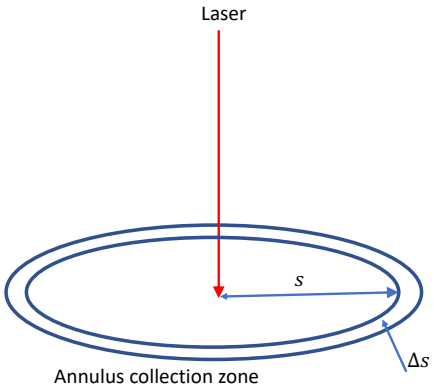


Figure 3.2: Definition of the collection annulus as used in this validation exercise with source detector separation s, and annulus width Δs. Figure adapted from the Matousek et al.[21] study used for benchmarking.

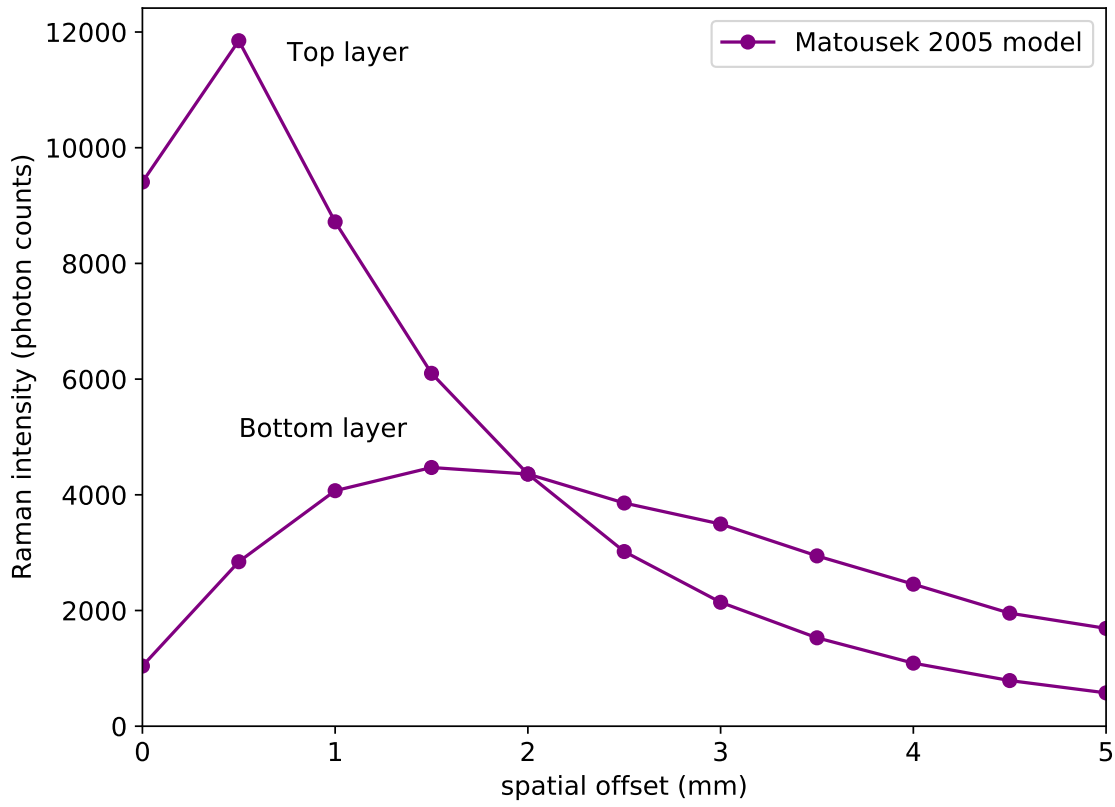


Figure 3.3: Digitised plot showing the number of Raman photons detected from the top and bottom layers, as a function of the radius of the annulus detector. Figure digitised from Matousek et al.[21]

layer as the spatial offset increases was selected. This is the figure digitised and shown in fig 3.3. This is useful and interesting to show the theory behind SORS; above a certain

source-detector separation the signal from the bottom layer is stronger than that from the top layer, thus allowing greater depths to be probed preferentially.

3.2 The importance of definitions

The same setup as described in the Matousek study was implemented in this code. The same source-detector separations were used in order to bin the Raman photons and directly compare the results. The initial output from our code returned a similar shape and trend, but very different absolute numbers were detected, as seen in figure 3.4.

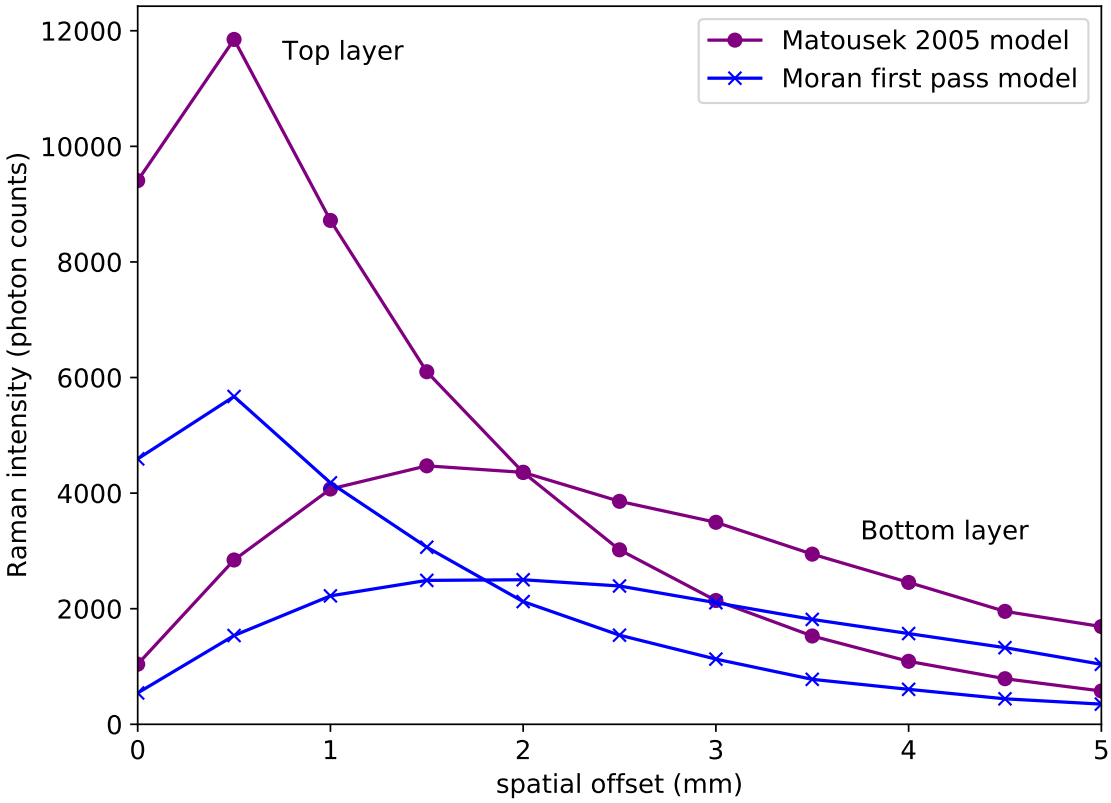


Figure 3.4: Plot showing the Raman photon signal from top and bottom layers, as a function of spatial offset of the detector. The purple lines are the results from the original paper, and the blue lines are the results from our first pass attempt to benchmark with our code.

It is clear from figure 3.4 that the overall behaviour has been replicated by our code, but there appears to be a vast difference in the detected numbers of packets. This implied that many more photon packets in our code were escaping before being detected, or that we were not generating enough Raman photon packets. This discrepancy also leads to a slight difference in the "cross over" point: the annulus radius where the signal from the bottom layer becomes stronger than the signal from the top layer. The difference between

these absolute numbers of photons being detected was ultimately determined to be a result of definitions.

Commonly in physics, **optical depth** is used to describe the amount of transmitted radiant power through a material. Conversely, this paper defined the power loss by the term **optical density**. These similar sounding, but mathematically distinct terms are defined below, and the difference is highlighted.

Optical depth is given by:

$$\tau = \ln\left(\frac{\Phi^i}{\Phi^t}\right). \tag{3.1}$$

Optical density is given by:

$$A = \log_{10}\left(\frac{\Phi^i}{\Phi^t}\right). \tag{3.2}$$

In equation 3.1, τ is the optical depth, Φ^i is the radiant flux received by the material and Φ^t is the radiant flux transmitted by the material. Crucially, this definition uses the **natural logarithm** to determine the ratio. In equation 3.2, A is the absorbance and the other symbols have the same meaning. Note that here, the ratio is defined as the **common logarithm**. Thus, the mathematical difference between these two definitions can be defined as:

$$A_{\text{diff}} = \frac{e}{\ln 10} \approx 2.303. \tag{3.3}$$

These expressions are used in the algorithm at the point of determining a Raman scattering event. When a scattering event is about to occur, a random number $\xi \in [0, 1)$ is generated. If $\xi - P(\text{R}) \geq 0$, then Raman scattering occurs, otherwise it is an elastic scattering. Here, $P(\text{R})$ is the probability of a Raman scattering event. In the optical depth definition, this is $e^{-\mu_{\text{R}} t}$; in the optical density definition this is $10^{-\mu_{\text{R}} t}$.

Our code had been set up and defined by the traditional physical definition of optical depth, whereas the paper code had been set up with the optical density metric. Therefore, every time the photon packet was stepped, and the probability of the probe packet being converted to a Raman packet was calculated, our code was 2.3× less likely to convert them. Thus, the large difference in the absolute number of Raman photons detected between the two codes, without a noticeable change in behaviour. By altering our code to use base 10 this was corrected, as shown in figure 3.5.

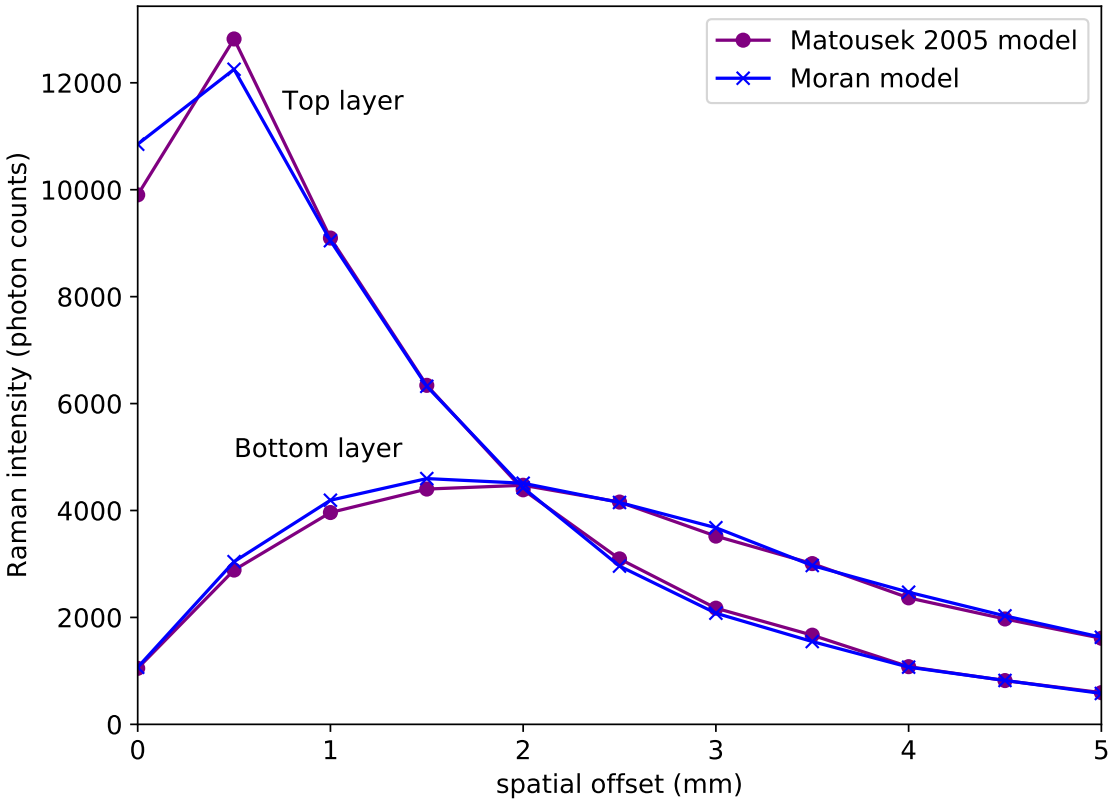


Figure 3.5: Plot showing the Raman photon signal from top and bottom layers, as a function of spatial offset of the detector. The purple lines are the results from the original paper, and the blue lines are the results from our second pass attempt to benchmark with our code, after using the correct definition of optical density.

Although the results in figure 3.5 are much improved, there remains a discrepancy in the top layer signal at the smallest spatial offsets. The results from our code have a smaller difference between these points than the paper data. Given the simulation setup is identical in each code, this difference should not exist. Additionally, there is also a very slight difference between the two bottom layers. The source of these small, but important, differences was in the scattering routine.

3.3 Isotropic scattering

Scattering isotropically requires ensuring that the direction is randomly generated correctly. In the code from the original paper, this direction was generated incorrectly by choosing values that were not linked to the solid angle.

Matousek method:

$x = \mathbf{Rand}[] - 0.5$

$y = \mathbf{Rand}[] - 0.5$

$z = \mathbf{Rand}[] - 0.5$

This generates three uniformly distributed random numbers, each in the range [-0.5, 0.5]. The direction is then found from:

$$\text{dir} = \frac{1}{\sqrt{x^2 + y^2 + z^2}} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{3.4}$$

This method of generating a direction may appear to be isotropic but is actually generating directions based on a cube, which are then mapped to a sphere. This results in a bias to one of the corners of the cube. Thus, the original Mathematica code was not truly isotropically scattering.

True isotropic scattering needs to be performed in polar (or spherical) co-ordinates, as mentioned in chapter 2, and uses the following definitions:

$x = r\sin(\theta)\cos(\phi)$

$y = r\sin(\theta)\sin(\phi)$

$z = r\cos(\theta)$.

In spherical co-ordinates, r is the radius of the sphere, θ is the polar angle and ϕ is the azimuthal angle: $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$.

The instinctive next step is to randomly generate θ and ϕ from a uniform distribution, then apply the above transformations, thus creating Cartesian points. However, to do this will result in a cluster of points around the poles ($\theta = 0$ and $\theta = \pi$) and a dearth of points around the middle ($\theta = \pi/2$).

This is caused by not taking into account the area of the spherical differential surface element:

$$\mathrm{d}A = r^2 \sin(\theta)\mathrm{d}\theta\mathrm{d}\phi. \tag{3.5}$$

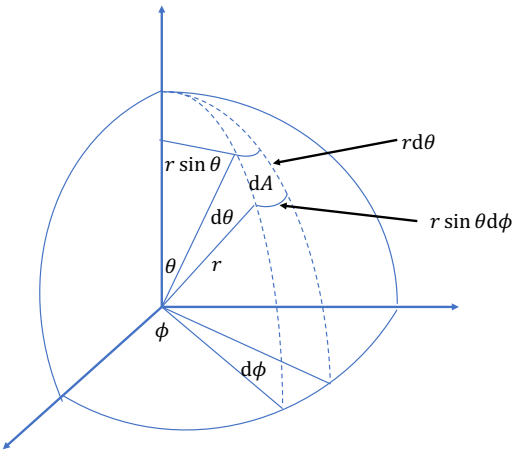


Figure 3.6: Diagram illustrating how the differential surface area element of a sphere is derived.

Equation 3.5 is derived from the trigonometry of the sphere, as can be seen in figure 3.6. Near the poles, the surface element $\mathrm{d}A$ gets closer to 0 since $\sin(\theta)$ tends to 0 as θ approaches 0 or π. We therefore want to find a way to collect samples from a probability distribution $f(\theta, \phi)$ that gives a uniform distribution on the surface of a sphere.

Let p be a point on a unit sphere S. The probability distribution $f(p) = \frac{1}{4\pi}$ since $\iint_S f(p)\,\mathrm{d}A = 1$ and $\iint_S \mathrm{d}A = 4\pi$. Since $f(p)\mathrm{d}A$ represents the probability of finding a

point in an area $\mathrm{d}A$ about the point p, we can get a uniform distribution by defining:

$$f(p)\mathrm{d}A = \frac{1}{4\pi}\mathrm{d}A = f(\theta,\phi)\mathrm{d}\theta\mathrm{d}\phi. \tag{3.6}$$

For a uniform sphere $\mathrm{d}A = \sin(\theta)\mathrm{d}\theta\mathrm{d}\phi$, so we can then normalise $f(\theta)$ over $[0, \pi]$ and $f(\phi)$ over $[0, 2\pi]$. This gives $f(\theta) = \frac{1}{2}\sin(\theta)$ and $f(\phi) = \frac{1}{2\pi}$.

Using the inverse transform sampling, which allows sampling of a general probability distribution using uniform random numbers, we can now generate isotropic directions. We need to create the cumulative distribution function (CDF) for θ and ϕ, where we use ξ to represent a uniform random number.

$$\begin{aligned}\xi_1 = \int_0^\theta f(\theta)\,\mathrm{d}\theta = \frac{1}{2}\int_0^\theta \sin(\theta)\,\mathrm{d}\theta = \frac{1}{2}(1-\cos(\theta))\\ \Rightarrow \theta = \cos^{-1}(2\xi_1 - 1)\end{aligned} \tag{3.7}$$

$$\begin{aligned}\xi_2 = \int_0^\phi f(\phi)\,\mathrm{d}\phi = \frac{1}{2\pi}\int_0^\phi \mathrm{d}\phi = \frac{\phi}{2\pi}\\ \Rightarrow \phi = 2\pi\xi_2\end{aligned} \tag{3.8}$$

Using equations 3.7 and 3.8 to generate correct polar and azimuthal angles, and then inserting these into the metric to get between Cartesian and polar co-ordinates will give correct isotropic scattering.

Correcting the scattering in the source code from Matousek and plotting the results against ours gives an excellent overlap, as seen in figure 3.7. The purple lines are from the original Matousek code, after we changed the scattering to be truly isotropic. The blue lines are the result from our code, using the optical density definition. Fixing these two differences has ensured that the two simulations are truly equivalent, and returned matching behaviours.

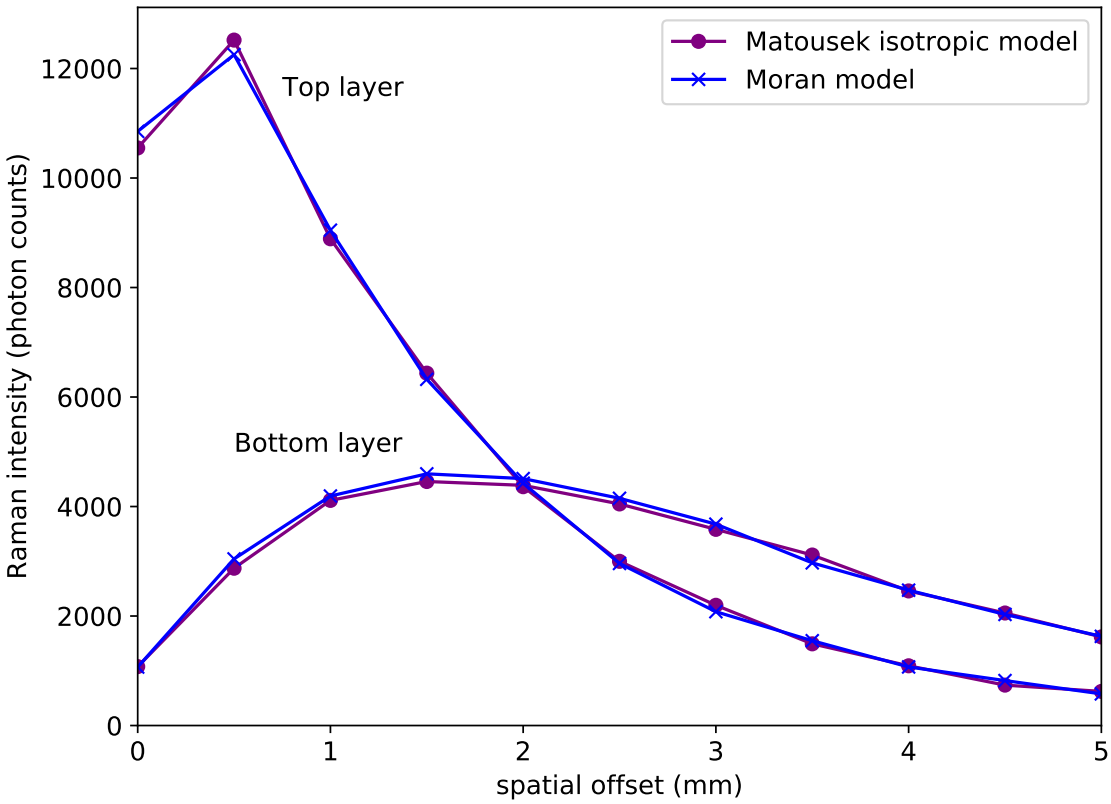


Figure 3.7: Plot showing the Raman photon signal from top and bottom layers, as a function of spatial offset of the detector. The purple lines are the results from the original code, with isotropic scattering implemented, and the blue lines are the final results from our code.

3.4 Conclusions

The work carried out in this chapter was to ensure that our code was validated against known, published Monte Carlo modelling results that included Raman scattering. This is the important first step to ensure that our code is fit for purpose, and returning expected results. By proving that it is as good as previous work, we can be confident that it is behaving as expected.

The paper that our code was benchmarked against in this chapter was published to show the theoretical basis for Spatially offset Raman spectroscopy (SORS). SORS is a spectroscopic technique which is useful for retrieving Raman signals at deeper layers without interference from a top layer, making it vital for use at depth in environments such as breast tissue. The results from the paper show that with a laser-detector separation of > 2 mm, the signal from a lower layer will be greater than that from the top layer, for this domain.

By implementing the same setup and conditions in our code, alongside some definition and maths checking, we have been able to replicate this result. This study was insightful in showing the different terms and definitions used between disciplines which are important to consider, moving forward. It was also interesting to see another, more basic, Monte Carlo code and be able to compare the two. Our success in returning the same results, and thus validating our code, means we can move to the next step with confidence: aiming to prove that our code can return and then predict experimental results.

Chapter 4

Monte Carlo modelling of Raman scattering in tissue phantoms

The work in this chapter is based on a paper published in the Royal Society of Chemistry journal, *Analyst*, published November 2021.[153] The paper is titled "An experimental and numerical modelling investigation of the optical properties of Intralipid using deep Raman spectroscopy". The paper was written by me: I also carried out the experimental work, set up and ran the Monte Carlo simulations, performed the analysis and created the visualisations. Dr Freddy Wordingham developed the initial software, in the form of the Rust library `arctk` as described in chapter 2; I aided in developing the Raman scattering capabilities. Dr Benjamin Gardner set up the experimental work, trained me in the equipment use and was involved in the experimental investigation.

In this chapter, Monte Carlo simulations were performed to investigate the distribution of Raman signals in tissue phantoms and to validate the `arctk` code. The aim was to show our code is capable of replicating experimental results in order to use it to inform similar future studies and to predict the outcomes. The experiment performed to benchmark our code used large volume liquid tissue phantoms to simulate the scattering properties of human tissue. The scattering agent used was Intralipid (IL), of various concentrations, filling a small quartz tank. A thin sample of PTFE was made to act as a distinct layer in the tank; this was the Raman signal source. The IL was a liquid tissue

phantom to represent breast tissue, while the PTFE was used to represent the microcalcification Raman signal. We studied experimentally, and then reproduced via simulations, the variation in Raman signal strength in a transmission geometry as a function of the optical properties of the scattering agent and the location of the Raman material in the volume. We have also found that a direct linear extrapolation of scattering coefficients between concentrations of Intralipid is an incorrect assumption at lower concentrations when determining the optical properties. By combining experimental and simulation results, we have calculated different estimates of these scattering coefficients. The results of this study give insight into light propagation and Raman transport in scattering media and show how the location of maximum Raman signal varies as the optical properties change.

4.1 Introduction

Designing a clinical diagnostic technique relies on a thorough understanding of the propagation of light (both laser and Raman photons) through tissue. There have been studies on both the strength of the Raman signal versus the depth of the tissue, and also the spatial distribution of Raman scattering throughout a sample volume.[22] These studies can tell us about the dependence of Raman signal detection on the optical properties of the sample. This type of work often uses tissue phantoms, where the optical properties have been specifically picked to match those of tissue. They also contain an inclusion with a known Raman spectrum when investigating signal origins and recovery of photons.[154–156] Despite the success of laboratory studies using tissue phantoms, there is a growing interest for *in silico* studies to better understand the behaviour of photon transport throughout the turbid media.

A significant advantage of a validated *in silico* code is the potential to simulate many more experimental setups than could be done in a laboratory. This is cost effective and allows for ineffective experiments to be eliminated before being attempted in a laboratory setting. Simulations can also offer a greater physical insight into the results seen experimentally by modelling the light transport throughout the domain.

In order to use a code as a predictive tool, we first need to be confident that it returns

known "real world" experimental results. Due to the biological nature of the end goal application, we chose to base this experimental work around tissue phantoms made of dilute Intralipid. There have been many experiments using direct measurements to characterise Intralipid at different wavelengths and concentrations.[157–164] Of the relevant studies, that are performed at a similar wavelength to this, the estimates of the optical properties were varied. We used the literature and the results found in this investigation to confirm our tool replicates the experimental response observed.

For clarity, we have split this chapter into two parts: laboratory based experimental work to provide data for comparison with the modelling, then the Monte Carlo method and its application here. First, we discuss the experiment performed to explore the variation in Raman signal strength of a thin layer at different depths. This experiment is performed for a variety of IL concentrations, to also investigate how changing the optical properties of the tissue phantom impacts the Raman signals.

We then move on to explain the details of the Monte Carlo numerical method in this circumstance, especially the handling of the Raman scattering method. We discuss how the experimental results feed into the Monte Carlo model to help constrain the model and then show that the derived optical properties allow the model to return the experimental data and thus the code validation is complete.

4.2 Experimental setup and results

4.2.1 Materials and methods

Intralipid (IL) was used to induce diffuse scattering in the tissue phantoms in this experiment; the IL here was from Sigma Aldrich and the bulk solution of 20% w/w was diluted to the required concentrations. Intralipid is a dilute mixture of emulsified fatty acids where the vast majority of the absorption comes from the water. The concentrations of IL were chosen based on previous work by Vardaki et al.[22] on the distribution of deep Raman signals in turbid media. This experimental study differs in that we have used a sheet of PTFE which acts as a "semi infinite" layer to see how this signal changes with depth compared to that of the finite sized inclusion used in Vardaki et al.'s work.

PTFE was selected for these experiments to represent the dominant Raman peak of the pathobiological material that is found in breast tissue that can be indicative of cancer. PTFE has a strong Raman peak at 734 cm^{-1} due to symmetric CF_2 stretching; this is spectrally similar to that of calcium hydroxyapatite (HAP: the material that makes up the microcalcifications that can be associated with cancer) which has a peak around 960 cm^{-1} due to phosphate stretching. Both PTFE and HAP have distinct signals from the surrounding material, be that IL or breast tissue; therefore they can be easily identified in tissue phantom studies such as this. The PTFE used here is in a distinct layer to allow a simple geometric setup in the simulations, in reality the HAP in the breast tissue appears as microcalcifications on the scale of 0.02–2 mm. The work performed with this experimental set-up allowed the collection of data in a well defined geometry enabling evaluation of the performance of our MC models.

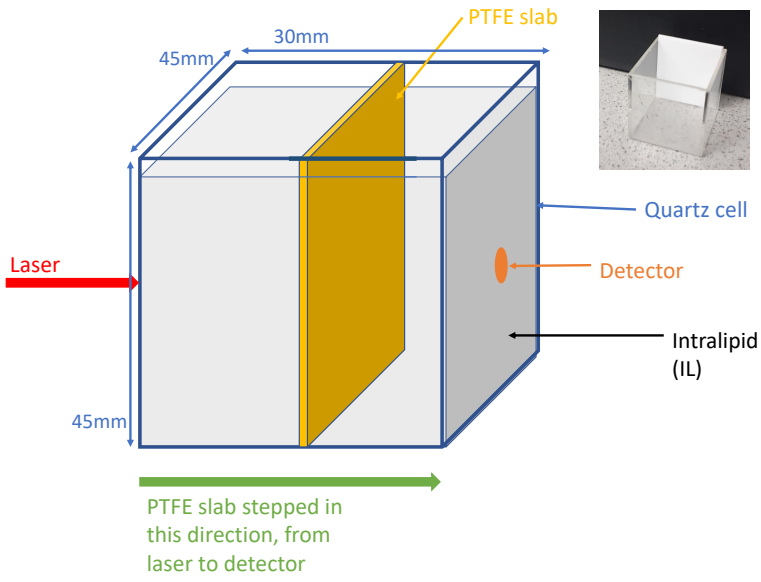


Figure 4.1: Liquid tissue phantom schematic with inset photo of empty quartz tank with PTFE slab inside.

The liquid phantoms consisted of a quartz tank (45 mm width × 30 mm depth × 45 mm height) which had an internal optical path length of 26 mm, as seen in figure 4.1. This tank contained the aqueous solutions of Intralipid at the various concentrations. A thin slab of PTFE (43 mm width × 2 mm depth × 48 mm height) was placed inside the tank such that the 2 mm path was aligned with the optical axis of the system, the 43 mm width was aligned perpendicular to this and is equal to the internal tank width, thus creating the

"semi infinite" effect desired. This setup can be seen in fig 4.1, along with an inset photo of the empty tank with the PTFE slab inside it. The PTFE used in our experiments was from "The Plastic Shop" (an online shop) and is described as "virgin grade PTFE sheet, 2mm thick". The slab of PTFE was attached to a motorised translation stage and moved to 23 different positions along the optical axis, moving from the laser side to the detector side in 1 mm steps. The quartz tank dimensions were chosen in order to provide an approximate equivalent to breast volume in mammographic screening (1.9–7.2 cm compressed breast thickness[165]).

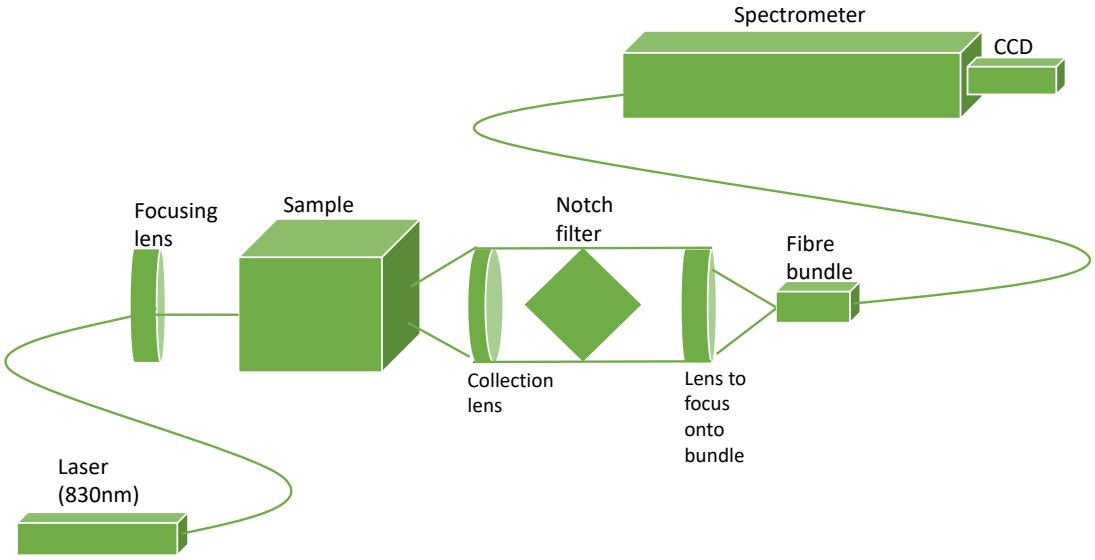


Figure 4.2: Diagram of the deep Raman setup used, in transmission mode. Based on a figure from Vardaki et al.[22]

The deep Raman setup at the University of Exeter was used in a transmission Raman configuration, as can be seen in figure 4.2. This setup has a spectrum-stabilised laser (Innovative Photonics Solutions) with laser emission at 830 nm and an output power of ~300 mW. This was coupled to a Thorlabs 105 μm multimode patch cord, then collimated and filtered by passing through two laser line filters (Semrock). These bandpass filters suppress the off-centre spectral emission from the laser line. The laser beam was brought onto the sample in a 1–2 mm diameter spot. The collimated beam passed through the sample, being scattered, absorbed and/or experiencing wavelength shifts according to the optical properties. The photons were then collected using an AR coated lens (f = 60 mm and diameter = 50 mm). The collimated light was then passed through a holographic

super notch filter (HSPF-830.0 AR-2.0, Kaiser Optical Systems) to remove the elastically scattered light (the laser photons) and imaged onto a fibre probe bundle by an identical lens to the one used for collection. The fibre bundle (CeramOptec, "spot to slit" line type bundle assembly, active area spot diameter approximately 2 mm, slit line approximately 0.25 mm x 14.95 mm) was connected to a Holospec VPH system spectrograph (Kaiser Optical Systems). This Holospec system had a focal length of 85 mm and a dispersion grating average reciprocal linear dispersion of 2.0 cm^{-1}/pixel. The spectra were recorded using a deep depletion CCD camera cooled to -75°C (Andor Technology, DU420A-BR-DD, 1024×255 pixels). The overall spectral resolution of the system was ~ 8 cm^{-1}. The signal was collected using 12 accumulations of 5 seconds and the system was calibrated using Raman bands of an aspirin tablet (acetylsalicyclic acid).

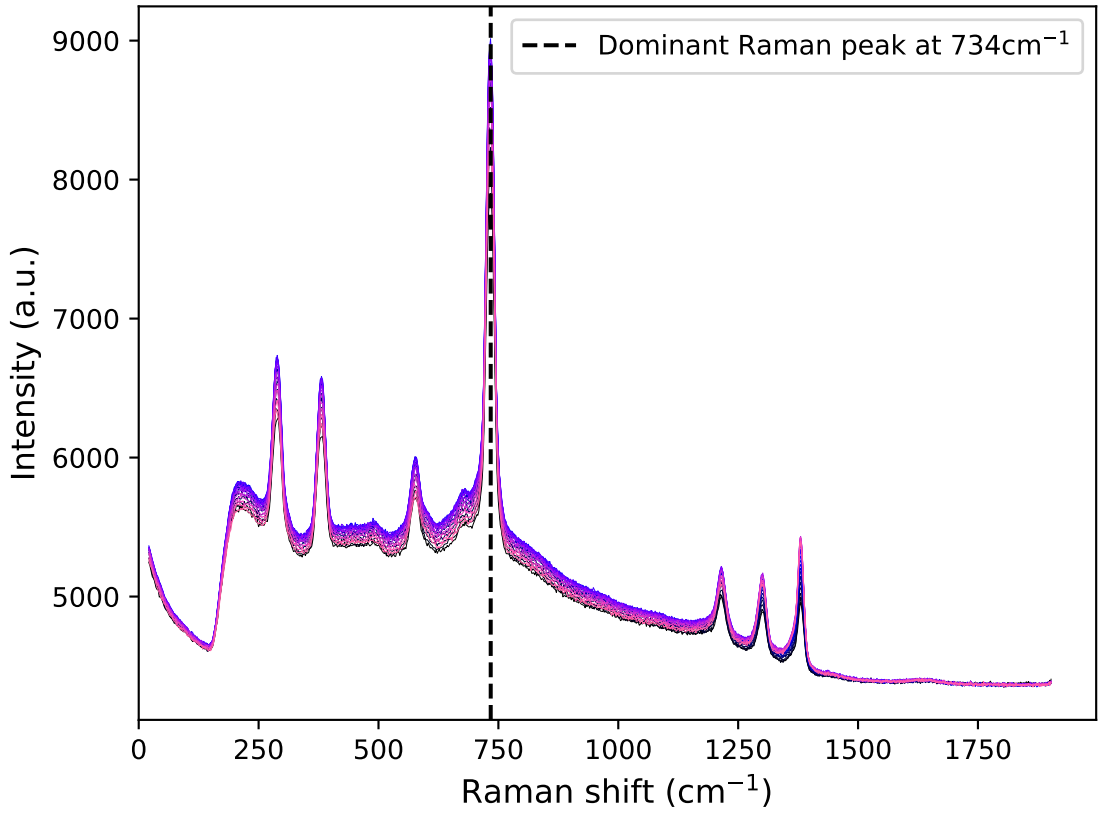


Figure 4.3: Raman spectra of PTFE for all positions in the tank with 0.25% Intralipid as the scattering agent. The dominant Raman peak at 734 cm^{-1} is indicated by the dashed line.

Raman spectra were recorded for tissue phantoms with different IL concentrations, and therefore different scattering coefficients. Parameters such as power, wavelength, beam size, and acquisition time were kept constant throughout and between mappings.

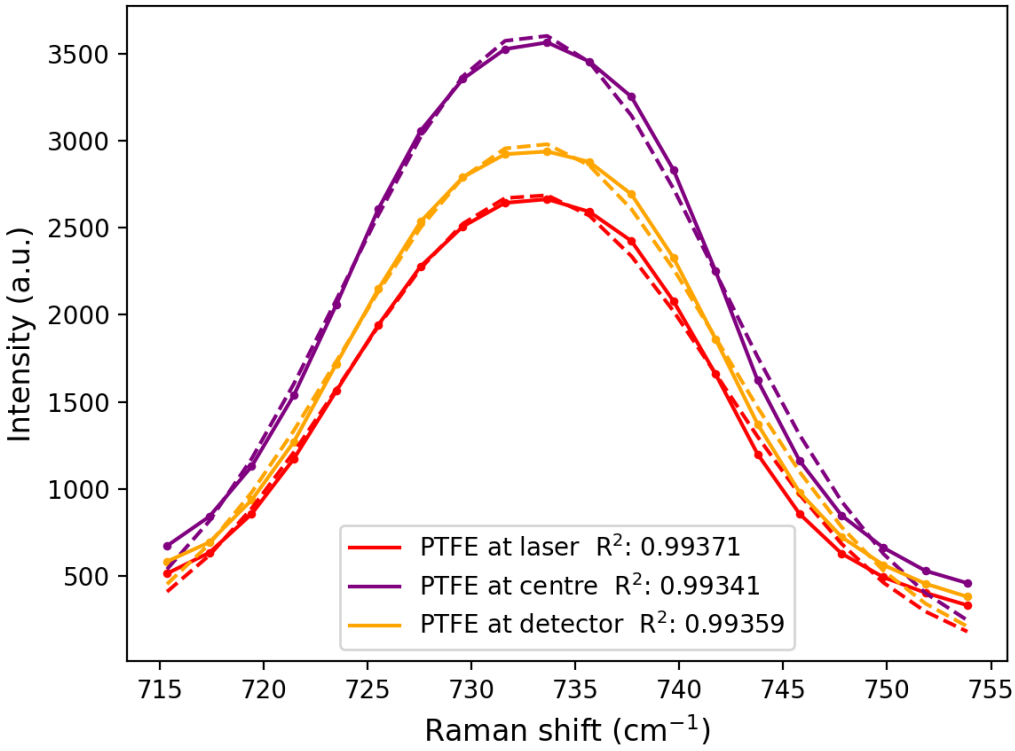


Figure 4.4: Raman peak at 734 cm^{-1} for three of the positions of the PTFE in the tank with 0.25% IL concentration: PTFE closest to the laser, in the centre, and closest to the detector. The dashed lines show the best fit Python Gaussian for each of these, and the legend contains the R^2 value for each of these fits.

The median PTFE Raman spectra for each position of the PTFE in the tank (from a total of 12 measurements taken at each position) are seen in figure 4.3, with 0.25% Intralipid present as the scattering agent. I wrote a Python code to perform a median filter on the raw data. This eliminates cosmic rays which were present in the unfiltered spectra, and prevents any skewing by these high signals.

It is clear that there is a background signal present in the spectra. Since the Monte Carlo code would only be shifting photon wavelengths according to the dominant peak (discussed further in section 4.3.1), there was no need to fit and subtract all of the background signal. Instead, we clipped each spectrum around the dominant Raman peak, drew a linear baseline through the troughs either side of the peak, and then subtracted this. Python software was then used to fit a Gaussian peak at each position along the optical axis, the results of which can be seen for three PTFE positions in figure 4.4. Using the standard integral for a Gaussian, the area under each was calculated and plotted to

show how the concentration of IL and the position of the Raman source influenced the intensity of the Raman signal detected.

4.2.2 Experimental outcomes

Figure 4.5 shows a plot of the total intensity under the dominant PTFE Raman peak for each of the concentrations of Intralipid used in the experiment: 0.25%, 0.5%, 1%, 2%, 3% and 4%. This intensity is plotted as a function of the location of the PTFE in the tank; 0 mm corresponds to the slab being closest to the laser, and 25 mm corresponds to the PTFE being closest to the detector. The units of the intensity are arbitrary as the calculation involved integrating under a Gaussian curve to the CCD output spectra. The large difference in the Raman intensities is due to higher concentrations of Intralipid decreasing the overall signal through the tank.

The error bars in figure 4.5 were calculated using the guidance from the Andor camera manual, which stated the overall noise was 1.41 times the shot noise. Each count recorded by the camera was 2.5 electrons, thus giving: noise = $\sqrt{2.5 \times count} \times 1.41$.

From figure 4.5, there is an evident maximum in the Raman intensity around the middle of the tank. The concentration of the Intralipid varies the peak of the Raman intensity and the shape of this peak. The lowest concentration of Intralipid gives the strongest signal, which is to be expected as the mean free path will be longest in this instance. It is interesting to note that the peak of the Raman intensity shifts closer to the detector as the concentration of Intralipid used increases. This behaviour is discussed in section 4.4.1.

These results support our understanding of how Raman intensity varies with depth when the source is a "semi-infinite" slab. The peak signal appearing around the centre of the scattering volume matches previous work done by Matousek et al.[166] and is in contrast to the finite Raman source results from Vardaki et al.[22] This is in line with physical expectations: the finite source is subjected to the same diffusion of laser photons, resulting in the maximum Raman photon production nearest to the laser and maximum Raman photon detection probability nearest to the detector. The difference arises from the fact that when in the centre, the finite source could be missed by the laser photons which can easily be scattered around it. Therefore the further from the laser that the finite Raman

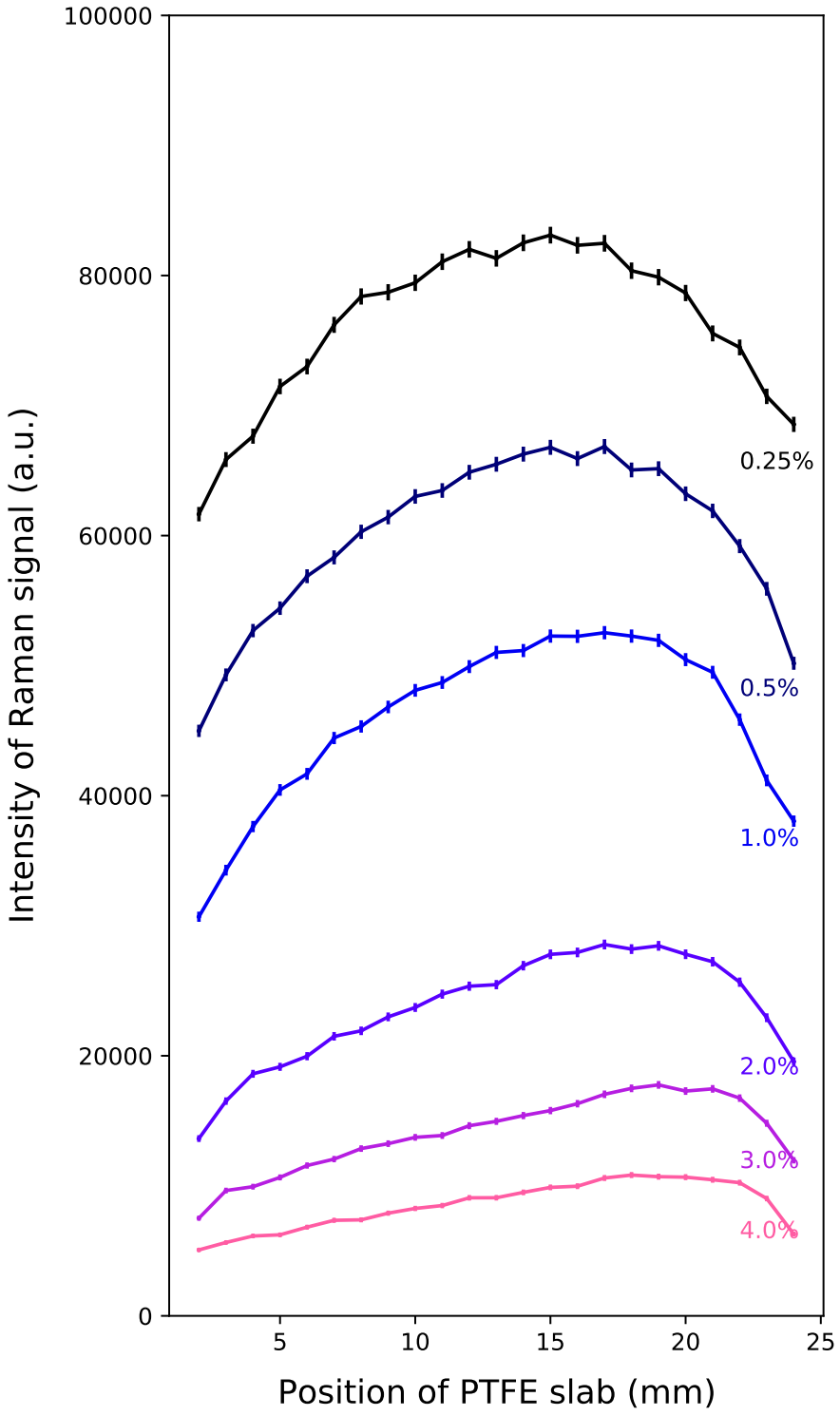


Figure 4.5: Intensity (measured as area under the peak) of dominant Raman peak in the PTFE spectrum for each of the tested concentrations of Intralipid. The laser is incident on the tank at 0 mm, and the collection is at 25 mm. The error bars represent the shot noise from the camera.

source is placed, the fewer laser photons will enter it and the fewer Raman photons are produced. This results in the Raman signal peaks for a finite source with the optics set up

in transmission geometry appearing at the laser side of the tank and also the detector side. This is the opposite result to what we have observed in this study, but both are physically correct and give insight into photon migration in turbid media and where we can expect to obtain maximum signals from.

Using the results of this experimental investigation and the understanding of how signal from a semi-infinite Raman layer behaves at different depths and optical properties, we can validate our code. The next section details the Monte Carlo numerical method employed, the optical properties required for accurate simulation, and how combining the experimental results found here with the code output can give us useful insights.

4.3 Monte Carlo setup and results

4.3.1 Monte Carlo method

An MCRT code was developed in Rust (a compiled, strongly typed, memory safe language similar to C++) to simulate the scattering and absorption processes present when light interacts with tissue. The code, when used here, returns the number of Raman photons generated and the number of those Raman photons that were detected. The incidence and detection regions are modelled to be the same as in the experimental section.

The presented simulation is a three-dimensional cuboid-shaped domain. The tank is modelled as $26 \times 41 \times 41$ mm (x,y,z) filled with the scattering agent. Into this, the Raman scattering layer is modelled as a $2 \times 41 \times 41$ mm (x,y,z) entity, with the optical properties of PTFE. Two circular triangle meshes of 1 mm radius at opposite sides of the tank entity are used to describe the laser and the detector. The laser light source (located at x = 0 mm) emits photon packets at a wavelength of 830 nm; the detector mesh (located at x = 25 mm) sums the weights of the photon packets that reach it and have a wavelength of 883.9 nm, which corresponds to a shift of 734 cm^{-1}. The optical properties of the IL are characterised by the scattering (μ_s) and absorption (μ_a) coefficients and are calculated from Grabtchak et al.[164] values for 1% IL, detailed in the optical properties section 4.3.2.

The Monte Carlo simulations were optimised to replicate the experimental behaviour while not taking too long to run; they were run with the slab at locations from 3 mm to

23 mm in step sizes of 2 mm. This was a fine enough grid to show the same results as the experiments, while being coarse enough to run in a reasonable time. Large numbers of photon packets were used in order to detect sufficient signal to noise ratio: 10^9 photon packets for each simulation, and 10 simulations for each PTFE slab location in the tank to obtain the variance. Each concentration, therefore, is comprised of 110 simulations and took approximately 1 week to run on our 32-thread computer, with the code parallelised.

A brief reminder of the Monte Carlo algorithm, plus the details of how Raman scattering is handled in this simulation, are outlined here. The initial direction of each packet is normal to the surface, facing into the tank along the x-direction to simulate a collimated laser beam. Following emission, the packet travels forward a random optical depth along its direction vector. As discussed in the Monte Carlo chapter 2, this step size is described by

$$l = \frac{-\ln(1-\xi)}{\mu} \tag{4.1}$$

where ξ is a uniform random number [0–1) and μ is the interaction coefficient of the medium the packet is currently in.

Each material in the simulation has an associated set of optical properties at the relevant wavelengths. The interaction coefficient for the medium is calculated by summation:

$$\mu = \mu_s + \mu_a. \tag{4.2}$$

For the PTFE material that is creating Raman photon packets, this scattering coefficient is calculated by summing the elastic scattering and Raman coefficients: $\mu_s = \mu_e + \mu_R$. The single scattering albedo is (again) defined as

$$a = \frac{\mu_s}{\mu}. \tag{4.3}$$

After travelling distance l, the packet then interacts with the medium through scattering. The other optical property of the medium that is required for a fully correct Monte Carlo model is the anisotropy factor (g). This variable is a value in the range [-1,1] where -1 describes entirely backward scattering, 1 is for fully forward scattering and 0 is equiv-

alent to isotropic scattering. It is used to calculate the deflection angle for the scattering packet, via the Henyey-Greenstein phase function[149]

$$P_\theta = \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 - 2g\cos(\theta))^{3/2}} \tag{4.4}$$

where P_θ is a probability density function and θ is the scattering angle in radians.

We are looking at a single wavelength shift in the PTFE Raman spectrum, therefore if the photon packet has not undergone a Raman shift we use a loop to determine if it will be shifted. This simulation is using a one-step process: a random number in [0,1) is generated, and if it is less than μ_R/μ then the wavelength of the packet is adjusted and the optical properties are updated accordingly. The packet's statistical weight is also reduced to reflect the fraction of photons in the packet that would have been absorbed in the interaction. The remaining fraction is given by

$$w' = w\,(1 - a). \tag{4.5}$$

The photon packet will continue to propagate and scatter until it exits the domain of the simulation. In the event that its statistical weight is so small that the computational power to simulate it outweighs the contribution to the output, the packet undergoes a roulette scheme.

The one-step process used here differs from that in the theoretical investigation in chapter 3 by retaining a laser packet at each Raman event. When a Raman scattering event occurs, the location at which it occurs is stored. This is in order to go back to this location, once the lifetime of the Raman photon has been simulated, to continue the re-weighted laser photon path as if the Raman event had not occurred. By doing this, the laser photons are not significantly depleted which is more akin to the physical experiment. This is an important consideration because if it is not implemented then the output of the Monte Carlo simulation does not match that of the experiment due to under-sampling the PTFE layer.

The Raman scattering coefficient, μ_R, was set to be 0.01 cm^{-1} as this gave a good pro-

duction rate of Raman packets without significantly increasing the computational time. This is several orders of magnitude greater than the realistic Raman scattering coefficient which is ~ 10^{-6} cm^{-1}. To counteract this "over generation" of Raman photons, all of the shifted photons had correspondingly decreased weights. The dominant peak in the PTFE Raman spectrum occurs at a Raman shift of 734 cm^{-1} which means a shifted wavelength from 830 nm (the laser wavelength) to 883.9 nm. The simulated packets were only shifted a maximum of once, and always to this wavelength. The experimental data was analysed at this peak only, and so comparisons between the relative Raman intensities from the MC simulations and the experiment can be found.

4.3.2 Optical properties

The absorption and scattering coefficients, along with the anisotropy factor, are the dominant optical properties for a material. These have a huge impact on the output of Monte Carlo simulations and so it is vital to get them correct. For the PTFE, the optical properties were determined experimentally by time-domain diffuse spectroscopy as used in Mosca et al.[25] : $\mu_a = 0.0477$ cm^{-1}, $\mu_s' = 9.997$ cm^{-1}. It is a very highly scattering medium with little absorption; this is in line with the physical properties we can observe by eye - it is a solid, white, opaque plastic.

The optical properties of Intralipid are more difficult to determine or locate in literature, especially at the low concentrations used in this study. Initially, the scattering coefficients used were taken from Vardaki et al.[22] as the same concentrations were used in both studies. The reduced scattering coefficient μ_s' was 9.2 cm^{-1} at 1% IL, and extrapolated linearly to the other concentrations. These were originally estimated by linear extrapolation from Michels[163] experiments on 10%, 20% and 30% Intralipid. The optical properties did not return the expected behaviour from the experiments and were deemed inappropriate for the concentrations we are working with.

A study by Grabtchak et al.[164] experimentally derived the scattering coefficient and anisotropy factor for 1% IL. The anisotropy factor of $g = 0.67$ is in close agreement with another study at similar concentrations by Aernouts et al.[157] Trying the scattering coefficients from each of these studies, only the Grabtchak value returned the behaviour from our ex-

perimental output for 1% concentration: $\mu_{s,\,\text{Grabtchak}} = 27.4\ \text{cm}^{-1}$ and $\mu_{s,\,\text{Aernouts}} = 5.6\ \text{cm}^{-1}$. Direct linear extrapolation from this value to the other concentrations did not return good results for the Monte Carlo output. Therefore, the code is validated against the experimental work carried out at 1% concentration of IL, and we can utilise the MC and experimental work together to derive estimates for the scattering coefficient of IL at the other concentrations. Despite not being a direct measurement of the scattering coefficient, this is a useful application of the results of our experimental and computational work.

The absorption coefficient measured for 1% concentration is approximately equal to that of water ($0.03\ \text{cm}^{-1}$), meaning the Intralipid absorption is negligible at these low concentrations compared to that of the water. The Monte Carlo simulation uses an arbitrary input laser power, and the recorded Raman signal is the sum of the weights of the detected photon packets. Therefore, to compare these outputs to detected power in our experiments, the relative behaviour of each was plotted and compared. Using both the relative behaviours and the differences between the Raman signals at each concentration allows us to estimate the scattering coefficients.

Scattering coefficient estimates were determined by first running a large suite of Monte Carlo models: from $12\ \text{cm}^{-1}$ to $57.5\ \text{cm}^{-1}$, in step sizes of $0.1\ \text{cm}^{-1}/0.2\ \text{cm}^{-1}/0.5\ \text{cm}^{-1}$ increasing as the scattering coefficient increased. Then, the average of each of the differences between the 1% experimental output and the other concentrations' experimental outputs were calculated. This "relative difference" was repeated for the Monte Carlo outputs, relative to the scattering coefficient value of $27.4\ \text{cm}^{-1}$ (the experimentally and now computationally validated value for 1% IL concentration[164]). These lists of relative differences were compared in order to find the scattering coefficient from the Monte Carlo suite of models that returned the corresponding relative fractional difference as seen in the experimental outputs, for each concentration. This yielded promising results, as seen in the top plot of figure 4.6, finding close matches in the fractional relative average differences which reproduce the experimental output behaviour. This fractional relative average difference means that the intensity of the dominant Raman peak for 1% Intralipid is approximately 0.8 times greater than the intensity for 4% Intralipid, as seen in figure 4.5.

The results of using the experimental differences as a basis to determine the scatter-

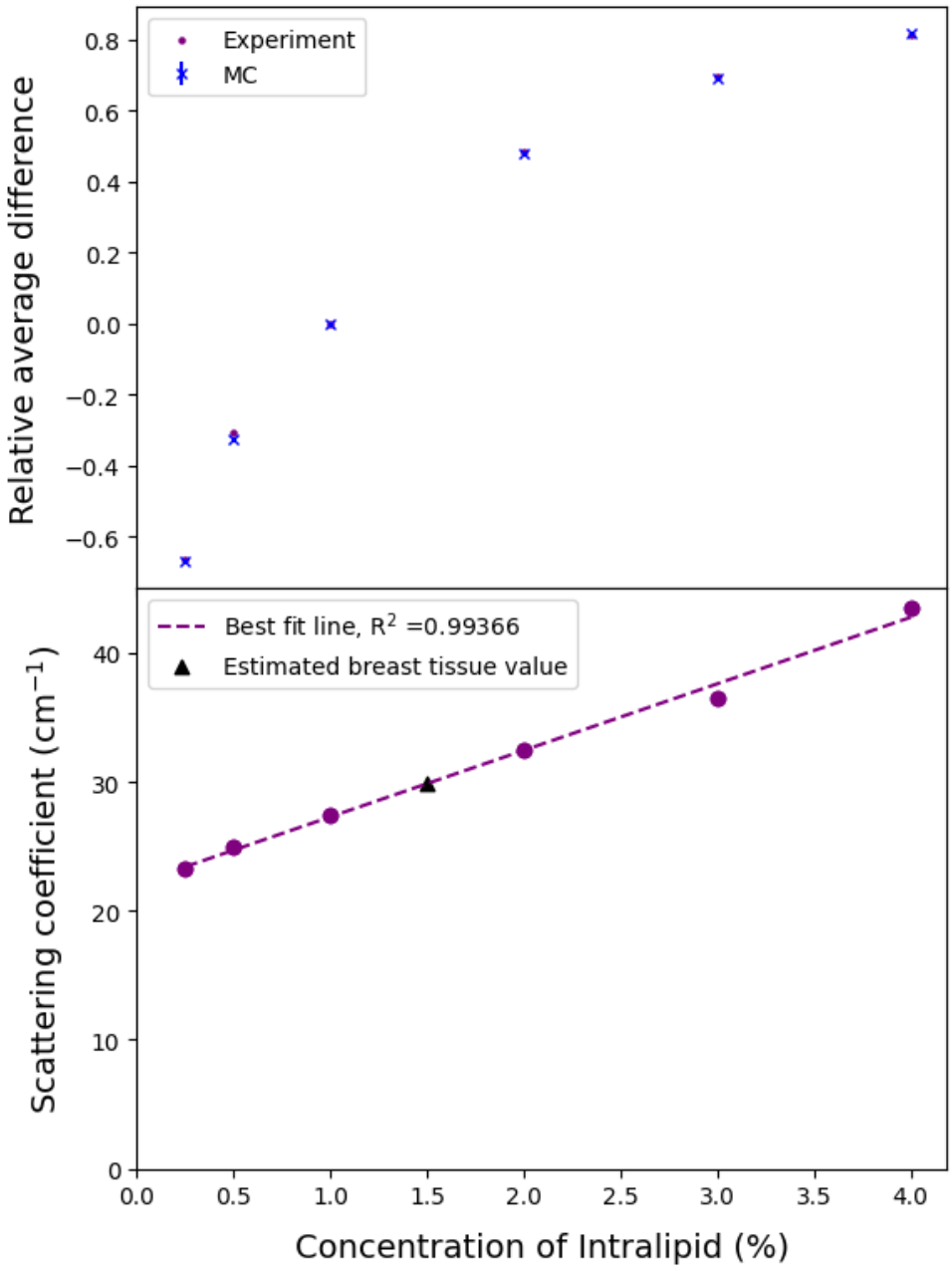


Figure 4.6: Upper: the fractional comparison of the experimental average difference (purple circles) and the Monte Carlo simulation average difference (blue crosses), both performed relative to 1% Intralipid values. Lower: plot is of our corresponding derived scattering coefficients against concentration. The variation in the scattering coefficient does not scale with the concentration directly as previously thought. The best fit line is the dashed line and has the formula $\mu_s = 5.16 \times \text{concentration}(\%) + 22.1$, and the R^2 value indicated in the legend. The black triangle indicates the estimated scattering coefficient for breast tissue based on our results.

ing coefficients can be seen in the lower plot of figure 4.6. The derived scattering coefficient found from the suite of models is plotted against the relevant concentration. Although the relationship is linear, it is clearly not a direct scaling from concentration to scattering coefficient: the scattering coefficient does not double when the concentration is doubled. The best linear fit yields a relationship of $\mu_s = 5.16(\pm 0.21) \times \text{concentration}(\%) + 22.1(\pm 0.5)$. This relationship cannot hold for all concentrations: consider a concentration of 0% as a water-filled tank, the scattering coefficient would presumably be approximately zero. However, this line appears to be a good fit for this range of Intralipid concentrations, which encompasses the scattering properties of biological tissues. The scattering coefficients were found from comparing the relative average differences of a suite of Monte Carlo models to those found in the experimental outputs and can be seen in Table 4.1. The reduced scattering coefficient, μ_s', is calculated by $\mu_s' = \mu_s(1-g)$ for each of the predicted Intralipid scattering coefficients in order to compare it to values for breast tissue. The reduced scattering coefficient in uncompressed breast tissue at 830 nm has been shown to be $\mu_s' = 9.84$ cm^{-1}.[24] The absorption coefficient of breast tissue has been estimated as $\mu_a = 0.068$–0.102 cm^{-1} for light of wavelength 825 nm in papillary breast cancer.[23] It should be noted that the optical properties have been shown to change with pressure applied and this change is dependent on patient BMI.[167] Using the reduced scattering coefficient allows for comparison between materials with different anisotropy values, as is the case in this study for Intralipid and breast tissue. From Table 4.1, it is evident that the concentration of Intralipid that most reflects breast tissue in its optical properties is 1.5%, as the breast tissue value lies between the 1% and 2% IL values derived here.

Table 4.1: Absorption (μ_a), scattering (μ_s), anisotropy (g), and reduced scattering coefficient (μ_s') used in the simulations at 830 nm, the laser wavelength. The values for breast tissue[23,24] are included in order to compare with the tissue phantoms. The absorption coefficient values for the diluted IL are that of water, as this is the dominant absorbing material in the tissue phantoms.

Material	μ_a (cm^{-1})	μ_s (cm^{-1})	g	μ_s' (cm^{-1})
PTFE	0.0477	99.97	0.9	9.997
Intralipid 0.25%	0.03	23.3	0.67	7.69
Intralipid 0.5%	0.03	25.0	0.67	8.25
Intralipid 1%	0.03	27.4	0.67	9.04
Intralipid 2%	0.03	32.5	0.67	10.7
Intralipid 3%	0.03	36.5	0.67	12.0
Intralipid 4%	0.03	43.5	0.67	14.4
Breast tissue	0.068–0.102	98.4	0.9	9.84

4.4 Comparison

4.4.1 Results and discussion

The experimental results can be seen compared to the Monte Carlo simulation output for each concentration in figure 4.7. The simulations match the experimental output faithfully; most importantly, 1.5% Intralipid concentration has the most similar scattering coefficient to breast tissue, and 1% and 2% IL experimental outputs are very well reproduced by the MC code. The behaviour of the Raman intensity peak moving closer to the detector with increasing concentration is still clear when plotting relative intensity, and is replicated in the Monte Carlo results. It is clear why if we examine the physical processes at play here.

In lower concentrations, the Raman intensity curve has symmetry between the illumination (left) and collection (right) sides. The highest number of Raman photons are generated when the PTFE block is at the illumination side as this is the point at which there are the most laser photons present in the tank, thus the highest probability for Raman scattering to occur. Once a laser photon has been converted however, it then needs to survive travelling the whole length of the tank in order to be detected. These Raman photons generated with the PTFE at the illumination side have the least chance of traversing the tank without being scattered out or being absorbed. On the other side, when the PTFE block is on the opposite side of tank (and closest to the detector), Raman photons are more efficiently detected because they have the least distance to travel to the detector. However, the fewest number of Raman photons are generated here as the laser photons need to survive traversing the tank of Intralipid.

The distribution of Raman signal intensity is therefore a balance between these phenomena, and the result is that the peak of this signal is somewhere near the middle of the tank. When the PTFE is at these locations, the number of laser photons reaching the block is sufficient to create plenty of Raman photons, and then many of these Raman photons are able to traverse the remainder of the tank to be collected. The balance shifts towards the detector as the concentration of the Intralipid increases because water absorbs Raman photons (i.e. longer wavelengths of light) more than laser photons as shown in Vardaki

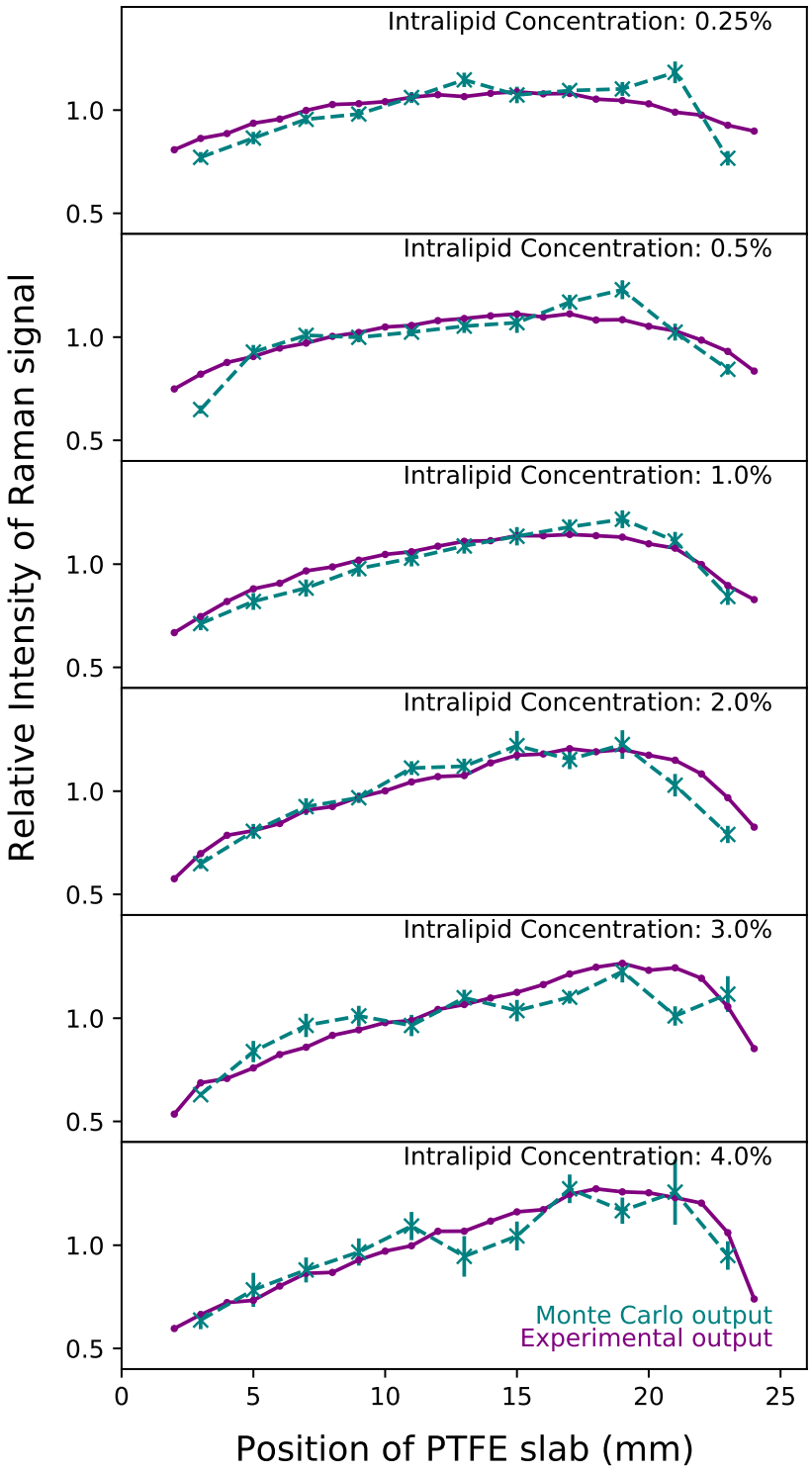


Figure 4.7: A comparison of the experimental output at each concentration of Intralipid (purple) against the results of the Monte Carlo simulations (dashed teal). The scattering coefficients for the Monte Carlo simulations were calculated from the relative difference ratio and are detailed in Table 4.1. The laser is incident at 0 mm and the collection point is at 25 mm. Monte Carlo simulations were run with 10^9 photon packets and repeated 10 times for each PTFE slab position in order to calculate the variance, shown by the error bars.

et al.[22] The increasing scattering coefficient means the total path lengths are longer and therefore Raman photons are more likely to be absorbed between generation in the PTFE and the detector compared to lower concentrations. Thus, the maximum point of Raman signal detection moves closer to the detector as the concentration of Intralipid increases because the Raman photons have a shorter total path length and are less likely to be absorbed or exit the tank before being detected. This result is similar to what would be expected from an analytical process to solve for diffuse fluorescence with two coupled diffusion processes giving the bi-phasic outcome.[168]

The scattering coefficients that we have determined here differ from the values found in other studies, except the value at the concentration of 1% as discussed earlier. There are limited studies, which do not always overlap (see Hohmann et al.[169]), and did not return the experimental behaviour we have observed. At 1% IL concentration, the values found in the literature were varied: μ_s = 2.7,[169] 5.6,[158] 27.9 cm^{-1} (adapted from Vardaki et al.[22]).

The exact reasons for the differences are outwith the scope of this study; the main difference noted here is the brand of the Intralipid used. The work by Grabtchak et al.[164] uses the same brand of IL (Sigma Aldritch) as used in this study, and returned a good result when used in our Monte Carlo simulations. Other studies did not return the expected results, and used different brands of IL. As IL is not manufactured specifically to have identical scattering coefficients, this could go some way to explaining the differences observed. Another important factor is the duration of the experiments. IL is a lipid emulsion, therefore over time there can be quality degradation due to the fat droplets coalescing. Our experimental work was completed in 2 days from opening the bottle in order to minimise this effect but it is unclear how long other studies have taken. Other measures taken to minimise the coalescence of the fat droplets were to store the IL in a fridge overnight, then ensuring it was at room temperature prior to starting the experiment. The room used for this work was a climate controlled laser laboratory. There was no noticeable degradation or change in the IL over the 2 days of the experimental work. The period of time to complete the work could have an effect on the optical properties determined by different studies. Additionally, the difference in the phantom volumes used could play a

role, especially pertaining to boundary conditions.

Our Monte Carlo numerical tool replicates the behaviour seen in the experimental results and is also capable of gaining information about the optical properties of the tissue phantoms by use of the experimental outputs. This information can then in turn be used to inform future experiments: this reduces costs by predicting experimental outcomes and simulating far more environments and Raman material distributions than currently readily available in a physical laboratory.

4.5 Conclusions

Experimentally, we have explored the Raman scattering distribution in liquid tissue phantoms in order to validate our Monte Carlo numerical tool. These tissue phantoms are useful because they approximately mimic the scattering properties of breast tissue. It has clearly been shown how the intensity of the Raman signal varies as the scattering coefficient of the liquid phantom changes. The results show that for a "semi infinite" layer of Raman scattering material, the highest signals are likely to be from the layers around the middle of the material using transmission Raman. This is contrary to earlier observations with a finite vial containing the Raman scattering material, which showed a stronger signal from the illumination and collection surfaces and the weakest signal in the middle. The finite vial experiments have a way for laser photons to bypass the inclusion; in the semi-infinite inclusion used here no photons can traverse the tank without passing through the Raman material. These results are in line with previous studies and strengthen our understanding of light transport in these situations.

In this chapter, we have shown that our Monte Carlo numerical tool with the library `arctk` is capable of reproducing experimental results. We have recovered the same behaviour in our simulated photon packets as was recorded in the experimental outcomes: a strongest signal in the middle of the phantom, with this peak shifting towards the detector for more turbid media. Using experimentally derived optical properties as our input, this has validated the code. This shows promise for future applications of Monte Carlo simulations in the journey to improve breast cancer detection.

Crucially, we have found that assuming direct linear extrapolation for the scattering coefficient between different concentrations of Intralipid may not be physically valid in the most dilute regimes. A linear relationship does exist, with a shallow gradient. By looking at the differences between the experimental outputs, we have found more realistic scattering coefficients for Intralipid at a variety of concentrations and proven that the new values return the observed signal behaviour.

Now that the code has been validated against theoretical studies and experimental work, we can use it as a predictive and insightful tool. The next chapter works towards more realistic environments with finite microcalcifications acting as a Raman source, to provide information on Raman photon propagation and detection in breast tissue.

Chapter 5

Simulations of Raman scattering from microcalcifications in breast tissue

The previous two chapters have shown that the `arctk` library, and specifically the MCRT tool, with added Raman physics reproduce the results of previous codes, and can reproduce experimental results. This chapter now looks at smaller, finite Raman sources rather than the "semi-infinite" layer used in the previous chapter.

Microcalcifications are represented by generated point sources, and these are used as the source of Raman-shifted photons. By moving the microcalcifications within a simulated tissue volume, and altering the collection geometries between SORS and TRS, we can see how different choices of detection method have different sensitivities to a variety of calcification locations.

5.1 Introduction

As discussed in chapter 1, microcalcifications are small calcium deposits that can be found in breast tissue. They come in two chemically distinct types and it is feasible to use Raman spectroscopy to differentiate between the types. There is potential in this technique

to determine whether a biopsy is medically necessary for a breast cancer diagnosis, by determining which type of calcification is present.

There are many variables to be taken into account: this work is focused on the different illumination and collection geometries for Raman spectroscopy. Both SORS and TRS are simulated, with different beam and collection radii. These simulations are also repeated for different distributions of calcifications, moving around inside the same volume. This allows predictions of how much Raman signal is collected when microcalcifications are located both on and off the optical axis. The decision was made to first run these simulations in a homogeneous volume to see the effects of changing the parameters discussed. Then, a more realistic breast tissue volume is modelled by creating a heterogeneous tissue volume with features representing human anatomy.

For the final section in this chapter, Jennifer Haskell (a PhD student at the University of Exeter) performed experimental work in the lab and the data analysis on the results from the experiment. This experimental work was to create a volume of bacon similar to an average breast volume, and move a small sample of HAP through the layers in the bacon volume. She took Raman spectra at each position of the HAP through the volume, in a TRS set-up, and calculated the HAP peak in order to see how the signal varied as the sample was moved through the volume. She also created a signal to noise plot to aid with the comparison. I then set up a simulation to mimic this experimental work, and compared the results from the two in order to find a reasonable estimate of a detection zone where HAP could realistically be detected for a given laser beam and detector.

5.2 Methods

The simulations were run in a domain of size 8 cm $\times$ 8 cm $\times$ 5 cm, giving a total volume of 320 cubic centimetres. This is within the normal ranges of breast volume measurements.[170] The 5 cm depth was selected to represent the normal depth that breasts are compressed to in a mammogram, as it would be advantageous to perform this Raman measurement at the same time as the mammogram in the clinical pathway. The other dimensions are large enough to have a total breast volume within a normal range, but not so large that the computation time is unnecessarily long, requiring the simulation of photon packets in

the outer regions that have very little chance of being detected.

The simulation was run as a two-step process: this is faster and less noisy than the one step process used for the work in chapter 4. The volume was set up with the desired optical properties, then the laser was placed at the middle of the "top" face. The laser illumination zone was modelled as a circle where packets had equal probability to be emitted from any location across the area of the circle, directed at a normal vector to the face in order to represent a collimated laser beam. The power of the laser was set as 1 W in every simulation to be high enough to get through the volume and approximately the same as those used in real experimental work. These packets were all given the same wavelength, 808 nm. The packets then propagated through the medium with the appropriate optical properties for the laser wavelength, and the energy density for the entire grid was saved. An example of this is visualised in figure 5.1 for a 1 cm diameter laser illumination spot.

The microcalcifications were then introduced by a Python script which was written to randomly generate n points in a sphere of radius r, making it very flexible. This set of points were then written to a CSV file, and used as the locations of the Raman light source.

The babbage tool in the `arctk` code was then run. This tool takes the locations of the randomly generated microcalcifications, and maps them onto the energy density map from the laser simulation. It extracts the energy density at each microcalcification point and returns the values. These values and locations are then used as the light powers and source locations respectively in a second Monte Carlo simulation. This second simulation has the same domain as the laser simulation but the packets are emitted from the generated calcification locations. A visualisation of two different groups of calcifications as emitters can be seen in figure 5.2.

Additionally, two detectors were placed in the simulation: a spectrometer covering each of the top and bottom faces to allow both TRS and SORS to be simulated. These spectrometers collected the location and the total weight of a packet multiplied by the power of that packet. The optical properties of the medium were updated for the Raman packets; the packets are emitted with a wavelength of 876 nm to correspond to the main Raman peak of HAP at 960 cm^{-1}. These packets are allowed to propagate throughout the

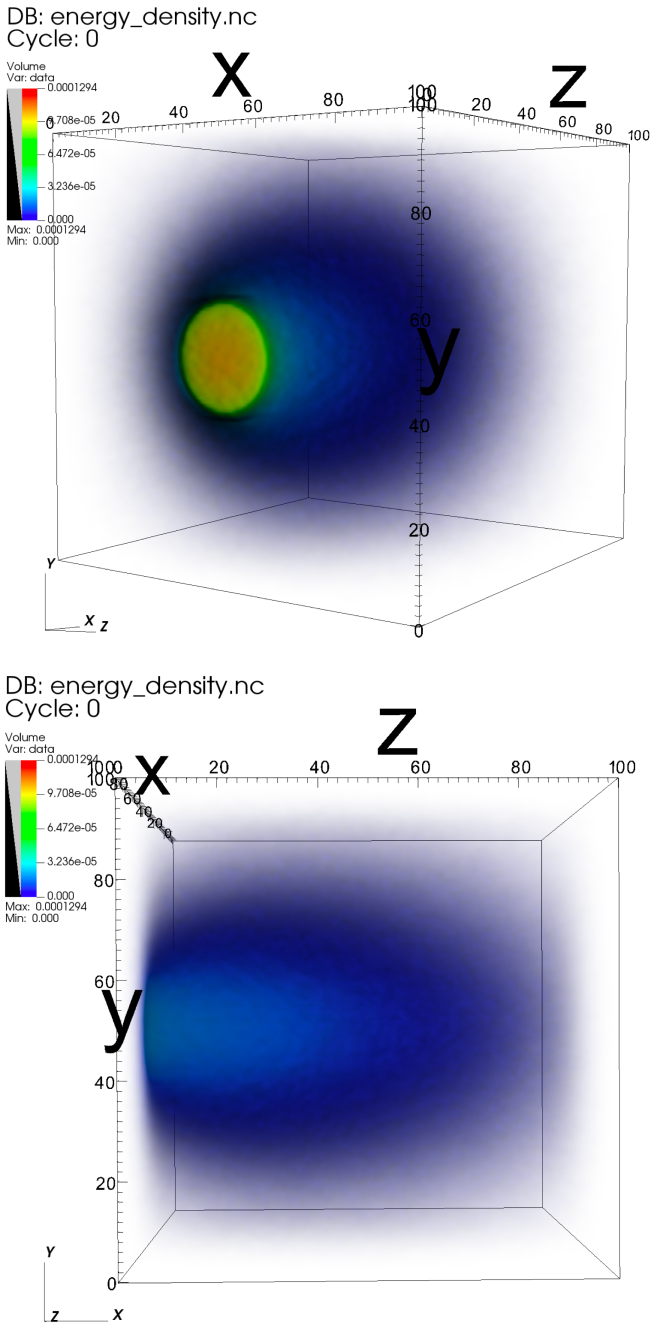


Figure 5.1: Visualisation of the energy density in a gridded volume for a 1 cm diameter, 1 W power laser beam. This is an example output saved from the first step in a two-step Raman scattering simulation.

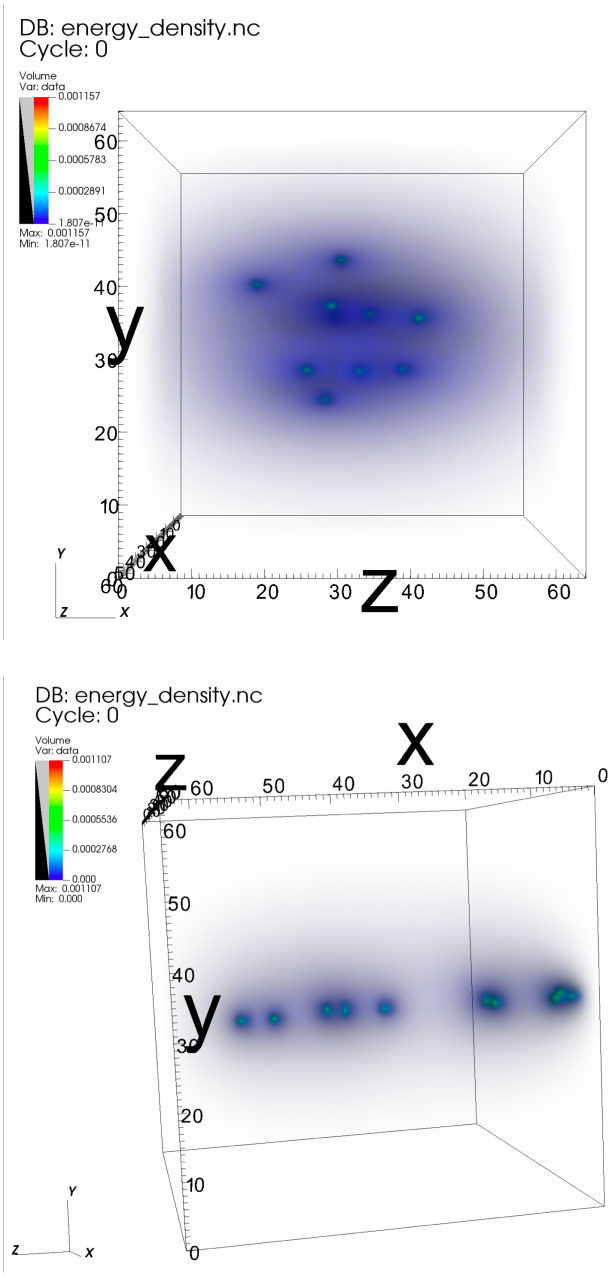


Figure 5.2: Visualisation of the energy density in a gridded volume for two different groups of 10 randomly generated microcalcifications. The power of each calcification Raman source is the energy density from the laser calculation at the location of the calcification. These are examples of light sources in the second stage of a two-step Raman simulation.

domain until they escape or are detected.

To explore the Raman signal variation when the microcalcifications are off of the optical axis, the distribution of calcifications was then moved in the domain. Therefore, the simulation process followed this format:

- Set up domain with laser wavelength optical properties, and laser light source on top face with appropriate radius.
- Run Monte Carlo simulation for an incident laser beam on a homogeneous medium and save the 3D energy density throughout the grid.
- Randomly generate n microcalcifications inside a sphere of radius r.
- Place the microcalcifications at appropriate locations in the domain.
- Run a babbage command to extract the energy density from the laser simulation at the locations of the generated microcalcifications.
- Run a Monte Carlo simulation where the calcifications are the light sources, with powers equal to the extracted energy densities from the laser simulation. Detect any Raman photons that leave from the top and bottom faces.
- Move the microcalcifications distribution to the next location in the domain and repeat.

Moving the distribution of calcifications was performed by moving the centre of the sphere that they were randomly generated in. By moving them with small steps through the domain, a 2D image can be built up to see how the Raman signal intensity varies for the same incident beam and collection geometry but at different location of Raman sources.

By collecting all of the Raman packets that leave via the top or bottom faces, a variety of "detectors" can be imposed in post-processing. Graphics to visualise this can be seen in figures 5.3 and 5.4. The yellow areas are an example of the areas inside which the packet weights would be totalled up, giving a detected signal for a given shape and size of detector. The flexibility of this method means running two Monte Carlo simulations but

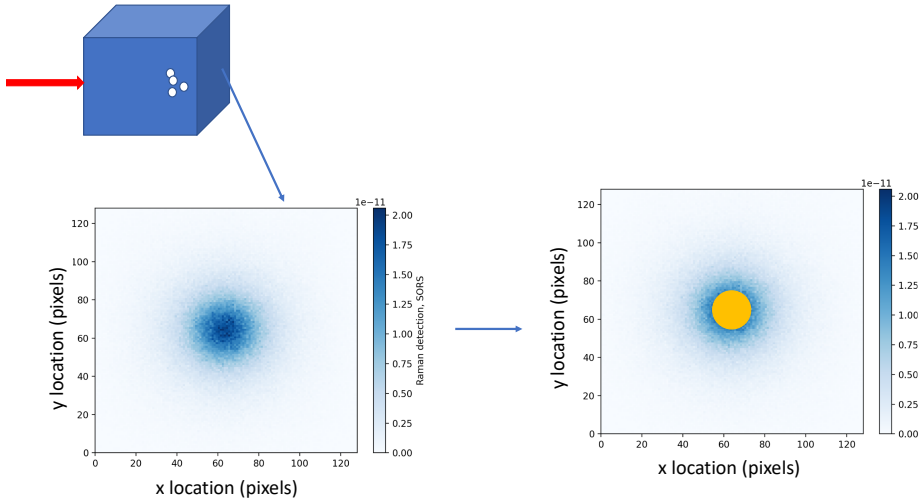


Figure 5.3: Graphic showing the post processing detector for TRS: a spectrometer in the simulation on the opposite face to the input laser collects the weights and locations of every shifted packet that exits that face. The image on the left is the output on the face opposite the laser at the end of a simulation; the image on the right shows a detector being added in post processing. All of the collected packet weights inside a circular area representing the detector are summed - thus one simulation can be used for a variety of detector radii. After the end of the simulation, the spectrometer data is processed in a Python script to sum the weights within an arbitrary circle.

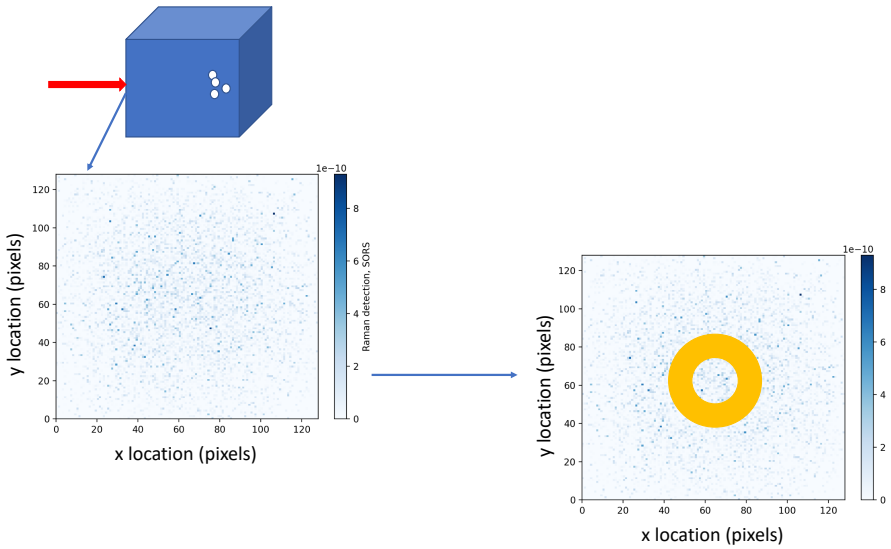


Figure 5.4: Graphic showing the post processing detector for SORS: a CCD in the simulation on the same face as the input laser collects the weights and locations of every shifted packet that exits there. The image on the left is the output on the same face as the input laser at the end of a simulation; the image on the right shows a detector being added in post processing. All of the collected packet weights inside an annulus representing the detector are summed - thus one simulation can be used for a variety of source-detector separations. After the end of the simulation, the spectrometer data is processed in a Python script to sum the weights within an annulus.

with any number of detectors, allowing SORS and TRS to be performed "simultaneously" which is unavailable in an experimental set up.

The simulations are initially run for pork tissue optical properties, for two main reasons. Firstly, it is common practice to use pork as a breast tissue phantom in Raman spectroscopy studies. Secondly, the optical properties of pork tissues have been well characterised as in Mosca et al.[25] Both of these factors made pork a good first choice for a material to simulate.

The same set of simulations are then run with bulk average breast optical properties, from Spinelli et al.[26] We then compare the results of the Raman signal detection in both materials and discuss the suitability of pork as a breast tissue phantom in experimental work.

Next, we move to a heterogeneous study by adding skin layers and gland tissue to the homogeneous model. We surround the microcalcifications in the gland material to see how this affects the detected signal.

Finally, we use experimental work which was carried by PhD student Jennifer Haskell, University of Exeter, to determine the limit of detection of HAP in pork tissue. We then use the results of this to impose on our homogeneous results a realistic contour of where it could be feasible to detect type II calcifications.

5.3 Results

5.3.1 Homogeneous tissue

In this section we will show the results of microcalcification distributions in a homogeneous pork volume, followed by a homogeneous breast volume.

5.3.1.1 Pork tissue volume

A volume was made with the dimensions 8 cm × 8 cm × 5 cm depth and given the optical properties of pork from Mosca et al.'s work[25] in time-resolved spectroscopy to determine the optical properties of various porcine tissues. The values used are listed in table 5.1.

Table 5.1: Absorption (μ_a) and reduced scattering coefficient (μ_s') used in the simulations at 808 nm (the laser wavelength) and 876 nm (the Raman wavelength). Values from Mosca et al.[25] for pork muscle.

Wavelength	μ_a (cm^{-1})	μ_s' (cm^{-1})
808 nm	0.13	4.13
876 nm	0.2	3.79

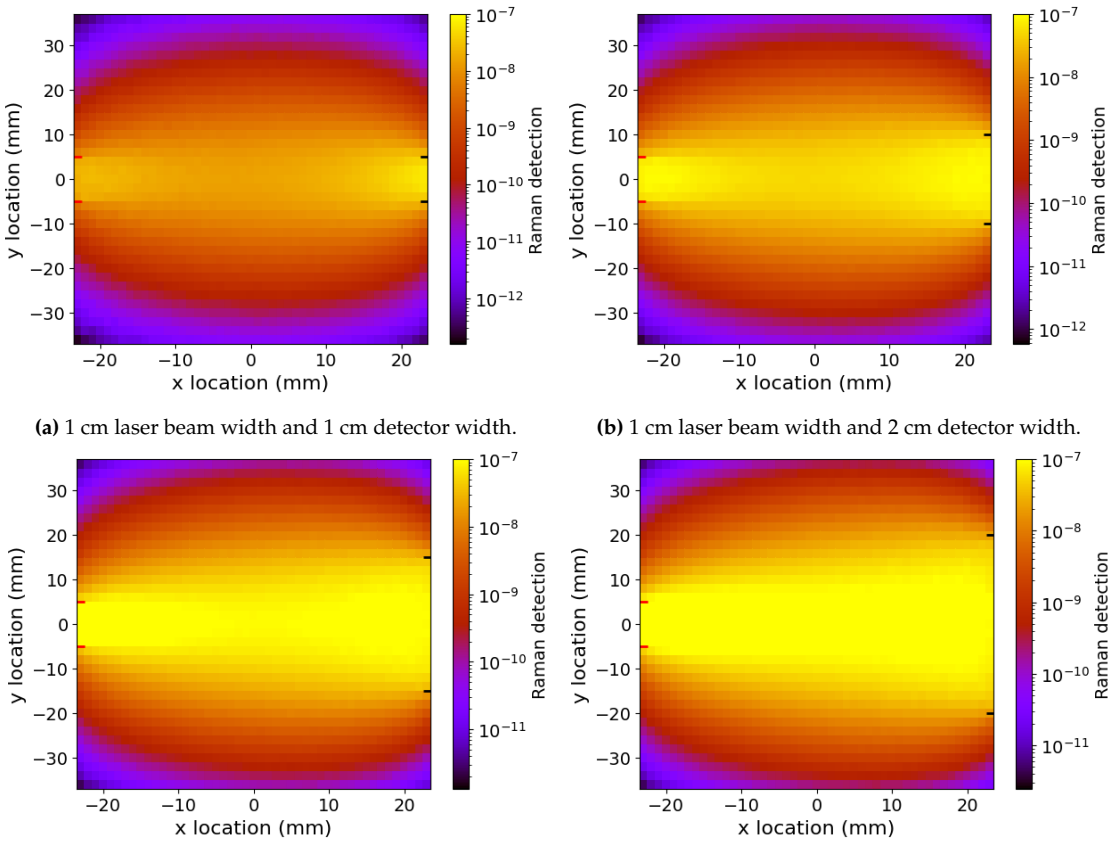
Two laser Monte Carlo simulations were run: one with a 1 mm diameter laser beam and one with a 1 cm diameter laser beam. Each of these took around one to two minutes to run. The energy densities of the grid for each of these scenarios was saved in order to determine the power of the Raman emission for required microcalcification distributions, as described in the Methods section 5.2.

5.3.1.1.1 Transmission Raman spectroscopy

Transmission Raman spectroscopy (TRS) is normally performed with a defocused laser beam, so these simulations were performed with the 1 cm incident laser beam diameter, and various circular collection apertures of diameters 1 cm, 2 cm, 3 cm and 4 cm. There were 10 microcalcifications generated within a 1 mm sphere volume. Figure 5.5 shows the effect of moving this collection of microcalcifications throughout the volume, for a variety of detector sizes.

The red lines on the left hand side indicate the width of the incident laser, the black lines on the right hand side show the width of the collection apertures. Each square on the plot has a value which is the detected Raman signal when the centre of the microcalcification distribution is in that location. Moreover, each square on the plot is a result of a Monte Carlo simulation, and each took around 30 seconds to run (on average). As the radius of the detection zone increases, the shape of the signal distribution changes. In figure 5.5a the signal nearest to the detector is comparatively greater than that nearest to the laser, however by figure 5.5d this is no longer the case.

The other note to make is that increasing the diameter of the collection geometry changes the value of the peak signal by up to an order of magnitude. Additionally it also changes the locations the calcifications need to be in order for this maximum to occur. These are idealised simulations where every packet that escapes into a detector is counted (so no numerical aperture consideration) and there are no competing signals, as every



(a) 1 cm laser beam width and 1 cm detector width.

(b) 1 cm laser beam width and 2 cm detector width.

(c) 1 cm laser beam width and 3 cm detector width.

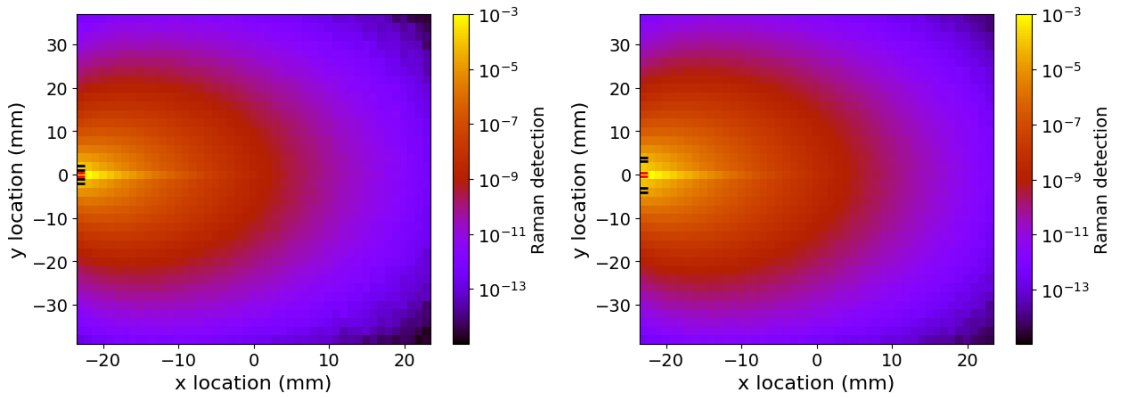
(d) 1 cm laser beam width and 4 cm detector width.

Figure 5.5: Raman signal strength in pork tissue for 10 microcalcifications randomly distributed in a 1 mm sphere, for transmission Raman spectroscopy geometry. Each rectangle in the 2D plot represents the signal detected when the microcalcification distribution is centred at that location. The laser width is shown by the red lines on the left of each subplot and the detector by the black lines on the opposite side.

packet is at the shifted wavelength. Although it appears that simply making the detector larger would increase the signal, this would need to outweigh the increased noise that would arise in a real-world scenario.

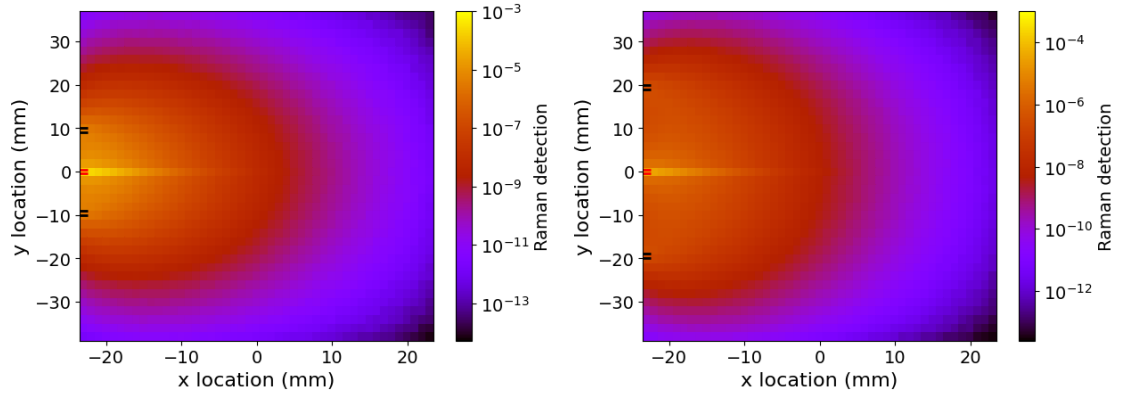
5.3.1.1.2 Spatially offset Raman spectroscopy

Spatially offset Raman spectroscopy (SORS) normally uses a narrower laser beam than TRS, so these simulations were performed with the same pork tissue volume, microcalcification generation and positioning. The detector here is an annulus of width 1 mm, on the top face where the incident laser beam was placed for the laser Monte Carlo simulation. A selection of the results for different annulus radii can be seen in figure 5.6. The radius of the annulus is measured from the centre where the laser is incident to the outer edge, and the radii displayed are 2 mm, 4 mm, 10 mm and 20 mm.



(a) 1 mm laser beam width and annulus detector with the outer edge 2 mm radius from the input beam.

(b) 1 mm laser beam width and annulus detector with the outer edge 4 mm radius from the input beam.



(c) 1 mm laser beam width and annulus detector with the outer edge 10 mm radius from the input beam.

(d) 1 mm laser beam width and annulus detector with the outer edge 20 mm radius from the input beam.

Figure 5.6: Raman signal strength in pork tissue for 10 microcalcifications randomly distributed in a 1 mm sphere, for spatially offset Raman spectroscopy geometry. Each rectangle in the 2D plot represents the signal detected when the microcalcification distribution is centred at that location. The laser width is shown by the red lines on the left of each subplot and the detector by the black lines on the same side.

In figure 5.6, the same descriptions apply as in the TRS figure: the red lines show the diameter of the incident laser beam, the black lines show the diameter and location of the annulus detector. Each detection annulus is one millimetre thick, with the outer edge at different distances from the centrally incident laser beam. Each square in figure 5.6 is the location of the centre of the microcalcification distribution and shows the Raman signal for that configuration of beam and detector.

We can see that the shape remains broadly the same for all four of the SORS detectors. Broadening the annulus (going from 5.6a to 5.6d) decreases the peak signal observed, but widens the region where the signal is comparatively large, i.e. in 5.6a there is a strong signal along the optical axis, but it quickly drops off as the calcifications are moved off-axis. In 5.6d the peak signal is two orders of magnitude lower but the drop off is also less

severe. In all of the cases, the peak signal is greater than that seen in the transmission Raman set-up, however sensitivity to calcifications at the opposite side from the laser is reduced in comparison.

5.3.1.2 Breast tissue results

Satisfied that there are sensible results in the homogeneous pork tissue simulations, we move on to a homogeneous breast tissue simulation. First we compare the results for identical breast and pork tissue simulations, to comment on the usefulness of pork muscle as a breast tissue phantom in laboratory work. Then, a similar TRS and SORS investigation is conducted for the homogeneous breast tissue volume.

5.3.1.2.1 Comparison to pork tissue

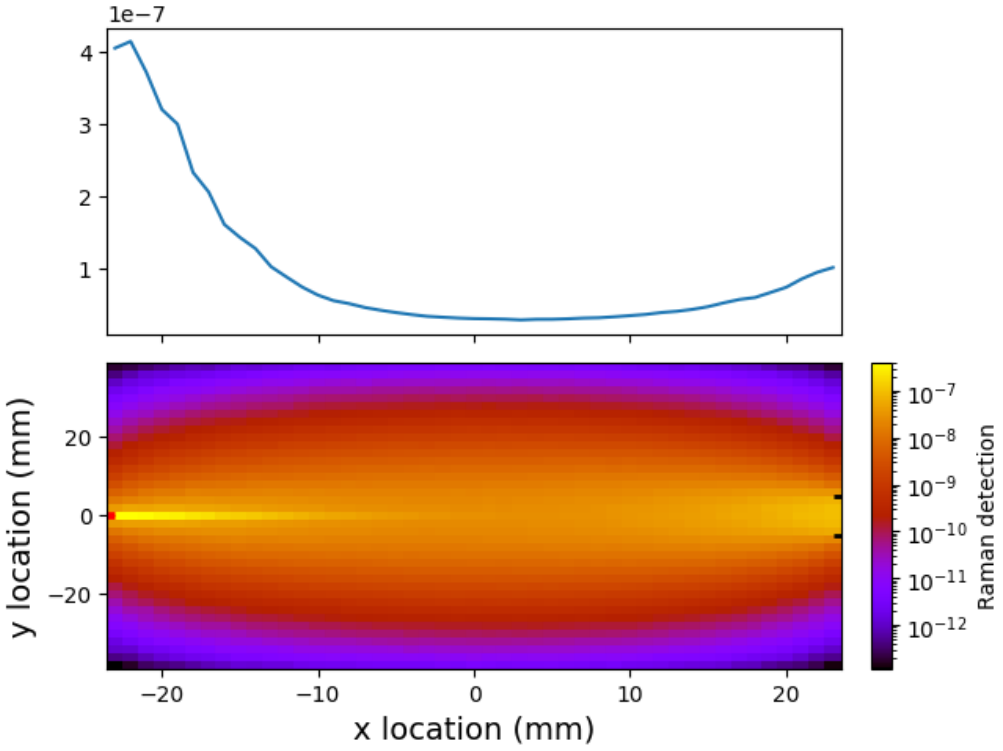
Two simulations were run: both a volume of 8 cm × 8 cm × 5 cm depth, one with pork muscle optical properties as defined in section 5.3.1.1 and the other with bulk average breast optical properties as defined in table 5.2 and taken from Spinelli et al.[26]

Table 5.2: Absorption (μ_a) and reduced scattering coefficient (μ_s') used in the simulations at 808 nm (the laser wavelength) and 876 nm (the Raman shifted wavelength). Values from Spinelli et al.[26]

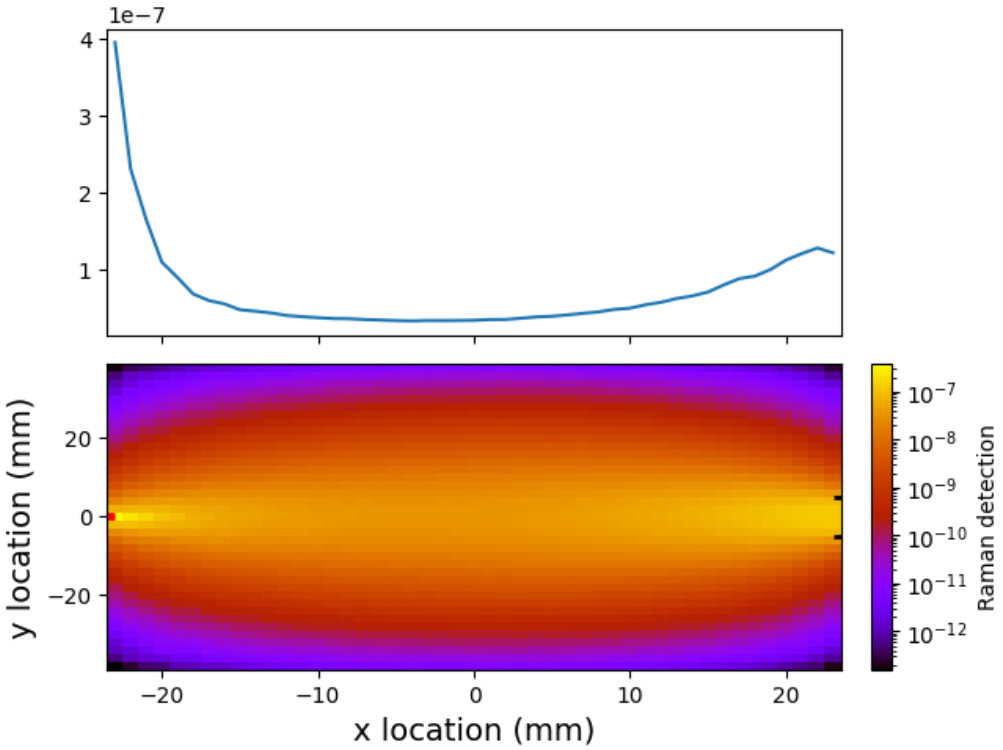
Wavelength	μ_a (cm^{-1})	$\mu_s{}'$ (cm^{-1})
808 nm	0.035	11.7
876 nm	0.052	10.99

The comparison of running the identical volumes, both with a 1 mm diameter incident laser beam and 1 cm collection diameter in TRS geometry can be seen in figure 5.7. The calcification distribution used was 10 points generated within a 1 mm sphere. Note that this pairing of beam diameter and collection diameter is not normally used in experiments, and so were used solely for interest and comparison purposes. A narrow laser beam is not generally suitable for transmission spectroscopy purposes due to reducing the penetration depths. The line plot in each scenario is taken along the optical axis.

Figure 5.7 shows the difference in using breast tissue values compared to pork muscle tissue. The main difference is in the signal drop-off as the calcifications move from the laser towards the detector along the optical axis, shown in the line plots. The decline in detected Raman signal is far steeper in the breast tissue case than in the pork. By examin-



(a) Lower plot is for a pork volume with a 1 mm laser beam width and a 1 cm collection diameter. The upper plot is the line plot taken along the optical axis, the y-axis 0 of the 2D plot.



(b) Lower plot is for a breast volume with a 1 mm laser beam width and a 1 cm collection diameter. The upper plot is the line plot taken along the optical axis, the y-axis 0 of the 2D plot.

Figure 5.7: Raman signal strength in pork and breast tissue for 10 microcalcifications randomly distributed in a 1 mm sphere, for transmission Raman spectroscopy geometry. Each rectangle in the 2D plot represents the signal detected when the microcalcification distribution is centred at that location. The laser width is shown by the red lines on the left of each subplot and the detector by the black lines on the opposite side. The laser diameter is 1 mm which is not typical for real world TRS but is done here for comparison purposes and curiosity.

ing the optical properties used, this is due to the scattering being an order of magnitude higher in the breast tissue, while the absorption is 10× lower. This is interesting because the overall signal detected is approximately the same, and the width of the higher signal regions in the 2D plot as a whole looks very similar.

Increasing the scattering while reducing the absorption means more packets survive for longer in the domain and have far longer total paths. The steep drop-off comes from these packets being less likely to make it from nearer the laser side of the tank (where they are generated in the microcalcifications) across to the detector side without being scattered away. However, this means that when the calcifications are off the optical axis, more laser light is likely to reach them as there is so much scattering present. Conversely, in the 2D plot for the pork tissue, there is a clear signal "hot spot" present directly in front of the laser on the optical axis. Due to the weaker scattering, it is easier for these packets to make it from generation in the calcifications to the detector on the opposite side. The increased absorption keeps the overall signal to the approximate level of that in the breast tissue.

The shallow "u-shape" pattern we observe along the optical axis is in line with other works on finite sources at depth in turbid media.[22] As discussed in the previous chapter, it is a consequence of the balance between shifted photon creation and detection. In physical real-world terms (rather than the specifics of these simulations): when the calcifications are in front of the laser they are more likely to generate Raman photons, when they are in front of the detector (in TRS) they are least likely. On the other hand, a Raman photon generated at the laser side of the tank has a long way to travel in a turbid medium prone to scattering it off axis or absorbing it before the Raman photon could be detected. Raman photons generated near the detector have a comparatively painless trip to being counted. Moreover, when the calcifications are near the centre of the volume, it is far easier for laser photons to have been scattered around them, and thus miss interacting and potentially experiencing a wavelength shift due to a Raman scattering event.

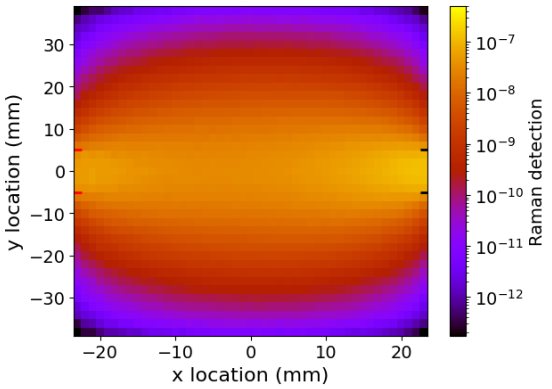
Overall, it appears that pork muscle is a good phantom for breast tissue. Although the difference in optical properties will have more of an impact in thinner samples, for depth work such as this the overall signal levels are similar. It appears that detecting Ra-

man signals at depth in pork muscle is a good estimator for being able to detect them in breast tissue. With this in mind, we now look at the impact that the density of microcalcifications has on the Raman signals.

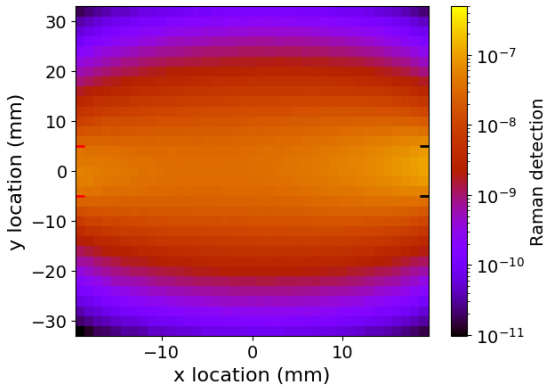
5.3.1.2.2 Transmission Raman spectroscopy

As discussed with the pork simulation results, transmission Raman spectroscopy is often performed with a broad laser beam so a 1 cm diameter laser was used for these simulations. The broader laser beam enables a lower power density, meaning that a higher powered laser can be used. This is required to achieve detectable signals through large volumes of tissue. The volume of the breast tissue was the same as before: 8 cm $\times$ 8 cm $\times$ 5 cm depth. Two detector sizes are shown in figure 5.8, and two calcification distributions; 10 microcalcifications each time but in either a 1 mm spherical volume or a 10 mm spherical volume. The 10 mm spherical volume can therefore not be moved as close to the boundaries of the simulation as the 1 mm volume. Additionally, the result of the random distribution has more of an effect on the resulting signal map as the points can easily be "off-centre" within the volume.

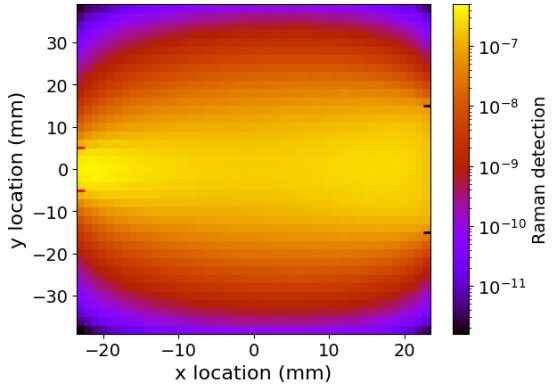
Comparing figure 5.8a to 5.8c shows the difference in broadening the detector width, as seen in figure 5.5. As expected, the differences here are similar to those seen for the pork tissue. The more interesting comparisons to make here are between figure 5.8a to 5.8b and figure 5.8c to 5.8d. These pairings have the same laser beam width and detector diameter, but a different calcification distribution volume. The differences are subtle and most easily seen in the region in front of the laser where the signal appears to be slightly stronger in the smaller distribution. This could be due to the calcifications all being in the area with the highest laser energy density, whereas in the 1 cm distribution some could be generated further out. Additionally, the 1 cm distribution is not able to be moved as close to any of the edges as the 1 mm one, including the edge by the laser so the calcifications in the larger distribution will never be in the higher energy density region. Overall, these maps do show expected behaviour in signal variation as predicted by the earlier work in this chapter.



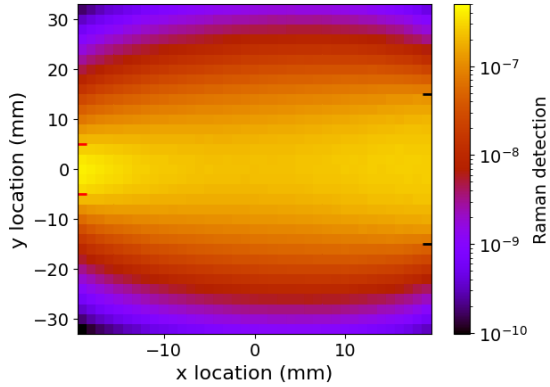
(a) 1 cm laser beam width and 1 cm detector width, with 10 microcalcifications randomly distributed in a 1 mm spherical volume.



(b) 1 cm laser beam width and 1 cm detector width, with 10 microcalcifications randomly distributed in a 10 mm spherical volume.



(c) 1 cm laser beam width and 3 cm detector width, with 10 microcalcifications randomly distributed in a 1 mm spherical volume.



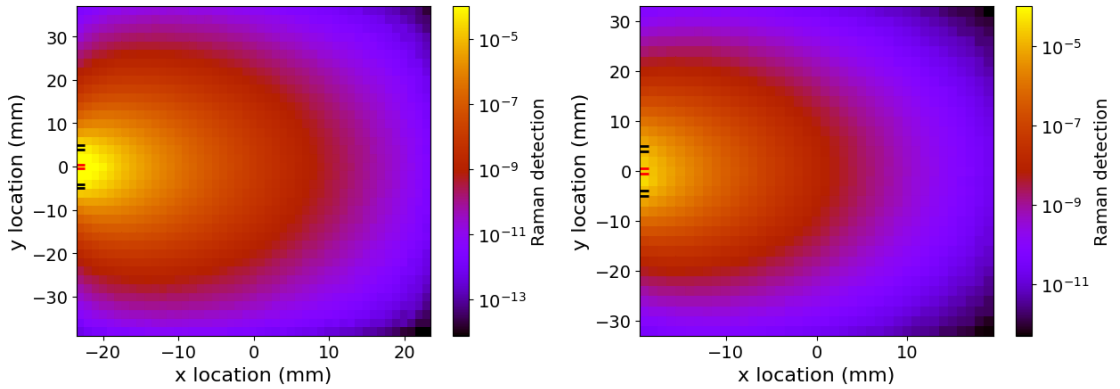
(d) 1 cm laser beam width and 3 cm detector width, with 10 microcalcifications randomly distributed in a 10 mm spherical volume.

Figure 5.8: Raman signal strength in breast tissue for 10 microcalcifications randomly distributed in different sphere volumes, for transmission Raman spectroscopy geometry. Each rectangle in the 2D plot represents the signal detected when the microcalcification distribution is centred at that location. The laser width is shown by the red lines on the left of each subplot and the detector by the black lines on the opposite side.

5.3.1.2.3 Spatially offset Raman spectroscopy

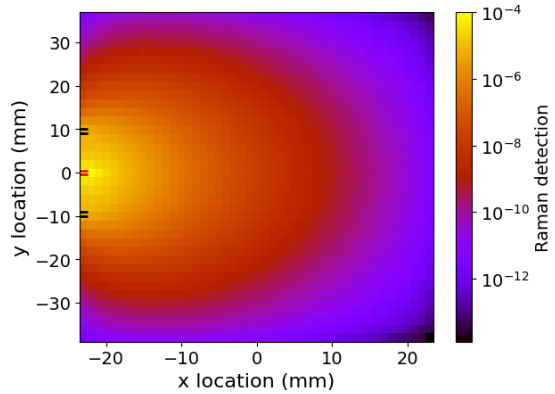
As discussed in the pork SORS results, a narrower laser beam is generally used, so these simulations were performed with the same breast tissue volume, microcalcification generation, and positioning. The detector here is an annulus of width 1 mm, on the top face where the incident laser beam was placed for the laser Monte Carlo simulation. A selection of the results for different annulus radii and different microcalcification distributions can be seen in figure 5.9. The radius of the annulus is measured from the centre where the laser is incident to the outer edge, and the radii displayed are 5 mm and 10 mm.

Comparing figures 5.9a and 5.9c shows the difference in increasing the source-detector distance for the same 1 mm calcification distribution. This is the same set up

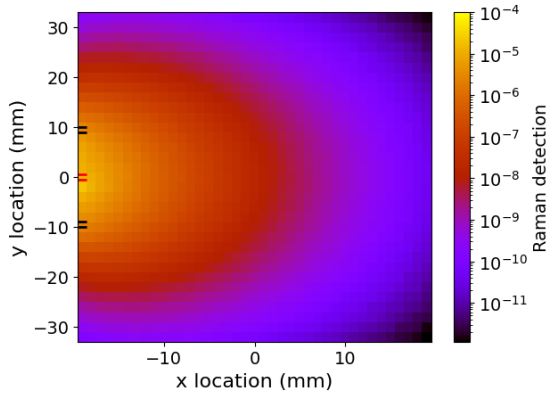


(a) 1 mm laser beam width and annulus detector with the outer edge 5 mm radius from the input beam; 10 microcalcifications randomly distributed in a 1 mm spherical volume.

(b) 1 mm laser beam width and annulus detector with the outer edge 5 mm radius from the input beam; 10 microcalcifications randomly distributed in a 10 mm spherical volume.



(c) 1 mm laser beam width and annulus detector with the outer edge 10 mm radius from the input beam; 10 microcalcifications randomly distributed in a 1 mm spherical volume.



(d) 1 mm laser beam width and annulus detector with the outer edge 10 mm radius from the input beam; 10 microcalcifications randomly distributed in a 10 mm spherical volume.

Figure 5.9: Raman signal strength in breast tissue for 10 microcalcifications randomly distributed in different sphere volumes, for spatially offset Raman spectroscopy geometry. Each rectangle in the 2D plot represents the signal detected when the microcalcification distribution is centred at that location. The laser width is shown by the red lines on the left of each subplot and the detector by the black lines on the sane side.

as seen before in figure 5.6. The change to bulk breast tissue optical properties has not vastly changed the behaviour, as expected from our comparison work. Also the signal maps observed match what would be expected, showing that a greater source-detector separation leads to collecting photons from greater depths in the volume. It is more interesting here to investigate how the calcification distribution volume affects the signal maps. By comparing figure 5.9a to 5.9b, and figure 5.9c to 5.9d, we can see how changing from 10 point sources inside a 1 mm spherical volume compares to 10 point sources in a 10 mm spherical volume.

From the colour bars, we can see that the simulations with the calcifications in a 1 mm volume have higher peak signal compared to those with the 1 cm volume. However,

the larger calcification distribution simulations show a deeper region of peak (or near to peak) signal, denoted by the yellow zones in the 2D maps. Since all of these simulations use the same laser energy density distribution (1 mm beam width in the breast tissue volume), this difference comes entirely from the calcifications. The point sources in the larger volume have potential to be more spread out, and thus can be in the regions of higher laser energy density for more positions in the mapping, leading to a larger region of highest signal. The small point source volume simulations will have higher total energy density by virtue of every point being in the high energy density region, but then once the centre of the distribution reaches a low energy density region every point will have a low value. Thus the 1 mm distribution has a higher peak, and a sharper fall, than the 1 cm distribution.

This work on homogeneous signal maps for calcification distributions have allowed us to see light transport in a way that experiments do not. For example, spreading the calcification distributions and moving them throughout the volumes both along the optical axis, and off, is far more logistically difficult in real life. The two detection methods of SORS and TRS are useful for sampling different aspects of a volume. SORS is more effective at giving depth information but is limited to a volume around the laser illumination zone. TRS is better for sampling the entire volume but with lower detected signals and no depth information. For the purposes of microcalcification classification, TRS is more useful: identifying the calcifications requires sampling the bulk of the volume, and it fits into the mammogram set-up more easily (since there are top and bottom plates).

The next, and final, stage in this investigation is to create a heterogeneous volume that more accurately reflects real world breast tissue, and see how this affects the detected signals.

5.3.2 Heterogeneous tissue

Imagining integrating Raman spectroscopy into the clinical pathway, it seems the best place would be to include it at the time of mammogram. This prevents adding an extra step to the diagnostic procedure, and the compression of the breast tissue means that we have a target depth of ~ 5 cm.

The investigations in the homogeneous section showed us clearly how the signal changes when moving the Raman source, the microcalcifications, along the optical axis and also off the optical axis in bulk average breast tissue. If Raman spectroscopy were to be added to the mammogram procedure, it would make sense for this to be in transmission geometry (TRS). The aim would be for the technique to be non-invasive so it is important to include skin layers in our model, and to have a more representative breast anatomy model. First we will examine the addition of skin, before moving on to have the microcalcifications inside gland tissue, as seen in ductal carcinoma scenarios.

5.3.2.1 Skin addition

Transmission Raman spectroscopy performed non-invasively will therefore require the laser light to penetrate skin at the top surface of the breast, and the Raman shifted photons to penetrate skin again at the bottom surface.

A simulation was run with a three layer skin model placed at the top and bottom surface of the bulk breast volume used in the homogeneous section. The volume of the simulation remained the same size and the skin layers' depths and optical properties are detailed in table 5.3, while a diagram of the set up can be seen in figure 5.10. The values for the skin layer optical properties were taken from Bhandari et al.[27]

The three layers of skin used in this simulation are the upper epidermis, lower epidermis and the dermis. They each have different constituent parts, and hence different optical properties. The upper epidermis, which represents the stratum corneum, is the outermost layer of skin that acts as a protective barrier to the lower layers and the body. This layer is mostly made of dead keratinocyte cells and melanin dust. Melanin dust absorbs and scatters light; keratinocyte cells were assumed to absorb only. The lower epidermis is a bloodless layer with living keratinocytes and melanocytes. The dermis is deeper again, and contains many different constituents: collagen, hemoglobin, plasma, fat and water. A lot of the absorption in this region is due to the blood (hemoglobin), and most of the scattering is by the collagen. It is important to note that darker skin contains more melanin, and thus this would have an effect on the optical properties in the upper and lower epidermis.[27]

Table 5.3: Absorption (μ_a), anisotropy (g) and scattering coefficient (μ_s) used in the simulations at 808 nm (the laser wavelength) and 876 nm (the Raman wavelength). Values from Bhandari et al.[27]

Skin layer, thickness	Wavelength (nm)	μ_a (cm^{-1})	g	μ_s (cm^{-1})
Upper epidermis, 0.02 mm	808	15.7	0.643	516.9
	876	11.2	0.610	404.5
Lower epidermis, 0.08 mm	808	15.7	0.852	1157.1
	876	11.2	0.838	945.1
Dermis, 1 mm	808	0.224	0.877	173.5
	876	0.224	0.891	141.6

Table 5.3 lists the optical properties of the skin layers; this is due to their biological components as discussed above. Consistently across the layers the scattering coefficient is reduced as the wavelength increases and the anisotropy is very forward scattering as expected for biological tissue. A diagram of the set up of the simulation domain can be seen in figure 5.10, with skin present at the top and bottom of the bulk breast volume. The input laser beam is at the top of the volume, and the detector at the bottom, in order to replicate the transmission Raman spectroscopy set up. The total depth is still 5 cm, with the skin layer thicknesses as outlined in table 5.3.

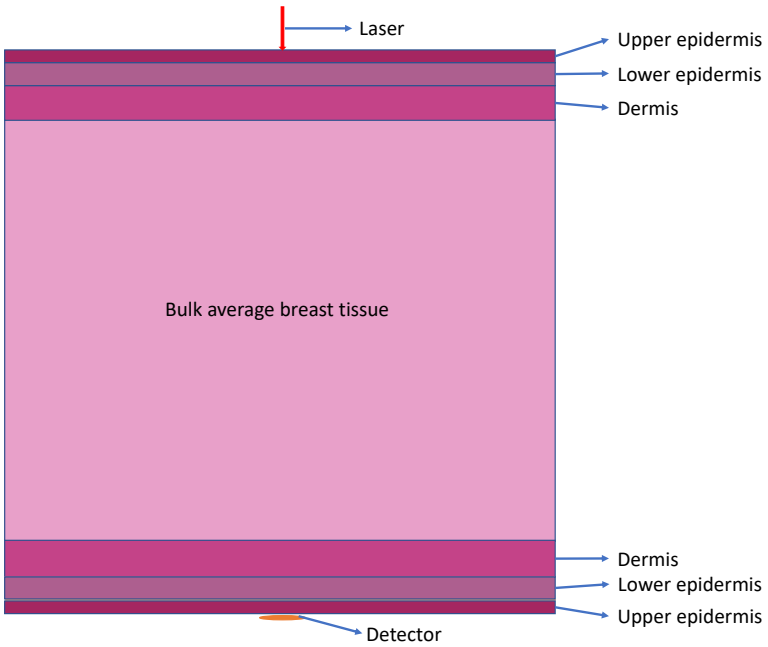


Figure 5.10: Graphic to demonstrate the simulation domain when skin layers were added to the top and bottom surface of the bulk average breast tissue volume. For transmission Raman geometry, the laser would be incident at the top, and the detector underneath.

The simulation was run in the same two-step process as the homogeneous volume: a laser beam simulation with bulk average breast tissue between the two sets of skin layers returning the energy density through the volume, followed by a Raman photon simulation

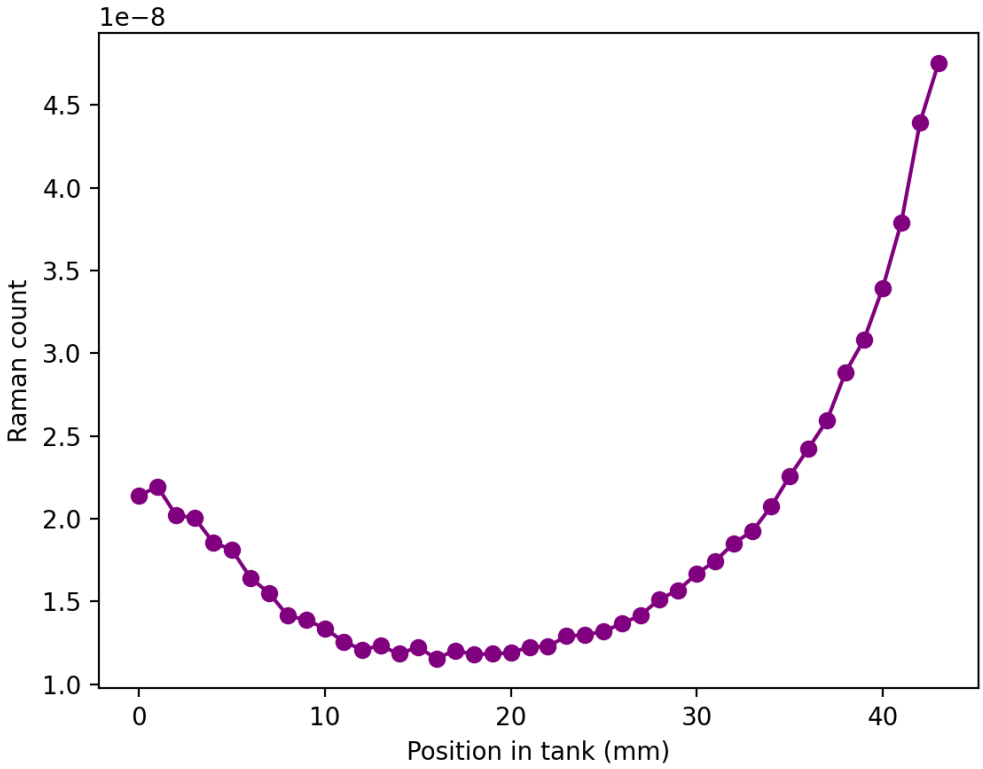


Figure 5.11: Plot of the detected Raman signal for a bulk breast volume with skin at the top and bottom (in front of the laser and detector). The laser and detector both had a diameter of 1 cm, there were 10 microcalcifications inside a 1 mm spherical volume and these were moved 1 mm at a time along the optical axis.

where the Raman packets are emitted from point sources representing the microcalcifications. The microcalcifications were generated in the same manner, randomly distributed within a given volume.

The simulation was run with a 1 W laser, 1 cm beam diameter, and 10 microcalcifications within a 1 mm sphere for the Raman sources. The microcalcifications were simulated along the optical axis only, to see the effect the skin has on the detected signal. By keeping them on the optical axis, we can compare the detected signal in the locations where we are most likely to be able to see them. The detector used a 1 cm diameter circular aperture; the number of photons run was $1{\times}10^6$. The output can be seen in figure 5.11, where 0 mm on the x-axis corresponds to the laser side of the tank.

A comparison of figure 5.11 to the line plot for breast tissue in figure 5.7, can be seen in figure 5.12 with the bulk breast volume model being reduced by a factor of 10 to show the shape difference. This is due to the detected signal once the skin is included in

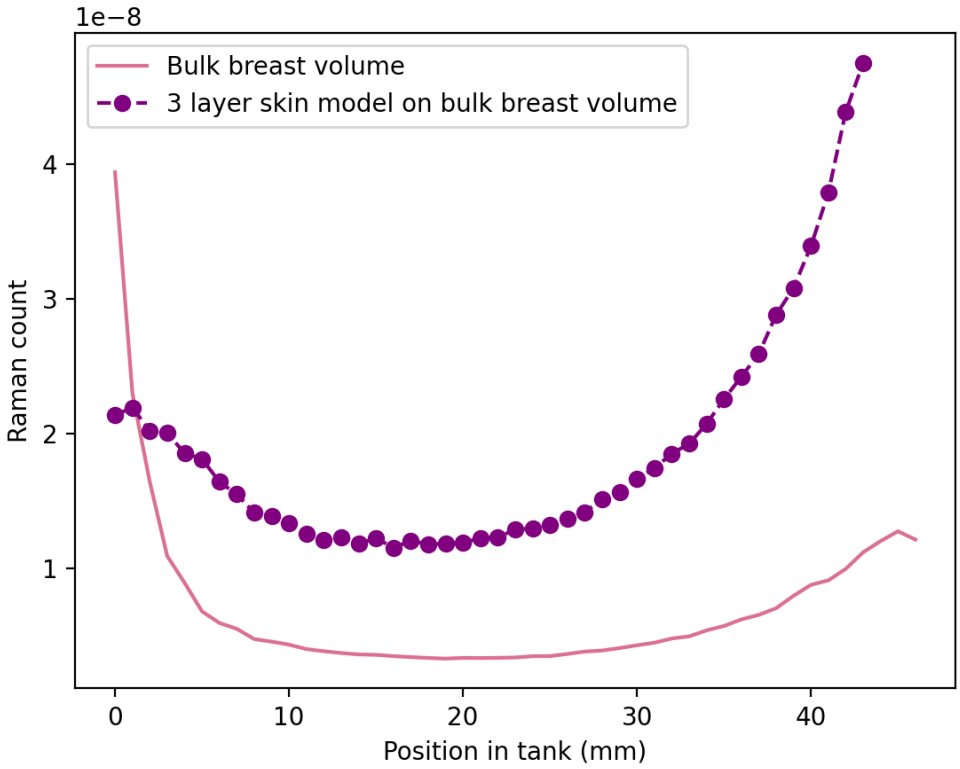


Figure 5.12: Plot of the detected Raman signal for a bulk breast volume with skin at the top and bottom (purple), compared to the shape of the same simulation but without skin (pink). The model without skin has been scaled down by a factor of 10 to be plotted on the same y-axis.

the simulation being much more attenuated. Although this was run along the optical axis only, it is reasonable to assume that this would apply off-axis as well. A notable difference is in the u-shape observed. The same behaviour of a minimum in the middle of the volume is observed here, which is promising as it is what we would expect. However, in the skin model this u-shape is uneven with a higher signal in front of the detector, as opposed to the case with no skin, where the peak signal appears in front of the laser. This change is due to the larger absorption in the skin layers, specifically the lower and upper epidermis, compared to the bulk average breast tissue.

For the final step in this study, we now look at how the signal changes if the bulk average breast tissue is instead adipose tissue by using optical properties from the literature, and the sphere containing the calcifications is given the optical properties of a gland.

5.3.2.2 Gland addition

Another step towards a more realistic model is to alter the homogeneous bulk average breast tissue material used in the previous section. A more realistic scenario is having adipose and glandular tissues in the breast volume. Based on the discussions of breast anatomy in the introduction, and the prevalence of ductal carcinoma, we decided to make the bulk of the material adipose (fatty) tissue, and then encase the calcifications inside a sphere with gland properties. This will demonstrate whether the gland boosts or hinders Raman signal in comparison to the homogeneous material.

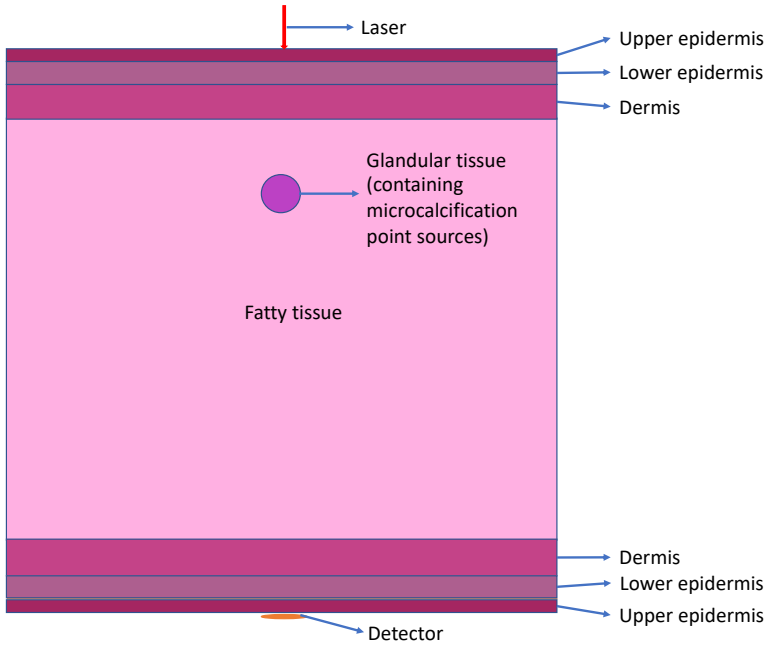


Figure 5.13: Graphic to demonstrate the simulation domain when skin layers were added to the top and bottom surface of a fatty tissue volume with a sphere of glandular material containing the microcalcification point sources. For transmission Raman geometry, the laser would be incident at the top, and the detector underneath.

The set up for the simulation is the same as in the skin section, and the two step process is performed again. As the calcification distribution was moved through the domain, so too was the sphere of gland material. A graphical representation of the set-up can be seen in figure 5.13. This means that in this case, a laser simulation was run before each Raman simulation to correctly model the interactions between the laser photons and the gland material sphere in the appropriate location. Some literature optical properties for the adipose and glandular tissues are detailed in Table 5.4 and are taken from Peters et al.[28]

Table 5.4: Absorption (μ_a), anisotropy (g) and scattering coefficient (μ_s) used as the basis for the simulations at 808 nm (the laser wavelength) and 876 nm (the Raman shifted wavelength). Values from Peters et al.[28]

Tissue type	Wavelength (nm)	μ_a (cm^{-1})	g	μ_s (cm^{-1})
Adipose	808	75.7	0.976	308.8
	876	76.8	0.976	303.6
Glandular	808	46.2	0.964	323.8
	876	52.4	0.964	288.8

The values in table 5.4 are taken from a study using breast samples to characterise optical properties in different tissues and lesions in the breast. Comparing these values to the ones used in the bulk breast example (Table 5.2), and in the Intralipid comparison work of chapter 4, the absorption coefficients here are around 10× greater. We therefore ran the simulation with the coefficients as stated in Table 5.4, but with the absorption coefficient reduced by a factor of 10 in both the adipose and the glandular tissue. When run with the optical properties as found in Peters et al.,[28] the signal was too low to be of any use, and would require a larger simulation with more photon packets to give meaningful output. Using values more in line with other literature work seemed like a sensible compromise, given the variety of values in the literature.

Figure 5.14 shows the "u-shape" that we have become familiar with; figure 5.15 shows the comparison between including only a skin model, and a more complete anatomical model of skin, fatty tissue and glandular tissue. The impact of differentiating fatty breast tissue from glandular tissue containing the calcifications rather than modelling bulk average breast tissue appears to make the "u" more even, and decreases the signal by around a factor of 4. The symmetry of the plot indicates that the microcalcifications might be better detected when in front of the laser than assumed from bulk tissue results. The regular "increase - decrease" pattern seen in sequential points on the left side are an artefact of the sampling of the grid: this would be reduced on a finer grid when modelling the laser energy density.

The work presented thus far in this chapter has been entirely simulated, while being based on real-world estimations of the optical properties of pork, bulk breast and specific breast tissue types. While useful for investigating the behaviour of light throughout the volume, and the effect of including more realistic geometries and anatomy such as skin, it would be useful to have some metric for linking the values seen in the simulation output

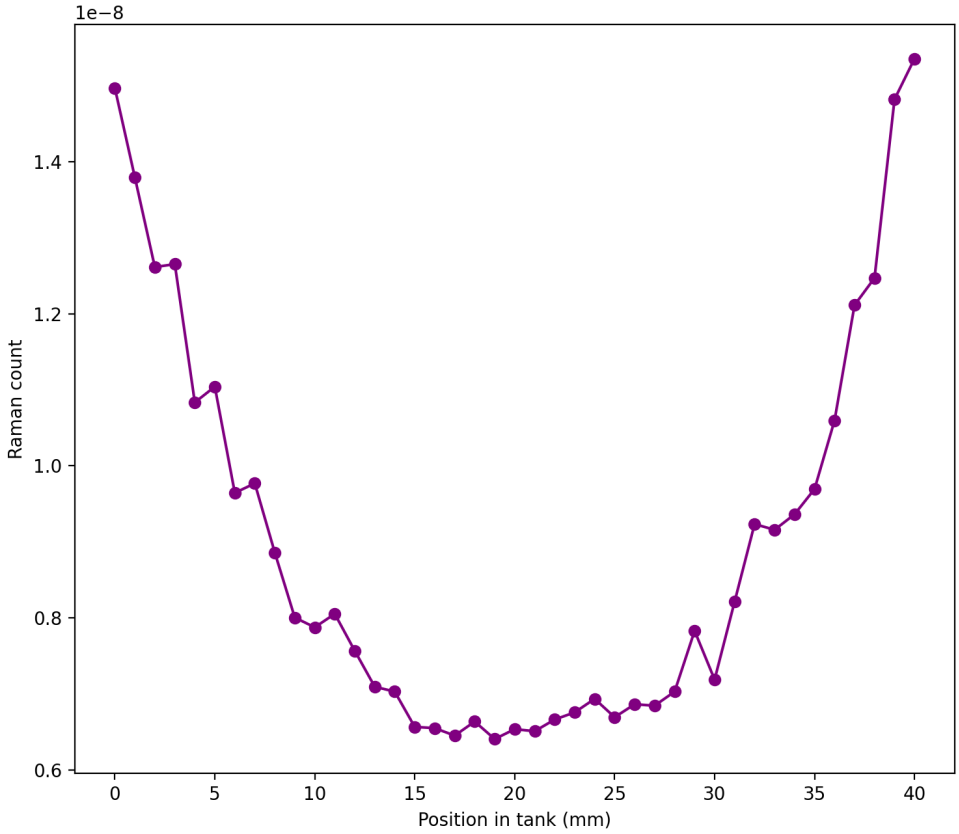


Figure 5.14: Plot showing the detected Raman signal for a fatty tissue volume with skin at the top and bottom (in front of the laser and detector), and the calcifications contained within a spherical gland. The scattering and absorption coefficients have been lowered by a factor of 10 to make them more similar to the bulk breast literature results from earlier. The laser and detector both had a diameter of 1 cm, there were 10 microcalcifications inside a 1 mm spherical volume and these were moved 1 mm at a time along the optical axis.

to laboratory work. The next section investigates this in some more detail.

5.4 Detection limit

The simulation results presented thus far have been useful for demonstrating the effects of increasing the complexity of the model, and how Raman signal varies throughout a large volume. The result of moving the Raman source off the optical axis is something that can be difficult to perform in a laboratory, as well as the spreading out of the calcifications and moving these throughout the volume. It is clear, therefore, that Monte Carlo simulations afford an insight that would be difficult to obtain through traditional experiment means. However, it would be beneficial to have some measure of what the simulated detected

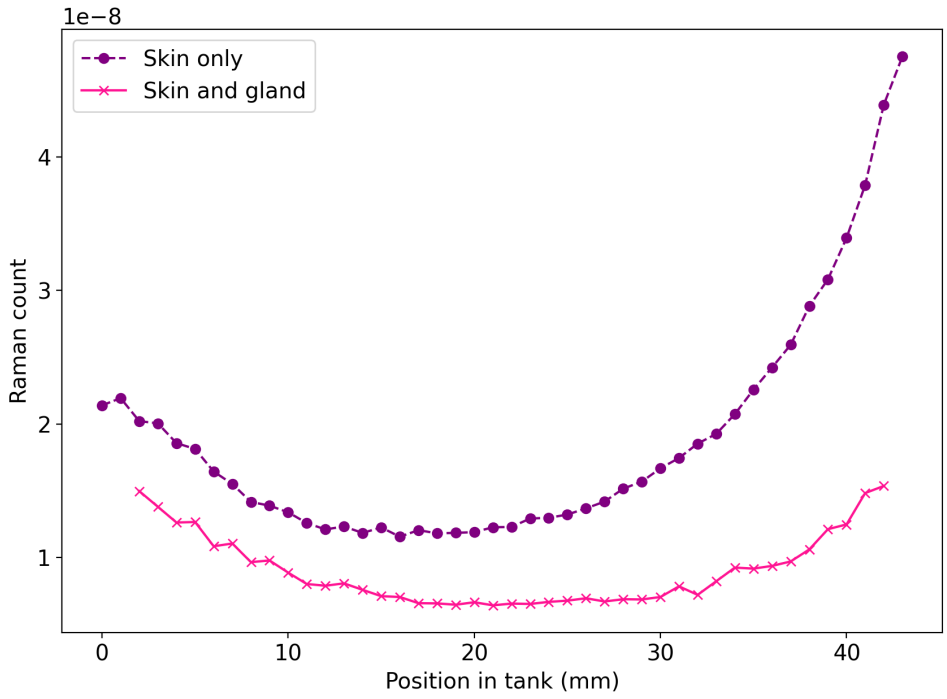


Figure 5.15: Plot comparing the results of a 3-layer skin model around bulk average breast tissue compared to the simulation with a 3-layer skin, fatty tissue and glandular tissue model.

Raman signal means in real-world terms. In order to get an estimate of where a detectable signal might be found, we again compare simulation output to laboratory results.

5.4.1 Laboratory work

Jennifer Haskell, a PhD student at the University of Exeter, performed the following experimental work in the lab and the data analysis on the results from the experiment. In order to get a volume similar to those simulated previously, based on an average breast volume, a large amount of bacon was shaped into a 8 cm × 5 cm × 4.5 cm, where the 4.5 cm dimension was aligned with the optical axis. This slab was made up of layers approximately 0.225 cm thick each. A 50 mg sample of HAP powder was wrapped in clingfilm in a ~4 mm diameter sphere and moved through each layer. The input laser had a power of 2 W and a 1 cm diameter, at 808 nm. The detector was a circular aperture of approximately 3 mm diameter. The Raman spectra were taken with 10×10 second acquisitions at each step, and then the maximum of the main HAP peak (at 960 cm^{-1}) taken and plotted against the depth of HAP in the bacon slab. The results of this experiment can be seen in

figure 5.16, where the errors were calculated according to the Andor CCD camera manual: noise = $\sqrt{2.5 \times count} \times 1.41$. The experiment could have been repeated several times to assess the standard deviation, however due to time using the CCD noise (as done in chapter 4) was deemed sufficient.

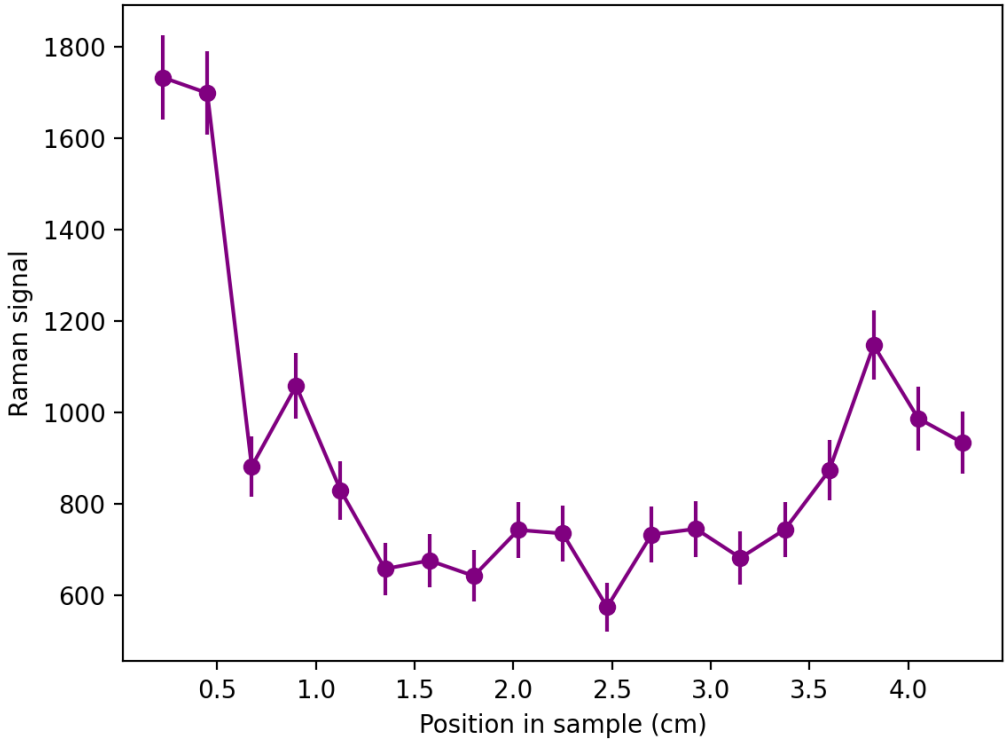


Figure 5.16: The result of moving a 4 mm sphere of 50 mg of HAP powder through a volume of bacon on the optical axis. The laser and detector were in a transmission Raman spectroscopy set-up and the acquisition 10× 10 seconds for each point.

Figure 5.16 returns the "u"-shape that is expected. The data is quite noisy but this is to be expected due to variation in the layers of the bacon used. Moreover, this work was completed with a small amount of HAP, in comparison to other studies which tend to use a larger mass. The small amount was chosen in order to help estimate a limit of detection, and be as close to the earlier simulation work (with only 10 microcalcifications) as possible. The next stage in this estimation was to replicate this experiment in a simulation.

5.4.2 Simulation work

A grid of the same dimensions (8 cm × 5 cm × 4.5 cm) was created, and a 2W, 1 cm diameter laser beam incident on the sample. Taking the average size of a microcalcification to be

0.5 mm, this means there were approximately 500 microcalcifications inside the 4 mm diameter sphere of HAP used in the experiment, by volume of spheres calculation. We therefore generated 500 points inside a 4 mm diameter sphere to use as the Raman source in the simulation. A discussion of how physically accurate this mass of HAP is compared to what is seen in the real world can be found in the next section, 5.4.3. The detector was set up to be a 3 mm diameter circular aperture, as in the real world experiment. The centre of the sphere was then moved in 0.225 cm steps through the volume from laser to detector along the optical axis. The optical properties used were those found in Mosca et al.[25] for pork muscle. The results of this can be seen in figure 5.17.

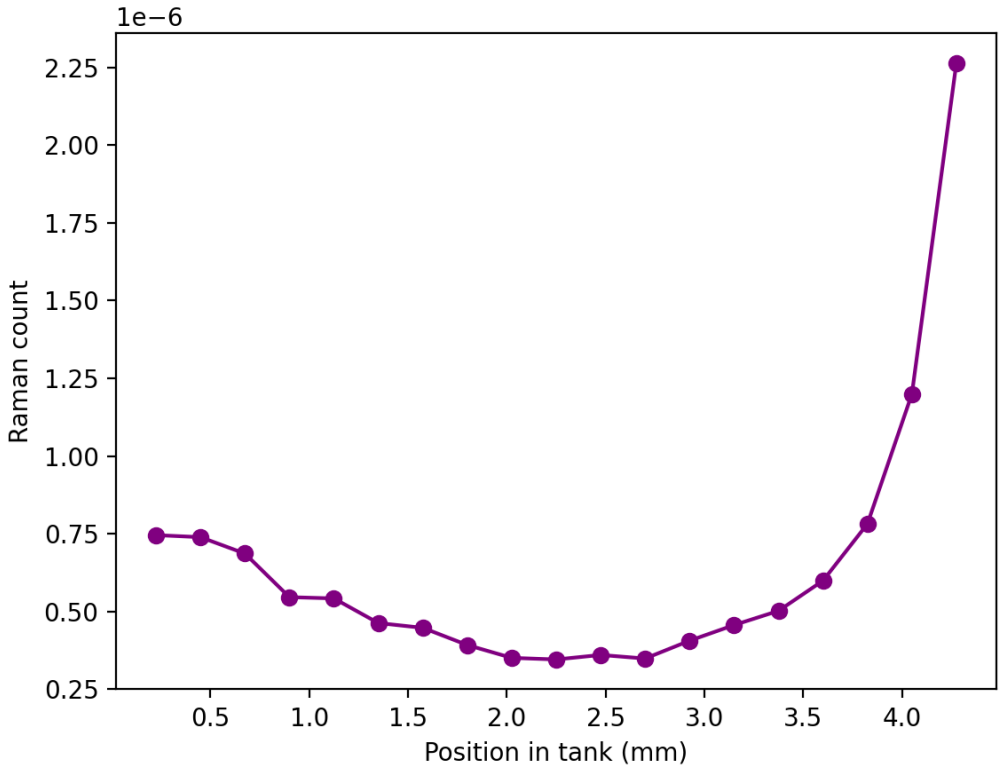


Figure 5.17: The result of simulating a 4 mm sphere of 500 microcalcifications through a volume of bacon on the optical axis. The laser and detector were in a transmission Raman spectroscopy set-up as in the experiment.

The same "u-shape" is apparent here, however the asymmetry of the curve is tipped the other way with higher signal in front of the detector. This could be down to the optical properties as bacon could be different to pork muscle, however we were limited by the literature values available. The difference in absorption between the laser wavelength photons and the Raman photons would affect this balance. However, the estimation of

the limit of detection is focused on the lowest signal and so we can compare the troughs of these plots.

5.4.3 Comparison

To compare the troughs of figures 5.16 and 5.17, the average value from depths of 1.8 – 2.7 mm (five data points) were calculated. The experimental result was 685 ± 75, which became 1710 ± 190 when converted to photons; the simulation result was (3.6 ± 0.19) × 10^{-7}. The errors in these values was taken as the standard deviation in the values used for the trough.

Since we want to estimate the limit of detection, and the HAP was detected without issue in the volume used in the experiment (50 mg in a 4 mm diameter sphere), Haskell created a plot of the Raman HAP count against the signal-to-noise ratio (SNR), as is seen in figure 5.18.

Using figure 5.18, we can find that for an SNR of 2 we would expect Raman HAP counts of 325 (or 812.5 photons). By using the ratios of the mean values of the troughs in the previous sections, and this estimate for limit of detection from SNR extrapolation, we can find an estimation for the limit of detection in the units of the simulation output, and see detection zones on the maps created earlier. The experimental numbers are on the left and the simulation numbers are on the right for the following calculation, with the standard deviation on the respective means being quoted as the initial error and then propagated through.

$$1710(\pm 190) : 3.60 \times 10^{-7}(\pm 0.19 \times 10^{-7})$$

$$1 : 2.1 \times 10^{-10}$$

$$810(\pm 90) : 1.71 \times 10^{-7}(\pm 0.09 \times 10^{-7})$$

This comparison gives the result that the limit of detection is **(1.71 ± 0.09)** $\times$**10^{-7}** in

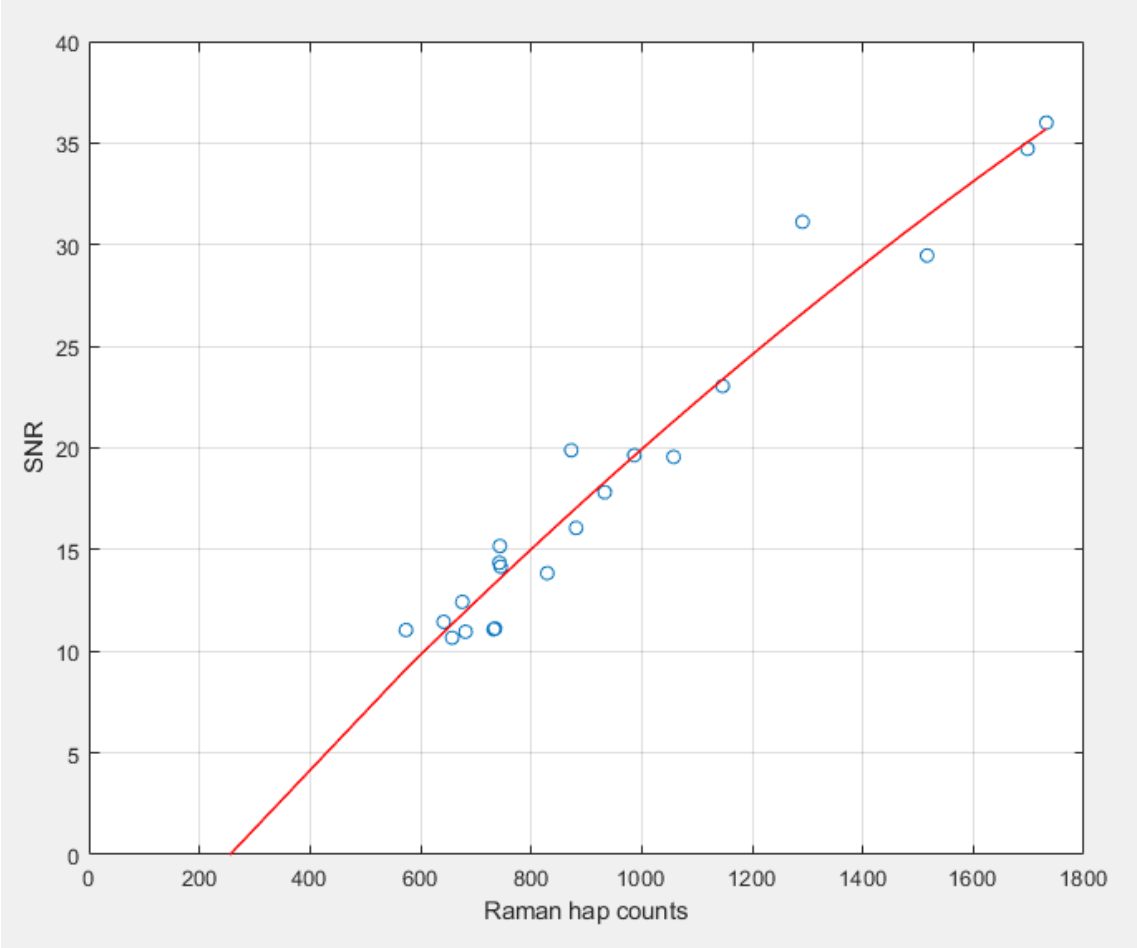


Figure 5.18: Plot made by Jennifer Haskell from her laboratory results of the Raman HAP counts against the signal to noise ratio, with a red line fitted through the data to allow extrapolation to lower SNR than the experiment afforded.

simulation output units. The detector in the simulation counts the weight of the detected packet multiplied by the power of the detected packet. This conversion factor takes into account all of the differences between the real world set-up and the simulation such as the optics, signal losses, edge effects.

Returning to the maps created from the simulation work earlier, we can apply this limit to see where we would expect to detect microcalcifications. The maps showed us how the signal varies as the Raman source moves along and perpendicular to the optical axis but were entirely in arbitrary simulation output units; now we can have an estimate of where calcifications need to be in order to be physically detected by a real world system. Figure 5.19 takes a map from earlier done in a pork tissue volume with a 1 mm sphere containing 10 microcalcifications.

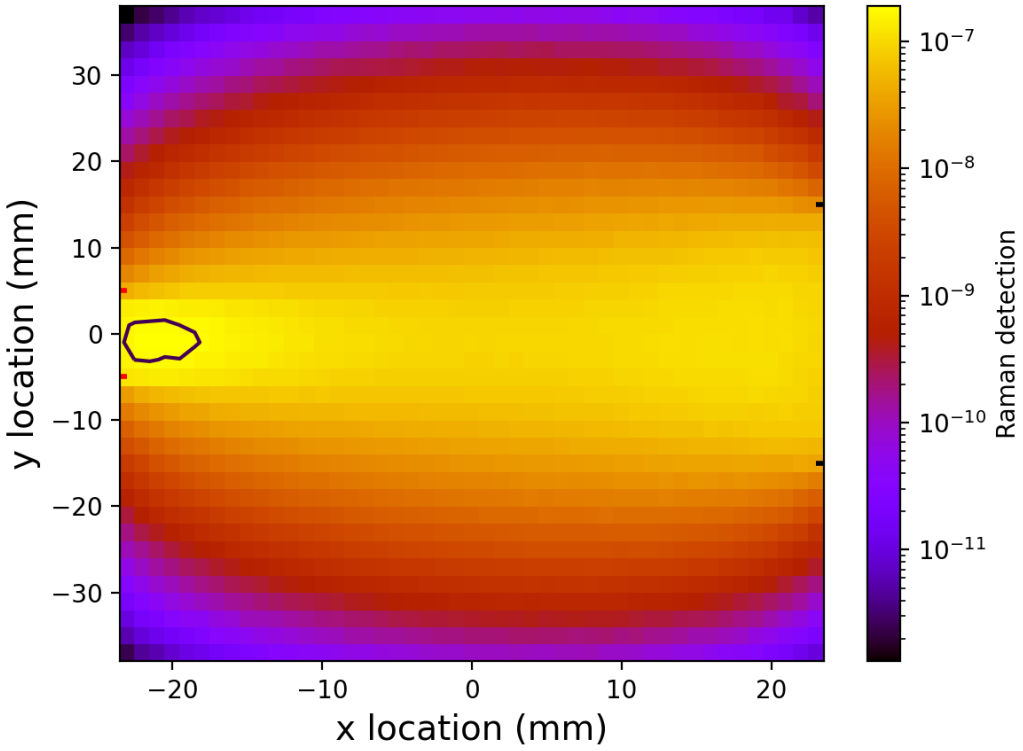


Figure 5.19: Map of Raman signal for a 1 mm sphere with 10 microcalcifications in it as it is moved through a volume of pork tissue. The input beam was a 1 W laser, 1 cm in diameter, with a circular detector opposite with a 4 cm diameter. The contour shows the region where the calcifications need to be in order to be detected.

The simulation for figure 5.19 differs from the experiment as it used a 1 W laser, and far fewer microcalcifications. This has resulted in requiring a large detector in order to have any zones where the microcalcifications could produce enough signal to be detected in a real world set-up. The estimation of the limit of detection is transferable to these other systems and so this result holds. It is to be expected, as these theoretical maps are based on having 50× fewer microcalcifications than the experiment. In reality, there could be a mass of HAP between our theoretical estimate here of 10 microcalcification point sources, and the experiment mass of 50 mg present so this contour plot represents a "worst-case scenario" or the minimum zone we could expect.

The mass of HAP used in the experimental work, 50 mg, is similar to what would be expected from a 10 mm tumour buried in 4 cm of breast tissue. From the Precision project, there was estimated to be approximately 3.3% HAP in samples selected from lesions. Using a sampling volume of ~3142 mm^3 with a lesion volume of 524 mm^3, this 3.3% HAP

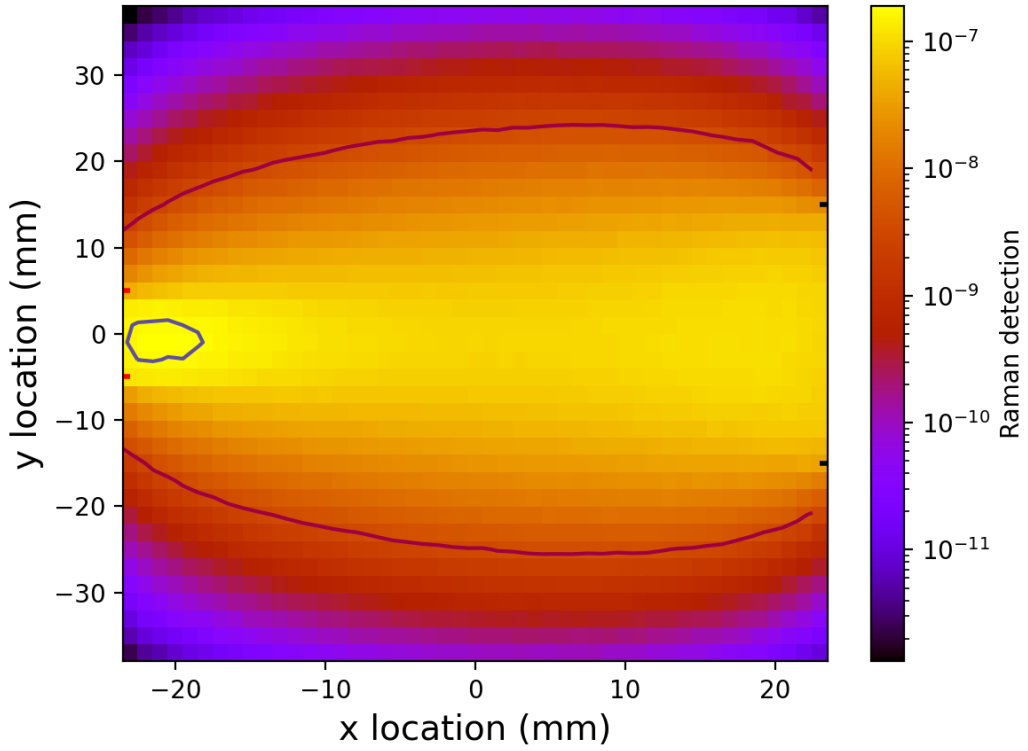


Figure 5.20: Map of Raman signal for a 1 mm sphere with 10 microcalcifications in it as it is moved through a volume of pork tissue. The input beam was a 1 W laser, 1 cm in diameter, with a circular detector opposite with a 4 cm diameter. The smaller contour shows the region where the group of 10 calcifications need to be in order to be detected; the larger contour demonstrates the region where a group of 500 calcifications would need to be in order to be detected.

gives a 17.3 mm^3 volume of microcalcification. The density of HAP is 3.2 $mg \cdot mm^{-3}$: thus there could be around 55 mg of HAP for a lesion of 10 mm diameter. This discussion and estimation is from private communications with Professor Nick Stone. Therefore, it is not unrealistic to consider the detection zone for this volume of calcification. We added this to the previous contour plot by reducing the limit of detection by a factor of 50, equivalent here to increasing the number of calcifications to 500. This was performed assuming that increasing the calcifications in a region will linearly increase the signal contribution, which is in accordance with simulation outputs.

Figure 5.20 shows two contours: a smaller one as seen earlier for a group of 10 calcifications, and a broader contour for a group of 500 calcifications. The smaller contour can be considered a minimum zone of detectability; the larger contour shows where a mass of HAP equivalent to those seen around 10 mm tumours could be detected. These results are

promising, showing that for larger volumes of HAP detection is possible off-axis. However, this map is for bulk pork tissue and for all of the HAP volume being contained within a 1 mm sphere. In reality, the HAP is likely to spread across a broader area which will lower the peak signal detected but show shallower decreases as calcifications move off-axis and towards the centre of the volume. Physiological features would have an impact on signal strength and the locations of peak signal, as demonstrated earlier in this chapter.

5.5 Conclusion

In this chapter, work was done to investigate how using point sources to replicate microcalcifications affect the detected Raman signal in a variety of optical geometries and physical environments. Previous work has characterised the signal profile on-axis in bulk tissue phantoms; we were interested in how the signal varies as a finite Raman source was moved off the optical axis. Using MCRT, we were also able to look into a variety of laser and detector geometries, and the impact of having calcifications clustered close together or further apart.

We began with the simplest scenario: a volume of pork tissue (optical properties taken from Mosca et al.[25]) similar to breast tissue. We then generated 10 microcalcifications (inside a sphere of 1 mm diameter) represented by point sources, as the calcifications are small on the scale of the volume we were simulating them in. These calcifications were moved throughout the pork volume and the Raman signal detected through the front and back faces were collected. This allowed different detectors to be imposed in post-processing, for both TRS and SORS geometries. In TRS, increasing the detector diameter increased the peak signal but in the real world this would be impacted by increased noise. The SORS results showed deeper light penetration with increasing the source-detector separation, and a higher peak signal than TRS, but SORS is unable to sample the entire volume.

Pork is often used in experiments as a breast tissue phantom material, so it was of interest to investigate how the results varied between the two. Two identical simulations were run, where only the optical properties were changed to reflect either pork or bulk average breast tissue. The 2D maps and the 1D plot along the optical axis were compared.

This demonstrated that, overall, pork tissue is a good approximation for breast tissue in these volumes, such as a depth along the optical axis of 5 cm.

Next, the bulk average breast tissue optical properties were sampled by TRS and SORS, in a similar fashion to the pork simulations. The investigation here focused on the difference between having 10 randomly generated calcifications contained within a 1 mm sphere, or a 10 mm sphere. The laser and detector diameters were kept the same in order to compare only the effect of spreading the calcifications out, in both optical geometry set-ups. The larger sphere containing the point sources resulted in a lower peak signal, due to fewer calcifications being in the laser "hot spots" and not being able to move the distribution as close to any of the edges.

Due to the real-world applications of this work, it was deemed appropriate to find literature values for physical breast features such as skin, adipose and glandular tissue and incorporate these into the models. A 3-layer skin model was added first, at the laser face and detector face. This still returned the expected "u-shape" along the optical axis. Then, the bulk average breast tissue was changed for adipose tissue, and a sphere of gland optical properties around the microcalcification point sources was generated. This was in order to see whether having Raman photons from the calcifications released into the gland material would help or hinder signal collection. The peak signal in this simulation was around four times lower than bulk breast with just skin, however the signal in front of the laser and in front of the detector were a lot more even. This implied that modelling the calcifications inside the gland material improved the relative signal from the laser side of the volume.

Finally, work was done to estimate how the simulation output units of detected Raman signal could be compared to real-world results. Jennifer Haskell performed an experiment with HAP powder and a large bacon volume, which was replicated in an MCRT simulation. By comparing the values of the output from each, a conversion factor was estimated for a limit of detection for HAP in a large bacon volume. The 2D plots made for the pork volume at the beginning of the chapter had contours added to show a worst-case and best-case scenario for volumes of HAP. These contours showed that for tiny volumes of HAP, the detectable region is small and directly in front of the laser. However,

for larger volumes of HAP, it could be possible to detect all along the optical axis and for some distance off-axis as well.

From the work to estimate the regions of detection for microcalcifications, it could be possible to identify a microcalcification up to 2 cm away from the optical axis. It could be optimal to implement some scanning of the breast tissue, were the Raman probe to be used in conjunction with mammography. By taking several measurements across the compressed breast, the entire bulk volume of the breast tissue could be sampled and the data analysed to check if type II calcifications are present.

For implementation of Raman spectroscopy into the clinical pathway, there are many considerations to be made. Current work with excised tissue is done with long integration times or higher powered lasers which could pose a safety risk for *in vivo* work. Raman spectroscopy is considered non-destructive but long term effects of repeated tissue exposure to laser radiation is not well quantified. Use of near infrared light is only expected to lead to thermal effects at high power densities; lower wavelengths could be more dangerous hence they are avoided for this application. Moreover, the clinical importance of this method of detection should be well examined: it must exceed the current routine and be cost-saving in order to be worthwhile implementing broadly. Finally, Raman spectroscopy is not well known to clinicians and so in order to develop a real-time *in vivo* tool, there would need to be consideration for easily understandable output and automated data analysis.

Overall, this chapter has demonstrated the power of MCRT simulations in showing signal strength on and off the optical axis. A variety of optical geometries can be modelled with ease and the Raman source can be easily moved around throughout the volume. These simulations can be used to predict useful experimental work and also show insight into what results the experiments should show. We have shown that pork is a useful breast tissue phantom, and that including more physically accurate breast anatomy can change the shape of the Raman signal curves.

Chapter 6

The Conclusion

6.1 Concluding remarks

The aim of this thesis was to develop a Monte Carlo code capable of simulating Raman scattering in heterogeneous media, and to run simulations to aid in the development of using Raman spectroscopy to improve breast cancer detection. We have managed to explore how moving a Raman source on and off the optical axis impacts the detected signal, as well as modelling a variety of input laser beam widths, detector sizes, and optical properties.

This thesis began by framing the problem: a large number of biopsies for breast cancer diagnosis are carried out and the majority are benign. By implementing a non-invasive diagnostic tool, this cost and stress could be reduced. Microcalcifications in the breast tissue are a good indicator of malignancy and Monte Carlo modelling is an effective method of simulating light transport in biological tissue. The history of Monte Carlo modelling and the numerical methods required to develop an algorithm for our purposes were discussed in detail.

The first step in developing a model for predictive purposes is to verify that our tool is as good as other published efforts. We chose to compare results with a paper by Pavel Matousek et al. where we discovered that optical absorption and absorbance are not defined the same way mathematically speaking, and that the Matousek code was not scattering isotropically. Once these corrections were applied, we verified that both codes

return the same results, and showed the benefits of spatially offset Raman spectroscopy where deeper Raman photons can be better collected by a wider detector-illumination separation.

The next stage was to check that our code returned the same results as an experiment. Lab work was performed to investigate how the signal from a "semi-infinite" slab of Raman scattering material moved along the optical axis in different concentrations of Intralipid varied. We saw that the maximum Raman signal in this situation is when the slab is in the centre of the volume. The same set-up was created in the `arctk` code and the results examined. In order to run the simulations correctly, the optical properties of Intralipid need to be known. Previous work in the field had assumed a linear extrapolation from concentrations of 10% down to lower concentrations (to simulate biological tissue optical properties), however we found that this assumption did not hold. By running a suite of Monte Carlo models, we found better estimates for the scattering coefficient of Intralipid at low concentrations ($< 4\%$). The outputs from both the Monte Carlo simulations and the lab work overlapped well, showing that the code was capable of replicating experimental output. This gave us the confidence to move on to more theoretical investigative work.

Our predictive work began with pork tissue volumes, as pork is often used as a breast tissue phantom in experimental work in this field. We simulated a volume similar to that of an average human breast and generated a random distribution of microcalcifications within a small sphere. We then moved this distribution throughout the volume in a "mapping" fashion. We modelled a variety of detectors to see how the signal for both TRS and SORS geometries changed. This showed the expected results of TRS being better for sampling the bulk volume but SORS detecting stronger signals. Then, we compared the results of pork to those of breast tissue, based on literature optical properties. This showed that pork is a good approximation for breast tissue in large tissue volume samples such as those modelled here, despite the difference in the optical property values.

We then moved on to examine how bulk average breast tissue maps would look, with microcalcifications randomly generated in two different sized spheres. This allowed us to see how the spread of calcifications impacts the detected signal, again in both SORS

and TRS geometries. The further apart the calcifications were distributed, the lower the peak signal was. These simulations to generate and map calcifications were new to the field, and would be extremely difficult to perform in a laboratory setting, thus giving new insight into how the Raman signal propagates at depth in breast tissue.

To create a more realistic model, we added skin, and then exchanged bulk average breast tissue for glandular and adipose tissues. This work was carried out in TRS only as that is the preferred geometry for sampling the entire volume. We saw that adding the more realistic physiological features reduced the maximum detected signal and changed the location where this maximum signal could be detected.

Finally, we moved to estimate how the output of the MCRT simulations could be tied to real-world results. Experiments carried out by another PhD student in the group to move a small sphere of HAP powder through a large volume of bacon were replicated in the `arctk` tool by me, and the results compared. By finding a rough limit of detection through comparing these outputs, contours were added to the previous mapping in pork work to gauge a reasonable detection zone where calcifications would need to be in order to be detectable by TRS. We saw that the detection zone varies depending on the mass of calcification present and that for typical masses of HAP present in real lesions, it should be possible to detect calcifications throughout the volume on the optical axis, and for up to 2 cm off axis, given a large enough detector diameter. This implies that to sufficiently sample an entire breast, a raster map or several illumination sources would be required to cover the full volume.

The results from the predictive mapping work show that TRS is the best optical set-up for sampling the entire breast volume, and that moving the work to be non-invasive will significantly reduce the signal collected given the effects of skin and other breast anatomy on photon transport. By increasing the laser power and the detector radius, or using a needle probe to bypass the impact of skin, could allow Raman spectroscopy to become an important part of the breast cancer diagnostic pathway.

We conclude that we have successfully developed a tool capable of predicting Raman photon transport through homogeneous tissue phantoms and heterogeneous breast

tissue. Our results have shown that the choice of input beam width and detector geometry impacts how the volume is sampled and that the optical properties are extremely important. The flexibility of `arctk` and the simplicity to install make it a useful tool in continuing this work, and for other scenarios where light transport simulation is required.

The work in this thesis is the first of its kind to numerically examine the Raman signal from calcifications at depth in tissue phantoms and breast tissue. We have used our Monte Carlo code as a predictive tool for the optical properties in the most dilute concentrations of Intralipid which have not been measured in the literature. We have then expanded to solid tissue phantoms and found important insights into how laser and Raman light propagate through the volume. This work has aided in learning how best to exploit this to allow Raman spectroscopy to become a non-invasive tool for the diagnosis of breast cancer.

Raman spectroscopy has a variety of applications, and verifying this Monte Carlo tool means that it can now be used for more of this work. In addition to building on the breast cancer detection work, other cancers and other types of Raman spectroscopy could benefit from predictive numerical modelling simulations. Some interesting directions for future work to go in are described in the next, and final, section.

6.2 Future work

The verification of `arctk` and the predictive work completed in this thesis are a strong beginning to showing the potential of using numerical modelling in applying Raman spectroscopy to biological tissue. Some further possibilities are outlined here, all of which are interesting and beneficial.

One of the big discrepancies between the tissue phantoms and simulation volumes used throughout this work and the reality of integrating Raman spectroscopy into the mammogram stage of the diagnostic pathway is the shape of the breast. Our work was done on rectangular volumes, whereas the human female breast has curved edges. Implementing more realistic boundaries on the simulation volume could lead to packets remaining inside the volume for longer due to refraction effects, as seen in Tran et al.[171]

Additionally, it would be interesting to expand on the rudimentary anatomy explored in the previous chapter with a model closer to the reality: a branching ductal system and perhaps the addition of a lesion and its optical properties. By creating a more realistic breast anatomy with a lesion and calcifications as seen in real-world breast cancer samples, a more accurate representation of Raman light transport in these environments can be obtained.

Microcalcifications present in the prostate can be a biomarker for cancer, and thus Raman spectroscopy could be used to identify these in a similar fashion to those in breast tissue. Transrectal probes are in use currently for investigating prostates and so it is feasible that a fibre-optic Raman probe could also be incorporated or created. Prostate cancer is the most common cancer to be diagnosed in men in the UK, and there are around 12,000 deaths every year from this disease.[172] Raman spectroscopy could be useful for probing this organ due to the existing probes and the size of the organ, at around 30–50 mm across.[173]

An advantage of using modelling software for medical research is the ability to quickly run many different scenarios, in a "virtual laboratory". In this way, a variety of incident beams and detector widths and geometries can be tested far more quickly than in a physical experiment, and easily change the locations and geometry of the Raman source distributions. The work in this thesis has begun to explore these parameter spaces, but more investigations could be done to fully investigate the impact of changing light sources. For example, is it better to use a more powerful laser with a broader beam, compared to raster scanning with a smaller and less powerful beam? How does calcification distribution and/or volume impact the illumination and detection choice? These simulation results can be used to find out which set-ups could be the most effective for detection of calcifications at depth, informing future experimental work.

More broadly, Monte Carlo modelling by the `arctk` library could be applied to a large variety of further medical uses. One example is using MRI images to base models on, then performing Monte Carlo simulations in those models; this has been done for brain scans to investigate the benefit of patient specific simulations.[174] Another useful application of this codebase would be to add the physics required for proton beam therapy

to be modelled. Useful calculations such as the dose delivered into tissues under different circumstances could be performed.[175] One advantage of `arctk` is how easily new engines and library sections can be added to the Rust crate.

The tool developed in this thesis has previously been used to model how nanoparticles affect heat deposition in photothermal skin cancer treatment.[115] There is also a Raman spectroscopy technique called Surface Enhanced Raman Spectroscopy (SERS)[176] which uses excitation of surface plasmon resonance of nanoparticles to enhance the Raman signal of molecules in the vicinity of the surface. SERS nanoparticles can be specifically optimised for particular biomarkers. This has been used for improving the penetration depth of SORS and can be exploited to determine the depth of an inclusion in a tissue phantom.[177] By incorporating this effect into the existing code, it is possible to model the impact of SESORS (the combination of SERS and SORS) on detection zones and optimise the laser-detector offset.

Another phenomenon that would be interesting to model is the observed differences in how the distribution of the Raman material affects the detected signal. It has been noted (in private communications with Professors Nick Stone and Pavel Matousek) that having the same mass of HAP powder in one packed volume and then split into several smaller volumes do not result in the same amount of Raman signal. It appears that the one densely packed volume creates some kind of "ringing" or trapping effect which generates more Raman photons than expected due to laser photons scattering inside the HAP for greater path lengths. Investigating this further and implementing the effect of this into simulations could give greater insight into how the distribution of the calcifications affects the detected signals.

The purpose of this thesis was to develop, benchmark, and use a code capable of simulating Raman spectroscopy in breast tissue. The codebase `arctk` that was developed is novel in its ability to model arbitrary geometries, giving rise to more realistic biological media than in previous works that have relied on plane parallel layers. By combining data from experiments and output from simulations, better estimation of optical properties in tissue phantoms was possible. The work on modelling microcalcifications at depth in breast tissue allowed a number of conclusions to be drawn. Firstly, TRS is the best method

for sampling the entire bulk tissue volume, including for finite calcifications (as opposed to the larger inclusions that are possible for laboratory work). Also, further simulations showed that skin has a huge impact on the signal strength for non-invasive detection of these structures, and including different optical properties for internal anatomical features also affected the signal. This highlights the importance of having heterogeneous, realistic models in this work. These are difficult to achieve in a laboratory, so Monte Carlo simulations are key for testing such scenarios. Finally, by comparing results from experimental work in a bulk tissue volume with Monte Carlo simulations of the same set-up, it was determined that the most likely detection zone for HAP in tissue is in the region nearest to the input laser beam. This work involving finite calcification Raman scattering material in the Monte Carlo simulations is the first of its kind and has provided invaluable insight.

Appendix : Output from PhD

Papers

- "An experimental and numerical modelling investigation of the optical properties of Intralipid using deep Raman spectroscopy" published in the Royal Society of Chemistry journal *Analyst*: Laura J. Moran, Freddy Wordingham, Benjamin Gardner, Nicholas Stone, and Tim J. Harries. An experimental and numerical modelling investigation of the optical properties of Intralipid using deep Raman spectroscopy. Analyst, 146:7601–7610, 2021.

- "Numerical simulations of SORS using an updated Monte Carlo routine" in preparation, based on the work in Chapter 3 of this thesis.

- "Numerical simulations of Raman scattering from microcalcifications in breast tissue" in preparation, based on the work in Chapter 5 of this thesis.

Key Achievements

- Won the Martin & Willis Prize for Best Oral Presentation from IRDG, September 2019.

- As part of the M&W prize, gave a presentation at the IRDG Christmas international meeting, December 2019.

- Won the University of Exeter EMPS Conference 2019 prize for best talk.

- Spoke at Pint of Science Exeter in 2018, and organised a three day Pint of Science event in 2019.

- Taught physics and maths modules at the University of Exeter to second year undergraduates, and was the lead demonstrator for two years.

Meetings attended

- University of St Andrews Monte Carlo Summer School, August 2017
- CREAM-1, Queen's University Belfast, April 2018
- University of Exeter EMPS Conference, January 2019
- IRDG 224, University of Lancaster, September 2019
- IRDG Christmas meeting, UCL, December 2019

Bibliography

[1] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, n/a(n/a).

[2] Cancer Research UK. CRUK: www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/mortalityheading-Two, Accessed November 2021.

[3] Encyclopedia Brittanica. Mammary Gland: https://www.britannica.com/science/mammary-gland, Accessed November 2021.

[4] Cancer Research UK. CRUK: https://www.cancerresearchuk.org/about-cancer/breast-cancer/about, Accessed May 2022.

[5] L. Frappart, M. Boudeulle, J. Boumendil, Hu Chi Lin, I. Martinon, C. Palayer, Y. Mallet-Guy, D. Raudrant, A. Bremond, Y. Rochet, and J. Feroldi. Structure and composition of microcalcifications in benign and malignant lesions of the breast: Study by light microscopy, transmission and scanning electron microscopy, microprobe analysis, and X-ray diffraction. *Human Pathology*, 15(9):880–889, 1984.

[6] Roberta De Angelis, Milena Sant, Michel P Coleman, Silvia Francisci, Paolo Baili, Daniela Pierannunzio, Annalisa Trama, Otto Visser, Hermann Brenner, Eva Ardanaz, Magdalena Bielska-Lasota, Gerda Engholm, Alice Nennecke, Sabine Siesling, Franco Berrino, and Riccardo Capocaccia. Cancer survival in Europe 1999–2007 by country

and age: results of EUROCARE-5—a population-based study. *The Lancet Oncology*, 15(1):23–34, 2014.

[7] Blausen Medical. Medical gallery of blausen medical 2014. *WikiJournal of Medicine*, 1(2):1–79, 2014.

[8] Mokbel K. Hindle, W. Diagnosis and Management of Benign Breast Diseases. *Glob. Libr. Women's Med.*, 2009.

[9] S.J. Withrow, D.M. Vail, and R.L. Page. *Withrow and MacEwen's Small Animal Clinical Oncology - E-Book*. Elsevier Health Sciences, 2013.

[10] Matthew J. Baker, Shawn R. Hussain, Lila Lovergne, Valérie Untereiner, Caryn Hughes, Roman A. Lukaszewski, Gérard Thiéfin, and Ganesh D. Sockalingum. Developing and understanding biofluid vibrational spectroscopy: a critical review. *Chem. Soc. Rev.*, 45:1803–1818, 2016.

[11] Eduarda Mendes and Noélia Duarte. Mid-Infrared Spectroscopy as a Valuable Tool to Tackle Food Analysis: A Literature Review on Coffee, Dairies, Honey, Olive Oil and Wine. *Foods*, 10(2), 2021.

[12] B. T. Cox. *Introduction to Laser-Tissue Interactions*. Optics in Medicine, 2007.

[13] D. B. Karl Sandeman. Dispersive Raman Spectroscopy. DoITPoMS, 2016.

[14] P. Matousek, N. Everall, M. Towrie, and A. W. Parker. Depth Profiling in Diffusely Scattering Media Using Raman Spectroscopy and Picosecond Kerr Gating. *Applied Spectroscopy*, 59(2):200–205, 2005. PMID: 15720761.

[15] Pavel Matousek and Nicholas Stone. Emerging concepts in deep Raman spectroscopy of biological tissue. *Analyst*, 134:1058–1066, 2009.

[16] Kenry, Yukun Duan, and Bin Liu. Recent Advances of Optical Imaging in the Second Near-Infrared Window. *Advanced Materials*, 30(47):1802394, 2018.

[17] F. Coccetti. The Fermiac or Fermi's Trolley. *Nuovo Cimento C (Online)*, 39(2):1–8, 2016.

[18] S. Chandrasekhar. *Radiative Transfer*. Dover Books on Physics. Dover Publications, 2013.

[19] L.V. Wang and H. Wu. *Biomedical Optics: Principles and Imaging*. Wiley, 2012.

[20] G. Yoon, A. Welch, M. Motamedi, and M.v. Gemert. Development and application of three-dimensional light distribution model for laser irradiated tissue. *IEEE Journal of Quantum Electronics*, 23(10):1721–1733, 1987.

[21] P. Matousek, M. D. Morris, N. Everall, I. P. Clark, M. Towrie, E. Draper, A. Goodship, and A. W. Parker. Numerical Simulations of Subsurface Probing in Diffusely Scattering Media Using Spatially Offset Raman Spectroscopy. *Appl. Spectrosc.*, 59(12):1485–1492, Dec 2005.

[22] Martha Z. Vardaki, Benjamin Gardner, Nicholas Stone, and Pavel Matousek. Studying the distribution of deep Raman spectroscopy signals using liquid tissue phantoms with varying optical properties. *Analyst*, 140:5112–5119, 2015.

[23] Sergio Fantini, Scott A. Walker, Maria Angela Franceschini, Michael Kaschke, Peter M. Schlag, and K. Thomas Moesta. Assessment of the size, position, and optical properties of breast tumors in vivo by noninvasive optical methods. *Appl. Opt.*, 37(10):1982–1989, Apr 1998.

[24] Steven L Jacques. Optical properties of biological tissues: a review. *Physics in Medicine and Biology*, 58(11):R37–R61, may 2013.

[25] Sara Mosca, Pranav Lanka, Nick Stone, Sanathana Konugolu Venkata Sekar, Pavel Matousek, Gianluca Valentini, and Antonio Pifferi. Optical characterization of porcine tissues from various organs in the 650-1100nm range using time-domain diffuse spectroscopy. *Biomed. Opt. Express*, 11(3):1697–1706, Mar 2020.

[26] Lorenzo Spinelli, Alessandro Torricelli, Antonio Pifferi, Paola Taroni, Gian Maria Danesini, and Rinaldo Cubeddu. Bulk optical properties and tissue components in the female breast from multiwavelength time-resolved optical mammography. *Journal of biomedical optics*, 9(6):1137—1142, 2004.

[27] A. Bhandari, B. Hamre, Ø. Frette, K. Stamnes, and J. J. Stamnes. Modeling optical properties of human skin using Mie theory for particles with different size distributions and refractive indices. *Opt. Express*, 19(15):14549–14567, Jul 2011.

[28] V G Peters, D R Wyman, M S Patterson, and G L Frank. Optical properties of normal and diseased human breast tissues in the visible and near infrared. *Physics in Medicine and Biology*, 35(9):1317–1334, sep 1990.

[29] C R Smittenaar, K A Petersen, K Stewart, and N Moitt. Cancer incidence and mortality projections in the UK until 2035. *Br J Cancer*, 115(9):1147–1155, Oct 2016.

[30] Cancer Research UK.

[31] C J Stein and G A Colditz. Modifiable risk factors for cancer. *British Journal of Cancer*, 90(2):299–303, 2004.

[32] Katrina F Brown, Harriet Rumgay, Casey Dunlop, Margaret Ryan, Frances Quartly, Alison Cox, Andrew Deas, Lucy Elliss-Brookes, Anna Gavin, Luke Hounsome, et al. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. *British Journal of Cancer*, 118(8):1130–1141, 2018.

[33] P Lichtenstein, N V Holm, P K Verkasalo, A Iliadou, J Kaprio, M Koskenvuo, E Pukkala, A Skytthe, and K Hemminki. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*, 343(2):78–85, Jul 2000.

[34] Cancer Collaborative Group on Hormonal Factors in Breast. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet*, 358(9291):1389–1399, Oct 2001.

[35] Bengt Zöller, Xinjun Li, Jan Sundquist, and Kristina Sundquist. Familial transmission of prostate, breast and colorectal cancer in adoptees is related to cancer in biological

but not in adoptive parents: a nationwide family study. *Eur J Cancer*, 50(13):2319–2327, Sep 2014.

[36] Nancy E. Hynes and Christine J. Watson. Mammary Gland Growth Factors: Roles in Normal Development and in Cancer. *Cold Spring Harbor Perspectives in Biology*, 2(8), 2010.

[37] A.P. Cooper. *On the Anatomy of the Breast*. Number v. 1 in On the Anatomy of the Breast. Longman, 1840.

[38] Susan M. Love and Sanford H. Barsky. Anatomy of the nipple and breast ducts revisited. *Cancer*, 101(9):1947–1957, 2004.

[39] J.N. Eble, F.A. Tavassoli, P. Devilee, World Health Organization, and International Agency for Research on Cancer. *Pathology and Genetics of Tumours of the Breast and Female Genital Organs*. Who/IARC Classification of Tumours. IARC Press, 2003.

[40] Meltem Gülsün, Figen Başaran Demirkazik, and Macit Ariyürek. Evaluation of breast microcalcifications according to Breast Imaging Reporting and Data System criteria and Le Gal's classification. *Eur J Radiol*, 47(3):227–231, Sep 2003.

[41] Solveig Hofvind, Barbro F Iversen, Liv Eriksen, Bodil M Styr, Kjell Kjellevold, and Kathinka D Kurz. Mammographic morphology and distribution of calcifications in ductal carcinoma in situ diagnosed in organized screening. *Acta Radiol*, 52(5):481–487, Jun 2011.

[42] E. Özer, T. Canda, P. Balcı, and Ö. Gökçe. Calcium Oxalate Crystals in Benign Cyst Fluid from the Breast. *Acta Cytologica*, 43(2):281–284, 1999.

[43] R Baker, K D Rogers, N Shepherd, and N Stone. New relationships between breast microcalcifications and cancer. *Br J Cancer*, 103(7):1034–1039, Sep 2010.

[44] Marleen M. Kerssens, Pavel Matousek, Keith Rogers, and Nicholas Stone. Towards a safe non-invasive method for evaluating the carbonate substitution levels of hydroxyapatite (HAP) in micro-calcifications found in breast tissue. *Analyst*, 135:3156–3161, 2010.

[45] Iain R. Gibson and William Bonfield. Novel synthesis and characterization of an AB-type carbonate-substituted hydroxyapatite. *Journal of Biomedical Materials Research*, 59(4):697–708, 2002.

[46] Abigail S Haka, Karen E Shafer-Peltier, Maryann Fitzmaurice, Joseph Crowe, Ramachandra R Dasari, and Michael S Feld. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy. *Cancer Res*, 62(18):5375–5380, Sep 2002.

[47] Willie Hamilton, Sally Stapley, Christine Campbell, Georgios Lyratzopoulos, Greg Rubin, and Richard D. Neal. For which cancers might patients benefit most from expedited symptomatic diagnosis? Construction of a ranking order by a modified Delphi technique. *BMC Cancer*, 15(1):820, 2015.

[48] World Health Organization. *Guide to cancer early diagnosis*. World Health Organization, 2017.

[49] MP Coleman, D Forman, H Bryant, J Butler, B Rachet, C Maringe, U Nur, E Tracey, M Coory, J Hatcher, CE McGahan, D Turner, L Marrett, ML Gjerstorff, TB Johannesen, J Adolfsson, M Lambe, G Lawrence, D Meechan, EJ Morris, R Middleton, J Steward, and MA Richards. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *The Lancet*, 377(9760):127–138, 2011.

[50] Arduino Verdecchia, Silvia Francisci, Hermann Brenner, Gemma Gatta, Andrea Micheli, Lucia Mangone, and Ian Kunkler. Recent cancer survival in Europe: a 2000–02 period analysis of EUROCARE-4 data. *The Lancet Oncology*, 8(9):784–796, 2007.

[51] Mike Richards. EUROCARE-4 studies bring new data on cancer survival. *The Lancet Oncology*, 8(9):752–753, 2007.

[52] Milena Sant, Claudia Allemani, Riccardo Capocaccia, Timo Hakulinen, Tiiu Aareleid, Jan Willem Coebergh, Michel P. Coleman, Pascale Grosclaude, Carmen Martinez, Janine Bell, Judith Youngson, Franco Berrino, and the EUROCARE Working Group. Stage

at diagnosis is a key explanation of differences in breast cancer survival across Europe. *International Journal of Cancer*, 106(3):416–422, 2003.

[53] S. Walters, C. Maringe, J. Butler, B. Rachet, P. Barrett-Lee, J. Bergh, J. Boyages, P. Christiansen, M. Lee, F. Wärnberg, C. Allemani, G. Engholm, T. Fornander, M L Gjerstorff, T B Johannesen, G. Lawrence, C E McGahan, R. Middleton, J. Steward, E. Tracey, D. Turner, M A Richards, M P Coleman, and The ICBP Module 1 Working Group20. Breast cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK, 2000-2007: a population-based study. *British Journal of Cancer*, 108(5):1195–1208, 2013.

[54] M A Richards. The size of the prize for earlier diagnosis of cancer in England. *British Journal of Cancer*, 101(2):S125–S129, 2009.

[55] MA Richards, AM Westcombe, SB Love, P Littlejohns, and AJ Ramirez. Influence of delay on survival in patients with breast cancer: a systematic review. *The Lancet*, 353(9159):1119–1126, 1999.

[56] Mauro Laudicella, Brendan Walsh, Elaine Burns, and Peter C Smith. Cost of care for cancer patients in England: evidence from population-based patient-level data. *British Journal of Cancer*, 114(11):1286–1292, 2016.

[57] National Institute for Health and Care Excellence. Breast Cancer Quality Standard [QS12], 2016.

[58] Cornelia J. Baines, Claus Wall Bsc, Harvey A. Risch, James K. Kuin, and Ian J. Fan. Changes in breast self-examination behavior in a cohort of 8214 women in the Canadian national breast screening study. *Cancer*, 57(6):1209–1216, 2021/09/10 1986.

[59] JP Kösters and PC Gøtzsche. Regular self-examination or clinical examination for early detection of breast cancer. *The Cochrane database of systematic reviews*, (2):CD003373, 2003.

[60] Rukaiya Malik, Numa Vera, Chandra Dayal, Abhay Choudhari, Jyotishna Mudaliar, Amanda Noovao Hill, Ilisapeci Kubuabola, and Ronny Gunnarsson. Factors associated

with breast cancer awareness and breast self-examination in Fiji and Kashmir India –a cross-sectional study. *BMC Cancer*, 20(1):1078, 2020.

[61] Tran Thu Ngan, Nga T Q Nguyen, Hoang Van Minh, Michael Donnelly, and Ciaran O'Neill. Effectiveness of clinical breast examination as a 'stand-alone' screening modality: an overview of systematic reviews. *BMC Cancer*, 20(1):1070, Nov 2020.

[62] Public Health England. Radiation risk with digital mammography in breast screening. *Research and analysis*, 2017.

[63] Patrick Forrest. Breast cancer screening: report to the Health Ministers of England, Wales, Scotland & Northern Ireland. *HM Stationery Office*, 1986.

[64] NHS England. Breast Screening Programme, England 2019-20. *NHS Digital*, 2021.

[65] Lydia E Pace and Nancy L Keating. A systematic assessment of benefits and risks to guide breast cancer screening decisions. *JAMA*, 311(13):1327–1335, Apr 2014.

[66] Solveig Hofvind, Antonio Ponti, Julietta Patnick, Nieves Ascunce, Sisse Njor, Mireille Broeders, Livia Giordano, Alfonso Frigerio, and Sven Törnberg. False-Positive Results in Mammographic Screening for Breast Cancer in Europe: A Literature Review and Survey of Service Screening Programmes. *Journal of Medical Screening*, 19(1_suppl):57–66, 2012. PMID: 22972811.

[67] P T Huynh, A M Jarolimek, and S Daye. The false-negative mammogram. *RadioGraphics*, 18(5):1137–1154, 1998. PMID: 9747612.

[68] Wendie A Berg, Zheng Zhang, Daniel Lehrer, Roberta A Jong, Etta D Pisano, Richard G Barr, Marcela Böhm-Vélez, Mary C Mahoney, W Phil Evans, Linda H Larsen, Marilyn J Morton, Ellen B Mendelson, Dione M Farria, Jean B Cormack, Helga S Marques, Amanda Adams, Nolin M Yeh, Glenna Gabrielli, and ACRIN 6666 Investigators. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA*, 307(13):1394—1404, April 2012.

[69] Farideh Gharekhanloo, Mostafa Morad Haseli, and Saadat Torabian. Value of Ultrasound in the Detection of Benign and Malignant Breast Diseases: A Diagnostic Accuracy Study. *Oman Med J*, 33(5):380–386, Sep 2018.

[70] Sudheer Gokhale. Ultrasound characterization of breast masses. *Indian J Radiol Imaging*, 19(3):242–247, Jul-Sep 2009.

[71] W Teh and A.R.M Wilson. The role of ultrasound in breast cancer screening. A consensus statement by the European Group for breast cancer screening. *European Journal of Cancer*, 34(4):449–450, 1998.

[72] Giovanna Trecate, Daniele Vergnaghi, Siranuosh Manoukian, Silvana Bergonzi, Gianfranco Scaperrotta, Monica Marchesini, Claudio Ferranti, Bernard Peissel, Gianbattista Spatti, Silvia Bohm, Alberto Conti, Claudia Costa, Manila Sporeni, Franca Podo, and Renato Musumeci. MRI in the Early Detection of Breast Cancer in Women with High Genetic Risk. *Tumori Journal*, 92(6):517–523, 2006. PMID: 17260493.

[73] Mieke Kriege, Cecile T. M. Brekelmans, Inge Marie Obdeijn, Carla Boetes, Harmine M. Zonderland, Sara H. Muller, Theo Kok, Radu A. Manoliu, A. Peter E. Besnard, Madeleine M. A. Tilanus-Linthorst, Caroline Seynaeve, Carina C. M. Bartels, Reini Kaas, Siebren Meijer, Jan C. Oosterwijk, Nicoline Hoogerbrugge, Rob A. E. M. Tollenaar, Emiel J. T. Rutgers, Harry J. de Koning, and Jan G. M. Klijn. Factors Affecting Sensitivity and Specificity of Screening Mammography and MRI in Women with an Inherited Risk for Breast Cancer. *Breast Cancer Research and Treatment*, 100(1):109–119, 2006.

[74] National Institute for Health Care and Excellence. Clinical guidance CG164. *UK*, 2019.

[75] Aziza Nassar. Core needle biopsy versus fine needle aspiration biopsy in breast—A historical perspective and opportunities in the modern era. *Diagnostic Cytopathology*, 39(5):380–388, 2011.

[76] I O Ellis, S Humphreys, M Michell, S E Pinder, C A Wells, and H D Zakhour. Best Practice No 179. *Journal of Clinical Pathology*, 57(9):897–902, 2004.

[77] Public Health England. NHS Breast Screening Programme: Clinical guidance for breast cancer screening assessment. *NHSBSP publication number 49*, 2016.

[78] Joann G. Elmore, Gary M. Longton, Patricia A. Carney, Berta M. Geller, Tracy Onega, Anna N. A. Tosteson, Heidi D. Nelson, Margaret S. Pepe, Kimberly H. Allison, Stuart J. Schnitt, Frances P. O'Malley, and Donald L. Weaver. Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens. *JAMA*, 313(11):1122–1132, 03 2015.

[79] NHS. Breast cancer treatment: https://www.nhs.uk/conditions/breast-cancer/treatment, Acessed October 2021.

[80] Cancer Research UK. CRUK: https://www.cancerresearchuk.org/about-cancer/breast-cancer/treatment, Accessed October 2021.

[81] NICE. NICE guidelines: https://bnf.nice.org.uk/treatment-summaries/breast-cancer, Accessed October 2021.

[82] D. Cozzolino, E. Corbella, and H. E. Smyth. Quality Control of Honey Using Infrared Spectroscopy: A Review. *Applied Spectroscopy Reviews*, 46(7):523–538, 2011.

[83] Radigya M. Correia, Eloilson Domingos, Vagne M. Cáo, Brenda R.F. Araujo, Sthefany Sena, Layla U. Pinheiro, André M. Fontes, Luiz Felipe M. Aquino, Ernesto C. Ferreira, Paulo R. Filgueiras, and Wanderson Romão. Portable near infrared spectroscopy applied to fuel quality control. *Talanta*, 176:26–33, 2018.

[84] Andrew V. Ewing and Sergei G. Kazarian. Infrared spectroscopy and spectroscopic imaging in forensic science. *Analyst*, 142:257–272, 2017.

[85] Christopher Scotter. Use of near infrared spectroscopy in the food industry with particular reference to its applications to on/in-line food processes. *Food Control*, 1(3):142–149, 1990.

[86] David Caffrey, Ainur Zhussupbekova, Rajani K. Vijayaraghavan, Ardak Ainabayev, Aitkazy Kaisha, Gulnar Sugurbekova, Igor V. Shvets, and Karsten Fleischer. Crystallographic Characterisation of Ultra-Thin, or Amorphous Transparent Conducting Oxides—The Case for Raman Spectroscopy. *Materials*, 13(2), 2020.

[87] Andrey V. Korsakov, Matthew J. Kohn, and Maria Perraki. Applications of Raman Spectroscopy in Metamorphic Petrology and Tectonics. *Elements*, 16(2):105–110, 9/28/2021 2020.

[88] Anastasia Rousaki and Peter Vandenabeele. In situ Raman spectroscopy for cultural heritage studies. *Journal of Raman Spectroscopy*, n/a(n/a).

[89] Yi Xu, Peng Zhong, Aimin Jiang, Xing Shen, Xiangmei Li, Zhenlin Xu, Yudong Shen, Yuanming Sun, and Hongtao Lei. Raman spectroscopy coupled with chemometrics for food authentication: A review. *TrAC Trends in Analytical Chemistry*, 131:116017, 2020.

[90] JoséIzo Santana da Silva de Jesus, Raimar Löbenberg, and Nádia Araci Bou-Chacra. Raman Spectroscopy for Quantitative Analysis in the Pharmaceutical Industry. *Journal of Pharmacy & Pharmaceutical Sciences*, 23(1):24–46, 2021/09/28 2020.

[91] M.R. Arnfield, J. Tulip, and M.S. McPhee. Optical propagation in tissue with anisotropic scattering. *IEEE Transactions on Biomedical Engineering*, 35(5):372–381, 1988.

[92] Julia L Sandell and Timothy C Zhu. A review of in-vivo optical properties of human tissues and its impact on PDT. *Journal of biophotonics*, 4(11-12):773–787, 11 2011.

[93] Dustin W. Shipp, Faris Sinjab, and Ioan Notingher. Raman spectroscopy: techniques and applications in the life sciences. *Adv. Opt. Photon.*, 9(2):315–428, Jun 2017.

[94] Adolf Smekal. Zur Quantentheorie der Dispersion. *Naturwissenschaften*, 11(43):873–875, October 1923.

[95] C. V. Raman and K. S. Krishnan. A New Type of Secondary Radiation. *Nature*, 121(3048):501–502, 1928.

[96] Simona Cintă Pînzaru and Wolfgang Kiefer. *Raman's Discovery in Historical Context*, pages 3–21. Springer International Publishing, Cham, 2018.

[97] GS Landsberg and LI Mandelstam. New phenomenon in scattering of light (preliminary report). *Journal of the Russian Physico-Chemical Society, Physics Section*, 60:335, 1928.

[98] I. L. Fabelinskii. Priority and the Raman effect. *Nature*, 343(6260):686–686, 1990.

[99] R.L. McCreery. *Raman Spectroscopy for Chemical Analysis*. Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications. Wiley, 2005.

[100] T. H. Maiman. Stimulated Optical Radiation in Ruby. *Nature*, 187(4736):493–494, 1960.

[101] S. Sasic and S. Ekins. *Pharmaceutical Applications of Raman Spectroscopy*. Wiley Series on Technologies for the Pharmaceutical Industry. Wiley, 2007.

[102] R.L. McCreery. *Raman Spectroscopy for Chemical Analysis*. Chemical Analysis: A Series of Monographs on Analytical Chemistry and Its Applications. Wiley, 2000.

[103] Neil Everall, Thomas Hahn, Pavel Matousek, Anthony W Parker, and Michael Towrie. Photon migration in Raman spectroscopy. *Appl Spectrosc*, 58(5):591–597, May 2004.

[104] P. Matousek, I. P. Clark, E. R. C. Draper, M. D. Morris, A. E. Goodship, N. Everall, M. Towrie, W. F. Finney, and A. W. Parker. Subsurface Probing in Diffusely Scattering Media Using Spatially Offset Raman Spectroscopy. *Appl. Spectrosc.*, 59(4):393–400, Apr 2005.

[105] Pavel Matousek and Nicholas Stone. Recent advances in the development of Raman spectroscopy for deep non-invasive medical diagnosis. *J Biophotonics*, 6(1):7–19, Jan 2013.

[106] P. Matousek and A. W. Parker. Non-invasive probing of pharmaceutical capsules using transmission Raman spectroscopy. *Journal of Raman Spectroscopy*, 38(5):563–567, 2007.

[107] Rachel Kast, Gregory Auner, Sally Yurgelevic, Brandy Broadbent, Aditya Raghunathan, Laila M Poisson, Tom Mikkelsen, Mark L Rosenblum, and Steven N Kalkanis. Identification of regions of normal grey matter and white matter from pathologic glioblastoma and necrosis in frozen sections using Raman imaging. *J Neurooncol*, 125(2):287–295, Nov 2015.

[108] Steven N Kalkanis, Rachel E Kast, Mark L Rosenblum, Tom Mikkelsen, Sally M Yurgelevic, Katrina M Nelson, Aditya Raghunathan, Laila M Poisson, and Gregory W Auner. Raman spectroscopy to distinguish grey matter, necrosis, and glioblastoma multiforme in frozen tissue sections. *J Neurooncol*, 116(3):477–485, Feb 2014.

[109] K Maheedhar, Rani A Bhat, R Malini, N B Prathima, Patil Keerthi, Pralhad Kushtagi, and C Murali Krishna. Diagnosis of ovarian cancer by Raman spectroscopy: a pilot study. *Photomed Laser Surg*, 26(2):83–90, Apr 2008.

[110] Stella Corsetti, Thomas Rabl, David McGloin, and Ghulam Nabi. Raman spectroscopy for accurately characterizing biomolecular changes in androgen-independent prostate cancer cells. *J Biophotonics*, 11(3), Mar 2018.

[111] Dinesh K R Medipally, Adrian Maguire, Jane Bryant, John Armstrong, Mary Dunne, Marie Finn, Fiona M Lyng, and Aidan D Meade. Development of a high throughput (HT) Raman spectroscopy method for rapid screening of liquid blood plasma from prostate cancer patients. *Analyst*, 142(8):1216–1226, Apr 2017.

[112] Abhilash K. Pandya, Gulay K. Serhatkulu, Alex Cao, Rachel E. Kast, Houbei Dai, Raja Rabah, Janet Poulik, Sanjeev Banerjee, Ratna Naik, Volkan Adsay, Gregory W. Auner, Michael D. Klein, Jagdish S. Thakur, and Fazlul H. Sarkar. Evaluation of Pancreatic Cancer With Raman Spectroscopy in a Mouse Model. *Pancreas*, 36(2), 2008.

[113] Adrian Ghita, Pavel Matousek, and Nick Stone. High sensitivity non-invasive detection of calcifications deep inside biological tissue using Transmission Raman Spectroscopy. *Journal of Biophotonics*, 11(1):e201600260, 2018.

[114] Nicholas Stone, Rebecca Baker, Keith Rogers, Anthony William Parker, and Pavel Matousek. Subsurface probing of calcifications with spatially offset Raman spectroscopy (SORS): future possibilities for the diagnosis of breast cancer. *Analyst*, 132:899–905, 2007.

[115] J. Charles G. Jeynes, Freddy Wordingham, Laura J. Moran, Alison Curnow, and Tim J. Harries. Monte Carlo Simulations of Heat Deposition during Photothermal Skin Cancer Therapy Using Nanoparticles. *Biomolecules*, 9(8), 2019.

[116] G.L.L. Buffon. *Histoire naturelle, générale et particulière, servant de suite à la Théorie de la Terre et d'introduction à l'Histoire des Minéraux...Supplément Tome premier [septième]*. Number v. 4 in Histoire naturelle, générale et particulière, servant de suite à la Théorie de

la Terre et d'introduction à l'Histoire des Minéraux...Supplément Tome premier [septième]. de l'imprimerie royale, 1777.

[117] P. Beckmann. *A History of Pi*. Griffin Books. St. Martin's Publishing Group, 2015.

[118] Asaph Hall. On an experimental determination of pi. *Messeng. Math*, 2:113–114, 1873.

[119] Shu Tezuka. *Financial Applications of Monte Carlo and Quasi-Monte Carlo Methods*, pages 303–332. Springer New York, New York, NY, 1998.

[120] Victor Platon and Andreea Constantinescu. Monte Carlo Method in Risk Analysis for Investment Projects. *Procedia Economics and Finance*, 15:393–400, 2014. Emerging Markets Queries in Finance and Business (EMQ 2013).

[121] Andrew J. Cassey and Ben O. Smith. Simulating confidence for the Ellison–Glaeser index. *Journal of Urban Economics*, 81:85–103, 2014.

[122] Tilmann Gneiting and Adrian E. Raftery. Weather Forecasting with Ensemble Methods. *Science*, 310(5746):248–249, 2005.

[123] Ahmad Ali, Tim J Harries, and Thomas A Douglas. Modelling massive star feedback with Monte Carlo radiation hydrodynamics: photoionization and radiation pressure in a turbulent cloud. *Monthly Notices of the Royal Astronomical Society*, 477(4):5422–5436, 04 2018.

[124] T-P Li and Y-Q Ma. Analysis methods for results in gamma-ray astronomy. *The Astrophysical Journal*, 272:317–324, 1983.

[125] Nicholas Metropolis. The beginning of the Monte Carlo method. *Los Alamos Science*, 15(584):125–130, 1987.

[126] Nick Metropolis and Jack Worlton. A Trilogy on Errors in the History of Computing. *Annals of the History of Computing*, 2:49–59, 1980.

[127] Nicholas Metropolis and S. Ulam. The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.

[128] Lorne W Avery and Lewis L House. An investigation of resonance-line scattering by the Monte Carlo technique. *The Astrophysical Journal*, 152:493, 1968.

[129] Brian C. Wilson and G. Adam. A Monte Carlo model for the absorption and flux distributions of light in tissue. *Medical physics*, 10 6:824–30, 1983.

[130] M Keijzer, S L Jacques, S A Prahl, and A J Welch. Light distributions in artery tissue: Monte Carlo simulations for finite-diameter laser beams. *Lasers Surg Med*, 9(2):148–154, 1989.

[131] SA Prahl, M Keijzer, SL Jacques, and AJ Welch. A Monte Carlo Model of Light Propagation in Tissue.

[132] Lihong Wang, Steven L. Jacques, and Liqiong Zheng. MCML—Monte Carlo modeling of light transport in multi-layered tissues. *Computer Methods and Programs in Biomedicine*, 47(2):131–146, 1995.

[133] Karthik Vishwanath and Mary-Ann Mycek. Time-resolved photon migration in bilayered tissue models. *Opt. Express*, 13(19):7466–7482, Sep 2005.

[134] Matthew V Schulmerich, Jacqueline H Cole, Kathryn A Dooley, Michael D Morris, Jaclynn M Kreider, Steven A Goldstein, Subhadra Srinivasan, and Brian W Pogue. Noninvasive Raman tomographic imaging of canine bone tissue. *J Biomed Opt*, 13(2):020506, Mar-Apr 2008.

[135] Annika M. K. Enejder, Tae-Woong Koo, Jeankun Oh, Martin Hunter, Slobodan Sasic, Michael S. Feld, and Gary L. Horowitz. Blood analysis by Raman spectroscopy. *Opt. Lett.*, 27(22):2004–2006, Nov 2002.

[136] Wei-Chuan Shih, Kate L. Bechtel, and Michael S. Feld. Intrinsic Raman spectroscopy for quantitative biological spectroscopy Part I: Theory and simulations. *Opt. Express*, 16(17):12726–12736, Aug 2008.

[137] Matthew D. Keller, Robert H. Wilson, Mary-Ann Mycek, and Anita Mahadevan-Jansen. Monte Carlo Model of Spatially Offset Raman Spectroscopy for Breast Tumor Margin Analysis. *Applied Spectroscopy*, 64(6):607–614, 2010. PMID: 20537228.

[138] Shuang Wang, Jianhua Zhao, Harvey Lui, Qingli He, Jintao Bai, and Haishan Zeng. Monte Carlo simulation of in vivo Raman spectral Measurements of human skin with a multi-layered tissue optical model. *Journal of Biophotonics*, 7(9):703–712, 2014.

[139] Alireza Akbarzadeh, Ehsan Edjlali, Guillaume Sheehy, Juliette J. Selb, Rajeev Agarwal, Jessie R. Weber, and Frédéric Leblond. Experimental validation of a spectroscopic Monte Carlo light transport simulation technique and Raman scattering depth sensing analysis in biological tissue. *Journal of Biomedical Optics*, 25(10):1 – 19, 2020.

[140] Vijitha Periyasamy, Humaira Bte Jaafar, and Manojit Pramanik. Raman Monte Carlo simulation for light propagation for tissue with embedded objects. In E. Duco Jansen and Hope Thomas Beier, editors, *Optical Interactions with Tissue and Cells XXIX*, volume 10492, pages 107 – 112. International Society for Optics and Photonics, SPIE, 2018.

[141] Vijitha Periyasamy, Sanchita Sil, Gagan Dhal, Freek Ariese, Siva Umapathy, and Manojit Pramanik. Experimentally validated Raman Monte Carlo simulation for a cuboid object to obtain Raman spectroscopic signatures for hidden material. *Journal of Raman Spectroscopy*, 46(7):669–676, 2015.

[142] Sara Mosca, Priyanka Dey, Marzieh Salimi, Francesca Palombo, Nick Stone, and Pavel Matousek. Non-invasive depth determination of inclusion in biological tissues using spatially offset Raman spectroscopy with external calibration. *Analyst*, 145(23):7623–7629, 2020.

[143] Sara Mosca, Priyanka Dey, Marzieh Salimi, Benjamin Gardner, Francesca Palombo, Nick Stone, and Pavel Matousek. Spatially Offset Raman Spectroscopy—How Deep? *Analytical Chemistry*, 93(17):6755–6762, 2021. PMID: 33886282.

[144] Alexander P Dumont, Qianqian Fang, and Chetan A Patil. A computationally efficient Monte-Carlo model for biomedical Raman spectroscopy. *J Biophotonics*, 14(7):e202000377, Jul 2021.

[145] T.J. Harries, T.J. Haworth, D. Acreman, A. Ali, and T. Douglas. The TORUS radiation transfer code. *Astronomy and Computing*, 27:63–95, 2019.

[146] Gilwon Yoon, Scott A. Prahl, and Ashley J. Welch. Accuracies of the diffusion approximation and its similarity relations for laser irradiated biological media. *Appl. Opt.*, 28(12):2250–2255, Jun 1989.

[147] Reindert Graaff, Jan G. Aarnoudse, Frits F. M. de Mul, and Henk W. Jentink. Similarity relations for anisotropic scattering in absorbing media. *Optical Engineering*, 32(2):244 – 252, 1993.

[148] Caigang Zhu and Quan Liu. Review of Monte Carlo modeling of light transport in tissues. *Journal of Biomedical Optics*, 18(5):1 – 13, 2013.

[149] Louis G Henyey and Jesse Leonard Greenstein. Diffuse radiation in the galaxy. *The Astrophysical Journal*, 93:70–83, 1941.

[150] SL Jacques, CA Alter, and SA Prahl. Angular dependence of HeNe laser light scattering by human dermis. *Lasers in the Life Sciences*, 2(4):309–333, 1988.

[151] L.B. Lucy. Computing radiative equilibria with Monte Carlo techniques. A&A, 344:282–288, April 1999.

[152] E.J. Kontoghiorghes. *Handbook of Parallel Computing and Statistics*. Statistics: A Series of Textbooks and Monographs. CRC Press, 2005.

[153] Laura J. Moran, Freddy Wordingham, Benjamin Gardner, Nicholas Stone, and Tim J. Harries. An experimental and numerical modelling investigation of the optical properties of Intralipid using deep Raman spectroscopy. *Analyst*, 146:7601–7610, 2021.

[154] Francis W. L. Esmonde-White, Karen A. Esmonde-White, Matthew R. Kole, Steven A. Goldstein, Blake J. Roessler, and Michael D. Morris. Biomedical tissue phantoms with controlled geometric and optical properties for Raman spectroscopy and tomography. *Analyst*, 136:4437–4446, 2011.

[155] Jennifer-Lynn H. Demers, Francis W.L. Esmonde-White, Karen A. Esmonde-White, Michael D. Morris, and Brian W. Pogue. Next-generation Raman tomography instrument for non-invasive in vivo bone imaging. *Biomed. Opt. Express*, 6(3):793–806, Mar 2015.

[156] Ingeborg E. Iping Petterson, Francis W. L. Esmonde-White, Wouter de Wilde, Michael D. Morris, and Freek Ariese. Tissue phantoms to compare spatial and temporal offset modes of deep Raman spectroscopy. *Analyst*, 140:2504–2512, 2015.

[157] Ben Aernouts, Eduardo Zamora-Rojas, Robbe Van Beers, Rodrigo Watté, Ling Wang, Mizuki Tsuta, Jeroen Lammertyn, and Wouter Saeys. Supercontinuum laser based optical characterization of Intralipid® phantoms in the 500-2250 nm range. *Opt. Express*, 21(26):32450–32467, Dec 2013.

[158] Ben Aernouts, Robbe Van Beers, Rodrigo Watté, Jeroen Lammertyn, and Wouter Saeys. Dependent scattering in Intralipid® phantoms in the 600-1850 nm range. *Opt. Express*, 22(5):6086–6098, Mar 2014.

[159] Paola Di Ninni, Fabrizio Martelli, and Giovanni Zaccanti. Intralipid: towards a diffusive reference standard for optical tissue phantoms. *Physics in Medicine and Biology*, 56(2):21, 2010.

[160] Paola Di Ninni, Fabrizio Martelli, and Giovanni Zaccanti. Effect of dependent scattering on the optical properties of Intralipid tissue phantoms. *Biomed. Opt. Express*, 2(8):2265–2278, Aug 2011.

[161] Giovanni Zaccanti, Samuele Del Bianco, and Fabrizio Martelli. Measurements of optical properties of high-density media. *Appl. Opt.*, 42(19):4023–4030, Jul 2003.

[162] Hugo J. van Staveren, Christian J. M. Moes, Jan van Marie, Scott A. Prahl, and Martin J. C. van Gemert. Light scattering in lntralipid-10% in the wavelength range of 400–1100 nm. *Appl. Opt.*, 30(31):4507–4514, Nov 1991.

[163] René Michels, Florian Foschum, and Alwin Kienle. Optical properties of fat emulsions. *Opt. Express*, 16(8):5907–5925, Apr 2008.

[164] Serge Grabtchak, Tyler Palmer, Florian Foschum, André Liemert, Alwin Kienle, and William Whelan. Experimental spectro-angular mapping of light distribution in turbid media. *Journal of biomedical optics*, 17:067007, 06 2012.

[165] Francesco Sardanelli, Franco Zandrino, Andrea Imperiale, Emma Bonaldo, Maria G. Quartini, and Nadia Cogorno. Breast Biphasic Compression versus Standard Monophasic Compression in X-ray Mammography. *Radiology*, 217(2):576–580, 2000. PMID: 11058663.

[166] P. Matousek and A. W. Parker. Bulk Raman Analysis of Pharmaceutical Tablets. *Applied Spectroscopy*, 60(12):1353–1357, 2006. PMID: 17217583.

[167] Shudong Jiang, Brian W. Pogue, Keith D. Paulsen, Christine Kogel, and StevenP. Poplack. In vivo near-infrared spectral detection of pressure-induced changes in breast tissue. *Opt. Lett.*, 28(14):1212–1214, Jul 2003.

[168] Daqing Piao, Anqi Zhang, and Guan Xu. Photon diffusion in a homogeneous medium bounded externally or internally by an infinitely long circular cylindrical applicator. V. Steady-state fluorescence. *J. Opt. Soc. Am. A*, 30(4):791–805, Apr 2013.

[169] Martin Hohmann, Benjamin Lengenfelder, Daniel Muhr, Moritz Späth, Maximilian Hauptkorn, Florian Klämpfl, and Michael Schmidt. Direct measurement of the scattering coefficient. *Biomed. Opt. Express*, 12(1):320–335, 2021.

[170] N. Bulstrode, E. Bellamy, and S. Shrotria. Breast volume assessment: comparing five different techniques. *The Breast*, 10(2):117–123, 2001.

[171] Anh Phong Tran and Steven L. Jacques. Modeling voxel-based Monte Carlo light transport with curved and oblique boundary surfaces. *Journal of Biomedical Optics*, 25(2):1 – 13, 2020.

[172] Cancer Research UK. CRUK: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer, Accessed June 2022.

[173] Shi-Jun Zhang, Hai-Ning Qian, Yan Zhao, Kai Sun, Hui-Qing Wang, Guo-Qing Liang, Feng-Hua Li, and Zheng Li. Relationship between age and prostate size. *Asian journal of andrology*, 15(1):116–120, 01 2013.

[174] Ching-Cheng Chuang, Yu-Tzu Lee, Chung-Ming Chen, Yao-Sheng Hsieh, Tsan-Chi Liu, and Chia-Wei Sun. Patient-oriented simulation based on Monte Carlo algorithm by using MRI data. *BioMedical Engineering OnLine*, 11(1):21, 2012.

[175] Harald Paganetti. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Physics in Medicine and Biology*, 57(11):R99–R117, may 2012.

[176] Paul L. Stiles, Jon A. Dieringer, Nilam C. Shah, and Richard P. Van Duyne. Surface-Enhanced Raman Spectroscopy. *Annual Review of Analytical Chemistry*, 1(1):601–626, 2008. PMID: 20636091.

[177] Sara Mosca, Priyanka Dey, Tanveer A. Tabish, Francesca Palombo, Nicholas Stone, and Pavel Matousek. Determination of inclusion depth in ex vivo animal tissues using surface enhanced deep Raman spectroscopy. *Journal of Biophotonics*, 13(1):e201960092, 2020.