Information limits of imaging through highly diffusive materials using spatiotemporal measurements of diffuse photons

Jack Radford

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

May 2023

Abstract

Conventional medical imaging instruments are bulky, expensive, and use harmful ionising radiation. Combining ultrafast single-photon detectors and pulsed laser sources at optical wavelengths has the potential to offer inexpensive, safe, and potentially wearable alternatives. However, photons at optical wavelengths are strongly scattered by biological tissue, which corrupts direct imaging information about regions of absorbing interactions below the tissue surface. The work in this thesis studies the potential of measuring indirect imaging information by resolving diffuse photon measurements in space and time. The practical limits of imaging through highly diffusive material, e.g., biological tissue, is explored and validated with experimental measurements. The ill-posed problem of using the information in diffuse photon measurements to reconstruct images at the limits of the highly diffusive regime is tackled using probabilistic machine learning, demonstrating the potential of migrating diffuse optical imaging techniques beyond the currently accepted limits and underlining the importance of uncertainty quantification in reconstructions. The thesis is concluded with a challenging biomedical optics experiment to transmit photons diametrically through an adult human head. This problem was tackled experimentally and numerically using an anatomically accurate Monte Carlo simulation which uncovered key practical considerations when detecting photons at the extreme limits of the highly diffusive regime. Although the experimental measurements were inconclusive, comparisons with the numerical results were promising. More in-depth numerical simulations indicated that light could be guided in regions of low scattering and absorption to reach deep areas inside the head, and photons can, in principle, be transmitted through the entire diameter of the head. The collective evidence presented in this thesis reveals the potential of diffuse optical imaging to extend beyond the currently accepted limits to non-invasively image deep regions of the human body and brain using optical wavelengths.

Contents

List of Figures

Acknowledgements

I would like to acknowledge Prof. Daniele Faccio for his supervisory guidance throughout the projects described in this thesis. I would also like to thank Dr. Samuel Nerenberg, Dr. Francesco Tonolini, Dr. Ashley Lyons, Dr. Valentin Kapitany, and Dr. Alex Turpin for their collaboration and guidance. The work in this thesis was supported by EPSRC Centre for Doctoral Training in Intelligent Sensing and Measurement, Grant No. EP/L016753/1 and EPSRC (UK, grant no. EP/T00097X/1).

Declaration

With the exception of chapter 2, which contains background material, all work in this thesis was carried out by the author unless otherwise explicitly stated.

Sections 4.2 and 5.1 were experiments performed as part of a collaboration with co-authors of the works cited in the main text. The contribution to these studies is explicitly stated in the chapter.

Chapter 1

Introduction

Medical imaging instruments are crucial to medical diagnostics, treatment and research. However, they are bulky, expensive and often use harmful ionising radiation. Optical imaging techniques offer a safe, cheap and potentially wearable alternative but they are currently hindered by the physical process of scattering which degrades imaging information and restricts most modalities to imaging through depths to 1–2cm of tissue.

To image deep into the human body and brain (e.g., >5cm), the problem of collecting imaging information from scattered light becomes exponentially more difficult. As such, there is a lack of research exploring potential optical medical imaging modalities in the highly diffusive regime. The work presented in this thesis investigates optical imaging through diffusive materials at extreme length scales and at the limits of photon detection, with a particular focus on imaging deep within the human brain.

Anyone who has tried covering a torch with their hand will find that the human body is translucent to red light, yet the trans-illuminated shadow of dense structures such as bone are not visible. Unlike X-rays, red and near-infrared light interacts with the body predominately by scattering events due to microscopic changes in refractive index within biological tissue. Scattered photons are categorised in three types [1–3]: *ballistic* photons propagate through tissue in a direct path to the detector (i.e. maintaining their spatial coherence) and therefore a trans-illuminated shadow of absorbing regions embedded inside the material can be directly resolved by accumulating many of these photons on an imaging detector array; *snake* photons are weakly scattered and lose some spatial coherence, resulting in a loss of some imaging information; and *diffuse* photons which undergo multiple scattering events and have a trajectory that no longer corresponds to their initial direction of propagation. As a consequence, diffuse photons appear in randomised locations at a detector array and the imaging information is considered to be corrupted. An illustration of these photon propagation regimes can be found in figure (2.8).

It is commonly accepted in current literature that diffuse photons contain little to no imaging information and, consequently, most techniques aim to remove the diffuse photons from the measured signal to isolate the ballistic and snake photons [2, 4–10]. However, more recent works show that including diffuse photons can enhance imaging through diffusive materials [11] and enable imaging through more scattering than ever thought possible [12]. This indicates that there is potential for new imaging regimes which do not attempt to remove diffuse photons from the signal but instead use this information to enhance image retrieval quality, or image in cases where ballistic and snake photons cannot be detected.

The work in presented in this thesis first investigates, empirically and theoretically, the information contained in statistical measurements of diffuse photons and the practical limits of image information transfer through highly diffusive materials. Using a recent machine learning framework for computational imaging inverse problems [13], simulated and experimental evidence of imaging near these derived limits is shown and the practical advantages of using probabilistic inversion models to determine the uncertainty in the solutions is discussed. Finally, a challenging but relevant biomedical optics experiment to detect photons transmitted through an entire adult human head is explored to demonstrate the detection of photons at the extreme limits of tissue scattering. Anatomically accurate simulations of photon migration in the human head are presented to conclude the thesis and justify the potential for a paradigm shift in diffuse optical imaging to explore deep regions of the human body and brain.

1.1 Thesis outline

Chapter 2 begins with a background theory discussion of the physical process of photon diffusion through scattering materials and introduces some key variables that are frequently used in the following chapters. A literature review of diffuse optical imaging provides context and motivation of the thesis and provides an overview of the methodologies and hardware that are used in the field. The chapter concludes with an intuitive example that gives a sense of the difficulties involved with determining accurate solutions to ill-posed inverse problems such as diffuse optical imaging.

In Chapter 3, the data and methods used in the work of Lyons et. al. [12], which demonstrates the reconstruction of images using the spatiotemporal distribution of highly diffuse photons, are used in an analysis study to distinguish imaging in the highly diffuse regime from early photon imaging methods. The analysis demonstrates the advantage of including diffuse photons for improving image reconstruction quality and underlines the importance of resolving diffuse photon measurements in both space and time when imaging in the highly diffusive regime.

Chapter 4 takes a more rigorous approach to studying the importance of fully resolving measurements by comparing the information content of the simulated diffuse photon measurements.

The principle of using information theory concepts to analyse measurements of multiple scattering is introduced with a case study in collaboration with the co-authors of [14]. These analysis concepts are then applied to the setting of diffuse optical imaging to compare the information resolved in different measurement domains and estimate the practical limits of imaging through highly diffusive materials. This numerical study of simulated measurements is then validated with an experiment that demonstrates that the estimated parameters used in the study are reasonable using today's technology. The impact of resolving information in different measurement domains on the reconstructed image quality is then assessed using a simple linear inversion model to indicate the ill-conditioning of the inverse problem.

Using the conclusions of Chapter 4, Chapter 5 demonstrates that the information content of measurements at the extreme limits of photon diffusion is sufficient to reconstruct images using a state-of-the-art machine learning inverse retrieval. The chapter begins by introducing the probablistic machine learning framework developed by Tonolini et al. [13], with a case study of a challenging inverse optics problem conducted in collaboration with the co-authors of [13]. An experimental design optimised using the conclusions of Chapter 4 is then simulated and a machine learning inverse model is used to demonstrate the potential to reconstruct images using measurements at the limits of the highly diffusive regime. The experiment was also implemented in the laboratory to test the inverse model with real-world data and take the first steps towards extending diffuse optical imaging beyond currently accepted limits.

Chapter 6 tackles an ambitious real-world biomedical optics challenge of transmitting photons diametrically though an adult human head. First the evolution of the laboratory experiments is introduced that lead to the measurement of plausible results. These results are then compared with anatomically accurate Monte Carlo simulations of the laboratory experiments, which were extended to examine the distribution of light in the head with different experimental conditions. A discussion of the possible inaccuracies of the Monte Carlo model leads to potential new research avenues that might enable imaging deep regions of the adult human brain.

Chapter 2

Background

2.1 Physics of photon transport in diffusive materials

Light is an electromagnetic (EM) wave which can interact with matter in two ways: absorption and scattering. Light-matter interaction encompasses the various pathways in which light can transfer its energy to charged particles in a material. It also describes how matter can transfer its energy by the creation of a photon by an electron transition to a lower energy state.

A photon is the elementary particle of light, a discrete wave packet of electromagnetic energy. Photons can be described as waves or as particles, i.e., the excitation of quantised mode of the electromagnetic field. The quantum nature of a photon refers to the phenomenon that a photon can only be created or destroyed as a whole entity and is indivisible. The energy of a photon is linearly related only to its frequency ν by Planck's constant [15]:

$$E = h\nu. \tag{2.1}$$

The term "photon" is used throughout this thesis and justified since, at the extreme limits of light detection, we employ single-photon counting detectors. Light detectors operate by the processes of absorbing quanta of electromagnetic energy by excitation of electrons to generate free charge carriers that produce a detectable change in electrical current at the detector output. However, the generation of a small number of charge carriers at low light levels produces an electrical current too weak to detect. Single-photon counting detectors typically amplify the number of free charge carriers to produce a detectable electrical current when the energy of just a single photon is absorbed in the photosensitive region.

Although a photon is a quantum particle, this thesis does not consider the quantum effects of light-matter interaction beyond discrete detection events as described above. The following sec-

tion describes the physical phenomenon of absorption and scattering, where *light* consists of numerous *photons*. Therefore the classical theory of electromagnetic (EM) wave interaction with dielectric materials is a sufficient description. In the subsequent section, the photon diffusion model used throughout this thesis is derived using radiative transfer theory, which is yet a simpler model, where all scattering contributions of light by a material are independent such that the interaction of EM fields can be neglected and only the addition of each scattered energy distribution may be considered.

However, photon propagation in a diffusive material can also be modelled numerically as a ray-tracing problem using geometrical optics. These models consider the complex path of a photon (or a "packet" of multiple photons) through a scattering material by independently sampling sequential steps of a random walk. Accumulating the distribution of trajectories with numerous random walks, and sampling the probability of absorption on each step in a Monte Carlo approach, is an accurate model of the macroscopic statistical distribution of photon transport in diffusive media [16–18]. Although these models are not used for simulations for the majority of work in this thesis, they provide an intuitive framework to describe the propagation of photons in a scattering material, where each random walk is commonly referred to as the trajectory of a single "photon".

2.1.1 Classical light-matter interaction

The interaction of light with a dielectric material can be considered classically by the Lorentz oscillator model [15, 19, 20]. In this description of the interaction, incoming (primary) EM radiation transfers energy to the material by displacing negatively charged electrons from positively charged nuclei causing temporary dipole moments throughout the bulk material. The driven oscillation of the dipole moments immediately re-radiates electromagnetic waves known as secondary radiation. The electrostatic attraction between positive nuclei and negative electron cloud also acts as a restoring force with a natural frequency ω_0. Since electrons have significantly less mass than protons, it can be assumed that only the cloud of electrons is displaced. Furthermore, the outer orbital electrons are displaced significantly more than the inner orbitals for visible light with relatively high-frequency EM oscillations. Therefore, the outer electron displacements from the nucleus can be modelled by a driven damped harmonic oscillator with a resonant frequency at the natural frequency $\omega_r = \omega_0$.

Suppose the energy of the incoming EM wave is close to the resonant frequency ω_r of the dipole moments. In that case, there is sufficient energy transferred to the atom or molecule to excite it to a higher energy state with high probability, e.g., by promotion of an electron to a higher energy orbital (mainly energies in the visible spectrum) or excitation of a vibration/rotation mode of molecules in the gas phase (predominantly energies in the infrared spectrum). This defines absorption of light as the transfer of incoming light energy to excite the electronic energy state

of a material, which is then either re-radiated by mechanical vibrations in the material (i.e., collisions of particles in a densely packed material leading to thermal energy) or re-emitted as an EM wave with frequency and direction of propagation which depends on the specifics of the dipole transition energy to the ground state (e.g., fluorescence). It is also possible for energy to be re-radiated as a combination of these processes so long as energy is conserved [15, 19].

However, if incoming light is not close to a resonance of the dipoles within the material, energy will still be transferred to cause the displacement of charges, but the oscillation of the displaced electrons will in turn cause immediate re-radiation of this energy as an EM wave which has a phase lag with respect to the primary wave. Each scattering centre (e.g., atom, molecule or solid/liquid particle) in the material can be considered as a dipole which radiates EM waves in all directions and the resulting superposition of both primary and secondary fields will interfere. In cases where the primary and secondary fields have the same energy, this is known as elastic scattering [15].

For densely packed scattering centres, where the spacing is much less than the wavelength, then neighbouring scattered fields will cancel each other in all directions apart from the forwards direction. This is a description of Rayleigh scattering in dense materials, where scattering particles are typically $\leq \frac{\lambda}{10}$ (e.g., individual atoms and molecules) [15].

This forward propagating scattering phenomena is not the case for loosely packed materials, which have interacting particles sparsely spaced greater the wavelength of the primary wave and the size of the scatterers [17]. In this case, the large spacing between dipole emitters means that light does not completely destructively interfere in lateral and backwards directions. This allows light to scatter away from the initial propagation direction where the probability to scatter in different directions depends on the microscopic local environment and configuration of the dipoles [15].

For larger particles $> \frac{\lambda}{10}$, which contain numerous molecules (such as fat globules and proteins), the dipole oscillations at opposite boundaries of the particle no longer produce secondary emission that constructively interferes in the lateral direction. Mie theory more accurately describes scattering for single particles in this regime, which results in more directional (anisotropic) scattering than in the Rayleigh case [15].

In the context of this thesis and photon diffusion in biomedical optics, scattering particles are loosely packed and elastically scattering such that the contributions of each particle can be assumed independent of all other scattering events. Mie scattering can be used to model scattering of a single spherical particle of any size [17], but when the scatterers are $> 10\lambda$ a simpler treatment using geometrical optics is sufficient [15].

The interaction of light with a single scattering particle can be described using light as an electro-

magnetic wave to determine the probability of interaction outcomes. The statistical description of single-particle scattering can then be summed for an ensemble of particles in loosely packed materials according to radiative transfer theory. Monte Carlo sampling models use the statistical description of single-particle interactions derived by treating light as a wave but then treat light as a classical particle to sample random walk trajectories between scattering events [17]. These models are suited to accurately simulate photon diffusion in materials with arbitrary complex and heterogeneous geometries but are computationally expensive. However, probabilistic descriptions for single-particle interactions can also be extended to statistical analytic expressions of photon diffusion, which can be derived using the principles of radiative transfer theory. These analytical models use approximations and are non-trivial for complex materials, but they are computationally inexpensive and have been shown to model simple homogeneous materials accurately [21]. These analytically derived models will be used in the majority of work presented in this thesis.

The absorption and scattering coefficients are introduced in the following sections of this chapter. These coefficients are statistical parameters which extend this microscopic description of scattering to the macroscopic radiative transfer theory by considering the summation of ensembles of absorption and scattering particles, where contributions are treated as independent of one another in loosely packed materials. Approximations to this theory that consider photon diffusion are then used to derive the photon diffusion approximation to the radiative transfer equation, the solution of which is at the core of the work in this thesis.

2.1.2 Absorption coefficient

Rather than characterise the various dissipative absorption pathways specific to each particle in a bulk material, a statistical approach can also be used to describe the absorptive properties of an ensemble of particles. This treatment first considers an absorption interaction as a loss of energy upon interacting with a single particle (e.g., atom or molecule) in the material irrespective of the way in which the energy is dissipated.

The efficiency of a single absorbing particle within a material to absorb light can be described the ratio of power absorbed P_{abs} compared with the incident power P_i [17, 19]:

$$Q_{abs} = \frac{P_{abs}}{P_i}. \tag{2.2}$$

For a particle with geometrical cross section σ_g, the absorption efficiency can be rewritten by substituting the incident power with the product of the incident fluence rate Φ_i (Wcm^{-2}) and the area of interaction. This leads to a relationship between the absorption efficiency and the geometrical cross section of a particle, and defines a parameter known as the absorption cross

section σ_a (cm^2) [19]:

$$\begin{aligned} Q_{abs} &= \frac{P_{abs}}{\Phi_i \sigma_g} \\ &= \frac{\sigma_a}{\sigma_g}. \end{aligned} \tag{2.3}$$

The absorption cross section is typically defined in units of cm^2 in biomedical optics literature, which will also be the units used in the context of this thesis.

To transition this single-particle description to bulk materials, a statistical approximation is required since the number, size and shape of every particle is unknown in practice. Instead of considering each particle's individual geometry, a bulk material is assumed to consist of distribution of spherical particles whereby the radius is determined by the average radius of the true geometrical shape, r_{avg}. Furthermore, each particle is considered to have an absorption efficiency which is independent of direction of incident light [19].

Considering the regime where the distances between particles is much greater than the wavelength, interference effects can be neglected which allows for a simplified calculation of the collective absorption cross section per volume of material. This assumption considers the absorption contribution of each particle to be independent and therefore the net effect of an ensemble of particle is simply the summation of each contribution. The macroscopic statistical characterisation of the absorption of an ensemble of particles is known as the absorption coefficient [19]:

$$\boxed{\mu_a = \int_0^\infty \sigma_a(r_{\text{avg}}) \xi(r_{\text{avg}})\, dr_{\text{avg}},} \tag{2.4}$$

where $dN = \xi(r_{\text{avg}}) dr_{\text{avg}}$ is the number of particles in a unit volume with radii in the interval dr_{avg}. The absorption coefficient describes the mean number of absorption events per unit length of propagation through the material. Note that for identical particles which all have the same average radius r_0, then $\xi = \rho\delta(r_{\text{avg}} - r_0) = \rho$, and the absorption coefficient reduces to $\mu_a = \rho\sigma_a$, where the density $\rho = N/V$ (cm^{-3}) is the number of absorbing particles N per volume V [19]. Since σ_a has units cm^2, μ_a has units cm^{-1}.

2.1.3 Scattering coefficient

The scattering of a light by an ensemble of particles can also be described statistically. Similar to determining the absorption cross section above, the scattering cross section of a single particle can found by considering the scattering efficiency i.e., taking the ratio of incident power compared with the power scattered by the particle.

Consider a time-independent incident monochromatic plane wave electric field, with wavenum-

ber $k = \frac{2\pi}{\lambda}$ propagating in direction $\hat{s}$ [19]:

$$E_i(\vec{r}) = |E_0| \exp(ik\hat{s} \cdot \vec{r}), \tag{2.5}$$

where the $|E_0|$ is the amplitude. The scattered electric field $E_s(\vec{r})$ by a particle in the far-field can be expressed by a spherical wave originating at the centre of a particle with a decaying amplitude inversely proportional to the radius r [19]:

$$E_s(\vec{r}) = |E_0| f(\hat{s}', \hat{s}) \frac{\exp(ikr)}{4\pi r}, \tag{2.6}$$

where $f(\hat{s}', \hat{s})$ is the scattering amplitude which determines the contribution of the incident field scattered from the $\hat{s}$ direction into the direction $\hat{s}'$. The scattering amplitude function encompasses the size, shape and spatially varying refractive index properties of a particle, and is normalised by the amplitude $|E_0|$ of the incident wave. Unlike the case of absorption, the scattering amplitude must be included to determine the scattered power as a function of direction.

The power per area of an electromagnetic wave is represented by the time-averaged Poynting vector $\langle \vec{S} \rangle$ (W/m^2), which for a plane-wave in the far-field is [15, 19]:

$$\langle \vec{S} \rangle = \frac{c\varepsilon_0}{8\pi} |\vec{E}|^2 \hat{s}, \tag{2.7}$$

where c is the speed of light and ε_0 is the permittivity of free space. Note that the time-averaged Poynting vector has the units of fluence rate (or intensity), but is a vector quantity. However, the magnitude of the Poynting vector is equivalent to the fluence rate:

$$\Phi = |\langle \vec{S} \rangle|. \tag{2.8}$$

Substituting (2.5) and (2.6) into (2.7), expressions for the incident $\langle \vec{S}_i \rangle$ and scattered $\langle \vec{S}_s \rangle$ Poynting vectors can be written by:

$$\langle \vec{S}_i \rangle = \frac{c\varepsilon_0}{8\pi} |E_0|^2 \hat{s}, \tag{2.9}$$

and

$$\langle \vec{S}_s \rangle = \frac{c\varepsilon_0}{8\pi} |E_0|^2 \frac{f(\hat{s}', \hat{s})}{r^2} \hat{s}', \tag{2.10}$$

respectively.

Furthermore, the power scattered by a particle P_s in the absence of any external contributing fields can be expressed as the divergence of the scattered Poynting vector $\langle \vec{S}_s \rangle$ integrated over

the volume of the particle:

$$P_s = \int_V \nabla \cdot \langle \vec{S}_s \rangle \, dV, \tag{2.11}$$

where $\nabla = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle$ and the scalar product represents the rate of change in power density out of the volume.

Similar to the derivation of the absorption cross section, the scattering cross section can be found by considering the scattering efficiency of the particle:

$$\begin{aligned} Q_s &= \frac{P_s}{P_i}, \\ &= \frac{P_s}{\Phi_i \sigma_g}, \\ &= \frac{\sigma_s}{\sigma_g}, \end{aligned} \tag{2.12}$$

where the incident power P_i on the denominator is substituted as a product of the incident fluence rate Φ_i and the geometrical cross section of the particle σ_g. The expression for the scattering cross section is then:

$$\sigma_s = \frac{P_s}{\Phi_i}. \tag{2.13}$$

Substituting (2.8) and (2.11) into the equation above yields,

$$\begin{aligned} \sigma_s &= \int_V \frac{\nabla \cdot \langle \vec{S}_s \rangle}{|\langle \vec{S}_i \rangle|} \, dV, \\ &= \int_S \frac{\langle \vec{S}_s \rangle \cdot \vec{n}}{|\langle \vec{S}_i \rangle|} \, dS, \end{aligned} \tag{2.14}$$

where the divergence theorem is used to express the volume integral as a surface integral. Substituting the expressions for the incident (2.9) and scattered (2.10) time-averaged Poynting vectors into (2.14) and using the definition of differential solid angle $d\Omega = \frac{dS}{r^2}$, the scattering cross section can be expressed in its most common form [19]:

$$\begin{aligned} \sigma_s &= \int_S \frac{|f(\hat{s}', \hat{s})|^2}{r^2} \, dS, \\ &= \int_S |f(\hat{s}', \hat{s})|^2 \, \frac{dS}{r^2}, \\ &= \int_{4\pi} |f(\hat{s}', \hat{s})|^2 \, d\Omega. \end{aligned} \tag{2.15}$$

To describe the scattering statistics of a macroscopic ensemble of particles, the same assumptions made for the absorption coefficient can be made whereby each particle in the volume is assumed to be spherical with a radius determined by the average radius of their true geometric shape, r_{avg}. Furthermore, if the interference of light is neglected then the contribution of scattering from each particle is additive. These assumptions allow for the integration of the scattering cross section for every particle, and defines the macroscopic quantity known as the scattering coefficient:

$$\boxed{\mu_s = \int_0^\infty \sigma_s(r_{\text{avg}})\xi(r_{\text{avg}})\, dr_{\text{avg}},} \tag{2.16}$$

where $dN = \xi(r_{\text{avg}})dr_{\text{avg}}$ is the number of particles in a unit volume with radii in the interval dr_{avg}. As discussed in the derivation of absorption case, for particles with identical radius, (2.16) reduces to $\mu_s = \rho\sigma_s$, where the density $\rho = N/V$ (cm^{-3}) is the number of scattering particles N per volume V, and because σ_s has units of cm^2, μ_s has units cm^{-1}.

2.1.4 Reduced scattering coefficient

The angular dependence of scattering is encapsulated by the scattering amplitude $f(\hat{s}', \hat{s})$ introduced in the previous section. Although the scattering amplitude depends on the geometry and optical properties of each individual particle in a material, in the context of biomedical optics it is sufficient to assume that all particles are statistically equivalent and a common generic function can be used. Furthermore, since the scattering of a single particle can be characterised by its scattering cross section, only the modulus-squared of the scattering amplitude needs consideration, as per equation (2.15). This introduces a new parameter, the phase function $p(\hat{s}', \hat{s})$, which represents the probability of incident light travelling in direction $\hat{s}'$ to scatter into the direction $\hat{s}$ [19, 22]:

$$p(\hat{s}', \hat{s}) = \frac{1}{(\sigma_a + \sigma_s)} |f(\hat{s}', \hat{s})|^2. \tag{2.17}$$

The normalisation factor is a requirement of a probability distribution. Note that the phase function has no relationship to the phase of the EM wave and is purely a naming convention originated in astronomy applications where it was first used [19].

Interestingly, the integration of the phase function over all angles describes the total probability of interacting light to scatter, which is commonly referred to as the albedo [19]:

$$\begin{aligned} W_0 &= \int_{4\pi} p(\hat{s}', \hat{s})\, d\Omega, \\ &= \frac{\sigma_s}{(\sigma_s + \sigma_a)}. \end{aligned} \tag{2.18}$$

When no absorption is present $W_0 = 1$ and all the interacting light will be scattered, whereas $W_0 <$

1 is an indication of the scattering to absorption ratio. This parameter is commonly used when defining the phase function for scattering regimes such as the trivial isotropic phase function [19]:

$$p(\hat{s}', \hat{s}) = \frac{W_0}{4\pi}, \tag{2.19}$$

or the Henyey-Greenstein phase function, which is more relevant for biomedical optics [17, 19]:

$$p(\hat{s}', \hat{s}) = \frac{1}{4\pi} \frac{W_0(1-g^2)}{(1+g^2-2g[\hat{s}' \cdot \hat{s}])^{3/2}}. \tag{2.20}$$

The Henyey-Greenstein phase function has been shown to accurately describe the statistical scattering bulk properties of biological tissue, however it is not general to all biological scattering, especially when considering scattering by a small number of single cells [23].

The Henyey-Greenstein phase function introduces an important parameter known as the anisotropy factor g, which is the expected value of the cosine scattering angle between incident $\hat{s}'$ and measured $\hat{s}$ directions [19]:

$$\begin{aligned} g &= \frac{1}{W_0} \int_{4\pi} p(\hat{s}', \hat{s})[\hat{s}' \cdot \hat{s}] \, d\Omega, \\ &= \frac{1}{\int_{4\pi} p(\hat{s}', \hat{s}) \, d\Omega} \int_{4\pi} p(\hat{s}', \hat{s})[\hat{s}' \cdot \hat{s}] \, d\Omega, \\ &= \langle \hat{s}' \cdot \hat{s} \rangle, \\ &= \langle \cos\theta \rangle, \end{aligned} \tag{2.21}$$

where $\langle . \rangle$ denotes the expected value. An anisotropy factor of $g = 0$ represents an isotropic material, since the forward and backward probability of scattering are equal and cancel. Most biological materials are highly forward scattering and have an anistropy factor $0.8 < g < 1$ [24, 25].

Note that the normalisation factor in (2.21) is required since g does not depend on absorption. To avoid confusion with various literature definitions, it is convenient to absorb this normalisation and define the probability density function (PDF) of scattering from direction $\hat{s}'$ into $\hat{s}$ with a separate parameter:

$$P(\hat{s}', \hat{s}) = \frac{p(\hat{s}', \hat{s})}{W_0}, \tag{2.22}$$

where $\int_{4\pi} P(\hat{s}', \hat{s}) \, d\Omega = 1$ by definition. Furthermore, since the assumption was made that the particles were spherical on average in Section (2.1.2), it is often a reasonable assumption that the PDF is independent of incident direction, and can be parameterised by just the cosine angle

between the incident and scattered directions:

$$P(\hat{s}', \hat{s}) = P(\hat{s}' \cdot \hat{s}). \tag{2.23}$$

The purpose of the g factor in the Henyey-Greenstein phase function is to account for the preferential direction of scattering as a function of the cosine angle of the incident and scattered directions. Therefore this phase function can also be expressed by this simplified parameterisation, $p(\hat{s}', \hat{s}) = p(\hat{s}' \cdot \hat{s})$ and, when normalised, can be used to determine the PDF of scattering, e.g., for biological tissue [17].

However, so far, only single-particle anisotropic scattering has been considered. For homogeneous materials (i.e., the density and nature of scatters does not change throughout the volume of the material) consisting of many anisotropically scattering particles, the scattering coefficient μ_s can be modified to the reduced scattering coefficient:

$$\boxed{\mu_s' = (1-g)\mu_s.} \tag{2.24}$$

For isotropic materials $g = 0$, and the μ_s' is equivalent to μ_s. However, μ_s' generalises to descriptions of anisotropic materials, e.g., biological tissues where $\mu_s' \sim 0.1\mu_s$, and reduces the number of scattering events per cm by an order of magnitude.

2.1.5 Extinction coefficient

The combined attenuation of a beam of light due to both absorption or scattering is known as extinction [17]. The extinction cross section of a particle within a material is the sum of the scattering and absorption cross sections $\sigma_{\text{ext}} = (\sigma_a + \sigma_s)$ and so the statistical value of extinction coefficient for the ensemble of particles in bulk material can be expressed by

$$\mu_{\text{ext}} = \mu_a + \mu_s. \tag{2.25}$$

This definition is useful when considering the extent of interaction a photon has with a material and in the Beer's law to describes the total attenuation of light through an axial distance z through a material [17]:

$$\Phi(z) = \Phi_0 \exp^{-\mu_{\text{ext}} z}, \tag{2.26}$$

where $\Phi(z)$ is the fluence rate (Wcm^{-2}) as a function of the distance and Φ_0 is the incident fluence rate on the material.

2.1.6 Transport mean free path

Since the work presented in this thesis considers materials with varying degrees of scattering, it is useful to use length scales in terms of the total number of scattering and absorption events experienced by light propagating through the material and remain agnostic to material thickness.

The scattering mean free path is reciprocal of the scattering coefficient $\ell_s = 1/\mu_s$ and describes the average distance light propagates in a medium before it interacts via a scattering event. Note that this is not the same as the distance between physical scattering particles in the material. The scattering coefficient μ_s (Eq. (2.16)) is defined by the total scattering cross section σ_s per volume. The scattering cross section (Eq. (2.15)) is the effective cross section area of the particle compared to the physical geometrical cross section when considering its scattering efficiency Q_s.

However, as discussed in the previous section, scattering is not exclusively isotropic, and it is important to consider the preferred scattering direction of the material. For example, a material for which there is a high probability of forward scattering will cause an incident light beam to retain information about its initial direction of propagation over a greater distance compared with an isotropic scattering material. In this case, the distance between scattering events is still ℓ_s, but the distance over which the direction of propagation of light is randomised is determined by the reduced scattering mean free path $\ell'_s = 1/(1-g)\mu_s = 1/\mu'_s$, which also accounts for the anisotropy of scattering from each particle. Figure (2.1) illustrates this equivalent path length for an isotropic and anisotropic example. In the isotropic case the length scale over which the direction is randomised is simply the scattering mean free path $\ell_s = \ell'_s$, whereas in the anisotropic case, the photon propagates further before losing memory of its initial direction. Therefore the reduced scattering mean free path is a more general length scale of material scattering statistics, which can encompass both isotropic and anisotropic scattering materials.

The (reduced) transport mean free path (TMFP) is yet another length scale, which combines the effects of both scattering and absorption and describes the average propagation length of light before it is either absorbed or loses information about its initial direction of propagation:

$$\boxed{\ell^* = \frac{1}{(\mu_a + \mu'_s)}.} \tag{2.27}$$

This parameter is the reciprocal of the extinction coefficient and is particularly useful when modelling the attenuation of light in a material. Moreover, when in the diffuse regime (i.e., $\mu_s \gg \mu_a$) and the material is not back-scattering (i.e., $g \geq 0$), such that $\mu'_s \leq \mu_s$, then the transport mean free path essentially describes the length scale travelled by a photon before its propagation direction is randomised. When light propagates through a thickness much greater than the transport mean free path length, $L \gg \ell^*$, then this is considered the diffusive regime.

Since the phenomenon of randomisation of the propagation direction of light is the cause of

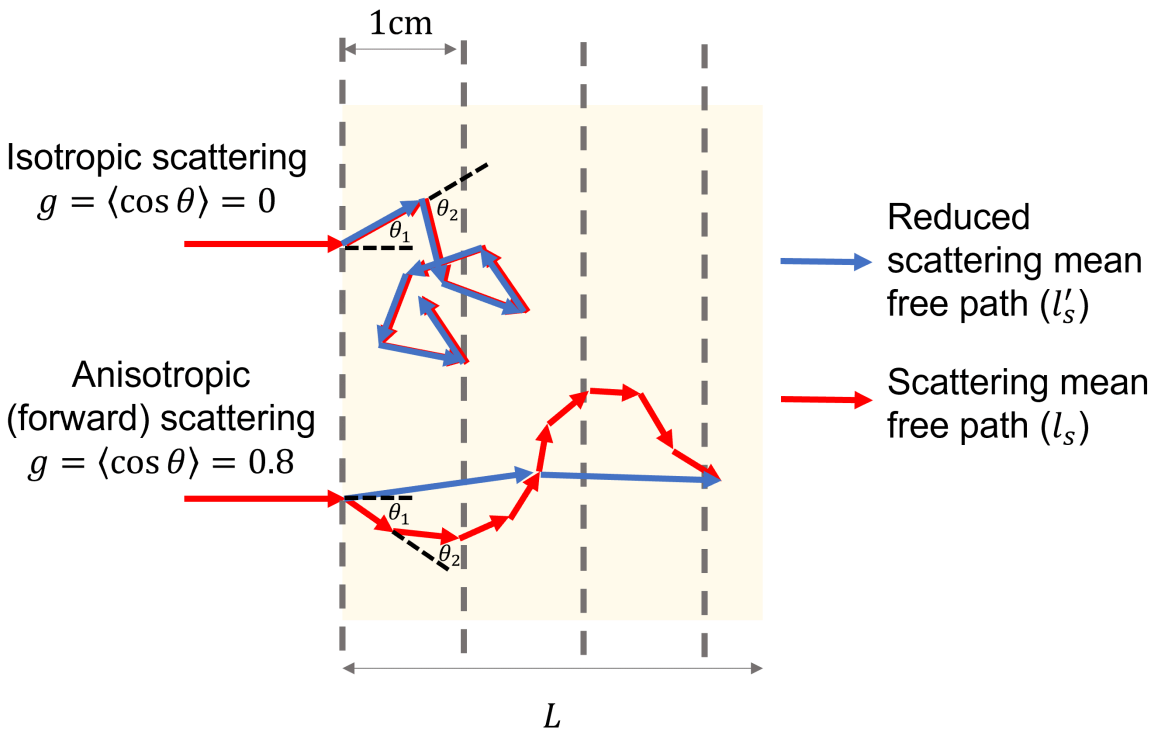


Figure 2.1: An illustration comparing the scattering mean free path (ℓ_s) (red arrows) and the reduced scattering mean free path (ℓ_s') (blue arrows) for an isotropic scattering material (upper) and a forward scattering (lower) material. The reduced transport mean free path is an equivalent scattering length which represents the distance of propagation experienced by a photon until it loses memory of its initial propagation direction.

the degradation to direct image quality, the number of TMFPs in the highly diffuse regime is an indicator of the level of imaging information corruption due to the random nature of light scattering.

2.2 Radiative transfer equation

A mathematically rigorous approach to photon transport in scattering media can be derived from Maxwell's equations which includes the effects of diffraction and interference effects when combining scattered EM-waves from multiple scattering particles. However, this approach is mathematically complicated for large numbers of scattering particles and a more tractable approach is to use radiative transport theory, which assumes that the energy distribution from each scattered wave is independent. This mitigates the problem of calculating the interaction of all EM fields, and allows for a simpler model which sums the scattered power contributions [26]. Although this alternative approach neglects the effects of polarisation, coherence and non-linear susceptibility, the context of this thesis is to investigate linear materials in the highly diffusive regime, in which coherence effects such as speckle interference patterns have unresolvable sub-wavelength grain

sizes and polarisation states are completely randomised.

The radiative transfer equation (RTE) is a heuristic model used to describe the propagation of light through a volume element of material with absorption coefficient μ_a and scattering coefficient μ_s. The RTE still remains challenging to solve analytically for practical problems and further approximations are made about photon transport in the diffusive regime (i.e., where $\mu_s \gg \mu_a$ as explained in Section (2.1.3)) to simplify the equation. The result of these assumptions leads to the photon diffusion approximation which has a Green's function solution that enables computationally efficient simulations of practical experimental configurations.

In this section, the RTE is derived from energy conservation, from which the photon diffusion approximation is subsequently derived. The following derivations in this thesis are an adaptation from Wang and Wu [17], which is a modern interpretation of the derivation by Ishimaru [22].

2.2.1 Derivation of the radiative transfer equation

Consider the volume element of scattering media pictured in figure (2.2), there are four possible physical processes which contribute to a change in radiant energy of a light beam within a solid angle $d\Omega$ around propagation direction $\hat{s}$:

1. *Divergence*: unless a beam of light is collimated or converging, it will diverge and spread radially outwards causing a loss in energy within solid angle $d\Omega$. This contribution is present even in vacuum.

2. *Extinction*: photons can be absorbed by the material or scattered out of the solid angle $d\Omega$.

3. *Scattering*: photons travelling in direction $\hat{s}'$ can be scattered into $d\Omega$ around direction $\hat{s}$.

4. *Source*: a source of photons inside the volume element can add to the radiant energy.

Each of the contributions above can be combined to find an expression for flow of energy through a material. Expressed mathematically, the radiant energy dE per time dt transported through area dA and within solid angle $d\Omega$ is

$$dE = L(\vec{r}, \hat{s}, t)(\hat{s} \cdot \hat{n})\, dA\, d\Omega\, dt. \tag{2.28}$$

Radiance $L(\vec{r}, \hat{s}, t)$ ($\mathrm{Wm^{-2}sr^{-1}}$) describes the energy flow per second per area per solid angle where $\vec{r} = \langle x, y, z \rangle$ is position in space, $\hat{s} = \langle \sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta \rangle$ is a unit vector describing direction of flow parameterised by spherical coordinates, and $\hat{n}$ is the unit vector normal to the area element. The scalar product $\hat{s} \cdot \hat{n}$ calculates the cosine angle between the direction $\hat{s}$ of the radiance with the surface normal $\hat{n}$ to determine the contribution of the energy flow through the surface.

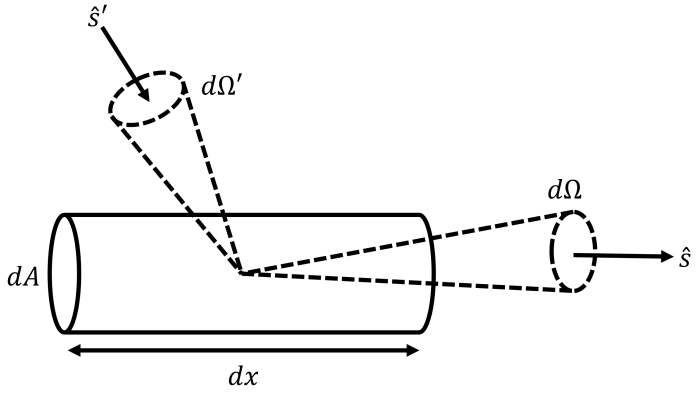


Figure 2.2: Scattering of light from $d\Omega'$ around direction $\hat{s}'$ into $d\Omega$ around direction $\hat{s}$ in a differential volume element (adapted from Ishimaru [22]).

Divergence term

The differential power dP_{div} is determined by the rate of change of the radiance traveling out of (divergence), or into (convergence), the differential volume or differential solid angle $d\Omega$:

$$dP_{\text{div}} = \frac{\partial L(\vec{r},\hat{s},t)}{\partial x} d\Omega dV, \tag{2.29}$$

where $dV = dAdx$ is the differential volume. More generally, the three-dimensional gradient vector projected onto the direction $\hat{s}$ can be used to express the divergence term by

$$dP_{\text{div}} = \hat{s} \cdot \nabla L(\vec{r},\hat{s},t)\, d\Omega\, dV, \tag{2.30}$$

where $\nabla = \left\langle \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right\rangle$ is the del operator.

Extinction term

The extinction of photons refers to the loss in energy due to scattering out of the solid angle and absorption within the volume element. The differential power dP_{ext} lost due to extinction is given by

$$dP_{\text{ext}} = (\mu_{\text{ext}} dx) L(\vec{r},\hat{s},t)\, dA\, d\Omega, \tag{2.31}$$

where $\mu_{\text{ext}} = (\mu_a + \mu_s)$ describes the combined number of scattering and absorption events per cm, as described in Section (2.1.5).

Scattering term

There will be an addition of power in the volume element due to scattering of photons from solid angle $d\Omega'$ around directions $\hat{s}'$ into solid angle $d\Omega$ around $\hat{s}$.

To determine the contribution from all possible incoming angles requires integrating the incoming radiance over all possible incoming solid angles and so the differential power gain due to scattering can be expressed by

$$dP_{\text{scat}} = (N_s dV)\left[\int_{4\pi} L(\vec{r},\hat{s},t)P(\hat{s}',\hat{s})\sigma_s\, d\Omega'\right] d\Omega, \tag{2.32}$$

where N_s is the number density of scatterers and σ_s is the scattering cross section. For identical scattering particles the scattering coefficient can be substituted for the scattering coefficient $\mu_s = N_s\sigma_s$. Note that in the case of non-identical scattering particles, the more general definition in equation (2.16) would be used instead. The introduction of a probability density function (PDF) $P(\hat{s}',\hat{s})$ describes how likely light is to scatter from direction $\hat{s}'$ to $\hat{s}$. In the context of biological optics, the assumption can be made that the scattering PDF depends only on the cosine of the angle of scattering, and can subsequently be written $P(\hat{s}'\cdot\hat{s})$ as discussed in Section (2.1.3).

The scattering term in equation (2.32) can then be rewritten:

$$dP_{\text{scat}} = (\mu_s dV)\left[\int_{4\pi} L(\vec{r},\hat{s},t)P(\hat{s}'\cdot\hat{s})\, d\Omega'\right] d\Omega. \tag{2.33}$$

Source term

A source $S(\vec{r},\hat{s},t)$ ($\text{Wm}^{-3}\text{sr}^{-1}$) within the differential volume may arise from e.g., fluorescence, bioluminescence or a fibre-optic guided source embedded in the material and will add to the power in the volume element within the solid angle element. The source term of differential power is given by

$$dP_{\text{src}} = S(\vec{r},\hat{s},t)dV\, d\Omega. \tag{2.34}$$

Combining the terms leads to the energy conservation equation:

$$dP_{\text{total}} = -dP_{\text{div}} - dP_{\text{ext}} + dP_{\text{scat}} + dP_{\text{src}}, \tag{2.35}$$

where the differential change in total power dP_{total} can be written in terms of radiance as

$$dP_{\text{total}} = c^{-1}\frac{\partial L(\vec{r},\hat{s},t)}{\partial t} dV\, d\Omega. \tag{2.36}$$

The introduction of the speed of light factor c is required since radiance is defined as energy per second per area per solid angle ($\text{Wm}^{-2}\text{sr}^{-1}$) and dividing by speed gives energy per volume per solid angle ($\text{Jm}^{-3}\text{sr}^{-1}$) which, when taking the derivative with respect to time, gives the units of power per volume per solid angle ($\text{Wm}^{-3}\text{sr}^{-1}$). Note that this is the speed of light in the material $c = \frac{c_0}{n}$ where c_0 is the speed of light in vacuum and n is refractive index of the material.

Substituting each contribution described above gives an expression for the radiative transport equation:

$$\boxed{c^{-1}\frac{\partial L(\vec{r},\hat{s},t)}{\partial t} = -\hat{s}\cdot\nabla L(\vec{r},\hat{s},t) - \mu_{\text{ext}} L(\vec{r},\hat{s},t) + \mu_s \int_{4\pi} L(\vec{r},\hat{s}',t) P(\hat{s}'\cdot\hat{s})\, d\Omega' + S(\vec{r},\hat{s},t).} \tag{2.37}$$

2.3 Diffusion approximation

Radiance is a general term which can be used to derive other useful physical quantities which are easier to measure in practice such as fluence rate (also known as intensity) $\Phi(\vec{r},t)$ (Wm^{-2}) which describes the total radiant energy flowing per second per area integrated over solid angle:

$$\Phi(\vec{r},t) = \int_{4\pi} L(\vec{r},\hat{s},t)\, d\Omega, \tag{2.38}$$

and current density $\vec{J}(\vec{r},t)$ (Wm^{-2}), which is the vector equivalent of fluence rate:

$$\vec{J}(\vec{r},t) = \int_{4\pi} L(\vec{r},\hat{s},t)\hat{s}\, d\Omega. \tag{2.39}$$

Since radiance depends on direction, each unit direction vector in this integration will be weighted by a radiance value. The result is a vector field which does not depend on direction but points in the direction of net difference in flow of energy. If the flow of energy is equal in all directions (i.e., isotropic) then the integral, and hence current density, will be zero. As shown later in this section, fluence rate and current density will be useful quantities for describing isotropic and anisotropic diffusion respectively. However, the goal of the diffusion approximation is to rewrite the RTE (2.37) in terms of fluence rate only. This reduces the problem of solving the RTE for a vector field with 6 variables (x,y,z,θ,φ,t), to solving for a scalar field with only 4 variables (x,y,z,t). There are two assumptions in the diffusion approximation:

1. *Directional broadening*: in the diffusive regime there is very little absorption events compared with scattering events (i.e., $\mu_s \gg \mu_a$). This means that after many scattering events, the radiance will be close to isotropic.

2. *Temporal broadening*: in materials where scattering dominates, each photon in a pulse of light incident on the material will experience a variety of different path lengths. This results in a broadening of the pulse in time. Concretely, this assumption states that a fractional change in current density over one transport mean free path must be very small (much less than one).

Both of the assumptions can be represented by the condition that $\mu_s' \gg \mu_a$.

Deriving a scalar differential equation in terms of only fluence rate which is valid for the diffusive regime involves five major steps:

1. Using the directional broadening assumption, the expression for radiance can be expanded using spherical harmonics and expressed in terms of isotropic and anisotropic contributions.

2. Integrating the radiance terms in the RTE (2.37) over solid angle as in equation (2.38) derives a scalar differential equation in terms of fluence rate and current density.

3. To write the equation in terms of only fluence rate, an expression for current density can be determined by multiplying each radiance term in the RTE by $\hat{s}$ and integrating over solid angle as in equation (2.39).

4. The resulting vector differential equation can be rearranged for current density in terms of fluence rate using the temporal broadening assumption and Fick's first law of diffusion.

5. The current density can be substituted into the scalar differential equation derived in step (2) which yields the photon diffusion approximation to the RTE.

Each of these steps will now be presented in more details in the following subsections, which will results in the solution the photon diffusion approximation to the RTE in terms of fluence rate. Although this derivation closely follows Chapter 5 of Wang and Wu [17], the reasoning for each step is my own interpretation, and many of the omitted mathematical steps left as an exercise to the reader have been written explicitly in this thesis adaptation.

This thesis can be understood without understanding how the diffusion approximation is derived (skip to Section (2.3.6)). However some key assumptions and parameters introduced in the following derivation will help to give context to the restricted regimes where this equation is accurate for modelling photon diffusion.

2.3.1 Expansion of radiance (directional broadening assumption)

The radiance is assumed to be mostly isotropic after many scattering events. This can be mathematically expressed by expanding radiance scalar field as the superposition of the spherical harmonics where only the isotropic and first order anisotropic terms are used:

$$L(\vec{r},\hat{s},t) \simeq \sum_{n=0}^{1}\sum_{m=-n}^{n} L_{n,m}(\vec{r},t)Y_{n,m}(\hat{s}), \tag{2.40}$$

where $L_{n,m}$ are the expansion coefficients of radiance and $Y_{n,m}$ are the spherical harmonics which form an orthonormal basis set:

$$\int_{4\pi} Y_{n,m}(\hat{s}) Y^*_{n',m'}(\hat{s})\, d\Omega = \delta_{nn',mm'}. \tag{2.41}$$

The spherical isotropic term is denoted by $n = 0$ and $m = 0$ and the first 3 isotropic terms are $n = 1$ and $m = 0 \pm 1$ which determine the Legendre polynomials and normalisation factors of each basis function:

$$Y_{0,0}(\theta, \varphi) = \frac{1}{\sqrt{4\pi}}, \tag{2.42}$$

$$Y_{1,-1}(\theta, \varphi) = \sqrt{\frac{3}{8\pi}} \sin\theta e^{-i\varphi}, \tag{2.43}$$

$$Y_{1,0}(\theta, \varphi) = \sqrt{\frac{3}{4\pi}} \cos\theta, \tag{2.44}$$

$$Y_{1,1}(\theta, \varphi) = -\sqrt{\frac{3}{8\pi}} \sin\theta e^{i\varphi}. \tag{2.45}$$

Using equation (2.38), fluence rate can be written in terms of spherical harmonics by

$$\begin{aligned}
\Phi(\vec{r}, t) &= \sum_{n=0}^{1} \sum_{m=-n}^{n} \int_{4\pi} L_{m,n}(\vec{r}, t) Y_{m,n}(\hat{s})\, d\Omega, \\
&= \int_0^{\pi} \int_0^{2\pi} L_{0,0}(\vec{r}, t) Y_{0,0}(\hat{s}) \sin\theta\, d\varphi\, d\theta, \\
&= L_{0,0}(\vec{r}, t) Y_{0,0}(\hat{s}) \left(\int_0^{\pi} \int_0^{2\pi} \sin\theta\, d\varphi\, d\theta \right), \\
&= L_{0,0}(\vec{r}, t) Y_{0,0}(\hat{s}) (4\pi),
\end{aligned} \tag{2.46}$$

where the $n = 1$ terms are zero when integrating over solid angle.

Likewise, the current density as defined in equation (2.39) can also be rewritten in spherical harmonics. Firstly, the radiance expansion must be multiplied by the unit vector $\hat{s}$. Using the definitions of spherical harmonics above and exponential expressions for sine and cosine, the unit vector can be expressed by

$$\begin{aligned}
\hat{s} &= \langle \sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta \rangle, \\
&= \left\langle \sin\theta \left(\frac{e^{i\varphi} + e^{-i\varphi}}{2} \right), \sin\theta \left(\frac{e^{i\varphi} - e^{-i\varphi}}{2i} \right), \cos\theta \right\rangle, \\
&= \left\langle \frac{1}{2} \sin\theta e^{i\varphi} + \sin\theta e^{-i\varphi}, -\frac{1}{2} i \left(\sin\theta e^{i\varphi} - \sin\theta e^{-i\varphi} \right), \cos\theta \right\rangle,
\end{aligned}$$

$$
\begin{aligned}
&= \left\langle \frac{1}{2}\sqrt{\frac{8\pi}{3}}\left(-Y_{1,1}(\hat{s})+Y_{1,-1}(\hat{s})\right), -\frac{1}{2}\sqrt{\frac{8\pi}{3}}i\left(-Y_{1,1}(\hat{s})-Y_{1,-1}(\hat{s})\right), \sqrt{\frac{4\pi}{3}}Y_{1,0}(\hat{s}) \right\rangle, \\
&= \left\langle \sqrt{\frac{2\pi}{3}}\left(Y_{1,-1}(\hat{s})-Y_{1,1}(\hat{s})\right), \sqrt{\frac{2\pi}{3}}i\left(Y_{1,1}(\hat{s})+Y_{1,-1}(\hat{s})\right), \sqrt{\frac{4\pi}{3}}Y_{1,0}(\hat{s}) \right\rangle, \\
&= \sqrt{\frac{2\pi}{3}}\left\langle \left(Y_{1,-1}(\hat{s})-Y_{1,1}(\hat{s})\right), i\left(Y_{1,-1}(\hat{s})+Y_{1,1}(\hat{s})\right), \sqrt{2}Y_{1,0}(\hat{s}) \right\rangle. \qquad (2.47)
\end{aligned}
$$

Using the following symmetry property of spherical harmonics:

$$
Y^*_{n,m}(\theta,\varphi) = (-1)^m Y_{n,-m}(\theta,\varphi), \qquad (2.48)
$$

where * denotes the complex conjugate, an equivalent expression for equation (2.47) can be determined:

$$
\begin{aligned}
\hat{s} &= \sqrt{\frac{2\pi}{3}}\left\langle \left(-Y^*_{1,1}(\hat{s})+Y^*_{1,-1}(\hat{s})\right), i\left(-Y^*_{1,1}(\hat{s})-Y^*_{1,-1}(\hat{s})\right), \sqrt{2}Y^*_{1,0}(\hat{s}) \right\rangle, \\
&= \sqrt{\frac{2\pi}{3}}\left\langle \left(Y^*_{1,-1}(\hat{s})-Y^*_{1,1}(\hat{s})\right), -i\left(Y^*_{1,1}(\hat{s})+Y^*_{1,-1}(\hat{s})\right), \sqrt{2}Y^*_{1,0}(\hat{s}) \right\rangle. \qquad (2.49)
\end{aligned}
$$

This form of the unit vector allows for a trivial derivation of the current density in terms of the expanded radiance by multiplying (2.40) with (2.49), substituting into (2.39) and using the orthonormal property of basis functions in (2.41):

$$
\vec{J}(\vec{r},t) = \sqrt{\frac{2\pi}{3}}\left\langle \left(L_{1,-1}(\vec{r},t)-L_{1,1}(\vec{r},t)\right), -i\left(L_{1,1}(\vec{r},t)+L_{1,-1}(\vec{r},t)\right), \sqrt{2}L_{1,0}(\vec{r},t) \right\rangle. \qquad (2.50)
$$

The component of current density in the direction $\hat{s}$ the scalar product between equations (2.47) and (2.50):

$$
\begin{aligned}
\vec{J}\cdot\hat{s} &= \frac{2\pi}{3}\left[\left(L_{1,-1}-L_{1,1}\right)\left(Y_{1,-1}-Y_{1,1}\right)+\left(L_{1,1}+L_{1,-1}\right)\left(Y_{1,-1}+Y_{1,1}\right)+2L_{1,0}Y_{1,0}\right], \\
&= \frac{2\pi}{3}\left[2L_{1,-1}Y_{1,-1}+2L_{1,1}Y_{1,1}+2L_{1,0}Y_{1,0}\right], \\
&= \frac{4\pi}{3}\sum_{m=-1}^{1} L_{1,m}Y_{1,m} \qquad (2.51)
\end{aligned}
$$

where the dependent variables are omitted to avoid cluttered presentation. Rearranging (2.51) and combining with the isotropic term derived in (2.46) gives an expression for the radiance in the diffusive regime in terms of fluence and current density:

$$
\boxed{L(\vec{r},\hat{s},t) = \frac{1}{4\pi}\Phi(\vec{r},t)+\frac{3}{4\pi}\vec{J}\cdot\hat{s}} \qquad (2.52)
$$

2.3.2 Scalar differential equation

To simplify the RTE and write it in terms of more manageable physical quantities such as fluence rate and current density, the expansion for radiance derived in (2.52) can be substituted into (2.37) and integrated over solid angle. This integration is best described on a term-by-term basis.

Left term

The left hand side of the RTE in (2.37) can be written as fluence rate as defined in equation (2.38):

$$\int_{4\pi} c^{-1} \frac{\partial L(\vec{r},\hat{s},t)}{\partial t}\, d\Omega = c^{-1} \frac{\partial \Phi(\vec{r},t)}{\partial t}. \tag{2.53}$$

First right term

The first term on the right of equation (2.37) integrated over solid angle can be written:

$$\int_{4\pi} \hat{s}\cdot\nabla L(\vec{r},\hat{s},t)\, d\Omega = \int_{4\pi} \nabla\cdot\left(L(\vec{r},\hat{s},t)\hat{s}\right)\, d\Omega, \tag{2.54}$$

where $\hat{s}\cdot\nabla L = \nabla\cdot(\hat{s}L) - L(\nabla\cdot\hat{s})$ using the identity (for an arbitrary function $f(x,y,z)$ and vector $\vec{a} = \langle a_1, a_2, a_3\rangle$ in Cartesian coordinates):

$$\begin{aligned}
\nabla\cdot\left(f\vec{a}\right) &= \left(\frac{\partial f}{\partial x}a_1\right) + \left(\frac{\partial f}{\partial y}a_2\right) + \left(\frac{\partial f}{\partial z}a_3\right), \\
&= \left(\frac{\partial f}{\partial x}a_1 + f\frac{\partial a_1}{\partial x}\right) + \left(\frac{\partial f}{\partial y}a_2 + f\frac{\partial a_2}{\partial y}\right) + \left(\frac{\partial f}{\partial z}a_3 + f\frac{\partial a_3}{\partial z}\right), \\
&= f\left(\frac{\partial a_1}{\partial x} + \frac{\partial a_2}{\partial y} + \frac{\partial a_2}{\partial z}\right) + \left(\frac{\partial f}{\partial x}a_1 + \frac{\partial f}{\partial y}a_2 + \frac{\partial f}{\partial z}a_3\right), \\
&= f(\nabla\cdot\vec{a}) + \vec{a}\cdot(\nabla f),
\end{aligned} \tag{2.55}$$

and $\nabla\cdot\hat{s} = 0$, since it does not depend on position and is a constant vector describing direction[1]:

$$\begin{aligned}
\nabla\cdot\hat{s} &= \frac{\partial}{\partial x}(\sin\theta\cos\varphi) + \frac{\partial}{\partial y}(\sin\theta\sin\varphi) + \frac{\partial}{\partial z}\cos\theta, \\
&= 0.
\end{aligned} \tag{2.56}$$

The intuition for equations (2.54), (2.55) and (2.56) requires revisiting the explanation of the divergence term in Section (2.2.1). Radiance depends on the variables (x,y,z,θ,φ,t), and the gradient of the radiance *w.r.t* position $\vec{r} = \langle x,y,z\rangle$ projected onto direction $\hat{s}$, produces a scalar

[1] Although in the same coordinate system, the direction unit vector $\hat{s}$ is parameterised by only θ and φ. Explicitly, the direction unit vector and position vector can be written as $\hat{s} = \sin\theta\cos\varphi\hat{i} + \sin\theta\sin\varphi\hat{j} + \cos\theta\hat{k}$ and $\vec{r} = x\hat{i} + y\hat{j} + z\hat{k}$. This means $\hat{s}$ remains constant for all positions $\vec{r}$ (i.e., they are independent of each other). Therefore, the $\nabla\cdot\hat{s}$ term, which arises solely from the identity in (2.55), describes the total gradient in all dimensions of a constant vector, which must be zero.

field describing the rate of change of power due to divergence of a beam out of a differential volume dV or solid angle $d\Omega$ at every position $\vec{r}$. This describes the left term of equation (2.54), where integrating yields another scalar field representing the total power loss due to divergence over all directions rather than only direction $\hat{s}$.

The right term of (2.54) uses the identity described in (2.55) to describe an equivalent scalar field produced when taking the sum of the rate of change in each spatial dimension (i.e., the divergence operator $\nabla\cdot$) of radiance in direction $\hat{s}$. Again, integrating finds the total contribution for all angles.

The integration of radiance in direction $\hat{s}$ over all angles yields a vector field which describes power per area with a net direction of flow. This is the definition of current density in (2.39) and hence the first right term of the integrated RTE can be written:

$$\int_{4\pi} \nabla \cdot \left(L(\vec{r}, \hat{s}, t)\hat{s} \right) d\Omega = \nabla \cdot \vec{J}(\vec{r}, t). \tag{2.57}$$

Second right term

Similar to the left term, the second right term can be written as

$$\mu_{\text{ext}} \int_{4\pi} L(\vec{r}, \hat{s}, t) = \mu_{\text{ext}} \Phi(\vec{r}, t), \tag{2.58}$$

using equation (2.38).

Third right term

Substituting radiance with its expanded form in (2.52), the third right term can be written as the sum of two integrals:

$$\mu_s \int_{4\pi} \int_{4\pi} L(\vec{r}, \hat{s}', t) P(\hat{s}' \cdot \hat{s})\, d\Omega'\, d\Omega = \frac{\mu_s}{4\pi} \int_{4\pi} \int_{4\pi} \left(\Phi(\vec{r}, t) + 3\vec{J}(\vec{r}, t) \cdot \hat{s}' \right) P(\hat{s}' \cdot \hat{s})\, d\Omega'\, d\Omega, \tag{2.59}$$

where the first integral is

$$\begin{aligned} \frac{\mu_s}{4\pi} \int_{4\pi} \int_{4\pi} \Phi(\vec{r}, t) P(\hat{s}' \cdot \hat{s})\, d\Omega'\, d\Omega &= \frac{\mu_s}{4\pi} \Phi(\vec{r}, t) \int_{4\pi} \int_{4\pi} P(\hat{s}' \cdot \hat{s})\, d\Omega'\, d\Omega, \\ &= \frac{\mu_s}{4\pi} \Phi(\vec{r}, t) \int_{4\pi} (1)\, d\Omega, \\ &= \frac{\mu_s}{4\pi} \Phi(\vec{r}, t)(4\pi), \\ &= \mu_s \Phi(\vec{r}, t), \end{aligned} \tag{2.60}$$

and the second integral is

$$\begin{aligned}
\int_{4\pi}\int_{4\pi}\left(\vec{J}(\vec{r},t)\cdot\hat{s}'\right)P(\hat{s}'\cdot\hat{s})\,d\Omega'\,d\Omega &= \int_{4\pi}\int_{4\pi}\left(|\vec{J}(\vec{r},t)||\hat{s}'|\cos\theta'\right)P(\hat{s}'\cdot\hat{s})\,d\Omega'\,d\Omega,\\
&= |\vec{J}(\vec{r},t)|\int_{4\pi}\int_{4\pi}\left((1)\cos\theta'\right)P(\hat{s}'\cdot\hat{s})\,d\Omega'\,d\Omega,\\
&= |\vec{J}(\vec{r},t)|\int_{4\pi}\cos\theta'\left[\int_{4\pi}P(\hat{s}'\cdot\hat{s})\,d\Omega\right]d\Omega',\\
&= |\vec{J}(\vec{r},t)|\int_{0}^{\pi}\int_{0}^{2\pi}\cos\theta'\sin\theta'\,d\varphi'\,d\theta',\\
&= |\vec{J}(\vec{r},t)|2\pi\int_{0}^{\pi}\cos\theta'\sin\theta'\,d\theta',\\
&= |\vec{J}(\vec{r},t)|2\pi\left[\frac{\cos^2\theta'}{2}\right]_0^{\pi},\\
&= |\vec{J}(\vec{r},t)|\pi\left[\cos^2\theta'\right]_0^{\pi},\\
&= 0. \qquad (2.61)
\end{aligned}$$

Note that the integrated probability density function $P(\hat{s}'\cdot\hat{s})$ is equal to 1 by definition, $d\Omega = \sin\theta d\varphi d\theta$ and $\vec{a}\cdot\vec{b} = |a||b|\cos\theta$. In this case, $\vec{J}(\vec{r},t)$ is aligned along the z-axis and has no φ component for simplicity. Therefore, the third term is given by the solution to the first integral:

$$\mu_s\int_{4\pi}\int_{4\pi}L(\vec{r},\hat{s}',t)P(\hat{s}\cdot\hat{s})\,d\Omega'\,d\Omega = \mu_s\Phi(\vec{r},t). \qquad (2.62)$$

Fourth right term

The final term in equation (2.37) is the source term. The source is assumed to be inside the material at the location where the scattering becomes isotropic. Under this assumption the source term can be expressed as only the isotropic component in spherical coordinates $S_{0,0}(\vec{r},t)Y_{0,0}(\hat{s}) = \frac{S(\vec{r},t)}{4\pi}$ ($\mathrm{Wm^{-2}}$) similar to (2.46). After integration the source term is simply

$$\begin{aligned}
\int_{4\pi}S(\vec{r},\hat{s},t)d\Omega &= \int_{4\pi}\frac{S(\vec{r},t)}{4\pi}\,d\Omega,\\
&= \frac{S(\vec{r},t)}{4\pi}\int_{4\pi}d\Omega,\\
&= \frac{S(\vec{r},t)}{4\pi}(4\pi),\\
&= S(\vec{r},t). \qquad (2.63)
\end{aligned}$$

Collecting terms

Collecting all the terms together, the scalar differential equation derived by substituting the ex-

panded radiance in (2.52) into the RTE (2.37) and integrating gives

$$\boxed{c^{-1}\frac{\partial\Phi(\vec{r},t)}{\partial t} = -\nabla\cdot\vec{J}(\vec{r},t) - \mu_{\text{ext}}\Phi(\vec{r},t) + \mu_s\Phi(\vec{r},t) + S(\vec{r},t).} \tag{2.64}$$

2.3.3 Vector differential equation

To write the differential equation (2.64) above in terms of only fluence rate, an expression for $\vec{J}(\vec{r},t)$ must be determined and substituted into the first right term. To do this, the process of deriving (2.64) is repeated but every radiance term must also be multiplied by the unit vector $\hat{s}$ before integrating as per the definition of current density in (2.39). As before, this integration is best presented on a term-by-term basis.

Left term

Using the definition of current density in (2.39), the left term is

$$\int_{4\pi} c^{-1}\frac{\partial L(\vec{r},\hat{s},t)}{\partial t}\hat{s}\,d\Omega = c^{-1}\frac{\partial\vec{J}(\vec{r},t)}{\partial t}. \tag{2.65}$$

First right term

The first right term of the RTE (2.37) can be split into two integrals by substituting the expansion of radiance (2.52):

$$\int_{4\pi}\left(\hat{s}\cdot\nabla L(\vec{r},\hat{s},t)\right)\hat{s}\,d\Omega = \frac{1}{4\pi}\int_{4\pi}\left(\hat{s}\cdot\nabla\Phi(\vec{r},t)\right)\hat{s}\,d\Omega + \frac{3}{4\pi}\int_{4\pi}\left[\hat{s}\cdot\nabla\left(\vec{J}(\vec{r},t)\cdot\hat{s}\right)\right]\hat{s}\,d\Omega, \tag{2.66}$$

where the first integral is

$$\begin{aligned}\int_{4\pi}(\hat{s}\cdot\nabla\Phi)\hat{s}\,d\Omega &= \int_{4\pi}(\Phi_x\sin\theta\cos\varphi + \Phi_y\sin\theta\sin\varphi + \Phi_z\cos\theta)\hat{s}\,d\Omega,\\ &= \int_{4\pi}F\hat{s}\,d\Omega,\end{aligned} \tag{2.67}$$

Note that the variables of fluence rate have been omitted for brevity and $F = f(x,y,z,\theta,\varphi,t)$ represents the function on the right in the brackets and will be used for efficient note-keeping in the subsequent steps. The subscripts of the fluence rate denote the gradient in each direction $\left(\nabla\Phi = \left\langle\Phi_x,\Phi_y,\Phi_z\right\rangle\right)$. Addressing the integrals of the three dimensions in (2.67) individually, the integral in the first dimension is

$$\begin{aligned}\int_{4\pi}F(\sin\theta\cos\varphi)\,d\Omega &= \int_{4\pi}(\Phi_x\sin\theta\cos\varphi + \Phi_y\sin\theta\sin\varphi + \Phi_z\cos\theta)(\sin\theta\cos\varphi)\,d\Omega,\\ &= \int_{4\pi}\left(\Phi_x\sin^2\theta\cos^2\varphi + \Phi_y\sin^2\theta\sin\varphi\cos\varphi + \Phi_z\cos\theta\sin\theta\cos\varphi\right)d\Omega,\end{aligned}$$

$$
\begin{aligned}
&= \int_0^{\pi}\int_0^{2\pi} \left(\Phi_x \sin^3\theta\cos^2\varphi + \Phi_y \sin^3\theta \sin\varphi\cos\varphi ...\right.\\
&\quad \left. + \Phi_z \cos\theta\sin^2\theta\cos\varphi\right) d\varphi\, d\theta,\\
&= \Phi_x \int_0^{\pi} \left(\sin^3\theta\left[\frac{\varphi}{2} + \sin\varphi\cos\varphi\right]_0^{2\pi} + [0] + [0]\right) d\theta,\\
&= \Phi_x \int_0^{\pi} \sin^3\theta[\pi]\, d\theta,\\
&= \pi\Phi_x \left[\frac{1}{12}(\cos 3\theta - 9\cos\theta)\right]_0^{\pi},\\
&= \frac{4\pi}{3}\Phi_x. \qquad (2.68)
\end{aligned}
$$

Following the same process, the integral in the second dimension is

$$
\begin{aligned}
\int_{4\pi} F(\sin\theta\sin\varphi)\, d\Omega &= \int_{4\pi} (\Phi_x \sin\theta\cos\varphi + \Phi_y \sin\theta\sin\varphi + \Phi_z\cos\theta)(\sin\theta\sin\varphi)\, d\Omega,\\
&= \int_{4\pi} \left(\Phi_x \sin^2\theta\cos\varphi\sin\varphi + \Phi_y \sin^2\theta\sin^2\varphi + \Phi_z\cos\theta\sin\theta\sin\varphi\right) d\Omega,\\
&= \int_0^{\pi}\int_0^{2\pi} \left(\Phi_x \sin^3\theta\cos\varphi\sin\varphi + \Phi_y \sin^3\theta\sin^2\varphi ...\right.\\
&\quad \left. + \Phi_z\cos\theta\sin^2\theta\sin\varphi\right) d\varphi\, d\theta,\\
&= \Phi_y \int_0^{\pi} \left([0] + \sin^3\theta\left[\frac{\varphi}{2} + \sin\varphi\cos\varphi\right]_0^{2\pi} + [0]\right) d\theta,\\
&= \Phi_y \int_0^{\pi} \sin^3\theta[\pi]\, d\theta,\\
&= \pi\Phi_y \left[\frac{1}{12}(\cos 3\theta - 9\cos\theta)\right]_0^{\pi},\\
&= \frac{4\pi}{3}\Phi_y. \qquad (2.69)
\end{aligned}
$$

Finally the integral in the third dimension is

$$
\begin{aligned}
\int_{4\pi} F(\cos\theta)\, d\Omega &= \int_{4\pi} (\Phi_x \sin\theta\cos\varphi + \Phi_y \sin\theta\sin\varphi + \Phi_z\cos\theta)(\cos\theta)\, d\Omega,\\
&= \int_{4\pi} \left(\Phi_x \sin\theta\cos\varphi\cos\theta + \Phi_y \sin\theta\sin\varphi\cos\theta + \Phi_z\cos^2\theta\right) d\Omega,\\
&= \int_0^{\pi}\int_0^{2\pi} \left(\Phi_x \sin^2\theta\cos\varphi\cos\theta + \Phi_y \sin^2\theta\sin\varphi\cos\theta + \Phi_z\cos^2\theta\sin\theta\right) d\varphi\, d\theta,\\
&= \Phi_z \int_0^{\pi} \left([0] + [0] + \cos^2\theta\sin\theta[\varphi]_0^{2\pi}\right) d\theta,\\
&= \Phi_z \int_0^{\pi} \left(\cos^2\theta\sin\theta[2\pi]\right) d\theta,\\
&= 2\pi\Phi_z \left(\left[\frac{-\cos^3\theta}{3}\right]_0^{\pi}\right),
\end{aligned}
$$

$$= \frac{4\pi}{3}\Phi_z. \tag{2.70}$$

Combining the terms gives the result the first integral of (2.66):

$$\begin{aligned}\int_{4\pi} \left(\hat{s}\cdot\nabla\Phi(\vec{r},t)\right)\hat{s}\,d\Omega &= \left\langle \frac{4\pi}{3}\Phi_x, \frac{4\pi}{3}\Phi_y, \frac{4\pi}{3}\Phi_z \right\rangle, \\ &= \frac{4\pi}{3}\nabla\Phi. \end{aligned} \tag{2.71}$$

The second integral of (2.66) can be solved in a similar way. First the coordinate system is orientated such that $\vec{J}(\vec{r},t)$ is aligned with the $\vec{k}$ direction. In this case the scalar product can be expressed using only azimuthal angle by $\vec{J}(\vec{r},t)\cdot\hat{s} = |\vec{J}(\vec{r},t)||\hat{s}|\cos\theta$, where $|\hat{s}| = 1$ by definition. The integral can then be written as

$$\begin{aligned}\int_{4\pi}\left[\hat{s}\cdot\nabla\left(\vec{J}\cdot\hat{s}\right)\right]\hat{s}\,d\Omega &= \int_{4\pi}\left[\hat{s}\cdot\nabla\left(|\vec{J}|\cos\theta\right)\right]\hat{s}\,d\Omega, \\ &= \int_{4\pi}\left[\hat{s}\cdot\left\langle|\vec{J}|_x, |\vec{J}|_y, |\vec{J}|_z\right\rangle\cos\theta\right]\hat{s}\,d\Omega, \\ &= \int_{4\pi}\left[\left(|\vec{J}|_x\sin\theta\cos\varphi + |\vec{J}|_y\sin\theta\sin\varphi + |\vec{J}|_z\cos\theta\right)\cos\theta\right]\hat{s}\,d\Omega, \\ &= \int_{4\pi}\left[G\cos\theta\right]\hat{s}\,d\Omega, \end{aligned} \tag{2.72}$$

where the variables of current density have been omitted for brevity and again $G = g(x,y,z,\theta,\varphi,t)$ is used to represent the function on the right of the equation in curved brackets. The integral in φ evaluated along each dimension is the same as (2.68),(2.69) and (2.70). Using these results, the remaining integral over θ for each dimension of (2.72) can be evaluated. Integrating the first dimension gives

$$\begin{aligned}\int_{4\pi} G\cos\theta(\sin\theta\cos\varphi)\,d\Omega &= \pi|J|_x\int_0^\pi \sin^3\theta\cos\theta\,d\theta, \\ &= \pi|J|_x\left[\frac{\sin^4\theta}{4}\right]_0^\pi \\ &= 0. \end{aligned} \tag{2.73}$$

Similarly for the second dimension:

$$\begin{aligned}\int_{4\pi} G\cos\theta(\sin\theta\cos\varphi)\,d\Omega &= \pi|J|_y\int_0^\pi \sin^3\theta\cos\theta\,d\theta, \\ &= \pi|J|_y\left[\frac{\sin^4\theta}{4}\right]_0^\pi \\ &= 0. \end{aligned} \tag{2.74}$$

Finally, the integral along the third dimension is

$$\begin{aligned}\int_{4\pi} G\cos\theta(\sin\theta\cos\varphi)\,d\Omega &= 2\pi|J|_z\int_0^{\pi}\cos^3\theta\sin\theta\,d\theta,\\ &= 2\pi|J|_z\left[\frac{-\cos^4\theta}{4}\right]_0^{\pi}\\ &= 0.\end{aligned} \tag{2.75}$$

Collecting the results in each dimension, the second integral of (2.66) is $\vec{0} = \langle 0,0,0\rangle$ which leaves the result of the first integral as the result for the first right term of vector differential equation:

$$\int_{4\pi}\left(\hat{s}\cdot\nabla L(\vec{r},\hat{s},t)\right)\hat{s}\,d\Omega = \frac{1}{3}\nabla\Phi(\vec{r},t). \tag{2.76}$$

Second right term

Using the definition of current density in (2.39) the second right term is

$$\mu_{\text{ext}}\int_{4\pi} L(\vec{r},\hat{s},t)\hat{s} = \mu_{\text{ext}}\vec{J}(\vec{r},t). \tag{2.77}$$

Third right term

Similar to the derivation of the third right term for the scalar differential equation in Section (2.3.2), the expansion for radiance in (2.40) can be substituted into the third term of (2.37) and the double integral can be split into two terms:

$$\mu_s\int_{4\pi}\int_{4\pi}\left[L(\vec{r},\hat{s},t)P(\hat{s}'\cdot\hat{s})\right]\hat{s}\,d\Omega'\,d\Omega = \frac{\mu_s}{4\pi}\int_{4\pi}\int_{4\pi}\left[\left(\Phi(\vec{r},t)+3\vec{J}(\vec{r},t)\cdot\hat{s}'\right)P(\hat{s}'\cdot\hat{s})\right]\hat{s}\,d\Omega'\,d\Omega. \tag{2.78}$$

The first integral reduces to

$$\begin{aligned}\Phi(\vec{r},t)\int_{4\pi}\hat{s}\int_{4\pi}P(\hat{s}'\cdot\hat{s})\,d\Omega'\,d\Omega &= \Phi(\vec{r},t)\int_{4\pi}\hat{s}\,d\Omega,\\ &= \vec{0},\end{aligned} \tag{2.79}$$

since the integration of the PDF is $\int_{4\pi}P(\hat{s}'\cdot\hat{s})\,d\Omega = 1$ by definition and

$$\begin{aligned}\int_{4\pi}\hat{s}\,d\Omega &= \int_0^{\pi}\int_0^{2\pi}\langle\sin\theta\cos\varphi,\sin\theta\sin\varphi,\cos\theta\rangle\sin\theta\,d\varphi\,d\theta,\\ &= \int_0^{\pi}\int_0^{2\pi}\left\langle\sin^2\theta\cos\varphi,\sin^2\theta\sin\varphi,\cos\theta\sin\theta\right\rangle d\varphi\,d\theta,\\ &= \int_0^{\pi}\left\langle 0,0,\cos\theta\sin\theta\,[\varphi]_0^{2\pi}\right\rangle d\theta,\end{aligned}$$

$$
\begin{aligned}
&= \left\langle 0, 0, \left[-\frac{1}{2}\cos^2\theta \right]_0^{\pi} 2\pi \right\rangle, \\
&= \langle 0, 0, 0 \rangle, \\
&= \vec{0}.
\end{aligned} \tag{2.80}
$$

The second integral of (2.78) can be expressed as

$$
\int_{4\pi} \hat{s} \int_{4\pi} \left(\vec{J}(\vec{r},t) \cdot \hat{s}' \right) P(\hat{s}' \cdot \hat{s})\, d\Omega'\, d\Omega = \int_{4\pi} \left[\int_{4\pi} P(\hat{s}' \cdot \hat{s}) \hat{s}\, d\Omega \right] \left(\vec{J}(\vec{r},t) \cdot \hat{s}' \right) d\Omega'. \tag{2.81}
$$

The inner integral in square brackets can be expanded and solved using the vector triple product identity $\vec{a} \times (\vec{b} \times \vec{c}) = (\vec{a} \cdot \vec{c})\vec{b} - (\vec{a} \cdot \vec{b})\vec{c}$, where $\vec{a} = \vec{c} = \hat{s}'$ and $\vec{b} = \hat{s}$. In which case the unit vector can be written $\hat{s} = (\hat{s}' \cdot \hat{s})\hat{s}' + \hat{s}' \times (\hat{s} \times \hat{s}')$, where $(\hat{s}' \cdot \hat{s}') = 1$. Substituting this identity form of $\hat{s}$ into the inner integral of (2.81) gives

$$
\begin{aligned}
\int_{4\pi} P(\hat{s}' \cdot \hat{s}) \hat{s}\, d\Omega &= \hat{s}' \int_{4\pi} (\hat{s} \cdot \hat{s}') P(\hat{s}' \cdot \hat{s})\, d\Omega + \hat{s}' \times \left[\int_{4\pi} \hat{s} P(\hat{s}' \cdot \hat{s})\, d\Omega \right] \times \hat{s}', \\
&= \hat{s}' g,
\end{aligned} \tag{2.82}
$$

where the first term uses the definition of g as shown in (2.21) and the second term is zero since $P(\hat{s}' \cdot \hat{s})$ is symmetric about $\hat{s}$. This means that when $\hat{s} P(\hat{s}' \cdot \hat{s})$ is integrated over all possible directions $\hat{s}$, the summation of opposing vectors around $\hat{s}'$ cancel and the result is a vector which must be parallel with $\hat{s}'$. In general, the cross product of two vectors separated by angle θ is $\vec{a} \times \vec{b} = |\vec{a}||\vec{b}| \sin\theta$, hence the cross product is zero for two parallel vectors and the second term disappears.

This allows the right term of (2.78) to be rewritten:

$$
\int_{4\pi} \int_{4\pi} \left[\left(\Phi(\vec{r},t) + 3\vec{J}(\vec{r},t) \cdot \hat{s}' \right) P(\hat{s}' \cdot \hat{s}) \right] \hat{s}\, d\Omega'\, d\Omega = 3g \int_{4\pi} \hat{s}' \left(\vec{J}(\vec{r},t) \cdot \vec{s}' \right) d\Omega', \tag{2.83}
$$

where the right of the equation is of the same form of the integral in (2.71):

$$
\int_{4\pi} \hat{s}' \left(\vec{J}(\vec{r},t) \cdot \hat{s}' \right) d\Omega' = \frac{4\pi}{3} \vec{J}(\vec{r},t). \tag{2.84}
$$

Therefore the final form of (2.78) is:

$$
\begin{aligned}
\mu_s \int_{4\pi} \int_{4\pi} \left[L(\vec{r},\hat{s},t) P(\hat{s}' \cdot \hat{s}) \right] \hat{s}\, d\Omega'\, d\Omega &= \frac{\mu_s}{4\pi} \int_{4\pi} \int_{4\pi} \left[\left(\Phi(\vec{r},t) + 3\vec{J}(\vec{r},t) \cdot \hat{s}' \right) P(\hat{s}' \cdot \hat{s}) \right] \hat{s}\, d\Omega'\, d\Omega \\
&= \frac{\mu_s}{4\pi} \left[3g \left(\frac{4\pi}{3} \vec{J}(\vec{r},t) \right) \right], \\
&= \mu_s g \vec{J}(\vec{r},t).
\end{aligned} \tag{2.85}
$$

Fourth right term

The source is again defined as isotropic and by following the derivation in (2.46), is equal $S_{0,0}(\vec{r},t)Y_{0,0}(\hat{s}) = S(\vec{r},t)/4\pi$ (Wm^{-2}). However when multiplied by $\hat{s}$ this term disappears:

$$\begin{aligned}\int_{4\pi} S(\vec{r},\hat{s},t)\hat{s}d\Omega &= \frac{S(\vec{r},t)}{4\pi}\int_{4\pi}\hat{s}\,d\Omega \\ &= \vec{0},\end{aligned} \tag{2.86}$$

since the integration of the unit vector is $\int_{4\pi}\hat{s}\,d\Omega = \vec{0}$ as shown in (2.80).

Combining terms

When combining the terms above the following vector differential equation can be formulated:

$$c^{-1}\frac{\partial\vec{J}(\vec{r},t)}{\partial t} = -\frac{1}{3}\nabla\Phi(\vec{r},t) - \mu_{\text{ext}}\vec{J}(\vec{r},t) + \mu_s g\vec{J}(\vec{r},t). \tag{2.87}$$

Rearranging yields

$$\begin{aligned}
c^{-1}\frac{\partial\vec{J}(\vec{r},t)}{\partial t} + \frac{1}{3}\nabla\Phi(\vec{r},t) + \mu_{\text{ext}}\vec{J}(\vec{r},t) - \mu_s g\vec{J}(\vec{r},t) &= 0,\\
c^{-1}\frac{\partial\vec{J}(\vec{r},t)}{\partial t} + \frac{1}{3}\nabla\Phi(\vec{r},t) + \vec{J}(\vec{r},t)(\mu_{\text{ext}} - \mu_s g) &= 0,\\
c^{-1}\frac{\partial\vec{J}(\vec{r},t)}{\partial t} + \frac{1}{3}\nabla\Phi(\vec{r},t) + \vec{J}(\vec{r},t)(\mu_a + \mu_s - \mu_s g) &= 0,\\
c^{-1}\frac{\partial\vec{J}(\vec{r},t)}{\partial t} + \frac{1}{3}\nabla\Phi(\vec{r},t) + \vec{J}(\vec{r},t)\left(\mu_a + \mu_s(1-g)\right) &= 0,\\
c^{-1}\frac{\partial\vec{J}(\vec{r},t)}{\partial t} + \frac{1}{3}\nabla\Phi(\vec{r},t) + \vec{J}(\vec{r},t)(\mu_a + \mu_s') &= 0,
\end{aligned} \tag{2.88}$$

where $\mu_{\text{ext}} = \mu_a + \mu_s$ is the extinction attenuation coefficient and $\mu_s' = \mu_s(1-g)$ is the reduced scattering coefficient. The final form of this differential equation when collecting terms can be expressed as

$$\boxed{c^{-1}\frac{\partial\vec{J}(\vec{r},t)}{\partial t} + \vec{J}(\vec{r},t)(\mu_a + \mu_s') = -\frac{1}{3}\nabla\Phi(\vec{r},t).} \tag{2.89}$$

2.3.4 Current density expression (temporal broadening assumption)

The temporal broadening assumption states that the change in fractional current density is very small compared with one transport mean free path. This is an important assumption which invalidates the diffusion approximation for weak scattering and thin materials, where a short laser pulse has not sufficiently spread in time compared with the time taken to propagate over one transport mean free path. The time taken for a photon to propagate over a transport mean free path

is calculated by dividing by the speed of light in the material (ℓ^*/c). The temporal broadening assumption expressed in mathematical terms is

$$\left(\frac{\left|\partial\vec{J}(\vec{r},t)/\partial t\right|}{\left|\vec{J}(\vec{r},t)\right|}\right)\left(\frac{\ell^*}{c}\right) \ll 1. \tag{2.90}$$

Rearranging this expression shows that the second term of equation (2.89) dominates over the first:

$$c^{-1}\left|\frac{\partial\vec{J}(\vec{r},t)}{\partial t}\right| \ll \left|\vec{J}(\vec{r},t)\right|\ell^*. \tag{2.91}$$

Under this assumption the rate of change of current density in equation (2.89) can be set to zero and the remaining expression can be written as

$$\vec{J}(\vec{r},t)\ell^* = -\frac{1}{3}\nabla\Phi(\vec{r},t). \tag{2.92}$$

The equation above can be directly compared with Fick's first law of diffusion which states that the diffusion flux (i.e., current density) goes from regions of high concentration (i.e., fluence) to low concentration proportional to the concentration gradient multiplied by a diffusion coefficient D:

$$\vec{J}(\vec{r},t) = -D\nabla\Phi(\vec{r},t) \tag{2.93}$$

Comparing equations (2.92) and (2.93) determines the diffusion coefficient to be $D = 1/3\ell^* = 1/[3\left(\mu_s' + \mu_a\right)]$.

2.3.5 Photon diffusion approximation

Finally, the equation for current density in (2.93) can be substituted into the scalar differential equation (2.64) to give:

$$\begin{aligned}
c^{-1}\frac{\partial\Phi(\vec{r},t)}{\partial t} &= -\nabla\cdot\vec{J}(\vec{r},t) - \mu_{\text{ext}}\Phi(\vec{r},t) + \mu_s\Phi(\vec{r},t) + S(\vec{r},t),\\
&= -\nabla\cdot\left[-D\nabla\Phi(\vec{r},t)\right] - \mu_{\text{ext}}\Phi(\vec{r},t) + \mu_s\Phi(\vec{r},t) + S(\vec{r},t),\\
&= D\nabla^2\Phi(\vec{r},t) + (-\mu_{\text{ext}} + \mu_s)\Phi(\vec{r},t) + S(\vec{r},t),\\
&= D\nabla^2\Phi(\vec{r},t) + (-\mu_a - \mu_s + \mu_s)\Phi(\vec{r},t) + S(\vec{r},t),\\
&= D\nabla^2\Phi(\vec{r},t) - \mu_a\Phi(\vec{r},t) + S(\vec{r},t).
\end{aligned} \tag{2.94}$$

The final form of the equation is commonly known as the photon diffusion equation:

$$\boxed{c^{-1}\frac{\partial\Phi(\vec{r},t)}{\partial t} = S(\vec{r},t) + D\nabla^2\Phi(\vec{r},t) - \mu_a\Phi(\vec{r},t).} \tag{2.95}$$

2.3.6 Solution to the photon diffusion equation

The solution to the photon diffusion approximation accurately describes photon statistics in the diffusive regime where $\mu_s' \gg \mu_a$ [21]:

$$\Phi(\vec{r},t) = \frac{c}{(4\pi Dct)^{3/2}}\exp\left(-\frac{|\vec{r}|^2}{4Dct} - \mu_a ct\right), \tag{2.96}$$

where $\Phi(\vec{r},t)$ (Wcm^{-2}) is the photon fluence rate resolved in space and time, $D = 1/[3(\mu_s' + \mu_a)]$ is the diffusion coefficient, $\vec{r}$ is the radial distance of the output surface of the material from a source/object pixel, t is time and c is the speed of light in the medium.

This solution, originally derived by Chandrasekhar [27], is for a short pulse from an isotropic point source in an infinite homogeneous medium. However, Patterson et al. [21] later validated the accuracy of this model for quantifying biological tissue properties when the measurement is made far from the source location at times long after the pulse incidence.

The time dependence of the solution is the product of an exponentially increasing component due to spatial and temporal broadening, and an exponentially decreasing component in the form of Beer's law (Eq.(2.26)), which considers only the absorption of the material. However, note that these exponential terms cannot be separated into independent scattering and absorption contributions, since the diffusion coefficient is dependent on both of these parameters: $D = 1/[3(\mu_s' + \mu_a)]$. The prefactor also contains the diffusion coefficient and its time dependency ensures the solution is localised in time as $\mu_a \to 0$.

Examples of normalised fluence rate using Eq. (2.96) are plotted in Fig. (2.3) to demonstrate the sensitivity of the equation to the absorbing and reduced scattering coefficients. The fluence rate is normalised by the area to represent the probability density of observing a photon in time at a radius of $\vec{r} = 5$cm. Interestingly, increasing absorption has a dramatic effect on the width and mean time-of-flight of the distribution. Intuitively, this reflects the physics of the problem, where the number of photons which have a long optical path length in the material due to multiple scattering are much more likely to be absorbed. Therefore, an increase in overall absorption of the material suppresses the probability of observing highly scattered photons. Note that both μ_a and μ_s' have an effect on the shape and mean of the normalised distribution.

Note that the photon diffusion equation is parameterised by the reduced scattering coefficient μ_s', not the underlying anisotropy factor g and scattering coefficient μ_s. When using this equation, it

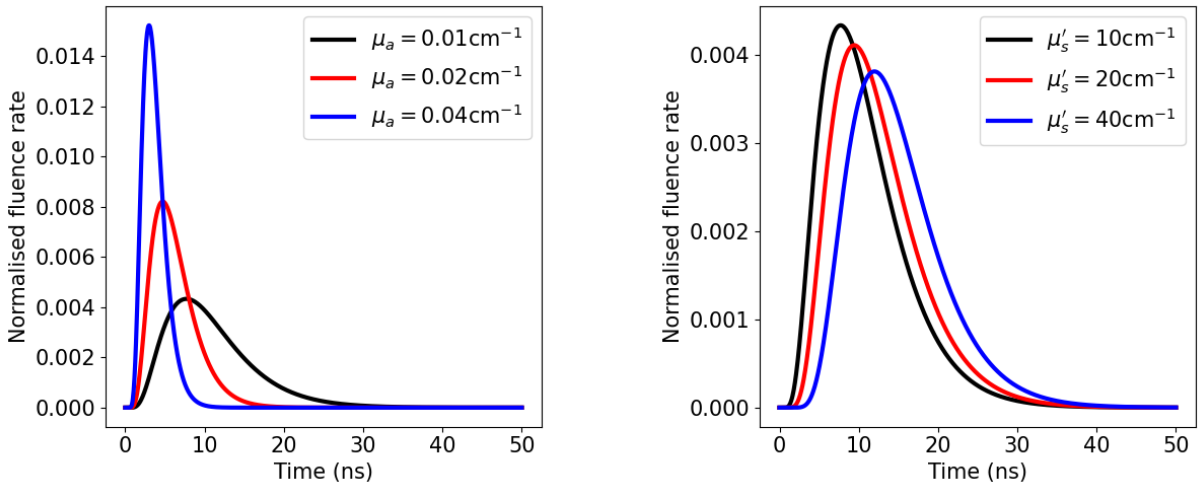


Figure 2.3: The solution to the photon diffusion approximation to the RTE (Eq. (2.96)) for a radius of $\vec{r} = 5\text{cm}$. Left) The normalised fluence rate versus time for a fixed reduced scattering coefficient $\mu_s' = 10\text{cm}^{-1}$ and varying absorption coefficient μ_a. Right) The normalised fluence rate versus time for varying reduced scattering coefficient μ_s' for a fixed absorption $\mu_a = 0.01\text{cm}^{-1}$.

is not possible to simulate (or extract by fitting) the anisotropy properties of the material. This degeneracy of the equation for different combinations of g and μ_s is known as the similarity relation [17]. However, these parameters are unimportant in this thesis since knowledge of μ_s' is sufficient to contextualise the results to biomedical applications by comparing with reported values for biological tissue types (see [25] for a review).

2.4 Time-correlated single photon counting

Most modern time-resolved diffuse optical imaging techniques use photon counting detectors and time-correlated single photon counting (TCSPC). Therefore, in this section an introduction to this technique is discussed before introducing the concepts of time-resolved diffuse optical imaging in the following section.

Time-correlated single photon counting is a technique used to build a histogram of the time delay between the emission and detection of a photon with picosecond resolution. In the context of diffuse optical imaging, the goal is to build a distribution of the time taken for a photon to propagate though a scattering material. Using the ray tracing analogy of photon diffusion, a photon which experiences more scattering will have a longer optical path length before reaching the detector than a less scattered photon. To measure the time taken for a photon to propagate through a material, the delay between the electronic reference signal of the source emission and the electronic signal indicating the arrival of a photon from a single-photon detector is determined. Note that this is a relative measurement and the absolute time delay of the laser pulse reaching the detector in the absence of a scattering sample must be known to realise an absolute measure of time

with this technique. By repeating this measurement for many photons, a histogram representing the distribution of the time-of-flight for every detected photon through the material is resolved. The various electronics which are required for accurately timing the difference between the signal from the detector and the reference pulse from the source will be referred to as the TCSPC module from hereon.

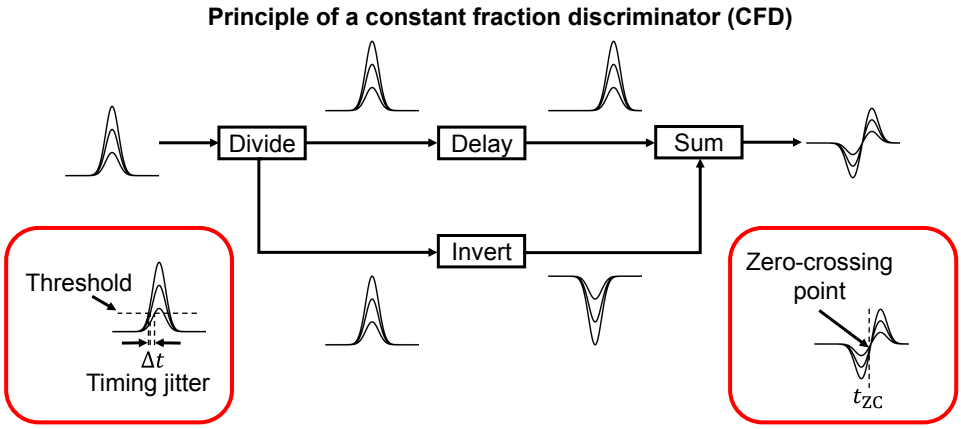


Figure 2.4: The basic principle of a constant fraction discriminator. An input pulse is replicated and divided. In the upper path the signal is delayed, and in the other the signal is inverted. The combination of these two signal results in a signal which crosses the zero amplitude (zero-crossing point) at the same time position (t_{ZC}), which is independent of initial amplitude of the signal (right box). Whereas, an amplitude-dependent method of timing (e.g., using a threshold) results in significant timing jitter (Δt) (left box).

The signal electric pulse is generated by the detection of a single photon on a single-photon sensitive detector, which produces a digital high from the detector in the form of a nuclear instrument module (NIM) or a transistor-transistor logic (TTL) signal. Unlike typical photodetectors which have a proportional electrical response to incident light, single-photon detectors usually operate in Geiger mode which, like a Geiger-Müller tube, register a single click regardless of the number of photons or energy incident on the active region. This signal is input into a TCSPC module which starts a timing process until it receives the subsequent signal from the source. This mode of operation is known as reversed start-stop, since the time measured is the time between the signal pulsed followed by its reference, rather than the more intuitive method of timing between the emission and detection. This causes the detection of photons with long time-of-flight to have the shortest delay times to the subsequent reference pulse, and requires the histogram to be reversed when interpreting the distribution of photon times through the material.

A requirement of the reversed start-stop mode is that the reference signal from the laser must be delayed such that it arrives after the detector signal. If this is not the case, or the signal pulse is delayed too much such that it is stopped by a later reference pulse, then this will introduce jitter caused by the variation in time between source emissions into the impulse response function of

the whole system. In cases where the electrical trigger from the source has an odd shape or is unstable, the time of emitted pulses from the source can be measured directly by sampling the beam directly after the emission output and using an avalanche photodiode or a fast photodiode.

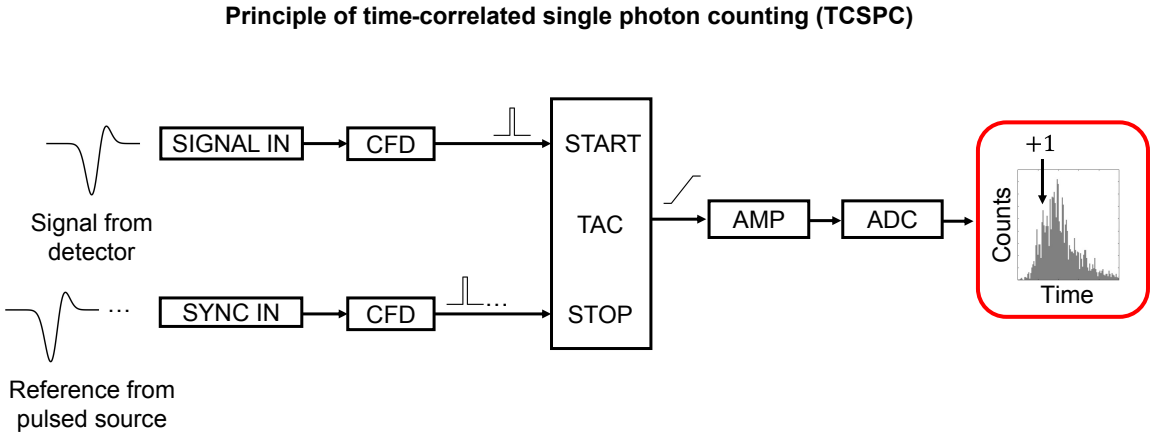


Figure 2.5: The basic principle of classical TCSPC. The single photon detector and laser synchronisation signals are NIM signals, which are input into a constant fraction discriminator (CFD) to accurately determine the time-of-arrival independent to the amplitude of the pulse. The TTL output of the CFD is then input to a time to analogue converter (TAC), which is a capacitor that starts charging when the detector pulse is input and stops charging when the next synchronisation pulse arrives (if operating in stop-start mode). The readout voltage of the capacitor is then amplified and quantised by an analogue to digital converter (ADC). The time-stamp value is input to memory adds a count to the corresponding time-bin of the time-of-flight histogram. Over many pulses, a distribution of arrival times is built up.

An important consideration for precise timing of the delay between electrical signals is to minimise jitter due to variations of the amplitude and precise shape of the electrical pulses (e.g., from electrical noise). To address this, a constant fraction discriminator (CFD) is used to reliably determine the time that the signal and reference pulses reach the TCSPC module. The basic principle of a constant fraction discriminator is shown in Fig. (2.4). A naïve discriminator to determine the time of arrival of a signal would simply use a threshold voltage, such that when the leading edge of the signal reached the threshold the timing is started or stopped. However, as illustrated in the left red box of Fig. (2.4), the varying amplitude or shape of the pulse will cause significant timing jitter, even if the discriminator had infinite timing resolution [28]. To overcome this with a constant fraction discriminator, the signal is first replicated, where in one arm a delay is applied, and in the other the signal is flipped. Since the replicated signals had the same shape and amplitude, when recombined the leading edge of the summed signal will always cross from negative to positive voltage at a constant fraction of the amplitude. The point at which this occurs is known as the zero crossing point, shown in the right red box of Fig. (2.4). In practice, using exactly zero as a threshold is not possible and moreover, the discriminator used to determine when the leading edge crosses zero has some intrinsic delays which should be com-

pensated for by adjusting the zero crossing level [28]. Therefore the zero crossing level is left as an adjustable parameter in TCSPC modules. This description of constant fraction discriminators is high level, and more sophisticated designs are used to achieve sub-ps resolution in modern TCSPC modules.

A TCSPC module uses a constant fraction discriminator for both the signal and reference (SYNC) channels, as shown in Fig. (2.5). The output TTL pulses of the CFDs are input into a time-to-amplitude converter (TAC) to measure the delay time. A TAC consists of a capacitor which begins charging when the signal pulse is input and stops charging when the reference arrives. The analogue voltage reached by the capacitor in this time is then amplified to increase the dynamic range when converting it to a digital value with an analogue-to-digital converter (ADC). Finally, the quantised value of voltage is stored in memory as the value "1" stored at an address corresponding to the magnitude of the ADC output.

The resolution of the histogram time-bins is set by minimum time difference which can be resolved by the TAC and ADC, however this is much smaller than the impulse response function of the hardware used in TCSPC experiments.

Time-correlated single photon counting is not the only way to time-resolve photons in experiments, but one of the benefits is that it can be used with a large variety of single-photon sensors, and it is the sole technique used for experiments in this thesis.

2.4.1 Time-resolved photon counting hardware

Early time-resolved diffuse imaging apparatus included optical Kerr gates or streak cameras. The optical Kerr effect is a nonlinear response of a material to intense polarised incident light, which proportionally modifies the refractive properties and causes temporary birefringence. When placed between two crossed polarising filters, an ultrafast pulse of light can induce birefringence for as short as 10ns. The re-emitted pulse of light coming from the sample can be gated such that only a short fraction of the pulse is transmitted [29]. The delay between the pump beam, which induces the birefringence, and the weaker signal beam from the sample, can be tuned to allow, e.g., only early photons to be transmitted through the Kerr gate and imaged by a sensitive detector array such as a cooled CCD [2].

On the other hand, streak cameras resolve the time-of-flight of the photons using a photocathode to generate a photoelectron beam proportional to the incident intensity of light. A fast-scanning electric field maps the photoelectron beam onto a fluorescent screen which can be imaged with a sensitive detector array. These devices are costly but can resolve the full time distribution of the arrival pulse of light with only a single pulse with sub-picosecond timing resolution. Kerr gate and streak camera techniques can be single-photon sensitive with an appropriate detector.

Detectors used for time-correlated single photon counting (TCSPC) must be suitable for operation in Geiger mode, whereby the output is not proportional to incident intensity but, upon detection of a single photon, produces a binary high or low represented by a transistor-transistor-logic (TTL) or a nuclear instrument module (NIM) pulse. A typical detector for TCSPC experiments is a photomultiplier tube (PMT) which works by the principle of the photoelectric effect. Light is incident on a photocathode material with a low ionisation energy corresponding to the energy of a single photon. A generated photoelectron is accelerated towards a dynode with a low work function and ionises secondary electrons. This process is repeated for successive dynodes until a readily detectable current is produced at the anode. A microchannel plate (MCP) works by the same principle as a PMT, but the dynode cascade is performed by a two-dimensional array of narrow tubes which spatially resolve detection events.

Recent advancements in solid-state hardware, such as single photon avalanche detectors (SPADs), has miniaturised large and delicate equipment such as PMTs/MCPs, which require rack-mounted TCSPC Modules [30,31]. The working principle of a SPAD is to reverse bias a p-n junction with a high electric field, such that when an electron-hole pair is produced from the absorption of a single photon, the electron is accelerated with enough kinetic energy to collide with other atoms. When collisions occur, more electrons in the active region are promoted to the conduction band in a process known as impact ionisation. The secondary electrons then collide with other atoms until an "avalanche" produces a detectable current in the circuit. A quenching resistor actively or passively suppresses the current to stop the avalanche. Passive quenching is achieved using a resistor in series with the diode that increases in voltage with increasing current and therefore reduces the bias voltage over the p-n junction. A reduced bias over the active region will reduce the kinetic energy of the free electrons below the impact ionisation energy until the avalanche stops. When all the free charge carriers have dissipated in the p-n junction, the voltage over the quench resistor is low again, and the high bias over the diode returns, ready to absorb another photon. Active quenching is the quicker process of using a discriminator as an electrical switch to reduce the bias voltage across the p-n junction.

Single photon avalanche diodes can be used as single pixels to make up arrays as large as one-megapixel [32]. Some benefits of SPADs over alternative devices are that the silicon camera industry can be leveraged to fabricate them at low cost, they are also more mechanically robust, orders of magnitude more compact than alternative detectors, and they are not as sensitive to magnetic fields or overexposure of light. However, SPADs typically have more coarse timing resolution and higher dark count noise compared with other single-photon detectors, but as SPAD research develops, this is expected to improve. An avalanche photodiode (APD) is a commonly used detector in non-time-resolved diffuse imaging as a sensitive detector. The operation principle is the same, but the bias voltage across the active region is lower so that the avalanche is controllable and proportional to the intensity of incident light.

One of the drawbacks of using single-photon detectors for TCSPC is the requirement for a dedicated TCSPC electronic module. However, some SPAD arrays have been fabricated with independent timing electronics integrated into every pixel to overcome this [33]. The timing electronics differs from a conventional TCSPC module which uses time-to-analogue capacitors followed by analogue-to-digital converters and instead uses a fully digital time-to-digital converter (TDC). The principle is to start a digital counter known as a ring oscillator upon detecting an electron avalanche. The basic principle of a ring oscillator is to create a digital clock using an odd number of NOT gates connected in a ring such that the output is continuously flipping between high and low states. When the reference pulse from the source (e.g., a pulsed laser) is received, the number of clock cycles since the detected photon is recorded as an integer value. Each integer can be converted to a timestamp using the knowledge of the clock period ($\sim$ 50ps). A histogram of all of the timestamps produces a reversed start-stop distribution of the time-of-arrival of the photons at the pixel, which can be flipped to represent the time-of-flight of the photons through the scattering material. Unfortunately, because of the space required for TDC electronics on the microchip, the active area for detecting light has a much lower fill factor than other single-photon detectors such as PMTs. Fortunately, new designs have improved the fill factor, but this is still an active research field [34]. Integrating independent TDCs for every pixel can realise the miniaturisation of single-photon TCSPC detectors, which can be incorporated into wearable biomedical imaging technologies.

One way to increase the fill factor and efficiency of SPAD arrays is a related detector known as a silicon photomultiplier (SiPM). These devices comprise an array of SPADs connected to a common output rather than timing every SPAD independently [35,36]. This decreases the spatial resolution since it is unknown which SPAD in the array produced the signal at the common output. However, two significant bottlenecks of conventional SPAD arrays can be overcome. Firstly, the time for a SPAD to quench and reset (known as dead time) is typically nanoseconds, which is much longer than the repetition rate of the source used in an experiment (e.g., 12.5ps for an 80MHz pulsed laser). Whilst one SPAD is resetting, a neighbouring SPAD can detect another photon and increase the overall detection efficiency by parallelising the detection.

The second major bottleneck is the active area; since timing electronics are only required for a group of SPADs rather than individually, the fill factor can be much higher for SiPMs. Furthermore, if spatial resolution is not needed, the number of SPADs can be scaled to cover a large area. Although each SPAD operates in Geiger mode, the output of the SiPM can be a voltage which is proportional to the number of active SPADs within a small time frame (analogue mode) or a train of short digital high pulses corresponding to detection from any of the SPADs (digital mode). The latter can be used for TCSPC experiments.

Superconducting nanowire single-photon counting detectors (SNSPDs) are another single-photon sensitive detector which can be used in TCSPC experiments. The principle is to cool a conduc-

tive nanowire to around one kelvin, which becomes superconductive. However, when the energy of a single photon is incident on the wire, it is absorbed by the splitting of electron cooper pairs, raising the energy state of the electrons and creating a thermal hot spot. This causes a non-zero resistivity in the nanowire and causes the current to flow through a shunt resistor of lower impedance, which produces a detectable voltage. These detectors have very high quantum efficiency for a broad range of wavelengths, but they are bulky, expensive, and restricted to coupling light to a single-mode fibre.

The pulsed source is the final piece of equipment to be considered in a time-domain diffuse optical imaging system. The ideal pulsed source is a near-IR laser source with powers of 10s of mW, < 200ps pulse duration and high repetition rates (> 20MHz) [30]. Although not explored in the scope of this thesis, vertical-cavity surface-emitting lasers (VCSELs) are a suitable candidate for a cost-effective way to miniaturise expensive and bulky ultrafast pulsed lasers that are predominantly used to date.

2.5 Diffuse optical imaging overview

Diffuse optical imaging is a term used in the context of this thesis to encompass a variety of techniques which involve resolving images of hidden absorbing or scattering inhomogeneities inside diffusive materials. Note that in this definition, there is a distinction between imaging through *diffusive* rather than *scattering* materials. Imaging through scattering/complex/turbid media is a broader research field, of which diffuse optical imaging is a sub-field that concerns the regimes where the diffusion approximation is valid (i.e., $\mu_s' \gg \mu_a$ and $L \gg \ell^*$).

The motivation for pursuing solutions to imaging through diffusive materials is mainly to realise biomedical imaging applications, namely monitoring brain activity and optical mammography. These applications exploit the low absorption of biological tissues in the near-infrared (near-IR) wavelengths (Fig. (2.6)) and hold promise for optical technologies to overcome the limitations of current medical imaging devices. For example, the transmissibility of near-IR light to the human skull overcomes the issue of high reflectivity to bone when using ultrasound to image the brain [37]. Optical wavelengths also present a safer alternative to instruments that use ionising radiation, such as X-ray computed tomography (CT) and positron emission tomography (PET). Another benefit is that optical hardware is much cheaper, easier to operate, and less bulky than CT, PET, or magnetic resonance imaging (MRI) devices. Optical biomedical imaging technologies also overcome contraindications that restrict the use of MRI in cases where the patient has metallic implants or is required to be monitored continuously at the bedside [30].

Diffuse optical imaging devices in the near-IR can also monitor brain activity at the outer cortical surface using a technique known as functional near-infrared spectroscopy (fNIRS), which indirectly measures brain activity by monitoring blood oxygenation levels similar to blood-oxygen-

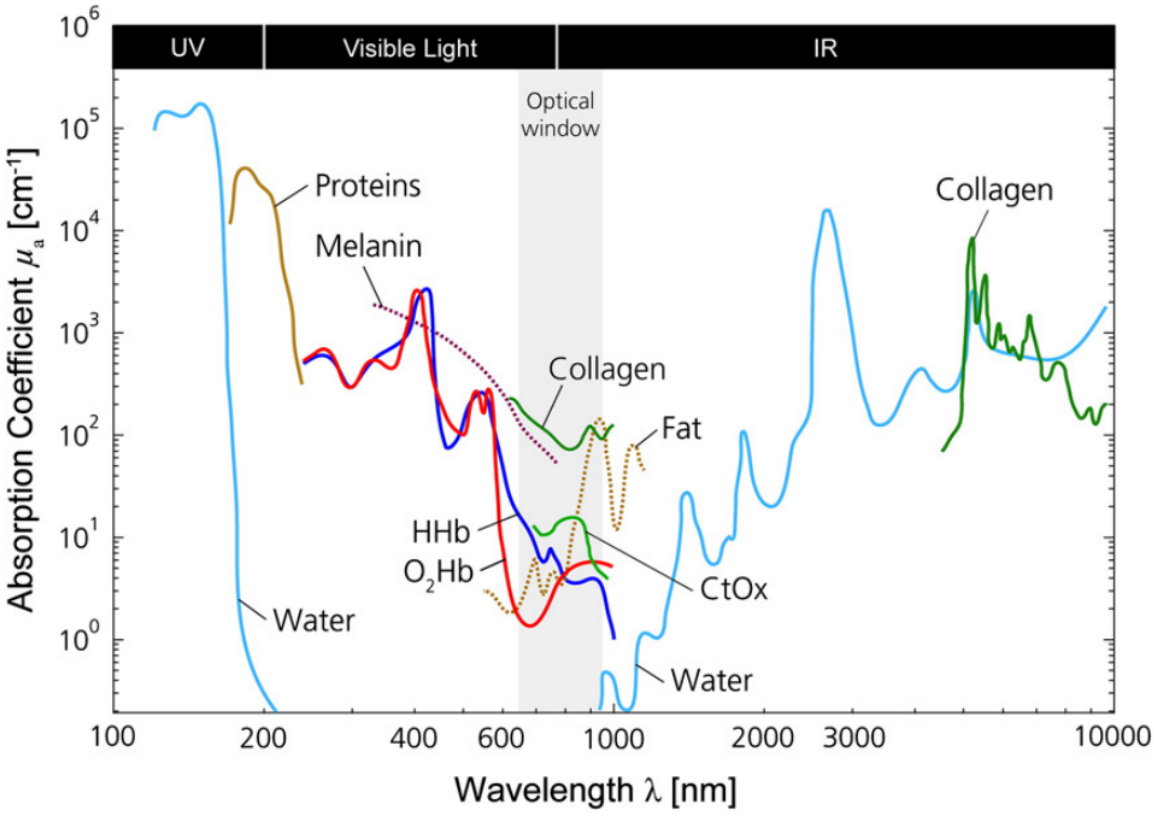


Figure 2.6: Absorption coefficient for various biological tissues highlighting the optical window for biological imaging in the near-infrared. The difference of oxyhaemoglobin (O_2Hb) and deoxyhaemoglobin (HHb) absorption spectra between 650nm-950nm is used in Near-infrared spectroscopy (NIRS) to measure blood oxygenation content (figure from Scholkmann et. al. [38]).

level-dependent (BOLD) contrast signals measured in functional MRI (fMRI) [39]. Therefore, fNIRS offers a portable alternative to fMRI for conducting neuroscience experiments and presents opportunities such as: enabling studies with multiple participants simultaneously, allowing participants to move freely in a natural environment, and negating unwanted brain activity associated with the loud sounds of MRI scanners [30, 40]. The potential presented by fNIRS devices is to enable affordable access to a wearable technologies [41] that can monitor whole brain function with a sample rate up to 200Hz [42], which is orders of magnitude faster than fMRI (1Hz-3Hz) or PET (<0.1Hz), and with a better spatial resolution (2cm-3cm) than comparable wearable techniques like electroencephalography or magnetoencephalography (5cm-9cm) [43].

However, fNIRS is limited in other respects. Unlike fMRI, the sensitivity of fNIRS devices to changes in blood oxygenation is restricted to the outer layer of cortical tissue and activity in the deep cortical or subcortical regions cannot be measured with current devices. Furthermore, fNIRS devices do not provide any anatomical information. Consequently, a structural MRI scan of the participant's head is usually required for precise localisation of activity. The precision of brain activity signals is also inherently limited by the sensitivity of fNIRS to all vasculature,

including arteries, veins, and capillaries, which makes signals harder to interpret [40].

Another hindrance of fNIRS is contamination of signals due to changes in blood chromophore concentration in superficial layers of tissue between the scalp and brain [40]. However, signal contamination can be mitigated by subtracting short source-detector from long source-detector channels [44] or using time-resolved techniques to isolate photons that have a high probability of exploring deeper regions [45].

Near-infrared spectroscopy

Near-infrared spectroscopy (NIRS) works by the same principle as pulse oximeters, which exploit the difference in spectral absorption for oxy- and deoxyhaemoglobin in the near-IR window (Fig. (2.6)) to extract the oxygen content of blood non-invasively. However, unlike conventional oximetry, which determines the percentage of arterial oxygen saturation of the blood (SaO_2), NIRS extends these measurements to monitor the dynamics of oxygen saturation in all vasculature of the tissue (StO_2), including capillaries and veins. Therefore, NIRS is sensitive to metabolic processes in tissue, leading to increased sensitivity compared with pulse oximeters in clinical settings, e.g., for measuring systemic hypoxia for bedside cardiorespiratory monitoring and early warning of airway obstructions in sedative dentistry [46].

Functional near-infrared spectroscopy is a technique which extends NIRS to monitor haemodynamics of the cerebral cortex. Specifically, measuring the differential measurement of absorption between oxyhaemoglobin (O_2Hb) and deoxyhaemoglobin (HHb) within different explored volumes under the scalp, provides an indirect way of localising brain activity. Active regions of the brain require oxygen to metabolise glucose for energy. Red blood cells each have 4 Fe^{2+} sub-units which bind with oxygen in the lungs and transport it around the body. Fortunately, the difference in the electronic band structure of oxygenated and deoxygenated blood leads to different absorption spectra. Functional near-infrared spectroscopy monitors the dynamics of oxygenated blood surges in the brain by measuring the change in intensity of reflected light around 800nm due to the total increase in blood volume compared with the change in intensity at some other wavelength where the absorption of light is more sensitive to oxygenated blood [19, 47, 48].

The technique of fNIRS dates back to the 1977, where Jobsis [49] reported correlations between induced hypoxia and near-IR absorption of a transilluminated cat head in-vivo. These initial transillumination experiments were extended to human subjects but only partial measurements were obtained. However, in the following decades, successful studies were reported to measure blood oxygenation using near-IR spectral absorption measurements in humans for transilluminated neonates [50] and in adult heads using reflected measurements [47, 51, 52].

There are three main techniques in fNIRS: Continuous-Wave (CW-NIRS), Frequency-Domain (FD-NIRS), and Time-Domain (TD-NIRS). The most successful is CW-NIRS which has been developed into a portable wearable devices for adult psychological studies [38, 41, 53], breast

tumour identification [54, 55], neonatal brain monitoring [56, 57], and has been demonstrated to have up to similar resolution (accuracy within 4mm) to functional MRI scans [39, 58]. In this configuration, compact inexpensive sources (e.g., LEDs, diode lasers) provide a constant amplitude light source in direct contact with the surface of the scalp using optical fibres. Likewise, light collection optical fibres are also placed on the scalp and used to guide light to detectors, producing an arrangement of source-detector channels. However, more recently commercial devices have the sources and detectors built-in to a fibreless wearable cap (e.g., [59]).

A large source-detector (SD) distance provides information about explored volumes deeper regions of the material (Fig. (2.7)). Shorter SD channels probe blood oxygenation levels in shallow layers and are used to mitigate the contribution of the signal from blood in superficial tissues and increase the sensitivity of the measurement to the cortical surface. Since there is an exponential decrease in light intensity for increasing SD distances, sensitive detectors such as avalanche photodiodes (APDs) are commonly used [30, 53].

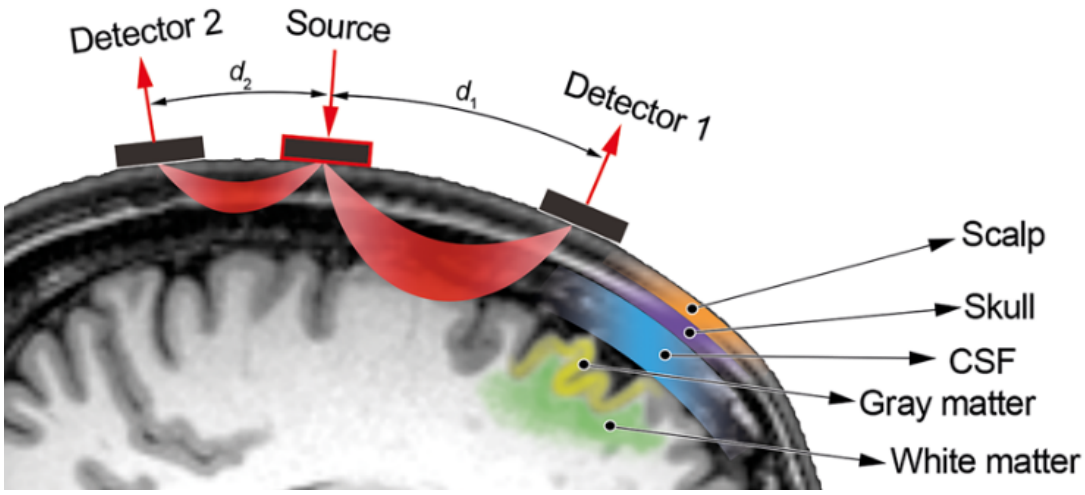


Figure 2.7: The source-detector distances d_1 and d_2 illustrate sensitivity to blood in different explored volumes in the brain (figure from Pinti et al. (2018) [43]).

Diffuse optical tomography (DOT) is the three-dimensional reconstruction of subsurface heterogeneities in optical properties using information from diffuse light re-emitted from a scattering medium. This technique was extended to fNIRS applications by placing channels close to one another such that numerous detectors receive light from the same source, but the detected photons will have explored different sub-surface volumes. More specifically, this is known as high-density DOT (HD-DOT) [60]. The multiple SD channels measured for a single source position enable superior spatial resolution and allow depth information to be resolved[2]. The sensitivity of

[2]Devices not in the HD-DOT regime are restricted to reconstructing only two-dimensional maps of blood oxygenation concentration and are not strictly tomographic techniques. For this reason, there is a distinction between diffuse optical *topography* and diffuse optical *tomography*, which are easily confused. Unfortunately, they share also the same DOT acronym.

NIRS to blood volume and oxygen content, combined with DOT techniques to image subsurface absorbing heterogeneities, has also found numerous other applications such as locating tumours in breast tissue [61] or identifying strokes [62].

The principle of CW-NIRS is to measure the extinction of light at multiple wavelengths and infer the concentration of oxy- and deoxyhaemoglobin. However, scattering also contributes to the extinction of light, and any changes in scattering during the measurement could be misinterpreted as absorption. This limits CW-NIRS to measuring relative haemodynamics rather than quantifying absolute chromophore concentrations. Furthermore, it must be known *a priori* the optical path length travelled by the photons detected, which not only requires knowledge of the geometry of the tissue layers, which varies between subjects, but is also dependent on the unknown scattering properties [63].

The second common fNIRS technique is frequency-Domain (FD-NIRS), which are an evolution of CW-NIRS techniques that use amplitude modulated sources (>100MHz) and measure both absorption and the phase of the re-emitted light [64, 65]. The additional phase information in this context allows an estimation of the average time taken for photons to reach the detector and hence the reduced scattering coefficient μ_s' of the material can be estimated independently from the absorption coefficient μ_a. The quantification of these parameters allows accurate estimation of absolute chromophore concentration. This technique is less commonly used due to the complexity of the post-processing and the size of equipment required [30].

Time-domain techniques (TD-NIRS) utilise time-of-flight information for photons to propagate through the medium to directly determine μ_a and μ_s' using a simulated photon diffusion forward model to perform parameter estimation with regression algorithms [38, 66–68]. Although FD-NIRS can also determine reduced scattering and absorption coefficients, time-domain techniques enable better depth discrimination and are sensitive to a larger sub-surface volume of tissue compared with FD-NIRS [69]. In a different approach known as null source-detector (null-SD) or quasi-null source-detector separation, the method of measuring many SD channels to quantify depth information (i.e., HD-DOT) is not used. Instead, the time-of-flight of the photons is used to separate photons which have explored shallow volumes versus deep volumes [30, 70, 71]. Using this null-SD approach, the limit of depth sensitivity can be extended by at least two-fold compared with continuous-wave techniques in the ideal case of detector efficiency and noise. This improvement corresponds to extending the maximum depth sensitivity to $\sim$ 6cm below the scalp, and experiments have demonstrated sensitivity up to $\sim$ 4cm with prototype devices [72].

Extending the depth of penetration using TD-NIRS systems could enable imaging of other organs to emerge that are currently difficult to image non-invasively, such as monitoring heart perfusion to detect infarction [73], and measuring optical properties of the lungs to determine physiological and morphological changes associated with cardio-pulmonary diseases [74]. Additionally,

DOT instruments may also be used to image bones [75], and diagnose bone pathologies such as osteoporosis [76].

Diffuse optical imaging techniques

The research field of NIRS is a crucial motivator for the evolution of diffuse optical imaging (DOI). Indeed many of the technological milestones in DOI were achieved in the sub-field of NIRS. This section will discuss a brief overview of the history and recent advancements in the broader field of DOI, which also references some literature that reports significant development of hardware and image reconstruction methods in NIRS.

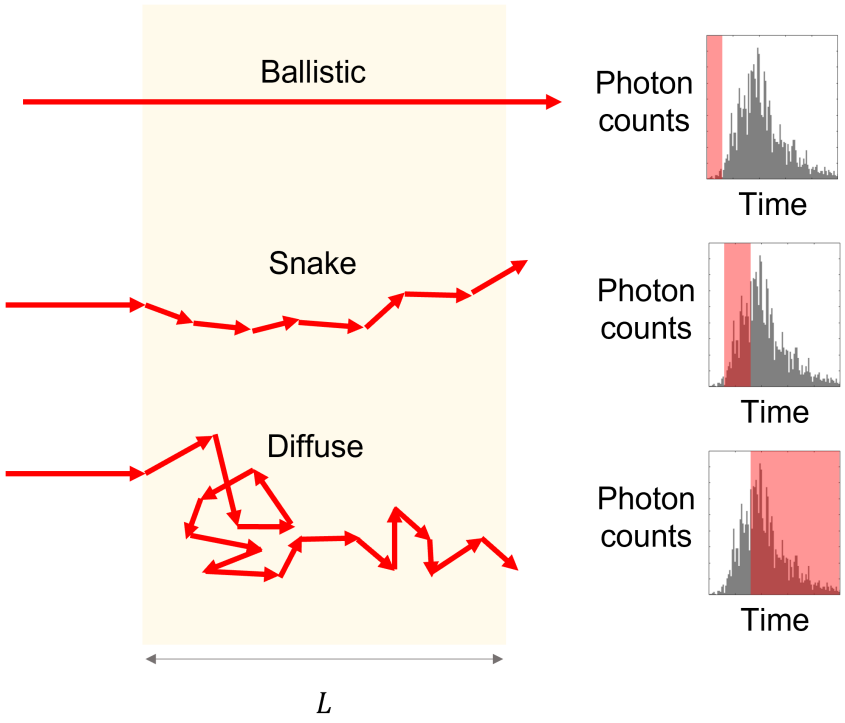


Figure 2.8: The three regimes of photon propagation in diffusive materials and their arrival times highlighted in a time-of-flight distribution.

The overarching methodologies in DOI stem from three different regimes of photon transport in diffusive materials: ballistic, snake, and diffuse propagation (Fig. (2.8)). *Ballistic* photon propagation describes light which does not interact with the material and maintains spatial coherence. Detecting ballistic photons is appealing since any spatial absorption information is directly cast as a shadowgram at the detector, similar to the principle of X-ray imaging. However, the probability of a non-interacting photon exponentially decreases with an increasing number of transport mean free path lengths and techniques which exploit ballistic photons are restricted to weakly scattering regimes. *Diffuse* photon transmission refers to photons which have scattered multiple times, and information about their initial propagation direction is randomised. Of course, if these photons have experienced an extreme number of scattering events, there is a very high proba-

bility of absorption, even in materials with a relatively low absorption coefficient. However, diffuse photons are orders of magnitude more likely to be transmitted through the material and detected for forward scattering materials such as biological tissue. Between these two extremes, there exist weakly scattered photons which are referred to as *snake* photons. These photons are loosely defined as those which do not deviate significantly from their initial direction of propagation [3, 77–79].

When a photon is scattered, the information about the initial direction of propagation is randomised, and consequently, the direct transilluminated spatial information about hidden absorbing objects is corrupted. Therefore, most techniques for imaging through scattering materials attempt to isolate coherent ballistic or snake photons to retrieve images. There are a variety of techniques to achieve this, the earliest of which are reported in the 1980s-1990s.

One method to isolate coherent photons is to exploit the definition that they retain their direction of propagation through the material using spatial filtering. The principle of a spatial filter in this context is to block uncollimated light by placing a pinhole at a position corresponding to the zero frequency component in the Fourier plane of a lens. One of the earliest demonstrations of this technique for biomedical imaging was reported by Jarry et al. [80], which successfully resolved images of metallic objects embedded in various mammalian organs using spatial filtering to isolate collimated transmitted light. This idea was explored further by Hebden et al. [81]. In this study, the attenuation of ballistic light with increasing thickness of pig adipose tissue is reported using a spatial filter. The findings show that isolating ballistic photons is unlikely to be useful for clinical applications due to the difficulty of detecting them in thick samples. However, in a separate experiment, they examined the image resolution of a thin plastic target embedded between relatively thick adipose tissue slices (2.4mm total thickness). They concluded that, even though strictly no ballistic photons are present, an image enhancement can be gained using spatial filtering methods due to the forward scattering nature of tissue. This is because, in forward scattering materials, photons deviate only slightly from their initial propagation direction over relatively small length scales and therefore retain most of their direct transillumination imaging information.

This conclusion that photons which only slightly deviate from their initial propagation direction (snake photons) can also be useful for imaging through biological tissue validate the findings of an earlier numerical study Maarek et al. [82]. The focus of this numerical study however, was to show that gating the earliest arriving photons in time has potential for biomedical imaging applications where the sample is too thick for spatial filtering. Interestingly, the authors of this study also envisioned that the use of distribution of all photons may also be used to reconstruct three-dimensional images with future imaging devices, an idea which was realised in the following decades and is the principle used in many of today's time-domain diffuse optical tomography devices.

Time-resolved measurements of re-emitted scattered light can also isolate ballistic photons. The photons which have a direct path through the material follow the shortest optical path length and arrive at the detector the earliest, whereas the multiple scattering of diffuse photons have a much longer path and arrive later (Fig. (2.8)). Using a time-gate to isolate the earliest arriving photons can produce direct images using ballistic photons in weakly scattering or thin materials. More commonly, quasi-coherent snake photons are the first photons to arrive in thicker samples which have not significantly deviated from their initial propagation direction and still provide significant direct imaging information. The technique of using the first photons to arrive is most commonly known as early photon imaging (discussed in more detail in Section (2.5.1)).

Time-resolved techniques can also be used to measure indirect imaging information about the internal structure of a diffusively scattering material. Similar to TD-NIRS discussed in the previous sub-section, the statistical moments of the time-of-flight distribution, or curve fitting the time-of-flight distribution to a photon diffusion forward model to retrieve the reduced scattering μ_s' and absorption μ_a coefficients, can add additional information to the image reconstruction process. This information can be useful when imaging in regimes where the direct intensity information is strongly attenuated in thick highly diffusive samples. Indirect time-resolved diffuse optical imaging techniques are discussed in more detail in Section (2.5.2).

2.5.1 Time-resolved early photon imaging

In the 1970s-1980s, the first time-resolved measurements of light-in-flight were reported. They used an optical Kerr gate as an ultrafast camera shutter to image a laser pulse transmitted through milky water [83], and topographic depth images of internal mammalian heart ventricles [84]. Since ballistic photons have the most direct path through a scattering material, the optical path length and time-of-flight are the shortest. Therefore, numerous studies in the following decade report the use of time-resolved measurements of light to time-gate the first arriving photons at the detector and access direct image information of absorbing hetereogeneities embedded inside scattering materials.

The first attempts to isolate ballistic transilluminated photons using a time-gated approach used a 8ps resolution streak camera to record the transmission of an 80fs laser pulse incident on a glass cell containing suspended latex beads [85]. The results show that as the concentration of latex beads increase, the coherent contribution of the signal diminishes, which was an early indicator of the limitations of ballistic imaging. Shortly following this study, an optical Kerr gate was also used to isolate the first transmitted photons from a fluorescent target when excited by a pulsed source [77]. This technique improved the spatial resolution of imaging fluorescent targets through samples of millimetre thickness, including *ex vivo* human breast tissue, *ex vivo* chicken breast tissue, and a phantom consisting of polystyrene balls suspended in water. However, these samples were also relatively weakly scattering, equivalent to $< 8\ell^*$.

Time-gating the earliest arriving photons can also be achieved with a time-resolved detector rather than the use of Kerr gates. Measurements using a ultra-short pulsed lasers and an ultra-fast single-photon detectors to enhance transilluminated images of absorbing occlusions in diffusive media using the first-arriving photons also data back to the early 1990s. Work by Hebden et al. [79] used a streak camera to resolve the time-of-flight of distribution, and show that isolating the first-arriving photons at the detector could image absorbing objects embedded inside the medium which were unresolvable when integrating photons from all times (i.e., using a continuous wave source). Similar approaches around the same time were demonstrated using photomultiplier tube (PMT) and microchannel plate (MCP) detectors, which use time-correlated single photon counting (TCSPC) to resolve the temporal statistics of photon time-of-flight [86, 87].

Following these pioneering studies, early photon imaging emerged as a technique which aimed to isolate the earliest arriving photons which may not strictly have perfect spatial coherence. Even though scattered, the first photons to arrive at the detector after propagating through a forward scattering material do not significantly deviate from their initial direction of propagation and retain some direct imaging information. From hereon, the term *early photon imaging* will be used to refer to techniques which aim to isolate ballistic and snake photons which retain some direct imaging information.

In an alternative use of time-of-flight information for early photon imaging, Benaron et. al [88] introduced a method which uses the absolute time-of-flight corresponding to the upper bound of integrating the first 1% of photons, and produces an image directly using these time-of-flight values. They demonstrated impressive results which resolve the locations of vital organs of a transilluminated deceased rat, and highlight the difference of imaging the total absorption through samples versus using the time-of-flight information of the earliest arriving photons.

Early photon imaging techniques continued to evolve with various techniques and applications. Reconstruction of complex shapes inside relatively weak scattering materials using early photons was enhanced when measuring from multiple angle projections similar to computed tomography [89]. In a similar projection approach, Chen et al. [90] show improved reconstruction quality when using time-gated pattern illumination to parallelise measurements without introducing increased spatial blurring. Early photon imaging was also demonstrated to enhance the quality of reconstructed tomographic images of fluorescent targets in scattering phantoms [91] and mice [92,93], and in another combined technique, it was shown that early time-gated off-axis holography could improve the sensitivity of imaging through turbid water and extend the depth sensitivity by 1.5 absorption lengths (i.e., $1.5\ell^*$) [8].

More recently, Tanner et al. [4] successfully located the distal-end of an optical fibre with 1cm precision through ventilated sheep lungs, the palm of a human hand, and through the entire human torso ($\sim$ 25cm). Further studies by McShane et al. [94] have shown to resolve point-like emitters

in various locations along the length of an optical fibre using early photons to improve location precision. These works have profound clinical relevance for locating endoscopes non-invasively through the body.

There are also coherence-gating methods to isolate coherent ballistic photons from the incoherent diffuse photons. Using interferometric methods to gate ballistic photons has been shown to enhance synthetic aperture imaging of three-dimensional objects hidden behind scattering materials [9]. In a similar experiment, Maruca et al. [10] used mode selective frequency up-conversion in a non-linear waveguide to filter coherent backscattered photons through a scattering phantom ($19\ell_s$), and a thin slice of chicken breast ($15\ell_s$). This technique was shown to resolve three-dimensional objects with diffraction limited lateral resolution and sub-mm depth resolution. However, these techniques as well as imaging through scattering materials using laser speckle imaging and wavefront shaping techniques [95] are better suited to non-diffusive regimes due to the rarity of strictly coherent light propagating through samples larger than 1cm of biological tissues.

2.5.2 Time-resolved all photons imaging

Around the same year (1990) as the first reports of isolating ballistic photons discussed in the previous sub-section, Singer et al. [96] reported the use of an image retrieval algorithm which could reconstruct the optical properties of a diffusive material using the intensity distribution of scattered photons. Importantly, method does not attempt to isolate coherent (ballistic) photons, but instead uses the distribution of the coherent and incoherent light at various detector locations to determine the internal absorbing structures. The proposed algorithm used a Monte Carlo random sampling simulation as a forward model, in which the absorption and scattering properties of voxels in the material can be updated by gradient descent of the least-squares problem between the output of the simulation and the measurements.

This early study of image retrieval in diffuse optical tomography was hindered by finding solutions in local minima and limited by the small computer memory capacities of its time. However, regularisation techniques such as truncated singular value and Tikhonov regularised least-squares algorithms (see Section (2.6) for overview of regularised least-squares algorithms) were applied in subsequent studies e.g., by Arridge et al. [97], to overcome local minima, and fortunately small memory capacities in computers are no longer a major issue. This lead to a parallel research effort to early photon imaging techniques, which instead used the distribution of diffused light resolved in both space and time as indirect imaging information used to reconstruct the internal structure of a material using inversion algorithms.

More sophisticated retrieval processes were introduced at the turn of the century which were used for tomography in cylindrical geometries with multi-channel time-resolved [98–100] and

frequency-resolved [101] apparatus. Reconstruction algorithms were introduced to resolve reconstructions in depth as well as laterally [102, 103], and simultaneously reconstructing scattering and absorption inhomogeneities was demonstrated [104, 105].

However, early time-resolved techniques did not directly use the full raw data in the reconstruction. Instead, these methods showed improved image reconstruction quality compared with frequency domain and continuous wave techniques by: time-gating early photons, improving the forward model by quantifying optical parameters from the fitted measurements, or using the statistical moments of the time-traces to incorporate additional information to the image retrieval process [105, 106]. However, in 2002, Gao et al. [107] developed an algorithm which exploits the full shape of the time traces and show a significant improvement in reconstruction quality compared with algorithms which use only the moments of the time-resolved data.

Time-resolved imaging, which used regression techniques to fit the full temporal distribution of the measurements with a photon diffusion model and determine tomographic projections of the absorption and scattering properties, was demonstrated beyond phantoms in the early 2000s. Reconstructions of a human forearm [108, 109] and a mouse lung *in vivo* [5] indicated the potential of leveraging the full time distribution. However, at the time, these experiments were hindered by the complexity of the necessary hardware and algorithms. They were also much more expensive to build than continuous wave or frequency domain devices. Fortunately, the introduction of powerful computers and cheaper, more compact pulsed lasers and single-photon detectors have sparked a revival of time-resolved all-photons diffuse optical imaging in the recent decade [30, 67].

The choice of forward model and inverse retrieval algorithm is task-specific and generally custom designed by each research group. However, the majority of algorithms are some form of regularised least-square retrieval that use an iterative convex optimisation solver (e.g., Levenberg–Marquardt algorithm, alternating direction method of multipliers (ADMM), conjugate-gradient method, etc.). Detailed reviews of the forward and inverse methods used in diffuse optical imaging can be found in [110–112].

All photons imaging is a technique introduced by Satat et al. [11] which measures the full spatiotemporal distribution of photon time-of-flight through a transilluminated scatterer using a streak camera. Compared with the methods mentioned above, the difference of this technique is that there is no requirement for scanning the source or using multiple point-like detector positions which sparsely sample the diffused light. Instead, the sample is uniformly illuminated, and the detector is wide-field. The retrieval algorithm consists of a regularised least-squares curve fitting between the measured normalised spatiotemporal profile and a simplified forward model using a convolution kernel derived from the photon diffusion approximation (Eq.(2.96)). The results show that this method dramatically improves the peak signal-to-noise and structural sim-

ilarity metrics with the target image compared with early photon imaging. However, in this study considers only weak scattering and a resemblance of the target is apparent in the early photon time-gated image, which was used as an initial guess to the retrieval.

More recently, Lyons et al. [12] demonstrated the use of the full measured spatio-temporal profile with a similar regularised least-squares retrieval algorithm which was shown to reconstruct images through more than $80\ell^*$. This experiment consisted of two 2.5cm polystyrene foam slabs, in which an absorbing black piece of tape is placed in various geometric shapes which is transilluminated by an expanded beam of an ultrafast pulsed laser. The diffuse photons are resolved in space and time using a single photon avalanche array and, unlike all photons imaging, there is no resemblance of the target object when time-gating for early photons. This is an indication that this demonstration of diffuse optical imaging is in a much stronger diffusive regime than previous studies. Aspects of this study will be discussed in more detail in the following chapters, as it is a precursor to the experiments and analysis of the work conducted in this thesis.

The most current time-resolved diffuse optical imaging techniques are exploring the use of machine learning for inverse retrieval algorithms [113, 114]. The use machine learning for the process of image reconstruction in diffuse optics is a relatively new field, but progress is discussed in more detail in Chapter 5.

2.6 Diffuse optical imaging as an inverse problem

Inverse problems are a set of mathematical problems for which the input and output (with error) are known, and we wish to estimate the model. Alternatively, the model and the output are known, and we wish to know the input. The latter is the most relevant for diffuse optical imaging and so this will be referred to as the inverse problem in this section. These inverse problems are considered *ill-posed* if: many solutions could be plausible for the same known information, there exist no solutions, or the solution is highly sensitive to perturbations of the given data.

More formally, an ill-posed problem is the opposite of a well-posed problem which meets three criteria defined by Jacques Hadamard [115]:

1. *Existence* - the problem has a solution.
2. *Uniqueness* - there is only one solution.
3. *Stability* - the solution depends continuously on the initial conditions.

For context, it is easiest to consider some simple examples of ways in which these criteria can be violated and how to reformulate the problem to one which is well-posed. A simple example which violates the *existence* criteria is an over-determined system of equations such as $a = 1$ and $2a = 3$ (adapted from [115]). There is no solution for a which satisfies both equations. This

may seem a trivial example, but is common in the case where experimental uncertainties cause repeated measurements to take different values. A common way to solve such problems is to add information by the introduction of reasonable constraints, such as the expectation that the magnitude of the residual for both equations is minimised for a given a. To do this, the problem can be reposed as a minimisation of the least-squares error:

$$\arg\min_a \; ||f(a)||_2^2 = \arg\min_a \left\{(a-1)^2 + (2a-3)^2\right\}, \tag{2.97}$$

where $||.||_2^2$ denotes the is the L_2-norm (also known as the Euclidean norm) squared. This problem is differentiable and has a solution which *exists*:

$$\begin{aligned} \frac{df(a)}{da} = 2(a-1) + 4(2a-3) &= 0, \\ 2a - 2 + 8a - 12 &= 0, \\ 10a - 14 &= 0, \\ a &= 1.4. \end{aligned} \tag{2.98}$$

Note that this is not equivalent to taking the mean of each independent solution ($a = 1.25$), which would result in a larger residual error. This method is of course commonly used to fit linear regression models to noisy experimental data-sets, such that the solution has smallest sum of distances from every measurement.

The *uniqueness* criteria is also simple to violate with the equation $a + b = 1$, which has infinite solutions when $a, b \in \mathbb{R}$ [115]. Contrary to the example given for *existence* above, this ill-posed equation is underdetermined since there are more unknowns than knowns. To solve this requires a prior assumption about the properties of a and b to constrain the space of solutions. The most common assumption is that the magnitude of each unknown variable are comparable, and therefore the reformulation of the problem is to find the solution which has the minimum L_2-norm for a and b which still satisfies the right hand side of the equation(s). However, minimising the L_2-norm has an equivalent solution to the more readily solved square of the L_2-norm i.e., $\arg\min_a \; ||f(a)||_2 = \arg\min_a \; ||f(a)||_2^2$, which is equivalent to the least-squares regression in Eq.(2.97). Therefore, this problem which had no unique solutions, can be reformulated as

$$\arg\min_a \; ||f(a)||_2^2 = \arg\min_a \left\{a^2 + (1-a)^2\right\}, \tag{2.99}$$

which leads to a simple calculation of the derivative and yields a *unique* solution to the problem:

$$\begin{aligned} \frac{df(a)}{da} = 2a + 2(1-a)(-1) &= 0, \\ 2a + 2a &= 2, \end{aligned}$$

$$a = \frac{1}{2}$$
$$= b. \tag{2.100}$$

Finally, the *stability* condition requires that a solution must be continuously dependent on the initial conditions. For example, an arbitrarily small change in the initial parameters of the problem should mean the solution also varies by a small amount. Conversely, if a small perturbation of initial parameters causes a large change in solution, the problem is unstable. Ill-posed problems are often also ill-conditioned. However, an ill-conditioned problem specifically considers when small error contributions in the forward problem become large ones in the inverse solution.

A problem can be well-posed but ill-conditioned, which is a numerically solvable problem that depends continuously on the data. Whereas, if it is an ill-posed problem due to the violation of the stability condition, then the problem is not numerically solvable due to the extent of sensitivity to perturbations in the input, and it needs to be reformulated as a well-posed problem, similar to the examples above, using a technique called regularisation.

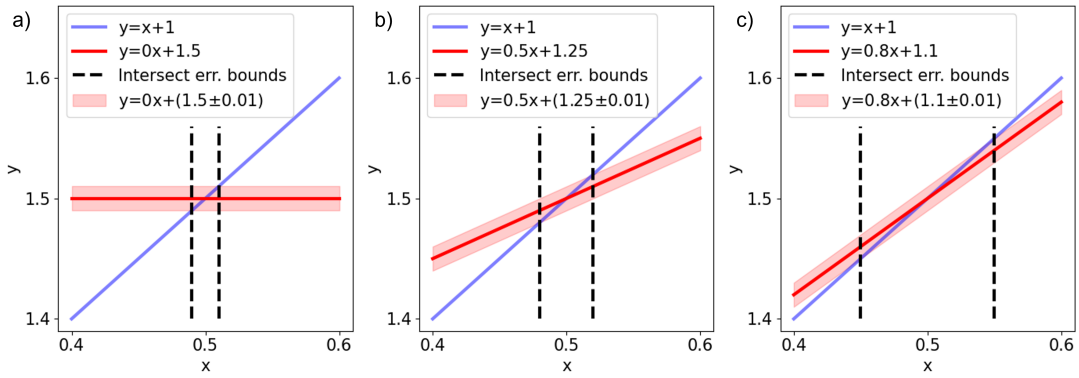


Figure 2.9: An example of an increasingly ill-conditioned problem of determining the point of intersect between two lines. A constant error $\varepsilon = 0.01$ applied to the y-intercept of the red line causes an increasingly large error in the point of intersect (bounds denoted by dashed lines) when the gradient (m) of the red line, a) $m = 0$, b) $m = 0.5$, and c) $m = 0.8$, becomes increasingly similar to the gradient of the blue line ($m = 1$).

An interesting example (inspired by [116]) which illustrates the stability conditions is solving the point of intersection between two almost parallel lines. Consider the examples shown in Fig. (2.9a-c), where the inverse problem is to determine the intersect between the blue line $y = x + 1$ and the red lines. For each of the three examples, there is a unique and exact solution for the point of intercept at $[0.5, 1.5]$. Furthermore, when a small perturbation is made $\varepsilon = \pm 0.01$ to the y-intercept of the red lines, the error in determining the intercept will increase proportionally, making this a well-posed problem.

However, as the gradient of the red line approaches the gradient of the blue line, the problem

becomes increasingly ill-conditioned. When the lines approach parallel, but still intersect at some point, then a small perturbation of the y-intercept in the red line will cause an arbitrary large change in the solution. When the solution varies by magnitudes which are not possible to compute, (i.e., beyond the dynamic range limited by the numerical precision of computers) then the problem is unstable and therefore ill-posed [117]. To solve this, regularisation can be introduced to constrain the possible range of solutions to a sub-set of *a priori* expected values. However, this inevitably introduces inaccuracies and a dependency on introducing information based on prior expertise.

To illustrate the implications of an ill-conditioned problem on the solution space, consider the example for determining the x-coordinate of the intersect in Fig. (2.9c) as an inverse problem. The equation of the blue line is $y = x + 1$, and the equation of the red line is $y = 0.8x + 1.1$. The problem of determining the coordinate of intersection can be written as a system of linear equations:

$$\begin{gathered} Ab = c, \\ \begin{bmatrix} 1 & -1 \\ 0.8 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 \\ -1.1 \end{bmatrix}, \end{gathered} \tag{2.101}$$

where A can be interpreted as a linear mapping from the coordinates $b = [x, y]^T$ to the y-intercept values c. To find the solution where the two lines have a common coordinate, we simply invert the matrix on the left and multiply by the right-hand side:

$$\begin{gathered} b = A^{-1}c, \\ \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0.8 & -1 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ -1.1 \end{bmatrix}. \end{gathered} \tag{2.102}$$

The inverse of 2×2 matrix is a well-known problem, which can easily be solved by hand. However, for any arbitrary real- or complex-valued matrix can be inverted using a Moore-Penrose pseudo-inverse by computing the singular value decomposition (SVD).

The SVD of a matrix is similar to an eigendecomposition of a matrix, but it generalises to non-square matrices by squaring it and then finding the eigenvalues and eigenvectors. The idea is to decompose a matrix A into an orthonormal set of left singular vectors U and right singular vectors V which are related by a diagonal matrix Σ composed of scalar singular values σ:

$$A = U\Sigma V^T. \tag{2.103}$$

To calculate these matrices there are two major steps which use the orthonormal properties $U^TU = V^TV = I$. Firstly, the matrix squared to find an expression for the right singular vectors

and the singular values by eigendecomposition:

$$\begin{aligned} A^T A &= V\Sigma^T U^T U\Sigma V^T, \\ &= V\Sigma^T \Sigma V^T, \end{aligned} \tag{2.104}$$

where the eigenvalues λ and eigenvectors $\vec{v}$ can be found using the condition:

$$(A^T A - \lambda I)\vec{v} = \vec{0}. \tag{2.105}$$

Once the eigenvalues are found then the corresponding singular values are $\sigma = \sqrt{\lambda}$, and the right singular vectors are the unit eigenvectors $\hat{v}$.

The second step is to determine the left singular vectors by simply rearranging the definition of the SVD to $AV = U\Sigma$. Multiplying A by V and normalising for the singular values yields the left singular vectors, and the SVD is complete.

The left singular vectors U can be treated as a set of eigenvectors which capture fundamental basis features of the problem. These basis vectors are weighted by the singular values. Importantly, the singular values are ordered by magnitude (i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ [115]), and give insight into how important each of the left singular values are for representing the column vectors of in the original matrix.

The right singular vectors V, indicate how much of each left singular vectors (weighted by their singular value) are required to reconstruct the matrix back into its original form A. As an intuitive example, consider the eigenfaces problem [118], whereby a matrix constructed of images of peoples faces arranged as column vectors is decomposed by SVD. Each column of the left singular vectors represents a basis image containing fundamental features of the faces in the provided data set. These are weighted by the singular values, ordered from most important to least important. The right singular vectors contain the weights required to reconstruct the original images from the set of singular-value-weighted left singular vectors.

The SVD of a matrix can be used to determine the pseudo inverse by [115]:

$$A^{-1} = V\Sigma^{-1}U^T. \tag{2.106}$$

Since the singular values matrix is diagonal, the inverse is calculated by simply taking the reciprocal of each element. However, the reciprocal of the small singular values associated unimportant features (e.g., noise) will become the strongest weighted contributions in the inverse matrix and cause inaccuracies. The ratio of the largest to the smallest singular vector is known as the condition number of the matrix $\kappa = \frac{\sigma_{max}}{\sigma_{min}}$, which quantifies the sensitivity of the inversion to small perturbations in noise contributions of an ill-conditioned problem.

Returning to the example of the inverse operation to determine the coordinates of intersection of two lines outlined in Eq.(2.102). Using the SVD to calculate the pseudo inverse yields:

$$\begin{aligned}\begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 1 & -1 \\ 0.8 & -1 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ -1.1 \end{bmatrix}, \\ &= \begin{bmatrix} 5 & -5 \\ 4 & -5 \end{bmatrix} \begin{bmatrix} -1 \\ -1.1 \end{bmatrix}, \\ &= \begin{bmatrix} 0.5 \\ 1.5 \end{bmatrix}. \end{aligned} \tag{2.107}$$

This exact solution is found since the problem is well-posed and relatively well-conditioned. To investigate the conditioning of the problem, consider again the slight perturbation of $\varepsilon = 0.01$ to the second element of the y-intercept:

$$\begin{aligned}\begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 5 & -5 \\ 4 & -5 \end{bmatrix} \begin{bmatrix} -1 \\ -1.1+\varepsilon \end{bmatrix}, \\ &= \begin{bmatrix} 5 & -5 \\ 4 & -5 \end{bmatrix} \begin{bmatrix} -1 \\ -1.09 \end{bmatrix}, \\ &= \begin{bmatrix} 0.45 \\ 1.45 \end{bmatrix}. \end{aligned} \tag{2.108}$$

As expected from the visual inspection of the right plot of Fig. (2.9), the x and y co-ordinates have an error of 10% and 3.3% compared with the exact values, for an introduction of only 0.9% in the y-intercept value. As the gradient of the red line becomes increasingly similar to the blue (but the y-intercept value is adjusted to maintain a true intersection of the lines at $[0.5, 1.5]^T$, e.g, $y = 0.9999x + 1.00005$), the addition of an error of the same small error $\varepsilon = 0.01$ causes the solution to be very far from the true solution:

$$\begin{aligned}\begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 1 & -1 \\ 0.9999 & -1 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ -1.00005+\varepsilon \end{bmatrix}, \\ &= \begin{bmatrix} 1000.5 \\ 1001.5 \end{bmatrix}. \end{aligned} \tag{2.109}$$

The sensitivity is due to the amplification of the small singular values in the pseudo inverse operation. The condition number of the example in (2.108) is $\kappa = 18.145$, whereas in (2.109) $\kappa = 39998$. To improve the condition of the problem, a regularisation can be introduced to stabilise the problem. One way to do this is to remove the small singular value in the inversion by setting it to zero. This technique is called a truncated SVD (TSVD). Applying a TSVD to

(2.109) lowers the condition number to $\kappa = 1$, since there is now only one non-zero singular value. The regularised problem yields a solution of $[-0.525, 0.525]^T$, which is much closer to the exact solution of $[0.5, 1.5]^T$ than the result of (2.109).

The truncated SVD finds a lower rank matrix which approximates the mapping between inputs and outputs, without including the small contributing singular vectors. However, if the rank is reduced too much as the problem becomes increasingly more ill-conditioned, then not enough linearly independent singular vectors remain to reconstruct the solution accurately. This is equivalent to the problem of an increasingly underdetermined linear system of equations, which also violates the uniqueness criteria of a well-posed problem. Therefore, the inversion has increasingly many solutions as the problem becomes more ill-conditioned. For this reason, increasingly ill-conditioned inverse problems are also increasingly ill-posed.

Another way to regularise an ill-conditioned problem is to add a penalty term to a least-squares optimisation, which favours solutions with a small norm. This suppresses the likelihood of finding a solution which has a large norm due to the amplification of high-frequency components. Such a technique is known as Tikhonov regularisation, and the magnitude of the amplification can be tuned using a regularisation parameter, often denoted by λ. When the regularisation parameter is chosen to be a scalar value, then the Tikhonov regularisation is equivalent to introducing an additional constraint to a least-squares minimisation problem that the L_2-norm of the parameter must also be minimised [115]:

$$\arg\min_b \left\{ ||Ab - c||_2^2 + \lambda^2 ||b||_2^2 \right\}. \tag{2.110}$$

Regularisation is used to improve the condition of an inverse problem. However, in cases where the ill-conditioning is extreme, the stability condition is violated, and the problem becomes ill-posed. In these cases, introducing regularisation is a method of reformulating the problem to become well-posed.

Numerous other regularisation techniques exist to improve the conditionality of the inverse problem. It is important to introduce regularisation, which constrains the solution based on prior expertise and reasonable assumptions about the problem. There is also a risk of over-regularising a problem, e.g., using a large regularisation parameter λ or truncating too many singular values, which can result in reducing the accuracy of the solution because there are not enough singular vectors to reproduce the solution, or the high-frequency components often associated with smaller singular values are suppressed and information is lost.

In the context of diffuse optical imaging, the aim is to invert the measured distribution of diffused light at the detectors to reproduce the optical parameters which were the underlying cause of the distribution. Since the photon diffusion approximation (Eq.(2.96)) is a smoothing kernel in space

and time. The resulting forward model, consisting of convolution integrals with this kernel, will cause a dampening of the high-frequency components of the direct imaging information. Upon inverting, the dampened high-frequency components will be associated with small singular vectors that cause the problem to become increasingly ill-conditioned for an increasing number of transport mean free path lengths, and the problem quickly becomes ill-posed. As a result, the inverse problem of image reconstruction in diffuse optical imaging in highly diffusive materials often has many plausible solutions which are highly sensitive to small noise perturbations in the measured data. To overcome this regularisation in the form of *a priori*, expectations of the solution need to be introduced to reformulate the problem as a well-posed problem with a constrained solution space.

Numerous computational techniques have been developed to improve the condition of the inverse problem by combining physical models of photon diffusion with regularisation, such as total-variation (TV-norm), Tikhonov, L1-norm and L2-norm, to dampen or remove the small contributing basis vectors during the inversion process [11, 12, 72, 75, 98, 119–121]. Information theoretical regularisation approaches have also been used to co-register *a priori* anatomical images from high-resolution modalities with diffuse reconstructions to improve the accuracy of the results [122]. More recently, the inverse problem has been tackled by fusing physical models with learned, data-driven priors in machine learning frameworks for computational imaging inverse retrievals [13, 123–125].

The following chapters of this thesis study the ill-posed problem of diffuse optical imaging using indirect imaging information in the spatial distribution of photons transmitted through a highly diffusive material. The concepts discussed here, such as the condition number, are used as indicators of the difficulty of image reconstruction for different regimes. The sensitivity of the solution to perturbations of the measurements is an important consideration for critically ill-posed problems when imaging beyond the regime of $100\ell^*$ and a probabilistic inverse model that can be used to assess the diversity of solutions for imaging in this extreme regime is introduced in Chapter 5.

Chapter 3

The role of late photons in diffuse optical imaging

Most techniques to image through diffusive materials aim to isolate ballistic or snake photons which maintain their initial direction of propagation and preserve spatial coherence. As discussed in Section (2.5), these photons take the most direct path to the detector and therefore arrive the earliest. Isolating the earliest arriving photons using time-resolved techniques removes the less coherent diffuse photons from the signal and a direct image of absorbing regions behind or embedded inside the material can be directly resolved.

However, ballistic photons are exponentially suppressed with an increasing number of transport mean free path lengths of propagation in the material, and when imaging through $> 80\ell^*$ the rate of detection is roughly equivalent to one ballistic photon per age of the universe. Therefore, when using time-resolved methods in the highly diffusive regime, early photon methods fail, since even the first photons to arrive are diffuse photons.

In this chapter, the role of late photons for image reconstructions in the highly diffuse regime $> 80\ell^*$ is studied using experimental data and the regularised least-squares image retrieval algorithm developed by the authors of [12]. Findings show that indeed early photon imaging methods fail to improve the quality of reconstructed images, and surprisingly, reconstructed images using exclusively late photons have comparable fidelity. The main conclusion of this study is that, in the highly diffusive regime, there are no coherent photons containing direct imaging information to perform early photon imaging techniques. Therefore, the inversion algorithm is instead sensitive to the indirect imaging information embedded in the full profile of the spatiotemporal distribution of photon counts.

3.1 Introduction

Most techniques to image through highly scattering materials aim to isolate ballistic or snake photons that maintain spatial coherence and preserve direct imaging information. These photons can be filtered using time-resolved equipment to perform early photon imaging, which time-gates the first arriving photons at the detector, as described in Section (2.5.1). However, early photon imaging is restricted to weakly scattering regimes since the number of photons that weakly interact with the material is exponentially suppressed with increasing transport mean free path lengths (TMFPs). This is illustrated in Fig. (3.1), which shows simulated examples of time-gating for early photons for increasing TMFPs using the model described in the following subsection (Section (3.2)). The enhancement of contrast using early time-gating of photons degrades significantly beyond $40\ell^*$ when filtering for the first 1% of photons as shown in Fig. (3.1a).

However, in practice, the noise contribution to the signal would prevent perfect filtering of the small fraction of photons which arrive this early. A more realistic situation is to integrate a larger portion of the early-arriving photons to improve the signal-to-noise ratio (SNR). Unfortunately, by including later-arriving photons, the contrast enhancement degrades significantly after only $20\ell^*$, as shown in Fig. (3.1b). Note that the illustration of this figure is for ideal signals, but in practice, noise contributions and the exponential extinction of photons for increasing TMFPs would further reduce the contrast enhancement and restricts most early photon imaging to regimes $< 10\ell^*$. Even in ideal conditions, it is clear that early photon imaging fails beyond $80\ell^*$. Most of the detected signal consists of incoherent diffuse photons which have a much greater SNR but remain unused in early photon imaging.

The internal absorbing and scattering structure of a diffusively scattering material can also be reconstructed by measuring the intensity distribution of incoherent light re-emitted from the material [38, 96, 121]. Furthermore, effort has been focused on time-resolved techniques that use the full spatiotemporal distribution of photon counts to improve the spatial resolution of reconstructed images [11, 12, 107, 126].

All photons imaging (API) is a technique for imaging through diffusive materials introduced by Satat et al. [11], which emphasises that the full spatiotemporal profile of the detected photons can enhance the spatial resolution of reconstructed images. However, this work is in the regime of relatively weak diffusion ($\sim 15\ell^*$), and as a result, there is a resemblance of the target image in both the time-integrated and early-gated images. The gated image of ballistic photons is also used to provide a strong initialisation prior for the retrieval algorithm, which would not be applicable in experiments in the highly diffusive regime, in which the early-gated images have little or no resemblance to the target object. The API technique also assumes that there is no spatial information in the temporal profile of the detected photons and subsequently normalises all the spatial pixels of the measurement by the time traces with the highest signal-to-noise ratio. This

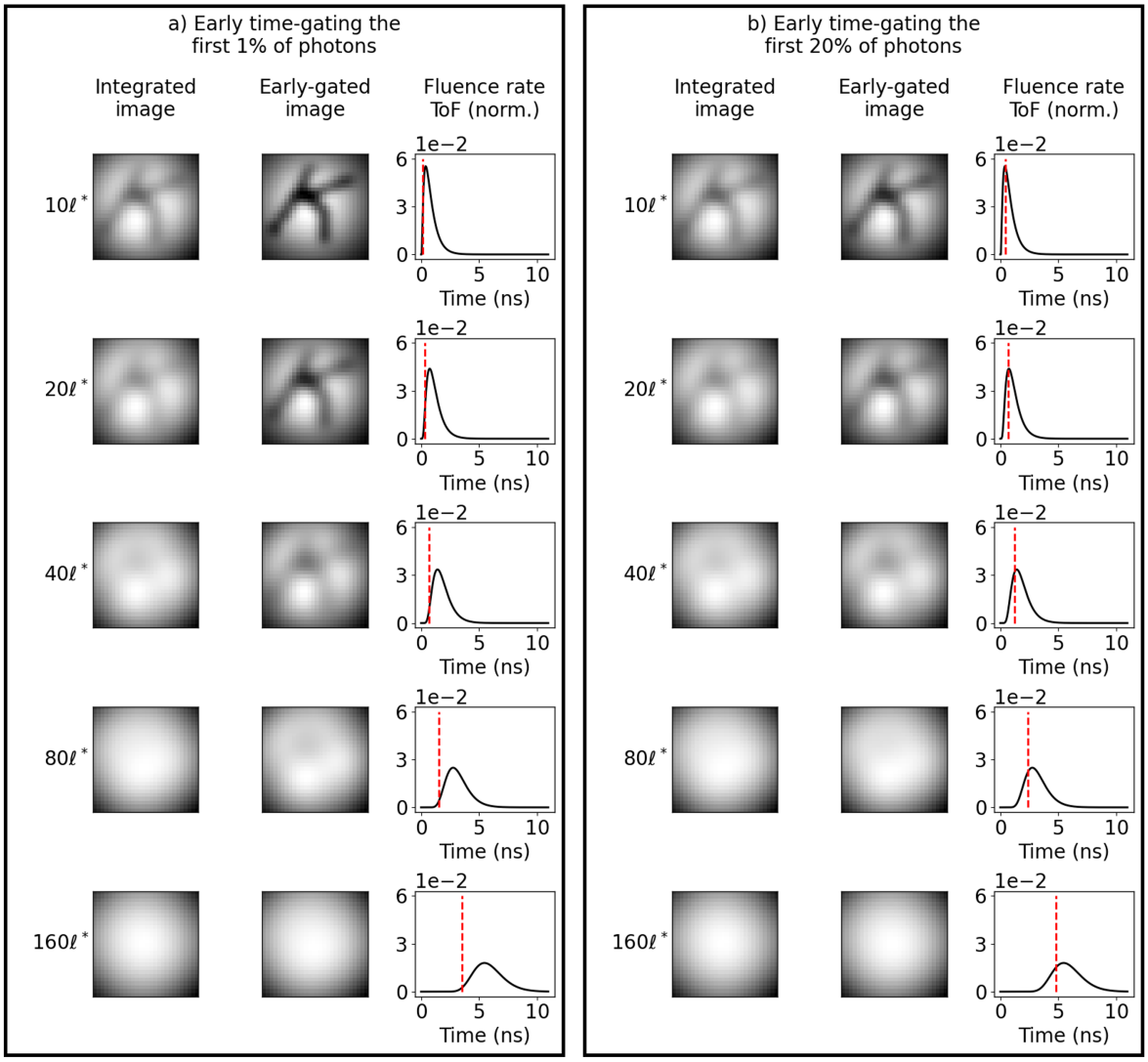


Figure 3.1: Simulated measurements of a hidden object in the shape of the letter "k" through increasing number of transport mean free paths. The middle column of each box shows the enhancement of contrast when early gating a) the first 1% of photons to arrive at the detector and b) the first 20% of photons to arrive at the detector. The right column in each figure shows the pixel aggregate time of flight distribution, where $t = 0$ integrated up to the red dashed line indicates the first 1% and first 20% of photons respectively

assumption overlooks the potential to use the diversity in the shape of the time traces at different detector locations, but admittedly the diversity is most likely negligible in this weakly scattering regime with a limited solid angle of collection.

Another study published by Proskurin [127] also highlights that there is valuable information in diffuse photons. The study concludes that late-arriving photons alone can locate embedded absorbing regions of an otherwise homogeneous scattering material without the aid of any inverse retrieval algorithm. This technique is demonstrated with a cylindrical phantom with a single source fibre and a ring of detection fibres arranged around the circumference. In the case of no

embedded absorber, each detector will measure the same rate of decay of late-arriving photons with respect to time. When an absorber is placed close to a detector fibre, there is a much faster decay rate of late-arriving photons since there is an increased probability that highly scattered photons will be absorbed before reaching the detector. Looking only at the rate of decay of the late-arriving photons in each detection channel, one can estimate the position of the absorber without the requirement of an inverse image retrieval algorithm. Unlike the technique of API, this work does not accurately spatially resolve the shape of a hidden absorbing object. Still, it indicated that the diversity of time traces in spatiotemporal measurements alone could provide information about the existence and whereabouts of hidden inhomogeneities. The experimental results were shown for phantoms corresponding to transmitting photons through $\sim 35\ell^*$ where the diversity of the time traces of measurements is probably more significant than in the work of Satat et al. [11].

Realising diffuse optical imaging in practical settings for imaging deep inside the human body with the goal of replacing modalities such as X-ray or MRI requires a paradigm shift of the field to image beyond $100\ell^*$ - where early photon imaging techniques fail. The challenge of working in this highly diffusive regime can be appreciated by considering the attenuation of transmitted photons through an increasing number of transport mean free path lengths.

The attenuation of ballistic photons is described by the Beer-Lambert law, which depends strictly on the average distance before either a scattering or absorption event [3, 128–130]:

$$\Phi = \Phi_0 \exp\left[-(\mu_a + \mu_s)L\right] = \Phi_0 \exp\left[-\frac{L}{\ell}\right]. \tag{3.1}$$

Note that the exponential trend depends on the number of transport mean free paths $\ell = 1/(\mu_a + \mu_s)$ and not the *reduced* number of transport mean free paths $\ell^* = 1/(\mu_a + \mu_s')$. This is because, by definition, a ballistic photon is transmitted through a material without any interaction and, therefore, cannot be described by the reduced scattering coefficient since there can be numerous scattering events in one reduced scattering length. The reduced scattering coefficient is a combination of the scattering coefficient and the expected angle of scattering ($\mu_s' = (1-g)\mu_s$), which is used to more accurately describe light diffusion in anisotropic materials (see Section (2.1.3)). However, it is strictly the scattering coefficient that determines a photon's mean free path before a scattering interaction. Therefore the scattering coefficient μ_s is the appropriate parameter to use in Eq. (3.1).

Early photon techniques also use quasi-coherent snake photons, which have weakly interacted with the material and only slightly deviate from their initial direction of propagation. The abundance of these photons transmitted to the detector is related to the anisotropy of the material's scattering properties. Determining the exact number of snake photons expected to be transmitted through a material is challenging since they have a loose definition with varying degrees of

interaction. A reasonable estimate can be found by replacing ℓ with ℓ^* in Eq. (3.1) [131], which for forward scattering materials such as biological tissue ($g \simeq 0.9$) would result in one order of magnitude reduction in the exponent compared with ballistic photons. This loosely defines the transition from ballistic to diffusive propagation between $1\ell^* - 10\ell^*$. However, the solid angle of collection of the detector must also be considered in practice since spatial filtering for spatially coherent light can extend this point of transition [132]. Note that for isotropic materials $g = 0$, then $\ell = \ell^*$ and the number of photons which maintain some coherence for early photon imaging techniques decays as per Eq. (3.1).

When in the diffusive regime (i.e., $\mu_s' \gg \mu_a$ and the thickness of material $L \gg \ell^*$) the fluence rate of diffuse photons Φ measured by a detector with a solid angle of collection $\delta\Omega$ can be modelled as [130],

$$\Phi = \Phi_0 \frac{\delta\Omega}{4\pi} \exp\left[-\sqrt{3\mu_a(\mu_s' + \mu_a)}L\right], \tag{3.2}$$

where Φ_0 is the source fluence rate incident on the material.

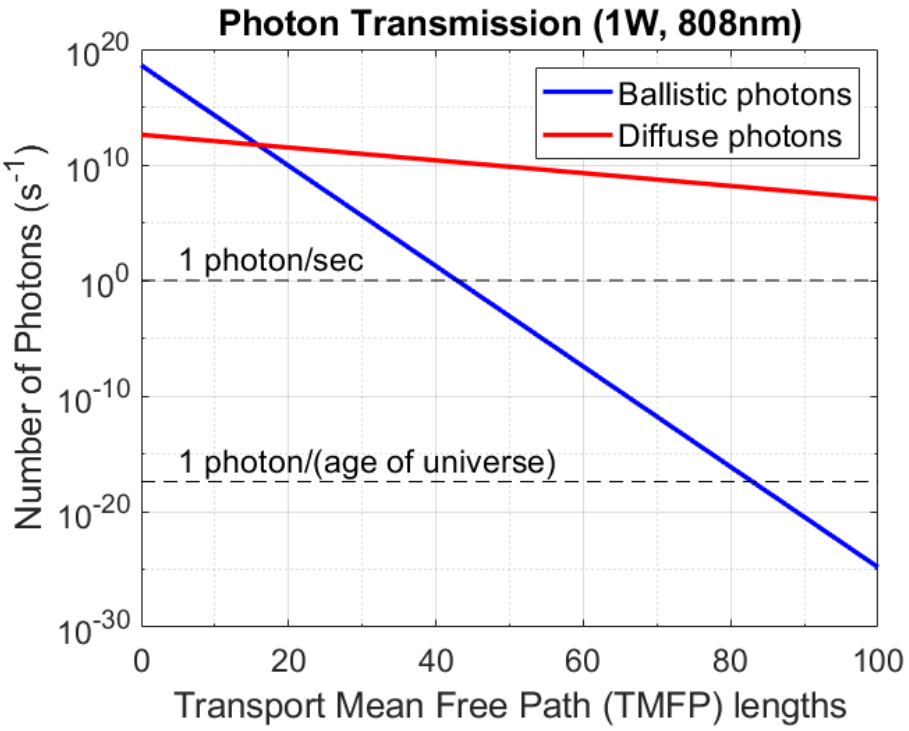


Figure 3.2: Graph published in [133] showing the number of transmitted ballistic photons transmitted per second (blue line) based on Eq. (3.1) for an isotropic scattering material where $\ell = \ell^*$, and the number of diffuse photons per second (red line) based on Eq. (3.2), as the transport mean free path length increases.

Using typical experimental parameters as in the configuration used by Lyons et al. [12] ($\lambda =$ 808nm, $\mu_a = 0.09\text{cm}^{-1}$, $\mu_a = 16.5\text{cm}^{-1}$, $\Phi_0 = 1\text{W/cm}^2$, $L = 5\text{cm}$, $\delta\Omega/4\pi = 10^{-6}$) the expected transmitted number of coherent (ballistic or snake) photons versus incoherent (diffuse) photons

can be calculated and compared as shown in Fig. (3.2). The trends show that, even with a relatively high power source of 1W, collecting ballistic photons is restricted to $< 50\ell^*$ for this experimental configuration, and in the regime of most literature $\sim 10\ell^*$ there are approximately 10^{14} ballistic photons per second corresponding to readily detected microwatt power levels at the detector. However, in the regime of Lyons et al. [12] beyond $80\ell^*$, it is clear that the only photons to reach the detector are incoherent diffuse photons. In fact, at $\sim 80\ell^*$, there exist around 10^{-17} ballistic photons per second, corresponding to one detection per age of the universe. For this reason, the definition of early photons in the context of this study presented in this chapter and [133] is defined not by the collection of ballistic or snake photons but as diffuse photons that arrive before the peak of the time-of-flight distribution.

Although it is known from Fig. (3.2) that there are no coherent photons detected for the measurements made in [12], the purpose of the study presented in this chapter is to demonstrate that the concept of early photon imaging does not aid the retrieval quality in the highly diffuse regime. Instead, the late photons contribute significantly to producing higher-quality image retrievals and sampling the full extent of the temporal statistics of photon counts results in the best image reconstructions.

3.2 Analytical forward model simulation

Using the solution to the diffusion approximation to the radiative transfer equation (Eq. (2.96)) derived in Section (2.2.1), a simulation of diffuse light resolved in space and time can be constructed for homogeneous scattering material with an embedded 2D absorbing object to emulate the experimental configuration used in [12] and shown in Fig. (3.3). The principle of the simulation is first to use the solution to the photon diffusion approximation Eq. (2.96) to determine a point spread function (PSF) given parameters $\left[\mu_a, \mu_s', \vec{r}\right]$. The plane $\vec{r}$ is chosen to be midway through the material where the absorber is located. Once the PSF is calculated, it is convolved with the source's spatiotemporal profile. The result is multiplied by zeros for every element in space and time that interacts with the absorbing object. The resulting array is treated as a new source located midway through the material, and convolving with a second PSF corresponding to the second slab, the distribution of the light in space and time at the output plane of the material can be calculated.

Mathematically, the model is an operator $\mathcal{A}(m)$ which simulates the output profile of diffuse photons in space and time for a given 2D perfect absorber $m(x,y)$ embedded inside a homogeneously scattering material. The laser pulse profile can be any arbitrary spatiotemporal data cube $I(x,y,t)$, but for a femtosecond pulsed laser, the time resolution of the system is determined by the detector ($\sim 10^{-11}$s) which is much longer than the pulse duration ($\sim 10^{-13}$s). Therefore, the source pulse can be approximated by a 2D spatial profile at a single time-bin width set by the

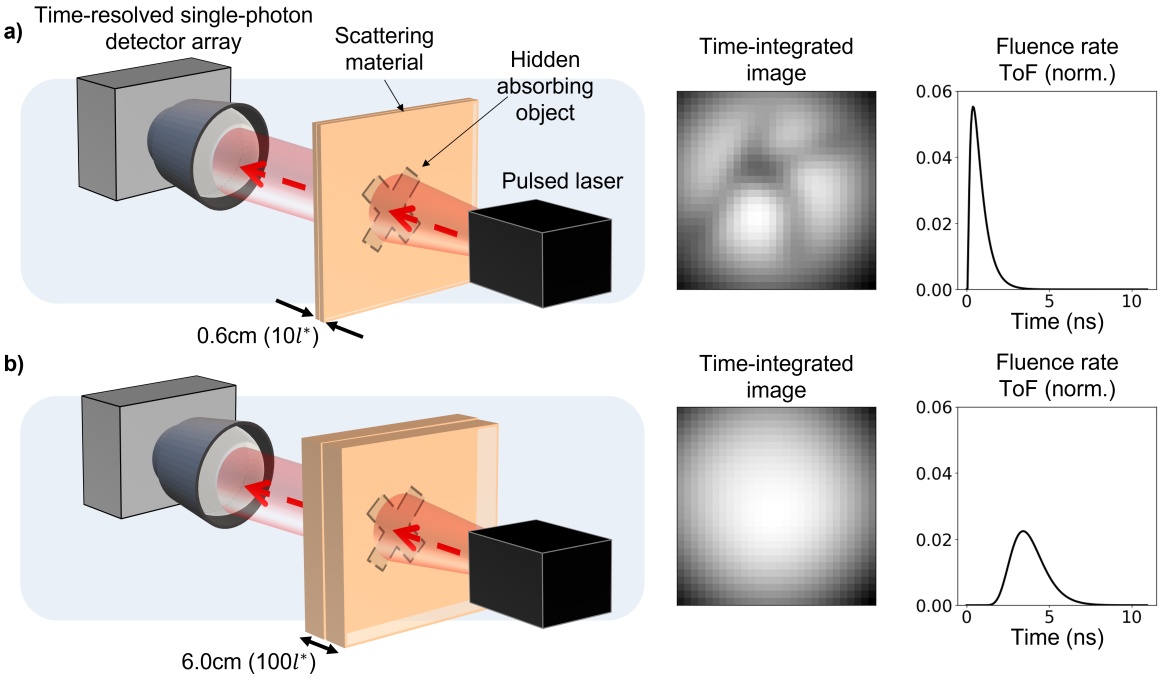


Figure 3.3: An illustration of the experimental configuration in [12] that is emulated by the simulations in this thesis. On the right is examples of simulated images and time traces for a) 0.6cm ($10\ell^*$) and b) 6.0cm ($100\ell^*$) of material ($\mu_a = 0.09\text{cm}^{-1}$, $\mu_s' = 16.5\text{cm}^{-1}$), for the hidden object in the shape of a letter "k" from the EMNIST dataset [134].

resolution of the detector. The laser pulse is convolved with the PSF $\Phi(\vec{r}_m, t)$ using Eq. (2.96) calculated at the plane of the absorber $\vec{r}_m = \sqrt{x^2 + y^2 + L_m^2}$, where L_m is the thickness from the incident surface to the plane of the absorber. The absorption is simulated by element-wise multiplying the result by a binary array in the shape of the embedded object $m(x,y)_T$ repeated for the number of time-bins T. The result is then considered a source located at the plane corresponding to the location of the absorber which is convolved with a second PSF $\Phi(\vec{r}_{out}, t)$ calculated from plane of the absorber to the exit plane $\vec{r}_{out}$ of the material:

$$\mathcal{A}(m) = \Phi(\vec{r}_{out}, t) * \left[\left[I(x,y,t) * \Phi(\vec{r}_m, t)\right] \odot m(x,y)_T\right], \tag{3.3}$$

where $*$ represents a convolution and $\odot$ is an element-wise multiplication.

Using the convolution theorem (i.e., $f * g = \mathcal{F}\{f\}\mathcal{F}\{g\}$) and appropriately applying zero-padding, the simulation can be made more computationally efficient:

$$\mathcal{A}(m) = \text{Re}\left[\mathcal{F}^{-1}\left\{\mathcal{F}\left\{\Phi(\vec{r}_{out}, t)\right\}\mathcal{F}\left\{\phi_m \odot m(x,y)_T\right\}\right\}\right], \tag{3.4}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote a Fourier transform and an inverse Fourier transform, respectively. The term ϕ_m is the result of the convolution of the laser intensity profile $I(x,y,t)$ with the point spread

function of the fluence rate at the object plane $\Phi(\vec{r}_m, t)$:

$$\phi_m = \mathrm{Re}\left[\mathcal{F}^{-1}\left\{\mathcal{F}\left\{I(x,y,t)\right\}\mathcal{F}\left\{\Phi(\vec{r}_m,t)\right\}\right\}\right]. \tag{3.5}$$

When normalised by the area, the final result is a probability mass function (PMF) of photon arrival at the detector resolved in space and time. To calculate the fluence or number of photons at the output, the PMF can be multiplied by the the expected quantities calculated using equations (3.1) and (3.2). When simulating photon counts it is also important to discretise e.g., by rounding down to the nearest integer or numerically sampling the PMF using a random number generator.

For the inverse retrievals presented in this chapter, the forward model included in the algorithm developed by Lyons et al. [12] is used. However, the simulations shown in Fig. (3.1) and Fig. (3.3), and all simulations in the remaining chapters of this thesis, use an amended version of the forward model in [12] that I have developed in Python. The forward model developed in [12] and the amended Python version use the same mathematical principles described in this section.

3.3 Methods

The methodology of this numerical study of early- and late-arriving photons is to use experimental data and inversion methods developed by Lyons et. al. [12] to investigate the contribution of different segments of time-of-flight data to the final image reconstruction. This is performed by truncating the data for different segments of time by replacing the actual values with zero and assessing the impact on the reconstructed image quality. The simulated data are truncated in the same way to avoid large gradients due to the residual between the zero'd time-bins and simulated values. The results are quantitatively compared using image comparison metrics which were chosen to indicate the accuracy of determining the location of the hidden object and the similarity in shape between the target and the reconstructions.

3.3.1 Experimental layout

The experiment described in [12] consists of a hidden absorbing piece of black tape in the shape of the letters "A" and "X" placed between two pieces of polystyrene foam as shown in Fig. (3.3). A pulsed laser beam (808nm, 1W, 130fs, 80MHz) was expanded using a diverging lens to 5cm diameter and aligned along the same optical axis as the centre of the hidden object and the detector array.

The detector is a single photon avalanche diode (SPAD) array developed by Richardson et. al. [33] and commercialised by Photon Force. There are 32x32 SPAD pixels in the array, each have a built-in time-to-digital (TDC) converter which produce a time-tag label with 55ps resolution. The

time-tags are used to produce a histogram of the time difference between the laser pulse emission and the arrival time of the photon for every pixel in the array. This results in a spatiotemporal distribution of photon counts at the detector.

The TDC of each SPAD is 10-bit and so has an upper limit of 1024 possible time-tag values. However, the first time-bin is used to record when no photon is detected within one period of the laser repetition rate which results in an upper limit of 56.265ns. Furthermore, since the repetition rate is 80MHz, the TAC is restricted to approximately 12.5ns and therefore only 230 time-bin values are used in the TPSF. For this experimental configuration, the total time spread of photon arrivals above the noise level is less than 6ns and so only 100 time-bins are used in the image reconstruction process to reduce computational memory requirements. An example of the preprocessed data when integrated in time is shown in Fig. (3.4) which demonstrates that there are no discernible features of the ground truth image present, even when time-gating the data for early photons. However, it is the full spatiotemporal profile that is used as input to the reconstruction algorithm. Note that the data presented in the bottom row of Fig. (3.4) is smoother due to increased filtering of the data during preprocessing performed by the authors of [12]. The time-gated unprocessed data can be found in [12], which also shows no resemblance of the target.

The material properties were determined to be $\mu_s' = 16.5\text{cm}^{-1}$ and $\mu_a = 0.09\text{cm}^{-1}$ by the authors of [12], corresponding to a transport mean free path length of $\ell^* = 603\mu\text{m}$. Since the thickness of each polystyrene slab is 2.5cm, the total length of material is equivalent to $83\ell^*$.

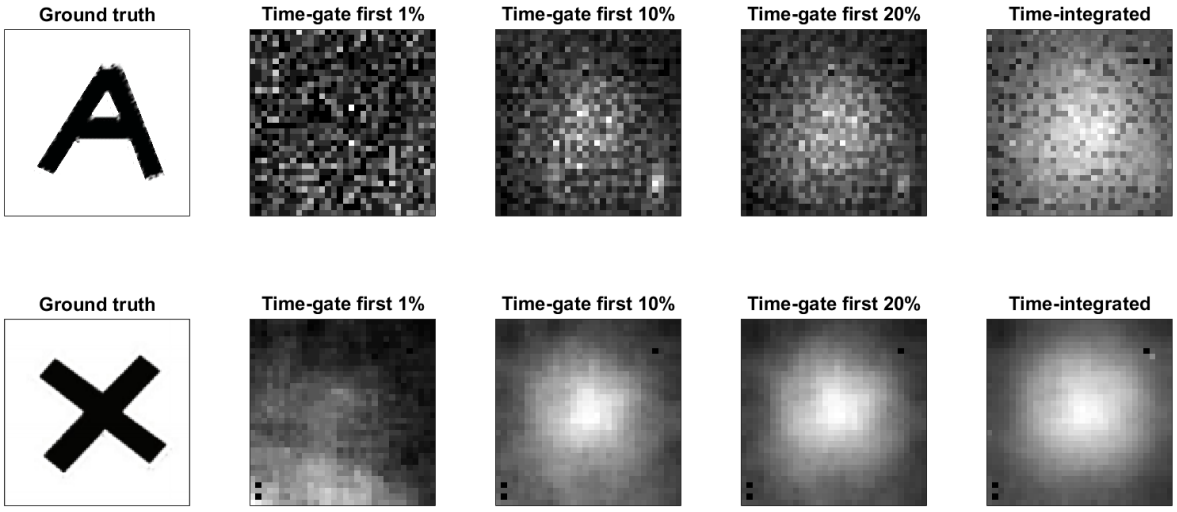


Figure 3.4: The ground truth images and early photon time-gated images of the preprocessed data used in [12] and [133]. From left to right shows integrating the first 1%, 10%, and 20% of photon arrival times as well as integrating over all arrival times. The data for the "X" in the bottom row looks smoother because it has been filtered more during preprocessing by the authors of [12]. The time-gated unprocessed data can be found in [12] which also shows no resemblance of the target.

3.3.2 Image retrieval algorithm

To reconstruct the hidden object from the measurements at the SPAD array, the iterative regularised least-squares algorithm developed by Lyons et al. [12] was used. A least-squares optimisation is a convex formulation of the problem as discussed in Section (2.6) whereby the square of the residual between the measured data Y and the output of the forward model $\mathcal{A}(m)$ (Eq. 3.3) is minimised to find the hidden object m:

$$\arg\min_m\{O(m)\} = \arg\min_m\Big\{\,||\mathcal{A}(m)-Y||_2^2 + \lambda_1||m||_{TV} + \lambda_2||m||_1\,\Big\}, \tag{3.6}$$

where $||m||_{TV}$ and $||m||_1$ are the total variation norm and L_1-norm regularisers weighted by the regularisation parameters λ_1 and λ_2 which are tunable parameters. The effect of the regularisation parameters is discussed in more detail in the following subsection (Section (3.4)).

This minimisation can be solved iteratively using gradient descent:

$$m^{(k+1)} = m^{(k)} - \alpha O'(m), \tag{3.7}$$

where k is the iteration number, α is the step size and $O'(m)$ is the gradient of the loss function $O(m)$ in Eq. (3.6).

Reconstructing an image of a hidden absorbing object from diffused light is an ill-posed problem and ill-conditioned. As a result, small changes due to small noise contributions in the input can lead to large errors in the solution that can be very far from the true target in the space of solutions but still satisfy the minimisation of the residual term in Eq. (3.6). To constrain the problem and achieve more accurate results, the possible solutions of the minimisation which are less likely are penalised using regularisation terms. The regularisation terms is the addition of *a priori* information which explicitly bias the solution to contain known physical characteristics of the target object.

The first regularisation term enforces the total variation norm (TV-norm) to be minimised. The TV-norm is the sum of the gradient of the image:

$$||m||_{TV} = \sum_{i,j}\sqrt{|m_{i,j}-m_{i+1,j}|^2 + |m_{i,j}-m_{i,j+1}|^2}. \tag{3.8}$$

A small value of the TV-norm promotes uniformity in the image whilst preserving sharp step-like features. The inclusion of this regularisation term is motivated by the expectation of the reconstructed image to contain a binary shape with a constant value in its interior and sharp boundaries around the edges [12].

The second regularisation term minimises the L_1-norm of the reconstruction. This norm is the

absolute sum of all of the pixel values in the image:

$$||m||_1 = \sum_{i,j} |m_{i,j}| \tag{3.9}$$

where |.| denotes the absolute value. The effect of minimising this norm is to promote sparsity; in other words, it increases the number of pixels with the value zero. The motivation for using this regularisation term in this context is because the reconstructed image is expected to be binary whereby the area of the object consists of value one on a background of zeros [12].

A derivation of the gradient $O'(m)$ of the loss function $O(m)$ described in Eq. (3.6) is presented in Appendix (A).

3.4 Image reconstructions with early and late photons

It is demonstrated in Fig. (3.2) that no ballistic photons exist in measurements beyond $80\ell^*$. Therefore, in the experimental data, we define early photons as arriving before the peak of the pixel-aggregated histogram and late photons as arriving after the peak. The experimental data used in [12] is cropped accordingly to study the sensitivity of the reconstructions to the early and late segments of the distribution compared with the full spatiotemporal measurement. This is performed by multiplying each pixel in the data with a unit step function which selects a subset of the full distribution. The same step function is applied in the forward model to avoid the retrieval fitting to the regions with zero counts in the measurements.

The image retrieval algorithm described in Section (3.3.2) is sensitive to hyperparameters. The step size α and regularisation parameters λ_1 and λ_2 can affect the quality of the reconstructed image. Adjusting each hyperparameter independently and comparing the retrieved images determined that only α and λ_1 were significantly sensitive and required optimisation. A linearly spaced coarse-scale parameter grid search was performed for the expected range of both sensitive parameters. The retrievals scoring the best match with the ground truth image were optimised further with a fine-scale parameter search. A combination of the mean squared error (MSE) and the multi-scale structural similarity (MS-SSIM) metrics were used to compare the retrieved images with the ground truth image.

The MSE is a standard image processing metric which sums the squared residual between the ground truth image X and the reconstructed image $\hat{X}$:

$$\text{MSE} = \frac{1}{N} \sum_{i}^{N} (X_i - \hat{X}_i)^2, \tag{3.10}$$

where N is the number of pixels. This metric determines whether the values in the reconstruction

directly overlap with the ground truth image. However, in practice, unaccounted experimental effects such as the incidence angle of the source beam not being exactly normal with the material surface or lens aberrations from a mispositioned lens could cause the reconstruction to have accurate structural features that do not perfectly overlap with the ground truth image.

The MS-SSIM metric is particularly suited to quantitatively assessing the structural content of images when they differ by noise or distortions. The structural similarity (SSIM) metric is derived by combining functions that compare luminance, contrast and structure between images [135]. First, the images are blurred by a pre-defined Gaussian blur kernel which is chosen based on the expected feature sizes and improves the robustness of the measure to noise and spurious artefacts. A scanning window of the same dimension as the blur kernel calculates the mean μ_x and $\mu_{\hat{x}}$, standard deviations σ_x and $\sigma_{\hat{x}}$, and covariance $\sigma_{x\hat{x}}$ for every scan position of the windows x and $\hat{x}$ of the ground truth (X) and predicted ($\hat{X}$) images. The result is a map of SSIM values can be produced using [135]:

$$\mathrm{SSIM}(x,\hat{x}) = \frac{\left(2\mu_x\mu_{\hat{x}} + C_1\right)\left(2\sigma_{x\hat{x}} + C_2\right)}{\left(\mu_x^2 + \mu_{\hat{x}}^2 + C_1\right)\left(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2\right)}, \tag{3.11}$$

where the arbitrary constants C_1 and C_2 are added for stability when the denominator approaches zero. Taking the mean of the SSIM map in Eq. (3.11) (referred to as the MSSIM) provides a global estimate of the similarity between images. Multi-scale SSIM is a more general metric than the single-scale MSSIM described above. In the calculation of multi-scale SSIM, the images are iteratively blurred and downsampled by a factor of two and the MSSIM of each iteration is combined by their geometric mean. Each iteration is also weighted, typically with a Gaussian distribution since the human visual system is optimal in the middle frequencies [135]. The MS-SSIM metric can have values $[-1, 1]$ where one corresponds to the maximum similarity between the images, zero corresponds to no similarity, and an MS-SSIM of negative one represents the maximum dissimilarity between the images. In this work, the Matlab function `multissim()` is used to calculate the MS-SSIM metric using the default arguments.

The purpose of comparing the retrieved images with the ground truth with the absolute error metric of the MSE is to give an insight into the accuracy of the object's location compared with the position of the ground truth. Whereas, the MS-SSIM metric quantitatively asses the similarities in the geometric features. If the images are identical, the MSE is zero, and the MS-SSIM is one.

3.4.1 Results

The results in Fig. (3.5) show that using late photons exclusively can reconstruct images with comparable fidelity compared to using early photons exclusively in the highly diffusive regime

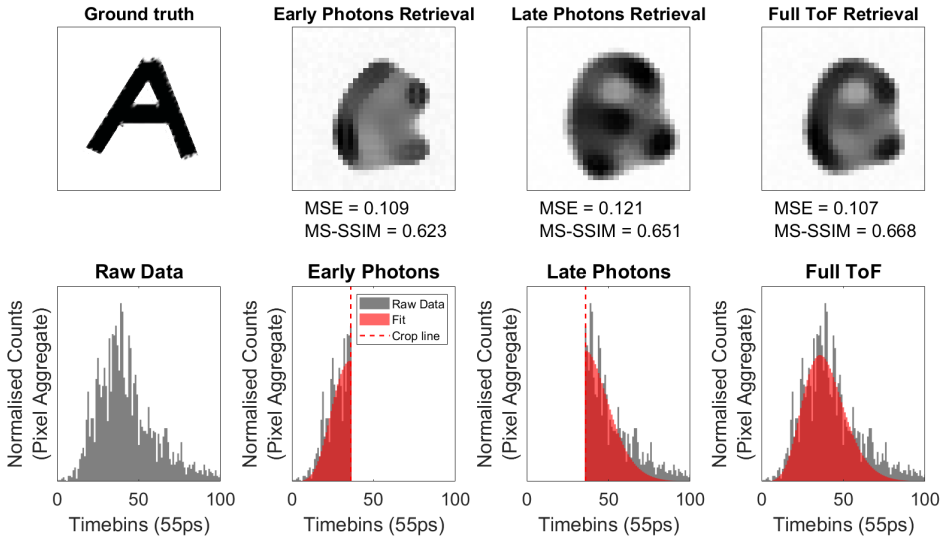


Figure 3.5: Figure published in [133] showing the ground truth of hidden object in the shape of the letter "A" and the optimised reconstructed images (top) for the same experimental data. The data is exclusively windowed for early- and late-arriving photons and the full time of flight (ToF) information (left to right). The mean squared error (MSE) and multi-scale structural similarity (MS-SSIM) metrics are with reference to the ground truth.

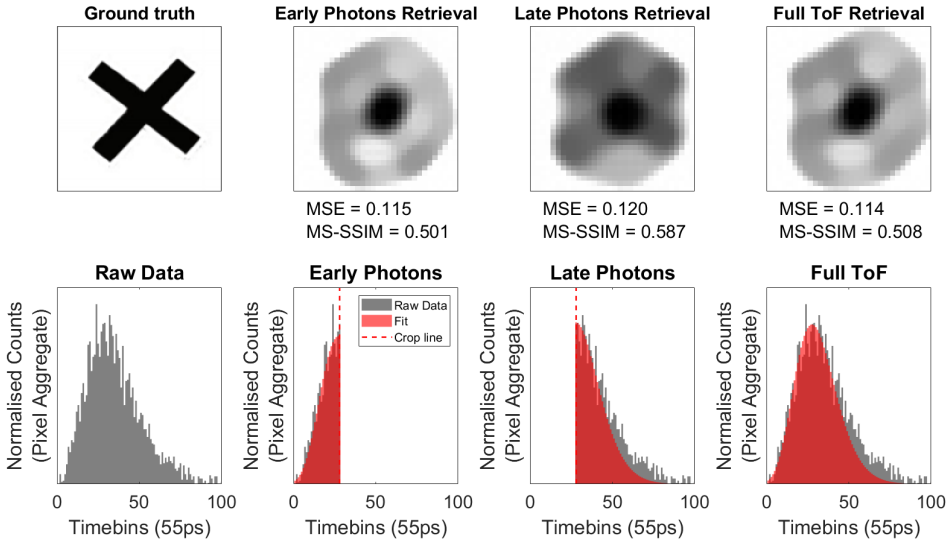


Figure 3.6: Figure published in [133] showing the ground truth of hidden object in the shape of the letter "X" and the optimised reconstructed images (top) for the same experimental data. The data is exclusively windowed for early and late arriving photons and the full time-of-flight (ToF) information (left to right). The mean squared error (MSE) and multi-scale structural similarity (MS-SSIM) metrics are with reference to the ground truth. The purpose of this figure is to show consistency and validation of the findings in Fig. (3.5) using a second independent measurement from the same experiment.

beyond $80\ell^*$ considered. This contradicts the concept of early photon imaging, which aims to remove the late photons since they have less direct information. However, as previously discussed, in the regime $> 80\ell^*$, there is no direct imaging information in the early or late photons. Instead, the diversity of the spatiotemporal distribution of photon arrivals at the detector array contains indirect statistical information about photon interactions with an absorber. For this reason, it is better to sample the full distribution to increase the sensitivity of the inverse retrieval algorithm to find diversity in the shape of time-of-flight (ToF) histograms.

A second independent measurement from the same experiment validates this conclusion is shown in Fig. (3.6) using a different absorber in the shape of the letter "X". Similar to the results in Fig. (3.6), the early photons have lower MSE but higher MS-SSIM than retrievals using late photons. This can be explained by comparing the common features in the reconstructions of both figures. Reconstructions with late photons have larger structural features that closely resemble the ground truth's geometry but have regions of blur spanning a large area. Whereas reconstructions with early photons have less blur, and the structure is less defined. In both cases, using the full distribution of photons gives the best trade-off between structural definition and blur. However, this discussion of features present in reconstructions using different time-gates is likely to be specific to the chosen image retrieval algorithm, and these conclusions cannot be expected to be general for all algorithms.

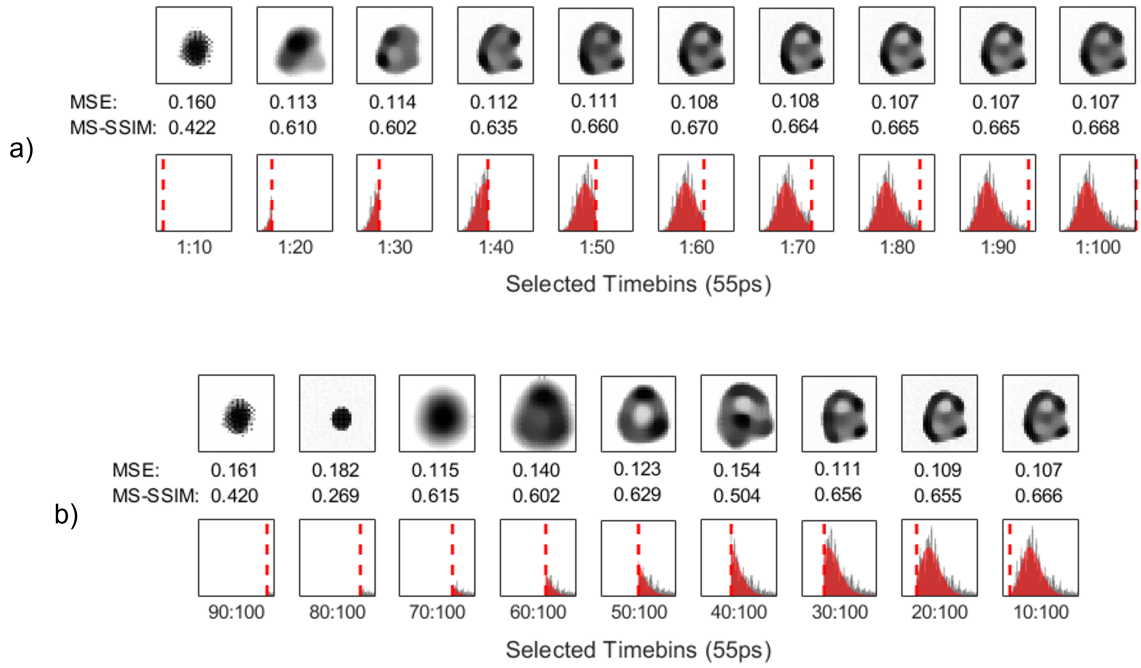


Figure 3.7: a) The optimal reconstructed images and image comparison metrics with the ground truth shown in Fig. (3.5) when including an increasing number of late time-bins in the measurement (bottom row). b) The optimised reconstruction and metrics when including an increasing number of early time-bins. The trend in both cases shows that including the full distribution of data provides the most faithful reconstruction.

The analysis was extended further to quantitatively assess the trend of the reconstruction fidelity when incrementally increasing the number of late photons included in the data. The results in Fig. (3.7a) show that increasing the number of late time-bins in the data improves the accuracy of image reconstructions. Interestingly, this is also true when including an increasing number of early time-bins Fig. (3.7b). In both cases, the trend in the MSE and MS-SSIM metrics indicate that including the full distribution of photons produces the most accurate image retrievals, which have the highest correlation in terms of absolute position and similarity of geometric features compared with the ground truth shown in Fig. (3.5). This result further validates the conclusion that the inverse retrieval algorithm is sensitive to the full spatiotemporal distribution of photon counts and demonstrates that the diversity of every time-bin in every pixel contributes to the final reconstructed image quality.

3.4.2 Hyperparameter optimisation

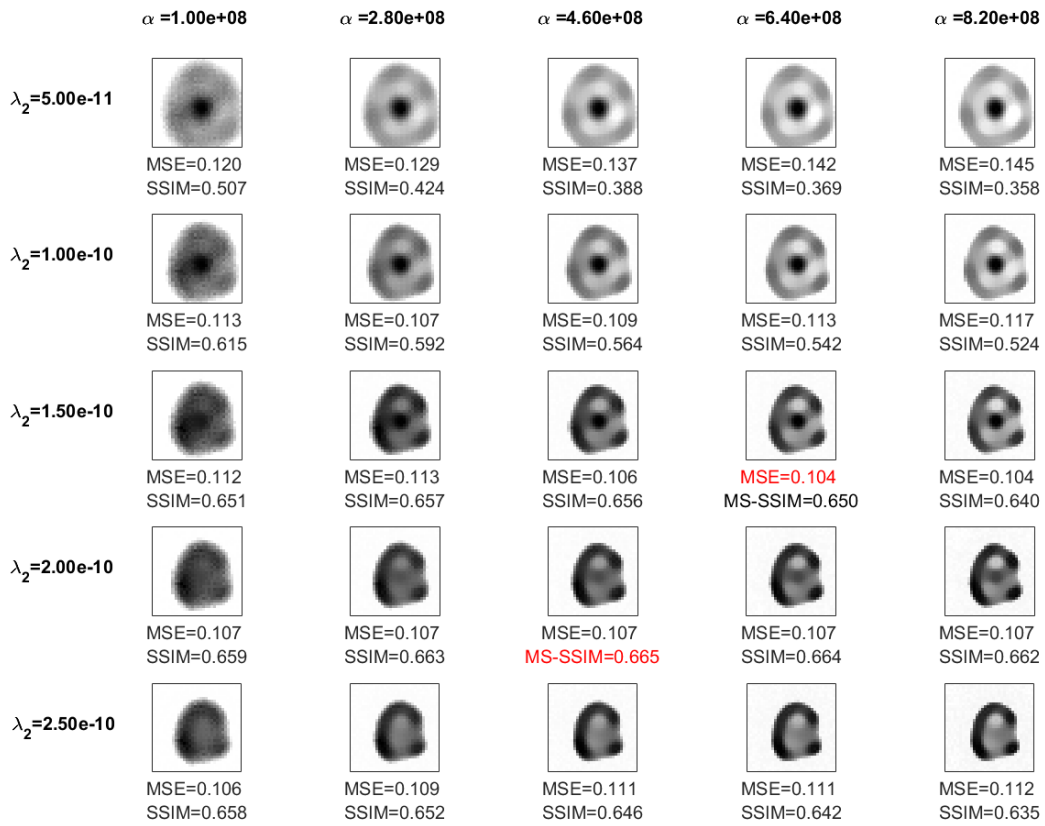


Figure 3.8: An example grid search of the step size α and the L_1 regularisation weight λ_2 which shows retrievals when using the full distribution of photon arrival times in the measurements corresponding to the ground truth hidden object "A" as shown in Fig. (3.5). The image comparison metrics with the ground truth are shown below each retrieval, where the minimum MSE and maximum MS-SSIM are highlighted in red. The optimised image reconstruction is found by further tuning the parameters around the best values indicated in the grid search.

As discussed above, the quality of the reconstructed images using the minimisation algorithm in Eq. (3.6) is significantly sensitive to the step size α and the L_1-norm regularisation weight λ_1. To optimise these hyperparameters for reconstructions in Fig. (3.7), a 10×10 grid search was performed for $10^8 < \alpha < 10^9$ in increments of 10^8, and $10^{-11} < \lambda_2 < (3 \times 10^{-10})$ in increments of 3.22×10^{-11} for a fixed number of 2000 iterations.

A smaller 5×5 grid search of the parameters used for the full ToF reconstruction presented in Fig. (3.5) is shown as an example in Fig. (3.8). The optimised retrieval is found by further tuning the hyperparameters around the values in the grid search that correspond to the best metric values highlighted by red text.

The parameter search in Fig. (3.8) also shows the effect of the L_1-norm regulariser to increase the number of zeros in the retrieval when the λ_2 weighting parameter is increased. Whereas, the increased step size progresses the minimisation towards a solution quicker for the same number of iterations. However, a step size too large can overshoot the minima and the solution is further from the ground truth.

Each retrieval takes around three minutes, which corresponds to > 100 hours computing time in total to perform all of the hyperparameter grid searches for the reconstructions shown in the figures of this chapter.

3.5 Conclusion

In the highly diffusive regime beyond $80\ell^*$, no ballistic or snake photons exist to perform early photon imaging techniques, and all photons measured by the detector are diffuse. The key contribution of this study is to show the importance of sampling of the full spatiotemporal distribution of detected photon counts when reconstructing images in the highly diffusive regime beyond $80\ell^*$ using the data and retrieval algorithm from Lyons et al. [12]. Contrary to the concept of early photon imaging, including the full distribution of photon arrival times in the retrieval process can improve the accuracy of the reconstructions. Furthermore, using late-arriving photons exclusively, which are commonly assumed detrimental to the image reconstruction process, can reconstruct images with similar fidelity to using early photons exclusively. Although none of the detected photons has direct image information about absorbing regions in the material, the diversity in the spatiotemporal statistics of photon arrival at the detector array can indirectly provide information which can be retrieved using the iterative regularised least-squares algorithm provided with an analytical forward model of diffuse photon propagation. This indirect information is critical for developing methods to image beyond $100\ell^*$ where early photon imaging fails, and can be leveraged to realise the potential of diffuse optical imaging in real-world applications such as imaging deep inside the human body and brain.

Chapter 4

Information analysis of diffuse photon measurements

Motivated by the sensitivity study of spatiotemporal measurements of photon diffusion in the previous chapter, this chapter examines the information contained in the data more closely using the quantitative information metric of Shannon entropy. Whereas the previous chapter considered the sensitivity of the inverse retrieval algorithm used in [12] to different segments of the spatiotemporal distribution of photon counts, the study presented in this chapter is agnostic to how an image can be retrieved from the measurement and instead considers only the extent of dependent relationships between input (the hidden object) and the output (spatially-resolved TCSPC measurements).

One of the most significant early contributions to the field of time-resolved diffuse optical imaging was the validation of the computationally inexpensive solution to the photon diffusion approximation to the radiative transport equation by Patterson et al. [21] (Eq. (2.96)). In the conclusions of this work, they speculate that there is much to be explored by studying the temporal evolution of light as it propagates in diffusive media, and that this temporal information is likely to have a key role in aiding image retrieval processes to locate absorbing inhomogeneities in the human breast and brain non-invasively. They claim that this computationally inexpensive alternative to Monte Carlo models could present an opportunity to create efficient simulations to allow studies to gain insight about the importance of the time evolution of light propagation through diffusive materials. However, aside from observation that using the moments of the time-of-flight distribution reduces the ill-conditioning of the inverse problem of image retrieval [97, 105], literature refers to time-resolved measurements as a method of gaining "superior information" compared with continuous-wave and frequency-domain techniques, but do not explicitly quantify the contribution of this additional information content in the time dimension for reconstructing images [98, 104, 107, 136].

The conclusion of the previous chapter was that measuring spatially-resolved photon counts for the full extent of the temporal impulse response of the material improved the reconstruction of an embedded absorbing object inside a diffusive material. However, it is interesting to compare the information resolved in space and time with measurements which are resolved only in space (i.e., continuous-wave configuration). The following information theory study presented in this chapter will compare measurements resolved in space, time and space-time to quantitatively evaluate the sensitivity of measuring photon counts in each domain for the task of reconstructing images.

Furthermore, the quantitative nature of using an information theory approach provides insight regarding the limits of detecting useful measurements about a hidden object embedded inside a diffusive material for increasing number of transport mean free path lengths. Although this is a numerical analysis, the studies presented in this chapter use typical experimental parameters and idealistic noise considerations to indicate the maximum number of transport mean free paths which could be imaged through using today's technology. An experimental validations of the assumptions used in the numerical models is presented in Section (4.5).

4.1 Quantifying information with Shannon entropy

To quantitatively asses the information content in diffuse light measurements, statistical tools such as Shannon entropy and mutual information which derive from information theory provide the most general approach compared with other correlation metrics[1].

Shannon entropy describes the expected reduction in uncertainty when observing the outcome of a random variable. It is the fundamental metric used in the mathematical field of information theory and best explained using an example. First consider a fair coin toss, revealing the value has reduced uncertainty by two-fold. To communicate the result of a coin toss, binary values of one and zero can be used. Therefore there is 1 bit of information in a fair coin toss. Now consider a fair die which, when rolled, there are six equally likely possibilities. By revealing the value, the uncertainty is reduced six-fold. To communicate the result of a die roll as a binarised message, at least three bits are needed since two bits can represent only the values from one to four. Mathematically, the number of bits required to represent six equiprobable outcomes is $\log_2 6 = 2.58$ bits. This defines Shannon information as the number of bits required to communicate the reduction in uncertainty given the outcome of a random variable. In the example of a fair die roll, there is a surplus of 0.42 bits which is redundant when using a 3-bit encoding

[1] E.g., other statistical linear metrics such as Pearson's correlation coefficient or non-linear approaches such as Spearman's rank correlation coefficient are restricted in their generalisation to complicated dependent relationships in data sets and make assumptions about the underlying statistics of the variables. Information theory metrics such as mutual information do not make prior assumptions about the distribution of variables and are sensitive to any dependent relationship type [137].

scheme. This highlights the difference between the information content due to the distribution of the random variable (2.58 bits) versus the information contained in the distribution used to encode and represent the variable (3 bits).

The trivial example of rolling a die above assumes that each roll has an equal probability. If a biased die is used which has two identical faces then the number of outcomes is reduced to five with probabilities $\{\frac{1}{3}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$. In which case, the expected value determines the number of bits [138]:

$$H(X) = -\sum_{i}^{N} p(x_i) \log_2 p(x_i), \tag{4.1}$$

where $X = \{x_1, x_2, ..., x_N\}$ denotes random variable (in this case the possible outcomes of each die roll) and $p(x_i)$ is the probability of sampling the i^{th} outcome. The expected value of the reduction in uncertainty measured in bits in Eq. (4.1) is known as Shannon entropy and is the general form of Shannon information used as the foundation for the majority of information theory. Note that the negative sign is due to using probability as an argument to the logarithmic function, which is the reciprocal of the reduction in uncertainty used in the definition of Shannon information above. Continuing with the example of the biased die, the entropy/information of the distribution of outcomes is 2.25 bits, this is lower than for the unbiased case since, although there are still six sides to the die, there is less surprise when observing the outcome that is labelled on two of the sides.

The examples above assume all events are independent; however, this is often not the case. If, by some magic, rolling a six on an unbiased die causes it to be modified such that it becomes more likely to roll a four on the subsequent roll, then this results in a dependency in the statistics between these outcomes. The governing statistical criteria to determine dependence between two random variables is that when the product of the marginal distributions is equal to their joint distribution, then the variables are independent, i.e., the independence criterion $P(X)P(Y) = P(X,Y)$ is only true if variables X and Y are statistically independent. In the context of the example, for a fair die, rolling a six or a four both have a probability $\frac{1}{6}$. If the product of these probabilities is equal to the joint probability of rolling a six followed by a four $\left(\frac{1}{36}\right)$, then these outcomes are independent. For clarity, the first variable X in this example is the outcome of the first throw, and the second variable Y is the outcome of the subsequent throw. If this independence condition is violated, then a relationship exists between the variables whereby knowing the outcome of one variable (e.g., the first roll) can reduce the uncertainty about the outcome of the second variable (e.g., the second roll).

When observing a random variable reduces the uncertainty of another random variable, there is mutual information between them. This is a measure of how much information is shared

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10
Roll 1 (X)	1	2	3	5	1	2	1	6	3	4
Roll 2 (Y)	4	2	6	4	3	3	4	4	5	3

Table 4.1: An example of data where a biased die is rolled two times for every trial and the objective is to determine if there is a dependency between the first (X) and second (Y) rolls. Since only ten trials are used, this data is under-sampled but it is sufficient to serve as an illustration for calculating independent and dependent information using Shannon entropy.

	X=1	X=2	X=3	X=4	X=5	X=6	P(Y)
Y=2		1/10					1/10
Y=3	1/10	1/10		1/10			3/10
Y=4	2/10				1/10	1/10	4/10
Y=5			1/10				1/10
Y=6			1/10				1/10
P(X)	3/10	2/10	2/10	1/10	1/10	1/10	

Table 4.2: The joint probability matrix $P(X,Y)$ and marginal distributions $P(X)$ and $P(Y)$ for the data in table (4.1). Note that the Y=1 row is omitted since this case does not occur in the data set.

between the two variables, which can be calculated by subtracting the amount of independent information from the total information of each variable. For example, take the fictitious results of multiple trials of consecutive rolls of a biased die in table (4.1)[2]. The first step is to determine the joint probability matrix by calculating the probability of each possible trial outcome. Then the marginal probabilities are found by calculating the number of times an observable occurs in the first roll divided by the number of trials, and likewise for the second roll. This information is displayed in table (4.2) and will serve as a reference for calculating the quantities for determining mutual information.

The Shannon entropy of each variable in table (4.1) can be calculated with equation (4.1) using the marginal probabilities in table (4.2), which yields $H(X) = 2.45$ bits and $H(Y) = 2.05$ bits. Likewise, the Shannon entropy of the joint probability distribution is $H(X,Y) = 3.12$. Using the rule of logs that $\log ab = \log a + \log b$, then the independence criterion becomes a summation rather than a product of marginals when considering the logarithm of the probabilities, i.e., $P(X)P(Y) = P(X,Y)$ becomes $\log P(X) + \log P(Y) = \log P(X,Y)$. Therefore, if the summation of the marginal entropies is equal to the joint entropy, then the information is independent. However, when the difference is non-zero, there must be shared information between the two variables. This leads to the definition of mutual information:

$$MI(X;Y) = H(X) + H(Y) - H(X,Y). \tag{4.2}$$

[2]Note that there are only ten trials in this data, and the probabilities are severely undersampled. Undersampling a distribution will cause mutual information to be overestimated. However, the toy data set in this table is sufficient as a simplified example to illustrate the calculation of independent and mutual information of two random variables.

For the example data in table (4.1) there is $H(X) = 2.45$ bits of information in the first roll and $H(Y) = 2.05$ bits of information in the second roll which is a combined total information of 4.5 bits in the data set. Of the total, $H(X,Y) = 3.12$ bits are independent, which leaves $MI(X;Y) = 1.38$ bits of shared information. This quantifies the relationship between the variables since knowing the observations of either the first roll or second roll exclusively can reduce the uncertainty about the values in the other roll and provide 1.38 bits of information (i.e., over half the information about one of the variables is contained in the other).

This is a useful measure of the compressibility of a variable, and this calculation can provide insight into redundant information in cases where the variables are measuring similar information (e.g., neighbouring pixels of a camera when imaging a blurry object). Note that in this example, the value of the second roll is a conditional probability based on the outcome of the first roll. If, instead, a biased die was used for every second roll which preferentially had outcome four, there would not be any mutual information since the product of the marginal distributions would equal the joint probability. In other words, the probability of rolling four is not affected by the outcome of the first roll.

The expression for mutual information in (4.2) is more commonly expressed as the Kullback-Liebler divergence (KL-divergence) between the product of the marginals and the joint distribution [139, 140]. The KL-divergence is a statistical measure of how similar distributions are. It can be used to calculate the statistical distance between the product of the marginal distributions and the joint probability distribution, which is equivalent to mutual information. This can be proved by rewriting (4.2) using the explicit form of entropy in (4.1). The derivation of the mutual information expressed as a KL-divergence below uses the rule of logarithms, a substitution trick of using $\sum_{X,Y} p(x|y) = \sum_{X,Y} p(y|x) = 1$ to get double sum operators, and the use of the product rule $p(x|y)p(y) = p(y|x)p(x) = p(x,y)$ to calculate the joint probability:

$$
\begin{aligned}
MI(X;Y) &= H(X) + H(Y) - H(X,Y) \\
&= -\sum_{i=1}^{M} p(x_i)\log_2 p(x_i) - \sum_{j=1}^{N} p(y_j)\log_2 p(y_j) - H(X,Y), \\
&= -\sum_{i=1}^{M} \left[p(x_i)\log_2 p(x_i)\times(1)\right] - \sum_{j=1}^{N} \left[p(y_j)\log_2 p(y_j)\times(1)\right] - H(X,Y), \\
&= -\sum_{i=1}^{M} \left[p(x_i)\log_2 p(x_i)\times\left(\sum_{j=1}^{N} p(y_j|x_i)\right)\right] \\
&\quad -\sum_{j=1}^{N} \left[p(y_j)\log_2 p(y_j)\times\left(\sum_{i=1}^{M} p(x_i|y_i)\right)\right] - H(X,Y), \\
&= -\sum_{i=1}^{M}\sum_{j=1}^{N} p(y_j|x_i)p(x_i)\log_2 p(x_i) - \sum_{j=1}^{N}\sum_{i=1}^{M} p(x_i|y_i)p(y_j)\log_2 p(y_j) - H(X,Y),
\end{aligned}
$$

$$
\begin{aligned}
&= -\sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log_2 p(x_i) - \sum_{j=1}^{N}\sum_{i=1}^{M} p(x_i,y_j)\log_2 p(y_j) - H(X,Y),\\
&= -\sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log_2 p(x_i)p(y_j) + \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log_2 p(x_i,y_j),\\
&= \sum_{i=1}^{M}\sum_{j=1}^{N} p(x_i,y_j)\log_2\left(\frac{p(x_i,y_j)}{p(x_i)p(y_j)}\right), \qquad (4.3)\\
&= D_{KL}(P(X,Y)||P(X)\otimes P(Y)). \qquad (4.4)
\end{aligned}
$$

Later in this chapter, the measures of Shannon entropy, mutual information and joint entropy are used to analyse the information content contained in measurements of photon diffusion. First, however, a study that uses these concepts to analyse the information in time-resolved measurements of multipath echoes is presented. Although this study is not of photon diffusion, it analyses the importance of sampling the full temporal distribution of the signal in macroscopic specular scattering settings. In the context of this thesis, it serves as an example of using information theory analysis to gain insights into multiply scattered signals.

4.2 Information in multipath echoes

Recently, Turpin et. al. [14] showed that supervised machine learning algorithms can reconstruct three-dimensional images using only the one-dimensional time-of-flight data of an optical pulse. In later studies, it was found that the high reflectivity of surrounding walls in the room when using radar and acoustic devices, the time-of-flight consisted of return echoes which experienced multiple reflections around the environment before reaching the detector (Fig. 4.1a-b). The principle of using joint entropy as a metric for quantifying the additional independent information was used in a separate study to analyse the importance of multipath radar information.

Comparing the information content of echoes which experience up to n bounces compared with up to $n-1$ bounces showed that this additional diversity of the data leads to improved image reconstructions (Fig. (4.1d)) [141]. This analysis helps to explain why the neural network proposed in [14] performs better when provided data containing later arrival times since the additional multipath information helps break the degeneracy of possible solutions.

The following study presented in this subsection was my contribution to the published work presented in [141]. Given an ensemble of simulated multipath echo data generated by the co-authors of this study, an information theory analysis was developed to analyse the additional information in consecutive echo bounces. The results correlate closely with a sensitivity analysis Fig. (4.3d) of the proposed machine learning image reconstruction algorithm performed by the co-authors using the same data set.

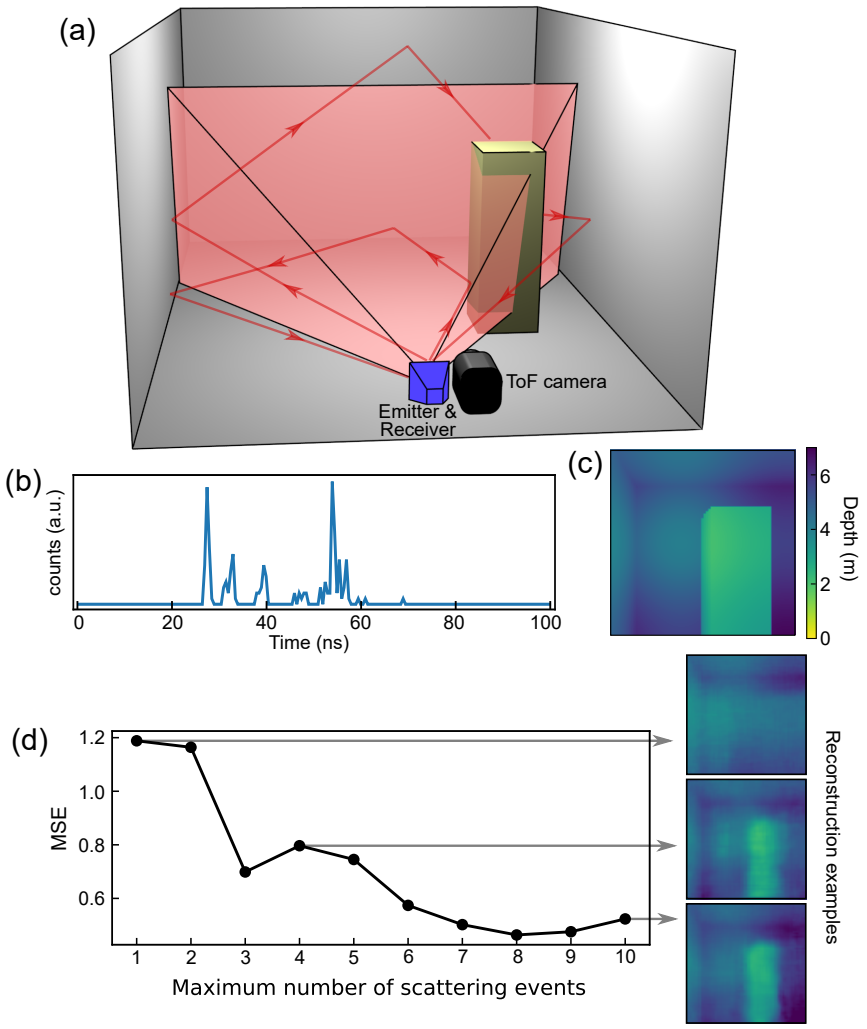


Figure 4.1: Figures produced by the co-authors of [141]. a) A diagram of the simulated experiment. b) An example time-of-flight distribution of a return echo. c) A depth image of the scene. d) The mean square error (MSE) between the reconstructed and true depth maps when the machine learning algorithm is given echo data with increasing multipath contributions. Example reconstructions are shown for input ToF traces containing up to one, four, and ten scattering events.

4.2.1 Method

The simulations produced by the co-authors of [141] used stochastic random sampling of a ray-tracing model for 2000 different environments. The environments were created by randomly positioning a cuboid pillar in a perfectly reflective square room (Fig. (4.1a)). A total of 10,000 samples were used to simulate time-of-flight (ToF) echoes (Fig. (4.1b)), which have scattered in the environment up to one time, two times, three times, etc.

The arrival time of an echo with one reflection could appear within a range of time values, so directly comparing the mutual information between the counts at each time-bin is non-trivial. Instead, the ToF distributions were simplified by binarising the counts such that any time-bin

containing a count to one and zero otherwise. Matching ToF distributions containing the echo of only a single bounce from the environment were then grouped into classes with unique integer identifiers.

A probability mass function (PMF) can then be determined by calculating a normalised histogram of the number of ToF distributions in each class. The PMF represents the probability of observing a binarised ToF distribution produced from echoes which have experienced only a single bounce before returning to the detector. This process is repeated for ToF distributions containing n bounces.

The analysis aims to determine the additional independent information in binarised ToF distributions containing n bounces (Y) compared with $n-1$ bounces (X). Using Shannon entropy as a measure of information, the conditional entropy $H(Y|X)$ of the PMF $P(Y|X)$ is the metric which addresses this aim. The interpretation of this conditional entropy is to consider the diversity of the ToF containing n bounces, which are otherwise identical for $n-1$ bounces.

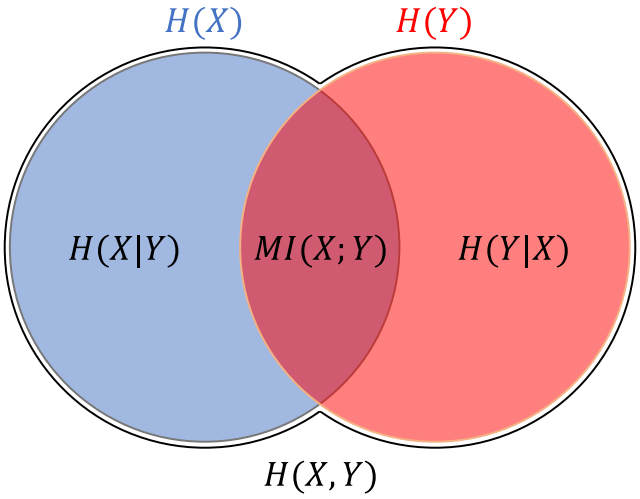


Figure 4.2: Venn diagram of the relationship between the marginal entropy of X (blue) and Y (red) with conditional entropies $H(X|Y)$ and $H(Y|X)$ and mutual information $MI(X;Y)$. The joint entropy $H(X,Y)$ is the combined area outlined in black of the intersecting circles.

Intuitively, data Y also contains information of X; therefore, to find the additional information exclusive to Y, the mutual information can be subtracted from the total information in Y: $H(Y) - MI(X;Y)$. Rearranging Eq.(4.2) gives the equivalent expression that the joint entropy (i.e., the independent information) subtracted by the entropy of X, $H(X,Y) - H(X)$ also leaves the additional independent information contained in Y. The Venn diagram shown in Fig. (4.2) is a helpful illustration of these relationships, where the objective, in this case, is to identify the independent information in Y given knowledge of X, $H(Y|X)$. The interpretation of the problem from the perspective of conditional entropy is to consider the diversity of the ToF containing n bounces, which are otherwise identical for $n-1$ bounces.

It is useful to note that the mutual information can go to zero if X and Y are completely independent, but the joint entropy has a lower bound of the marginal entropy with the largest value. This indicates that the second variable with less entropy shares all its information with the first variable. When the marginal entropy of a variable is equal to the mutual information, the additional independent information provided by the other variable is simply the difference in marginal entropies. This should be the regime for this study where the echo time traces for n bounces should contain all of the information of $n-1$ bounces. However, since the data was independently stochastically sampled for each scene to have up to n bounces, in some cases, the $n-1$ data could provide a minor improvement in independent information. The more robust method of calculating $H(Y|X) = H(X,Y) - H(X)$ was used to treat any additional information due to stochastic sampling as independent.

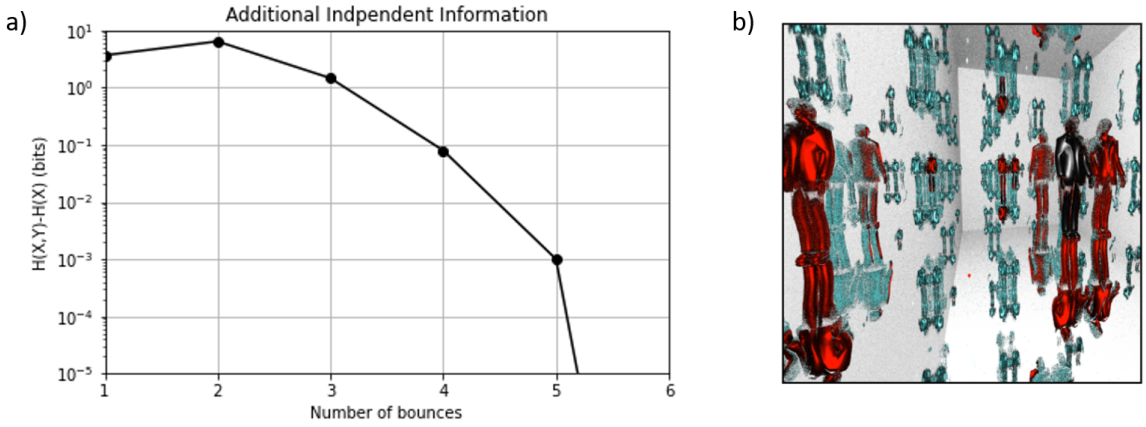


Figure 4.3: a) The additional independent information $H(Y|X) = H(X,Y) - H(X)$ when capturing multipath echo data which has experienced increasing number of bounces around the scene. Surprisingly, the gain in information when measuring echoes which have experienced up to two bounces is more than the direct reflection (1 bounce). b) A rendered scene produced by the co-authors of [141] showing a mannequin in a room showing the "hall of mirrors" effect where: black = one bounce, red = up to two bounces, light blue = more than two bounces. The additional bounces of an echo contain information about different perspectives of the scene.

4.2.2 Results

The results shown in Fig. (4.3a) indicate that, surprisingly, there is a larger gain in independent information when measuring echo ToF distributions that have experienced two bounces compared with the direct line-of-sight one-bounce reflections. This can be visualised as a "hall of mirrors" effect in Fig. (4.3b) where it is clear that echoes that have experienced more than one bounce contain information about different perspectives of the scene that would otherwise be inaccessible from the direct line-of-sight.

Echoes containing more than two bounces give decreasing independent information about the

scene. This suggests that the inclusion of later bounces adds increasingly redundant information, which can be explained by echoes that have experienced fewer bounces. Multipath echoes which experience more than five bounces do not significantly add any new independent information.

However, the ToF distributions were binarised to simplify the information calculations. The diversity in the number of counts at each time-bin would also add more independent information in practice. The machine learning reconstruction algorithm almost certainly uses the amplitude of the ToF distribution to break the degeneracy between similar ToF distributions further. Nevertheless, the simplified problem gives significant insight into the information content in multi-bounce time-of-flight data. Incorporating amplitude into the calculation would require many more samples to approximate the relevant population distributions.

Although this study is not an analysis of diffusion, it is a study of specular scattering in a macroscopic regime. In the context of this thesis, this analysis serves as an example of how information theory concepts can provide insight into how late-arriving scattered signals contribute to improving the quality of image retrieval algorithms. The following section applies the same information theory concepts to simulations from the diffuse light propagation model outlined in Section (3.2) to gain insight into resolving measurements in different domains and determine the practical limits of measuring information through highly scattering materials.

4.3 Information contained in different measurement domains

As concluded in Section (3.4), including the entire spatiotemporal measurement of diffuse photons improves the image reconstruction quality. However, it remains unclear how important it is to resolve photon counts in space-time compared with measurements exclusively in space or time domains. The analysis in this section quantitatively examines the measured information when resolving re-emitted diffuse light in different domains, which uncovers insights about how to maximise information collection about hidden absorbing objects inside highly diffusive materials.

4.3.1 Method

The forward model described in Section (3.2) is used to generate 1024 simulations using processed images from the EMNIST data set of handwritten letters [134]. The motivation for using this set of images as absorbing objects reflects the experimental work of Lyons et al., [12], in which absorbing black electrical tape was used to create occluded targets between two pieces of polystyrene foam. This image data set consists of 80,000 28×28 greyscale images of handwritten letters, binarised using a threshold such that the background is ones and the interior of the letter is zeros. This image set provided an efficient way of producing a sufficient distribution of relevant non-trivial targets and simulated photon counting measurements from which entropy values can

be calculated. Note that a canonical basis set could have been used, such as a pixel-wise raster scan or the Hadamard basis. However, these sets have sharp edges and common feature sizes, whereas the chosen image set has a more random set of asymmetric features and curved edges which is more representative of real-world applications, e.g., identifying boundaries of a tumour.

Each simulation used a uniform "flood" illumination as a source for all 28×28 pixels. Instead of zero padding, the forward model used padding with all values set to one to prevent high-frequency edge artefacts in the Fourier transforms. The simulation was configured such that there is an equal length of scattering material either side of the hidden object plane. The optical parameters of the diffusive material were chosen to be $\mu_a = 0.09\text{cm}^{-1}$ and $\mu_s' = 16.5\text{cm}^{-1}$ to match the work in [12]. Simulations for the first 1024 images of the binarised EMNIST image set were calculated from $50\ell^*$ to $1000\ell^*$ in increments of $20\ell^*$ which takes around 12hrs computing time using just-in-time (JIT) compiled python functions and parallel CPU processing (18-core). The timing resolution of 55ps was also chosen to reflect the work in [12] and a total length of 500 time-bins ensured that the width of the distribution at $1000\ell^*$ was not clipped (after shifting to avoid the zero bins at the start of the histogram due to the time delay of the first-arriving photons).

Once simulated as spatiotemporal photon counting distributions, each distribution is normalised to sum to one, producing a set of probability mass functions (PMF) of resolving a photon in space and time for each target binary image. To generate an equivalent data set resolved exclusively in space and exclusively in time, the spatiotemporal PMFs are integrated in time and space, respectively, and re-normalised (i.e., the marginal distributions are computed). For simplicity, each pixel in the space domain, and time-bin in the time domain, will be collectively referred to as “pixels” of the relevant PMF in the remainder of this chapter.

The entropy calculation in each domain is calculated by:

$$H(C) = -\sum_{i}^{K} p(c_i|x,y,t)\log_2 p(c_i|x,y,t), \tag{4.5}$$

where $c_i \in C$ is the number of photons with count value c_i for every pixel coordinate (x,y,t), and K is the number of bins in the probability mass function $P(C|x,y,t)$. In other words, a histogram is calculated for every pixel coordinate representing the distribution of photon counts for the ensemble of 1024 EMNIST hidden targets.

Representing distributions with histograms requires a careful consideration of bin width. After normalisation, each PMF is multiplied by 10^8 and discretised by converting from a floating-point array to an integer array which causes numbers to round down to the nearest whole number. This is to avoid rounding upwards, which produces more diversity in the data and increased entropy due to the discretisation process. The integer array can be interpreted as the number of photon counts in each pixel which can conveniently be histogrammed with a bin width of one. For a

fair comparison, the PMF in each domain for every number of transport mean free path lengths are consistently multiplied by 10^8, which is chosen heuristically by examining the histograms for random pixels to ensure there is no under-sampling nor oversampling due to the bin width. When considering real-world experimental parameters, 10^8 photons would correspond to a reasonable detected power of 20pW at $\lambda = 800$nm.

To determine the total information in each domain, the entropy $H(C)$ for every pixel is summed and the process is repeated for simulated data through increasing number of transport mean free path lengths.

4.3.2 Results

The results shown in Fig. (4.4) and published in [142] show the Shannon entropy trends on a logarithmic scale for increasing transport mean free path lengths (TMFPs) for each domain. The first conclusion of the results in Fig. (4.4) is that measurements resolved in space-time have orders of magnitude more information than resolving in each domain exclusively. The sharp increase in space-time information at low TMFPs is due to the broadening of the distribution in the time domain which is sampled by an increasing number of time-bins. However, since the total number of pixels remains fixed for increasing TMFPs, each pixel in the signal has a decreasing average number of photons which results in less diversity, and hence a decrease in information content evident in the downward trend after $200\ell^*$. Note that choosing to increase the number of photons $> 10^8$ would shift the peak of this trend to longer TMFPs and vise-versa when using less photons.

In practice, experimental parameters such as collection efficiency, exposure time and acquisition time of the camera determines the collected number of photons and will increase the information of the measured signal. Although space-time has the highest information content, the distribution is spread over more pixels compared with resolving in the alternative domains. The trade-off of gathering more information with an additional dimension is that there is a lower signal-to-noise ratio (SNR). This could be problematic in practice when the attenuation of photons is very high or the measurement is required to have a short exposure time.

This photon budget trade-off is also a limitation when using detectors with higher timing resolution. It was shown in [12] that increasing the timing resolution of the detector results in better spatial resolution of the reconstruction. Unfortunately decreasing the bin width in the time domain will result in an under-sampled histogram in regimes with a low number of collected photons. Regimes with low photon budget it may benefit by sacrificing resolution in the time domain to increase SNR and produce a sample distributions which closer resemble the true population distribution.

The same argument can be made for an increased number of pixels in the space domain (given the same field of view), since the same number of photons will be spread across a higher number of

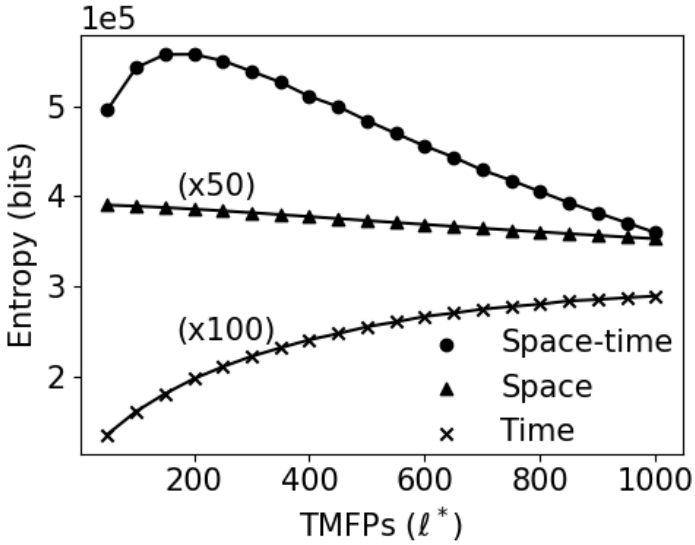


Figure 4.4: Figure published in [142] showing summed Shannon entropy of every pixel in the space-time, space-only (multiplied by 50) and time-only (multiplied by 100) domains for increasing number of transport mean free path lengths (TMFPS). A fixed acquisition of 10^8 photon counts was used for all simulated measurements.

pixels. Interestingly, it is expected that the number of pixels, and field of view of the array should be chosen to maximise the diversity in the spatial distribution. Since the spatial distribution becomes increasingly smooth (as can be seen in Fig. (3.1)), the field of view should be increased such that each pixel samples more independent information compared with the neighbouring pixels. In other words, oversampling the spatial distribution of photons in the highly diffuse regime may not add significant independent information and cause the signal-to-noise ratio to decrease.

As expected, Fig. (4.4) shows that the information in space decreases as the spatial features become increasingly blurred such that the number of photon counts at each pixel becomes less diverse between hidden object. Interestingly, the information in time increases with TMFPs since the distribution of photon counts is spread across more time-bins. These arguments can be visualised from the examples shown in Fig. (3.1) which demonstrates that the integrated space-resolved photon distributions become increasingly like a two-dimensional Gaussian distribution, whereas the distribution in time is sampled across an increasing number of time-bins.

The information in time also begins to plateau and is expected to eventually decrease since the distribution of photon counts in time will become less diverse (tending towards identical Gaussian distributions) at extreme number of transport mean free path lengths. However, even at $1000\ell^*$ the information in the time domain continues to increase.

Unfortunately, it is unrealistic to assume a collection of 10^8 photons in the extreme regimes around $1000\ell^*$ since light is exponentially attenuated with increased thickness, scattering or

absorption as per Eq. (3.2). This is the topic of the following section which considers practical experimental parameters and the mutual information of the signals with the pixels of the hidden object.

4.4 Limits of information transport in diffuse optical imaging

The previous section calculated the absolute information content of signals with a fixed number of photon counts and ideal noise-free signals. However, to determine how much shared information there is between the measurement and the pixel values on the hidden target, a mutual information analysis is required. By also introducing typical experimental parameters and considering noise contributions in the signal, a prediction of practical limits of transmitting information about a hidden absorbing target through a scattering material can be deduced.

4.4.1 Method

The mutual information between the pixels of a two-dimensional hidden absorption target and an n-dimensional measurement is easier to calculate by first transforming the data into some shared domain. Since the forward model in Eq. (3.3) is a linear operator acting on the image of the hidden object, a basis set of impulse response functions (IRFs) can be constructed in the measurement domain which maps to every pixel in the object image to a corresponding distribution of photon counts as shown in Fig. (4.5a). To do this, a set of measurements are simulated for a pixel-wise raster scan of the value one on a background of zeros. Careful consideration is needed for the zero-padding boundaries in the forward model. Since the source is a uniform flood illumination, non-zero contributions of transmitted light in the padding can cause edge artefacts in the simulated measurement which can be strong contributions that wash out the intended impulse response from a single pixel in the hidden object plane. For this reason, zero-padding was used for these IRF simulations rather than padding with ones. Each of the $28 \times 28 = 784$ raster scan non-absorbing pixel images were simulated from $10\ell^*$ to $250\ell^*$ in increments of $10\ell^*$.

The same 1024 binarised EMNIST images used in the previous study (Section (4.3)) are used in this study to create distribution of hidden absorbing objects. However, the simulations were amended to have identical parameters as the IRF basis simulations.

When the image set of a pixel-wise raster scan is reshaped such that each image is a column of the matrix, it is simply the identity, and so projecting the target images of the hidden objects onto these basis images does nothing since they already exist in this basis. However, the simulated measurements shown in Fig. (4.5b) need to be mapped into the 28×28 image domain, which is achieved by projecting them onto the matrix of IRFs.

More concretely, every impulse response is normalised to sum to one and arranged as a column

in the matrix $Q \in \mathbb{R}^{m \times n}$ (Fig. (4.5a)), where m is the number of elements in the measurement domain (i.e., number of pixels and timebins) and n is the number of elements in the hidden object domain (i.e., $28 \times 28 = 784$). Each of the 1024 simulated measurements of EMNIST images is then also normalised such that their sum is equal to one, and they are arranged as one-dimensional vectors in a matrix $R \in \mathbb{R}^{m \times N}$ (Fig. (4.5b)), where N is the number of hidden objects (1024). The inner product of these matrices $Y = \langle R, Q \rangle$ is a set of projection coefficients which determine the contribution of each basis function that make up the measurements. The projection coefficients can then be binarised to a high/low value using the mean value as a threshold (Fig. (4.5c)).

After this projection mapping, the problem can be considered a binary symmetric communication channel whereby every high/low value transmitted by pixels in the hidden object propagates through a noisy channel of photon diffusion where cross-talk occurs between neighbouring pixels. The measurement of photon count distributions resolved in space and time at the detector is then projected onto a basis, and the mean projection coefficient determines which detection channels are high/low. This is repeated for each of the 1024 EMNIST objects such that each channel has a Bernoulli distribution $P(X)$ of high/low input (i.e., the distribution of high/low values for a single pixel over the set of 1024 EMNIST images, as shown by the blue distribution in Fig. (4.5b)). This input distribution can be directly compared with the corresponding Bernoulli distribution $P(Y)$ or binarised projection coefficients after the information is propagated through the noisy channel, as shown by the red distribution in Fig. (4.5c). In this case, the extent of cross-talk of the noisy channel is determined by the number of transport mean free paths used in the simulation.

The mutual information between every input and output channel can then be calculated by constructing the joint probability distributions $P(X,Y) = P(X|Y)P(Y)$ and using the KL-divergence as derived in (4.3). This calculation is repeated for every pixel and averaged to find the mutual information between hidden images and output measurements for simulations corresponding to $10\ell^*$ to $220\ell^*$ in increments of $10\ell^*$.

4.4.2 Results

The solid blue curve in Fig. (4.6a) shows the mutual information for increasing TMFPs for space-time resolved simulations. When repeating the process for space-only measurements (i.e., integrating all simulations in time), the trends show that mutual information is lower when resolving in space-only (solid red line) compared with space-time. However, the advantage of measuring in space-time diminishes towards $100\ell^*$. The mutual information resolved in time-only was negligible in comparison for all transport mean free path lengths for the simulated configuration, and therefore only the space and space-time domains are studied in this section.

To account for the attenuation of photons with increasing TMFPs, each PMF is multiplied by

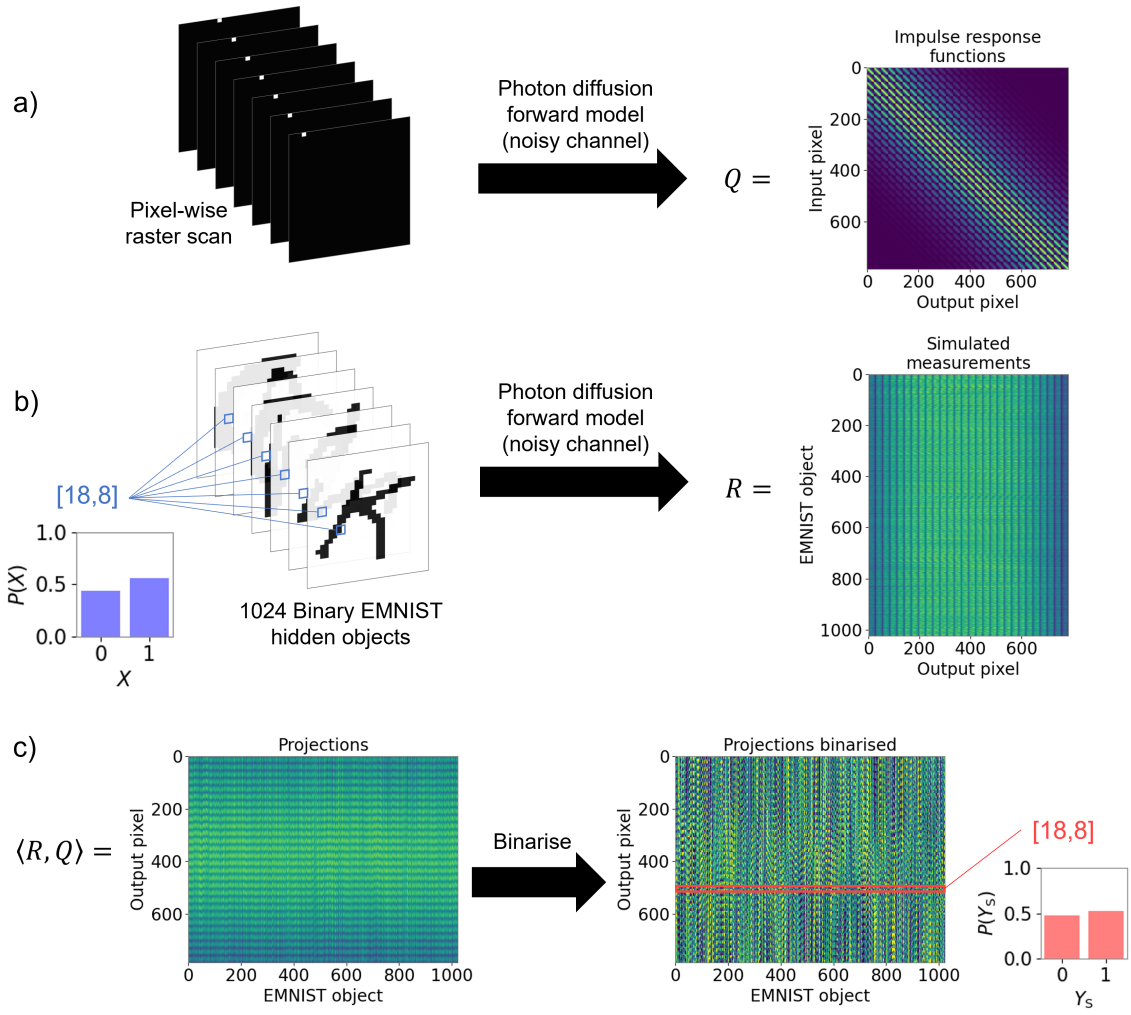


Figure 4.5: The pipeline for calculating the mutual information between hidden objects and diffuse photon measurements resolved in the space-only domain. a) First a matrix of basis functions Q is constructed using the point spread function associated with every pixel in the input. b) A matrix of measurements R is constructed by simulating 1024 binary EMNIST hidden objects. c) The inner product of the measurements and the basis functions $\langle R, Q \rangle$ produces a matrix of projection coefficients, which are binarised using the mean value in each column as a threshold. The mutual information between a pixel at the input X (e.g., [18,8]) (blue) and the associated projections at the output Y_S (red) can be calculated using the KL-divergence between the product of the marginal distributions $P(X)$ and $P(Y_S)$ and the joint probability $P(X, Y_S)$ distribution (not shown).

the expected number of photons N_{ph} given by Eq.(3.2), where the solid angle of collection is $\frac{\delta\Omega}{4\pi} = 10^{-5}$ for consistency with previous work [12,133] and the input intensity is a modest 4W/m^2 (10mW for a $(5\times5)\text{cm}^2$ area) which is three orders of magnitude below the British safety limit for exposure to human skin [143].

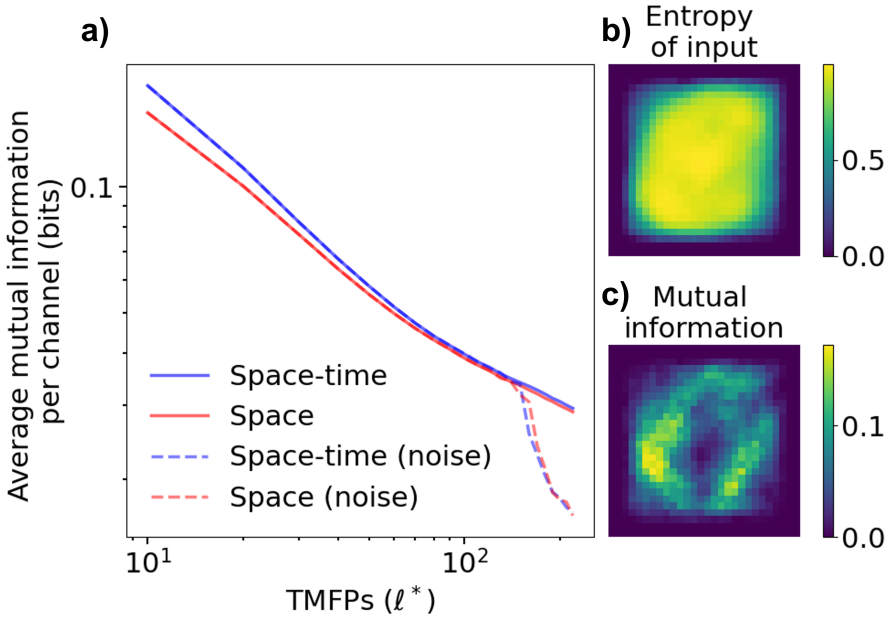


Figure 4.6: Figure published in [142]. a) The average mutual information between hidden object and measurements resolved in space-time (blue) and space (red) for increasing number of transport mean free paths (TMFPs). b) The entropy of each pixel in the set of hidden object images. c) The mutual information between each pixel in the hidden image and the output resolved in space-time for $50\ell^*$.

However, the total number of photons is underestimated when rounded to the nearest integer to quantise the data into photon counts. To overcome this, a gradient descent algorithm was implemented in python to multiply the PMF by a dummy multiplicative factor until the sum of photons after quantisation N_{q} meets the condition $N_{\text{ph}} < N_{\text{q}} < (N_{\text{ph}}+20)$. Although in practice, larger absorbing objects will cause increased extinction of photons compared with smaller objects for the same experimental conditions, it is simpler to instead consider an experiment where the acquisition time is adjusted to collect same number of photons for each measurement. This is a conservative assumption which reduces the diversity in the measurements compared with a fixed acquisition rate but ensures that each measurement has equivalent signal-to-noise ratio.

Quantisation noise due to the discrete nature of photon counting will reduce the diversity of the photon counts between measurements. However, this only has a significant effect when a low number of photons is collected. The larger contribution of noise is due to Shot noise since the variance of any counting statistics will increase noise as the square root of the number of counts. To include Shot noise contributions, each quantised measurement is replaced with a value sampled from a random number generator which samples a Poisson distribution $P(X =$

$k) = \frac{\lambda^k e^{-\lambda}}{k!}$, which determines the probability of observing k counts given the expected value $\lambda = \mathbb{E}[X]$. In this context, the number of counts in each data pixel is the expected value λ of the Poisson distribution. Unlike the case of quantisation noise, the random sampling of Shot noise causes an increase in diversity of the measurements and the absolute information in terms of Shannon entropy will increase. However, the variations due to noise will also be independent of the distribution of pixel values of the hidden object. Therefore, this information will not lead to an increase in mutual information.

Figure (4.6a) shows that including these practical considerations, the mutual information resolved in both space-time (blue dashed) and space (red dashed) deviates from the ideal curve only for regimes $> 100\ell^*$. Furthermore, there exists no mutual information in the data beyond $200\ell^*$ which corresponds to ~ 12cm of tissue ($\mu_a = 0.09\text{cm}^{-1}$, $\mu_s' = 16.5\text{cm}^{-1}$).

Another key insight is that the number of photons calculated with Eq.(3.2) also goes to zero around the same number of transport mean free path lengths at the limit of mutual information shown in Fig. (4.6a) (the quantised data has an average of only 16 photons at $200\ell^*$). Furthermore, the theoretical trends (solid lines) show that mutual information continues beyond $200\ell^*$, this suggests the surprising conclusion that so long as there is a detected photon, there exists mutual information at the detector about a hidden object. The impact this has for real-world experiments is that the attenuation of photons is the limitation to imaging in the highly diffusive regime rather than the loss of information due to cross-talk noise in the process of random photon scattering.

Figure (4.6b) shows the entropy contained in each pixel of the image set used as the absorbing objects. An illustration of this for a single pixel is shown by the blue histogram in Fig. (4.5b). The entropy is relatively uniform in the centre of the image and can be interpreted as where the information is distributed in space for the given input image distribution. The information map in Fig. (4.6c) shows the mutual information between the pixels in the hidden objects and the binarised projection coefficients for space-time resolved measurements through $50\ell^*$. The distribution of mutual information indicates which pixels preserve the most information when transmitted through $50\ell^*$ of diffusion. Interestingly, the dark region in the centre corresponds to losing almost all of the information for these pixel values. This is most likely due to the cross-talk of neighbouring pixels around the centre region. However, it is clear from the entropy of the hidden object pixels in Fig. (4.5b) that there is no information transmitted by border regions. This reduces the blurring cross-talk for the pixels at the edges of the hidden objects, and the information from these pixels is better preserved.

This analysis is continued in Fig. (4.7d-e) which compares the mutual information maps for measurements resolved in space-time and space-only for increasing number of transport mean free path lengths. For weakly scattering materials equivalent to $10\ell^*$, the information about the

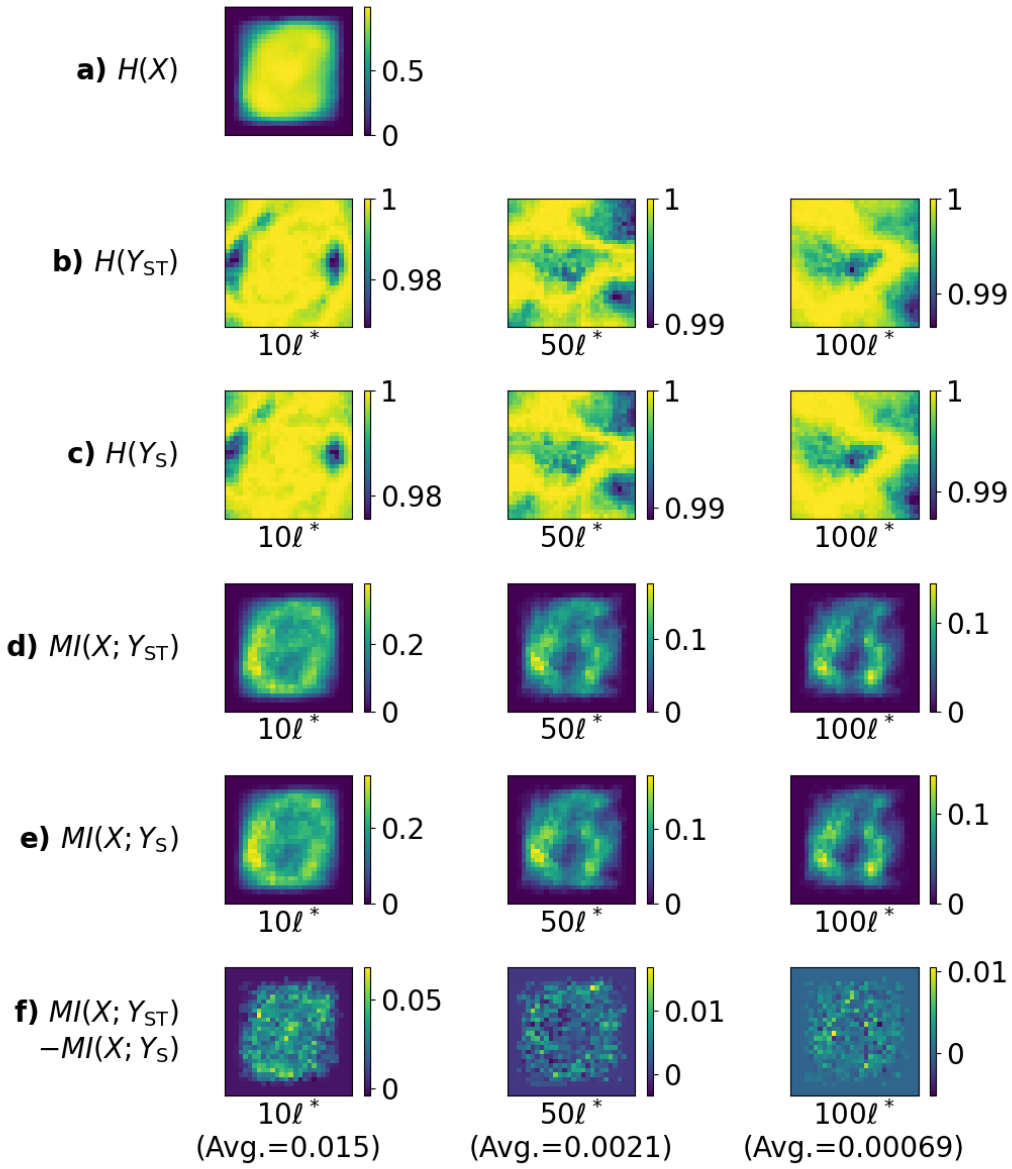


Figure 4.7: a) The entropy of each pixel in the set of 1024 hidden objects. b) The entropy of the binarised projection coefficients when resolving in the space-time domain for increasing transport mean free path (TMFP) lengths. c) The same as (b), but using measurements resolved only in space. d) A map of the mutual information between each pixel at the input and the corresponding binarised projection coefficients when using measurements resolved in the space-time domain for increasing TMFP lengths. e) The same as (d) but using measurements resolved only in space. f) The difference between the mutual information maps presented in (d-e).

centre pixels of the hidden objects is relatively well preserved.

However, as the number of transport mean free paths tends towards $100\ell^*$ regions of high information "hot spots" emerge. This mutual information analysis could be used to guide the design of experiments to sample information about different regions of hidden absorbers using patterned illumination or patterned detection, e.g., similar to experimental designs of Chen et al. [90], Konecky et al. [121] and Manohar et al. [144]. There is not a significant difference in the distribution of mutual information when measuring in space-time compared with space-only as shown in Fig. (4.7f). The entropy of each basis at the output is also consistently uniform for both space-time and space as shown in Fig. (4.7b-c) which indicates that absolute measure of information at the detector is spread evenly over all channels.

The currently accepted theoretical limit of diffuse optical imaging was presented in a study by Mora et al. [72], that concluded a maximum penetration depth of a time domain diffuse optical tomography device used in a null source-detector configuration (discussed in Section (2.5)), is expected to be around 6cm in human tissue, corresponding to roughly $121\ell^*$ round trip ($\mu_a = 0.1\text{cm}^{-1}$, $\mu_s' = 10\text{cm}^{-1}$). The analysis performed in [72], determined the maximum depth of sensitivity to detect a 1cm^3 perfectly absorbing object embedded in a $(6\times6\times6)$ cm^3 homogeneous scattering volume using a Monte Carlo photon diffusion simulation. Experimental results of this study also show a trivial reconstruction of a black plastic cylinder inside a phantom scattering volume with a return trip corresponding to around $80\ell^*$.

The limits of diffuse optical imaging in [72] were determined by comparing the difference in photon counts in a time-gated window which isolated late-arriving photons between the simulations with and without an absorbing object. When the relative contrast is less than 1% then it is assumed there is no information. However, in this reflection configuration, photons which take a long optical path in the medium due to multiple scattering in shallow regions of the material will also be detected in the same time-gate and contribute to a decrease in contrast. Whereas, in the transmission geometry studied in this thesis, all of the detected photons must have travelled through the entire medium and are therefore more likely to interact with the absorbing object. This may be the reason that the limit of information concluded in this thesis (and published in [142]) is that, so long as there is a photons detected then there is information about internal absorbing objects embedded inside the material.

The most current experimental limit of detecting photons in transillumination configuration was demonstrated by Lyons et al. [12], which presented evidence of reconstructing hidden absorbing objects with a non-trivial geometry through more than $80\ell^*$. In this study, the measurements were used to reconstruct images rather than the location of a point-like inclusion.

The information study presented in this thesis and published in [142] show that imaging information can be detected at more than double the number of transport mean free path lengths

demonstrated by the currently accepted extreme depth sensitivity in [72] and [12].

However, to validate the claim of transmitting photons through $> 200\ell^*$, it must be shown experimentally that it is possible to detect photons through such scattering materials. This is the aim of the experiment performed in the next section.

4.5 Evidence of photon transmission beyond $200\ell^*$

To validate the claims made in the previous section that information can be transmitted through $200\ell^*$ using realistic parameters, it is important to demonstrate experimental proof that it is possible to detect photons in this regime.

4.5.1 Method

An experiment was performed by transilluminating polystyrene foam with an ultrafast pulsed laser (140fs, 80MHz, 800nm). The laser was coupled to a 5mm diameter liquid light guide and the output was expanded to 1-inch diameter using a collimator. The diffuse light re-emitted from the foam was collected by a large aperture lens (Meike MK-35mm, f/1.4) and a single-photon sensitive photomultiplier tube (PMT) (Hamamatsu H7422P-50) was used to detect photons in Geiger mode. The solid angle of collection was calculated to be $\delta\Omega = 0.277$.

The experiment was performed using TCSPC (Becker and Hickl SPC-150N TCSPC module) such that the correlation of time-of-arrival of the photons with the laser repetition rate made it convenient to isolate signal photons from noise, and the time traces could be used to fit the time distribution to the forward model described in Section (3.2) and extract the optical coefficients.

The instrument response function (IRF) was recorded by replacing the polystyrene foam with a piece of paper to act as a diffuser which spreads the light in all directions without significantly broadening the pulse in time. This is a standard technique to ensure that the angular distribution of collected light is as close as possible to the distribution of collection for the measurements [145, 146].

Measurements were taken for increasing thicknesses of foam L by stacking $d = 2.58\text{cm} \pm 0.01\text{cm}$ slabs between the source and detector from $L = 5.16\text{cm}$ to $L = 15.48\text{cm}$. The power of the source was adjusted for every measurement to avoid saturating the detector.

The measured IRF was convolved with the forward model described in Section (3.2) and a least-squares optimisation was performed to extract the absorption μ_a and reduced scattering coefficient μ_s' for each thickness of foam using the function `optimize.curve_fit()` from the SciPy Python module. Using these fitted parameters the number of transport mean free paths can be calculated by $\frac{L}{\ell^*} = (\mu_a + \mu_s')L$.

The number of signal photon counts was determined as the total counts minus a background measurement obtained by turning off the laser and acquiring data for the same conditions. For each thickness of polystyrene, a measurement and a repeat measurement were taken with the laser on and off. The entire experiment was then repeated twice more to average the effects of background light changes in the lab (e.g., door opening and closing).

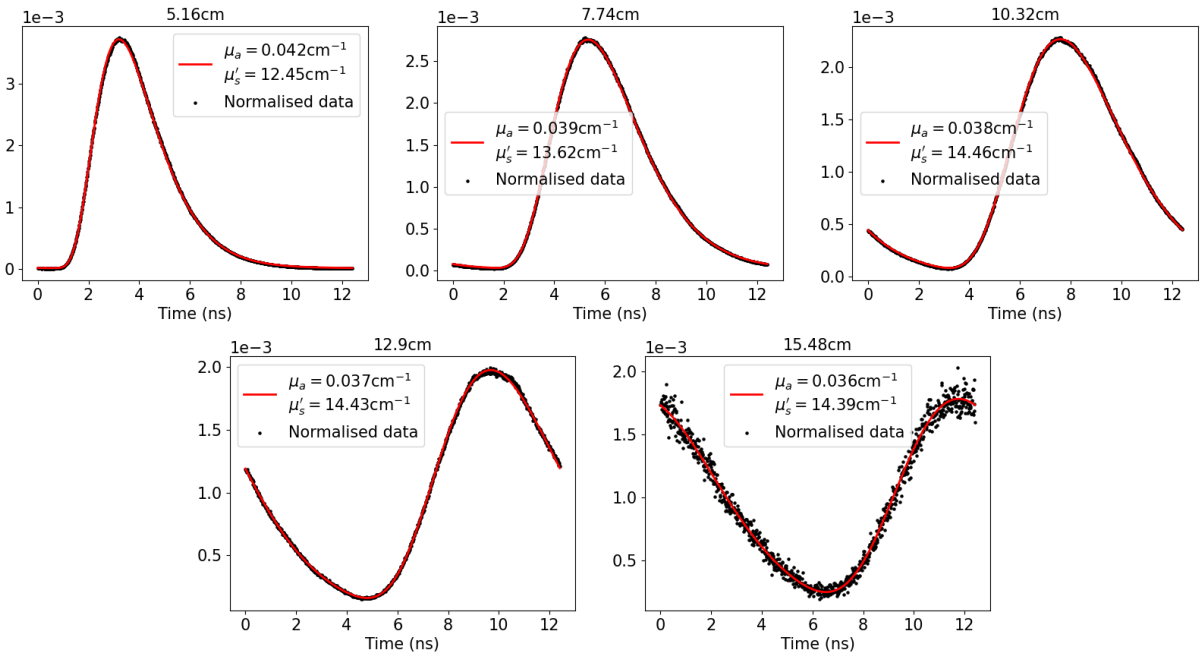


Figure 4.8: Normalised experimental data averaged over two trials (black dots) and forward model fits (red) for increasing thickness of polystyrene foam. The same 12.5ns time window was cropped for each set of data. The tail of the distribution from the previous pulse begins to wrap around and add to the rising edge as the thickness increases. This is due to the fixed period of the laser repetition rate which is shorter than the total temporal broadening of the pulse.

The mean of the background-subtracted signals for each thickness of foam was used to extract the optical coefficients μ_a and μ_s'. However, first the signal is cropped to select only 1024 time bins starting from the time bin corresponding to the peak of the IRF. This will accurately determine the relative delay caused only by the diffusion of light and removes delay introduced by the experimental set up (e.g., difference in cable lengths used for the reference and signal electronics or optical path length of the laser in the absence of the sample). Since the absolute delay is unknown before the experiment, the TCSPC module was set to use only every fourth reference pulse from the laser.

Furthermore, the laser repetition rate is fixed at 80MHz which corresponds to 12.5ns. As the time distribution of the photons broadens longer than the laser repetition period, photons which take longer than 12.5ns to arrive at the detector will be counted as early photons in the timing window

of the following pulse. Wrapping causes consecutive pulses to become increasingly overlapped for greater thicknesses, which can be seen in the experimental data in Fig. (4.8). This wrapping is unavoidable since the repetition rate of the laser is fixed, however this was accounted for when fitting for the optical coefficients by adding repeat pulses before and after the timing window in the forward model.

To determine the extinction ratio, the detected fluence rate Φ is normalised by the initial fluence rate Φ_0. This ratio was found by first converting the measured incident power on the foam to the number of incident photons by divided by the energy of a photon at 800nm. The incident number of photons was then divided by the total number of background-subtracted photons at the detector.

A line of best fit can be used when plotting the number of signal photons in log-scale against transport mean free path lengths to determine the rate of extinction which can be extrapolated to estimate the extreme limits of photon detection.

4.5.2 Results

The results shown in Fig. (4.9) show that photons were detected well beyond $200\ell^*$. The black dashed lines indicate the extinction corresponding to a detection of one photon per second for a 10mW source and a 1W source. The intersection of these dashed lines with the extrapolated line of best fit indicate that this increase can extend the detection limit by around $50\ell^*$. Furthermore, extrapolation shows that using a 1W source, an experiment with the same quantum efficiency and collection can readily detect photons transmitted beyond $300\ell^*$, and improvements in quantum efficiency of detectors at 800nm could extend this limit even further.

It is important to validate the practical assumptions used in the numerical experiments of the previous sections using these experimental results. To do this, the experimental best fit to the data shown by the black line in Fig. (4.9) was normalised by first dividing the experimental solid angle of collection factor $\frac{\delta\Omega}{4\pi} = 0.022$ and quantum efficiency (15%), and then multiplying by the solid angle of collection factor used in the previous numerical studies $\frac{\delta\Omega}{4\pi} = 10^{-5}$. This new extinction ratio for the simulated parameters is shown by the red line in Fig. (4.9), where the uncertainty in the best fit is shown in pale red.

For a clear comparison with the simulation parameters, Eq. (3.2) was used to determine the number of transport mean free paths corresponding to a transmission rate of one photon per second for 10mW source power, $\frac{\delta\Omega}{4\pi} = 10^{-5}$, $\mu_a = 0.09\text{cm}^{-1}$, and $\mu_s' = 16.5\text{cm}^{-1}$ used in the previous numerical studies. The calculated value is denoted by the blue cross in Fig. (4.9), which overlaps almost exactly with the estimated extinction rate using the experimental data. This validates the assumptions made in the previous studies and is evidence that the numerical forward model described in Section (3.2) accurately emulates photon diffusion propagation for homogeneous

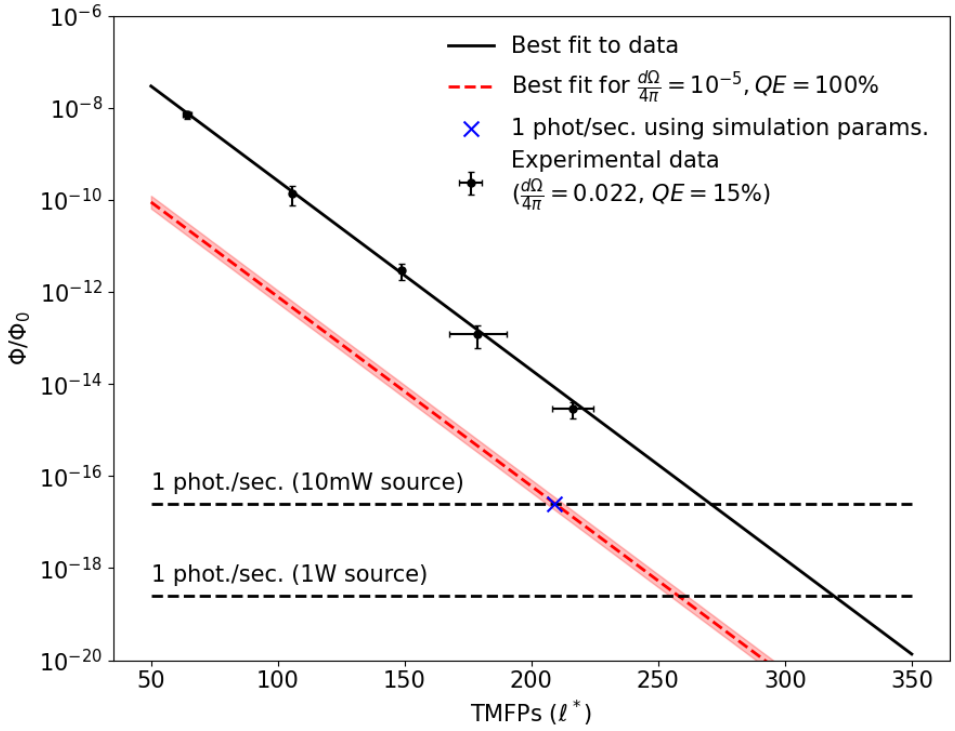


Figure 4.9: The experimental extinction Φ/Φ_0 for increasing transport mean free paths (TMFPs) and the optimised line of best fit. The experimental solid angle of collection factor is $\delta\Omega/4\pi = 0.022$ and the detector had a quantum efficiency (QE) of 15%. The black dashed lines show the extinction rate corresponding to transmitting one photon per second for an incident power 10mW and 1W. The red dashed line is the estimated rate of extinction for the simulated parameters (10mW, $\delta\Omega/4\pi = 10^{-5}$, QE=100%) used in the numerical studies in the previous sections of this chapter. The pale red area is the uncertainty of the estimated rate. The blue cross represents the number of TMFPs corresponding to a transmission of one photon per second for the simulated parameters.

materials such as the polystyrene foam used in this experiment.

Although the limits of mutual information using the simulated parameters was found to be around $200\ell^*$ in Section (4.4). The results of this experimental validation suggest that in practice, larger solid angle of collection, source power and quantum efficiency could extend this limit well beyond $250\ell^*$ which corresponds to around 15cm of biological tissue ($\mu_a = 0.09\text{cm}^{-1}$, $\mu_s' = 16.5\text{cm}^{-1}$). Although there exist information in these extreme regimes, it is still unclear how to use this information to produce images. This topic will be discussed further in the Chapter 5.

4.5.3 Estimating uncertainty

The uncertainty for each thickness L was found by taking the combined error when adding n slab thicknesses d together: $\Delta L_n = \sqrt{(\Delta d_1)^2 + (\Delta d_2)^2 + \ldots (\Delta d_n)^2}$, where $\Delta d_1 = \Delta d_2 = \Delta d_n =$

0.01cm is the measurement error found by taking the variance of digital calipers readings for each slab. The variance of the estimated optical parameters returned from the fitting function were negligible and so the error in the number of transport mean free paths (TMFPs) was found by multiplying by the fractional error of the length of material $\Delta\text{TMFPs} = \text{TMFPs}\frac{\Delta L}{L}$.

The uncertainty in the extinction ratio $\frac{\Phi}{\Phi_0}$ was found by calculating the norm of the standard deviation of measured power fluctuations over ten seconds and the standard deviation of the background-subtracted signal counts from six trials (two trials for each experiment).

The line of best fit was weighted by the norm of the relative errors of transport mean free path and extinction ratio. Since the fit was performed to the extinction ratio values in log-scale, the log error of the extinction ratio values was first calculated by:

$$\Delta \ln(\Phi/\Phi_0) = \frac{\Delta(\Phi/\Phi_0)}{(\Phi/\Phi_0)}. \tag{4.6}$$

The uncertainty shown in pale red in Fig. (4.9) for the estimated extinction rate using simulated parameters is determined by the standard deviation of the y-intercept returned by the least-squares fitting function.

4.6 Improved ill-conditioning of the inverse problem

Although quantifying the information collected at the detector gives insights into the limits of imaging in the highly diffuse regime and how best to maximise this information, it is also important to consider the impact of collecting more information when reconstructing images. In this context, the inverse problem of image retrieval is ill-posed, or more specifically, ill-conditioned, such that a small perturbation in the measurements can cause a significant change in the reconstructions. To quantify this, the condition number $\kappa = \frac{\sigma_{\max}}{\sigma_{\min}}$ introduced in Section (2.6) can be used as an indication to the stability of the inverse problem, where $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum singular values of the forward operator.

The singular value decomposition (SVD) analysis has been used in other studies of diffuse light to quantify the improvement of the conditioning of the inverse problem of image reconstruction with early photon imaging. Leblond et al. [147] showed that early time-gating of detected photons from fluorescent targets inside diffusive materials improved the condition number of the inverse problem by an order of magnitude. However, these experiments were demonstrated in phantoms $< 13\ell^*$, where early photon imaging is viable. Zhang et al. also used SVD analysis to demonstrate that early photon time-gating improves the inverse problem of locating fluorescent targets. In this analysis a fixed condition number was used as a threshold to truncate singular values. They show that time-gating for early photons preserves more singular vectors and therefore

higher frequency features can be more reliably retrieved. However, neither of these studies considered the highly diffusive regime and using an approach which utilised the full spatiotemporal distribution of the photon counts.

Aside from time-gating early photons, inverse retrieval algorithms in the highly diffusive regime typically add *a priori* information in the form of regularisers to improve the condition of the inverse problem. However, in this section, the improved conditioning of the problem is analysed by adding information solely by resolving the measurements in both space and time rather than only in space. Therefore, to quantify the effect of resolving measurements in a different domain and remain agnostic to *a priori* information, a simple linear inversion operator is constructed from which the condition number can be calculated: $Ax = y$, where the operator A is a matrix operation which maps from the images x (1024 binarised EMNIST images used in the previous information study sections) to the measurements y (space-time, space and time-resolved photon counting simulations using the forward model described in Section (3.2)).

The forward operator A for each transport mean free path length of material were constructed in the same way as the raster-scanned impulse response functions used in Section (4.4). The inverse operators are then found by calculating the pseudo-inverse:

$$A^{-1} = V\Sigma^{-1}U^T, \tag{4.7}$$

where U, V and Σ are the left and right singular vectors of the singular value decomposition of the forward operator A, respectively.

The condition number $\kappa = \frac{\sigma_{\max}}{\sigma_{\min}}$ uses the ratio maximum ($\sigma_{\max}$) and minimum ($\sigma_{\min}$) singular values of the forward operator A, which indicates the amplification of insignificant singular vectors, often associated with noise contributions when finding the reciprocal of their respective singular values in Eq.(4.7).

The singular values shown in Fig. (4.10a) are for the forward operators constructed using noise-free simulated measurements for $50\ell^*$ resolved in space-time, space and time domains, where each measurement is normalised to sum to one such that it is a probability mass function. The most obvious observation from the trends shown in graph is that the operator constructed exclusively in the time domain is severely ill-conditioned. Typically matrices are considered non-invertible when the condition number approaches $\sim 10^{16}$ [117].

The normalised singular values in Fig. (4.10b) show that the space-time operator singular values decay at a slightly lower rate than the space operator. This indicates that the relative weighting associated with the decomposed singular vectors is similar, and hence the respective contributions in the inverse mapping will produce more accurate solutions [115, 117]. Although the difference in the decay rate of singular values between space-time and space is relatively small,

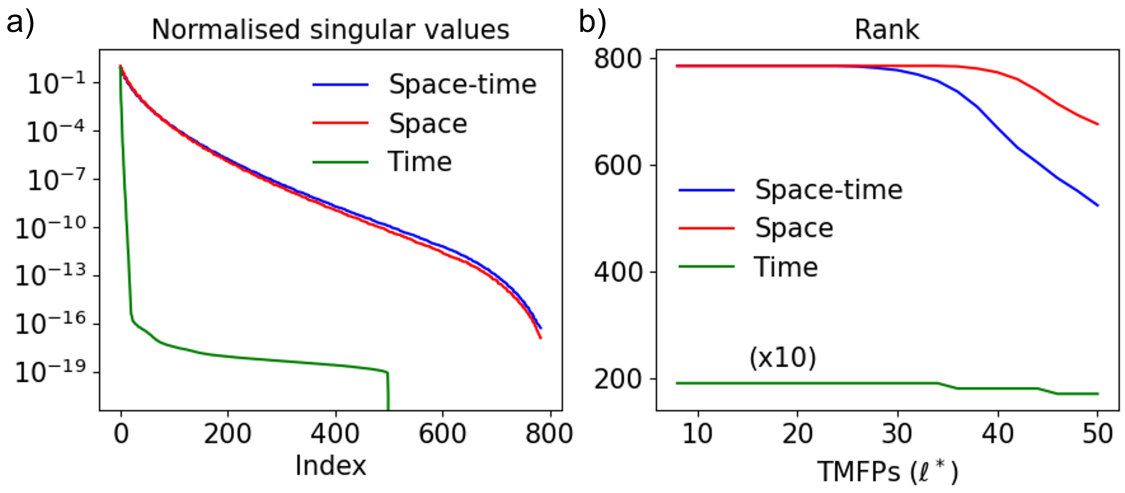


Figure 4.10: a) The normalised singular values for the linear operator A resolved in space-time (blue), space (red) and time (green) for $50\ell^*$. b) The rank of the forward operator A resolved in space-time, space and time for increasing transport mean free path lengths.

both are severely ill-conditioned and subtle differences in the magnitude of the smallest singular values will translate to large differences when inverted. This difference will result in a more robust inversion for the space-time operator when small noise contributions are present in the measurements since they will not be amplified as much as when using the space operator.

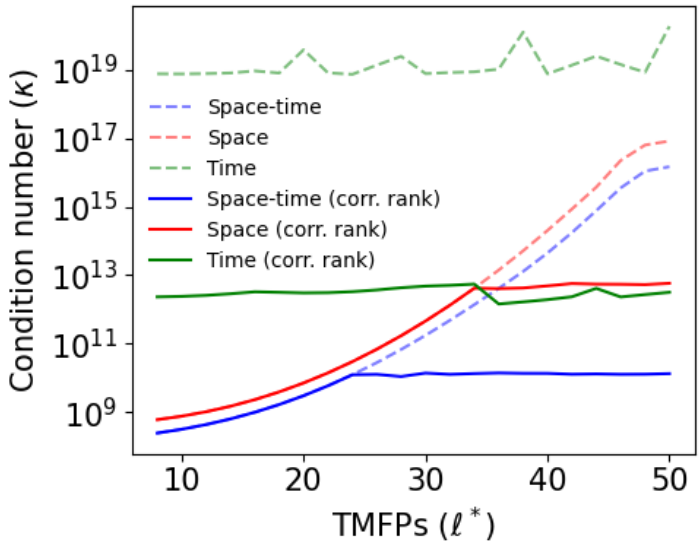


Figure 4.11: Figure published in the supplementary material of [142] showing the condition number κ of the forward operator A resolved in different domains after adjusting the number of singular values based on the rank of the operator. The dashed lines show that the space-time resolved operator is still consistently less ill-conditioned when including all of the singular values.

Although the singular values of the space-time and space operators appear to be at full rank (i.e.,

full rank corresponds to the smallest dimension of the operator, in this case, $28 \times 28 = 784$ which is the dimensionality of the solution space), the numerical precision of computing the pseudo-inverse limits the rank of the inverse operator [117]. Since the number of pixels is greater in space-time, the values of the probability mass functions will be smaller compared with the space domain after normalisation. Since the size of the space-time operator is larger and consists of values with smaller magnitude than the space operator, it has a lower rank beyond $24\ell^*$. When taking the pseudo-inverse, the minimum singular value, which is accurately inverted, does not correspond to the smallest singular value shown in Fig. (4.10a) but instead corresponds to the first value above the numerical precision boundary.

Determining the numerical precision boundary requires taking the product of the size of the matrix and the floating-point precision of the matrix norm [117]. Matlab has a built-in function `rank()`, which conveniently calculates this when given a matrix as an argument. The forward operator for each domain was calculated from $10\ell^*$ to $50\ell^*$ in increments of $1\ell^*$ and was input into the `rank()` function. The return value was then used to determine the cut-off threshold of absolute singular values. The rank of the operators is shown in Fig. (4.10b), which shows that the space-time operator suffers low rank due to numerical precision inaccuracies for increasing transport mean free path lengths. It is also clear that the time domain operator can only be reliably decomposed into < 20 singular vectors.

The condition number κ shown in Fig. (4.11) was calculated for the matrices resolved in each domain using the rank to threshold for the singular values. However, to ensure this truncation of singular values did not impact the conclusions of the condition number comparisons, the condition number of the operators before truncation is also included in Fig. (4.11) as dashed lines. The condition number of the space-time operator is consistently lower than the space operator for all cases, which indicates a more accurate inversion which is more robust to noise.

This can be directly observed in Fig. (4.12a), which shows truncated SVD solutions when adding a small noise perturbation to the data. When truncating too many singular values, only the low-frequency features are resolved. However, when increasing the number of singular vectors to $k = 150$ there is still a stronger resemblance of the target in the space-time domain compared with space due to the increased robustness to noise.

The noise perturbation was added by randomly sampling a Poisson distribution using the NumPy `np.random.poisson()` function, where the value of every element in the data is given as the mean. However, before random sampling, the data were normalised by the sum and multiplied by 10^{10} to achieve a high enough signal-to-noise to ensure a computable reconstruction. Since there are more pixels in the space domain than in the space-time domain, the maximum value of counts will be higher, and the signal-to-noise ratio (SNR) will be better when introducing Shot noise by sampling a Poisson distribution in this way. This reflects the case in real-world settings,

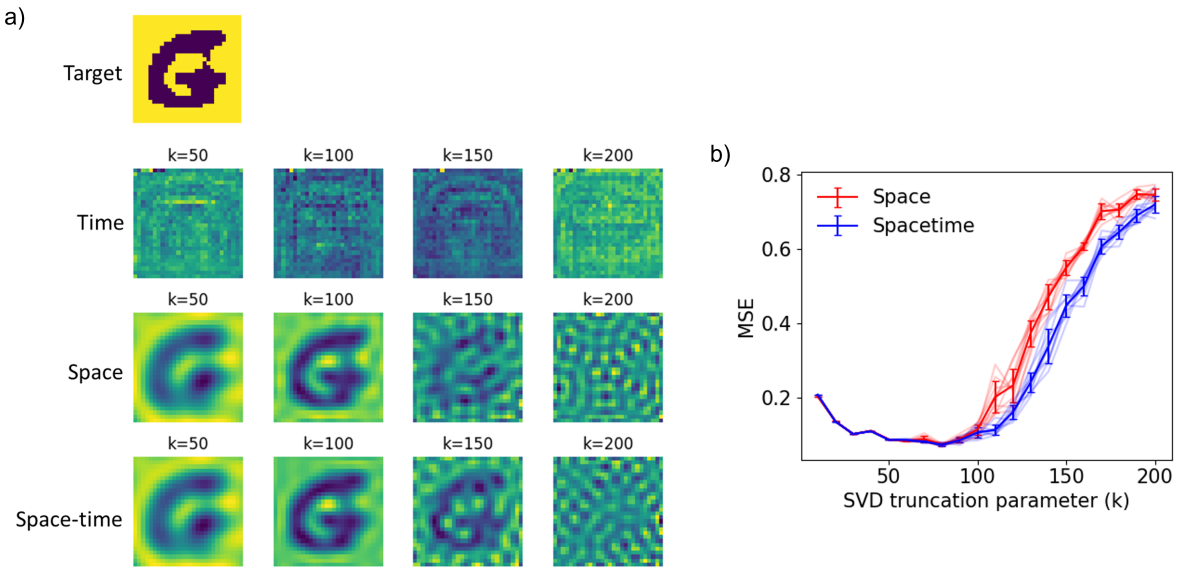


Figure 4.12: Figures published in the supplementary material of [142]. a) Truncated SVD reconstructions using the inverse operators A^{-1} applied to data with Shot noise resolved in different domains. Increasing number of singular vectors (truncated by the parameter k) shows an increased sensitivity to noise in the inversion. There is a stronger resemblance of the ground truth image when using $k = 150$ singular vectors in the space-time domain indicating a better robustness to noise in the data. b) The mean squared error (MSE) between the ground truth and the TSVD reconstructed images for increasing truncation parameter k using data with 10 different Shot noise sampling instances resolved in space-time (blue) and space (red). The mean is indicated by the bold line and error bars are the standard deviation.

where binning the same number of photon counts into fewer pixels will improve the SNR. Using the maximum count number of the simulated measurement N as signal and $\sqrt{N}$ as noise, this corresponds to introducing a $5\times10^{-3}\%$ error in the time domain, 0.2% error in the space domain and a 1% error in the space-time domain.

Simulating through $50\ell^*$ was a non-trivial inverse problem that was not too ill-posed such that a truncated SVD was still possible for $k \sim 100$. Since the problem is very sensitive to noise perturbations, measurements with different noise realisations were compared with the ground truth image using the mean squared error for increasing truncation parameter k. The mean squared error between the ground truth and ten different realisations of noise are shown in Fig. (4.12b), where the mean and standard deviation show a consistent improvement when using space-time resolved measurements even though the SNR is five times worse than resolving only in space.

4.7 Conclusion

Following the sensitivity study in Chapter 3, the experiments in this chapter present an information analysis of diffuse photon measurements that is agnostic to the chosen inversion algorithm. First, a case study using information theory concepts to analyse multipath echoes was introduced which provided insights into the increased accuracy of a machine learning image reconstruction algorithm developed by the authors of [141] when given echo measurements with increasing number of scattering events. The findings show that, surprisingly, there is more gain in independent information from echoes containing up to two bounces than direct reflections. This indicates that, although scattered signals appear to have less direct imaging information, when using inverse algorithms that utilise the full extent of the indirect imaging information, including the scattered signal can improve image reconstruction quality.

The following study analysed the information content of diffuse photon measurements resolved in space, time and space-time domains. For the same number of photons, it was clear that space-time resolved measurements have orders of magnitude more information than resolving in only space since there are more pixels in the distribution and the diversity in both the space and time dimensions is captured. Interestingly, the information in the time domain increases with increasing transport mean free paths (TMFPs) since the distribution of the photon counts is spread over an increasing number of time bins. The information content of the signals continued beyond $1000\ell^*$. However, assuming a large signal-to-noise ratio at these limits would be unrealistic.

The study was continued to show the mutual information captured in each domain between the pixel values of the hidden object and the diffuse photon measurements. It was clear that there is a slight increase in average mutual information when resolving photon counts in space-time compared with exclusively resolving in space. However, this advantage diminishes beyond $100\ell^*$ for the experimental configuration considered, i.e., uniform source illumination, $(5\times5)\text{cm}^2$ field of view, 28×28 spatial pixels and a time resolution of 55ps. The advantage of measuring photon counts in different domains is expected to differ for different experimental parameters, but the chosen parameters are typical and reflect the experimental conditions of [12].

Furthermore, the mutual information between pixels of the hidden object and measurements can be plotted as a map to highlight areas of preserved information. These mutual information maps show that the information about central regions of the hidden objects is less preserved for increasing TMFPs. This is due to cross-talk from neighbouring pixels, which causes increasing noise in the measurements. Information in pixels at the edges of the hidden objects is better preserved since there is no information transmission (i.e., diversity in pixel values) at the borders of the chosen image set. Therefore, to retain as much information as possible about pixels, it is vital to reduce the cross-talk between neighbouring pixels by carefully considering source and detector placement, e.g., a scanning source or patterned illumination would decrease cross-talk

and enhance mutual information.

When considering practical photon collection and incorporating noise into the simulations, the mutual information extended beyond $200\ell^*$. This well exceeds the currently accepted limits for diffuse optical imaging and corresponds to around 12cm depth penetration in human tissue. This limit also correlates closely with the expected number of collected photons, indicating that information about occluded absorbing objects in highly diffuse material remains as long as photons are detected.

An experiment was performed in the lab to measure the extinction of photons for increasing transport mean free path lengths to validate the numerical estimates of information transmission in the highly diffusive regime. It was shown that for the experimental configuration used, photons were collected well beyond $200\ell^*$. Furthermore, when normalising the data and multiplying by the collection efficiency used in the numerical studies, the experimental data precisely coincided with the expected number of TMFPs corresponding to a detection rate of one photon per second using the numerical parameters. This validates the assumptions used in the simulations for the numerical studies and proves that it is reasonable to expect photon transmission through highly diffusive materials beyond $200\ell^*$. The number of photons detected in the experimental measurements was also much higher than expected for the number of TMFPs used in the numerical studies. This highlights the importance of collection efficiency and source power for extending the limits of information transport through highly diffusive materials.

A simple linear operator also assessed the impact of resolving measurements in different domains for improving the ill-posed image retrieval inverse problem. It was shown in Fig. (4.12) that in the absence of *a priori* information used to regularise the problem, resolving in space-time improved the condition number of the inverse problem despite having a lower signal-to-noise ratio than resolving measurements exclusively in space. This results in more accurate image reconstructions, which are more robust to noise perturbations when using measurements resolved in space-time.

Overall, the studies presented in this chapter provide insight into the limits of information transport through highly diffusive materials and identify sensitive experimental parameters that can be optimised to preserve as much mutual information about hidden absorbing objects as possible. For reasonable experimental considerations with today's technology, it was shown to be numerically plausible to collect imaging information in diffuse photons well beyond the regime of $100\ell^*$, up to a limit of around $200\ell^*$, which can be extended by experimental considerations. However, the studies in this chapter do not indicate how to efficiently extract the information contained in the measurements to reconstruct images. Guided by the conclusions of this chapter, the next chapter presents an experiment designed to image beyond $100\ell^*$ using a probabilistic machine learning inverse reconstruction algorithm.

Chapter 5

Imaging beyond 100 transport mean free paths

The analysis of the previous chapters indicate that there exist indirect imaging information in diffuse photon measurements beyond $100\ell^*$. Furthermore, the information can be enhanced by consideration of experimental design such as source power, collection efficiency and source and detector arrangement. Although information theory was used to prove the existence of imaging information beyond $100\ell^*$, it does not indicate if there is enough information to accurately reconstruct images in this regime, or how this information should be used to reconstruct images.

As described in section (2.6), solving an ill-posed inverse problem is possible using a regularised least-squares error, where the choice of prior is based on domain expertise. Indeed this is done in many applications for diffuse optical tomography [11, 12, 111, 148]. However, recent machine learning methods have demonstrated capabilities beyond standard SVD techniques and regularised least-squares optimisation algorithms for solving the inverse problem of diffuse optical imaging [13, 113, 123–125].

In this chapter an experiment is designed using the conclusions of the previous chapter to maximise information through more than $100\ell^*$, and a probabilistic machine learning image reconstruction algorithm developed by the co-authors of [13] is used to asses the potential of imaging at the extreme limits of the highly diffusive regime.

Before discussing the experiment to image beyond $100\ell^*$, the first section of this chapter introduces the machine learning framework with a case study of an ill-posed holographic image reconstruction problem used in [13]. The machine learning framework was developed by the co-authors of the study and my contribution to the holographic image reconstruction experiment was the simulation and collection of experimental data, as well as the training and testing of the measurement in the proposed framework. The image reconstruction quality was compared

with reconstructions from a conventional holographic reconstruction algorithm performed by the co-authors, as well as a conventional deep neural network that I designed, trained, and tested.

The relevance of the experiment to this thesis is to demonstrate an example of current state-of-the-art image reconstruction algorithm that considers the stability of the ill-posed problem in its reconstructions. This framework was also demonstrated by the co-authors of [13] to improve the diffuse optical imaging results of Lyons et al. [12] with measurements corresponding to $> 80\ell^*$, which makes this proposed inversion algorithm a promising option to reconstruct images beyond $100\ell^*$.

5.1 Holographic image reconstruction using variational inference

Phase retrieval of a complex light field is a well-studied problem in optical physics since directly measuring the amplitude and phase of a diffraction pattern requires expensive and delicate interferometric instruments [149]. Instead, measuring the intensity (i.e., the square of the amplitude of the light field) with a conventional digital camera and using algorithms to reconstruct the phase and amplitude information of the light field is a more practical, inexpensive alternative technique. This is an ill-posed problem since many different phase and amplitude profiles can have identical diffracted intensity profiles at the measurement plane. Therefore, this example serves as a simple design of a challenging ill-posed inverse imaging problem to validate a machine learning variational inference framework developed by the co-authors of [13].

Typically deep neural networks output a single deterministic solution given some input data. Since the learned mapping between input and output can be arbitrarily complex, deep neural networks are powerful tools to directly solve inverse problems. Furthermore, because measurements can be learned to map directly to the targets, there is no additional instabilities associated with numerical inversions [150]. The training set of images also introduces an implicit bias to regularise the possible set of solutions to be more relevant to the problem compared with using general assumptions, e.g., introducing an L_1-norm regularisation to promote sparsity [13].

However, if the inverse problem is significantly ill-conditioned then the reconstructed solutions will still be very sensitive small changes in the inputs. Deterministic machine learning models will predict only one plausible solution for a given measurement, whereas in reality there are many plausible solutions. The objective of the framework developed in [13] is to instead use a generative model conditioned on the measurements to predict the probability density of solutions. This framework allows for a prediction of many plausible solutions for the same measurement, which can be used to estimate the pixel-wise mean and variance of plausible image reconstructions as well as estimate the most likely reconstruction.

The goal of the proposed image retrieval algorithm is to reformulate the inverse problem in terms of probabilities using Bayes rule:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}, \tag{5.1}$$

where the task is to approximate the posterior density $p(x|y)$ of the image x corresponding to the measurement y. The data likelihood $p(y|x)$ is the probability of a observing a measurement given an image, and samples of this distribution can be obtained by observing real-world measurements. The distribution $p(x)$ is the prior knowledge of the probability density of all possible images and $p(y) = \int p(x)p(y|x)\,dx$ is the density of all possible measurements.

In practice, we do not have all the required ingredients to produce the true posterior, but it can be approximated by combining all sources of information that can observed:

1. A sampled set of high-fidelity real measurements drawn from the true data likelihood distribution $y \sim p(y|x)$. These samples are often time-consuming and/or expensive to gather.

2. A set of low-fidelity simulated measurements using a forward model based on the physics of the problem. These are samples drawn from an approximate data likelihood $\widetilde{y} \sim p(\widetilde{y}|x)$ and are easy to gather but less accurate.

3. A large set of example target images from an open-source database chosen based on prior expertise to be representative of the true target images $x \sim p(x)$.

Of course, gathering more samples better approximates the true distributions and leads to a closer estimate of the posterior in Eq. (5.1). The proposed framework in [13] describes how these available data sets can be combined to accurately estimate the posterior using conditional variational autoencoders (CVAEs). The full details are not required to understand the work presented in this thesis, however, it is worthwhile summarising the framework to contextualise how these pieces of information lead to the results. The following subsections describing the multi-fidelity model and the inference model used in the framework is a summary of the work conducted by the coauthors in [13] to contextualise the results presented later in this chapter.

Both the multi-fidelity and inverse models of the framework are implemented using conditional variational autoencoders (CVAEs). An autoencoder is a type of deep neural network which first encodes an input to a lower dimensional latent variable. Using only the low-dimensional encoding, a decoder neural network must reconstruct the original input. Like most neural networks, the predicted reconstruction of the input is compared with the true target with a differentiable objective function (e.g., the mean squared error), and the network weights and biases are updated using mini-batch gradient descent using backpropagation [151].

At first glance, reconstructing an input seems like a redundant task. However, it is the lower

dimensional latent representation of the input that is useful. Since the number of learnable parameters (i.e., weights and biases) in the bottleneck of the autoencoder is relatively small, there is a limited capacity to capture all of the information of the original data. Therefore, the network learns to represent only the most information-rich features of the input in the encoding to reconstruct the input accurately. Once trained, the low-dimensional latent vector can be explored by manipulating the latent variable and observing the decoded output. This is a useful method of uncovering the important features in data, much like finding the singular vectors using SVD. However, unlike SVD, which is a linear decomposition of the data, autoencoders can deal with more complex data by harnessing the ability of deep neural networks to learn complex highly non-nonlinear mappings [151].

A variational autoencoder is an extension of an autoencoder architecture introduced by Kingma and Welling [152] that instead learns to map an input to a low-dimensional probability distribution in latent space rather than a deterministic vector. Samples from a highly complex distribution of inputs (e.g., a distribution of handwritten number images) can be mapped to the moments of a much simpler distribution (e.g., a multidimensional Gaussian distribution) using a deep neural network as a non-linear mapping. Once trained, the simpler latent distribution can be efficiently sampled to generate images that are approximately from the same distribution of images as the training set.

A conditional variational autoencoder (CVAE) learns to map to conditional distributions in latent space given additional input conditions. For example, instead of learning the full distribution of handwritten numbers, the latent distribution is conditioned by the label of the class that the image belongs to. This model is useful for inverse problems since the conditional probability of the input can be inferred given an output measurement of a system [153].

An overview of how CVAEs are used in the framework proposed in [13] to find the relevant distributions that approximate the posterior distribution described in Eq. (5.1) will be explained in the following subsections.

5.1.1 Multi-fidelity model

The first consideration of building a model of the posterior in Eq. (5.1) is to gather many samples from the data likelihood $y \sim p(y|x)$. Unfortunately, obtaining real-world experimental measurements is challenging and leads to the undersampling of this distribution. The first step of the framework proposed in [13] is to build a generative model which can be used to non-iteratively sample many measurements that are as close as possible to experimental measurements.

A generative model in this context is a probabilistic model of high-dimensional data that can be stochastically sampled to generate new data points. A simple example of a parametric probabilistic model is modelling a person's height from an ensemble of sample measurements. A

Gaussian distribution can be used to model that data parameterised by the mean and variance, which can be determined using maximum likelihood estimation. Once the parameters are known, a random number generator can generate new data from the distribution. This simple example is one-dimensional, and there is an easy analytical solution. However, in most imaging inverse problems, building a model representing the distribution is much more complex and usually analytically intractable. Therefore, deep neural networks can be used to model these distributions instead.

Usually, in inverse problems, a forward model simulation can be used to generate low-fidelity measurements $\widetilde{y} \sim p(\widetilde{y}|x)$, which approximate real-world measurements but with some error. In the context of the proposed framework, the multi-fidelity model aims to map low-fidelity measurements to high-fidelity measurements using a generative model. Once trained, a large set of high-fidelity training data can be generated for training the inverse model. Mathematically, the multi-fidelity model $p_\alpha(y|x,\widetilde{y})$ that generates high-fidelity measurements y, given target images x and the low-fidelity simulations $\widetilde{y}$, can be used to make an approximation $p_\alpha(y|x)$ to the true data likelihood:

$$p_\alpha(y|x) = \int p(\widetilde{y}|x) p_\alpha(y|x,\widetilde{y})\, d\widetilde{y}, \tag{5.2}$$

where the model is parameterised by α.

To capture the complexity in the distribution $p_\alpha(y|x,\widetilde{y})$, a latent variable model is used. In other words, the goal is to transform the inputs into a lower dimensional latent space, where a simple distribution can represent the data (e.g., a multivariate Gaussian distribution). A deep neural network's parameters α are optimised to output latent variables w that correspond to the moments of a multivariate Gaussian distribution that encodes the input data distribution. Using the concept of a less-restrictive latent variable model leads to an expression for $p_\alpha(y|x,\widetilde{y})$ which is composed of two neural networks with parameters (i.e., weights and biases) α_1 and α_2:

$$p_\alpha(y|x,\widetilde{y}) = \int p_{\alpha_1}(w|x,\widetilde{y}) p_{\alpha_2}(y|x,\widetilde{y},w)\, dw, \tag{5.3}$$

where the outputs of the models are the moments of multivariate Gaussian distributions. The models' parameters α_1 and α_2 are optimised by maximising the likelihood of the log of the expression in Eq. (5.3).

However, Eq. (5.3) requires an integral over latent variables, which is computationally intractable. Therefore a variational lower bound is used as a surrogate optimisation problem [152]. This also introduces a recognition model $q_\beta(w|x,y,\widetilde{y})$, which is also chosen to be a multivariate Gaussian distribution parameterised by the moments output by a neural network, and leads to a tractable

objective function [13, 152, 153]:

$$\underset{\alpha_1,\alpha_2,\beta}{\arg\max} \sum_{k=1}^{K}\sum_{v=1}^{V}\left[\sum_{s=1}^{S}\log p_{\alpha_2}(y_k|x_k,\widetilde{y}_v,w_s) - D_{KL}(q_\beta(w|x_k,y_k,\widetilde{y}_v)||p_{\alpha_1}(w|x_k,\widetilde{y}_v))\right], \quad (5.4)$$

where s denotes independent samples from the latent distribution, v denotes different independent noise realisations used in the simulations to emulate the noise in experiments, and k denotes different images and associated measurements drawn from the distribution of targets $x_k \sim p(x)$. The mathematical derivation to get to Eq. (5.4) from Eq. (5.3) can be found in the appendix of [13].

The first term of the objective function in Eq. (5.4) maximises the log-likelihood of the predictions. Because a normal distribution is chosen for the parametric distributions, the mean squared error between the predictions and the targets can be used - a standard loss function for machine learning models that can be easily differentiated.

The second term in Eq. (5.4) aims to ensure that the recognition model $q_\beta(w|x,y,\widetilde{y})$, which maps all the input variables to the latent variable, is close to the model which maps only the low-fidelity simulations and the target images to the latent variable $p_{\alpha_1}(w|x_k,\widetilde{y}_v)$. The latter generates images when the high-fidelity is unknown, but both are vital for training. The KL-divergence quantifies the statistical distance between the distributions, which has a differentiable closed-form solution.

Once trained, a large set of simulated measurements and associated targets can be used to efficiently generate high-fidelity samples that can be used to train the inverse model and produce a more accurate posterior.

5.1.2 Inverse model

The true posterior $p(x|y)$ of Eq. (5.1) can be closely approximated using a parametric model $r_\theta(x|y)$. Therefore, the objective is to make $r_\theta(x|y)$ as close as possible to the true posterior. This is achieved by minimising the expectation of the cross-entropy between the distributions with respect to the parameters of the model θ [13]:

$$\underset{\theta}{\arg\min}\ \mathbb{E}_{p(y)}\left[H(p(x|y),r_\theta(x|y)\right] = \underset{\theta}{\arg\max}\ \mathbb{E}_{p(y)}\int p(x|y)\log r_\theta(x|y)\,dx. \quad (5.5)$$

The cross-entropy between two distributions P and Q is $H(P,Q) = -\int p(x)\log_2 q(x)\,dx$. The expression in Eq. (5.5) can be rewritten using Bayes rule defined in Eq. (5.1) by [13]:

$$\mathbb{E}_{p(y)}\int p(x|y)\log r_\theta(x|y)\,dx = \iint p(y)\frac{p(y|x)p(x)}{p(y)}\log r_\theta(x|y)\,dy\,dx,$$

$$= \int p(x) \int p(y|x) \log r_\theta(x|y)\, dy\, dx, \tag{5.6}$$

where the expectation is over the distribution of input measurements $p(y)$.

The expression in Eq. (5.6) reformulates the problem to contain tractable distributions. The distribution $p(x)$ is the distribution of targets, $p(y|x)$ is the multi-fidelity model defined in the previous subsection. All that remains is to define a model for the approximate posterior $r_\theta(x|y)$.

Similar to the multi-fidelity model described previously, a CVAE is used as a latent variable model for $r_\theta(x|y)$ such that neural networks can be used to capture the complexity of the distribution [13]:

$$r_\theta(x|y) = \int r_{\theta_1}(z|y) r_{\theta_2}(x|z,y)\, dz, \tag{5.7}$$

where, in this case, the latent variable is represented by z. Substituting (5.7) into (5.6) gives an expression for the objective function in Eq. (5.5):

$$\arg\max_{\theta_1,\theta_2} \int p(x) \int p(y|x) \log \int r_{\theta_1}(z|y) r_{\theta_2}(x|z,y)\, dz\, dy\, dx. \tag{5.8}$$

However, just as the case with the multi-fidelity model, this objective function is intractable to compute since it integrates over a high-dimensional latent variable. Therefore the variational lower bound is optimised instead [13]:

$$\arg\max_{\theta_1,\theta_2,\phi} \sum_{l=1}^{L} \sum_{t=1}^{T} \left[\sum_{s=1}^{S} \log r_{\theta_2}(x_l|z_s, y_{l,t}) - D_{KL}(q_\phi(z|x_l, y_{l,t}) || r_{\theta_1}(z|y_{l,t})) \right], \tag{5.9}$$

where s denotes samples from the latent distribution, t denotes different samples from the multi-fidelity forward model, and l denotes different images and associated measurements drawn from the distribution of targets $x_l \sim p(x)$. The mathematical derivation of Eq. (5.9) from Eq. (5.8) can be found in the appendix of [13].

An illustration of the architecture used in training and testing of the inverse model is shown in Fig. (5.1). In the training phase, three neural networks are trained end-to-end using the objective function defined in Eq. (5.9). The first term in the objective function is to maximise the log-likelihood of the target images. the output distribution $r_{\theta_2}(x|z,y)$ is chosen to be an isotropic multivariate Gaussian with moments determined by the network such that the minimisation of the log-likelihood is simply the mean squared error between predictions and targets. This ensures the decoder faithfully reconstructs the target images.

Similar to the multi-fidelity model, the second term of the inverse model objective function is

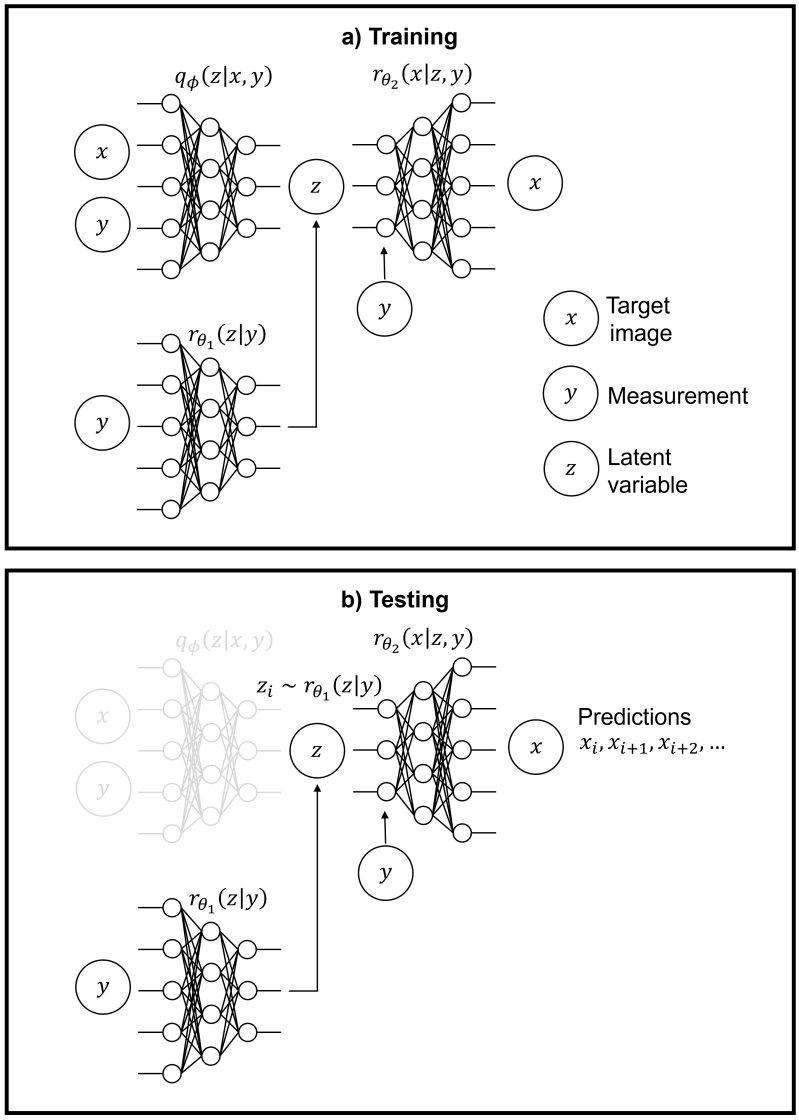


Figure 5.1: a) An illustration of the conditional variational autoencoder (CVAE) used in training inverse model. The objective function defined in Eq. (5.9) optimises the parameters of the neural networks ϕ, θ_1, and θ_2 to minimise the difference between the recognition model $q_\theta(z|x,y)$ and $r_{\theta_1}(z|y)$ by penalising the statistical distance between them using the KL-divergence whilst simultaneously maximising the log-likelihood of the generated outputs of $r_{\theta_2}(x|y,z)$ (e.g., minimising the mean squared error between predictions and targets). b) When testing the trained network, the recognition model is no longer required and samples can be drawn from the latent distribution conditioned only on the measurement. By sampling the latent distribution many times, the ensemble of generated images can be used to calculate the mean and variance of the reconstructions. The fully connected nodes represent neural networks of any type which have the capacity to accurately learn mappings to and from the latent space.

to minimise the statistical distance between the recognition model $q_\theta(z|x,y)$ and the conditional distribution $r_{\theta_1}(z|y)$ using the KL-divergence. In the testing phase shown in Fig. (5.1b) the recognition model can be ignored since these distributions are optimised to be very close in training,

which means the moments of the conditional latent distribution $r_{\theta_1}(z|y)$ can be determined using only the measurement y. This is useful since the target image that we wish to estimate is not required as an input to the model when testing.

For a given measurement, many samples can be drawn from the latent distribution $z_i \sim r_{\theta_1}(z|y)$ and used to generate numerous samples from the approximate posterior of the inverse problem $r_\theta(x|y) = \int r_{\theta_2}(x|z,y)\,dz$. Each sample can be considered different plausible image reconstructions for the given measurement and a pixel-wise mean and variance can be calculated form the ensemble of generated images [13].

Furthermore, generating a reconstruction using the maximum of the latent distribution, i.e., using the mean vector predicted by $r_{\theta_1}(z|y)$, produces an efficient way to determine the near-optimal maximum likelihood of $r_{\theta_2}(x|z,y)$. This reconstruction can be interpreted as the "pseudo-maximum likelihood," which estimates the most likely reconstruction given the measurement and is analogous to the solution predicted by a conventional deterministic neural network optimised using maximum likelihood [13].

The multi-fidelity CVAE has the same architecture shown in Fig. (5.1) but with the inputs, targets and loss function described in Section (5.1.1).

5.1.3 Method

To demonstrate the full potential of the variational framework proposed in [13] for solving inverse problems, a simple ill-posed optics experiment was constructed in the lab (Fig.(5.2)). An expanded laser beam was incident on a digital micromirror device (DMD) which projected amplitude images of binary handwritten numbers from the MNIST open-source data base [154] with a constant phase. A DMD is an array of micron-sized mirrors which can be angled in one of two possible directions. Pixels on the DMD which direct light through the optical system have an "on" state, whereas pixels configured in the other direction can be used as an "off" state. A lens placed at the focal distance from the display maps the image into the Fourier plane, where an iris was used to isolate the first-order diffraction pattern to avoid the high power direct reflections of the laser damaging the camera pixels. Using two lenses in a 4f configuration, a demagnified image of the Fourier plane was projected onto a digital camera.

To exactly reconstruct the amplitude image displayed by the DMD using the information captured in the Fourier plane, both the phase and amplitude is required. However, only the intensity information is detected at the camera, which makes this problem ill-posed since there are multiple amplitude images with various phases that could result in the same intensity image at the detector. Furthermore, the camera was deliberately saturated to further make the inverse problem underdetermined, and therefore more ill-conditioned. This was to encourage the trained model to produce images with increased uncertainty and more diversity in the possible valid solutions

for the given data.

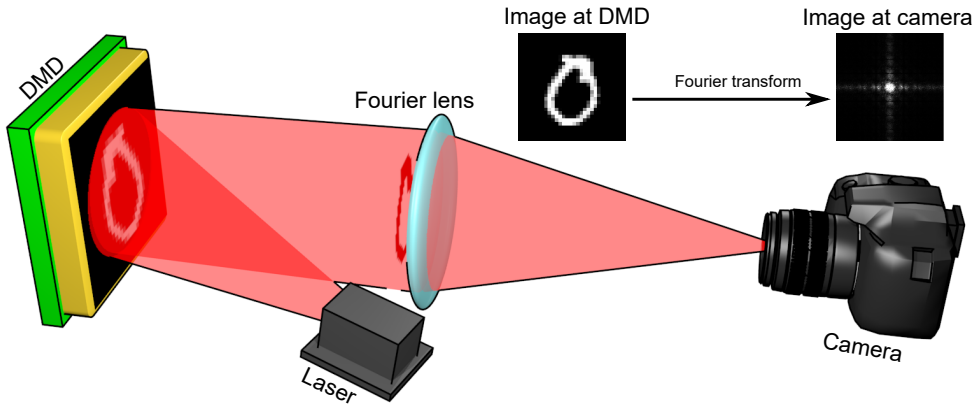


Figure 5.2: Simplified diagram of the experimental setup produced by the co-authors of [13]. A binary amplitude image is projected by the digital micromirror device (DMD) and a lens placed at the focal distance from the DMD display produces the corresponding Fourier image at the camera.

Since the buffer on the DMD can only hold 96 images at one time, the process of gathering experimental data is time-consuming and measurements from only the first 9600 images from the MNIST handwritten digits data set were obtained. This experimental data is samples from the true data likelihood distribution $y \sim p(y|x)$. However, as explained previously, this is not enough samples to accurately approximate the posterior distribution of Eq. (5.1).

To increase the training set, the multi-fidelity forward model described in Section (5.1.1) was used to generate 60,000 high-fidelity measurements using samples from a low-fidelity simulation of the forward problem $\widetilde{y} \sim p(\widetilde{y}|x)$. The low-fidelity forward model created images by taking the real part of the squared modulus of the two-dimensional fast-Fourier transform MNIST digits to emulate the real measurements. The images were then artificially saturated and qualitative resizing and co-registration was performed to match the experimental data. Both the high and low fidelity targets were then down-sampled to 28×28 using nearest neighbour interpolation to reduce computational memory requirements.

The multi-fidelity model was trained using the 9600 high-fidelity experimental observations and corresponding low-fidelity simulations. Note that this model is emulating the forward problem which is well-posed and does not require as much training data as the inverse model.

Once trained, the multi-fidelity model was used to generate high-fidelity measurements for the remaining 50,400 MNIST images and low-fidelity simulations. These multi-fidelity samples were used as training for the inverse model. Additional experimental measurements can optionally be included by concatenating these examples to the samples drawn from the multi-fidelity model, but in this case it was not required.

Both the multi-fidelity model and the inverse model consist of fully-connected neural networks.

The two encoder networks and the decoder network of the multi-fidelity CVAE model were constructed with two initial layers followed by two independent layers that predict the mean and log-variance of a 20-dimensional multivariate Gaussian latent distribution. All layers consist of 500 nodes and use ReLu activation functions. The inverse model CVAE has identical structure but each layer has 800 nodes.

For comparison of performance against other phase retrieval techniques, a hybrid input-output (HIO) complex light-field retrieval algorithm [155] and a conventional deep artificial neural network (deep ANN) (trained only with the 9600 experimental observations) were given the same test examples. The HIO algorithm is a commonly used iterative algorithm for complex light-field retrieval in the field of optics and provides a baseline reconstruction of the target images to compare with. The co-authors of [13] implemented the HIO on test data. A simple supervised deep ANN with four fully-connected layers and a mean squared error objective function was also tested on the same test data to compare the framework with a conventional deep learning approach.

5.1.4 Results

The results in Fig. (5.3) show that performance of the proposed variational framework performance far exceeded other techniques such as hybrid input-output (HIO) and a deep artificial neural networks (deep ANN). The HIO algorithm fails after only one iteration due to the extreme saturation and down-sampling. On the other hand, the deep ANN can make reasonably accurate image reconstructions due the capacity to learn complex non-linear mappings even with such corrupted data. However, the deep ANN is limited in performance due to the relatively small number of training data samples collected experimentally.

The pseudo-maximum images in Fig. (5.3d) show much more accurate reconstructions of the target images. The other benefit of the framework is to take many samples from the latent distribution and reconstruct other possible solutions. In the last row of Fig. (5.3d) it is clear that, although the pseudo-maximum likelihood image closely resembles the target image of a four, some of the sampled reconstructions are of nines with similar structure to the target. This is also reflected in the pixel-wise mean and standard deviation of the images which are non-zero in the top region between the two vertical lines of the number four. The pixel-wise mean and standard deviation is evaluated by drawing numerous samples from the latent distribution and taking the mean and standard deviation of each pixel over the ensemble of decoded images. In all of the examples, the standard deviation shows that the edges of the image are more uncertain than the interior and exterior regions of the shapes. This is exactly as expected for binary images since the regions inside and outside the structures are always a constant value and the borders are what define the shape of the object in the image.

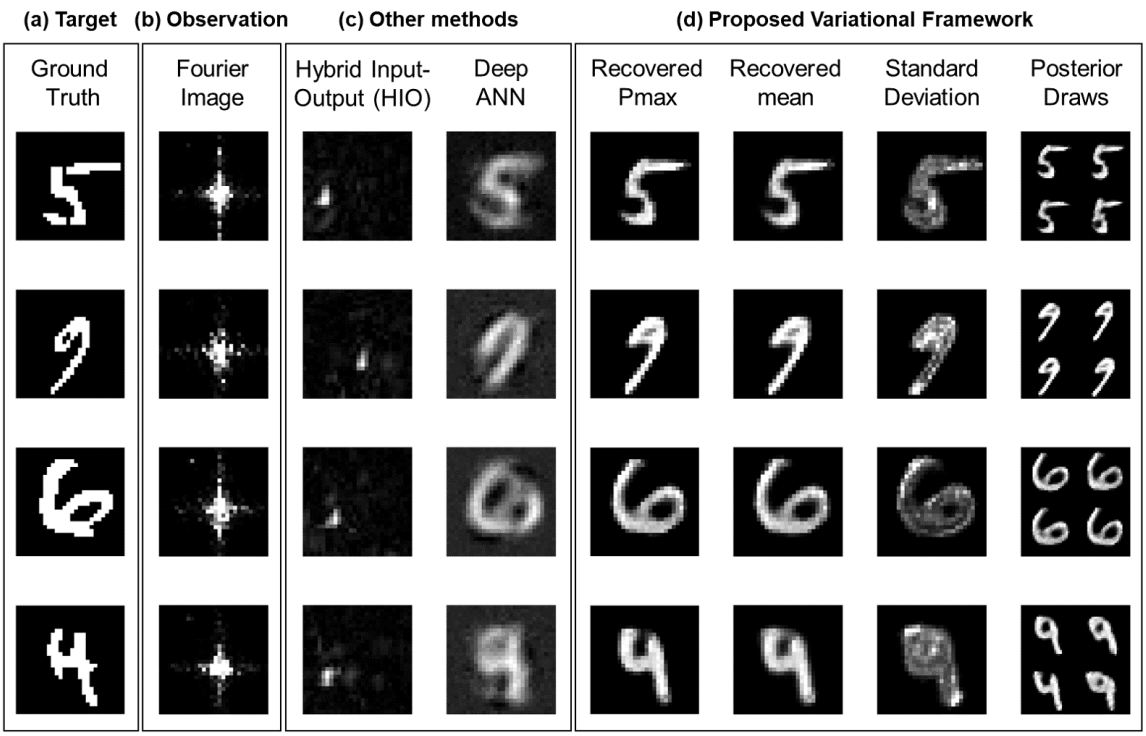


Figure 5.3: Figure published in [13]. a) The target amplitude image displayed on the DMD. b) The saturated down-sampled intensity image of the Fourier plane captured with the digital camera. c) The hybrid input-output algorithm reconstructions which fail due to the instability of the algorithm after one iteration because of the extreme saturation and down-sapling of the measurement. d) The proposed variational framework using multi-fidelity and inverse CVAE models. The pseudo-maximum (Pmax) likelihood is sampled using the mean values of the latent distribution and represents the most likely reconstruction given the measurement. The posterior draws are generated reconstructions from random sampling of the latent distribution. The mean and standard deviation of many samples is also shown.

To test the framework further, the problem was made increasingly more ill-posed by down-sampling the resolution of the Fourier images. The mean and standard deviation of reconstructions Fig. (5.4) reflect the increase uncertainty for lower resolution measurements which have increased diversity in the sampled reconstructed images.

Unlike most inverse image reconstruction methods, this CVAE framework harnesses the capacity of deep neural networks to learn complex non-linear mappings in the inverse direction directly avoiding error introduced by numerical inversion. The capacity of neural networks can produce higher quality reconstructions than conventional image reconstruction approaches and combined with the proposed variational inference framework, can be used to model the conditional distribution of the possible solutions given a measurement. This allows access other possible solutions and an estimation of the pixel-wise uncertainty of the reconstructions.

This framework has also been shown by the co-authors of [13] to work in diffuse optical tomog-

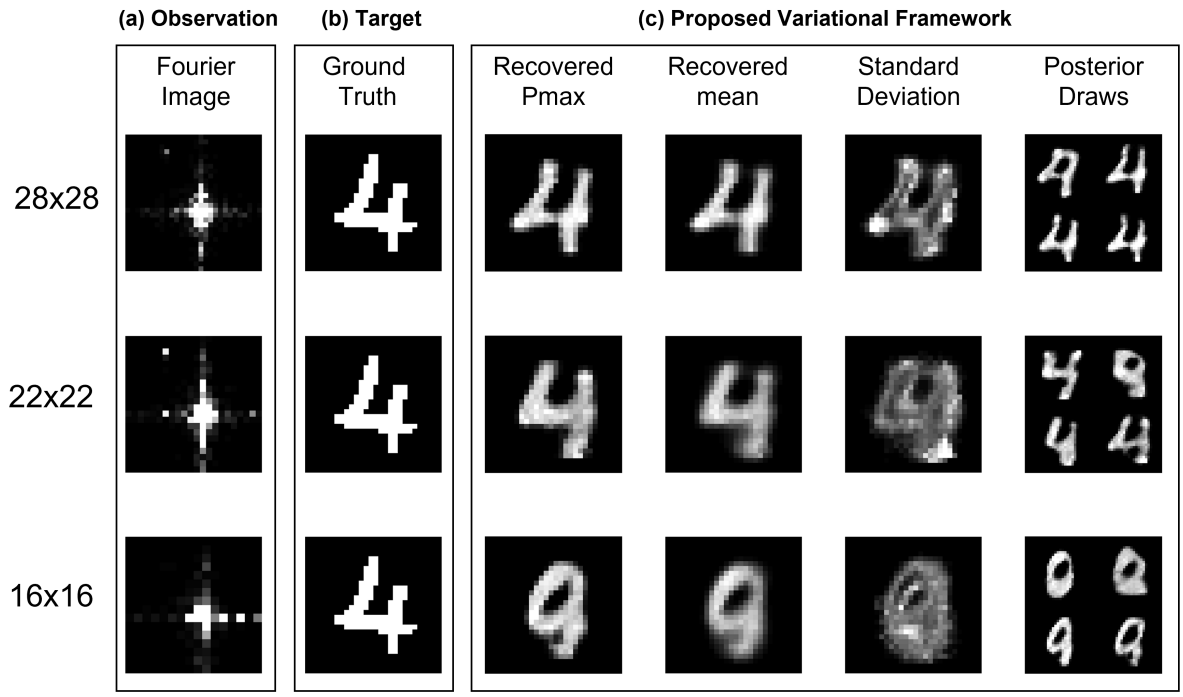


Figure 5.4: Figure published in [13]. a) Experimental Fourier intensity image data downsampled to 28×28, 22×22 and 16×16 for the same target image (b). c) The proposed variational framework which shows the reconstructed image quality degrades with decreasing resolution of the measured data. As expected, the standard deviation and samples from the recovered posterior show high variability to the solution when reaching the critically ill-posed resolution limit of 16×16.

raphy experiments which dramatically outperform the previous algorithm proposed by Lyons et al. [12], and make it a promising candidate inversion algorithm to reconstruct images beyond $100\ell^*$.

5.2 Imaging beyond $100\ell^*$ using variational autoencoding

An experiment was designed to maximise the transmission of information through diffusive materials beyond $100\ell^*$ guided by the results of Chapter4.

The first consideration is to optimise the collection of photons through the material. The selected photomultiplier tube (PMT) has a higher quantum efficiency (15% at 800nm), lower dark count rate ($\sim$15Hz) and larger photosensitive area (5mm diameter cathode) than comparable photon counting detectors such as single photon avalanche diode arrays. The sensitivity and low-noise of this detector ensures a high detection rate of photons, and the large area maximises the solid angle of collection.

However, using a single-pixel detector conflicts with the findings in Chapter4 that there is more information when combining the space and time domains. Therefore, the area of incidence of

the source was chosen to be small, and the position is raster scanned across the material. The advantages are two-fold: firstly, this introduces spatial information to the measurements, and secondly, using a source which covers only a small localised area of incidence will mitigate cross-talk noise compared with sources expanded over a large area of incidence.

This intuition is derived from the analysis of the mutual information maps in Fig. (4.7) that show information is better preserved if neighbouring regions remain at a constant light level. Using a source with a small area of incidence reduces the probability of a photon to interact with regions of the hidden object with large lateral separation from the position of direct line-of-sight as illustrated in Fig. (5.5a). Note that in highly diffusive materials, the source beam will still be significantly broadened in space before reaching the plane of the hidden object and photons are likely to illuminate all regions of the hidden object to some extent. However, the extinction of photons is related to the number of transport mean free paths a photon experience in a material as per the Beer-Lambert law introduced in Eq. (2.26). Regions of the hidden object that are separated laterally from the direct line-of-sight of the source are further from the point of incidence and light will be attenuated more. In the case of uniform illumination, all regions of the hidden object have an equal probability of illumination and information of neighbouring pixels will strongly interfere. The trade-off is that the point source needs to be scanned to gather information from different regions of the hidden object.

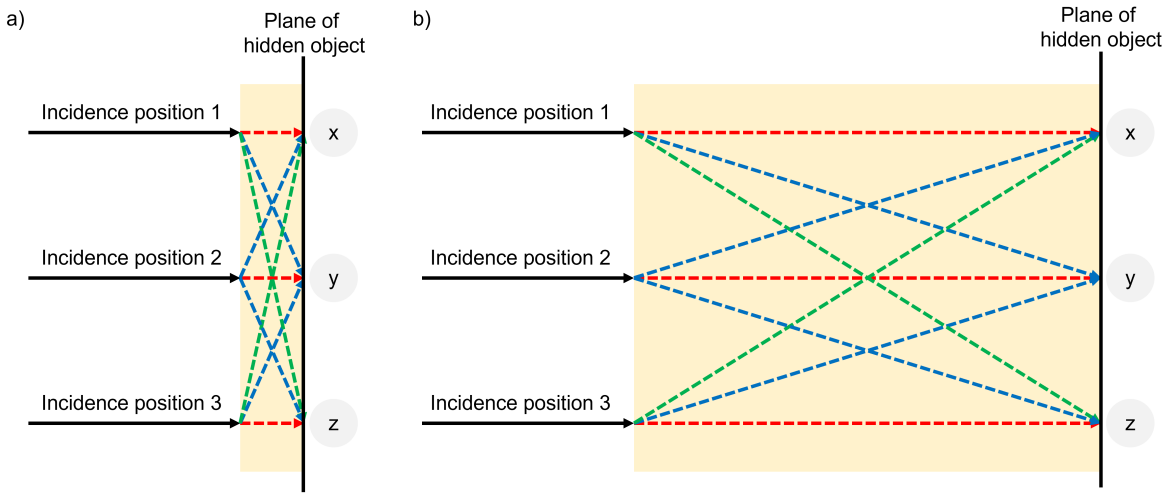


Figure 5.5: An illustration demonstrating short direct optical paths (red), longer optical paths (blue) and the longest optical path (green) from incident source positions to three laterally separated regions at the hidden object plane (x,y,z) for regimes with a) a low number of transport mean free paths (TMFPs) and b) a large number of TMFPs. Scanning the incident source positions sequentially will suppress the cross-talk, because the illumination of the hidden object inside the material is concentrated over a small area compared with uniform illumination. As the number of TMFPs increases, the difference between the path lengths decreases and the information between scanning points will become less diverse.

For materials corresponding to a greater number of transport mean free paths, the distance trav-

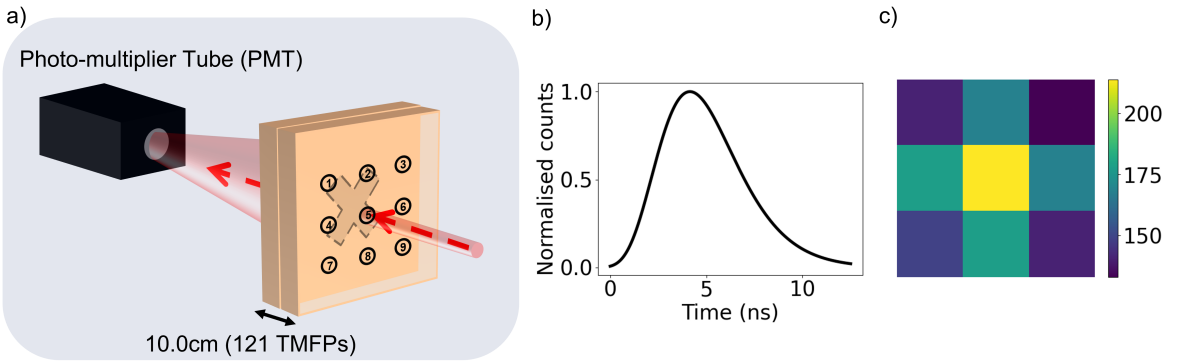


Figure 5.6: Figure used in the conference proceeding [156]. a) Experimental configuration where a pulsed laser is incident on one of 9 different locations on a diffusive material ($\mu_a = 0.04\text{cm}^{-1}$, $\mu_s' = 12.1\text{cm}^{-1}$) in which an absorbing object is hidden (e.g. EMNIST handwritten letter). The transmitted photons are counted using a photomultiplier tube (PMT) operated in Geiger mode and a distribution of the time-of-flight of the photons is recorded using TCSPC. b) The simulated time-of-flight histogram for source position 5 using the analytical forward model. c) The integrated photon counts at each position of the source.

elled by photons to reach the line-of-sight and laterally offset regions of the hidden objects will become increasingly similar as shown in Fig. (5.5b), and the advantage of cross-talk mitigation due to the extinction of photons decreases.

Similarly, if the scanning points are too close together, then the geometry is equivalent to Fig. (5.5b) even for a small number of transport mean free paths and the sampled information of the hidden object is less diverse between the measurements. Therefore, it is redundant to have sequentially scan the incident position with a small pitch separation when imaging through a large number of transport mean free paths. For this reason, a 3×3 grid scan of incident positions with a separation of 2.5cm was chosen to be a reasonable estimate of sampling independent spatial information of objects 2–3cm in size through 10cm of foam (i.e., $\sim 120\ell^*$). An illustration of the proposed experiment is shown in Fig. (5.6a).

Another consideration is to maximise the power of the source. However, if the power per area is too great then the material can be damaged. This is a vital consideration in the context of medical imaging where the maximum permissible exposure of human tissue limits the maximum irradiance that can be used.

Due to the highly ill-posed nature of the problem, the CVAE inverse model described Section (5.1.2) is used to reconstruct images. The CVAE model used in [13] was rewritten using the Keras library and the deep neural networks were modified to use three-dimensional convolutional layers which capture the local structure in both space and time of the $3\times3\times T$ measurement data cube, where $T = 512$ is the number of time-bins in the diffuse photon measurement (Fig. 5.5). The full details of the CVAE network can be found in Appendix B.

First, the experiment was simulated using a modification of the forward model described in Section (3.2) to determine the performance of a CVAE inverse model in the extreme limits of photon diffusion beyond $100\ell^*$ using ideal signals. Since the training and test measurements are generated from an analytical forward model, there is no requirement for a multi-fidelity model since all of the data are samples of the true data likelihood distribution $y \sim p(y|x)$.

The trained inverse model is then tested with real data acquired in the lab. Unlike the holographic reconstruction example described in Section (5.1), gathering measurements of different hidden objects in the lab requires placing 1000s of different configurations of black tape between two foam slabs by hand, which is not practical. Therefore, the real-world measurements are used only for testing the model which is trained using only simulations generated by the low-fidelity forward model $p(\widetilde{y}|x)$.

5.2.1 Simulations

To simulate the experiment in Fig. (5.6) and described in the previous subsection, the forward model introduced in Section (3.2) was adapted to account for a scanning source. Since the forward model uses two-dimensional fast Fourier transforms that require zero padding to avoid high-frequency artefacts at the boundaries, the source beam was always incident in the centre of the field of view and the hidden object was displaced to emulate the signal obtained from the source in different positions as shown in Fig. (5.7a). The result of the model at the output plane was cropped to match the field of view of the detector to ensure the hidden object remained in a fixed position (Fig. (5.7b)). The simulations at the output plane of the second slab (Fig. (5.7c)) were integrated in space to produce the single-pixel TCSPC histograms for each position as shown in Fig. (5.7d). For simplicity in the figures in the remainder of this chapter, the total integration of counts at each source position (as displayed in Fig. (5.6c)) is used to represent the measurements. However, it is the result shown in Fig. (5.7d) which emulates the experimental measurements and is used as the input to the inverse model.

Four instances of the CVAE inverse model were trained using simulations from 5cm, 10cm, 15cm, and 20cm using optical parameters determined by fitting the modified forward model to the experimental results (discussed in the next section). These thicknesses correspond to $60.5\ell^*$, $121\ell^*$, $181.5\ell^*$, and $242\ell^*$. Once trained, the model was tested on unseen simulated measurements and an example of the results are shown in Fig. (5.8).

A qualitative comparison of the reconstructions of the inverse model show that the accuracy of the reconstructed target decreases for increased number of transport mean free path lengths. This is also reflected in the increased standard deviation of the pixel estimates. Interestingly, although the pseudo-maximum likelihood of the posterior $p(x|y)$ of the images x given the measurements y is accurate up to $121\ell^*$, at $242\ell^*$ other plausible reconstructions generated by random sam-

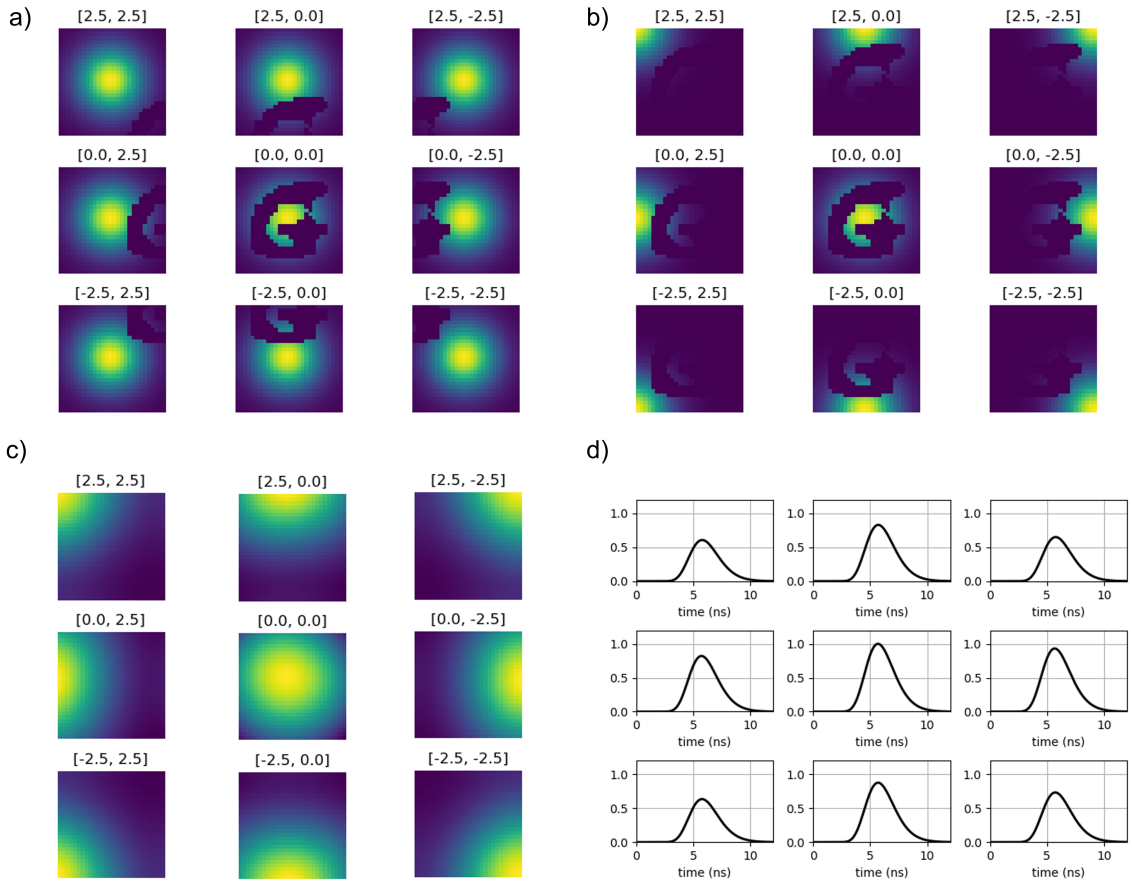


Figure 5.7: a) The forward model simulations of the proposed experiment displaces the hidden absorbing object (e.g., the letter "G") and fixes the position of the source to the centre of the field-of-view (FOV) to prevent boundary artefacts when taking two-dimensional Fourier transforms. b) The simulations at the output plane of the first slab were cropped to match the FOV of the detector which keeps the hidden object in a fixed position and emulates a scanning source. c) The simulations at the output plane of the second slab are integrated in space to produce a single TCSPC histogram (d). The co-ordinate position of the source is indicated above each image in units of centimetres. All images are normalised and the time traces shown in (c) are normalised by the maximum counts of all nine positions.

pling of the latent distribution show a closer qualitative resemblance of the structure of the target. The access to other plausible reconstructions given the measurement is a key advantage of using a probabilistic machine learning model compared with conventional deterministic machine learning models.

Reconstructing images in the highly diffusive regime beyond $200\ell^*$ is clearly possible using machine learning inverse models when given ideal measurements. This supports the claims of Chapter 4 that there exist imaging information beyond $200\ell^*$. Furthermore, the experimental design guided by the previous information theory analysis likely enhanced the collection of in-

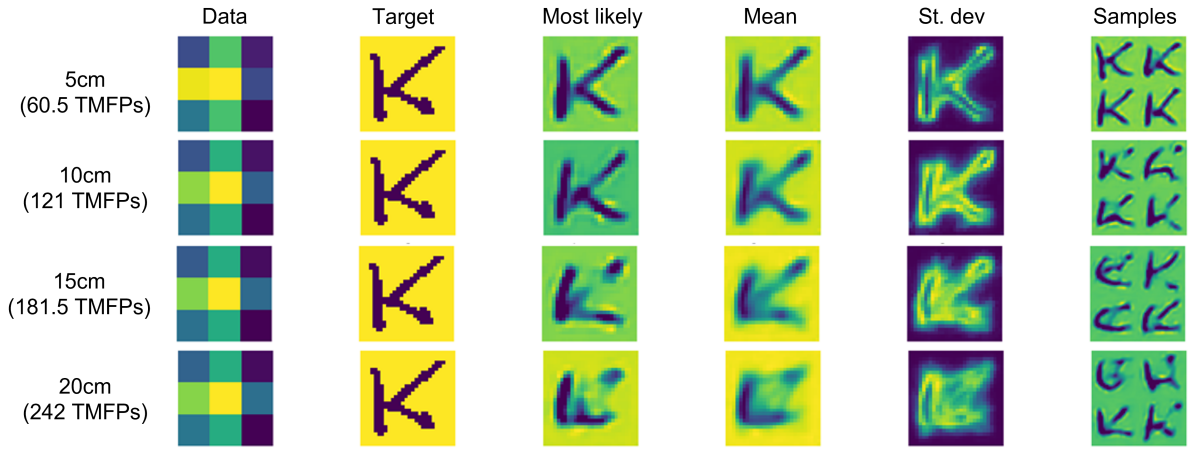


Figure 5.8: Figure used in the conference proceeding [156]. The data, shown as the sum of counts for each detector location, is simulated by the analytical forward model for increasing transport mean free paths (TMFPs). The most likely reconstruction is the pseudo-maximum likelihood of the estimated posterior distribution $p(x|y)$, and generating many plausible reconstructions by sampling the latent distribution gives an estimate mean and standard deviation of each pixel. A close resemblance of the target object can still be seen in the sampled reconstructions at $242\ell^*$. All images are normalised.

formation and enabled reconstructions through as much as $242\ell^*$.

However, as discussed in the results of Chapter 4, the challenge of imaging in the extreme regimes of highly diffusive materials is to overcome the exponential extinction of photons. The next section discusses real-world measurements that were collected in the lab using the same experimental configuration to test the inverse model with data that does not have an ideal SNR.

5.2.2 Experiment

The experiment in Fig. (5.6a) was built in the lab to gather real-world data to test on the CVAE inverse model described above. An ultrafast pulsed laser (800nm, 140fs, 1.12W) is coupled to a liquid light guide (LLG) (5mm diameter) which is placed into one of nine possible positions in a 3D printed 3×3 LLG mount pressed against a polystyrene slab. At the opposing side of the slab, a hidden absorbing object is embedded with black tape and obscured from the detector by a second polystyrene slab. The output fluence of the transilluminated slabs is imaged using a fish-eye lens (Samyang 8mm f/3.5) and focused onto the photocathode of a photomultiplier tube (PMT) (Hamamatsu H7422P-50). A 3D printed lens cover with an $(6.5\times6.5)\text{cm}^2$ cutout was placed over the lens to stop background light from entering, and the area of polystyrene imaged onto the detector. This area was measured to be slightly larger than the cutout ($(6.6\times6.6)\text{cm}^2$) by raster scanning the LLG at the plane being imaged using a reduced laser power and low PMT gain to avoid saturating the detector.

When a photon is detected, the PMT sends an electrical pulse to a TCSPC module which measures the time delay from the laser trigger signal and creates a histogram of photon time-of-flight through the material over many pulses. Although the resolution of the TCSPC module was 12.215ps, the data was re-binned by summing the counts in neighbouring time bins decreasing the resolution to 24.43ps to reduce the memory requirement of the training data for the inverse model. A five second exposure for each position of the LLG is used to create a $3 \times 3 \times 512$ measurement for each of the three hidden objects used.

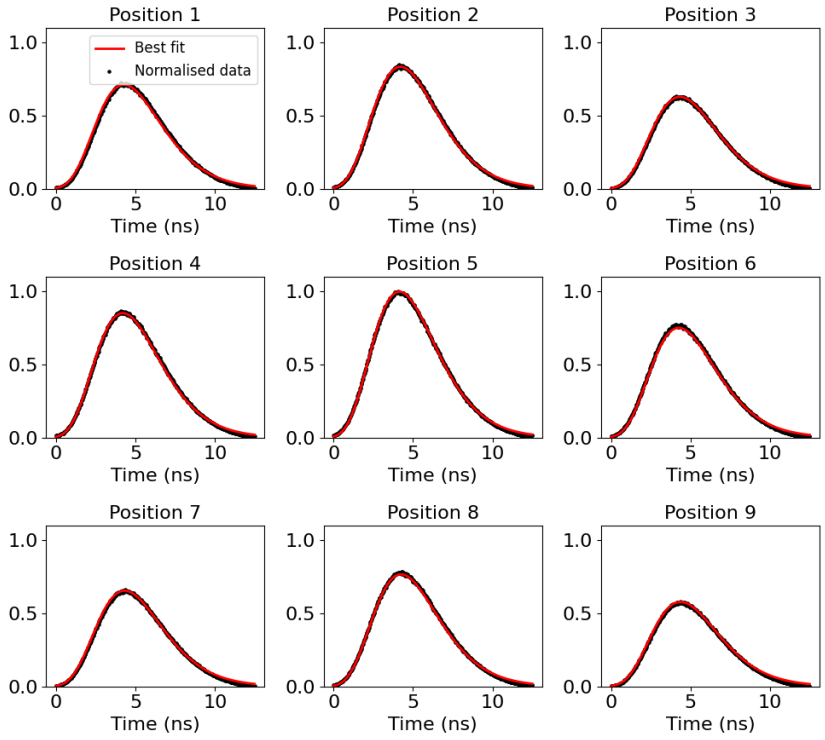


Figure 5.9: The normalised experimental measurements and simulations corresponding to the best fit optical parameters of $\mu_a = 0.04\text{cm}^{-1}$ and $\mu_s' = 12.1\text{cm}^{-1}$ when no absorbing object is present. The position numbers correspond to the labelled source positions in Fig. (5.6).

However, first a measurement of the material without any absorbing object was performed and optical parameters were extracted by fitting the forward model simulation to the measured data using the SciPy library `optimize.curve_fit()` function. Since there is little spatial information collected in the measurements, co-registering the simulations and measurements required calibration. The cropping windows of the simulation output were shifted horizontally and vertically by parameters optimised as part of the fitting process.

The extracted parameters were found to be $\mu_a = 0.04\text{cm}^{-1}$ and $\mu_s' = 12.1\text{cm}^{-1}$ and the best fit simulation overlaid with the experimental data is shown in Fig. (5.9). The total thickness of foam was measured to be 10cm, which is equivalent to $121\ell^*$.

Each measurement was background subtracted and a Savitzky-Golay filter was used to remove

high-frequency noise from the data before testing with the trained inverse model. The results of the inverse model tested on experimental data are shown in Fig. (5.10). There is a qualitative resemblance of the shape of the object in the pseudo-maximum reconstructions, e.g., the "x" target reconstruction has diagonal elongations, the "T" object has vertically elongated features, which are horizontally stretched at the top, and the "=" object has horizontally elongated features.

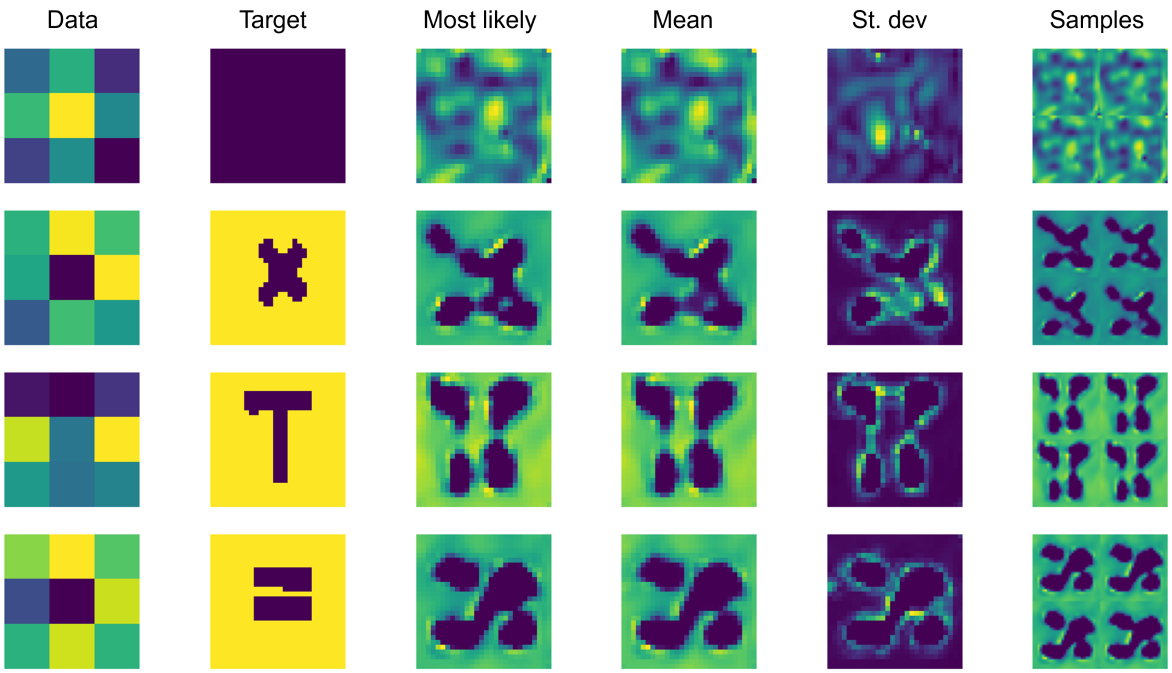


Figure 5.10: Reconstructed images using an inverse model CVAE trained on simulated data and tested with experimental measurements. The most likely image is the pseudo-maximum likelihood estimate. The mean and standard deviation are calculated from drawing many samples from the latent distribution and generating reconstructions. All images are normalised.

The low variability of the sampled reconstructions compared with the simulated reconstructions at $121\ell^*$ in Fig. (5.8) is an indication that the underlying posterior distribution of has not been accurately approximated by the model. In cases where variational autoencoder decoders learn to ignore the samples from the latent distribution and generate degenerate solutions, this is known as "posterior collapse" [157]. This is a common pitfall of VAEs and indicates that the model is not performing as required for this intended application. However, since the variability of the reconstructed images when testing on simulated measurements is performing as expected (Fig. 5.8), but using real experimental measurements results in low variability, this provides insight that the CVAE model is sufficiently trained but the learned latent distribution is not close enough to the true underlying posterior for the real-world measurements.

Ideally, the inverse model should be trained using measurements acquired experimentally. Unfortunately, it is not practical to collect this data in the lab as it is too time-consuming. Instead,

simulations from a more accurate photon transport forward model could be used to better approximate draws from the data likelihood distribution $p(y|x)$. However, more accurate forward models such as Monte Carlo ray tracing or finite element methods is that they are more computationally expensive simulations and take longer to generate simulated measurements.

Using the framework described in [13], a multi-fidelity CVAE forward model could be trained to map from inexpensive low-fidelity simulations to high-fidelity simulations from a more accurate forward model, which requires less training data than the inverse CVAE. Once trained, a CVAE multi-fidelity model could be used to efficiently generate samples from a closer data likelihood approximation $\widetilde{y} \sim p(\widetilde{y}|x)$ that can be used to train the inverse model and better approximate the posterior distribution $p(x|y)$. Following this framework, it can be expected that the fidelity reconstructions would improve and the posterior distribution would be better approximated and produce more diversity between reconstruction samples.

5.3 Conclusion

A state-of-the-art probabilistic machine learning framework for computational imaging inverse problems developed by Tonolini et al. [13] was demonstrated with a challenging ill-posed holographic reconstruction problem. The problem was critically ill-posed due to the extreme downsampling and saturation at the camera such that a hybrid input-output phase retrieval algorithm, commonly used by the optics community, could not converge on a solution.

A conventional four-layer deep artificial neural network to directly learn the inverse problem retrieved images that resemble the targets but the performance was inhibited by the limited number of training examples due to the time-consuming acquisition of experimental data.

Using a low-fidelity simulation of the experiment, a conditional variational autoencoder (CVAE) was used as a multi-fidelity forward model trained to learn the well-posed problem of generating high-fidelity measurements from low-fidelity simulations and targets. A large number of high-fidelity measurements were then sampled from the trained forward model, which closely approximated real-world experimental data. The large ensemble of generated images were used to train a CVAE inverse model to reconstruct target images from measurements. Since a statistical inverse model was used, random sampling of the latent distribution generated other plausible reconstructions to provide an indication of the uncertainty in the predictions.

To increase the difficulty of the inverse problem, the measurements were successively downsampled to decrease their information content. As expected, the inverse model showed increased diversity of the reconstructions for decreasing resolution of the measurements. This demonstrates the framework's ability to accurately determine the uncertainty in predictions of an ill-posed imaging inverse problem, which is not only informative in research settings but is critical when

using machine learning for applications such as medical imaging.

Guided by the results of the information theory analysis of imaging in the highly diffusive regime in Chapter 4, an experiment was designed to maximise imaging information collection from diffuse photon measurements. A CVAE inverse model was used to confirm that there is sufficient imaging information in diffuse photons beyond $200\ell^*$ to reconstruct images of hidden absorbing objects for ideal measurements. Furthermore, a resemblance of the target structure could still be identified in the reconstruction samples up to $242\ell^*$.

The performance of the model trained on ideal signals was then qualitatively assessed on real-world experimental data obtained through $121\ell^*$. The results show that the CVAE inverse model could reconstruct some identifiable features of the target objects. However, the accuracy and diversity of the reconstructions is expected to improve when incorporating a multi-fidelity forward model trained to closely approximate the true data likelihood distribution of the real-world measurements using more accurate simulations of diffuse photon propagation.

The experimental paradigm used in this chapter focuses on non-trivial absorbing objects occluded by a highly scattering slab of polystyrene introduced in Fig. (3.3). However, this toy diffuse imaging problem is a simplified approximation of real-world biological scenarios such as imaging regions of blood oxygenation at the surface of the human brain. In the next chapter, photon propagation in an adult human head is explored experimentally and numerically to examine the feasibility of imaging a biologically relevant sample at the limits of the highly diffuse regime.

Chapter 6

Transmitting photons through an adult human head

Transmitting light diametrically through an entire human head is a challenging biomedical optics problem which, if solved, would have a significant impact for medical and research applications. Functional near-infrared spectroscopy (fNIRS) has been shown to outperform the sampling speed of fMRI and the spatial resolution of EEG for monitoring brain activity at a relatively inexpensive cost [30, 40]. The portability, robustness to motion artefacts, and real-time brain activity monitoring capability of fNIRS devices has enabled neuroscience studies which were previously not possible e.g., due to the limited mobility, loud noise, and cost of fMRI, or the low spatial resolution and sensitivity to motion artefacts associated with EEG [41].

The field of fNIRS indirectly monitors neural activity of the brain by measuring cerebral blood oxygenation. The human head has a relatively complex geometry compared with imaging tumours in breast tissue or identifying regions of absorption in a homogeneous scattering phantom. The heterogeneous superficial layers of tissue and bone not only cause added complications for the modelling of photon propagation into and out of the head, but the vasculature in these layers also transport oxy- and deoxy-haemoglobin which interferes with the measurement in the deeper volume of interest at the surface of the brain.

Despite these difficulties, fNIRS has overcome the complexity of the heterogeneous modeling using Monte-Carlo ray tracing models of anatomically correct MRI scans of subjects and, using multiple source detectors or time-gating methods can mitigate the sensitivity of the measurements to superficial layers [70, 72]. However, the challenge of extending the depth of sensitivity beyond 4cm from the scalp remains unsolved and would enable higher fidelity image reconstructions, access to monitoring activity in cortical folds (sulci), and would extend the use of fNIRS to many clinical applications which require sensitivity to deeper regions of the brain, e.g., imaging

deep intracranial haemorrhages [62, 158].

An attempt to transmit light diametrically through the adult human head from one temple to another was presented in one of the pioneering experiments of fNIRS led by Jobsis [49]. The experiment aimed to detect a signal corresponding to an increase in transmission of near-IR light due to a decrease in cerebral blood volume during hyperventilation where a measurement of the transmitted number of photons was recorded every ten seconds. However, the results were incomplete due to stopping the experiment early before the signal could return to baseline. Since then, the only studies which detect light diametrically through the head use neonatal or infant subjects [50] who have more transparent skulls and smaller diameter head compared with adults.

The theoretical attenuation ratio order of magnitude for transmission of photons through the diameter of the head can be estimated using Eq. (3.2) using reported global optical coefficients for the human head. Typical values of human brain optical coefficients in the near-IR are of the order $\mu_a = 0.1\text{cm}^{-1}$ and $\mu_s' = 10\text{cm}^{-1}$ [25]. The typical human head has a diameter of around 15cm. This provides a rough estimate of an attenuation of $\sim 10^{-12}$.

However, more precise average estimates of the global parameters for the human head have been reported in fNIRS studies [159, 160] to be $\mu_a = 0.13\text{–}0.14\text{cm}^{-1}$ and $\mu_s' = 7.2\text{–}12\text{cm}^{-1}$ at a wavelength of 830nm. Using these reported estimated gives an attenuation between 10^{11} and 10^{14}. For context, this corresponds to a transmission rate between 10^4–10^7 photons per second for an input power of 1W at 800nm ($\sim 10^{18}$ photons per second). However, in practice the quantum efficiency ($\sim 10\%$ at 800nm) and collection efficiency of the detector ($\sim 10^{-5}$) reduces the detection rate to around one photon per second or less. Moreover, the noise of photon counting detectors is not ideal which makes it hard to distinguish if a photon count is from the head or due to the intrinsic dark count rate associated with the detector.

Unsurprisingly, the fNIRS community use a reflection geometry rather than transmitting light through the head. However, the relatively large signal-to-noise ratio of back-scattered photons makes it challenging to post-select photons that have interacted with deeper regions of the brain beyond the outermost surface (gyri). Detecting photons in a transmission geometry ensures that the photons have interacted with deeper regions of the brain.

The conclusions of the work presented in previous chapters of this thesis indicate that the limit of transmitting information in the highly diffuse regime extends well beyond what is currently acknowledged as the limits of imaging through diffusive media. It was also concluded in Chapter 4 that if photons can be detected, then there exist information for imaging. In this chapter, the limit of photon transmission is tested with an ambitious and relevant biomedical experimental challenge of transmitting photons diametrically through the adult human head.

6.1 Experimental methods

Using time-correlated single photon counting (TCSPC), signal measurements in low-photon number experiments are identifiable by the correlation between the time of photon detection and a pulsed source. Furthermore, the time-of-flight distribution of photons transmitted through the head is expected to resemble the characteristics of photon diffusion. This is an information-rich signal which can directly be compared with simulations of time-resolved photon migration that can aid in the process of determining the plausibility that the detected light was transmitted through the head.

First a description of the evolution of the experimental work is introduced which uncovered important considerations when detecting light through extreme length scales. The analysis pipeline is then discussed which lead to the result of plausible time-correlated signals. The following section then compares features of the signal with Monte Carlo simulations and elaborates on the sensitive parameters which could lead to variation in the signal.

6.1.1 First steps using SNSPDs

The first attempt to transmit photons diametrically through an adult human head used superconducting nanowire single-photon detectors (SNSPDs). These detectors have very high quantum efficiency ($\geq 85\%$ at 800nm) and low dark count rate ($\leq$ 10Hz) which make them one of the most sensitive single-photon detectors in the near-infrared [161].

The experiment was set up as shown in Fig. (6.1). The laser power was set to 20mW using a half-wave plate and polarising beamsplitter (Fig. (6.1a)) which was coupled into a multimode fibre to direct the laser light to a collimator pressed against the side of the head. The light was collected using a collimator placed on the diametrically opposite side of the head and guided to an SNSPD using a single mode fibre. Upon the detection of a photon the SNSPD sends an electrical NIM pulse to the TCSPC card which is synchronised with the laser trigger to produce a histogram of the time-of-flight of photons through the head.

Despite the high sensitivity of the detectors used in this experiment, it is fundamentally flawed by the requirement of coupling the detector to a single-mode fibre. This limits the detection of photons to a small angle of collection over a small area on the side of the head.

Nonetheless, convincing signals of photon diffusion were obtained. However, using an infrared viewing scope, it was clear that the multimode fibre used to guide light to the head from the laser was leaking light through the outer plastic jacket of the patch cord. When the optical fibre was brought closer to the collection collimator, the signal dramatically increased, which confirmed the cause of the signal. Interestingly, moving the fibre close to the collection collimator without the presence of a human head in the set up did not cause and increase in the signal detection. This

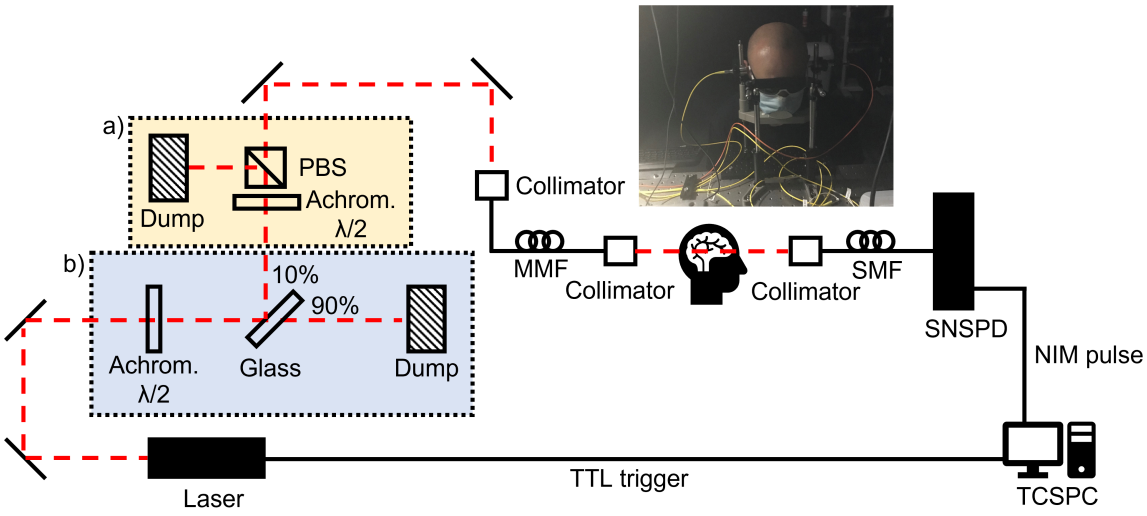


Figure 6.1: The experimental set up for head transmission experiments with superconducting nanowire single-photon detectors (SNSPDs). a) Power control using an achromatic half-wave plate and a polarising beam splitter (PBS). b) Before power control, 90% of the light from the laser transmitted using an anti-reflection coated beam sampler which is absorbed by a beam dump as a safety precaution to prevent high powers reaching the subject. After power control, the light is collimated into a multimode fibre (MMF) and a collimator projects the light directly onto the head. A collimator on the opposing side couples light into a single mode fibre (SMF) which is input to an SNSPD. For longer wavelengths the layout is identical, but the laser is used to pump an optical parametric oscillator (Coherent Levante IR fs OPO) which can be tuned from 1300nm-2000nm and the SNSPD detector is swapped (Single Quantum Eos 1350nm). (Laser = Coherent Chameleon Discovery (Tunable 650nm-1300nm, 100fs, 80MHz); SNSPD = Superconducting nanowire single photon detector (Single Quantum Eos 900nm); TCSPC = Time-correlated single photon counting module (Becker and Hickl SPC-150N)

is because the light leaked from the fibre was not in direct line-of-sight with the collection optics, and the presence of an object immediately in front of the collection collimator was required to reflect the light into the detector.

Although the experiment did not achieve the original goal of transmitting photons through the human head, it highlighted the importance of the collection of photons compared with detector sensitivity. The measurement of a convincing diffuse light signal of light leakage through the jacket of an optical fibre also reinforced the importance of isolating background sources of signal from the experiment.

6.1.2 Increasing angle of collection using SPADs

Light exiting a highly diffusive material is emitted over a large area with a uniform angular distribution. Therefore, to maximise the collection of diffuse light, the area and solid angle of collection must be as large as possible. The solid angle of collection of a fibre Ω is related

to the numerical aperture (NA) by $\Omega = 4\pi \sin^2\theta = 4\pi(\mathrm{NA})^2$, where the numerical aperture is $NA = n\sin\theta$, $n \simeq 1$ in air, and θ is the half-angle. However, the collection of light from a fibre also scales with the area of the fibre core. The product of the solid angle of collection and the area of collection $A\Omega$ is known as etendue (or throughput) of radiative transfer. In air, the radiance L and $A\Omega$ product is conserved [162]:

$$P = LA\Omega = L'A'\Omega', \tag{6.1}$$

where P is power (W), L is the radiance (W/m^{2}sr) from the object with area A (m^2) collected by the optics of the system with solid angle Ω (sr), and the primed variables are for the corresponding image plane as shown in Fig. (6.2).

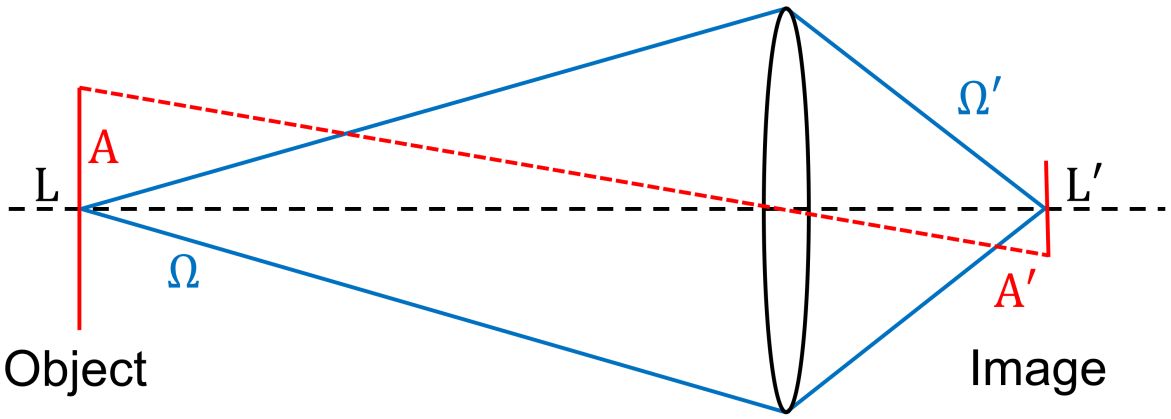


Figure 6.2: Figure adapted from [162] illustrating the conservation of etendue. The power, i.e., the product of the radiance L from an object with area A collected by a lens with a solid angle of collection Ω, must be conserved in the image plane.

When using a fibre, the area and solid angle of collection is fixed by the core size and numerical aperture. Introducing a collimating lens before the fibre (as per the experiment shown in Fig. (6.1)) will collect light from a larger area in the object plane but must sacrifice the solid angle of collection due to the conservation of etendue (Eq. (6.1)). The collection of photons is therefore inherently restricted by the $A\Omega$ product of the single-mode fibre required to couple to the SNSPD, irrespective of the optics used to collect the light.

Silicon photomultipliers (SiPMs) do not have a requirement to be fibre coupled and the $A\Omega$ product can be made much larger. In the next evolution of the experiment the SiPM was pressed directly against the temple region of the head and held in place using velcro straps (Fig. (6.3a)). The active area of the SiPM used in experiments was $A = (1.3 \times 1.3)\mathrm{mm}^2$ and the sensor is exposed with only a minor recess into the protective casing. This means the solid angle of collection can be increased when placed in close proximity to the head. However, due to the increased angle of collection, this prototype was very sensitive to light entering the detector from gaps between

the contact of the detector and the skin. Unsurprisingly, there was clear evidence of scattered background light from around the room in measurements.

To reduce reflection from the surrounding environment in the measurements, a rugby scrum cap was modified to accommodate the source fibre and SiPM detector to block light from entering in the gaps between the detector and the skin. An image of this prototype indicating light leakage through gaps in the cap is shown in Fig. (6.3b). It is apparent from this image that the diffusive signals from the background environment most likely originates from back-reflections of tissue immediately surrounding the source position. To suppress the background light further, black cloth was used to block light from regions surrounding the source.

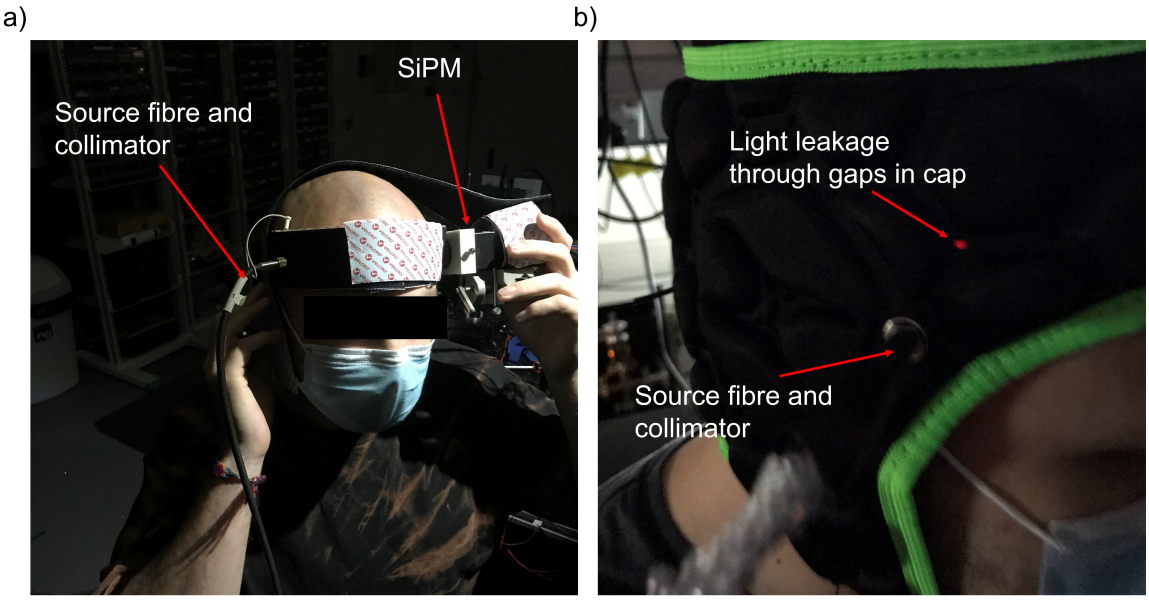


Figure 6.3: a) The silicon photomultiplier (SiPM) was placed directly against the head to maximise the solid angle of collection. b) In the following prototype, a rugby scrum cap was modified to accommodate the source fibre and detector in an attempt to suppress background light from entering the detector. However, it is evident that the source of diffuse light most likely originates from back-reflections of the tissue immediately surrounding the source, as indicated by the light leakage through gaps in the cap.

When carefully blocking background light sources, a two-minute acquisition showed no time-correlated photon detection synchronised with the laser. This is assumed to be because of the relatively low quantum efficiency of the SiPM at near-infrared wavelengths ($\sim 8\%$ at 800nm). Furthermore, the active area $(1.33 \times 1.33)\text{mm}^2$ is small compared with the area of the side of the head, and the modifications to suppress background light increased the distance between the detector and the head from 0.5cm to around 3cm, significantly reducing the solid angle of collection from $\Omega \simeq 0.8$ to $\Omega \simeq 0.02$. Multiple two-minute recordings could have been taken and averaged, but the main limiting factor of using the SiPM is the high dark count rate (~ 2.5kHz),

which would result in a low signal-to-noise ratio compared with other detectors.

Guided by the consideration of $A\Omega$ product, quantum efficiency, and dark count rate, numerous other detectors were trialled. Measurements at 1720nm were attempted with an InGaAs single photon avalanche diode (SPAD) (IDQ ID230), but no time-correlated signal was apparent. This InGaAs SPAD is also fibre coupled, which restricts the $A\Omega$ product.

Using a large NA (0.52) and core diameter (5mm) liquid light guide to collect light which was projected directly onto a 32×32 SPAD array (MF32, commercialised by Photon Force) also yielded no time-correlated signal. The disadvantages of the SPAD array are low quantum efficiency at the near-IR wavelengths ($\sim 6.5\%$ at 780nm), and the on-chip TDC electronics reduce the fill-factor of the detector to around 2% [163].

The SPAD array was swapped for an electron-multiplying charge-coupled device (EMCCD) which has a pixel array area two orders of magnitude higher (177mm^2), a high quantum efficiency (75% at 800nm). However, EMCCDs do not have ultrafast timing capabilities for TCSPC experiments. Therefore, the measured photon counts are not correlated with the laser and cannot be distinguished from noise and background counts.

6.1.3 Optimising diffuse light collection from the head

An experiment was designed to optimise the sensitive parameters discussed in the previous subsection, namely maximising the $A\Omega$ product at the detector plane whilst minimising back-reflected light from regions of tissue immediately surrounding the source and suppressing any time-correlated diffuse light in the room from entering the collection optics.

The most suitable detector for measuring diffuse photons in the near-IR was determined to be a Hamamatsu H7422P-50 photomultiplier tube (PMT). This detector has a 5mm diameter photocathode optimised for near-IR wavelengths ($\sim 15\%$ at 800nm) and a low dark count rate (~ 15Hz). A fish-eye lens (Samyang 8mm f/3.5) was placed in front of the PMT to image a large area of approximately $(6.6 \times 6.6)\text{cm}^2$ onto the photocathode as shown in Fig. (6.4a). A light baffle constructed from sponges wrapped in layers of black cloth was used to create a light-tight seal around the head and prevent stray light from entering the optics. A 3D printed black lens cover with a $(6.5 \times 6.5)\text{cm}^2$ area cutout was placed over the lens to suppress background light further.

The PMT was confined in a black foamboard enclosure filled with black cloth. The detection area shown in Fig. (6.4b) was then also filled with black cloth and enclosed. The laser coupling area and head area were also enclosed, leaving only one side for access in and out of the central head compartment. As a final measure, black cloth was layered over all the enclosures and a laser safety curtain was placed as an outer layer.

The source laser (80MHz, 1.2W, 140fs, 800nm) was directed to the head area with a liquid light

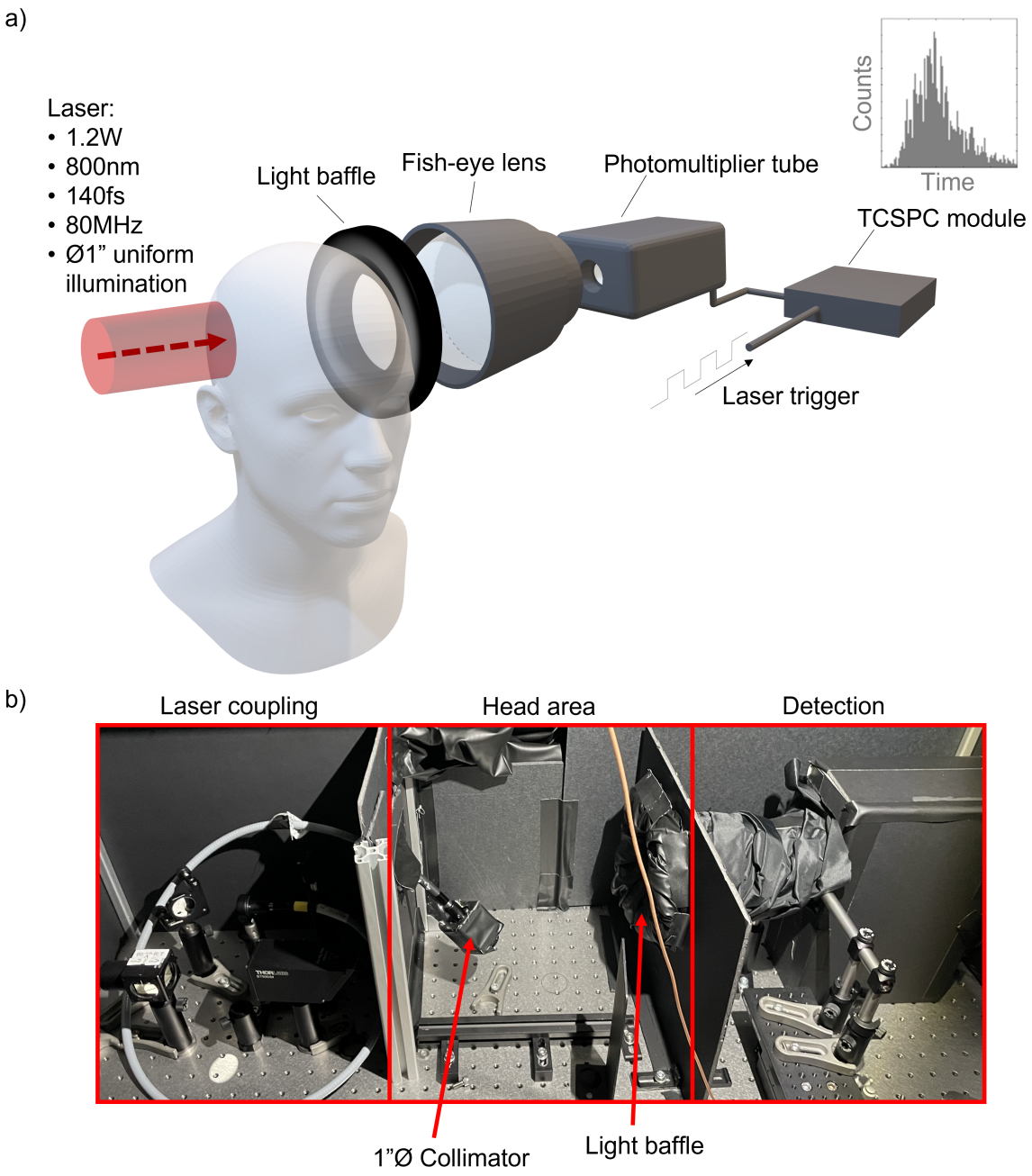


Figure 6.4: a) Experimental design to optimise the collection of diffuse light transmitted through the adult human head. A pulsed laser is expanded to one-inch diameter using a collimator which uniformly illuminates the side of the head. A light baffle made from black cloth and sponges creates a light-tight seal that conforms with the shape of the head and prevents background light from entering the collection optics. A fish-eye lens is used to image an area of $(6.63 \times 6.63)\text{cm}^2$ onto the photocathode of the photomultiplier tube. A TCSPC module time-correlates the detected photon counts with the laser synchronisation trigger from an optical constant fraction discriminator. b) An image of the set up in the lab with the outer enclosures removed showing the laser coupling, head area, and detection compartments.

guide (LLG) attached to a one-inch collimator which expanded the beam uniformly. During an experiment, the participant presses the collimator directly against the side of the head above and behind the ear while maintaining pressure on the sponge light baffle. This corresponds to an irradiance of 2368Wm^{-2}, which is 75% of the maximum permissible exposure for skin outlined in the British laser safety standards for (3169Wm^{-2}) [143]. An experiment collected up to 30 two-minute acquisitions with the laser on and a two-minute background reading with the laser switched off every three measurements. The TCSPC module was configured to trigger every fourth synchronisation pulse from an optical constant fraction discriminator since the signal's expected shape and time delay was unknown.

The participant was informed that they could stop the experiment or take breaks at any time by alerting the laser operator to switch off the laser. The status of the laser was communicated loudly to the participant with a countdown any time the laser was switched on or off. Participants provided informed consent before experiments and the experiments were approved by the University of Glasgow College of Science and Engineering Ethics Committee (application number: 300180292).

The experiment was repeated thirty-one times across two experimental setups. The first setup is pictured in Fig. (6.4) and described above. However, this was disassembled, and a second modified setup was built in a new lab to enhance the light collection efficiency and reduce light leakage.

To reduce light leakage from the area around the source, a silicone mould made to make a tight seal with the skin (Fig. (6.5a)). This was made by pouring liquid silicone into the bottom half of an aluminium can around a threaded lens tube placed in the centre. Once set, the circular ridge of the mould made from the bottom of the can conformed to the shape of the head and prevented gaps for light to escape. The lens tube was glued to the silicone so that it could be screwed onto the collimator as a single piece.

To isolate the head area (Fig. (6.4b)) from background light, a black cloth body wrap was made to surround the participant as shown in Fig. (6.5b). This alleviated the requirement for placing black cloth around the participant for every trial which was quicker, more reproducible, and made room for a fan and a foam cube the participant can use to rest their forehead, which improved the comfort of experiments. to improve the comfort of the experiments (Fig. (6.5c)).

In an attempt to improve light collection, the fish-eye lens with solid angle of collection $\Omega = 3.3 \times 10^{-3}$ was replaced with a larger aperture lens with $\Omega = 0.366$ but less magnification. This was a trade-off between the area of the head imaged onto the detector compared with the range of angles which could be collected. Since the PMT photocathode was recessed into the casing, this restricted the aperture to $\Omega \simeq 0.2$. Transilluminating polystyrene foam with a constant power showed around a twenty-fold improvement of the collection of photons when using the lens with

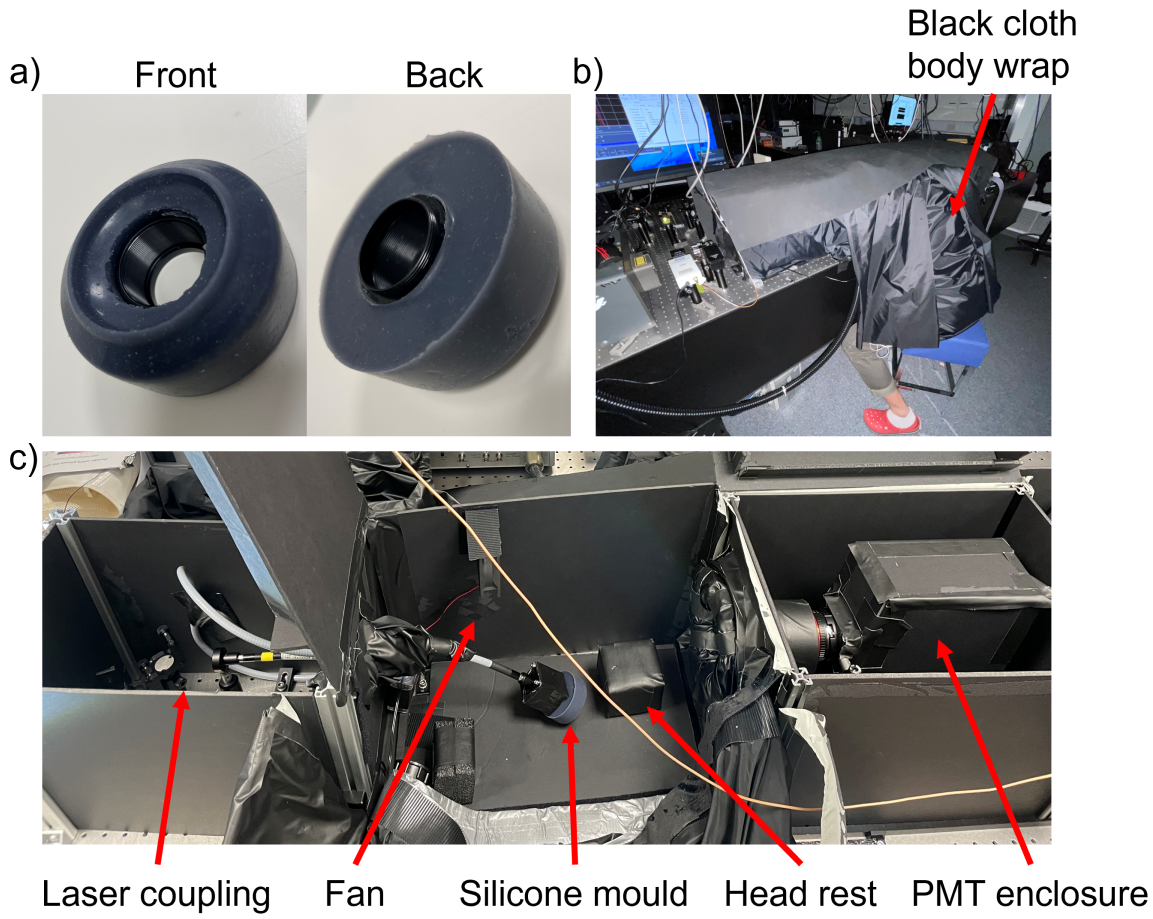


Figure 6.5: The modified experimental setup. a) A silicone mould conformed to the shape of the head and prevented reflected light leakage from the area around the source. b) A black cloth body wrap was attached to the head enclosure with velcro and was a more reproducible way of suppressing background light from entering the setup. c) A fan and a head rest were introduced to increase the comfort of experiments for participants.

greater solid angle of collection.

However, this was improved further when removing the lens and placing the PMT directly against the foam. Although the collected photons are from a smaller area, using a bare sensor also removed any possible losses due to Fresnel reflection from lens surfaces and absorption due to anti-reflection coatings that are not optimised for near-IR wavelengths. Therefore, the lens shown in Fig. (6.5) was removed.

6.2 Results

An example of a raw and a filtered signal taken with participant one using the first experimental set up (Fig. (6.4)) for increasing number of experimental trials is shown in Fig. (6.6). A third-order Savitzky-Golay filter was used with a window of 251 time-bins to filter the data. Although

there is no measurement of the impulse response function (IRF) for this measurement, the curve resembles a solution to the photon diffusion approximation and has an average full-width half-maximum (FWHM) of 1.99 ± 0.01ns.

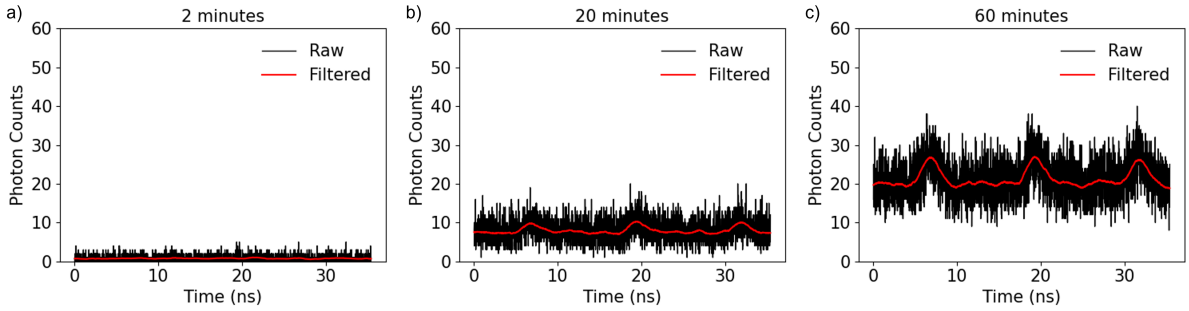


Figure 6.6: An example of raw and filtered data for participant one using the first experimental setup. There is no signal after a) a two-minute experiment trial, but after b) twenty minutes a time-correlated signal emerges, which is clear after c) sixty minutes of experimental acquisition.

Results from the second experimental setup (Fig. (6.5)) show similar signals and their averaged IRF. The peak-to-peak time delay from the IRF to the signals is 2.1 ± 0.2ns and the FWHM is 1.7 ± 0.2ns. The transparent signals shown in Fig. (6.7) are the averaged filtered background measurements.

The rate of detected photons above the background for the signals in Fig. (6.6) is 2.0 ± 0.9 photons per second. This was found by cropping the raw signal data and raw background data to include two complete periods of the laser repetition rate (9.16ns to 34.18ns). The mean over all two-minute trials was calculated and the difference between the sum of the averaged signal and background measurements was found. The difference was multiplied by two since the TCSPC module triggered on every fourth pulse from the laser. The result was then divided by the acquisition time to find the photon rate for the three experiments, and the mean and standard deviation was calculated.

The experimental attenuation was determined to be $2 \times 10^{13} \pm 1 \times 10^{13}$. This estimate was found using the ratio of the photon detection rate (2 ± 1 photons per second), with the input photon rate calculated using the input power (1.2W) and wavelength (800nm). This ratio was then normalised by the quantum efficiency (15%) and the $A\Omega$ product ($\sim 4 \times 10^{-5}\text{m}^2$) of the detector. The experimental photon detection rate is the same order of magnitude (1 photon per second) estimated by the theoretical arguments made in the introduction to this chapter which is a promising indication that the experimental attenuation of photons matches the attenuation expected for photons transmitted through the entire head.

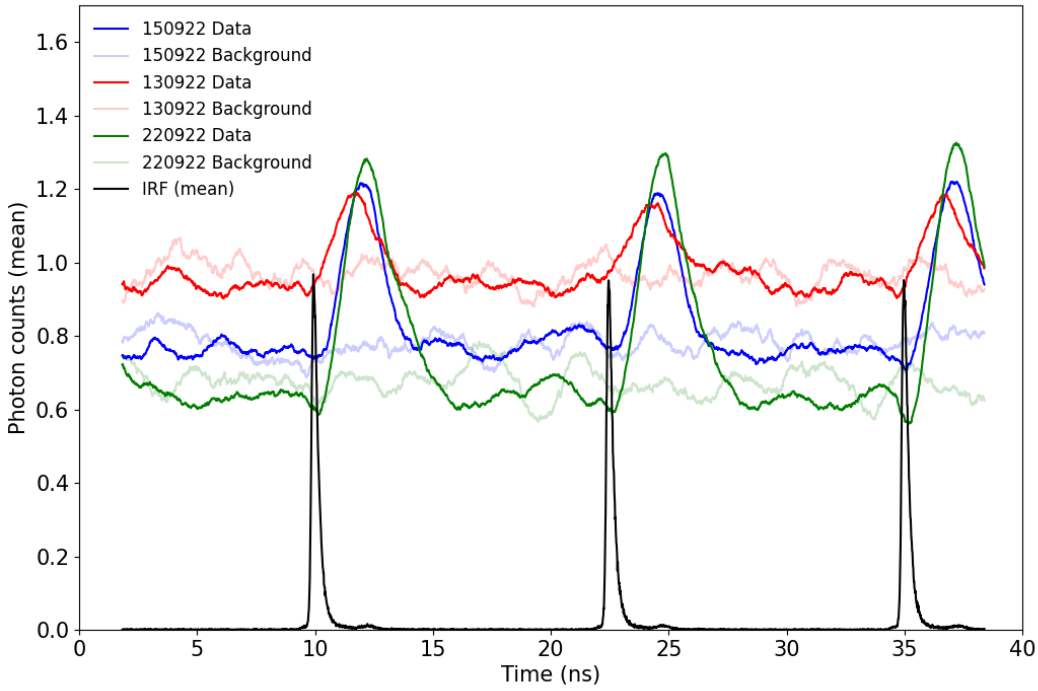


Figure 6.7: The filtered mean photon counts for a two-minute exposure for three separate thirty-minute experiments conducted on different dates using the second experimental setup. The semi-transparent traces are the filtered mean counts of the background measurements. The mean of the normalised impulse response (IRF) can be used to estimate the absolute delay of the photons due to photon diffusion.

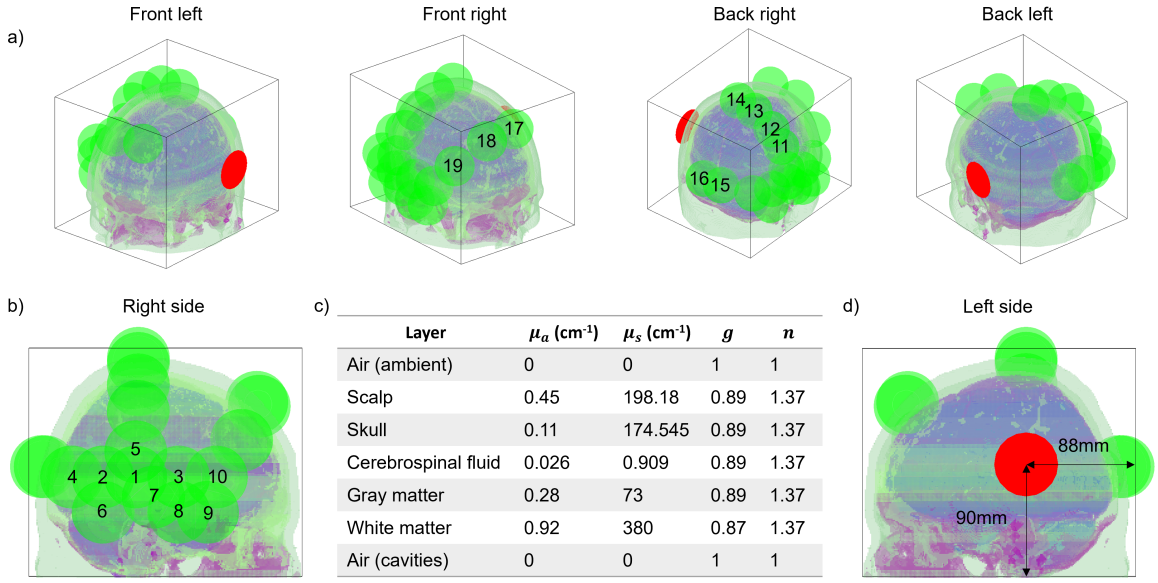
6.3 Simulations of light propagation in the human head

6.3.1 Method

A GPU-accelerated Monte Carlo program (MCX) was developed by Fang and Boas [164] which was motivated by the field of functional near-infrared spectroscopy (fNIRS) to accurately simulate photon propagation through anatomically accurate human head models remeshed from MRI scan data. This program is an ideal tool to estimate the expected signal at the detector for photons that are diametrically transmitted through the head, and can also be used to determine the distribution of light inside the head.

The simulation pipeline of MCX is described in [164]. To summarise, a photon packet is launched from the position of the source in a specified direction towards a 3D volume mesh object. The mesh can be any arbitrary geometry and different segmented volumes within the mesh can be assigned any optical properties. Using the specified scattering coefficient of the current voxel, the distance to the next scattering location is randomly sampled from an exponential distribution, and the azimuth and zenith scattering directions are sampled from a distribution specified by the

Henyey-Greenstein phase function (Eq. (2.20)). The photon packet is stepped one voxel length in the direction of the trajectory, decreasing in weight by the absorption coefficient in the step, until the scattering length is complete. A new scattering length and direction is sampled based on the optical properties of the current voxel. The process is repeated until the photon exits the volume, reaches a minimum weight threshold, or reaches the maximum time-gate. The process described above for one photon packet can then be repeated for a specified number of photons packets.



Layer	μ_a (cm^{-1})	μ_s (cm^{-1})	g	n
Air (ambient)	0	0	1	1
Scalp	0.45	198.18	0.89	1.37
Skull	0.11	174.545	0.89	1.37
Cerebrospinal fluid	0.026	0.909	0.89	1.37
Gray matter	0.28	73	0.89	1.37
White matter	0.92	380	0.87	1.37
Air (cavities)	0	0	1	1

Figure 6.8: a) Various viewing angles of the multi-layered head volume used for MCX Monte Carlo simulations. The green spheres represent detectors with diameter 25mm and the red circle is a uniformly distributed collimated disk source with diameter 25.4mm. b) The right side of the head showing detector labels. Obscured labels are visualised in part (a). c) The optical properties used in the simulations for each layer of the head from [165] for 810nm. The scalp collectively describes superficial tissues and the optical parameters for skin were used. d) The left side of the head noting the position of the source relative to the boundaries of the simulation volume.

Detectors can be placed near the surface of a mesh to collect photons that exit the mesh at a given position. Since the surface geometry of the mesh can be complex, a detector is specified by a sphere of a chosen radius which intersects the surface of the mesh. If a photon packet exits the mesh within the region encompassed by the detector sphere then its information is saved, e.g., direction of propagation, total path length, and final weight value.

The fluence rate can be calculated in space and time for the mesh by tracking the weight of each photon packet through each voxel and creating a distribution in time. The time-bin length can be determined by the user and the final result is normalised by the total number of photon packets such that the calculated fluence rate is in units of 1/mm^{2}s, i.e., the fluence rate for 1J of energy at the input. The full details of normalising the weights of the mesh to determine fluence rate can

be found in [166].

The simulation was defined using the geometry and optical properties outlined in Fig. (6.8). The head model used in the study is a modification of the open source Colin27 segmented MRI phantom [167] that reduces the number of segmented layers. The mesh is a part of the open-source MCX package [168] and the only modifications made for the simulations in this chapter are the optical parameters specified for each segmented volume. Although the optical properties remain uncertain for each tissue type in the head [160], a careful estimate of these values for the Colin27 segmented MRI was determined in [165]. The optical properties specified for 810nm in [165] were used and the source was configured to be a one-inch diameter uniform illumination placed above and behind the ear to emulate the experiments discussed in the previous section. A total of nineteen detectors were configured around the surface of the head, the first ten detectors were placed in close proximity to the position of the detector used in lab experiments.

A total of 647 simulations were run on two NVIDIA RTX 3090 and two NVIDIA RTX 4090 GPUs for 2×10^{10} photon packets, and took around 52hrs computation time. The results from all simulations were aggregated which is equivalent to simulating 1.294×10^{13} photon packets.

6.3.2 Results

A direct comparison of the simulation results and the experimental results obtained in the lab is shown in Fig. (6.9a). The time-resolved photon counts in detectors, shown in Fig. (6.8b) labelled 1–10, were binned into 0.782ns time-bins which was chosen as a multiple of the time-bins used in experiments (12.215ps) to avoid undersampling the distribution. The total weight in each detector was normalised by the total number of input photon packets to determine the attenuation factor, which was then multiplied by the number of photons corresponding to the 800nm 1.2W source used in laboratory experiments. The results in each detector were aggregated to determine the time-resolved photon rate shown in Fig. (6.9a).

The aggregated attenuation factor of detectors 1–10 is 2.8×10^{18} corresponding to around two photons per second for a 1.2W 800nm source. This is the same as the measured photon rate of the laboratory experiments, however, the additional attenuation due to $A\Omega$ product and quantum efficiency are not accounted for. Therefore, the attenuation factor of the experimental results is around five orders of magnitude less than estimated in simulation. Despite a high number of launched photon packets, the number of photon packets reaching each detector with a non-zero weight is low. Using the weight values of a relatively small number of photon packets at the detector to calculate the attenuation value is likely to give an inaccurate estimation.

The full-width half maximum of the Monte Carlo results ($\sim$ 2ns) is comparable to the width of the experimental measurements. Increasing the bin size of a histogram containing few samples better approximates the true population distribution [169] and therefore this result is likely to

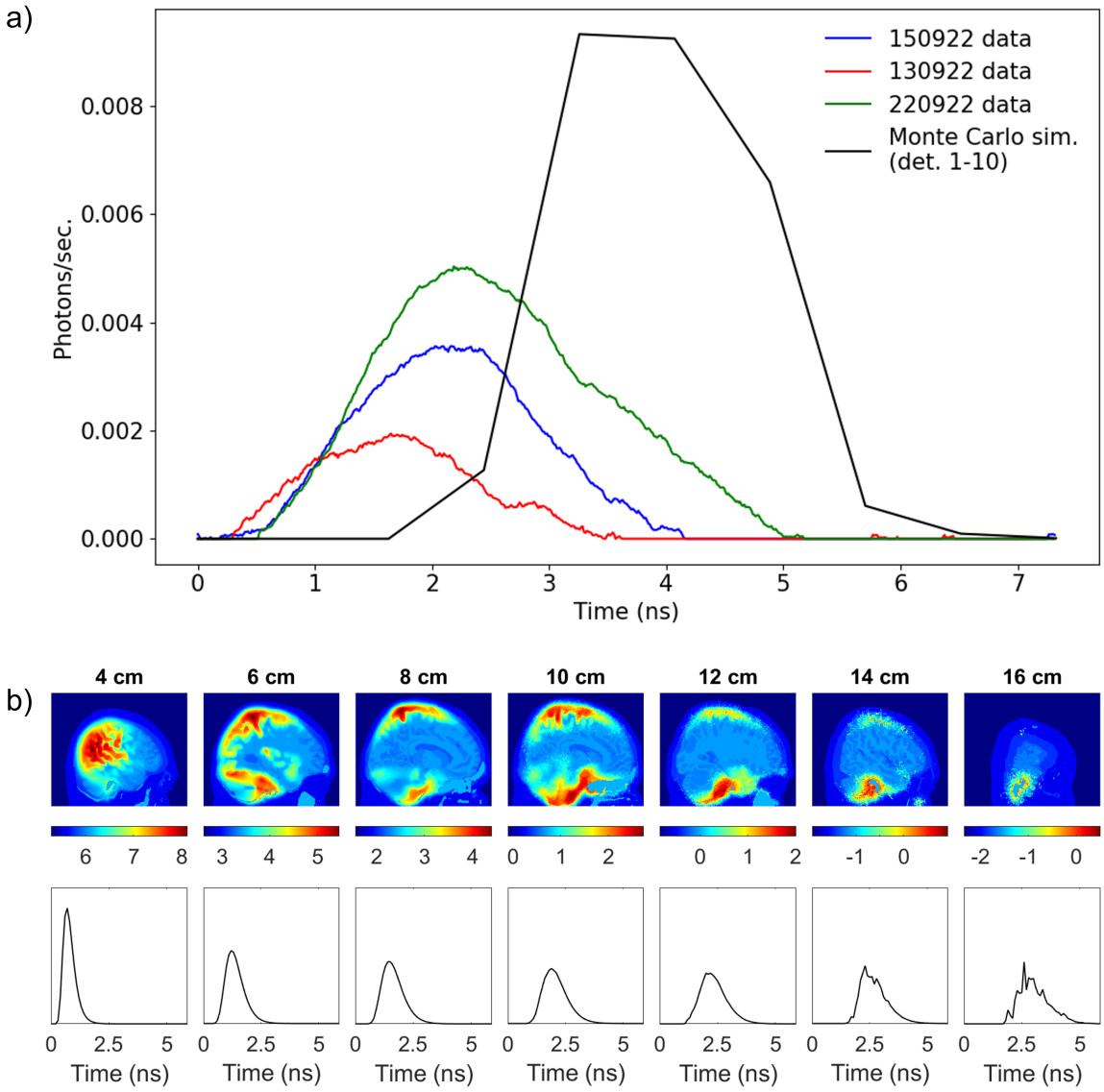


Figure 6.9: a) The background-subtracted and filtered mean of the experimental data normalised to photons per second and the sum of the Monte Carlo simulated time-resolved photon counts at detectors 1–10 (Fig. 6.8b)). b) Sagittal planes of the segmented head mesh overlayed with the energy-normalised fluence rate ($1/m^2s$) plotted in log-10 scale (top) and the corresponding normalised time distribution (bottom).

be accurate despite detecting a relatively small number of samples. It can also be assumed that the simulated distribution is unimodal given the trend of the normalised time-resolved fluence rate for sagittal planes of the head for increasing thickness tending towards the position of the detectors as shown in Fig. (6.9b). This trend also shows an increasing width of the time-resolved distribution towards 2ns.

The peak-to-peak difference of the simulated and experimental distributions is around 1.8ns.

This difference can partially be accounted for by the difference in head diameter. The measured head diameter of the participant in experimental work is 15.5cm, whereas, the corresponding source/detector distance in the Colin27 mesh is 16.9cm. A first-order approximation to the delay shift of a 200ps difference can be estimated using the peaks of the 4cm (0.7ns) and 16cm (2.6ns) traces in Fig. (6.9b) to determine a rate of delay with respect to thickness.

Another sensitive parameter that can lead to a difference in the delay is the source placement. The spatial distribution of light through the head shown in Fig. (6.9b) highlights two main pathways for light to transmit though the head. Guided by the channels of cerebrospinal fluid (CSF) surrounding the brain, light travels above or below the cerebral cortex and most likely exits the head from the region of porous bone of the skull behind the ear. This is illustrated more clearly when visualising different plane dimensions scanned in time in Fig. (6.10). Because of the high extinction properties of the brain, the distribution of light migrates from the source position above and below the cerebral cortex over time. This is an effective light guiding mechanism through the regions of low scattering and absorption.

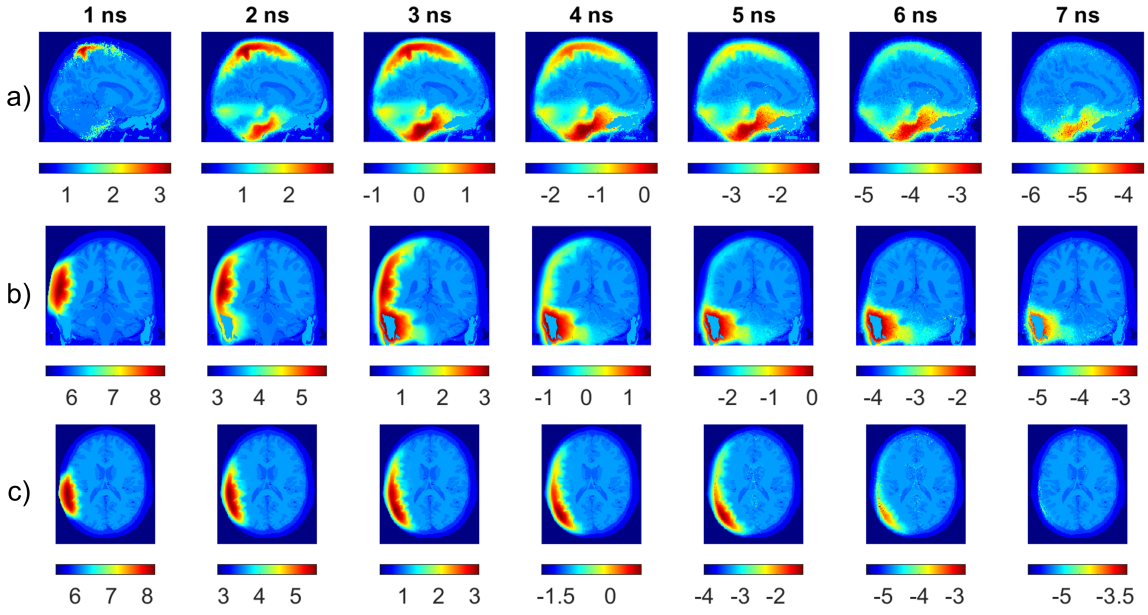


Figure 6.10: The log-10 energy-normalised fluence rate ($1/m^2s$) for increasing time at a) the sagittal plane 80mm from the source, b) the coronal plane 88mm from the back of the simulation volume, and c) the transverse plane 90mm from the bottom of the simulation volume.

Interestingly, the mechanism of light guiding in this context is not due to conventional reflective waveguides or total internal reflections due to a difference in refractive index in the material, but a statistical process whereby light preferentially migrates in regions of low extinction. The phenomenon of light guiding in CSF is an active research topic in fNRIS [170–172] which originated at the turn of the century [173–175] to improve the accuracy of the inverse retrieval models for estimation of regions of blood oxygenation in the brain. However, there has yet to be any rigorous

explanation of the underlying physics of the mechanism in diffusive media. A study by Kienle and Hibst [176] discusses guiding of light in arbitrary directions by a microscopic tubule structure of dentin. In this study, the authors show that the light guiding mechanism is not due to reflections or wave effects but the preferential scattering direction of aligned cylindrical microstructure of the material. Although the light guiding in the Monte Carlo simulations presented in this chapter do not have aligned structures on the microscopic scale, this study hints that scattering materials can act as light guides.

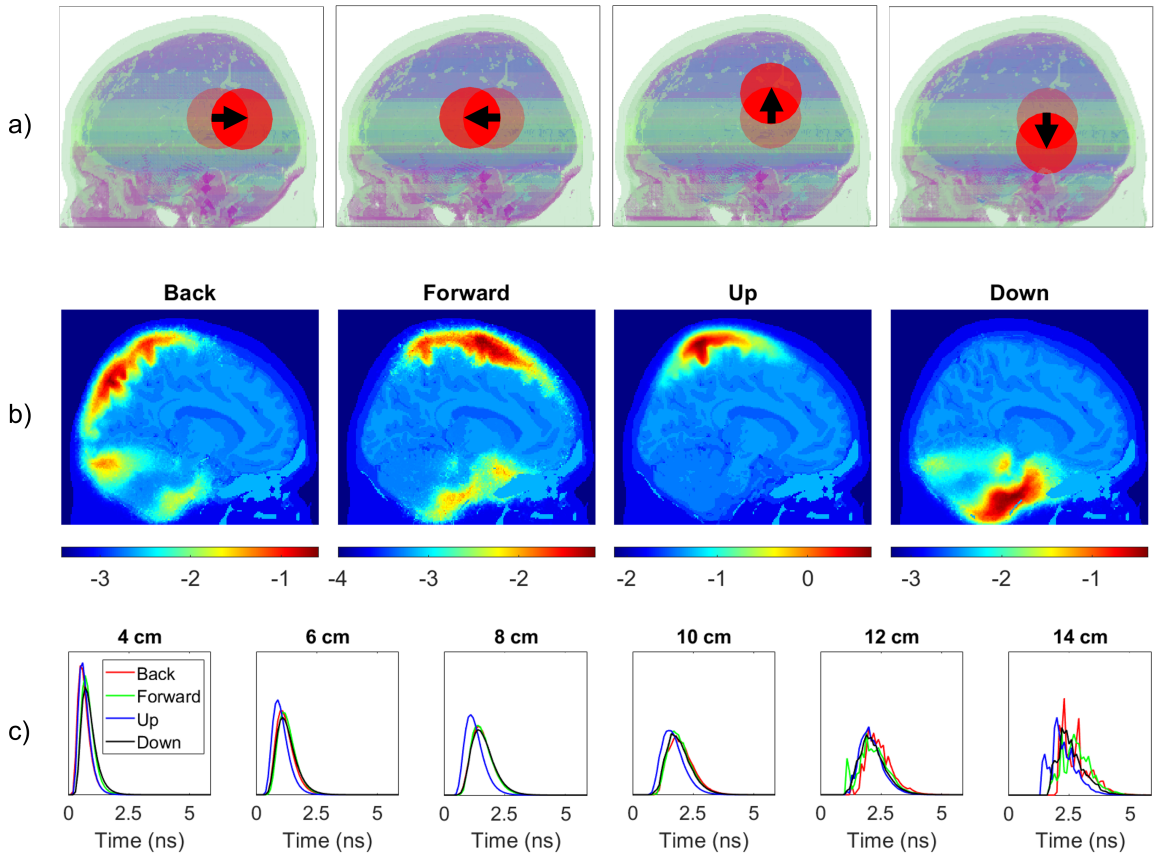


Figure 6.11: a) An illustration showing the exact displacement of the source compared with the initial position in Fig. (6.8d). The modified source position is shifted by 2cm in the direction of the arrow. b) The corresponding normalised fluence rate ($1/m^2s$) distribution in log-10 scale at the sagittal plane of the head 8cm from the source. c) A comparison if the time distribution at different sagittal planes coloured corresponding to the source position that was used.

There is clearly also scope for more studies into the guiding of light through heterogeneous models which also include the microscopic structure of tissues to more accurately simulate light transport through biological tissues beyond the assumptions of made in the diffusion approximation to the radiative transfer equation. For example, the fibrous structure of white matter could preferentially guide light through the brain which may be overlooked in current Monte Carlo photon transport models that are designed with applications that study the outer layer of gray matter. Careful consideration of the location of air cavities in the interior structure of skull

and regions of CSF could also be used to preferentially guide light to target areas by configuring source placement for photo-dynamic therapy applications [165] or to sample functional and anatomical information by detecting photons which have explored deep regions of the head.

To estimate the sensitivity of source placement for guiding light through different regions of the head, four additional Monte Carlo simulations were performed using the same configuration outlined in Fig. (6.8) using a source position which was translated 2cm above, below, backward, and forward from the location used in the previous larger simulation. The goal of this simulation is not to analyse the transmitted photons at the opposite side of the head but to visualise the preferential guiding of light due to scattering to different volumes in the head. Therefore, a reduced number of 2×10^{11} initial photon packets was used to reduce the simulation time (a total of 4hrs computation time using a NVIDIA RTX 4090).

The results shown for a sagittal plane midway through the head (8cm from the source) in Fig. (6.11b) indicate that the explored volume of the light dramatically varies depending on source position. Moving the source up or down can selectively isolate the path of migration around the top of the head in the CSF layer between brain and skull, or through the lower regions of the brain in CSF surrounding the cerebellum. Furthermore, the normalised time-resolved fluence rate shown in Fig. (6.11c) for different sagittal planes through the head show that displacing the source can have an effect on the time-delay of the photon distribution. A peak-to-peak time difference of 300ps was observed midway through the head (8cm from the source) between simulations that have a source 2cm above and below the original placement.

A time difference between the top and bottom guided paths of 220ps was also observed for the same plane midway through the head in the results of the original larger simulation. The cropped regions shown in Fig. (6.12) clearly show this difference even though the source is fixed to one position. This indicates that it is not necessarily the difference in source position, but the difference in guided photon paths which causes a difference in the delay and width of the time distribution.

The sensitivity of light coupling due to different source placement was also observed experimentally by illuminating a human skull in the lab with a 20mW 532nm laser as shown in Fig. (6.13). The light diffused by the illumination side of the skull projects onto the interior walls and the transilluminated light was photographed using a digital camera. There is clear difference in density and thickness of the side wall of the skull which have drastically different transmission properties. Light is also more intense close to the coronal suture at the top of the skull that could be used to indicate regions where light may be better coupled in and out of the brain.

Although the effects outlined above highlight the sensitivity to source position to the time-of-flight distribution of photons in the head, this still does not account for the delay between the experimental and simulated results. However, the differences between experimental results and

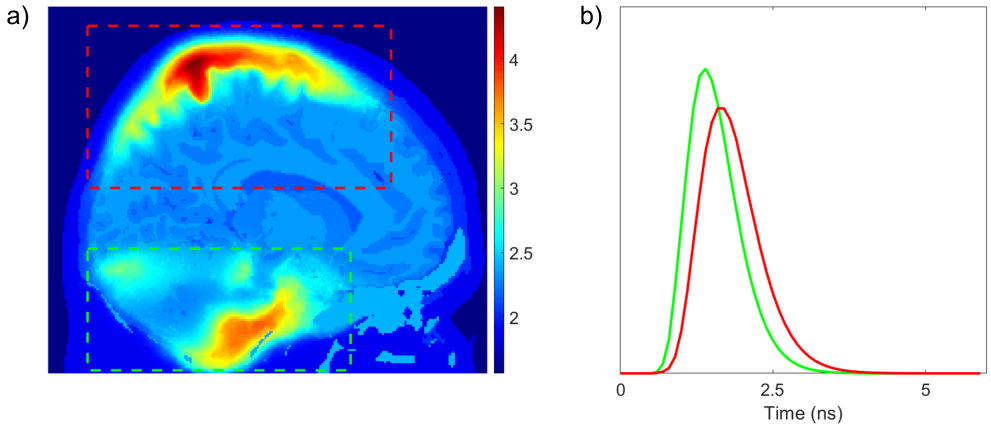


Figure 6.12: a) The normalised fluence rate ($1/m^2s$) distribution in log-10 scale at the sagittal plane of the head 8cm from the source. This result is from the simulation outlined in Fig. (6.8) using 1.294×10^{13} photon packets. b) The time distribution of the fluence rate of the cropped regions coloured to indicate the upper (red) region and the lower (green) region.

simulations due to the cerebrospinal fluid guiding mechanism throughout the head is still an active field of research for fNIRS. The differences studied in fNIRS are over relatively short distances which are likely to be amplified when considering the length scales of propagation through an entire head studied in this chapter.

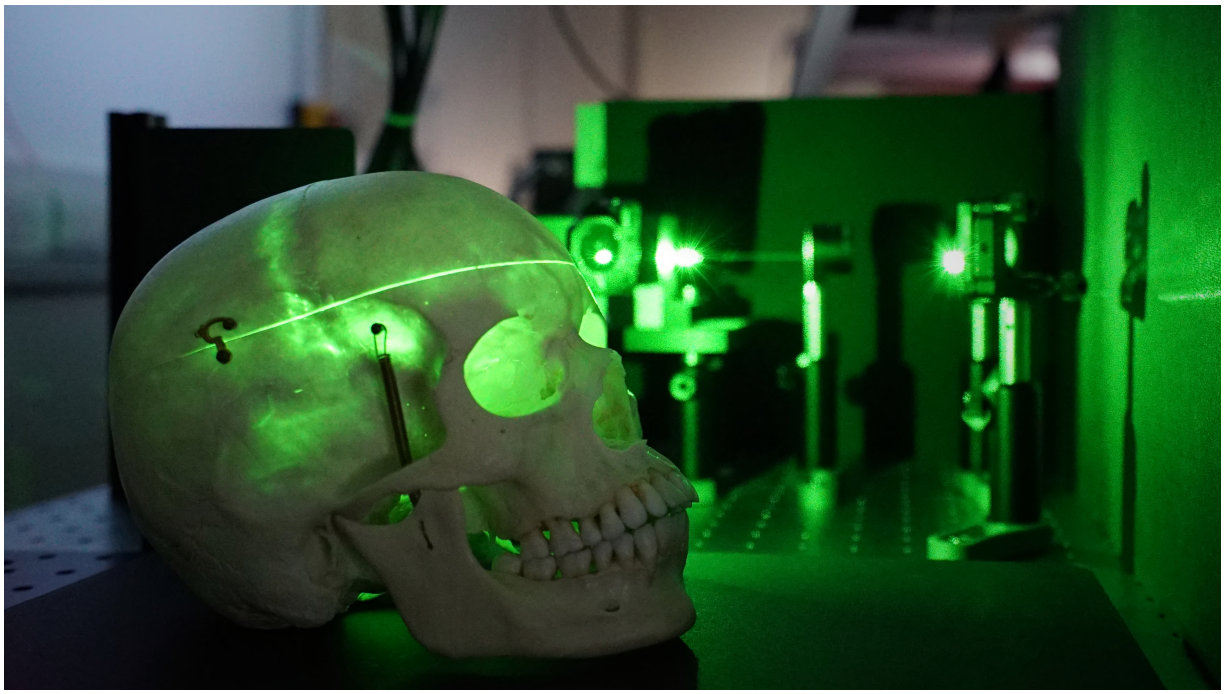


Figure 6.13: A photograph of a transilluminated human skull using a 20mW 532nm laser demonstrating regions of bone with higher transmissivity.

Furthermore, the optical properties of the human head *in-vivo* are still not accurately known and will have inter-subject variability e.g., hair thickness, hair colour, and skull density [177]. For

example, the reported reduced scattering coefficient of the human brain differ by around 100% [25, 160], and in a more recent study, inter-subject absorption and reduced scattering coefficient was shown to differ by around 50% [160]. Using different optical parameters would have a significant effect of the time distribution of transmitted light over such large distances.

Lastly, the microscopic structures of different tissue layers in the brain (e.g., fibrous white matter structure) could be unaccounted for by the diffusion approximation to the radiative transfer equation, which may cause light to be guided in other tissues other than the CSF, such as white matter, that could also lead to differences in the time distribution of transmitted photons for these length scales.

6.4 Conclusion

Transmitting photons through the head is a challenging, yet relevant, unsolved biomedical optics problem. In this chapter, the various experimental approaches were discussed, which uncovered important considerations when maximising the collection of photons in the extreme limits of the highly diffuse regime. Since there is no reference for the precise features that the signal transmitted through the head should have, an estimate of the signal was simulated using a detailed anatomically accurate multi-layer model of the human head.

From the analysis presented in this chapter it is inconclusive whether the detected light in experiments was transmitted through the head. However, the time distribution is not entirely different to the Monte Carlo simulated signal considering the difference in head diameter, sensitivity to source placement and potential inaccuracies of the model over large distances. It is promising that light can be guided to different regions of the brain in simulation, which could be exploited to selectively isolate regions of interest in deep brain regions currently unexplored in the field of fNIRS. Furthermore, it was shown in simulation that photons transmitted diametrically through the head can be collected by the detectors at a rate of around two photons per second which indicates that it is not unfathomable to have sensitivity to deep regions of the brain using near-IR wavelengths. Differential measurements using different source placement and detector could lead to isolating information from different deep volumes of the brain which may unlock the potential of diffuse optical imaging to reconstruct anatomical or functional information in currently inaccessible regions.

The work of this chapter indicates some future avenues of research towards enabling deep diffuse optical imaging of the human brain. The first of which is experimental considerations to optimise collection efficiency by maximising the etendue of the detector. The ideal sensor has a large solid angle of collection and area covering the entire side of the head. Hypothetically, this could be realised with 100% fill factor large area detectors that are placed directly against the skin and have sensitivity to photon detection at high angles of incidence. Multiple of these sensors could

be arranged on the interior of a highly absorbing wearable cap which would extend the total area of collection without requiring collection optics that are constrained by the conservation of etendue. A wearable design in close contact with the skin is also easier to isolate from background light entering the system. If such a wide-field modular detector also had low noise and high quantum efficiency in the near-IR, towards the ideal specifications of superconducting nanowire single-photon detectors, this would also dramatically increase the collection efficiency of diffuse photons. The next evolution of the experimental prototype is currently under construction. The design uses a black rubber-lined bicycle helmet with a PMT detector directly attached. Using 3D-printed mounts, the PMT can be brought into proximity to the skin sealed with a silicone mould that conforms to the shape of the head and prevents background light from entering the detector.

The challenge of determining whether the detected signal truly propagated through the head could be tackled by changing the experimental paradigm to gradually move the source and detector further apart from one another around the circumference of the scalp. These measurements would help determine the change of signal features as the separation increases so that the numerical model could be validated using experimental signals with high SNR. Understanding how the signal changes with increased propagation distance and tuning the numerical model with real-world data would increase the accuracy of extrapolation of these signals for photon propagation distances through the entire diameter of the head. Furthermore, knowledge of how the signal changes through short to medium distances through the head could also provide a more accurate framework for explaining the origin of experimentally measured signals.

There is also scope for studies into the mechanism of light guiding by scattering which could uncover more accurate descriptions of light-matter interaction, especially for tissues such as white matter which have microscopic fibrous structure. Incorporating this into Monte Carlo light transport models will be critical for developing techniques that deliver and collect photons in deep brain regions.

There are also some recent studies showing the potential of using longer wavelengths such as 1300nm and 1750nm that have favourable scattering and absorption coefficients for probing deep regions of the brain [178–180]. The different optical properties of the chromophores in the head and brain tissues at these wavelengths may also present opportunities to monitor functional processes other than blood oxygenation. The difference in optical coefficients may also lead to different guiding pathways through the head compared with 800nm light studied in this chapter. Multispectral approaches may provide complimentary independent information about anatomy and physiological processes in the brain.

Chapter 7

Conclusion

Diffuse optical imaging beyond $100\ell^*$ is an important step towards the development of non-invasive medical imaging technologies that are low-cost, compact, easily portable, and do not use harmful ionising radiation. This could enable clinical applications currently restricted by contraindications of alternative medical imaging technologies and present new opportunities for medical and neuropsychology research.

The physical problem of diffuse optical imaging to overcome is the process of photon scattering that randomises the direction of light propagation and degrades direct imaging information. Time-resolved methods such as early photon imaging can recover images through scattering materials using the first-arriving photons that maintain some spatial coherence. Unfortunately, the detection of these photons is exponentially suppressed with an increasing number of transport mean free path lengths. Therefore, early photon imaging methods fail in the highly diffuse regime beyond $80\ell^*$ where the detection rate of a ballistic photon is roughly one per age of the universe. However, it was shown in Section (3.4) that indirect imaging information remains in the full spatiotemporal distribution of detected photons, which enables image reconstructions of embedded absorbing objects in highly diffusive materials.

The information content of measurements was quantified using Shannon entropy and mutual information in Chapter 4. First a case study was presented using information theory concepts to analyse multipath echo signals used to reconstruct three-dimensional scenes. The analysis uncovered a surprising insight that return echoes from ray paths which have bounced twice around the environment contain more information than direct reflections. This can intuitively be interpreted as a "hall of mirrors" effect where including signals that have scattered multiple times contain information about the scene from many different perspectives.

Information theory analysis concepts were then applied to diffuse photon measurements. The results showed that imaging information can be measured beyond $200\ell^*$ for simulated measure-

ments including practical experimental conditions, which corresponds to more than 10cm of human tissue. Surprisingly, the estimated limit of detecting mutual information was not due to the physical process of photon diffusion but the exponential extinction rate of transmitted photons. Experimental results show that this limit can extend further by increasing collection efficiency or using higher source power.

Furthermore, it is clear from inspection of the spatial maps of mutual information in Section (4.4) that information about the input is preserved for regions with low cross-talk between neighbouring regions. Therefore, information throughput can be enhanced by minimising the cross-talk of the signals due to spatial broadening of the pulse after the object plane, e.g., by using a scanning source rather than uniformly illuminating the material.

It was also shown that resolving in both space and time improved the condition of the ill-posed problem. Resolving measurements in space-time decreases the sensitivity of the inverse problem to noise perturbations and is more likely to produce accurate image reconstructions. The extra dimension of the data helps to diversify the measurements and provide additional independent information.

The problem of imaging thorough $> 100\ell^*$ can be broken down into maximising the information collection and designing inverse image reconstruction algorithms to utilise all available collected information. A state-of-the-art machine learning algorithm for computational imaging inverse problems [13] was introduced in Chapter 5, which harnesses the capacity of neural networks to learn arbitrarily complex non-linear mappings between measurements and targets, and employs a probabilistic framework that can be used to assess the diversity of predictions for critically ill-posed inverse problems.

The machine learning framework designed by the co-authors of [13] was demonstrated using a simple optics experiment with a challenging ill-posed inverse image retrieval problem. Using only a limited number of experimentally acquired measurements and a low-fidelity simulation of the system, a multi-fidelity forward model was built using a conditional variational autoencoder (CVAE) to generate a large training set that could be used to train an inverse image retrieval model. A probabilistic inverse model was then constructed, also using CVAE, to accurately reconstruct the amplitude image from critically down-sampled and saturated experimental measurements such that a conventional holographic phase retrieval algorithm failed to converge on a solution. An advantage of using a probabilistic inverse model is the access to alternative plausible solutions which were shown to become more diverse when further degrading the quality of the measurements.

To assess the potential of a probabilistic machine learning model to reconstruct images of hidden objects embedded in highly diffusive materials, a CVAE inverse model was trained and tested with ideal diffuse photon measurements simulated through more than $100\ell^*$. The proposed ex-

perimental configuration was guided by the results in Chapter 3 and 4, which incorporated a raster scanning source and a detector with high collection efficiency to maximise information throughput. The results verify that there is enough information contained in measurements up to $242\ell^*$ to qualitatively reconstruct images. Estimates of other possible solutions were also predicted which accurately reflected the increased instability of the ill-conditioned problem for measurements through a greater number of transport mean free path lengths. Quantifying the variability in reconstructed images is critical for some real-world applications, e.g., medical imaging of tumours.

The inverse CVAE model trained on ideal simulations was then tested on experimental data and the predicted reconstructions contained resembling features of the target images. Future work is to incorporate a multi-fidelity approach using more accurate simulations that reflect the experimental conditions and produce a training data distribution more representative of the true distribution of experimental measurements.

As a real-world experiment at the extreme limits of the highly diffusive regime with relevance to medical and neuroscience applications, experiments conducted towards transmitting photons through an adult human head were presented in Chapter 6. The evolution of experiments uncovered key considerations of collection efficiency of diffuse light, such as maximising etendue and quantum efficiency whilst minimising the detection of background light and dark count rate. These trade-offs are considered in the most recent experimental configuration, and although promising signals of diffuse light have been measured, future work is to reliably reproduce these signals to investigate further if the detected photons have been transmitted through the head.

Simulations were performed using a Monte Carlo photon propagation solver to estimate the trajectory of photons inside the head using an anatomically correct mesh. Comparisons with the experimental data show that there are similarities such as the width of the distribution, however, there is not enough evidence to conclude if the detected light propagated through the head. Findings show that regions of cerebral spinal fluid guide light around the cerebral cortex have different temporal propagation statistics and can be coupled preferentially by changing the source location. The underlying mechanism of light guiding in the head is also an interesting topic to be explored in more detail. Understanding the principles of light guiding in the head could lead to optimised experimental configurations that couple into guiding pathways that efficiently deliver to deep regions of the brain, or optimised sensor locations that detect light that has interacted with deep regions.

In this thesis, it is clear that there is promising potential for using time-resolved diffuse optical imaging to reconstruct images deep inside regions of the human body and brain. However, to enable these technologies to image beyond $100\ell^*$, the focus of research must migrate from early photon imaging concepts to utilising the full available information of the detected photons.

The potential for diffuse optical imaging to be extended to deep regions of the brain, as explored in this thesis, could inspire the development of novel functional imaging techniques and applications e.g., incorporating signals from cortical folds to provide complimentary information about brain activity with cortical surface signals measured using existing fNIRS. Furthermore, the current progress of creating wearable time-domain multi-spectral fNIRS instruments enables imaging activity from other interesting physiological processes such as directly measuring metabolic activity using the spectral properties of cytochrome-c-oxidase. This modality extends the clinical applications of fNIRS to monitoring stroke recovery or the progression of neurological diseases [177, 181].

Beyond functional imaging, extending the depth of sensitivity of diffuse optical imaging could open opportunities for anatomical imaging anywhere in the body if the sensitivity of the imaging algorithms to accurately determine subtle differences in absorption coefficients is improved beyond the binary absorption samples in this thesis. Additionally, extending the work of this thesis to depth-resolved reconstructions of absorbing objects e.g., extending the machine learning algorithm in Section (5.2) to use 3D convolutional layers in the Decoder network such that the model maps from diffuse photon measurements to 3D mesh structures from simulation or MRI scans, could realise depth-resolved imaging of anatomical structures in deep tissue regions.

The future of diffuse optical imaging is likely a combination of these ideas which harnesses advances in time-resolved multi-spectral single-photon detector hardware to further increase the diversity of information captured at the detector by simultaneously sampling different spectral regions and resolving functional and anatomical information in space and depth.

Information can also be enhanced by resolving measurements in other domains and considering the experimental configuration, such as collection efficiency and source and detector placement. Beyond resolving photon counts in space, space-time, and multi-spectral domains, other plenoptic information could further increase the independent information content of the measurement, e.g., using a lenslet array to measure the angle of incidence of photons at the detector. Furthermore, imaging retrieval algorithms that consider the instability of the critically ill-posed problem of imaging in the highly diffusive regime are key to estimate the uncertainties of the reconstructions, which is especially important for medical imaging applications. However, in principle, other imaging applications currently limited by photon diffusion, such as using diffuse X-ray photons to enhance X-ray image quality, or imaging through fog and water, can apply the same conclusions of this thesis.

Appendix A

The derivative of the regularised least-squares algorithm

The regularised least-squares inverse retrieval algorithm developed by Lyons et al. [12] can be formulated as a minimisation of the operator $O(m)$, where the argument m is the hidden two-dimensional absorbing object we wish to reconstruct:

$$\arg\min_{m}\{O(m)\} = \arg\min_{m}\left\{ ||\mathcal{A}(m) - Y||_2^2 + \lambda_1 ||m||_{TV} + \lambda_2 ||m||_1 \right\}. \tag{A.1}$$

The regularisers $||m||_{TV}$ and $||m||_1$ are the total variation norm and L_1-norm regularisers weighted by the regularisation parameters λ_1 and λ_2.

The gradient $O'(m)$ of the loss function $O(m)$ defined in Eq.(A.1) can be found by considering the derivative of each term separately. The least-squares fidelity term can be found using the chain rule:

$$\frac{d}{dm}||\mathcal{A}(m) - Y||_2^2 = \sum_{t\in T} \mathcal{A}^*(\mathcal{A}(m) - Y), \tag{A.2}$$

where $\mathcal{A}^*(m)$ is the adjoint operator of Eq. (3.3). The summation of time-bins t for the full set of time-bins T in the histograms results in a two-dimensional gradient map that is used to update each pixel independently in the image estimate.

The forward operator photon diffusion model described in Eq.(3.3) can be rewritten using the convolution theorem:

$$\mathcal{A}(m) = \mathrm{Re}\left[\mathcal{F}^{-1}\left\{\mathcal{F}\left\{\Phi(\vec{r}_{out}, t)\right\}\mathcal{F}\left\{\phi_{\mathrm{m}} \odot m(x,y)_T\right\}\right\}\right]. \tag{A.3}$$

The term ϕ_{m} is the result of the convolution of the laser intensity profile $I(x,y,t)$ with the point

spread function of the fluence rate at the object plane $\Phi(\vec{r}_{\mathrm{m}}, t)$:

$$\phi_{\mathrm{m}} = \mathrm{Re}\left[\mathcal{F}^{-1}\left\{\mathcal{F}\left\{I(x,y,t)\right\}\mathcal{F}\left\{\Phi(\vec{r}_{m},t)\right\}\right\}\right]. \tag{A.4}$$

To find the adjoint operator in terms of Fourier transforms, the conjugate is found for each operator in Eq.(A.3) acting on the argument m in reverse order from the last operation to the first operation. Starting with the conjugate of the final operation in the forward model (an inverse Fourier transform), and ending with the first operation (multiplication by ϕ_{m}), the adjoint operator $\mathcal{A}^{*}(m)$ can be found by,

1. Conjugate of last operation: $\mathcal{F}\{m\}$,
2. Conjugate of last two operations: $\mathcal{F}^{*}\{\Phi(\vec{r}_{\mathrm{m}},t)\}\mathcal{F}\{m\}$,
3. Conjugate of last three operations: $\mathcal{F}^{-1}\{\mathcal{F}^{*}\{\Phi(\vec{r}_{\mathrm{m}},t)\}\mathcal{F}\{m\}\}$,
4. Conjugate of all operations: $\phi^{*}_{\mathrm{m}}\mathcal{F}^{-1}\{\mathcal{F}^{*}\{\Phi(\vec{r}_{\mathrm{m}},t)\}\mathcal{F}\{m\}\}$,

$$\mathcal{A}^{*}(m) = \mathrm{Re}\left[\phi^{*}_{\mathrm{m}}\mathcal{F}^{-1}\{\mathcal{F}^{*}\{\Phi(\vec{r}_{\mathrm{m}},t)\}\mathcal{F}\{m\}\}\right], \tag{A.5}$$

Where $\mathcal{F}^{*}$ denotes the conjugate results of the Fourier transform, rather than the conjugate of the Fourier transform itself, which is the inverse Fourier transform.

The second term of the gradient of the loss function is the derivative of the TV-norm [182]:

$$\nabla||m||_{TV} = -\nabla\cdot\left[\frac{\nabla m(x,y)}{|\nabla m(x,y)|}\right]. \tag{A.6}$$

Finally, the gradient of the L_1-norm is just the sign (positive or negative) of the pixel values:

$$\nabla||m||_1 = \mathrm{sign}(m). \tag{A.7}$$

Appendix B

CVAE model architecture: DOI beyond $100\ell^*$

The inverse conditional variational autoencoder (CVAE) model architecture for diffuse optical imaging beyond 100 transport mean free paths experiment described in section 5.2 is explained in more detail in this appendix.

Figure (B.1) shows a high-level block diagram of the inputs and outputs of the CVAE for training and testing. To understand the motivation between the connections and outputs of the model as a whole, consider the objective function [13]:

$$\underset{\phi,\theta_1,\theta_2}{\arg\max} \log p_{\theta_2}(x|z,y) - D_{KL}(q_\phi(z|x,y)||p_{\theta_1}(z|y)), \tag{B.1}$$

where the minimisation is performed with respect to the learn-able parameters of Encoder 1 (ϕ), Encoder 2 (θ_1), and the Decoder (θ_2). The first term of Eq. (B.1) ensures the output reconstruction is accurate by maximising the log-likelihood of predictions. Assuming that the output distributions for each pixel are Gaussian, then this can be realised by minimising mean squared error between the prediction. The second term of Eq. (B.1) can intuitively be interpreted as minimising the difference between the distributions modeled by the first and second encoders. Each encoder maps to the moments of a 10-dimensional multivariate Gaussian (i.e., multivariate normal) and so the KL-divergence can be explicitly defined. The objective function in Eq. (B.1) can be rewritten in terms of inputs and outputs of the model as

$$\underset{\phi,\theta_1,\theta_2}{\arg\min} \frac{1}{N}\sum_{i=1}^{N}(\hat{\mathbf{x}}^{(i)} - \mathbf{x})^2 + D_{KL}(\mathcal{N}(\boldsymbol{\mu_{E1}}, \boldsymbol{\sigma_{E1}})||\mathcal{N}(\boldsymbol{\mu_{E2}}, \boldsymbol{\sigma_{E2}})), \tag{B.2}$$

where $\hat{\mathbf{x}}^{(i)}$ is the i^{th} prediction from N samples from the latent distribution (for this case $N = 1$),

x is the target image, and $\mathcal{N}(\boldsymbol{\mu_{E1}}, \boldsymbol{\sigma_{E1}})$ and $\mathcal{N}(\boldsymbol{\mu_{E2}}, \boldsymbol{\sigma_{E2}})$ are multivariate normal distributions parameterised by the predicted moments of Encoder 1 (E1) and Encoder 2 (E2).

When training Fig. (B.1a), all three networks are used to minimise the objective function as per Eq. (B.2) and the samples are taken using the output moments of Encoder 1. However, when testing, Encoder 1 is redundant and samples are taken using the moments from Encoder 2.

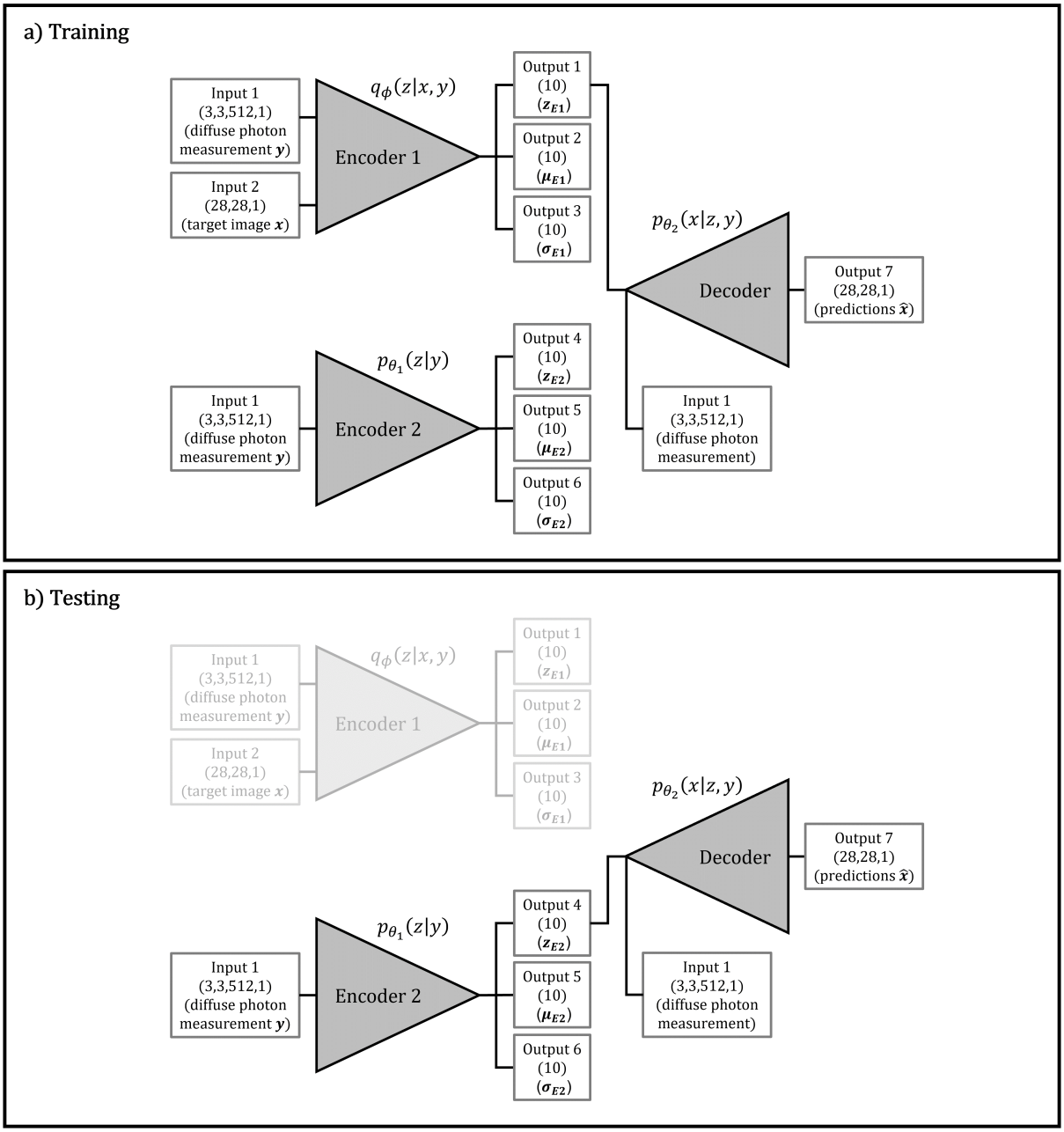


Figure B.1: The high-level block diagram of the CVAE architecture when a) training and b) testing. The shapes of the inputs/outputs are displayed under the variable names.

The layers and connections used in each of the neural networks in Fig. (B.1a) is illustrated in Fig. (B.2). Encoder 1 (Fig. (B.2a)), the measurements input to the model are spatially resolved

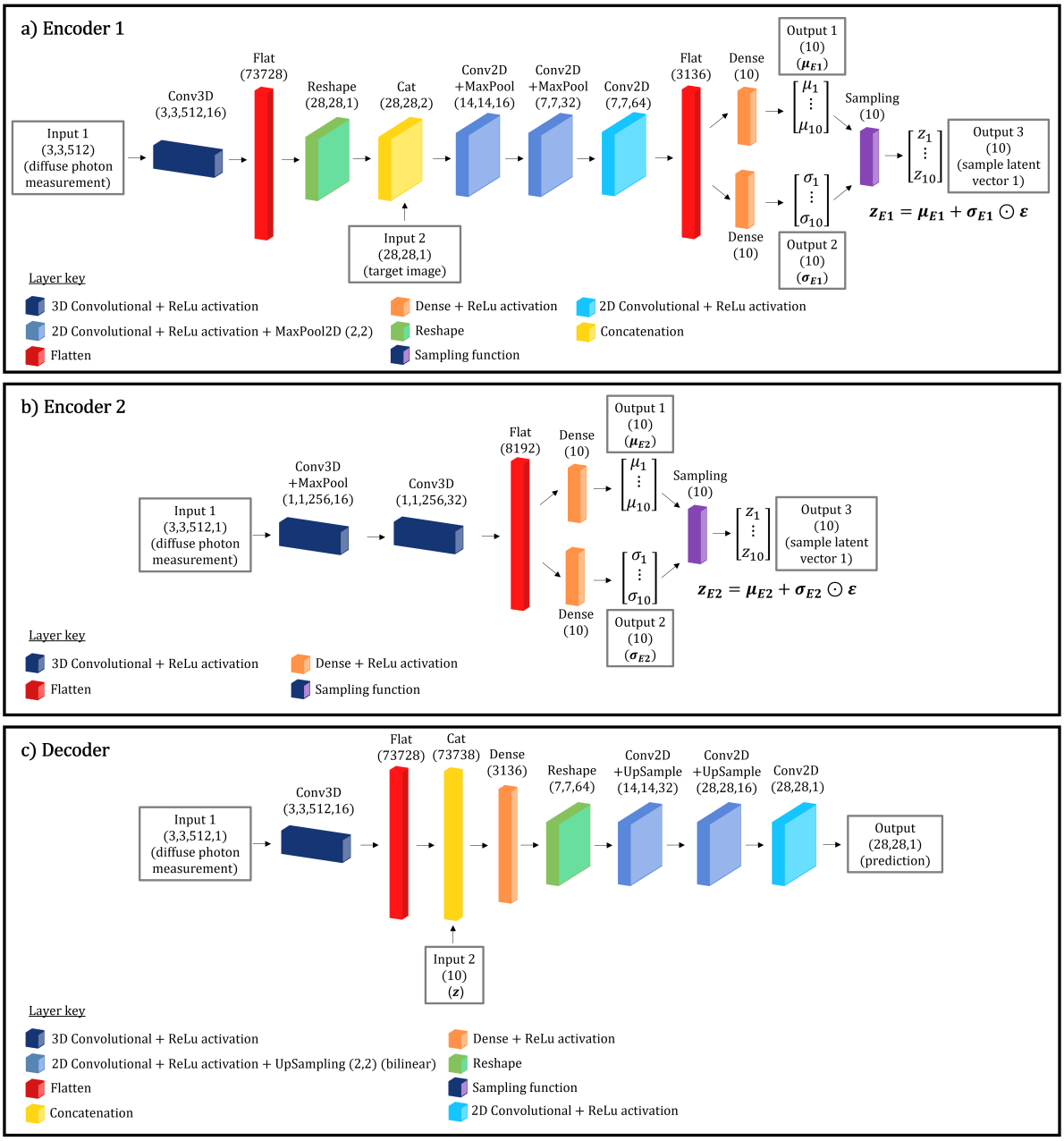


Figure B.2: The layers and connections used for a) Encoder 1, b) Encoder 2, and c) Decoder presented in Fig. (B.1a). Each layer type is represented as a colour block which is described by the key at the bottom of the figure. The output shape of each layer is displayed under the layer name and the inputs/outputs are highlighted using a box outline. Sampling the latent distribution is performed using $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\varepsilon}$, where $\odot$ is an element-wise product and $\boldsymbol{\varepsilon} \sim U(0,1)$ is a random number between zero and one from a uniform distribution.

time-of-flight diffuse photon measurements $\mathbf{x}$, therefore 3-dimensional convolutional filters (kernel size (3,3,3)) are used to extract local spatial and temporal features. The resulting feature maps are then reshaped to match the dimensions of the target images $\mathbf{y}$ followed by 2D convolutional layers (kernel size (3,3)) which extract features from both the measurement and the target image. The feature maps are successively down-sampled using max pooling until the features are flat-

tened to produce a column vector which is input to fully-connected dense layers that predict the moments of the latent distribution.

Similarly Encoder 2 (Fig. (B.2b)) uses 3-dimensional convolutional layers for feature extraction of the diffuse photon measurements $\mathbf{y}$ and dense layers to predict the moments of the latent distribution.

The Decoder (Fig. (B.2c)) has both the diffuse photon measurement $\mathbf{x}$ and the sampled latent vector $\mathbf{z}$ as input. Again the features of the measurements are extracted using a 3-dimensional convolutional layer and the resulting feature map is concatenated with the latent vector $\mathbf{z}$. To produce a predicted image at the output of the decoder, the feature vector is reshaped to 2-dimensions and consecutively up-sampled to match the shape of the target images $\mathbf{y}$. The use of 2-dimensional convolutional layers in this context is to increase the number of learn-able parameters to map between the low-dimensional reshaped feature maps ((7,7,64) i.e., 64 feature maps with shape 7×7) to the high dimensional output shape (28,28,1).

Encoder 1 has 57,890,164 trainable parameters, Encoder 2 has 178,164 trainable parameters, and the Decoder has 231,269,185 trainable parameters. The total number of trainable parameters for the combined networks in Fig.(B.1a) is 289,337,513.

The model was trained using an Adam optimiser for 1000 epochs using a mini-batch size of 200.

Bibliography

[1] K. M. Yoo, F. Liu, and R. R. Alfano, "When does the diffusion approximation fail to describe photon transport in random media?," *Physical Review Letters*, vol. 64, no. 22, pp. 2647–2650, 1990.

[2] L. Wang, P. P. Ho, C. Liu, G. Zhang, and R. R. Alfano, "Ballistic 2-D imaging through scattering walls using an ultrafast optical Kerr gate," *Science*, vol. 253, no. 5021, pp. 769–771, 1991.

[3] C. Dunsby and P. M. French, "Techniques for depth-resolved imaging through turbid media including coherence-gated imaging," *Journal of Physics D: Applied Physics*, vol. 36, no. 14, 2003.

[4] M. G. Tanner, T. R. Choudhary, T. H. Craven, B. Mills, M. Bradley, R. K. Henderson, K. Dhaliwal, and R. R. Thomson, "Ballistic and snake photon imaging for locating optical endomicroscopy fibres," *Biomedical Optics Express*, vol. 8, no. 9, pp. 4077–4095, 2017.

[5] M. J. Niedre, G. M. Turner, and V. Ntziachristos, "Time-resolved imaging of optical coefficients through murine chest cavities," *Journal of Biomedical Optics*, vol. 11, no. 6, p. 064017, 2006.

[6] V. Gopal, S. Mujumdar, H. Ramachandran, and A. K. Sood, "Imaging in turbid media using quasi-ballistic photons," *Optics Communications*, vol. 170, pp. 331–345, 1999.

[7] B. Brezner, S. Cahen, Z. Glasser, S. Sternklar, and E. Granot, "Ballistic imaging of biological media with collimated illumination and focal plane detection," *Journal of Biomedical Optics*, vol. 20, no. 7, p. 076006, 2015.

[8] A. V. Kanaev, A. T. Watnik, D. F. Gardner, C. Metzler, K. P. Judd, P. Lebow, K. M. Novak, and J. R. Lindle, "Imaging through extreme scattering in extended dynamic media," *Optics Letters*, vol. 43, no. 13, p. 3088, 2018.

[9] S. Woo, M. Kang, C. Yoon, T. D. Yang, Y. Choi, and W. Choi, "Three-dimensional imaging of macroscopic objects hidden behind scattering media using time-gated aperture synthesis," *Optics Express*, vol. 25, no. 26, p. 32722, 2017.

[10] S. Maruca, P. Rehain, Y. M. Sua, S. Zhu, and Y. Huang, "Non-invasive single photon imaging through strongly scattering media," *Optics Express*, vol. 29, no. 7, p. 9981, 2021.

[11] G. Satat, B. Heshmat, D. Raviv, and R. Raskar, "All Photons Imaging Through Volumetric Scattering," *Scientific Reports*, vol. 6, no. 1, p. 33946, 2016.

[12] A. Lyons, F. Tonolini, A. Boccolini, A. Repetti, R. Henderson, Y. Wiaux, and D. Faccio, "Computational time-of-flight diffuse optical tomography," *Nature Photonics*, vol. 13, no. 8, pp. 575–579, 2019.

[13] F. Tonolini, J. Radford, A. Turpin, D. Faccio, and R. Murray-Smith, "Variational inference for computational imaging inverse problems," *Journal of Machine Learning Research*, vol. 21, pp. 1–46, 2020.

[14] A. Turpin, G. Musarra, V. Kapitany, F. Tonolini, A. Lyons, I. Starshynov, F. Villa, E. Conca, F. Fioranelli, R. Murray-Smith, and D. Faccio, "Spatial images from temporal data," *Optica*, vol. 7, no. 8, pp. 900–905, 2020.

[15] E. Hecht and P. (Firm), *Optics*. Boston: Pearson Education, Inc, 5 global ed., 2017.

[16] C. Zhu and Q. Liu, "Review of Monte Carlo modeling of light transport in tissues," *Journal of Biomedical Optics*, vol. 18, no. 5, p. 050902, 2013.

[17] L. V. Wang and H.-i. Wu, *Biomedical optics: principles and imaging*. Hoboken, N.J: Wiley-Interscience, 2009.

[18] A. J. Welch and M. J. C. Van Gemert, *Optical-thermal response of laser-irradiated tissue*. Springer Dordrecht, 2nd ed., 2011.

[19] J. R. Lorenzo, *Principles of diffuse light propagation: light propagation in tissues with applications in biology and medicine*. London: World Scientific, 2012.

[20] L. Shi and R. R. Alfano, *Deep Imaging in Tissue and Biomedical Materials: Using Linear and Nonlinear Optical Methods*. Jenny Stanford Publishing, 2017.

[21] M. S. Patterson, B. Chance, and B. C. Wilson, "Time resolved reflectance and transmittance for the noninvasive measurement of tissue optical properties," *Applied Optics*, vol. 28, no. 12, p. 2331, 1989.

[22] A. Ishimaru, *Wave propagation and scattering in random media*. IEEE Press, 1997.

[23] A. Dunn and R. Richards-Kortum, "Three-Dimensional Computation of Light Scattering From Cells," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 2, no. 4, 1996.

[24] W. F. Cheong, S. A. Prahl, and A. J. Welch, "A Review of the Optical Properties of Biological Tissues," *IEEE Journal of Quantum Electronics*, vol. 26, no. 12, pp. 2166–2185, 1990.

[25] S. L. Jacques, "Optical properties of biological tissues: a review," *Physics in Medicine and Biology*, vol. 58, no. 11, pp. R37–R61, 2013.

[26] F. S. Marzano, "Radiative Transfer, Theory," in *Encyclopedia of Remote Sensing* (E. G. Njoku, ed.), pp. 624–634, New York, NY: Springer New York, 2014.

[27] S. Chandrasekhar, "Stochastic problems in physics and astronomy," *Reviews of Modern Physics*, vol. 15, no. 1, pp. 1–89, 1943.

[28] W. Becker, *Advanced time-correlated single photon counting techniques*. Springer Berlin, Heidelberg, 1 ed., 2005.

[29] P. P. Ho and R. R. Alfano, "Optical Kerr effect in liquids," *Physical Review A*, vol. 20, no. 5, pp. 2170–2187, 1979.

[30] A. Pifferi, D. Contini, A. D. Mora, A. Farina, L. Spinelli, and A. Torricelli, "New frontiers in time-domain diffuse optics, a review," *Journal of Biomedical Optics*, vol. 21, no. 9, p. 091310, 2016.

[31] M. Buttafava, E. Martinenghi, D. Tamborini, D. Contini, A. D. Mora, M. Renna, A. Torricelli, A. Pifferi, F. Zappa, and A. Tosi, "A Compact Two-Wavelength Time-Domain NIRS System Based on SiPM and Pulsed Diode Lasers," *IEEE Photonics Journal*, vol. 9, no. 1, 2017.

[32] K. Morimoto, A. Ardelean, M.-L. Wu, A. C. Ulku, I. M. Antolovic, C. Bruschini, and E. Charbon, "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications," *Optica*, vol. 7, no. 4, p. 346, 2020.

[33] J. Richardson, R. Walker, L. Grant, D. Stoppa, F. Borghetti, E. Charbon, M. Gersbach, and R. K. Henderson, "A 32x32 50ps Resolution 10 bit Time to Digital Converter Array in 130nm CMOS for time Correlated Imaging," in *International Image Sensor Workshop*, 2009.

[34] I. Gyongy, N. Calder, A. Davies, N. A. Dutton, R. R. Duncan, C. Rickman, P. Dalgarno, and R. K. Henderson, "A 256×256 , 100-kfps, 61% Fill-Factor SPAD Image Sensor for Time-Resolved Microscopy Applications," *IEEE Transactions on Electron Devices*, vol. 65, no. 2, pp. 547–554, 2018.

[35] A. D. Mora, E. Martinenghi, D. Contini, A. Tosi, G. Boso, T. Durduran, S. Arridge, F. Martelli, A. Farina, A. Torricelli, and A. Pifferi, "Fast silicon photomultiplier improves signal harvesting and reduces complexity in time-domain diffuse optics," *Optics Express*, vol. 23, no. 11, p. 13937, 2015.

[36] S. I. Fujisaka, T. Ozaki, T. Suzuki, T. Kamada, K. Kitazawa, M. Nishizawa, A. Takahashi, and S. Suzuki, "A clinical tissue oximeter using nir time-resolved spectroscopy," in *Advances in Experimental Medicine and Biology*, vol. 876, pp. 427–433, Springer, New York, NY, 2016.

[37] A. Yodh and B. Chance, "Spectroscopy and Imaging with Diffusing Light," *Physics Today*, vol. 48, no. 3, pp. 34–40, 1995.

[38] F. Scholkmann, S. Kleiser, A. J. Metz, R. Zimmermann, J. Mata Pavia, U. Wolf, and M. Wolf, "A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology," *NeuroImage*, vol. 85, pp. 6–27, 2014.

[39] A. T. Eggebrecht, B. R. White, S. L. Ferradal, C. Chen, Y. Zhan, A. Z. Snyder, H. Dehghani, and J. P. Culver, "A quantitative spatial comparison of high-density diffuse optical tomography and fMRI cortical mapping," *NeuroImage*, vol. 61, no. 4, pp. 1120–1128, 2012.

[40] V. Scarapicchia, C. Brown, C. Mayo, and J. R. Gawryluk, "Functional magnetic resonance imaging and functional near-infrared spectroscopy: Insights from combined recording studies," *Frontiers in Human Neuroscience*, vol. 11, 8 2017.

[41] P. Pinti, C. Aichelburg, S. Gilbert, A. Hamilton, J. Hirsch, P. Burgess, and I. Tachtsidis, "A Review on the Use of Wearable Functional Near-Infrared Spectroscopy in Naturalistic Environments," *Japanese Psychological Research*, vol. 60, no. 4, pp. 347–373, 2018.

[42] H. Y. Ban, G. M. Barrett, A. Borisevich, A. Chaturvedi, J. L. Dahle, H. Dehghani, J. Dubois, R. M. Field, V. Gopalakrishnan, A. Gundran, M. Henninger, W. C. Ho, H. D. Hughes, R. Jin, J. Kates-Harbeck, T. Landy, M. Leggiero, G. Lerner, Z. M. Aghajan, M. Moon, I. Olvera, S. Park, M. J. Patel, K. L. Perdue, B. Siepser, S. Sorgenfrei, N. Sun, V. Szczepanski, M. Zhang, and Z. Zhu, "Kernel Flow: a high channel count scalable time-domain functional near-infrared spectroscopy system," *Journal of Biomedical Optics*, vol. 27, no. 7, 2022.

[43] P. Pinti, I. Tachtsidis, A. Hamilton, J. Hirsch, C. Aichelburg, S. Gilbert, and P. W. Burgess, "The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience," *Annals of the New York Academy of Sciences*, vol. 1464, no. 1, pp. 5–29, 2018.

[44] L. Gagnon, M. A. Yücel, D. A. Boas, and R. J. Cooper, "Further improvement in reducing superficial contamination in NIRS using double short separation measurements," *NeuroImage*, vol. 85, pp. 127–135, 1 2014.

[45] J. Selb, D. K. Joseph, and D. A. Boas, "Time-gated optical system for depth-resolved functional brain imaging," *Journal of Biomedical Optics*, vol. 11, no. 4, p. 044008, 2006.

[46] A. Cheung, L. Tu, A. Macnab, B. K. Kwon, and B. Shadgan, "Detection of hypoxia by near-infrared spectroscopy and pulse oximetry: a comparative study," *Journal of Biomedical Optics*, vol. 27, no. 7, 2022.

[47] C. E. Elwell, M. Cope, A. D. Edwards, J. S. Wyatt, D. T. Delpy, and R. Reynolds, "Quantification of adult cerebral hemodynamics by near-infrared spectroscopy," *Journal of Applied Physiology*, vol. 77, no. 6, pp. 2753–2760, 1994.

[48] Y. Yamada and S. Okawa, "Diffuse optical tomography: Present status and its future," *Optical Review*, vol. 21, no. 3, pp. 185–205, 2014.

[49] F. Jobsis, "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science*, vol. 198, no. 4323, pp. 1264–1267, 1977.

[50] J. H. Meek, M. Firbank, C. E. Elwell, J. Atkinson, O. Braddick, and J. S. Wyatt, "Regional Hemodynamic Responses to Visual Stimulation in Awake Infants," *Pediatric Research*, vol. 43, no. 6, pp. 840–843, 1998.

[51] Y. Hoshi and A. Mamoru, "Dynamic multichannel near-infrared optical imaging of human brain activity," *Journal of applied physiology*, vol. 75, no. 4, pp. 1842–1846, 1993.

[52] A. Villringer, J. Planck, C. Hock, L. Schleinkofer, and U. Dirnagl, "Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults," *Neuroscience Letters*, vol. 154, pp. 101–104, 1993.

[53] H. Zhao and R. J. Cooper, "Review of recent progress toward a fiberless, whole-scalp diffuse optical tomography system," *Neurophotonics*, vol. 5, no. 1, p. 011012, 2017.

[54] J. Z. Wang, X. Liang, Q. Zhang, L. L. Fajardo, and H. Jiang, "Automated breast cancer classification using near-infrared optical tomographic images," *Journal of Biomedical Optics*, vol. 13, no. 4, p. 044001, 2008.

[55] N. Iftimia, X. Gu, Y. Xu, and H. Jiang, "A compact, parallel-detection diffuse optical mammography system," *Review of Scientific Instruments*, vol. 74, no. 5, pp. 2836–2842, 2003.

[56] G. Bale, S. Mitra, J. Meek, N. Robertson, and I. Tachtsidis, "A new broadband near-infrared spectroscopy system for in-vivo measurements of cerebral cytochrome-c-oxidase changes in neonatal brain injury," *Biomedical Optics Express*, vol. 5, no. 10, p. 3450, 2014.

[57] H. Singh, R. J. Cooper, C. Wai Lee, L. Dempsey, A. Edwards, S. Brigadoi, D. Airantzis, N. Everdell, A. Michell, D. Holder, J. C. Hebden, and T. Austin, "Mapping cortical haemodynamics during neonatal seizures using diffuse optical tomography: A case study," *NeuroImage: Clinical*, vol. 5, pp. 256–265, 2014.

[58] A. T. Eggebrecht, S. L. Ferradal, A. Robichaux-Viehoever, M. S. Hassanpour, H. Dehghani, A. Z. Snyder, T. Hershey, and J. P. Culver, "Mapping distributed brain function and networks with diffuse optical tomography," *Nature Photonics*, vol. 8, 2014.

[59] H. Zhao, E. M. Frijia, E. Vidal Rosas, L. Collins-Jones, G. Smith, R. Nixon-Hill, S. Powell, N. L. Everdell, and R. J. Cooper, "Design and validation of a mechanically flexible and ultra-lightweight high-density diffuse optical tomography system for functional neuroimaging of newborns," *Neurophotonics*, vol. 8, no. 01, 2021.

[60] M. D. Wheelock, J. P. Culver, and A. T. Eggebrecht, "High-density diffuse optical tomography for imaging human brain function," *Review of Scientific Instruments*, vol. 90, no. 5, p. 051101, 2019.

[61] D. Grosenick, H. Rinneberg, R. Cubeddu, and P. Taroni, "Review of optical breast imaging and spectroscopy," *Journal of Biomedical Optics*, vol. 21, no. 9, p. 091311, 2016.

[62] D. Viderman, A. Ayapbergenov, N. Abilman, and Y. G. Abdildin, "Near-infrared spectroscopy for intracranial hemorrhage detection in traumatic brain injury patients: A systematic review," *American Journal of Emergency Medicine*, vol. 50, pp. 758–764, 2021.

[63] P. Bloomfield, S. Brigadoi, G. Rizzo, and M. Veronese, *Basic Neuroimaging: A Guide to the Methods and Their Applications*. CreateSpace Independent Publishing Platform, 2017.

[64] M. A. Franceschini, S. Thaker, G. Themelis, K. K. Krishnamoorthy, H. Bortfeld, S. G. Diamond, D. A. Boas, K. Arvin, and P. E. Grant, "Assessment of Infant Brain Development With Frequency-Domain Near-Infrared Spectroscopy," *Pediatric Research*, vol. 61, no. 5, pp. 546–551, 2007.

[65] T. Durduran, R. Choe, J. P. Culver, L. Zubkov, M. J. Holboke, J. Giammarco, B. Chance, and A. G. Yodh, "Bulk optical properties of healthy female breast tissue," *Physics in Medicine and Biology*, vol. 47, no. 16, pp. 2847–2861, 2002.

[66] A. Torricelli, D. Contini, A. Pifferi, M. Caffini, R. Re, L. Zucchelli, and L. Spinelli, "Time domain functional NIRS imaging for human brain mapping," *NeuroImage*, vol. 85, pp. 28–50, 2014.

[67] M. Alayed and M. J. Deen, "Time-resolved diffuse optical spectroscopy and imaging using solid-state detectors: Characteristics, present status, and research challenges," *Sensors (Switzerland)*, vol. 17, no. 9, p. 2115, 2017.

[68] M. Schweiger and S. R. Arridge, "Optical tomographic reconstruction in a complex head model using a priori region boundary information," *Physics in Medicine & Biology*, vol. 44, pp. 2703–2721, 1999.

[69] X. Zhou, L. Chen, C. Y. Tse, T. B. Penney, and N. Chen, "Theoretical investigation of near-infrared light path in multi-layer brain models for three DOT systems," in *2012 Photonics Global Conference (PCG)*, 2012.

[70] A. Torricelli, A. Pifferi, L. Spinelli, R. Cubeddu, F. Martelli, S. Del Bianco, and G. Zaccanti, "Time-resolved reflectance at null source-detector separation: Improving contrast and resolution in diffuse optical imaging," *Physical Review Letters*, vol. 95, no. 7, 2005.

[71] M. Pagliazzi, L. Colombo, E. E. Vidal-Rosas, T. Dragojević, V. Parfentyeva, J. P. Culver, S. Konugolu Venkata Sekar, L. Di Sieno, D. Contini, A. Torricelli, A. Pifferi, A. Dalla Mora, and T. Durduran, "Time resolved speckle contrast optical spectroscopy at quasi-null source-detector separation for non-invasive measurement of microvascular blood flow," *Biomedical Optics Express*, vol. 12, no. 3, p. 1499, 2021.

[72] A. D. Mora, D. Contini, S. Arridge, F. Martelli, A. Tosi, G. Boso, A. Farina, T. Durduran, E. Martinenghi, A. Torricelli, and A. Pifferi, "Towards next-generation time-domain diffuse optics for extreme depth penetration and sensitivity," *Biomedical Optics Express*, vol. 6, no. 5, p. 1749, 2015.

[73] L. Spinelli, D. Contini, A. Farina, A. Torricelli, A. Pifferi, R. Cubeddu, L. Ascari, L. Potì, M. G. Trivella, A. L'Abbate, and S. Puzzuoli, "In vivo swine myocardial tissue characterization and monitoring during open chest surgery by time-resolved diffuse near-infrared spectroscopy," in *Photonic Therapeutics and Diagnostics VII*, vol. 7883, p. 78833D, SPIE, 2 2011.

[74] G. Quarto, A. Farina, A. Pifferi, P. Taroni, and M. Miniati, "Time-resolved optical spectroscopy of the chest: is it possible to probe the lung?," *European Conferences on Biomedical Optics*, p. 87990Q, 2013.

[75] Y. Xu, N. Iftimia, H. Jiang, L. L. Key, and M. B. Bolster, "Three-dimensional diffuse optical tomography of bones and joints," *Journal of Biomedical Optics*, vol. 7, no. 1, p. 88, 2002.

[76] A. Pifferi, A. Torricelli, P. Taroni, A. Bassi, E. Chikoidze, E. Giambattistelli, and R. Cubeddu, "Optical biopsy of bone tissue: a step toward the diagnosis of bone pathologies," *Journal of Biomedical Optics*, vol. 9, no. 3, p. 474, 2004.

[77] L. Wang, P. P. Ho, C. Liu, G. Zhang, and R. R. Alfano, "Ballistic 2-D Imaging Through Scattering Walls Using an Ultrafast Optical Kerr Gate," *Science*, vol. 253, no. 5021, pp. 769–771, 1991.

[78] X. Ni, Q. Xing, W. Cai, and R. R. Alfano, "Time-resolved polarization to extract coded information from early ballistic and snake signals through turbid media," *Optics Letters*, vol. 28, no. 5, p. 343, 2003.

[79] J. C. Hebden, R. A. Kruger, and K. S. Wong, "Time resolved imaging through a highly scattering medium," *Applied Optics*, vol. 30, no. 7, pp. 788–794, 1991.

[80] G. Jarry, S. Ghesquiere, J. Maarek, F. Fraysse, S. Debray, Bui-Mong-Hung, and D. Laurent, "Imaging mammalian tissues and organs using laser collimated transillumination," *Journal of Biomedical Engineering*, vol. 6, pp. 70–74, 1984.

[81] J. C. Hebden, D. R. Reinecke, and R. A. Kruger, "Examining The Feasibility Of Laser Collimated Transillumination," in *Medical Imaging III: Image Formation*, vol. 1090, p. 35, SPIE, 1989.

[82] J. M. Maarek, G. Jarry, J. Crowe, M.-H. Bui, and D. Laurent, "Simulation of laser tomoscopy in a heterogeneous biological medium," *Medical & Biological Engineering & Computing*, vol. 24, pp. 407–414, 1986.

[83] M. A. Duguay and A. T. Mattick, "Ultrahigh Speed Photography of Picosecond Light Pulses and Echoes," *Applied Optics*, vol. 10, no. 9, pp. 2162–2170, 1971.

[84] J. L. Martin, Y. Lecarpentier, A. Antonetti, and G. Grillon, "Picosecond laser stereometry light scattering measurements on biological material," *Medical & Biological Engineering & Computing*, vol. 18, pp. 250–252, 1980.

[85] K. M. Yoo and R. R. Alfano, "Time-resolved coherent and incoherent components of forward light scattering in random media," *Optics Letters*, vol. 15, no. 6, pp. 320–322, 1990.

[86] R. Berg, O. Jarlman, and S. Svanberg, "Medical transillumination imaging using short-pulse diode lasers," *Applied Optics*, vol. 32, no. 4, pp. 574–579, 1993.

[87] S. Andersson-Engels, R. Berg, and S. Svanberg, "Time-resolved transillumination for medical diagnostics," *Optics Letters*, vol. 15, no. 21, pp. 1179–1181, 1990.

[88] D. A. Benaron and D. K. Stevenson, "Optical Time-of-Flight and Absorbance Imaging of Biologic Media," *Science*, vol. 259, no. 5100, pp. 1463–1466, 1993.

[89] G. M. Turner, G. Zacharakis, A. Soubret, J. Ripoll, and V. Ntziachristos, "Complete-angle projection diffuse optical tomography by use of early photons," *Optics Letters*, vol. 30, no. 4, p. 409, 2005.

[90] J. Chen, V. Venugopal, F. Lesage, and X. Intes, "Time-resolved diffuse optical tomography with patterned-light illumination and detection," *Optics Letters*, vol. 35, no. 13, p. 2121, 2010.

[91] J. Pichette, J. B. Domínguez, and Y. Bérubé-Lauzière, "Time-domain geometrical localization of point-like fluorescence inclusions in turbid media with early photon arrival times," *Applied Optics*, vol. 52, no. 24, pp. 5985–5999, 2013.

[92] M. J. Niedre, R. H. De Kleine, E. Aikawa, D. G. Kirsch, R. Weissleder, and V. Ntziachristos, "Early photon tomography allows fluorescence detection of lung carcinomas and disease progression in mice in vivo," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 49, pp. 19126–19131, 2008.

[93] B. Zhang, X. Cao, F. Liu, X. Liu, X. Wang, and J. Bai, "Early-photon fluorescence tomography of a heterogeneous mouse model with the telegraph equation," *Applied Optics*, vol. 50, no. 28, pp. 5397–5407, 2011.

[94] E. P. McShane, H. K. Chandrasekharan, A. Kufcsák, N. Finlayson, A. T. Erdogan, R. K. Henderson, K. Dhaliwal, R. R. Thomson, and M. G. Tanner, "High resolution TCSPC imaging of diffuse light with a one-dimensional SPAD array scanning system," *Optics Express*, vol. 30, no. 15, p. 27926, 2022.

[95] J. Bertolotti and O. Katz, "Imaging in complex media," *Nature Physics*, vol. 18, no. 9, pp. 1008–1017, 2022.

[96] J. R. Singer, F. A. Grünbaum, P. Kohn, and J. P. Zubelli, "Image Reconstruction of the Interior of Bodies That Diffuse Radiation," *Science*, vol. 248, no. 4958, pp. 990–993, 1990.

[97] S. R. Arridge, P. van der Zee, M. Cope, and D. T. Delpy, "Reconstruction methods for infrared absorption imaging," in *Proceedings of SPIE: Time-Resolved Spectroscopy and Imaging of Tissues*, vol. 1431, pp. 204–215, SPIE, 1991.

[98] W. Cai, B. B. Das, F. Liu, M. Zevallos, M. Lax, and R. R. Alfano, "Time-resolved optical diffusion tomographic image reconstruction in highly scattering turbid media," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 24, pp. 13561–13564, 1996.

[99] H. Eda, I. Oda, Y. Ito, Y. Wada, Y. Oikawa, Y. Tsunazawa, M. Takada, Y. Tsuchiya, Y. Yamashita, M. Oda, A. Sassaroli, Y. Yamada, and M. Tamura, "Multichannel time-resolved optical tomographic imaging system," *Review of Scientific Instruments*, vol. 70, no. 9, pp. 3595–3602, 1999.

[100] A. H. Gandjbakhche, V. Chernomordik, J. C. Hebden, and R. Nossal, "Time-dependent contrast functions for quantitative imaging in time-resolved transillumination experiments," *Applied Optics*, vol. 37, no. 10, p. 1973, 1998.

[101] H. Jiang, N. V. Iftimia, Y. Xu, J. A. Eggert, L. L. Fajardo, and K. L. Klove, "Near-Infrared Optical Imaging of the Breast with Model-based Reconstruction," *Academic Radiology*, vol. 9, no. 2, pp. 186–194, 2002.

[102] W. Cai, S. K. Gayen, M. Xu, M. Zevallos, M. Alrubaiee, M. Lax, and R. R. Alfano, "Optical tomographic image reconstruction from ultrafast time-sliced transmission measurements," *Applied Optics*, vol. 38, no. 19, p. 4237, 1999.

[103] S. R. Arridge, J. C. Hebden, M. Schweiger, F. E. Schmidt, M. E. Fry, E. M. Hillman, H. Dehghani, and D. T. Delpy, "Method for three-dimensional time-resolved optical tomography," *International Journal of Imaging Systems and Technology*, vol. 11, no. 1, pp. 2–11, 2000.

[104] R. Cubeddu, A. Pifferi, P. Taroni, A. Torricelli, and G. Valentini, "Imaging of optical inhomogeneities in highly diffusive media: Discrimination between scattering and absorption contributions," *Applied Physics Letters*, vol. 69, no. 27, pp. 4162–4164, 1996.

[105] J. C. Hebden, F. E. W. Schmidt, M. E. Fry, M. Schweiger, E. M. C. Hillman, D. T. Delpy, and S. R. Arridge, "Simultaneous reconstruction of absorption and scattering images by multichannel measurement of purely temporal data," *Optics Letters*, vol. 24, no. 8, 1999.

[106] S. Lam, F. Lesage, and X. Intes, "Time Domain Fluorescent Diffuse Optical Tomography: analytical expressions," *Optics Express*, vol. 13, no. 7, p. 2263, 2005.

[107] F. Gao, H. Zhao, and Y. Yamada, "Improvement of image quality in diffuse optical tomography by use of full time-resolved data," *Applied Optics*, vol. 41, no. 4, pp. 778–791, 2002.

[108] E. M. C Hillman, J. C. Hebden, M. Schweiger, H. Dehghani, F. E. W Schmidt, D. T. Delpy, and S. R. Arridge, "Time resolved optical tomography of the human forearm," *Physics in Medicine and Biology*, vol. 46, pp. 1117–1130, 2001.

[109] H. Zhao, F. Gao, Y. Tanikawa, K. Homma, and Y. Yamada, "NIR time domain diffuse optical tomography experiments on human forearm," in *Optical Tomography and Spectroscopy of Tissue V* (B. Chance, R. R. Alfano, B. J. Tromberg, M. Tamura, and E. M. Sevick-Muraca, eds.), vol. 4955, pp. 437–446, SPIE, 2003.

[110] S. Okawa and Y. Hoshi, "A Review of Image Reconstruction Algorithms for Diffuse Optical Tomography," *Applied Sciences*, vol. 13, p. 5016, 4 2023.

[111] H. Dehghani, S. Sri Nivasan, B. W. Pogue, and A. Gibson, "Numerical modelling and image reconstruction in diffuse optical tomography," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1900, pp. 3073–3093, 2009.

[112] S. R. Arridge, "Methods in diffuse optical imaging," *Trans. R. Soc. A*, vol. 369, pp. 4558–4576, 2011.

[113] G. M. Balasubramaniam, B. Wiesel, N. Biton, R. Kumar, J. Kupferman, and S. Arnon, "Tutorial on the Use of Deep Learning in Diffuse Optical Tomography," *Electronics (Switzerland)*, vol. 11, no. 3, 2022.

[114] S. Arridge, P. Maass, O. Öktem, and C. B. Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, pp. 1–174, 2019.

[115] P. C. Hansen, *Discrete Inverse Problems*. Society for Industrial and Applied Mathematics, 2010.

[116] D. L. Myres, "Ill-conditioned equations," 1964.

[117] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. USA: Cambridge University Press, 3 ed., 2007.

[118] S. L. Brunton and J. N. Kutz, "Singular Value Decomposition (SVD)," in *Data-Driven Science and Engineering*, pp. 3–46, Cambridge University Press, 2019.

[119] S. R. Arridge, J. P. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, and M. Vauhkonen, "Approximation errors and model reduction with an application in optical diffusion tomography," *Inverse Problems*, vol. 22, no. 1, pp. 175–195, 2006.

[120] J. C. Hebden and D. T. Delpy, "Enhanced time-resolved imaging with a diffusion model of photon transport," *Optics Letters*, vol. 19, no. 5, p. 311, 1994.

[121] S. D. Konecky, G. Y. Panasyuk, K. Lee, V. Markel, A. G. Yodh, and J. C. Schotland, "Imaging complex structures with diffuse light," *Optics Express*, vol. 16, no. 7, p. 5048, 2008.

[122] C. Panagiotou, S. Somayajula, A. P. Gibson, M. Schweiger, R. M. Leahy, and S. R. Arridge, "Information theoretic regularization in diffuse optical tomography," *Journal of the Optical Society of America A*, vol. 26, no. 5, p. 1277, 2009.

[123] F. Schiffers, L. Fiske, P. Ruiz, A. K. Katsaggelos, and O. Cossairt, "Imaging through Scattering Media with a Learning Based Prior," *Electronic Imaging*, vol. 32, no. 14, pp. 306–1, 2020.

[124] J. Yoo, S. Sabir, D. Heo, K. H. Kim, A. Wahab, Y. Choi, S. I. Lee, E. Y. Chae, H. H. Kim, Y. M. Bae, Y. W. Choi, S. Cho, and J. C. Ye, "Deep Learning Diffuse Optical Tomography," *IEEE Transactions on Medical Imaging*, vol. 39, no. 4, pp. 877–887, 2020.

[125] J. Han, S. Zhu, E. Guo, J. Gu, and L. Bai, "Imaging through unknown scattering media based on physics-informed learning," *Photonics Research*, vol. 9, no. 5, pp. 210–219, 2021.

[126] J. C. Hebden, D. J. Hall, and D. T. Delpy, "The spatial resolution performance of a time-resolved optical imaging system using temporal extrapolation," *Medical Physics*, vol. 22, no. 2, pp. 201–208, 1995.

[127] S. G. Proskurin, "Using late arriving photons for diffuse optical tomography of biological objects," *Quantum Electronics*, vol. 41, no. 5, pp. 402–406, 2011.

[128] S. Farsiu, J. Christofferson, B. Eriksson, P. Milanfar, B. Friedlander, A. Shakouri, and R. Nowak, "Statistical detection and imaging of objects hidden in turbid media using ballistic photons," *Applied Optics*, vol. 46, no. 23, pp. 5805–5822, 2007.

[129] D. C. Sordillo, L. A. Sordillo, P. P. Sordillo, L. Shi, and R. R. Alfano, "Short wavelength infrared optical windows for evaluation of benign and malignant tissues," *Journal of Biomedical Optics*, vol. 22, no. 4, p. 045002, 2017.

[130] A. Yaroshevsky, Z. Glasser, E. Granot, and S. Sternklar, "Transition from the ballistic to the diffusive regime in a turbid medium," *Optics Letters*, vol. 36, no. 8, p. 1395, 2011.

[131] F. Liu, K. M. Yoo, and R. R. Alfano, "Transmitted photon intensity through biological tissues within various time windows," *Optics Letters*, vol. 19, no. 10, pp. 740–742, 1994.

[132] Z. Glasser, A. Yaroshevsky, B. Barak, E. Granot, and S. Sternklar, "Effect of measurement on the ballistic–diffusive transition in turbid media," *Journal of Biomedical Optics*, vol. 18, no. 10, p. 106006, 2013.

[133] J. Radford, A. Lyons, F. Tonolini, and D. Faccio, "Role of late photons in diffuse optical imaging," *Optics Express*, vol. 28, no. 20, 2020.

[134] G. Cohen, S. Afshar, J. Tapson, and A. Van Schaik, "EMNIST: Extending MNIST to handwritten letters," in *Proceedings of the International Joint Conference on Neural Networks*, vol. 2017-May, (Anchorage), pp. 2921–2926, IEEE, 2017.

[135] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, vol. 2, no. Ki L, pp. 1398–1402, 2003.

[136] I. Delfino, M. Lepore, and P. L. Indovina, "Experimental tests of different solutions to the diffusion equation for optical characterization of scattering media by time-resolved transmittance," *Applied Optics*, vol. 38, no. 19, p. 4228, 1999.

[137] R. A. Ince, B. L. Giordano, C. Kayser, G. A. Rousselet, J. Gross, and P. G. Schyns, "A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula," *Human Brain Mapping*, vol. 38, no. 3, pp. 1541–1573, 2017.

[138] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, 1948.

[139] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.

[140] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[141] A. Turpin, V. Kapitany, J. Radford, D. Rovelli, K. Mitchell, A. Lyons, I. Starshynov, and D. Faccio, "3D Imaging from Multipath Temporal Echoes," *Physical Review Letters*, vol. 126, no. 17, p. 174301, 2021.

[142] J. Radford and D. Faccio, "Information transport and limits of optical imaging in the highly diffusive regime," *Physical Review Research*, vol. 5, no. 2, p. L022008, 2023.

[143] British Standards Institution, "Safety of laser products - Part 14: A user's guide," tech. rep., London, 2004.

[144] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, "Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns," *IEEE Control Systems*, vol. 38, no. 3, pp. 63–86, 2018.

[145] H. Wabnitz, D. R. Taubert, M. Mazurenka, O. Steinkellner, A. Jelzow, R. Macdonald, D. Milej, P. Sawosz, M. Kacprzak, A. Liebert, R. Cooper, J. Hebden, A. Pifferi, A. Farina, I. Bargigia, D. Contini, M. Caffini, L. Zucchelli, L. Spinelli, R. Cubeddu, and A. Torricelli,

"Performance assessment of time-domain optical brain imagers, part 1: basic instrumental performance protocol," *Journal of Biomedical Optics*, vol. 19, no. 8, p. 086010, 2014.

[146] A. Liebert, H. Wabnitz, D. Grosenick, and R. Macdonald, "Fiber dispersion in time domain measurements compromising the accuracy of determination of optical properties of strongly scattering media," *Journal of Biomedical Optics*, vol. 8, no. 3, p. 512, 2003.

[147] F. Leblond, H. Dehghani, D. Kepshire, and B. W. Pogue, "Early-photon fluorescence tomography: spatial resolution improvements and noise stability considerations," *Journal of the Optical Society of America A*, vol. 26, no. 6, p. 1444, 2009.

[148] C. B. Shaw, "Effective contrast recovery in rapid dynamic near-infrared diffuse optical tomography using L1-norm-based linear image reconstruction method," *Journal of Biomedical Optics*, vol. 17, no. 8, p. 086009, 2012.

[149] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase Retrieval with Application to Optical Imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, 2015.

[150] O. Moran, P. Caramazza, D. Faccio, and R. Murray-Smith, "Deep, complex, invertible networks for inversion of transmission effects in multimode optical fibres," in *Advances in Neural Information Processing Systems*, vol. 2018-Decem, pp. 3280–3291, 2018.

[151] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[152] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, 2014.

[153] K. Sohn, X. Yan, and H. Lee, "Learning Structured Output Representation using Deep Conditional Generative Models," in *Advances in Neural Information Processing Systems*, pp. 3483–3491, 2015.

[154] L. Deng, "The MNIST database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012.

[155] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[156] J. Radford, S. Nerenberg, A. Shvarts, L. Kaczmarcyk, and D. Faccio, "Imaging Beyond 100 Transport Mean Free Paths using Variational Autoencoding," in *Imaging and Applied Optics Congress 2022 (3D, AOA, COSI, ISA, pcAOP)*, p. CM3A.5, Optica Publishing Group, 2022.

[157] B. Dai, Z. Wang, and D. Wipf, "The Usual Suspects? Reassessing Blame for VAE Posterior Collapse," in *Proceedings of the 37th International Conference on Machine Learning* (H. D. III and A. Singh, eds.), vol. 119 of *Proceedings of Machine Learning Research*, pp. 2313–2322, PMLR, 7 2020.

[158] B. Kessel, I. Jeroukhimov, I. Ashkenazi, T. Khashan, M. Oren, J. Haspel, M. Medvedev, V. Nesterenko, A. Halevy, and R. Alfici, "Early detection of life-threatening intracranial haemorrhage using a portable near-infrared spectroscopy device," *Injury*, vol. 38, no. 9, pp. 1065–1068, 2007.

[159] G. Giacalone, M. Zanoletti, D. Contini, R. Re, L. Spinelli, L. Roveri, and A. Torricelli, "Cerebral time domain-NIRS: reproducibility analysis, optical properties, hemoglobin species and tissue oxygen saturation in a cohort of adult subjects," *Biomedical Optics Express*, vol. 8, no. 11, p. 4987, 2017.

[160] A. Farina, A. Torricelli, I. Bargigia, L. Spinelli, R. Cubeddu, F. Foschum, M. Jäger, E. Simon, O. Fugger, A. Kienle, F. Martelli, P. D. Ninni, G. Zaccanti, D. Milej, P. Sawosz, M. Kacprzak, A. Liebert, and A. Pifferi, "In-vivo multilaboratory investigation of the optical properties of the human head," *Biomedical Optics Express*, vol. 6, pp. 2609–2623, 7 2015.

[161] V. Damagatla, P. Lanka, A. Brodu, N. Noordzij, J. Qin-Dregely, A. Farina, and A. Pifferi, "Interstitial null-distance time-domain diffuse optical spectroscopy using a superconducting nanowire detector," *Journal of Biomedical Optics*, vol. 28, 4 2023.

[162] J. E. Greivenkamp, *Field guide to geometrical optics*. SPIE Press, 2004.

[163] G. Gariepy, *Imaging light in motion and its application to tracking hidden objects*. PhD thesis, Heriot-Watt University, Edinburgh, 2016.

[164] Q. Fang and D. A. Boas, "Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units," *Optics Express*, vol. 17, no. 22, pp. 20178–20190, 2009.

[165] P. Cassano, A. P. Tran, H. Katnani, B. S. Bleier, M. R. Hamblin, Y. Yuan, and Q. Fang, "Selective photobiomodulation for emotion regulation: model-based dosimetry study," *Neurophotonics*, vol. 6, no. 1, p. 015004, 2019.

[166] D. A. Boas, J. P. Culver, J. J. Stott, A. K. Dunn, M. J. Holboke, B. J. Tromberg, X. Li, N. Shah, J. Fishkin, D. Kidney, J. Butler, B. Chance, A. G. Yodh, D. A. Benaron, W. F. Cheong, D. K. Stevenson, S. R. Hintz, A. Villringer, D. Boas, A. Kleinschmidt, J. Frahm, C. Hirth, H. Obrig, J. C. van Houten, E. L. Kermit, B. W. Pogue, S. P. Poplack, T. O. McBride, W. A. Wells, K. S. Osterman, U. L. Osterberg, and K. D. Paulsen, "Three di-

mensional Monte Carlo code for photon migration through complex heterogeneous media including the adult human head," *Optics Express*, vol. 10, no. 3, pp. 159–170, 2002.

[167] B. Aubert-Broche, A. C. Evans, and L. Collins, "A new improved version of the realistic digital brain phantom," *NeuroImage*, vol. 32, no. 1, pp. 138–145, 2006.

[168] A. P. Tran, S. Yan, and Q. Fang, "Improving model-based functional near-infrared spectroscopy analysis using mesh-based anatomical and light-transport models," *Neurophotonics*, vol. 7, no. 01, p. 1, 2020.

[169] D. Freedman and P. Diaconis, "On the histogram as a density estimator:L2 theory," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 57, no. 4, pp. 453–476, 1981.

[170] M. Zarei, M. A. Ansari, and K. Zare, "The temporal confounding effects of extra-cerebral contamination factors on the hemodynamic signal measured by functional near-infrared spectroscopy," *Journal of Lasers in Medical Sciences*, vol. 10, no. 4, pp. S73–S81, 2019.

[171] G. E. Strangman, Q. Zhang, and Z. Li, "Scalp and skull influence on near infrared photon propagation in the Colin27 brain template," *NeuroImage*, vol. 85, pp. 136–149, 2014.

[172] N. Eleveld, D. C. Esquivel-Franco, G. Drost, A. R. Absalom, C. J. Zeebregts, J. P. P. de Vries, J. W. J. Elting, and N. M. Maurits, "The Influence of Extracerebral Tissue on Continuous Wave Near-Infrared Spectroscopy in Adults: A Systematic Review of In Vivo Studies," *Journal of Clinical Medicine*, vol. 12, no. 8, 2023.

[173] Y. Fukui, Y. Ajichi, and E. Okada, "Monte Carlo prediction of near-infrared light propagation in realistic adult and neonatal head models," *Applied Optics*, vol. 42, no. 16, pp. 2881–2887, 2003.

[174] A. Custo, W. M. Wells Iii, A. H. Barnett, E. M. C. Hillman, and D. A. Boas, "Effective scattering coefficient of the cerebral spinal fluid in adult head models for diffuse optical imaging," *Applied Optics*, vol. 45, no. 19, pp. 4747–4755, 2006.

[175] E. Okada and D. T. Delpy, "Near-infrared light propagation in an adult head model. I. Modeling of low-level scattering in the cerebrospinal fluid layer," *Applied Optics*, vol. 42, no. 16, pp. 2906–2914, 2003.

[176] A. Kienle and R. Hibst, "Light guiding in biological tissue due to scattering," *Physical Review Letters*, vol. 97, no. 1, 2006.

[177] W. L. Chen, J. Wagner, N. Heugel, J. Sugar, Y. W. Lee, L. Conant, M. Malloy, J. Heffernan, B. Quirk, A. Zinos, S. A. Beardsley, R. Prost, and H. T. Whelan, "Functional Near-Infrared

Spectroscopy and Its Clinical Application in the Field of Neuroscience: Advances and Future Directions," *Frontiers in Neuroscience*, vol. 14, pp. 1–15, 7 2020.

[178] L. A. Sordillo, Y. Pu, S. Pratavieira, Y. Budansky, and R. R. Alfano, "Deep optical imaging of tissue using the second and third near-infrared spectral windows," *Journal of Biomedical Optics*, vol. 19, no. 5, p. 056004, 2014.

[179] L. Shi, L. A. Sordillo, A. Rodríguez-Contreras, and R. Alfano, "Transmission in near-infrared optical windows for deep brain imaging," *Journal of Biophotonics*, vol. 9, no. 1-2, pp. 38–43, 2016.

[180] N. Ozana, A. I. Zavriyev, D. Mazumder, M. Robinson, K. Kaya, M. Blackwell, S. A. Carp, and M. A. Franceschini, "Superconducting nanowire single-photon sensing of cerebral blood flow," *Neurophotonics*, vol. 8, no. 3, 2021.

[181] G. Bale, C. E. Elwell, and I. Tachtsidis, "From Jöbsis to the present day: a review of clinical near-infrared spectroscopy measurements of cerebral cytochrome-c-oxidase," *Journal of Biomedical Optics*, vol. 21, p. 091307, 5 2016.

[182] K. Khare, *Fourier Optics and Computational Imaging*. John Wiley & Sons Ltd., 2016.

List of publications and proceedings

Published manuscripts:

- F. Tonolini, **J. Radford**, A.Turpin, D. Faccio and R. Murray-Smith, "Variational Inference for Computational Imaging Inverse Problems," *Journal of Machine Learning Research* vol. 21, no. 179, pp. 1-46, 2020
- **J. Radford**, A. Lyons, F. Tonolini, and D. Faccio, "Role of late photons in diffuse optical imaging," *Optics Express*, vol. 28, no. 20, pp. 29486-29495, 2020.
- A. Turpin, V. Kapitany, **J. Radford**, D. Rovelli, K. Mitchell, A. Lyons, I Starshynov, and D. Faccio, "3D Imaging from Multipath Temporal Echoes," *Physical Review Letters*, vol. 126, no. 17, pp. 174301, 2021.
- M. G. Smith, **J. Radford**, E. Febrianto, J. Ramírez, H. O'Mahony, A. B. Matheson, G. M. Gibson, D. Faccio, and M. Tassieri, "Machine learning opens a doorway for microrheology with optical tweezers in living systems," *AIP Advances*, vol. 13, no. 7, pp. 075315, 2023.
- **J. Radford** and D. Faccio, "Information transport and limits of optical imaging in the highly diffusive regime," *Physical Review Research*, vol. 5, no. 2, 2023.

Published proceedings:

- **J. Radford**, A. Lyons, F. Tonolini, and D. Faccio, "The Role of Late Photons in Time-of-Flight Diffuse Optical Tomography," in Imaging and Applied Optics Congress, OSA Technical Digest (Optica Publishing Group, 2020), IF2E.5., 2020
- **J. Radford**, S. Nerenberg, A. Shvarts, L. Kaczmarcyk, and D. Faccio, "Imaging Beyond 100 Transport Mean Free Paths using Variational Autoencoding," *Imaging and Applied Optics Congress 2022 (3D, AOA, COSI, ISA, PcAOP)*, CM3A.5, 2022.
- **J. Radford**, S. Nerenberg, and D. Faccio, "Transmission of Photons Through an Adult Human Head," *in Imaging and Applied Optics Congress 2022 (3D, AOA, COSI, ISA, pcAOP)*, CM3A.2, 2022.

- V. Kapitany, A. Turpin, **J. Radford**, D. Rovelli, A. Lyons, I Starshynov, and D. Faccio, "Inferring spatial scenes from their time-resolved multipath echoes," *in 2021 Conference on Lasers and Electro-Optics Europe and European Quantum Electronics Conference, CLEO/Europe-EQEC*, 2021.

- D. Faccio, A. Turpin, **J. Radford**, and I. Starshynov, "Deep-diffuse, time-resolved imaging from an information perspective," *in Advanced Photon Counting Techniques XV*, 2021.

- **J. Radford** and D. Faccio, "Information Theory Analysis of Space-time Measurements and the Limits of Diffuse Imaging," *in OSA Imaging and Applied Optics Congress 2021 (3D, COSI, DH, ISA, pcAOP)*, CW3B.4, 2021.

- **J. Radford** and D. Faccio, "Information Analysis and Limits of Imaging through Complex Media," *in 2021 Conference on Lasers and Electro-Optics Europe and European Quantum Electronics Conference, CLEO/Europe-EQEC*, 2021.

- V. Kapitany, A. Turpin, **J. Radford**, D. Rovelli, A. Lyons, I. Starshynov, and D. Faccio, "Inferring spatial scenes from their time-resolved multipath echoes,"*in 2021 Conference on Lasers and Electro-Optics Europe and European Quantum Electronics Conference, CLEO/Europe-EQEC*, 2021.