Imperial College London

Monte Carlo Modelling of QED Interactions in Laser-Plasma Experiments

Robbie Alexander Watt

April 2021

Submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Physics of Imperial College London and for the Diploma of Imperial College London

Department of Physics
Imperial College London
Prince Consort Road
SW7 2AZ
United Kingdom

Declaration of Originality

I, the author, declare that the material presented in this thesis is my own work. Exceptions to this are acknowledged or referenced within the text.

Copyright Declaration

To my parents

Abstract

This thesis is concerned with the development of a start-to-end Monte Carlo simulation capability for laser based QED experiments.

A physics package has been developed for Geant4 which models particle-photon interactions, including the Breit-Wheeler process and photon-photon scattering. The interactions are treated as a beam of particles travelling through a static photon field. Gaussian process regression has been used to increase the event calculation rate by a factor of ~ 1000.

A strong field QED event generator that models the nonlinear Breit-Wheeler process and nonlinear Compton scattering has been developed and integrated into Geant4. Deep learning has been used to emulate this package and increases the calculation rate by a factor of $\sim$ 1000, allowing the package to be used as a forward model for Bayesian inference to aid the design and analysis of strong field QED experiments.

The tools developed were used to design and analyse a photon-photon collider experiment at the Gemini laser facility. This collided a $\sim 50\,\mathrm{fs}$ beam of gamma rays ($> 100\,\mathrm{MeV}$) with a $\sim 40\,\mathrm{ps}$ beam of x-rays ($\sim$ 1.5 keV). Using optimum beam conditions, simulations predict a signal-to-noise ratio of ~ 0.1, meaning a statistically significant measurement of the Breit-Wheeler process could be made with $\sim$ 100 shots. However, during the experiment the electron beam obtained was sub-optimal, reducing the signal-to-noise ratio to $\sim 2 \times 10^5$, making it unlikely that a measurement was possible.

No signatures of elastic photon-photon scattering were detected in the experiment, enabling a bound of $4.9 \times 10^{12}\,\mu\mathrm{b}$ to be placed on the photon-photon scattering cross-section at a centre of mass energy $\omega_{\mathrm{CM}} \approx 0.66\,m_e$. While still far above the predicted QED cross section, this is the most stringent bound placed to date, improving upon previous work by three orders of magnitude.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost I would to thank my supervisor, Stuart, for his continued support and advice throughout my four year struggle. I would also like to thank Zulf and Steve for all the insightful chats that have been pivotal to this work. Many thanks also go out to the whole of the TexMex and laser plasma groups.

The past four years (bar the write up) have been thoroughly enjoyable, made all the better by some of the great people I've met. I'd like to give special thanks to my office and in particular Nic and Luke (despite the absence of my name in his acknowledgements). I'm going to miss the nights spent in $\hbar$ and all the (sometimes overly) heated debates.

On a personal note, I would like to thank some of my close friends (the 63 Searles Close household) and family (Jamie, Katie, Mum and Dad). Finally I would like to thank Laura. Without your love, help and support (and proofreading) this work would not have been possible.

Chapter 1

Introduction

Electromagnetic phenomena is a subject that has captivated scientists for millennia. Since the technological revolution of the 19th century, gaining the ability to harness electromagnetism has had an immeasurable impact on society. Given its omnipresence today, one would be forgiven for believing the theory is fully understood. However, there are still unanswered questions regarding electromagnetic phenomena at the extremes, such as how a charge interacts with a strong electromagnetic field. These types of interaction are the subject of this work, and the theoretical tool we will use to study them is quantum electrodynamics (QED).

We will begin by exploring how QED came to fruition, starting with a short history of classical electromagnetism. We will see that inadequacies in the theory lead to the quantum revolution which dominated physics in the 20th century. Out of this revolution came QED, which is the most stringently tested theory in physics to date.

1.1 A Brief History of Classical Electromagnetism and Quantum Mechanics

Electromagnetism has not always been understood as a universal theory, with electric and magnetic forces traditionally viewed as two distinct phenomena. Although it has been suggested that its discovery occurred as early as 1000 BC [1], the first mention of magnetism in literature dates back to a 4th-century BC Chinese book [2]. This writing simply refers to the ability of magnetised rocks (known as lodestones) to attract iron, however, in the following centuries these rocks started to be used as a navigational tools through the invention of the compass. The fascination with electric phenomena likely dates back much later due to naturally occurring events such as lightning

Beyond its application for navigation, electricity and magnetism were seen as little more than a curiosity. However, this changed during the 17th century as scientists discovered ways of generating electricity. This quickly lead to studies in electrostatics, with

Joseph Priestley realising that the force between electric charges exhibits the same behaviour as the gravitational force between massive objects (*i.e.* the force diminishes with the inverse square of the distance between the charges or masses) [3]. This was proved quantitatively by Charles-Augustin de Coulomb in 1785 [4].

The first insight that there may be a link between electric and magnetic phenomena, came in 1820, when Hans Christian Ørsted discovered that the act of moving a current carrying wire near a compass would temporarily deflect the needle [5]. This was further studied by André-Marie Ampère, culminating in the development of a mathematical description, known today as Ampère's law. Not long after, the opposite process (converting magnetism into electricity) was confirmed by Michael Faraday, and in doing so he created the first electric dynamo [6]. Faraday went on to perform over 16,000 experiments throughout his career [5], making him one of the greatest experimental physicists of the 19th century.

With the groundwork set, it would take one of the greatest theoretical physicists of the 19th century, James Clerk Maxwell, to fully unite electricity and magnetism with a mathematical theory. Applying his knowledge of fluid mechanics, Maxwell proposed the existence of the electromagnetic field. He described the source and dynamics of this field by a set of equations, known today as Maxwell's equations. This led Maxwell to introduce the revolutionary concept of electromagnetic waves and calculated that they travel at the speed of light, c. From this, he concluded that light itself was an electromagnetic wave, stating that [7]

> "we can scarcely avoid the inference that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena."

By the late 19th century, Maxwell's theory of electromagnetism had become well established. At this time, Lord Rayleigh and Sir James Jeans were applying electromagnetic theory to study the radiation emitted from a body (known as black-body radiation). They knew the source of this radiation was oscillating electric charges within the object's material and, by applying thermodynamical principles, derived the following equation for the spectrum of emitted radiation

$$B_\lambda(T) = \frac{2c\,k_\mathrm{B}T}{\lambda^4} \tag{1.1}$$

where λ is the wavelength of the radiation, T is the temperature of the body and k_B is Boltzmann's constant. This is known as the Rayleigh-Jeans law and is plotted in figure 1.1. This spectrum agreed with experimental observations at low frequencies, however, diverges at high frequencies. This was clearly unphysical and later became known as the Rayleigh-Jeans catastrophe.

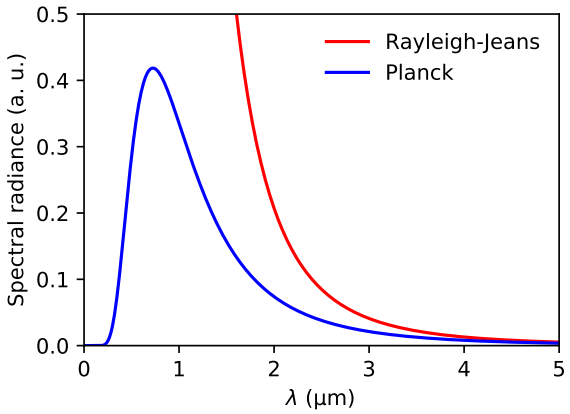


Figure 1.1: Plots of the spectra radiance of a black body, predicted by Rayleigh-Jeans (red) and Planck (blue).

At the same time Max Planck was also studying the thermal radiation from a black-body. However, rather than using a theoretically approach, he tackled the problem heuristically, by attempting to fit functions to experimental data. From this he obtained the following law

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5} \frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right) - 1} \tag{1.2}$$

where h is Planck's constant. This is also plotted in figure 1.1 and does not exhibit divergent behaviour.

Following the success of his experimental fit, Planck set to develop a theoretical understanding behind it. He started by attempting to apply thermodynamic principles, however, reluctantly had to accept a statistical mechanics approach to succeed. The theory came with the unexpected consequence that the radiation was emitted in discrete quanta (photons) of energy, $E = h\nu$. Soon after, Albert Einstein developed the idea of photons, using it to explain the photoelectric effect [8].

This led to the rapid development of quantum mechanics during the 1920s. Given the wave-particle duality exhibited by photons, Louis de Broglie proposed that matter particles (*e.g.* electrons) may exhibit wave-like properties. If this was the case, then there should be a wave equation describing their motion. In 1926, Austrian physicist Erwin Schrödinger was the first to derive such an equation, which described the behaviour of the wavefunction, Ψ [9]. Initially, the nature of Ψ was not fully understood, until Max Born interpreted it as the probability amplitude, whose modulus squared is equal to probability density. With this, quantum mechanics was Born.

1.2 Quantum Electrodynamics

Schrödinger applied his equation to calculate the spectral energies of the hydrogen atom, and his results agreed well with experimental measurements [9]. However, the equation was not consistent with special relativity. The solution to this problem came in 1928, when Paul Dirac derived a relativistically consistent wave equation for the electron [10]. When solving his equation, Dirac obtained both physical positive energy states and unphysical negative energy states. To get around this problem, he proposed the concept of the Dirac sea [11]. Here, he treated the vacuum as the state in which all negative energy states are filled in accordance with Pauli's exclusion principle. A hole in a negative energy state would respond to an electromagnetic fields as if it were a positively charged particle. Dirac initially believed these holes to be protons, however, he soon realised that they must have the same mass as the electron, giving rise to the 'anti-electron'. These anti-electrons were then discovered experimentally by Carl Anderson in 1932, who later named them positrons [12].

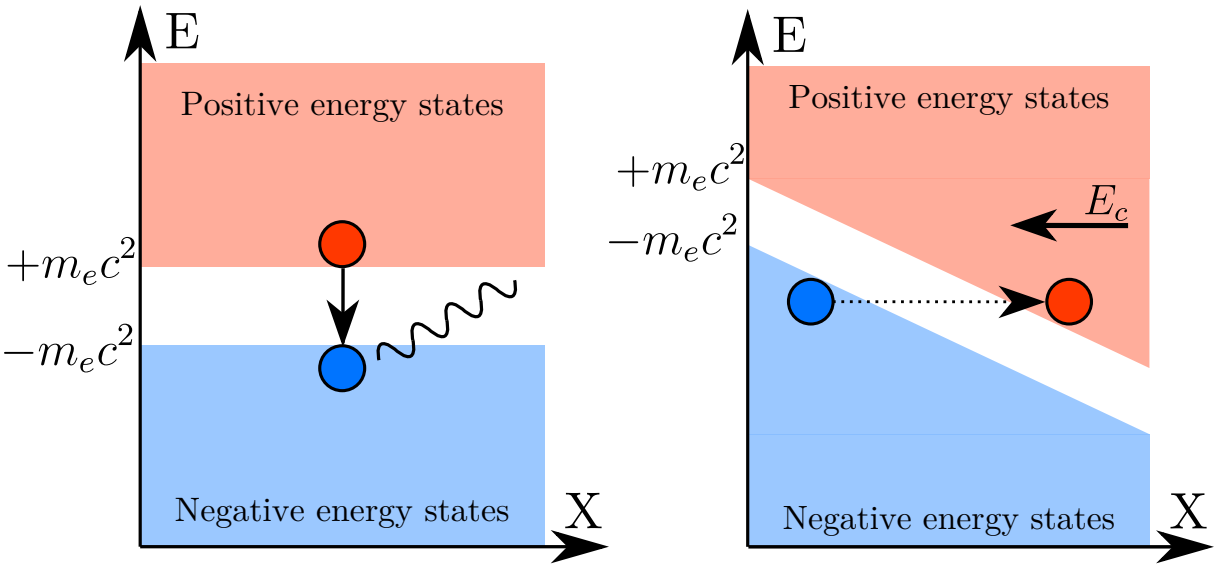
Dirac's theory predicted that an electron in a positive energy state could transition into an unoccupied negative energy state, resulting in the emission of electromagnetic radiation. This corresponds to electron-positron annihilation and is demonstrated in figure 1.2 (a). Following this, came a flurry of activity into more processes involving the electron. Work carried out by Hans Bethe and Walter Heitler [13] and others [14, 15, 16] showed that electron positron pairs could be produced through the annihilation of a photon in the nuclear field of an atom ($Z\gamma \rightarrow Ze^+e^-$). At the same time, Lev Landau and Evgeny Lifshitz [17] shown that the pairs could also be produced by the collision of two nuclei ($ZZ \rightarrow ZZe^+e^-$). However, perhaps the most striking prediction came from work carried out by Gregory Breit and John Wheeler [18]. By studying the impact of two light quanta on an electron in a negative energy state, they showed the possibility of creating electron-positron pairs from the vacuum.

During the same period, Fritz Sauter was studying the effect of a constant electric field on the Dirac sea [19]. He suggested that if the field was sufficiently strong, then a negative energy state electron could tunnel into a positive energy state, as demonstrated in figure 1.2 (b). The field strength at which this tunnelling would become significant is

$$E_{\mathrm{c}} = \frac{m_e^2 c^3}{e\hbar} = 1.326 \times 10^{18}\,\mathrm{Vm}^{-1} \tag{1.3}$$

and is known as the Schwinger field after Julian Schwinger who gave a complete theoretical description of the process in 1951 [20]. This was the first description of a nonperturbative QED interaction, with more nonperturbative phenomena, such as vacuum polarisation, studied soon after by Heisenberg and Euler [21, 22].

Despite the early success of QED, issues quickly arose in the form of infinities when



(a) Dirac annihilation

(b) Schwinger mechanism

Figure 1.2: Diagrams of the Dirac sea, showing the positive and negative energy states separated by a gap of $2m_ec^2$. (a) shows Dirac annihilation as a electron fills a negative energy state, resulting in the emission of photons. (b) shows a strong field applied to the sea, resulting in the tunnelling of an electron in a negative energy state to a positive energy state.

attempting to perform perturbative calculations beyond first order. On top of this, in the 1940s more precise measurements of the Lamb shift and magnetic moment of the electron showed discrepancies with the theory. It was not until the end of the 1940s that the problem was solved through the process of renormalisation, with significant contributions by Shin'ichiro Tomonaga [23], Richard Feynman [24, 25, 26], Schwinger [27, 28, 29] and Freeman Dyson [30, 31]. This enabled perturbative QED calculations to be performed to any order. Since then QED has become one of the most successful theories in physics, acting as a blueprint for other quantum field theories (*e.g.* quantum chromodynamics) and providing the most accurate predictions of experimental measurements (*e.g.* electron magnetic moment [32]) to date.

Despite certain aspects of QED being incredibly well tested, there remains little to no experimental evidence of other QED predictions. Examples include the Breit-Wheeler process, photon-photon scattering and Schwinger pair production. The reason for this is the extreme energies or field strengths required for the interactions to take place. However, such conditions are routinely found in many astrophysical environments.

For example, field strengths close to or exceeding that of the Schwinger field are thought to occur in both extremely short lived events, such as gamma ray bursts [33] and supernova explosions [34], as well as more stable objects such as the magnetosphere of strongly magnetised neutron stars [35] and the accretion disks of black holes [36].

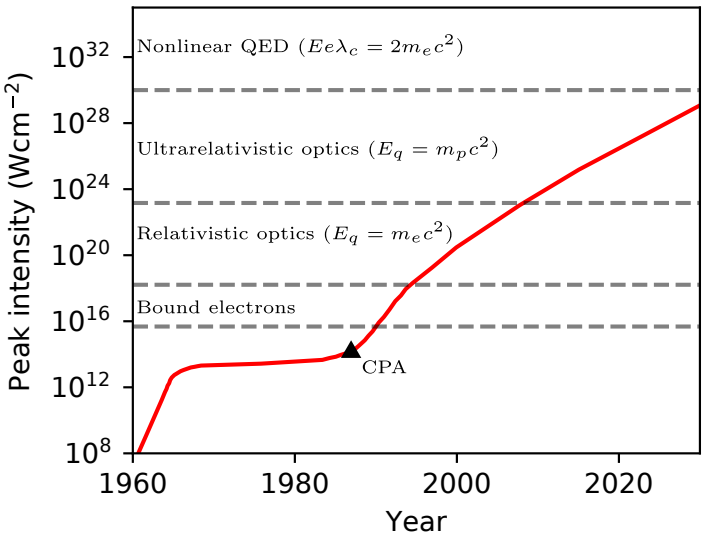


Figure 1.3: Peak laser intensity against year. Adapted from [45].

This gives rise a number of nonlinear QED effects, including pair production [37, 38, 39], quantum synchrotron emission [35] and vacuum birefringence [40].

Another aspect of astrophysics that gives rise to exotic QED interactions is the massive scale of the universe. Particles propagating through the intergalactic medium travel vast distances, giving time for even highly unlikely interactions to become significant. An example of this is the high energy ($< 100\,\text{GeV}$) cut-off in the cosmic gamma ray spectrum observed at Earth [41, 42] . This is due to these gamma rays annihilation with the cosmic microwave background. This mechanism was thought to be well understood, however, recent observations have found a larger level of high energy photons from quasar 3C 279 reaching Earth that expected from QED calculations [43]. This demonstrates the need for more experimental work, to better understand these astrophysical environments.

1.3 Laser Based QED Experiments

To perform a laboratory study of exotic QED interactions found in astrophysics requires a method of producing strong electromagnetic fields. These can be achieved using a high power laser, which have seen an exponential growth in peak intensity over the past few decades, as shown in figure 1.3. This was driven by the invention of chirped-pulse amplification (CPA), for which Donna Strickland and Gerard Mourou were awarded the 2018 Nobel prize [44].

A useful parameter when discussing high power lasers is the normalised vector potential

$$a_0 = \frac{eE_0}{m_e c\omega} = 0.60\left(\frac{I}{10^{18}\text{Wcm}^{-2}}\right)^{1/2}\left(\frac{\lambda}{\mu\text{m}}\right). \tag{1.4}$$

The significance of this parameter is that if $a_0 = 1$ (corresponding to an intensity of $\sim 10^{18}\,\mathrm{Wcm}^{-2}$ for a wavelength of $1\,\mu$m), an electron in the field will gain an energy of $m_e c^2$ over the wavelength of the laser. This leads to the nonlinear motion of the electron as the magnetic force becomes comparable to the electric force. Reaching this regime has led to a number of all-optical, high energy radiation sources, that can be used to study QED interactions. These include: electron beams through laser wakefield acceleration (LWFA) [46, 47, 48]; ion beams through laser-solid interactions [49, 50, 51]; x-ray betatron radiation from a LWFA [52]; and gamma ray radiation from Compton scattering [53] and bremsstrahlung emission [54].

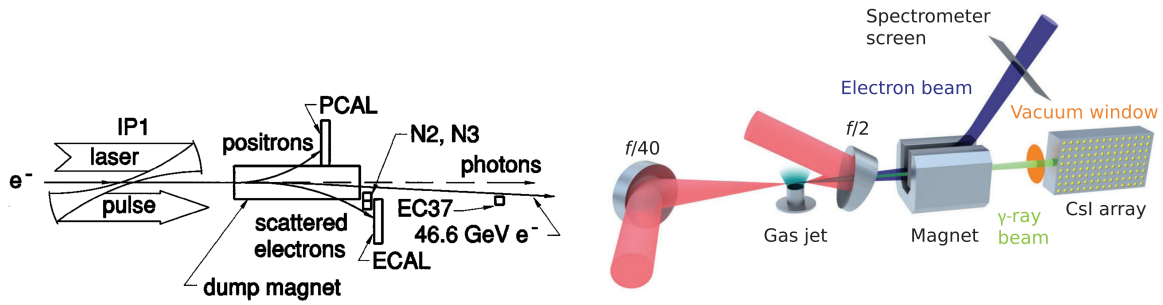
Despise these advances, we can see from figure 1.3 that laser intensities are still several orders of magnitude off the regime where nonlinear QED effects become important. At current trends we are unlikely to reach this for some years to come. However, the electromagnetic field is not a Lorentz invariant. It is therefore useful to define the following parameters

$$\eta = \frac{\sqrt{-(F_{\nu\mu}p^{\mu})^2}}{mcE_c} \qquad \chi = \frac{\hbar\sqrt{-(F_{\nu\mu}k^{\mu})^2}}{2mcE_c} \tag{1.5}$$

for electrons/positrons and photons respectively. Here, $F_{\nu\mu}$ is the electromagnetic tensor, p^{μ} is the four momentum of an electron and k^{μ} is the four wavevector of a photon. These are often referred to as quantum parameters, with $\eta, \chi > 1$ characterising the onset of nonlinear QED interactions becoming dominant. We can view η as the observed field strength in the rest frame of an electron in units of E_c. Although there is no rest frame for a photon, χ is defined in a similar way. Both η and χ depend linearly on the field strength and momentum of the particle. Therefore, by firing a relativistic beam of particles towards a counter-propagating laser, it is possible to observe strong field QED phenomena with a laboratory field strength several orders of magnitude lower than E_c.

The first attempt to probe the nonlinear QED regime occurred during the E-144 experiment at the Stanford Linear Accelerator (SLAC) in 1996. Here, they accelerated a beam of electrons to 46.6 GeV before colliding it with a laser of strength $a_0 \approx 0.3$, resulting in a quantum parameter of $\eta \approx 0.2$. A diagram of the experimental setup is shown in figure 1.4 (a). As the electrons interacted with the field, a beam of high energy gamma rays was emitted, providing the first evidence of nonlinear Compton scattering [55]. The experiment was also able to provide the first evidence of the nonlinear Breit-Wheeler process [56], as some of the emitted gamma rays interacted with the field to produce electron-positron pairs. On average $n \approx 6$ laser photons were absorbed per scattering or pair production event. As the experiment operated with $a_0, \chi < 1$, these nonlinear effects were highly suppressed, resulting in only ≈ 100 pairs generated over 22000 shots.

We can see from figure 1.3, that peak laser intensities have increased significantly since the mid 90s. This has lead to a renewed interest in studying strong field QED interactions, with recent experiments carried out by Cole *et al.* (2018) [57] and Poder *et.al* (2018) [58]



(a) Burke *et al.*, SLAC (1997) [56] (b) Cole *et al.*, Gemini (2018) [57]

Figure 1.4: Diagrams of strong field QED experiment setups. (a) shows the setup of the E-144 experiment at SLAC. (b) shows the setup of the radiation reaction experiment at Gemini. Although the quantum parameter of experiments has not improved over two decades, diagram making software clearly has.

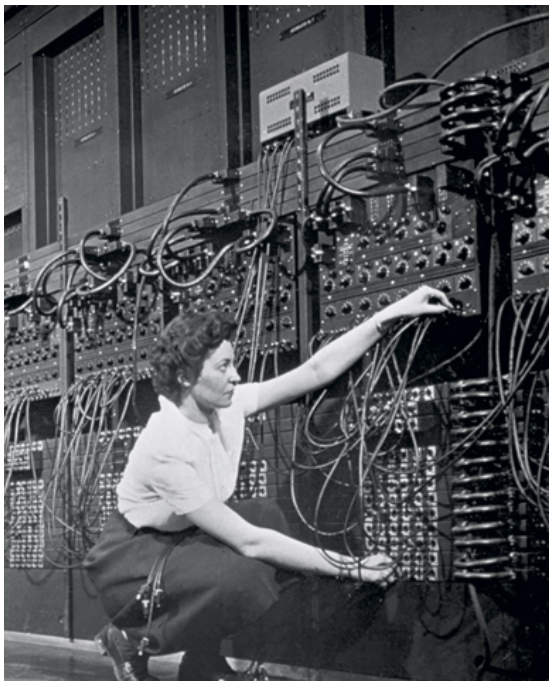
using the Gemini laser facility. These experiments were focused on the back reaction that a charge experiences due to the emission of radiation (known as radiation reaction) and a diagram of the setup can be found in figure 1.4 (b). To do this, they used a LWFA to accelerate an electron beam to $\sim 2\,\text{GeV}$ (significantly lower than E-144) and collide it with a tightly focused laser pulse of of strength $a_0 \approx 10$ (significantly higher than E-144). This resulted in a similar nonlinear quantum parameter of $\eta \approx 0.2$. In both experiments radiation reaction was observed, however, large uncertainties were experienced due to shot-to-shot fluctuations.

Following the success of E-144 and recent Gemini experiments, new strong field QED campaigns are being planned which will take advantage of higher energy facilities, including FACET-II at SLAC [59], and LUXE at the European X-FEL [60, 61]. With high energy electrons and more intense lasers, these experiments hope to probe closer to the regime where strong field QED processes become dominant.

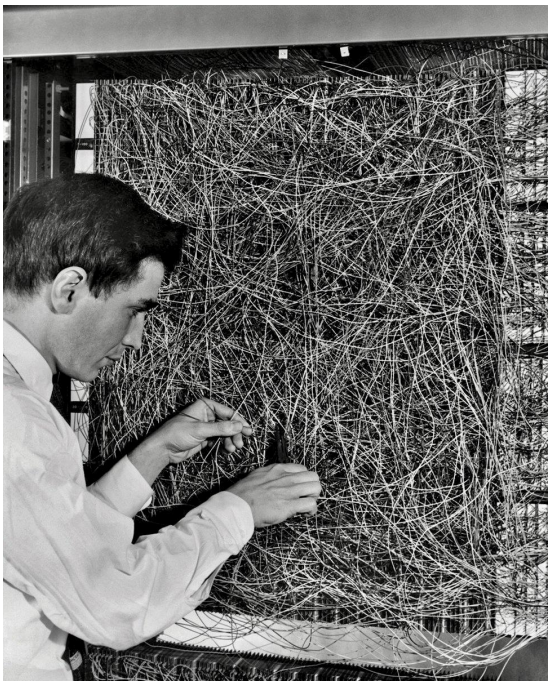
1.4 Monte Carlo and Machine Learning

A typical laser based QED experiment will have a low expected signal (*e.g.* E-144 with 100 pairs over 22000 shots) and due to the presence of high energy gamma rays, the background noise levels will be high. This makes carrying out detailed numerical modelling of such experiments vital, to maximise both the signal-to-noise ratio and chance of a successful result. Due to the large scale of the simulations, and the inherent randomness of the interactions, the Monte Carlo method is ideal for performing such calculations.

Monte Carlo methods are a group of algorithms that use repeated random sampling to obtain results. The first person to realise their potential to carry out simulations of physical systems was the Polish-American scientist Stanislaw Ulam. During the 1940s, whilst playing games of solitaire, Ulam attempted to calculate the probability of winning



(a) ENIAC



(b) Perceptron

Figure 1.5: (a) shows Fran Bilas programming the first electronic general-purpose digital computer named ENIAC [65]. (b) shows Frank Rosenblatt with the Mark I Perceptron machine [66].

given a randomly shuffled deck of cards. Through his own calculations, he was unable to find the answer, however, his genius was to envision the ability of new electronic computers to rapidly simulate games and use this to calculate probabilities [62]. Ulam quickly realised that this same method could be applied to a physical problem he was working on, neutron diffusion. After discussions with John von Neumann, the first Monte Carlo simulation was carried out using the ENIAC computer at Los Alamos [63], shown in figure 1.5 (a). By solving neutron diffusion, the goal of the simulation was to predict the explosive power of a fission device. Due to the secretive nature of their work, the project was given the code name Monte Carlo, in reference to the Monte Carlo casino in Monaco [64].

Since its invention, the Monte Carlo method has become ubiquitous in nuclear and particle physics. This had lead to the development of a number of open source, Monte Carlo simulation packages with examples including Geant4 [67, 68, 69], FLUKA [70, 71] and MCNP [72]. Many of the simulations performed in this work have been carried out using the package Geant4. Standing for 'GEometry ANd Tracking', Geant4 models the passage of high energy particles through matter. It is built using C++ and heavily relies on the object-oriented approach, making it easy to extend the package to our problem.

Performing Geant4 simulations of QED experiments is time consuming and compu-

tationally expensive. Therefore, in this work we have used machine learning algorithms to greatly increase the efficiency of these calculations. Like Monte Carlo methods, machine learning falls under the topic of computational statistics, and also has its roots in scientists building machines to play games.

The term machine learning was coined in 1959 by Arthur Samuel [73], using it to describe an algorithm he developed to play checkers. However, Samuel was not the first to contemplate machine intelligence, with Alan Turing discussing the topic in his 1950 seminal paper 'Computing Machinery and Intelligence' [74]. In this work he proposed the Turing test, a set of questions a computer must pass to be considered intelligent. This gave rise to the topic of artificial intelligence (AI), which covers the attributes of computer learning, reasoning and problem solving. Machine learning is considered a subset of this, which specifically focuses on learning trends from data, without explicit programming.

Not long after Samuel's work, the first example of an artificial neural network, called the perceptron, was designed by Frank Rosenblatt at Cornell Aeronautical Laboratory [75]. The name perceptron was initially used to describe the custom built machine that the neural network algorithm ran on, which is shown in figure 1.5 (b). It contained an array of photocells, used to detect a 20×20 pixel image, randomly connected to 'neurons' which act like neurons within the brain. Each neuron has a weight in the form of a resistance, which is learned by rotary variable resistors called potentiometer, driven by electric motors [76]. This machine was used as a linear classifier and was capable of distinguishing between simple shapes or characters.

Beyond the initial success of the perceptron, there were few advances in the area of machine learning over the following decades. This was partially due to lack of funding and criticism by Minksy and Papert (1969) [77] claiming that multi-layer perceptrons, required to learn nonlinear mappings, would be difficult to train. However, this changed in the 80s and 90s, coinciding with the rapid increased in computational power and the rediscovery of the backpropagation algorithm [78].

Today, machine learning algorithms are typically divided into three distinct sets: supervised learning, unsupervised learning and reinforcement learning. A discussion of these can be found in ref. [76]. In this work, we will focus on regression algorithms which are a subgroup of supervised learning. Given a data set comprising of input vectors with corresponding target vectors, the goal of regression is to learn a mapping from the input to the target which generalises to unseen inputs. The two algorithms we will use are Gaussian process regression [79] and neural network regression [76].

1.5 Thesis Outline

The work in this thesis is concerned with the development of software capable of modelling laser based QED experiments. The contents of the subsequent chapters are as follows.

Chapter 2: Theory of Electrodynamics

We will review the theory of charged particles interacting with electromagnetic fields. Starting with a review of classical electrodynamics, we will find inconsistencies which require QED to solve. Finally, we will introduce QED effects in strong background electromagnetic fields.

Chapter 3: Machine learning and Monte Carlo Methods

We will review the computational methods used in this work. We will start by discussing the two common approaches to statistical inference which are maximum likelihood estimation and Bayesian inference. These methods will then be used to discuss both the Monte Carlo method and machine learning algorithms used throughout in this work.

Chapter 4: Modelling Particle-Photon Processes in Geant4

We will present the development of a new photon physics package for Geant4. This package includes the Breit-Wheeler process, photon-photon scattering and Compton scattering in a slowly varying, radiation field. We will show how a Gaussian process regression can be used to greatly increase the efficiency of the package. Finally, we will demonstrate an application, using the package to study the efficacy of a photon collider experiment using the radiation from a laser heated hohlraum.

Chapter 5: Modelling Strong Field QED Interactions

In this chapter we will present the development of a new strong field QED package which models the nonlinear Breit-Wheeler process and nonlinear Compton scattering in a background field. Applications of this package include pre-experimental design and optimisation and post-experimental inference. Through the development of a deep-learning emulator, we use the package to perform Bayesian inference on a mock radiation reaction experiment.

Chapter 6: Design and Analysis of a Photon-Photon Collider Experiment

We will use the photon physics package discussed in chapter 4 to model a photon-photon collider experiment carried out at the Gemini laser facility. Here, we will present calculations carried out both before and after the experiment, discussing how the setup was optimised and inference made on the experimental measurements.

Chapter 2

Theory of Electrodynamics

In this chapter we will discuss the theory of electrodynamics, which describes the interaction between charges and the electromagnetic field. We will take a historical path through the topic, starting with a discussion of the classical theory. This is capable of explaining most of the electromagnetic phenomena we observe, however, breaks down at small scales. Therefore, quantum electrodynamics (QED) is required to explain phenomena at the extremes. In the final part of this chapter, we will study QED in a strong field background, where nonlinear effects become dominant.

Before beginning, we will briefly mention the formalism and units system used throughout this chapter. As both classical and quantum electrodynamics are relativistic theories, they can be expressed in a manifestly covariant form. However, this adds an extra layer of complexity, when discussing classical electrodynamics. Therefore, in section 2.1 we will work predominately with the three-vector form of equations.

However, as we move on to discussing QED, we will present equations in a manifestly covariant form. Here we will use the same conventions as Jackson [80] and Peskin [81], in which a three-vector quantity is denoted using a boldface variable and a four-vector quantity is denoted using an italic variable with a Greek index, *e.g.*

$$\mathbf{a} = (a^1, a^2, a^3) \qquad a^\mu = (a^0, a^1, a^2, a^3). \tag{2.1}$$

The metric tensor is given by

$$\eta^{\mu\nu} = \mathrm{diag}(1, -1, -1, -1) \tag{2.2}$$

and the scalar product of two four-vectors is

$$a \cdot b = \eta^{\mu\nu} a_\mu b_\nu. \tag{2.3}$$

The four-vector derivative is

$$\partial_\mu = \Big(\frac{1}{c}\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\Big). \tag{2.4}$$

Finally, in section 2.1 we will use SI and all constants will be expressed explicitly. However, as we move to QED, to prevent equations become unwieldy a natural units system will be used. Here, we will set the following

$$\hbar = c = 1. \tag{2.5}$$

2.1 Classical Electrodynamics

The two fundamental building blocks of classical electrodynamics are the Lorentz force and Maxwell's equations. These describe a coupled system in which the action of the field on the charges is governed by the Lorentz force and the action of the charges on the field is governed by Maxwell's equations.

Any discussion of classical electrodynamics would of course be incomplete without a review of Maxwell's equations, therefore this is where this section will start. We will then proceed to study the dynamics of a charged particle in an external electromagnetic field. Here we will use the Lagrangian and Hamiltonian formalisms, as a similar approach will be taken when discussing QED. Having reviewed the foundations of classical electrodynamics, we will move on to discuss more advanced topics, including the emission of radiation from an accelerating charge and the back reaction that the charge experiences as a result. We will see hows this leads to a fundamental flaw in classical electromagnetism, one which can only be fixed by moving to a quantum theory.

2.1.1 Maxwell's Equations

If we define the electromagnetic field in terms of its electric, $\mathbf{E}$, and magnetic, $\mathbf{B}$, components, Maxwell's equations are given by

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\ \nabla \times \mathbf{B} &= \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \end{aligned} \tag{2.6}$$

where ρ is the charge density, $\mathbf{J}$ is the current density, and ϵ_0 and μ_0 are the permittivity and permeability of free space respectively. Alternatively, we can define the field in terms of the four-potential, $A^\mu = (\phi/c, \mathbf{A})$. Here, ϕ is the electric potential and $\mathbf{A}$ is the

magnetic vector potential. These are linked to the electric and magnetic fields through

$$\begin{aligned} \mathbf{B} &= \nabla \times \mathbf{A} \\ \mathbf{E} &= -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}. \end{aligned} \tag{2.7}$$

Using A^μ, we can write Maxwell's equations in a more condense, and covariant form

$$\partial^2 A^\mu = \mu_0 J^\mu \tag{2.8}$$

where $J^\mu = (c\rho, \mathbf{J})$ is the four-current.

2.1.2 Charged Particle Dynamics

A charge located inside an electromagnetic field will feel a force from the field and act back on it. In this section, we are only concerned with the dynamics of the particle, so will assume the charge is small and the back-action is of negligible effect. The equations of motion for the charge can be derived by applying Hamilton's principle of least action. This states that the path taken by a system, going from configuration 1 at time t_1 to configuration 2 at time t_2, is the one in which the action,

$$A = \int_{t_1}^{t_2} L[Q_i(t),\, \dot{Q}_i(t),\, t]\mathrm{d}t \tag{2.9}$$

is minimised ($\delta A = 0$). Here, L is the Lagrangian and depends on the generalised coordinates of the system, $Q_i(t)$, their time-derivatives, $\dot{Q}_i(t)$, and time explicitly. By applying the principle of least action to equation 2.9, we can derive the following [82, 80]

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{Q}_i}\right) - \frac{\partial L}{\partial Q_i} = 0 \tag{2.10}$$

which are known as the Euler-Lagrange equations of motion.

An equivalent description of the system can be obtained using the Hamiltonian formalism. Here, rather than defining the system using $\dot{Q}_i$, we use the canonical momentum,

$$P_i = \frac{\partial L}{\partial \dot{Q}_i}\,. \tag{2.11}$$

The Hamiltonian is defined through the Lagrangian

$$H = \mathbf{Q} \cdot \mathbf{P} - L \tag{2.12}$$

and is the total energy of the system. The motion of the system is now given by two,

first-order differential equations

$$\frac{\mathrm{d}P_i}{\mathrm{d}t} = -\frac{\partial H}{\partial Q_i} \qquad \frac{\mathrm{d}Q_i}{\mathrm{d}t} = \frac{\partial H}{\partial P_i} \tag{2.13}$$

which are known as Hamilton's equations. From them, we can see that if a generalised coordinate does not appear in the Hamiltonian, then it is a conserved property of the motion. This is also true of the Lagrangian and is a consequence of Noether's theorem.

We can now apply these methods to the system of interest, a relativistic particle in an external electromagnetic field (ϕ_{ext}, $\mathbf{A}_{\text{ext}}$). The generalised coordinates are the components of the particle's position, $\mathbf{Q} = (x, y, z)$, and their time derivatives are the components of the particle's velocity, $\dot{\mathbf{Q}} = (v_x, v_y, v_z)$. For this system, the Lagrangian is given by [80]

$$L = -\gamma^{-1}mc^2 + q\mathbf{v}\cdot\mathbf{A}_{\text{ext}} - q\phi_{\text{ext}} \tag{2.14}$$

where $\gamma = (1 - v^2/c^2)^{-0.5}$ is the Lorentz factor and q and m are the charge and mass of the particle respectively. The first term in equation 2.14 is the Lagrangian for a free particle and the second two terms account for the interaction with the field. Using equation 2.11 the conjugate momentum of $\mathbf{Q}$ is

$$\mathbf{P} = \gamma m\mathbf{v} + q\mathbf{A}_{\text{ext}} = \mathbf{p} + q\mathbf{A}_{\text{ext}} \tag{2.15}$$

where $\mathbf{p}$ is the ordinary kinetic momentum. By applying equation 2.12, and rewriting the result in terms of $\mathbf{P}$, the Hamiltonian of the system is given by [80, 83]

$$H = \sqrt{m^2c^4 + (\mathbf{P} - q\mathbf{A}_{\text{ext}})^2} + q\phi_{\text{ext}}. \tag{2.16}$$

The motion of the particle can now be obtained through either Euler-Lagrange equations or Hamilton's equations. This will not be carried out here but can be found in Landau and Lifshitz (1971) [83]. The final result, written in terms of the electric and magnetic fields, is the familiar Lorentz force

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = q(\mathbf{E} + \mathbf{v}\times\mathbf{B}). \tag{2.17}$$

From the Lorentz force we can also derive the rate of change of the particles energy, $E = \gamma mc^2$, due to the interaction with the field, through the definition of work

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \mathbf{v}\cdot\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = q\mathbf{E}\cdot\mathbf{v} \tag{2.18}$$

where we have used the vector property $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{a} = 0$.

The four-momentum of the particle is $p^\mu = (\frac{E}{c}, \mathbf{p})$, and allows us to combine equations

2.17 and 2.18 into a single manifestly covariant form

$$\begin{aligned}\frac{\mathrm{d}p^\mu}{\mathrm{d}\tau} &= \frac{q}{mc}(\partial^\mu A^\nu - \partial^\beta A^\mu)p_\nu \\ &= \frac{q}{mc}F^{\mu\nu}p_\nu\end{aligned} \tag{2.19}$$

where $\mathrm{d}\tau = \gamma \mathrm{d}t$ is the infinitesimal proper time and $F^{\mu\nu}$ is the electromagnetic field tensor.

2.1.3 Electron Motion in a Monochromatic Plane Wave

Having derived the general equations of motion for a charged particle in an arbitrary electromagnetic field, we will now look at the specific case of an electron in a monochromatic plane wave. This is a system which will come up multiple times throughout this chapter. Here, we will assume the wave is linearly polarised and propagating in the $+z$ direction. By making use of the Coulomb gauge ($\boldsymbol{\nabla} \cdot \mathbf{A} = 0$), in a vacuum we have $\phi = 0$. Therefore, the wave is fully defined by the vector-potential, $\mathbf{A} = (A_x,\, A_y,\, 0)$, which we will assume is a twice-differentiable function of the wave phase $\psi = kz - \omega t$. Here k is the wavenumber and ω is the wave frequency.

As the wave does not vary in either x or y, from equation 2.16 we can see the same is also true for the Hamiltonian. Therefore, the conjugate momenta P_x and P_y are constants of the electron's motion

$$\frac{\mathrm{d}}{\mathrm{d}t}(p_x - eA_x) = \frac{\mathrm{d}}{\mathrm{d}t}(p_y - eA_y) = 0. \tag{2.20}$$

If we set the boundary condition that the electron is initially at rest before the arrival of the wave, we can solve these equations to give the x and y components of the electron's momentum

$$p_x = eA_x \qquad p_y = eA_y. \tag{2.21}$$

The field does vary in z however, and the conjugate momentum, P_z, is not a constant. To find the motion of the electron along this direction we can solve the Euler-Lagrange equation, giving

$$\frac{\mathrm{d}p_z}{\mathrm{d}t} = e\mathbf{v} \cdot \frac{\partial \mathbf{A}}{\partial z} \tag{2.22}$$

where we have used $A_z = 0$. By applying the chain rule to the right hand side of this equation, we can write it as

$$e\mathbf{v} \cdot \frac{\partial \mathbf{A}}{\partial z} = e\mathbf{v} \cdot \frac{\partial \mathbf{A}}{\partial t}\frac{\partial t}{\partial z} = \frac{e}{c}\mathbf{v} \cdot \mathbf{E} = \frac{1}{c}\frac{\mathrm{d}E}{\mathrm{d}t} \tag{2.23}$$

where we have used equation 2.7 and equation 2.18 in the third step. This reveals a

second constant of the electron's motion

$$\frac{\mathrm{d}}{\mathrm{d}t}(E - cp_z) = 0. \tag{2.24}$$

As the electron is stationary before the arrival of the wave, its initial energy and z component of momentum are $E = mc^2$ and $p_z = 0$ respectively. Using this we can solve equation 2.24 to give

$$p_z = \frac{p_x^2 + p_y^2}{2mc} = \frac{e\,\mathbf{A}^2}{2mc} \tag{2.25}$$

where we have used $\gamma = \sqrt{1 + (\frac{p_x}{mc})^2 + (\frac{p_y}{mc})^2 + (\frac{p_z}{mc})^2}$. By introducing the parameter $\mathbf{a} = e\mathbf{A}/mc$, known as the normalised vector potential, we can rewrite the three components of the electron's momentum as

$$p_x = a_x\, mc \qquad p_y = a_y\, mc \qquad p_z = \frac{\mathbf{a}^2}{2}\, mc. \tag{2.26}$$

We can see this parameter plays a significant role in the dynamics of the electron. If $|\mathbf{a}| > 1$, the electron gains and loses an energy of $E > mc^2$ every wave period and its motion is dominated by relativistic effects.

The electron gains a momentum component both perpendicular (x and y) and parallel (z) to the direction of the wave. The perpendicular component is proportional to $\mathbf{a}$, resulting in the electron quivering in the field, but the time averaged motion is zero. However, the parallel component is proportional to $\mathbf{a}^2$, which results in the electron drifting in the direction of the field. The average velocity which the electron drifts at is given by

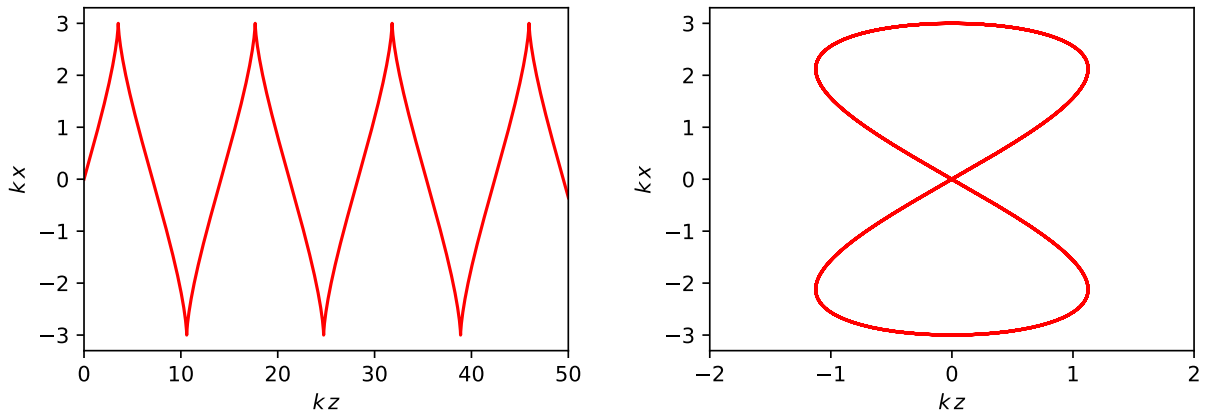
$$\frac{v_D}{c} = \left\langle \frac{p_z}{\gamma} \right\rangle = \frac{\mathbf{a}^2}{4 + \mathbf{a}^2}. \tag{2.27}$$

The full trajectory of the electron can be obtained by integrating the equations in 2.26. If we assume the field is sinusoidal and polarised in the x direction, it can be written as

$$\mathbf{a} = a_0 \cos\psi\, \hat{\mathbf{x}} \tag{2.28}$$

and there is no motion in the y direction. The trajectory in the x and z directions are given by the following

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= \frac{c}{\gamma}\, a_0 \cos\psi \\ \frac{\mathrm{d}z}{\mathrm{d}t} &= \frac{c}{2\gamma}\, a_0^2 \cos^2\psi \end{aligned} \tag{2.29}$$



(a) Laboratory frame

(b) Drift frame

Figure 2.1: Electron motion in a plane wave with $a_0 = 3$ as observed in the (a) laboratory frame and (b) average rest frame.

which, upon integrating yields

$$\begin{aligned} x &= \frac{a_0 c}{\omega} \sin\psi \\ z &= \frac{a_0^2 c}{4\omega}\left(\psi + \frac{1}{2}\sin 2\psi\right) \end{aligned} \tag{2.30}$$

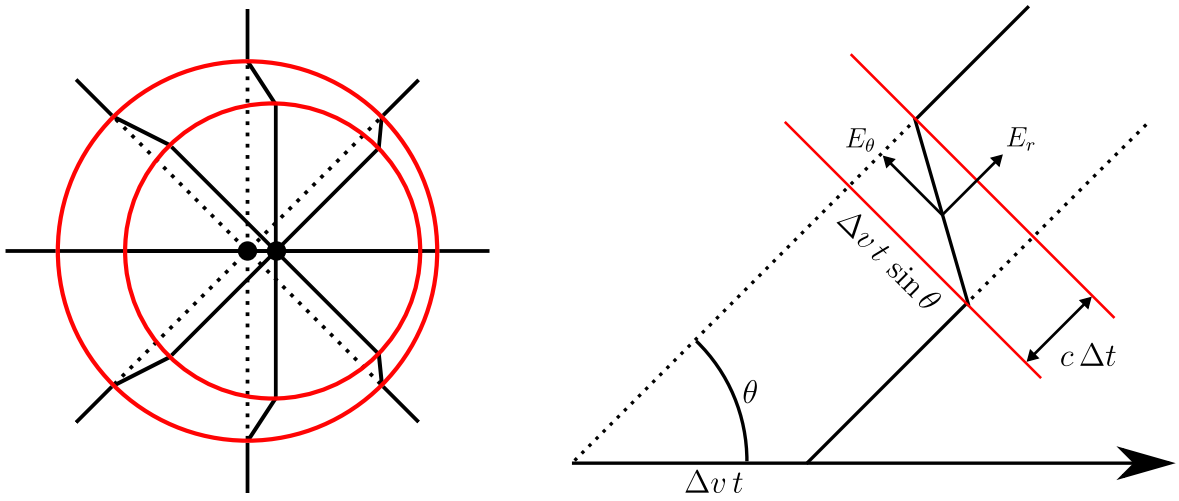
and is plotted in figure 2.1 (a). This shows both the quiver and drift motion of the electron in the laboratory frame. The average rest frame of the electron is obtained by boosting with a velocity v_D in the z direction. The trajectory of the electron in this frame is shown in figure 2.1 (b) with both components exhibiting simple sinusoidal motion. However, the z-component oscillates at twice the frequency, resulting in a figure-of-eight track.

2.1.4 Power Radiated from an Accelerating Charge

In the previous section, we assumed the charge was small and had no effect on the electromagnetic field. However, this is not always the case, as is evident from Maxwell's equations. In this section we will examine the action of a charge on the electromagnetic field and deriving a formula for the power emitted. We will use two approaches to tackle this problem. In the non-relativistic limit, an elegant pictorial description can be used, however, in the relativistic limit, a more mathematically rigorous approach is required.

Non-Relativistic Radiated Power

This pictorial derivation, which is shown in figure 2.2, was originally presented by J. J. Thomson, however, here we will follow the discussion given by M. Longair [84]. The diagram shows a stationary charge at the origin at time $t = 0$. The initial electric field configuration due to the charge is represented by the dashed lines. The charge then



(a) Accelerating charge (b) Field components

Figure 2.2: Diagram demonstrating the emission of radiation from an accelerating charge. (a) shows the field configurations of a charge initially at rest at the origin (dashed lines), and moving to the right at δv (solid lines). The red rings show the boundaries between the field configurations before, during and after the acceleration. (b) shows an enlarged picture of the field configuration inside the shell. The field is decomposed into radial and tangential components.

experiences an acceleration to a velocity $\delta v \ll c$ over a time interval Δt. After some time t, we can distinguish between the field configuration inside and outside of a sphere of radius $r = ct$. Inside the sphere, the field lines point radially out from the charge. However, outside the sphere the field configuration is not yet aware that the charge has moved, as information cannot travel faster than the speed of light. Therefore, the field lines still point radially out from the origin. Between the two configurations there is a shell of thickness $c\Delta t$ where the field lines must join up. Within this shell, there is a tangential component of the electric field, propagating out at the speed of light. This corresponds to the emission of a pulse of electromagnetic radiation.

In figure 2.2 (b) we can see the two components of the field inside the shell. Using this, the ratio of the tangential and radial components of the electric field is

$$\frac{E_\theta}{E_r} = \frac{\Delta v\, t\, \sin\theta}{c\Delta t}. \tag{2.31}$$

By replacing the radial component with Coulomb's law, the tangential component can be written as

$$E_\theta = \frac{q|\dot{\mathbf{v}}|\, \sin\theta}{4\pi\epsilon_0 c^2 r} \tag{2.32}$$

where $|\dot{\mathbf{v}}| = \Delta v/\Delta t$ is the acceleration of the charge. As there is a tangential electric field which changes over the pulse duration, Maxwell's equations tell us there must also be an azimuthal magnetic field, $B_\phi = E_\theta/c$. The energy flux in the field is given the Poynting

vector

$$\mathbf{S} = \frac{\mathbf{E} \times \mathbf{B}}{\mu_0} = \frac{q^2 |\dot{\mathbf{v}}|^2 \sin^2\theta}{16\pi^2 \epsilon_0 c^3 r^2} \hat{\mathbf{r}} \tag{2.33}$$

which, upon integrating over solid angle, gives the formula for total power emitted

$$P = \frac{q^2 |\dot{\mathbf{v}}|^2}{6\pi\epsilon_0 c^3} \tag{2.34}$$

and is known as the Larmor formula [85].

Relativistic Power

For a more mathematically rigorous approach, which correctly predicts the emission from a relativistic charge, we should solve Maxwell's equations for the system. Doing so is a fairly involved calculation and can be found in Jackson [80]. The result is the Liénard-Wiechert fields [86]

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \frac{q}{4\pi\epsilon_0} \left[\frac{\hat{\mathbf{n}} - \boldsymbol{\beta}}{\gamma^2 (1 - \boldsymbol{\beta} \cdot \hat{\mathbf{n}}) R^2} + \frac{\hat{\mathbf{n}} \times [(\hat{\mathbf{n}} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}}]}{c(1 - \boldsymbol{\beta} \cdot \mathbf{n})^3 R} \right]_{\text{ret}} \\ \mathbf{B}(\mathbf{r}, t) &= \frac{\hat{\mathbf{n}} \times \mathbf{E}(\mathbf{r}, t)}{c}. \end{aligned} \tag{2.35}$$

where $\boldsymbol{\beta} = \mathbf{v}/c$ is the normalised velocity of the charge, $\mathbf{r}_0$ is the location of the charge, $\mathbf{r}$ is the location of the observer, $\hat{\mathbf{n}}$ is the unit vector in the $\mathbf{r} - \mathbf{r}_0$ direction, $R = |\mathbf{r} - \mathbf{r}_0|$ and the right hand side of the equation is evaluated at the retarded time $t_r = t - R/c$.

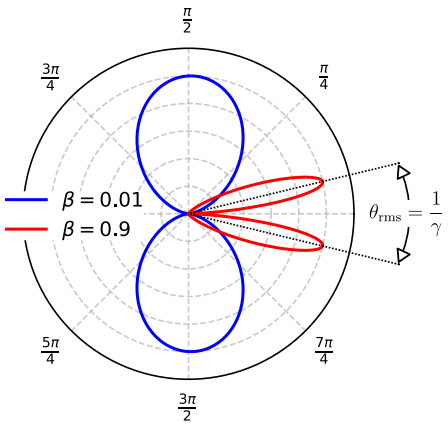
The first term in equation 2.35, which only depends on $\boldsymbol{\beta}$, is known as the velocity field and is proportional to R^{-2}. The second term, which depends on both $\boldsymbol{\beta}$ and $\dot{\boldsymbol{\beta}}$, is known as the acceleration field and is proportional to R^{-1}. Far away from the charge the velocity field will vanish and only the acceleration field will remain. Therefore, the energy flux is given by the radial component of Poynting's vector [80]

$$[\mathbf{S} \cdot \hat{\mathbf{n}}]_{\text{ret}} = \left[\frac{\mathbf{E} \times \mathbf{B}}{\mu_0} \cdot \hat{\mathbf{n}} \right]_{\text{ret}} = \frac{q^2}{16\pi^2 c \epsilon_0} \left[\frac{1}{R^2} \left| \frac{\hat{\mathbf{n}} \times \left[(\hat{\mathbf{n}} - \boldsymbol{\beta}) \times \dot{\boldsymbol{\beta}} \right]}{(1 - \boldsymbol{\beta} \cdot \hat{\mathbf{n}})^3} \right|^2 \right]_{\text{ret}}. \tag{2.36}$$

This equation gives the power per unit area, detected by an observer at time t and emitted by the charge at the retarded time $t_r = t - R(t_r)/c$. However, the important quantity is the power emitted by the charge with respect to its own time. This can be obtained by considering the total energy emitted by the charge as it accelerates between times $t_r = T_0$ to $t_r = T_1$

$$E = \int_{t=T_0+R(T_0)/c}^{t=T_1+R(T_1)/c} [\mathbf{s} \cdot \hat{\mathbf{n}}]_{\text{ret}} \mathrm{d}t = \int_{t_r=T_0}^{t_r=T_1} \mathbf{s} \cdot \hat{\mathbf{n}} \, \frac{\mathrm{d}t}{\mathrm{d}t_r} \mathrm{d}t_r \tag{2.37}$$

and is given by the integrand of the equation on the right-hand side. We can then define



(a) Laboratory frame

Figure 2.3: Normalised angular distribution of radiation emitted by a non-relativistic (blue) and a relativistic (red) charge with an acceleration parallel to its velocity. At high velocities the distribution collapses onto the motion axis.

the power radiated per unit solid angle as

$$\frac{\mathrm{d}P(t_r)}{\mathrm{d}\Omega} = R^2\,\mathbf{s}\cdot\hat{\mathbf{n}}\,\frac{\mathrm{d}t}{\mathrm{d}t_r} = \frac{q^2}{16\pi^2 c\epsilon_0}\frac{\left|\hat{\mathbf{n}}\times\left[(\hat{\mathbf{n}}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}\right]\right|^2}{(1-\hat{\mathbf{n}}\cdot\boldsymbol{\beta})^5}. \tag{2.38}$$

This equation is plotted in figure 2.3 for both a non-relativistic ($|\boldsymbol{\beta}| = 0.01$) and a relativistic ($|\boldsymbol{\beta}| = 0.9$) charge where $\boldsymbol{\beta} \parallel \dot{\boldsymbol{\beta}}$. In the non-relativistic case, we can see a double lobe characteristic of a $\sin^2\theta$ distribution. This is the same angle dependence that appeared in equation 2.33 during the pictorial derivation. In the relativistic case, the emission becomes beamed along the axis of motion. It can be shown that the root mean square of the emission angle is [80]

$$\langle\theta^2\rangle^{1/2} = \frac{1}{\gamma} \tag{2.39}$$

Therefore, at higher velocities the beaming will become more apparent as the cone angle shrinks.

By integrating equation 2.38 over solid angle we retrieve Liénard's result

$$P = \frac{q^2\gamma^6}{6\pi c\epsilon_0}\left[\dot{\boldsymbol{\beta}}^2 - (\boldsymbol{\beta}\times\dot{\boldsymbol{\beta}})^2\right]. \tag{2.40}$$

for the total power emitted by a relativistic charge. This increases rapidly with γ and suggests radiation effects can dominate a particle's motion for ultra-relativistic charges.

2.1.5 Radiation Spectrum From an Accelerating Charge

Having found the total power emitted from an accelerating charge, we will now turn our attention to the spectrum of the radiation. Following Jackson [80], we will define the radiation spectrum such that its integral over the frequencies measured by an observer, ω, gives the energy emitted, W, per unit solid angle, *i.e.*

$$\frac{\mathrm{d}W}{\mathrm{d}\Omega} = \int_0^\infty \frac{\mathrm{d}^2 I}{\mathrm{d}\omega \mathrm{d}\Omega} \mathrm{d}\omega \tag{2.41}$$

where I is the intensity. As we wish to work with the observer's frequencies, we will start by rewriting equation 2.38 in terms of the observer's time

$$\frac{\mathrm{d}P(t)}{\mathrm{d}\Omega} = \frac{c}{\epsilon_0}\Big[R^2\mathbf{E}^2\Big]_{\mathrm{ret}} = |\mathbf{A}(t)|^2. \tag{2.42}$$

If we assume the particle is accelerated over a finite duration, the total energy emitted is also finite. We can then write the total energy emitted per unit solid angle as

$$\frac{\mathrm{d}W}{\mathrm{d}\Omega} = \int_{-\infty}^{\infty} |\mathbf{A}(t)|^2 \mathrm{d}t. \tag{2.43}$$

By applying a Fourier transform, this equation can be expressed as an integral over frequency rather than time

$$\frac{\mathrm{d}W}{\mathrm{d}\Omega} = \int_{-\infty}^{\infty} |\mathbf{A}(\omega)|^2 \mathrm{d}\omega \tag{2.44}$$

where

$$\mathbf{A}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathbf{A}(t) e^{i\omega t} \mathrm{d}t \tag{2.45}$$

is the Fourier transform of $\mathbf{A}(t)$. Therefore, by comparing equations 2.41 and 2.44, we can write the radiation spectrum as

$$\frac{\mathrm{d}^2 I}{\mathrm{d}\omega \mathrm{d}\Omega} = 2|A(\omega)|^2 \tag{2.46}$$

where the factor of 2 has arisen due to only integrating over positive frequencies and assuming $|\mathbf{A}(\omega)|^2$ is symmetric[1]. To obtain the spectrum as a function of the charge's motion, we can insert the electric field from equation 2.35 which yields [80]

$$\frac{\mathrm{d}^2 I}{\mathrm{d}\omega\, \mathrm{d}\Omega} = \frac{q^2\omega^2}{16\pi^3 c\epsilon_0} \left| \int_{-\infty}^{\infty} \hat{\mathbf{n}} \times \Big(\hat{\mathbf{n}} \times \boldsymbol{\beta}(t_r)\Big) e^{i\omega(t_r - \hat{\mathbf{n}}\cdot\mathbf{r}(t_r)/c)} \mathrm{d}t_r \right|^2 \tag{2.47}$$

where integration by parts has been performed to simplify the final form.

[1] This must be the case if $\mathbf{A}(t)$ is real, see Jackson [80] for details.

2.1.6 Thomson Scattering and Synchrotron Spectra

Equation 2.47 shows that all we require to calculate the radiation spectrum from the motion of a charge, is a time record of $\mathbf{r}$ and $\boldsymbol{\beta}$. However, even for relatively simple motion, the analytical calculation is involved. In this section we will briefly discuss the radiation spectrum from two different systems. The first will be the emission from an electron interacting with a plane wave. The second will be the emission from an electron performing instantaneous circular motion. We will come back to both these systems in the final part of this chapter when discussing the quantum motion of charges in strong fields.

Thomson Scattering

The emission of radiation from an electron interacting with a plane wave is commonly referred to as Thomson scattering. For this system, we can simplify the calculation of equation 2.47 by noting that the electron exhibits periodic motion at the frequency of the wave. Therefore, we can write[2]

$$\boldsymbol{\beta}(t + mT) = \boldsymbol{\beta}(t) \qquad \mathbf{r}(t + mT) = m\mathbf{r}_0 + \mathbf{r}(t) \tag{2.48}$$

where m is an integer, $T = 2\pi/\omega_0$ is the wave period and $\mathbf{r}_0$ is the total displacement gained over one cycle. By inserting this into equation 2.47 and using the definition of the Dirac delta function, $\sum_m e^{imx} = \sum_m 2\pi\delta(x - 2m\pi)$, we arrive at the following [87]

$$\frac{\mathrm{d}^2 I}{\mathrm{d}\omega \mathrm{d}\Omega} = \frac{e^2\omega^2}{4\pi^2 c}\left|\hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \sum_{m=-\infty}^{\infty} \mathbf{F}_m \delta(\omega - \omega_m))\right|^2 \tag{2.49}$$

where

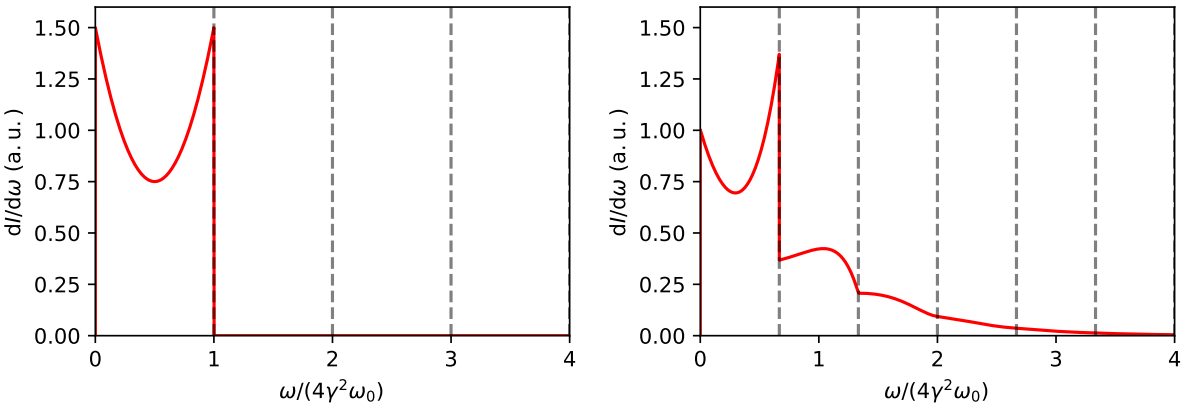
$$\omega_m = \frac{2\pi\, m}{T - \hat{\mathbf{n}} \cdot \mathbf{r}_0/c} \qquad \mathbf{F}_m = \frac{\omega_m}{2\pi}\int_0^T \boldsymbol{\beta}(t) e^{i\omega_m(t - \hat{\mathbf{n}}\cdot\mathbf{r}(t)/c)}. \tag{2.50}$$

Therefore, the spectrum exhibits harmonic behaviour, with a fundamental frequency that is not simply the frequency of the wave, but also depends on the net displacement of the electron.

The calculation of $\mathbf{F}_m$ is involved, so it will not be carried out here, however, has been discussed by multiple authors. The case of an initially static electron has been studied by Sarachik *et. al* [88]. Lau *et. al.* [87] has provided a review of the back-scattering radiation. Finally, a full review of the general interaction has been carried out by Esarey *et. al.* [89].

Figures 2.4 (a) and (b) show plots of the Thompson scattering spectrum from a 50 MeV

[2] In this equation, t is the retarded electron time, however, for the following discussion we have dropped the subscript r.



(a) $a_0 = 0.3$ (b) $a_0 = 1.0$

Figure 2.4: Normalised Thompson scattering spectrum. (a) shows the spectrum for a weak field with only one harmonic present. (b) shows the spectrum for a strong field in which multiple harmonics are present. The grey dashed lines show the boundaries of the harmonics.

electron interacting with a weak ($a_0 = 0.3$) and strong ($a_0 = 1.0$) field respectively. In the weak field case, only one harmonic is present with a maximum fundamental frequency of $\omega_1 = 4\gamma^2\omega_0$. Although this is a classical result, the up-shift in energy can be explained by boosting into the average rest frame of the electron and considering a photon elastically scattering off the electron. If the photon scatters back along the boost axis, its energy will be Doppler shifted by a factor of $4\gamma^2$. The up-shift in energy is reduced for larger scattering angles.

In the strong field case, a significant proportion of the radiation is emitted at higher harmonics. We can use the same heuristic, quantum argument to explain these higher harmonics. They result from multiple wave photons scattering off the electron at the same time, producing a single high energy photon. Therefore, the intensity of the higher harmonic will have a strong dependence of a_0 which results in them being highly suppressed for $a_0 < 1$. When higher harmonics are present in the spectrum it is often referred to as nonlinear-Thompson scattering.

Synchrotron Radiation

Later in this chapter, when studying the quantum emission from an accelerating charge, we will see there are few systems for which calculations of the radiation spectrum can be performed exactly. Therefore, approximations must be made, such as the quasi-stationary and weak field approximations. If these approximations are valid, we can assume the charge experiences instantaneously circular motion, simplifying the calculations. Here, we will review the classical equivalent to this, known as synchrotron radiation.

Even for this simple instantaneously circular motion, solving equation 2.47 is chal-

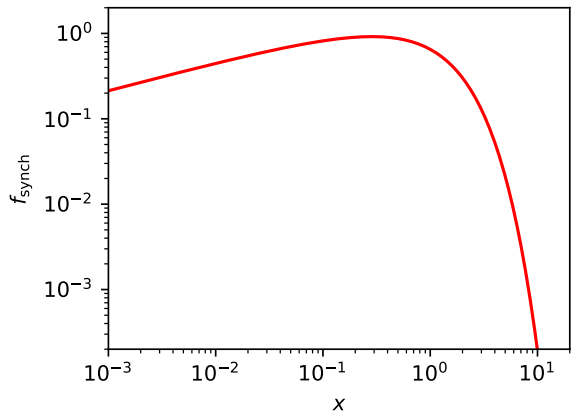


Figure 2.5: Plot of synchrotron function.

lenging. The full calculation can be found in Jackson [80], with the final result for an electron giving

$$\frac{\mathrm{d}^2 I}{\mathrm{d}\omega \mathrm{d}\Omega} = \frac{e^2}{12\pi^3 c\epsilon_0}\left(\frac{\omega\rho}{c}\right)^2\left(\frac{1}{\gamma^2}+\theta^2\right)^2\left[K_{2/3}^2(\psi)+\frac{\theta^2}{(\gamma^{-2}+\theta^2)}K_{1/3}^2(\psi)\right] \tag{2.51}$$

where ρ is the radius of curvature, $K_{2/3}$ and $K_{1/3}$ are modified Bessel functions of the second kind and their argument is given by

$$\psi = \frac{\omega\rho}{3c}\left(\frac{1}{\gamma^2}+\theta^2\right)^{3/2}. \tag{2.52}$$

The modified Bessel functions is exponentially suppressed for $\psi > 1$. Therefore, we can define a critical frequency beyond which there is little energy in the spectrum at any angle. Using $\psi = 1/2$ and $\theta = 0$, this is given by

$$\omega_c = \frac{3}{2}\gamma^3\frac{c}{\rho}. \tag{2.53}$$

The total intensity spectrum, integrated over solid angle is

$$\frac{\mathrm{d}I}{\mathrm{d}\omega} = \frac{\sqrt{3}}{4}\frac{e^2}{\pi c\epsilon_0}\gamma\, f_{\text{synch}}\left(\frac{\omega}{\omega_c}\right) \tag{2.54}$$

where

$$f_{\text{synch}}(x) = x\int_x^\infty K_{5/3}(x)\mathrm{d}x \tag{2.55}$$

is the classical synchrotron function and is plotted in figure 2.5.

At a synchrotron, an oscillating magnetic field device (*e.g.* a wiggler or undulator) is often inserted into the beam path, causing the electrons to emit radiation. With a magnetic field strength of $|\mathbf{B}| = 1\,\mathrm{T}$ and an electron beam of energy $E = 3\,\mathrm{GeV}$, the

critical frequency is $\omega_c \sim 6\,\text{keV}$, where we have used $\rho = p/e|\mathbf{B}|$. This hard source of x-rays has become an invaluable tool in many areas of scientific research.

If we look at a more extreme example from astrophysics, the magnetic field of pulsars can reach in excess of $|\mathbf{B}| = 10^9\,\text{T}$ [90]. The critical frequency for the same electron beam is $\omega_c \sim 6 \times 10^3\,\text{GeV}$, several orders of magnitude larger than the energy of the electron itself. This is clearly unphysical and suggests a breakdown in the theory. We will discuss the solution to this problem later in this chapter.

2.1.7 Radiation Reaction

When an accelerating charge radiates, energy and momentum are taken away by the field, a process known as radiation reaction. In section 2.1.2, when deriving the Lorentz force, we did not consider the self-interaction between the charge and its radiated field. Therefore, radiation reaction is not accounted for in equation 2.17. However, as we already know the power that is lost due to radiation (equation 2.34 for $|\boldsymbol{\beta}| \ll 1$ and equation 2.40 for $|\boldsymbol{\beta}| \lesssim 1$), adding an extra damping term to the equations of motion, *i.e.*

$$\begin{aligned} \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} &= q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) + \mathbf{F}_{\text{RR}} \\ \frac{\mathrm{d}p^\mu}{\mathrm{d}\tau} &= \frac{q}{mc} F^{\mu\nu} p_\nu + \mathcal{F}^\mu_{\text{RR}} \end{aligned} \tag{2.56}$$

should be a relatively simple task.

Lorentz-Abraham Force

In the nonrelativistic limit, from the definition of work done over the time interval $t_1 < t < t_2$, we can write

$$\int_{t_1}^{t_2} \mathbf{F}_{\text{RR}} \cdot \mathbf{v} \mathrm{d}t = -\int_{t_1}^{t_2} P \mathrm{d}t = -\frac{q^2}{6\pi\epsilon_0 c^3} \int_{t_1}^{t_2} \dot{\mathbf{v}}^2 \mathrm{d}t \tag{2.57}$$

where equation 2.34 has been used. By applying integration by parts, the right-hand side of this equation can be written as

$$\int_{t_1}^{t_2} \mathbf{F}_{\text{RR}} \cdot \mathbf{v} \mathrm{d}t = -\frac{q^2}{6\pi\epsilon_0 c^3} \left([\dot{\mathbf{v}} \cdot \mathbf{v}]_{t_1}^{t_2} - \int_{t_1}^{t_2} \ddot{\mathbf{v}} \cdot \mathbf{v} \mathrm{d}t \right). \tag{2.58}$$

If the motion is periodic, the first term will vanish and we can identify the radiation reaction force as

$$\mathbf{F}_{\text{RR}} = m\tau_{\text{rad}} \ddot{\mathbf{v}} \tag{2.59}$$

where $\tau_{rad} = q^2/6\pi\epsilon_0 mc^3$. This is the Lorentz-Abraham force [91] and has the following relativistic generalisation

$$\mathcal{F}^{\mu}_{\mathrm{RR}} = \tau_{\mathrm{rad}}\left(\frac{\mathrm{d}^2 p^{\mu}}{\mathrm{d}\tau^2} - \frac{p^{\mu}}{m^2c^2}\frac{\mathrm{d}p^{\nu}}{\mathrm{d}\tau}\frac{\mathrm{d}p_{\nu}}{\mathrm{d}\tau}\right) \tag{2.60}$$

known as the Lorentz-Abraham-Dirac (LAD) force [92].

The radiation reaction force has the unusual property that it is proportional to the third derivative of the particle's position. This leads to unphysical behaviour. To examine this, we can solve the full equation of motion for a radiating charge in an external field

$$m\dot{\mathbf{v}} = \mathbf{F}_{\mathrm{ext}} + m\tau_{\mathrm{rad}}\ddot{\mathbf{v}} \tag{2.61}$$

where $\mathbf{F}_{\mathrm{ext}} = q(\mathbf{E}_{\mathrm{ext}} + \mathbf{v}\times\mathbf{B}_{\mathrm{ext}})$ is the Lorentz force. The general solution to this equation is [93]

$$m\dot{\mathbf{v}}(t) = m\dot{\mathbf{v}}(t_0)e^{(t-t_0)/\tau_{\mathrm{rad}}} - \frac{e^{t/\tau_{\mathrm{rad}}}}{\tau_{\mathrm{rad}}}\int_{t_0}^{t}\mathbf{F}_{\mathrm{ext}}(t')e^{-t'/\tau_{\mathrm{rad}}}\mathrm{d}t' \tag{2.62}$$

where t_0 is a constant and $\dot{\mathbf{v}}(t_0)$ is the acceleration at this time. If there are no external fields, the second term will vanish. We are left with the first term, which suggests the charge will exponentially accelerate over a timescale τ_{rad} (with $\tau_{\mathrm{rad}} = 6.26\times10^{-24}\,\mathrm{s}$ for an electron). This is clearly unphysical and is known as a runaway solution.

One approach to remove runaway solutions is to ensure the acceleration tends to zero once the forces have finished acting. This is achieved by applying the boundary conditions $t_0 \to \infty$ and $\dot{\mathbf{v}}(t_0) \to 0$. By making the substitution $s = (t'-t)/\tau_{\mathrm{rad}}$, equation 2.62 becomes

$$m\dot{\mathbf{v}}(t) = \int_0^{\infty}\mathbf{F}_{\mathrm{L}}(t+\tau_{\mathrm{rad}}s)e^{-s}\mathrm{d}s \tag{2.63}$$

and does not exhibit exponential growth. However, it suggests that the acceleration of the charge at time t, depends on the force it experiences at all times in the future. This clearly breaks causality and should also be discarded.

Landau-Lifshitz Force

The unphysical solutions, which plague the nonrelativistic Lorentz-Abraham force, are still present in the fully relativistic LAD equation as they stem from the third derivative of position [93]. Therefore, to remove them, the order of the equation of motion should be reduced. Landau and Lifshitz [83] proposed a method for doing this by assuming the radiation reaction term is always much smaller than the Lorentz term, such that

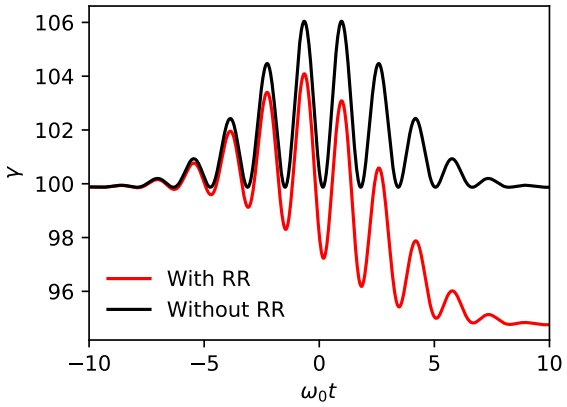


Figure 2.6: Demonstration of the radiation reaction force on an electron interacting with an intense laser pulse. Shown is the energy of the electron against time without radiation reaction included (black) and with radiation reaction included (red).

$m\dot{\mathbf{v}} = \mathbf{F}_{\text{ext}} + \mathcal{O}(\tau_{\text{rad}})$. By substituting this into equation 2.61 we get

$$\begin{aligned} m\dot{\mathbf{v}} &= \mathbf{F}_{\text{ext}} + \tau_{\text{rad}}\frac{d\mathbf{F}_{\text{ext}}}{\mathrm{d}t} + \mathcal{O}(\tau_{\text{rad}}^2) \\ &= q(\mathbf{E} + \mathbf{v}\times\mathbf{B}) + \tau_{\text{rad}}\left[q\Big(\frac{\mathrm{d}\mathbf{E}}{\mathrm{d}t} + \mathbf{v}\times\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t}\Big) + \frac{q}{m}\Big(\mathbf{E} + \mathbf{v}\times\mathbf{B}\Big)\times\mathbf{B}\right] + \mathcal{O}(\tau_{\text{rad}}^2). \end{aligned} \tag{2.64}$$

If we ignore contributions of order $\mathcal{O}(\tau_{\text{rad}}^2)$ we can identify the Landau-Lifshitz force as

$$\mathbf{F}_{\text{LL}} = \tau_{\text{rad}}\left[q\Big(\frac{\mathrm{d}\mathbf{E}}{\mathrm{d}t} + \mathbf{v}\times\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t}\Big) + \frac{q}{m}\Big(\mathbf{E} + \mathbf{v}\times\mathbf{B}\Big)\times\mathbf{B}\right]. \tag{2.65}$$

A similar method can be applied to the LAD equation to give the Landau-Lifshitz force for a relativistic charge [83]

$$\mathcal{F}^{\mu}_{\text{LL}} = \frac{q\,\tau_{\text{rad}}}{mc}\left[(\partial_\sigma F^{\mu\nu})p_\nu p^\sigma + \frac{q}{c}F^{\mu\nu}F_{\nu\sigma}p^\sigma + \frac{q}{m^2c^3}(F^{\nu\sigma}p_\sigma)^2 p^\mu\right]. \tag{2.66}$$

The Landau-Lifshitz force is generally accepted as the correct classical description of radiation reaction as it appears in the classical limit of fully quantum calculations [94]. The result of radiation reaction on an electron interacting with a laser pulse is demonstrated in figure 2.6. Without the force included, the electron's energy oscillates but ultimately leaves the laser unaffected. However, when the force is included, the oscillations cause the electron to radiate and it leaves with a lower energy.

2.2 Quantum Electrodynamics

In the previous section, we found situations in which classical electrodynamics predict unphysical behaviour. Specifically, for the case of radiation reaction we found runaway solutions or non-causal forces. Both these effects occur over a characteristic time-scale given by τ_{rad}, so we can define a characteristic length-scale as

$$c\,\tau_{\text{rad}} = \frac{e^2}{6\pi\epsilon_0 m c^2} = \frac{2}{3}\alpha\lambda\!\!\!^{-}_c \tag{2.67}$$

where $\alpha \approx 1/137$ is the fine structure constant and $\lambda\!\!\!^{-}_c$ is the reduced Compton wavelength. The reduced Compton wavelength is a length which is well within the quantum realm and $c\,\tau_{\text{RR}}$ is considerably smaller. Therefore, it should come as no surprise that QED is required to explain these phenomena.

In classical electrodynamics, both the electromagnetic field and charged particles are treated as distinctly different objects. However, in QED both the charged particles (*i.e.* electrons and positrons) and electromagnetic radiation (*i.e.* photons) are described by fields. The field used to represent the particles is known as the Dirac field, denoted by ψ whereas electromagnetic radiation is represented by the Maxwell field, denoted by A^μ. Through quantisation of these fields, particles appear as excited states. This gives rise to a theory of electrodynamics that is both quantum and relativistic.

In this section we will give a brief overview of QED, starting as we did for classical electrodynamics, with the Lagrangian for the system. From this, interaction probabilities can be calculated through perturbation theory, however, such calculations are not the aim of this work so will not be presented here. After reviewing the method of calculating interaction probabilities, we will discuss some specific two body quantum processes that are of importance to this work.

2.2.1 Calculating QED Cross-Sections

Most of the experiments linked to this work can be described as scattering experiments. Here, we interact two beams of particles (denoted by subscripts 1 and 2) and observe the particles that come out. Between experiments, there will be a number of parameters which change. Therefore, if we want to compare the number of scattering events, N, that occurred in one experiment to another, we should divide by these quantities. This gives rise to the total cross-section for the interaction

$$\sigma = \frac{N}{\rho_1\, l_1\, \rho_2\, l_2\, A} \tag{2.68}$$

where $\rho_{1,2}$ is the beam density, $l_{1,2}$ is the beam length and A is the overlap area between the beams.

We are also interested in the momentum distribution of the particles which scatter out. It is therefore useful to define the differential cross-section. In the case where there are only two outgoing particles (denoted by subscripts 3 and 4), the differential cross-section gives the probability of scattering into final states with momenta $\mathbf{p}_3$ and $\mathbf{p}_4$. Due to conservation laws in the centre-of-mass (CM) frame, only two of the six components of the momentum are not fixed. If we use the polar, θ, and azimuthal, ϕ, angles of the scattering axis as the free parameters, the differential cross-section is given by $\mathrm{d}\sigma/\mathrm{d}\Omega$, where $\mathrm{d}\Omega = \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi$.

Following Peskin and Schroeder [81], to perform a calculation of the cross-section for a process, we should start from the Lagrangian of the system. As we are now interested in interacting fields, it is more appropriate to work with the Lagrangian density, $\mathcal{L}$, rather than the Lagrangian itself. However, to be more succinct, we will also refer to $\mathcal{L}$ as the Lagrangian. $\mathcal{L}$ is a function of one or more fields, $\phi(x)$, and their derivatives, $\partial_\mu\phi(x)$[3]. Therefore, the action is given by

$$A = \int \mathcal{L}(\phi(x), \partial_\mu\phi(x))\mathrm{d}^4x \tag{2.69}$$

which we can apply the principle of least action to, resulting in the Euler–Lagrange equations for a field

$$\frac{\partial\mathcal{L}}{\partial\phi} - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right) = 0. \tag{2.70}$$

The total Lagrangian for QED can be decomposed into three parts

$$\begin{aligned}\mathcal{L}_{\mathrm{QED}} &= \mathcal{L}_{\mathrm{Dirac}} + \mathcal{L}_{\mathrm{Maxwell}} + \mathcal{L}_{\mathrm{int}} \\ &= \bar{\psi}(i\not{\partial} - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - e\bar{\psi}\gamma^\mu\psi A_\mu\end{aligned} \tag{2.71}$$

where $\psi(x)$ is the Dirac field operator $A^\mu(x)$ is the Maxwell field operator, $\not{\partial} = \gamma^\mu\partial_\mu$ where γ^μ are Dirac matrices and natural units ($\hbar = c = 1$) have been used. $\mathcal{L}_{\mathrm{Dirac}}$ is the free Lagrangian for electrons and positrons, $\mathcal{L}_{\mathrm{Maxwell}}$ is the free Lagrangian for the electromagnetic field and $\mathcal{L}_{\mathrm{int}}$, describes the interaction between the Dirac and Maxwell fields.

We can obtain the equations of motion for these fields by inserting $\mathcal{L}_{\mathrm{QED}}$ into equation 2.70. By setting $\phi = A_\mu$ and applying the Lorenz gauge we obtain the following [81]

$$\partial_\nu\partial^\nu A^\mu = e\bar{\psi}\gamma^\mu\psi \tag{2.72}$$

which, by comparing to equation 2.8 and identifying $j^\mu = e\bar{\psi}\gamma^\mu\psi$ as the current density operator, gives the QED equivalent of Maxwell's equations. By setting $\phi = \psi$ we obtain

[3]Here x is a four vector but we have dropped the index when it is the argument of a function.

the equations of motion for the Dirac field

$$(i\partial\!\!\!/ - m)\psi = eA\!\!\!/_\mu\psi \tag{2.73}$$

which is known as the Dirac equation.

For an arbitrary field, solving these equations of motion is extremely difficult, if not impossible. Therefore, to proceed, approximations have to be made through perturbation theory. To apply perturbation theory, it is assumed that the interaction term in the Lagrangian is much smaller than the free terms. However, it is more convenient to work with the interaction Hamiltonian rather than interaction Lagrangian of the system. We can write the Hamiltonian as

$$H = H_0 + H_{\text{int}} \tag{2.74}$$

where the interaction term is[81]

$$H_{\text{int}} = -\int \mathcal{L}_{\text{int}} \mathrm{d}^3x = \int e\bar{\psi}\gamma^\mu\psi A_\mu \mathrm{d}^3x \tag{2.75}$$

.

We will define a process as an interaction in which the system starts in state $\langle i|$ at time t_i and evolves to a final state $|f\rangle$ at time t_f. This evolution occurs according to the following operator

$$U(t, t_0) = \mathcal{T} e^{-i\int_{t_0}^{t} H_{\text{int}}(t')dt'} \tag{2.76}$$

known as Dyson's formula, where $\mathcal{T}$ is the time-ordering operator. The initial and final states are assumed to be eigenstates of the free Hamiltonian in the asymptotic limit $t_i \to -\infty$ and $t_f \to \infty$. The probability amplitude for the evolution from $|i\rangle$ to $|f\rangle$ is

$$S_{fi} = \lim_{t_i\to-\infty} \lim_{t_f\to\infty} \langle f| U(t_i, t_f) |i\rangle \tag{2.77}$$

which is known as the S-matrix. Calculations of this matrix can be performed for a given process by expanding in powers of $H_{\text{int}} \propto \sqrt{\alpha}$. Each term in this expansion can be expressed as a Feynman diagram. These diagrams are drawn according to the Feynman rules where the number of vertices corresponds to the order of the term. This greatly reduces the complexity of the calculation, however, it is beyond the scope of this work and will not be carried out here.

The structure of the S-matrix is such that it contains the probability of no interaction occurring and a delta function corresponding to the conservation of momentum. Therefore, we can extract these to define the scattering amplitude

$$S_{fi} = \delta_{fi} + i(2\pi)^4\delta^4(p_f^\mu - p_i^\mu)\mathcal{A}_{fi} \tag{2.78}$$

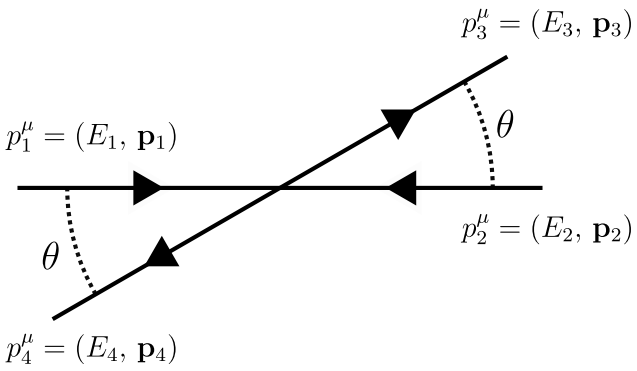


Figure 2.7: Diagram of binary collision in the CM frame of the system.

where p_f^μ and p_i^μ are the initial and final total four-momentum of the system. In terms of $\mathcal{A}_{fi}$, the differential cross-section for a binary collision in the CM frame is given by [81, 95]

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\mathrm{CM}} = \frac{1}{64\pi^2 E_{\mathrm{cm}}^2}\frac{|\mathbf{p}_1|}{|\mathbf{p}_3|}|\mathcal{A}_{if}|^2 \tag{2.79}$$

where E_{cm} is the centre of mass energy and $\mathbf{p}_1$ and $\mathbf{p}_3$ are the momentum of an initial and final state particle respectively (see figure 2.7).

2.2.2 Two-Body Scattering Processes

The lowest order scattering process one could imagine would involve two particles colliding to produce a third. However, such a process is unphysical as energy and momentum cannot be conserved. Therefore, the lowest order physical scattering process involves two incoming particles and two outgoing particles. A diagram of such a process is shown in figure 2.7. When discussing the cross-sections of two-body interactions, is it useful to write them as functions of Lorentz-invariant variables. Therefore, we will use the Mandelstam variables

$$\begin{aligned} s &= (p_1 + p_2)^2 = (p_3 + p_4)^2 \\ t &= (p_1 - p_3)^2 \;= (p_4 - p_2)^2 \\ u &= (p_1 - p_4)^2 = (p_3 - p_2)^2 \end{aligned} \tag{2.80}$$

which encode all the important parameters of the interaction, including the particles' energy, momentum and angles.

Table 2.1 gives a summary of all the two-body QED processes which have been grouped based on the initial state particle types. Here, we are particularly interested in processes which have a photon in the initial state (*i.e.* the electron-photon and photon-photon systems). Part of this work has been to develop a new QED package for Geant4 which includes the Breit-Wheeler process, Compton scattering and photon-photon scattering.

Table 2.1: Two-body QED processes.

Electron-Electron	Electron-Photon	Photon-Photon
Møller scattering	Compton scattering	Breit-Wheeler process
$e^{\pm}e^{\pm} \longrightarrow e^{\pm}e^{\pm}$	$e^{\pm}\gamma \longrightarrow e^{\pm}\gamma$	$\gamma\gamma \longrightarrow e^{+}e^{-}$
Bhabha scattering		Photon-photon scattering
$e^{+}e^{-} \longrightarrow e^{+}e^{-}$		$\gamma\gamma \longrightarrow \gamma\gamma$
Dirac annihilation		
$e^{+}\ e^{-} \longrightarrow \gamma\gamma$		

Table 2.2: Low order QED processes which take place in the nuclear field of an atom.

Nuclear-Electron	Nuclear-Photon
Mott scattering	Bethe-Heitler process
$Ze^{\pm} \longrightarrow Ze^{\pm}$	$Z\gamma \longrightarrow Ze^{+}e^{-}$
Bremsstrahlung	Photon splitting
$Ze^{\pm} \longrightarrow Ze^{\pm}\gamma$	$Z\gamma \longrightarrow Z\gamma\gamma$
	Delbrück scattering
	$Z\gamma \longrightarrow Z\gamma$

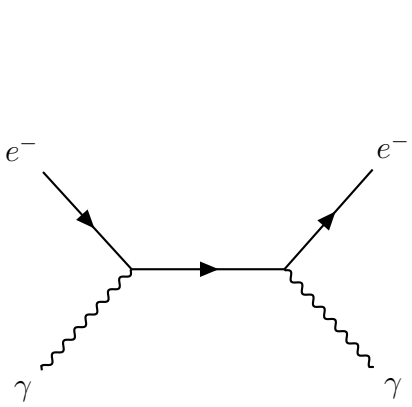
Therefore, in the remainder of this section we will provided a more detailed discussion of each. For a full review of the whole of table 2.1, see Jauch 1980 [96].

However, before moving on, it is important to mention some interactions which take place inside the nuclear field of an atom, Z, that are also important to this work. These are summarised in table 2.2. Apart from Mott scattering, each process has a counterpart from table 2.1 with a similar lowest order Feynman diagram. However, in the nuclear field processes, one or more of the real photons are replaced by a virtual photon. The nucleus is still present after the interaction so these processes cannot be described as two body interactions.

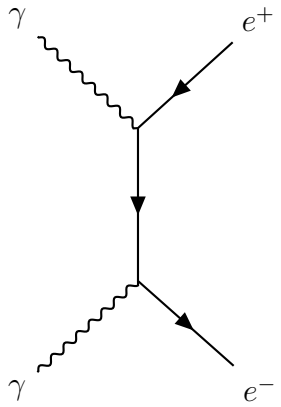
Compton Scattering

Compton scattering describes the interaction between a photon and a charged particle ($e^{\pm}\gamma \rightarrow e^{\pm}\gamma$). It was first observed by Arthur Compton in 1923 [97] when he measured the transfer of energy from a high energy photon to an electron. Therefore, the process is often described as inelastic scattering. However, in the CM frame it is always elastic as there is no transfer of energy. The lowest order Feynman diagrams which contribute to this process are of second order ($\propto \alpha^2$) and an example is shown in figure 2.8 (a).

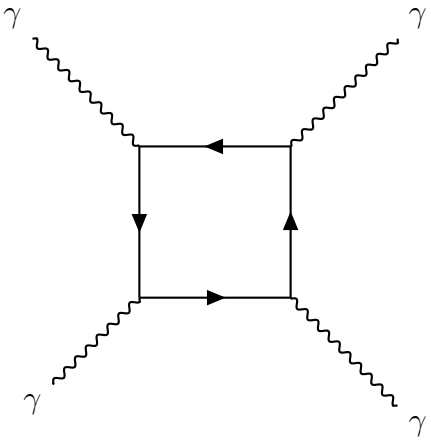
The frame of reference we will use for calculations in this work is the CM frame.



(a) Compton scattering



(b) Breit-Wheeler process



(c) Photon-photon scattering

Figure 2.8: Example of lowest order Feynman diagrams which contribute to (a) Compton scattering (b) the Breit-Wheeler process and (c) photon-photon scattering.

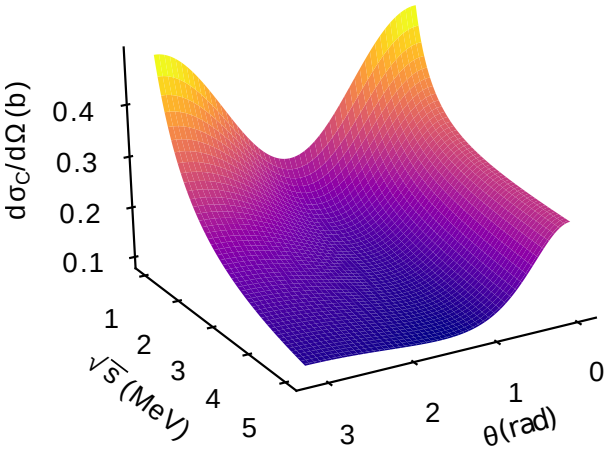
Here, the differential cross-section for Compton scattering is complex but can be found in Jauch 1980 [96]. However, it is plotted along with the total cross-section in figures 2.9 (a) and (b) respectively. Due to the difference in the mass of the scattering particles, the differential cross-section is asymmetric with a strong preference for back-scattering.

Although complex in the CM frame, in the rest frame of the electron, the differential cross-section simplifies significantly to [98]

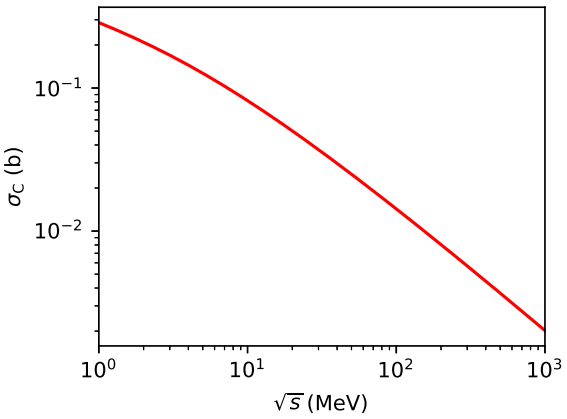
$$\frac{\mathrm{d}\sigma_{\mathrm{CS}}}{\mathrm{d}\Omega} = \frac{\alpha^2}{2m^2}\left(\frac{\omega'}{\omega}\right)^2\left(\frac{\omega}{\omega'} + \frac{\omega'}{\omega} - \sin^2\theta\right) \tag{2.81}$$

where ω is the frequency of the incoming photon, ω' is the frequency of the outgoing photon and θ is the scattering angle. Through conservation of momentum and energy we can relate the outgoing frequency to the incoming frequency

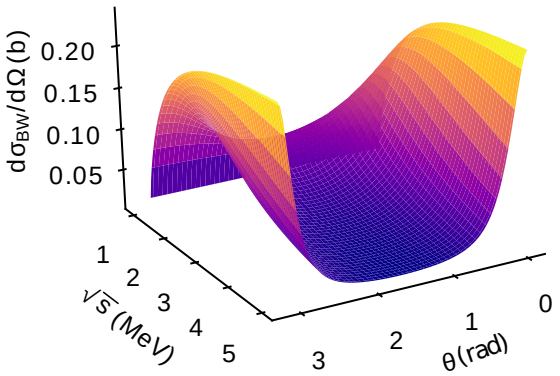
$$\omega' = \frac{\omega}{1 + \frac{\omega}{m}(1 - \cos\theta)}. \tag{2.82}$$



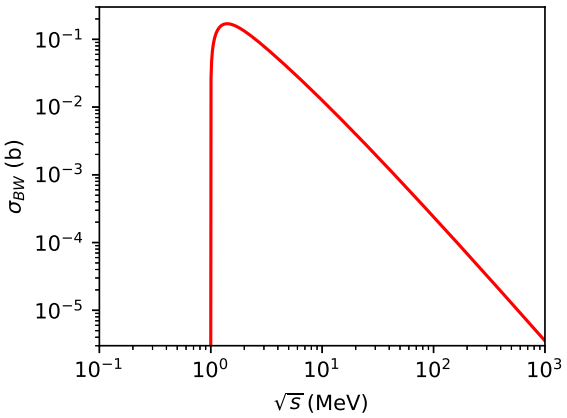
(a) Differential Compton scattering cross-section



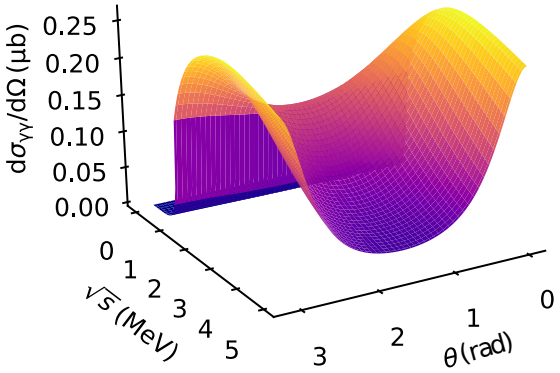
(b) Total Compton scattering cross-section



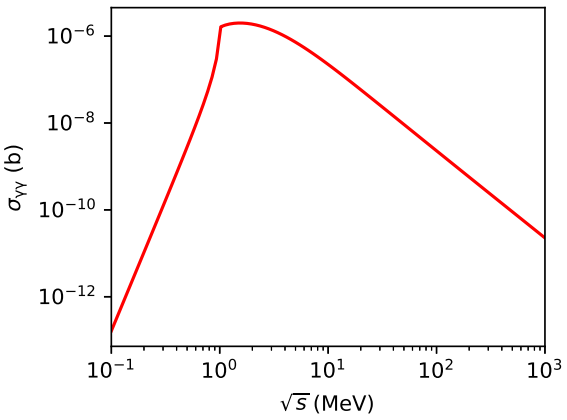
(c) Differential Breit-Wheeler cross-section



(d) Total Breit-Wheeler cross-section



(e) Differential photon-photon scattering cross-section



(f) Total photon-photon scattering cross-section

Figure 2.9: Plots of the differential cross-section (left column) and the total cross-section (right column) for Compton scattering, the Breit-Wheeler process and photon-photon scattering.

Integrating equation 2.81 over solid angle gives the total Compton cross-section [96]

$$\sigma_{\mathrm{CS}} = \frac{2\pi\alpha^2}{m^2}\left\{\frac{1+\omega}{\omega^3}\left[\frac{2\omega(1+\omega)}{1+2\omega} - \log(1+2\omega)\right] + \frac{\log(1+2\omega)}{2\omega} - \frac{1+3\omega}{(1+2\omega)^2}\right\}. \quad (2.83)$$

In the non-relativistic limit, $\omega \ll m$, from equation 2.82 we can see $\omega' \to \omega$. Therefore, the interaction in the electron's rest frame is elastic. Equations 2.81 and 2.83 then reduce to

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{\alpha^2}{m^2}(1+\cos^2\theta) \qquad \text{and} \qquad \sigma = \frac{8\pi\alpha^2}{3m^2} \quad (2.84)$$

respectively, which are the differential and total cross-sections for Thomson scattering.

The Breit-Wheeler Process

The Breit-Wheeler process is the annihilation of two photons to produce an electron positron pair ($\gamma\gamma \to e^+e^-$). It was first predicted in 1934 [18], however, has never been directly detected to date. By enabling two photons to interact, the Breit-Wheeler process demonstrates that the electromagnetic field is nonlinear. This shows a profound difference between QED and classical electromagnetism. The lowest order Feynman diagrams which contribute to the Breit-Wheeler process are again of second order and an example is shown in figure 2.8 (b).

The differential cross-section for the Breit-Wheeler process is given by [96]

$$\frac{\mathrm{d}\sigma_{\mathrm{BW}}}{\mathrm{d}\Omega} = \frac{\alpha^2\beta}{m^2 s}\left[\frac{1+2\beta\sin^2\theta - \beta^4 - \beta^4\sin^4\theta}{(1-\beta^2\cos^2\theta)^2}\right] \quad (2.85)$$

where $\beta = \sqrt{1-s^{-1}}$ is the velocity of the pair in the CM frame and θ is the angle between the collision and scattering axes. Integrating over solid angle we get the total Breit-Wheeler cross-section

$$\sigma_{\mathrm{BW}} = \frac{\pi\alpha^2(1-\beta^2)}{2m^2}\left[(3-\beta^4)\log\frac{1+\beta}{1-\beta} - 2\beta(2-\beta^2)\right]. \quad (2.86)$$

Equations 2.85 and 2.86 are plotted in figure 2.9 (c) and (d) respectively. The cross-section exhibits a threshold at $\sqrt{s} = 2mc^2$, below which the value is zero. This is due to the rest-mass of the electron-positron pair. For the process to occur the energy of the incoming photons must at least be large enough to create the pair at rest.

Comparing figures 2.9 (b) and (d), we can see that the total cross-section for Compton scattering and the Breit-Wheeler process are of the same order of magnitude when above the threshold. Therefore, it is not the smallness of the cross-section that has prevented the Breit-Wheeler process from being detected but the threshold behaviour. Creating the high energy photon sources, required to get over the threshold, that are also dense enough

to produce a detectable number of pairs is challenging.

Photon-Photon Scattering

In QED a photon cannot couple directly to another photon. Therefore, perturbing the S-matrix to second order yields no process where two incoming photons scatter into two outgoing photons. However, this is not the case if we expand to higher orders. Figure 2.8 (c) shows a diagram of a fourth order term ($\propto \alpha^4$) describing photon-photon scattering ($\gamma\gamma \to \gamma\gamma$). This was first discussed by Euler and Kockel in 1935 [21] and again demonstrates the non-linearity of the electromagnetic field.

The differential cross-section for photon-photon scattering is again complex and can be found in Tollis (1965) [99]. It is plotted along with the total cross-section in figures 2.9 (e) and (f). The overall shape of the differential cross-section is similar to that of the Breit-Wheeler process, however, is $\approx 10^4$ times smaller due to the factor of α^2. Therefore, photon-photon scattering has also never been directly observed in the laboratory.

Another difference between photon-photon scattering and the Breit-Wheeler process is the absence of the threshold. Taking advantage of this, experiments have been proposed to detect photon-photon scattering in the non-relativistic limit using IR lasers [100]. In this limit, the differential and total cross-sections simplify considerably to

$$\frac{\mathrm{d}\sigma_{\gamma\gamma}}{\mathrm{d}\Omega} = \frac{139}{8100}\frac{\alpha^4}{2\pi m^2}\left(\frac{\omega}{m}\right)^6(3+\cos^2\theta)^2 \tag{2.87}$$

and

$$\sigma_{\gamma\gamma} = \frac{973}{10125}\frac{\alpha^4}{\pi^2 m^2}\left(\frac{\omega}{m}\right)^6 \tag{2.88}$$

respectively. These decay rapidly with ω and along with the low value for the total cross-section, makes such experiments challenging.

2.3 Strong-Field Quantum Electrodynamics

Perturbative QED is one of the most successful theories to date, with theoretical calculations agreeing with experimental measurements to extreme precision (*e.g.* the electron magnetic moment [101]). In section 2.2.1 we made the assumption that the interaction Hamiltonian was small compared to the free Hamiltonian. This allowed us to expand in powers of H_{int}. However, we can see from equation 2.75, that if the field is large, such that the normalised vector potential $a_0 \gtrsim 1$, the perturbative calculation is destined to fail. This is demonstrated on the left of figure 2.10 which shows an electron emitting and absorbing an arbitrarily large number of photons n. Calculating the probability of this event would involve summing over all n which is not feasible.

To solve this issue we take a new approach to QED known as the Furry picture [103].

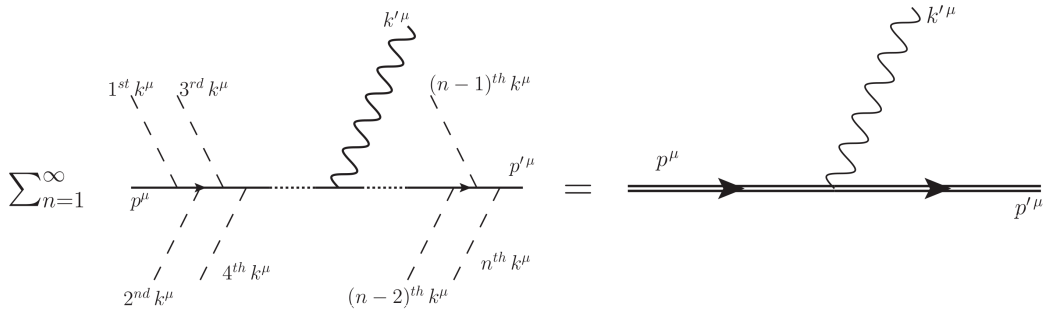


Figure 2.10: Feynman diagram of nonlinear Compton scattering in a strong field background. The left diagram demonstrates the perturbative calculation with an arbitrarily large number of photons absorbed and re-emitted. The right diagram demonstrates the equivalent process in the Furry picture with the double line of the electron representing the strong field background. Diagram from Mackenroth and Piazza 2011 [102]

This makes use of the correspondence principle [104] which allows the electromagnetic field to be treated classically, when the number of photons is large. Therefore, we can split the field into a large external background, which is treated classically, and a small quantum perturbation

$$A^\mu = A^\mu_{\text{ext}} + A^\mu_{\text{int}}. \tag{2.89}$$

This gives rise to the following Hamiltonian

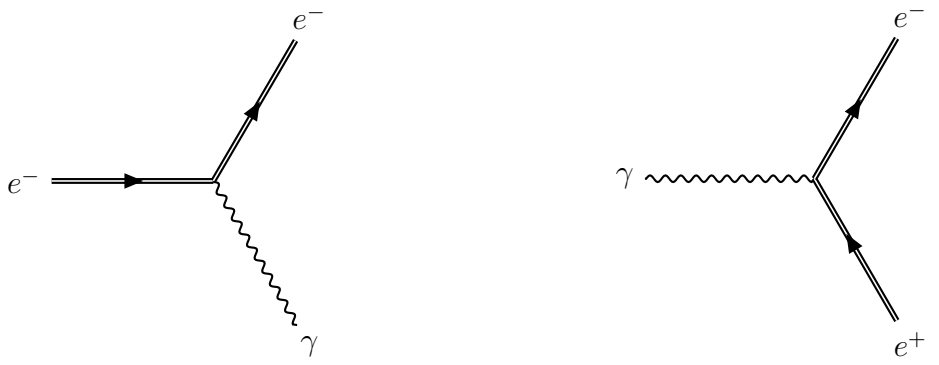
$$H = H_0 + H_{\text{ext}} + H_{\text{int}} = H'_0 + H_{\text{int}} \tag{2.90}$$

where we have absorbed the external field Hamiltonian into the free Hamiltonian. If the classical field is a plane wave, the eigenstates of H'_0 are known as Volkov states [105]. To calculate the probability of an interaction, the same method discussed in section 2.2.1 is used, however, now we assume the initial and final states are Volkov states in the asymptotic limit. To represent an electron or positron in a strong background field, a double line is used in Feynman diagrams as shown on the right of figure 2.10.

We can form a heuristic argument as to why a_0 controls nonlinear effects by expressing it in terms of $\hbar$

$$a_0 = \frac{eE}{\omega mc} = \frac{\lambda\!\!\!^{-}_C Ee}{\hbar\omega}. \tag{2.91}$$

Therefore, we can view a_0 as the energy the field exerts on an electron, over the Compton wavelength, in units of the photon's energy. If $a_0 \geq 1$ an electron is likely to absorb multiple photons during an interaction and nonlinear effects become dominant. In a similar way, we can define a field which exerts an energy of mc^2 on the electron, over the



(a) Nonlinear Compton Scattering (b) Nonlinear Breit-Wheeler process

Figure 2.11: Feynman diagrams for (a) nonlinear Compton scattering and (b) the nonlinear Breit-Wheeler process in a strong background field.

Compton wavelength

$$E_c = \frac{mc^2}{e\lambda\!\!\!^{-}} = \frac{m^2c^3}{e\hbar} \approx 1.32 \times 10^{18}\, V/m. \tag{2.92}$$

This is known as the Schwinger limit and is the field at which nonlinear quantum effects become dominant [106]. For example, a field of this strength is capable of producing pairs from the vacuum, a process known as vacuum polarisation [19, 22, 20]. It is also the field strength at which nonlinear QED effects, such as the recoil from the emission of a high energy photon, start to dominate the motion of an electron. However, the electric field is not a Lorentz invariant. Therefore, it is useful to define the following Lorentz invariant parameters

$$\eta = \frac{\sqrt{-(F_{\nu\mu}p^\mu)^2}}{mcE_c} \qquad \chi = \frac{\hbar\sqrt{-(F_{\nu\mu}k^\mu)^2}}{2mcE_c} \tag{2.93}$$

where p^μ is the electron's four momentum and k^ν is the photon's four wave vector. These are often referred to as the electron and photon quantum parameters respectively. We can view η as the field strength in the rest frame of the electron in units of E_c. Although there is no rest frame for a photon, χ is defined in a similar way. Therefore, when χ, $\eta > 1$ nonlinear quantum effects must be considered.

In this section we will study two nonlinear QED interactions, nonlinear Compton scattering and the nonlinear Breit-Wheeler process. Unlike in the weak field case the lowest order Feynman diagrams which contribute to these processes are of order α and are shown in figure 2.11. This is because momentum conservation is provided by the background field. There are of course many other nonlinear QED processes one could study. These include trident pair production [107, 108], double nonlinear Compton scattering [109, 110] and photon splitting in a strong field [111]. However, these are higher order processes and of less importance to this work.

2.3.1 Nonlinear Compton Scattering In a Plane Wave

Calculating the nonlinear Compton scattering rate, with a field configuration from a high intensity laser, is analytically challenging. Therefore, to simplify the calculations, the laser is often approximated as an infinite, monochromatic plane wave [112, 113, 114, 115, 116].

A nonlinear Compton scattering event in such a wave involves the electron absorbing n wave photons and emitting a single high energy photon ($e^- + n\gamma \rightarrow e^- + \gamma$). We can calculate the energy of the emitted photon using conservation of momentum

$$p^\mu + nk^\mu = p'^\mu + k'^\mu \tag{2.94}$$

where p^μ and p'^μ are the initial and final momentum of the electron, k^μ is the wavevector of a laser photon and k'^μ is the wavevector of the emitted photon. In the Furry picture, the motion of the electron in the background field is treated classically and can be obtained by solving the Lorentz force (equation 2.19) which gives

$$p^\mu = p_0^\mu + \frac{a_0^2 m^2}{2k \cdot p_0} k^\mu \tag{2.95}$$

where p_0^μ is the momentum of the electron before it interacts with the wave. Upon substituting this into equation 2.94 and rearranging, we get[116]

$$nk \cdot p_0 = k' \cdot p_0 + \left(n + \frac{a_0^2 m^2}{2k \cdot p_0}\right) k \cdot k' \tag{2.96}$$

where we have used $p \cdot k = p_0 \cdot k$, because k^μ is lightlike. If we assume the electron and wave collide head-on, we can write their momenta as

$$p_0^\mu = \gamma m(1, \beta\hat{\mathbf{n}}) \qquad k^\mu = \omega(1, -\hat{\mathbf{n}}) \tag{2.97}$$

and the momentum of the emitted photon as

$$k'^\mu = \omega'(1, \hat{\mathbf{n}}') \tag{2.98}$$

where $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}' = \cos\theta$. Using these definitions, we can rearrange equation 2.96 to give the frequency of the scattered photon

$$\omega' = \frac{n\omega}{1 + \kappa_n(a_0)\gamma(1-\beta)(1+\cos\theta)} \tag{2.99}$$

where

$$\kappa_n(a_0) = n\frac{\omega}{m} - \beta\gamma + \frac{a_0^2\gamma(1-\beta)}{2}. \tag{2.100}$$

Equation 2.99 gives a one-to-one mapping of the scattering angle to the energy of the

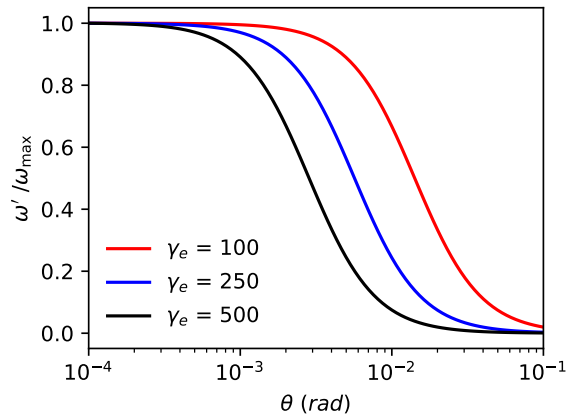


Figure 2.12: Normalised photon frequency against scattering angle for nonlinear Compton scattering in a plane wave with $a_0 = 1$.

emitted photons. It is plotted in figure 2.12 and shows the emission is highly beamed along the electron axis. This beaming becomes more pronounced for higher energy electrons. We can use equation 2.99 to investigate what happens in the low intensity limit. Letting $a_0 \to 0$, and assuming the electron is static ($\gamma = 1$), we have $\kappa_n(a_0) = n\omega/m$, which results in equation 2.82 for $n = 1$. This shows the equation is consistent with linear Compton scattering.

Having studied the kinematics, the next step is to calculate the interaction probability. In the low field case, this involves calculating a cross-section. However, a more sensible quantity when discussing an electron interacting with an infinite wave is the emission rate. A full calculation of this rate is performed by Ritus [106]. The final result for a circularly polarised wave, as a function of $x = \omega'/E_0$ where E_0 is the energy of the electron, is [116, 117]

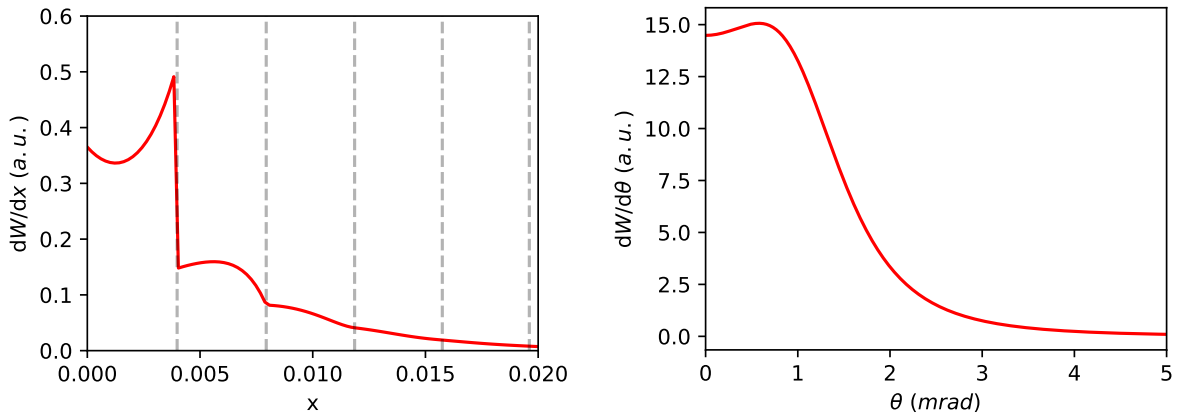
$$\frac{\mathrm{d}W_n}{\mathrm{d}x} = \frac{\alpha^2 m^2}{4E_0}\left\{ -4J_n^2(z) + a_0^2\left(1 - x + \frac{1}{1-x}\right)\left[J_{n-1}^2(z) + J_{n+1}^2(z) - 2J_n^2(z)\right]\right\} \quad (2.101)$$

where J_n is a Bessel function of the first kind and

$$z = \frac{2a_0}{\sqrt{1+a_0^2}}\frac{\sqrt{x\left(nu - (1+nu)\,x\right)}}{u(1-x)} \qquad u = \frac{4E_0\,\omega}{m^2(1+a_0^2)}. \quad (2.102)$$

Figure 2.13 (a) shows equation 2.101 summed over n for an $E_0 = 500\,\mathrm{MeV}$ electron in a field with $a_0 = 1$. Comparing this figure with figure 2.4 we can see the same harmonic structure that appears in nonlinear Thomson scattering. A formal proof that the nonlinear Compton rate tends to the nonlinear Thomson rate in the low energy limits is provided by Harvey and Heinzl [116].

As equation 2.99 provides a one-to-one mapping between ω' and θ, we can use it to



(a) Photon energy distribution

(b) Photon angle distribution

Figure 2.13: Normalised photon energy (a) and angle (b) distributions from nonlinear Compton scattering of an 500 MeV electron in a plane wave field with $a_0 = 1.0$.

derive the scattering rate as a function of angle, through the change of variables formula

$$\frac{\mathrm{d}W_n}{\mathrm{d}\Omega} = \frac{\mathrm{d}W}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}\Omega}. \tag{2.103}$$

This is plotted in figure 2.13 (b). The photons are emitted into a cone of angle $\sqrt{\langle\theta^2\rangle} \approx 1/\gamma$, the same relationship we found for classical emission.

2.3.2 Strong Field QED processes in a Quasi-Stationary Field

In recent work, calculations of nonlinear Compton scattering have been performed which go beyond the infinite wave case to include pulse effects [118]. However, to perform the calculation of a realistic experiment, there are additional effects that must be considered. To reach the high field conditions ($a_0 \gg 1$, $\chi \sim 1$) required to study strong field QED process, tightly focused laser pulses are required. This introduces spatial effects which are not accounted for in a plane wave. To overcome this, two approximations are often made.

The first is the quasi-stationary approximation, which assumes the field is static over the duration of the interaction. This allows the scattering rate to be calculated within a static field. For the field to be considered static, we require that the coherence time of the interaction, $t_{\mathrm{coh}} \approx \lambda_C E_{\mathrm{crit}}/cE_0$, is much smaller than the wave period [106]. Therefore, the quasi-stationary approximation applies if $a_0 \gg 1$.

The second approximation is that the fields are relatively weak. As we require the probability to be Lorentz invariant, there are only four invariant parameters it can depend

on. Two are given in equation 2.93 and the other two are

$$f = -\frac{F_{\nu\mu}F^{\nu\mu}}{2E_{\text{crit}}^2} = \frac{|E^2 - B^2|}{E_{\text{crit}}^2} \qquad g = -\frac{F^*_{\nu\mu}F^{\nu\mu}}{4E_{\text{crit}}^2} = \frac{|\mathbf{E}\cdot\mathbf{B}|}{E_{\text{crit}}^2} \tag{2.104}$$

where

$$F^*_{\nu\mu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F^{\alpha\beta} \tag{2.105}$$

and $\epsilon_{\mu\nu\alpha\beta}$ is the Levi-Civita symbol. For the field to be considered weak, we require that [119, 120]

$$f, g \ll 1 \qquad \eta^2 \gg \text{Max}(f, g). \tag{2.106}$$

If true, the electron/photon interaction probabilities can be written as a function of η/χ only.

As noted by Kirk et. al. [119], these approximations are likely to be appropriate in strong field QED experiments for the foreseeable future. Next generation laser facilities will reach intensities of $I \approx 10^{24}\,\text{Wcm}^{-2}$ at a wavelength of $\lambda \approx 1\,\mu\text{m}$. We can write the normalised vector potential as $a = 855\,I^{1/2}[10^{24}\text{W cm}^{-2}]\,\lambda[\mu\text{m}]$, showing the quasi-stationary condition is easily met. As for the weak field approximation, it will always hold true for a single plane wave as $g = f = 0$. For two counter-propagating waves f and g do not vanish, however, we can write $f, g < 1.6 \times 10^{-5} I[10^{24}\text{W cm}^{-2}]$. Therefore, laser intensities will have to increase by several orders of magnitude for the weak field approximation not to hold.

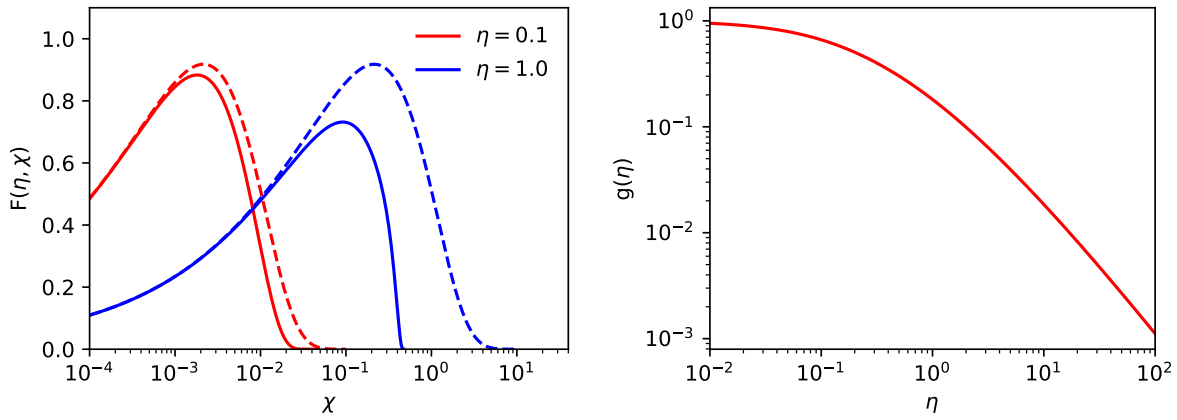
Quantum Synchrotron Radiation

Although we are interested in nonlinear Compton scattering (*i.e.* the interaction between an electron and wave), if the approximations above hold, the interaction is equivalent to quantum synchrotron emission. The differential rate of quantum synchrotron emission for a relativistic electron was found by Erber to be [39]

$$\frac{\mathrm{d}W}{\mathrm{d}\chi} = \sqrt{3}\alpha\frac{\eta}{\gamma}\frac{F(\eta,\chi)}{\chi} \tag{2.107}$$

where

$$F(\eta,\chi) = \frac{4\chi}{3\eta^2}\left[\left(1 - \frac{2\chi}{\eta} + \frac{1}{1-2\chi/\eta}\right)K_{2/3}(\delta) - \int_\delta^\infty K_{1/3}(t)\mathrm{d}t\right] \tag{2.108}$$



(a) Synchrotron functions

(b) Gaunt factor

Figure 2.14: Synchrotron emission plots. (a) shows the quantum synchrotron function (solid line) and the classical synchrotron function (dashed line) in both the low energy case ($\eta = 0.1$, red) and the high energy case ($\eta = 1.0$, blue). (b) shows the Gaunt factor as a function of η.

is the quantum synchrotron spectrum and

$$\delta = \frac{4\chi}{3\eta^2}\left(\frac{2\chi}{\eta} - 1\right)^{-1}. \tag{2.109}$$

In the classical limit, $\hbar \to 0$, this function becomes [119]

$$F(\eta, \chi) \to f_{sync}\left(\frac{4\chi}{3\eta^2}\right) \tag{2.110}$$

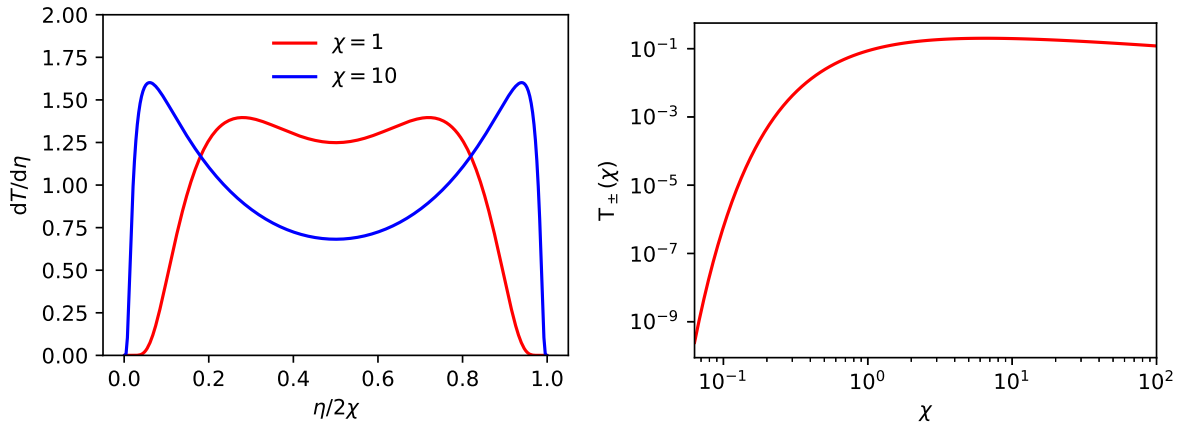
where $f_{sync}(x)$ is the classical function given by equation 2.55.

Both the classical and quantum synchrotron functions are plotted in figure 2.14 (a). For low energy photon emission, there is good agreement between the classical and quantum emission. However, at high fields strengths, the classical function gives a non-zero probability of emitting a photon with a higher energy than the electron itself. This unphysical behaviour does not appear for the quantum function.

Radiation Reaction

To obtain the average power emitted by an electron in a strong field, we can multiply equation 2.107 by the emitted photon energy and integrate, giving

$$P = \frac{2}{3}\alpha\eta^2 g(\eta) \tag{2.111}$$



(a) Differential pair emissivity

(b) Total pair emissivity

Figure 2.15: Nonlinear Breit-Wheeler plots. (a) shows the differential pair emission function for a low energy ($\chi = 1$, red) and a high energy ($\chi = 10$, blue) interaction. (b) shows the total pair emission function

where

$$g(\eta) = \frac{3\sqrt{3}}{2\pi\eta^2}\int_0^{\eta/2} F(\eta,\chi)\mathrm{d}\chi \tag{2.112}$$

which is plotted in figure 2.14 (b). As we saw for both classical synchrotron radiation (see section 2.1.6) and nonlinear Compton scattering in a plane wave (see section 2.3.1), the radiation is emitted into a cone of angle $\sqrt{\langle\theta^2\rangle} \approx 1/\gamma$. This is effectively parallel to the direction of motion for an ultra-relativistic electron. Therefore, using equation 2.111, the force on the electron is

$$\mathbf{F} = -\frac{2}{3}\alpha\eta^2 g(\eta)\hat{\mathbf{p}} \tag{2.113}$$

where $\hat{\mathbf{p}}$ is the direction of the electron's motion. By setting $g(\eta) = 1$ we get the ultra-relativistic limit of the Landau-Lifshitz force [117]. Therefore, we can view $g(\eta)$ as a quantum correction and is known as the Gaunt factor.

Equation 2.113 still misses a key aspect of the quantum motion of an electron in a strong field. By calculating a continuous, average force, we are neglecting the stochastic behaviour of emission. This leads to an effect known as straggling [121] and allows the electron to enter classically forbidden regions.

The Nonlinear Breit-Wheeler Process

Just as Compton scattering can occur in a strong background field, so can the Breit-Wheeler process ($\gamma' + n\gamma \to e^-e^+$). Again using the quasi-static, weak-field approximations Erber found the differential Breit-Wheeler decay rate, as a function of the photon's

quantum parameter χ, to be [39]

$$\frac{\mathrm{d}W_{\pm}}{\mathrm{d}\eta_{-}} = \frac{\alpha m}{\omega}\chi\frac{\mathrm{d}T(\chi)}{\mathrm{d}\eta_{-}} \tag{2.114}$$

where η_{-} is the quantum parameter for the electron and [122]

$$\frac{\mathrm{d}T(\chi)}{\mathrm{d}\eta_{-}} = \frac{1}{2\sqrt{3}\pi\chi^2}\left[\left(\frac{2\chi}{\eta_{-}} - 1 + \frac{1}{2\chi/\eta_{-} - 1}\right)K_{2/3}(\delta) - \int_{\delta}^{\infty} K_{1/3}(t)\mathrm{d}t\right] \tag{2.115}$$

is the differential pair emissivity function, and δ is given by equation 2.109. Upon normalising $\mathrm{d}T_{\pm}(\chi)/\mathrm{d}\eta_{-}$, we obtain the probability distribution for the emission of an electron with quantum parameter η_{-}. As the electron and positron have equal mass, the probability of scattering at a specific energy is equal. Therefore, $\mathrm{d}T(\chi)/\mathrm{d}\eta_{-} = \mathrm{d}T(\chi)/\mathrm{d}\eta_{+}$, where η_{+} is the quantum parameter of the positron, and the distribution is symmetric around $\eta_{\pm} = \chi$. A plot of $\mathrm{d}T(\chi)/\mathrm{d}\eta_{\pm}$ is shown in Figure 2.15 (a). For low energy photons, the energy is spread evenly between the pair. However, for higher energy photons, one particle is likely to receive most of the energy.

Upon integrating equation 2.114, we get the total rate for pair production

$$W_{\pm} = \frac{\alpha m}{\omega}\chi T(\chi) \tag{2.116}$$

where $T(\chi)$ is the total pair emissivity function and is plotted in figure 2.15 (b). In the limits of small and large χ, $T_{\pm}(\chi)$ can be approximated as

$$T_{\pm}(\chi) \propto \begin{cases} \exp\left(-\frac{4}{3\chi}\right) & \chi \ll 1 \\ \chi^{-1/3} & \chi \gg 1 \end{cases}. \tag{2.117}$$

At small χ, the probability of pair production is highly suppressed. $T_{\pm}(\chi)$ reaches a maximum at $\chi \approx 8$ before slowly decaying at higher values.

Chapter 3

Machine learning and Monte Carlo Methods

The aim of this work has been to develop the capabilities to carry out large scale simulations of quantum systems. Due to the inherent random nature of quantum systems, statistical methods are an invaluable tool for their study. Statistical methods also form the basis of all machine learning techniques, which have been used throughout this work. Therefore, We will start this chapter with a brief discussion of statistical inference. We will then move on to discuss the machine learning methods used in this work to reduce the computational cost of the large scale simulations. Finally, we will discuss the Monte Carlo method which is used to study quantum systems.

3.1 Maximum Likelihood Estimation and Bayesian Inference

As physicists, our goal is to observe some physical system and derive a model that predicts how the system behaves. Given a set of observations of the system, statistical inference is a useful tool for solving this problem. It allows us to make predictions about unknown parameters in our model or test whether our model is to be believed given the observations. Two commonly used methods for statistical inference are maximum likelihood estimation (MLE) and Bayesian inference (BI). Here we will give a quick overview of both methods and discuss when one might be chosen over the other. However, before starting it is useful to quickly review the framework of statistical inference. We will assume we have a series of observations which form an independent and identically distributed (i.i.d.) data set $\mathcal{D} = \{\mathbf{x}_i | i = 1, \ldots, N\}$ where N is the number of samples. To keep the discussion general, each observation, $\mathbf{x}_i$, is a multidimensional random variable of size D. We will also assume $\mathbf{x}_i$ is generated by some unobserved probability density function $p(\mathbf{x} \mid \boldsymbol{\theta})$. Here, $\boldsymbol{\theta}$ is a vector of the model parameters and our goal is to infer it from the data set.

3.1.1 Maximum Likelihood Estimation

MLE is a method for finding the value of $\boldsymbol{\theta}$ that gives the highest probability of generating the observed data. For a given value of $\boldsymbol{\theta}$, the joint probability of observing the data set is

$$L(\boldsymbol{\theta}) = p(\mathbf{X} \mid \boldsymbol{\theta}) = \prod_{i=1}^{N} p(\mathbf{x}_i \mid \boldsymbol{\theta}) \tag{3.1}$$

where $\mathbf{X}$ is a matrix formed by stacking the vectors $\mathbf{x}_i$ row-wise. $L(\boldsymbol{\theta})$ is known as the likelihood function and the value of $\boldsymbol{\theta}$ which maximises it is known as the maximum likelihood estimate. However, because equation 3.1 is the product of probabilities, finding the maximum is prone to arithmetic underflow errors when performed numerically. This problem can be solved by maximising the logarithm of $L(\boldsymbol{\theta})$ instead, which occurs at the same value of $\boldsymbol{\theta}$, as the logarithm function is monotonically increasing. Therefore, the MLE is given by

$$\boldsymbol{\theta}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta}} \sum_{i=1}^{N} \log\, p(\mathbf{x}_i \mid \boldsymbol{\theta}) \tag{3.2}$$

To carry out MLE we must first assume the functional form of $p(\mathbf{x} \mid \boldsymbol{\theta})$. The process of doing so is is known as *model selection*, which is discussed in detail in sections 3.2.1 and 3.2.3. However, it can be an expensive process and it is often more convenient to simply assume the data is Gaussian distributed

$$p(\mathbf{x} \mid \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-k/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\Big(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \Big) \tag{3.3}$$

where the two model parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance respectively and k is the number of dimensions. Using equation 3.2 we can find the MLE for the model parameters

$$\boldsymbol{\mu}_{\mathrm{ML}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x_i} \qquad\qquad \boldsymbol{\Sigma}_{\mathrm{ML}} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x_i} - \boldsymbol{\mu}_{\mathrm{ML}})(\mathbf{x_i} - \boldsymbol{\mu}_{\mathrm{ML}})^T. \tag{3.4}$$

There are a variety of reasons for assuming a Gaussian distribution. For example, in systems with a large number of particles, a Gaussian distribution can appear due to the central limit theorem. It may also just be mathematical convenience to assume one. Many problems can be solved analytically with a Gaussian distribution whereas a numerical approach would be required with other distributions [76].

As we will see in the next section, BI is an integral problem (equation 3.5) whereas MLE is an optimisation problem. Integral problems are more computationally expensive than optimisation problems and this is the main reason one might choose MLE over BI. However, being an optimisation problem, MLE suffers from the well known issue of finding a local minimum rather than the global minimum if $L(\boldsymbol{\theta})$ is non-convex. Another issue

with MLE is that it can be heavily biased for small sample sizes. As an example, if we flip a coin three times and observe three heads, using MLE we would assume the probability of landing heads is 1. Both these issues can be removed by moving to a Bayesian treatment.

3.1.2 Bayesian Inference

In BI, instead of assuming $\boldsymbol{\theta}$ has a true value, we treat it as random variable with an associated probability distribution. Our goal now shifts from estimating the most probable value of $\boldsymbol{\theta}$ to estimating the distribution over $\boldsymbol{\theta}$ given our data $p(\boldsymbol{\theta} \mid \mathbf{X})$. This gives an intuitive framework for understanding the uncertainty in our model parameters. To calculate $p(\boldsymbol{\theta} \mid \mathbf{X})$ we relate it back to $L(\boldsymbol{\theta})$ through Bayes' theorem

$$p(\boldsymbol{\theta} \mid \mathbf{X}) = \frac{p(\mathbf{X} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int p(\mathbf{X} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta}) \mathrm{d}\, \boldsymbol{\theta}}. \tag{3.5}$$

Here, $p(\boldsymbol{\theta})$ reflects our knowledge of $\boldsymbol{\theta}$ before observing the data and is known as the prior distribution. Using the observed data, equation 3.5 updates our knowledge of $\boldsymbol{\theta}$ to give the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{X})$. The denominator of equation 3.5 is known as the marginal likelihood or *evidence* and is a D dimensional integral over the model parameters.

The appearance of a prior in equation 3.5 is a key aspect of Bayesian methods. It solves the bias problem of MLE when the sample size is small. Using BI on the coin flip example with a non-informative prior (a uniform distribution between 0 and 1) we will obtain a less extreme estimate of the probability of heads. However, some people regard the inclusion of a prior in Bayesian methods as a flaw. If an inappropriate prior is used, BI can also lead to a biased estimate. Setting the prior is also a subjective choice and is often chosen for analytical convenience rather than being based on knowledge of the system [123].

The computationally expensive part of performing BI is calculating the marginal in equation 3.5. Before the development of Markov chain Monte Carlo sampling methods (see section 3.3.1) and the recent increased in computational power, BI was only practical when equation 3.5 can be solvable analytically. This occurs for a given likelihood function if the form of the prior is chosen such that the posterior is in the same probability distribution family. The prior which has this property is called the conjugate prior of the likelihood function [124]. A convenient mathematical property of Gaussian distributions is that the conjugate prior, if the covariance is fixed, is itself Gaussian. If we use a Gaussian prior on the mean $p(\boldsymbol{\mu}) = \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, we obtain a Gaussian posterior $p(\boldsymbol{\mu} \,|\, \mathbf{X}) = \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*)$ where

$$\begin{aligned} \boldsymbol{\mu}_* &= \boldsymbol{\Sigma}_* \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 + N \boldsymbol{\Sigma}_* \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{\mathrm{ML}} \\ \boldsymbol{\Sigma}_*^{-1} &= \boldsymbol{\Sigma}_0^{-1} + N \boldsymbol{\Sigma}^{-1}. \end{aligned} \tag{3.6}$$

As $N \to \infty$, $\boldsymbol{\Sigma}_* \to O$ and $\boldsymbol{\mu}_* \to \boldsymbol{\mu}_{\text{ML}}$ and the posterior becomes infinitely peaked around the MLE solution. BI and MLE are therefore equivalent for an arbitrarily large data set.

For a full Bayesian approach we should not fix $\boldsymbol{\Sigma}$ and treat it as a random variable. There is no direct conjugate prior for $\boldsymbol{\Sigma}$, however, we can define a conjugate prior for its inverse, known as the precision matrix, and is given by the Wishart distribution [125]. When both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown, the conjugate prior is simply the product of their individual conjugate priors, known as the Gaussian-Wishart distribution. In section 3.2.3 we will see this analytical method of solving equation 3.5 is useful for Bayesian regression.

Even if equation 3.5 is too computationally expensive to solve and it is not appropriate to assume a conjugate prior, it can still be useful to define a prior on the model parameters. Using the same method as MLE, we can make a point estimate of $\boldsymbol{\theta}$ by maximising equation 3.5. This is known as the maximum a posterior (MAP) estimate. The marginal does not depend on $\boldsymbol{\theta}$ so the MAP estimate is

$$\boldsymbol{\theta}_{\text{MAP}} = \arg\max_{\boldsymbol{\theta}} \; p(\mathbf{X} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta}). \tag{3.7}$$

This is useful to constrain the estimate of $\boldsymbol{\theta}$ when the data set is small [76].

3.2 Machine Learning Methods

Machine learning algorithms can be broadly divided into three categories, supervised learning, unsupervised learning and reinforcement learning [76]. In this work, we are predominantly interested in regression methods, which fall under the supervised learning category. In regression, our data set is replaced with a set of N input-target pairs, $\mathcal{D} = \{\mathbf{x}_i, t_i | i = 1, ..., N\}$, where $\mathbf{x}_i$ is the same D dimensional vector as before, and t_i is a single valued variable. We will assume the data has been generated from the following model

$$t(\mathbf{x}) = f(\mathbf{x}) + \epsilon \tag{3.8}$$

where $f(\mathbf{x})$ is some unknown function and ϵ is a random noise term. Our goal is to learn the mapping function, $f(\mathbf{x})$, from our data set $\mathcal{D}$.

Regression can be approached from either a maximum likelihood or Bayesian stance. This section will start by discussing the maximum likelihood approach, and the basic method of linear regression. We will then build up to more powerful methods such as neural network regression. The same approach will be taken when discussing Bayesian regression, starting with linear Bayesian regression and building up to Gaussian process regression.

3.2.1 Linear Regression

In linear regression we assume $f(\mathbf{x})$ can be written as a linear function of the model parameters

$$f(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \ldots + \theta_D x_D = \boldsymbol{\theta}^T \mathbf{x} \tag{3.9}$$

where $\boldsymbol{\theta}$ are the model parameters (sometimes referred to as weights), and $\mathbf{x} = (1, x_1, ..., x_D)^T$ has been extended to account for a fixed offset. Using MLE, we can find the value of $\boldsymbol{\theta}$ that is most likely to generate the data set. If we assume ϵ is Gaussian with zero mean and variance σ^2, the likelihood is given by

$$p(\mathbf{t}|\mathbf{X}, \boldsymbol{\theta}, \sigma) = \prod_{n=i}^{N} \mathcal{N}(t_n|\boldsymbol{\theta}^T \mathbf{x}_n, \sigma^2) \tag{3.10}$$

and the log-likelihood is

$$\log p(\mathbf{t}|\mathbf{X}, \boldsymbol{\theta}, \sigma) = \frac{N}{2} \log\left(\frac{1}{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2}(\mathbf{t} - \boldsymbol{\theta}^T\mathbf{X})^T(\mathbf{t} - \boldsymbol{\theta}^T\mathbf{X}) \tag{3.11}$$

where and $\mathbf{t}$ is a vector of the target variables. Inserting equation 3.11 into equation 3.2 and maximising for $\boldsymbol{\theta}$ and σ^2 separately gives

$$\begin{aligned} \boldsymbol{\theta}_{\text{LR}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{t} \\ \sigma^2_{\text{LR}} &= \frac{1}{N}(\mathbf{t} - \boldsymbol{\theta}^T_{\text{ML}}\mathbf{X})^T(\mathbf{t} - \boldsymbol{\theta}^T_{\text{ML}}\mathbf{X}). \end{aligned} \tag{3.12}$$

An example of linear regression applied to a synthetic data set is shown in figure 3.1. A different approach which yields the same result as equation 3.12 involves minimising the sum of the squared residuals and is known as least-squares regression. However, MLE is a more flexible framework as it is not limited to Gaussian data. For example, Poisson regression is often used for discrete data sets. However, there is no closed form solution to equation 3.2 in this case and a numerical treatment is required [126].

Feature Mapping

The linear model given by equation 3.9 is useful if the data itself is linear. If this is not the case, linear regression has a poor predictive performance. The top left plot of figure 3.2 shows data which has been generated by the function $t(x) = 0.6\sin(2\pi x) + \mathcal{N}(0, 0.25)$ and the top right plot shows a linear fit to this data set. We can clearly see the linear fit does not capture the features of the data set. However, this problem can be solved through feature mapping.

The term linear regression is used to refer to a model which is a linear function of the weights. However, it does not have to be a linear function of $\mathbf{x}$ [76]. A more generic linear

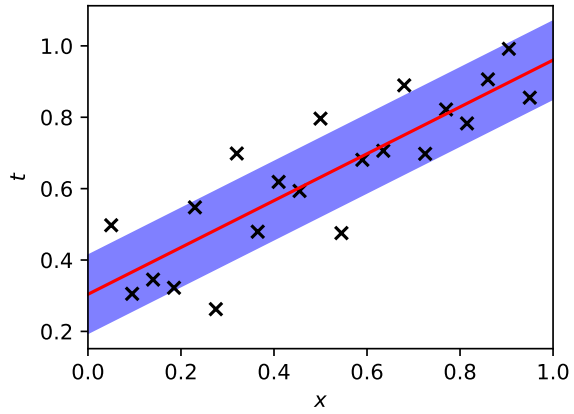


Figure 3.1: A linear regression fitted to synthetic data. The data was generated by the function $t(x) = 0.3 + 0.7x + \mathcal{N}(0, 0.1)$. The red line shows the MLE solution with $\boldsymbol{\theta}_{\text{LR}} = [0.304, 0.656]$. The blue shaded area is $\pm 1\sigma$ from the MLE solution with $\sigma_{\text{LR}} = 0.109$.

model than equation 3.9 can be formed by taking the sum of M nonlinear functions of $\mathbf{x}$

$$\begin{aligned} f(\mathbf{x}) &= \theta_0 + \theta_1\phi_1(\mathbf{x}) + \ldots + \theta_M\phi_M(\mathbf{x}) \\ &= \boldsymbol{\theta}^T\boldsymbol{\phi}(\mathbf{x}) \end{aligned} \tag{3.13}$$

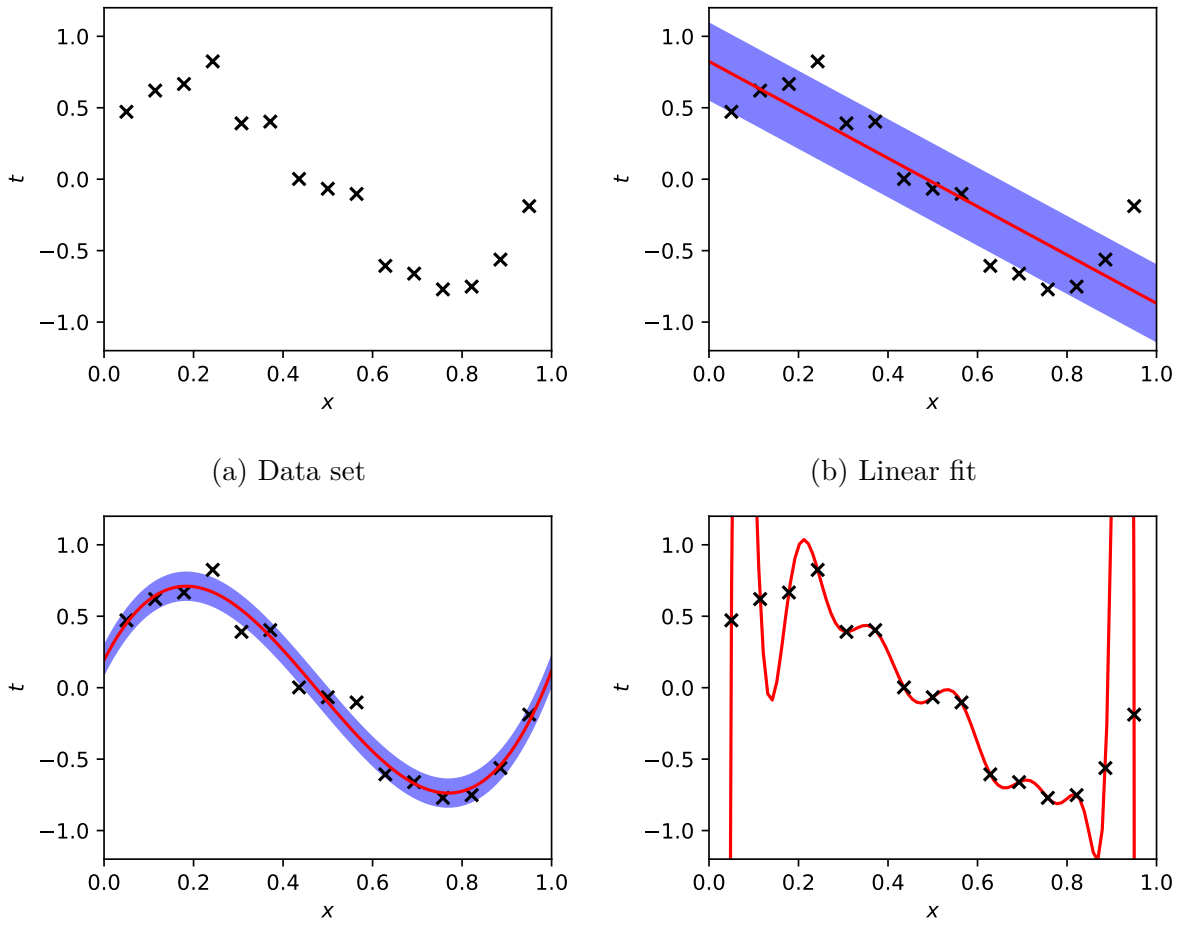
where $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_M)^T$ and $\boldsymbol{\phi} = (1, \phi_1, \ldots, \phi_M)^T$. The set of nonlinear functions $\{\phi_j(\mathbf{x})\}$ are known as basis functions. We can view the basis functions as mapping $\mathbf{x}$ from a D dimensional space into a M dimensional space. This is known as feature mapping. Performing MLE for equation 3.13, again assuming a Gaussian likelihood, gives

$$\begin{aligned} \boldsymbol{\theta}_{\text{FM}} &= (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{t} \\ \sigma^2_{\text{FM}} &= \frac{1}{N}(\mathbf{t} - \boldsymbol{\theta}^T_{\text{ML}}\boldsymbol{\Phi})^T(\mathbf{t} - \boldsymbol{\theta}^T_{\text{ML}}\boldsymbol{\Phi}^T) \end{aligned} \tag{3.14}$$

where $\boldsymbol{\Phi}$ is the $N \times D$ design matrix, formed by applying $\boldsymbol{\phi}(\mathbf{x})$ to each data point and stacking the output row-wise. Replacing $\boldsymbol{\Phi}$ with $\mathbf{X}$ in equation 3.14 yield the same result as equation 3.12.

Polynomial regression is an example where feature mapping is used. Assuming we have a single input dimension ($D = 1$), the basis functions for polynomial regression are $\{\phi_j(x) = x^j\}$. We can apply polynomial regression to the data set in 3.2 (a) for a better fit. The data has two stationary points, therefore, it is reasonable to assume $M = 3$ will give a good fit. This is shown in figure 3.2 (c) and indeed the fit better captures the features of the data. The estimate of σ^2 is also significantly reduced for the polynomial fit.

We can learn a more complex model by increasing M. A more complex model can pass closer to the data points and reduce the estimated variance. If we increase M above the number of data points, our model has enough freedom to pass through all the points



(a) Data set

(b) Linear fit

(c) $M = 3$

(d) $M = 15$

Figure 3.2: Linear regression with feature mapping, fitted to synthetic data. (a) shows the data set generated by the function $t(x) = 0.6\sin(2\pi x) + \mathcal{N}(0, 0.25)$, (b) shows a linear fit to the data set, (c) shows an $M = 3$ order fit, and (d) shows an $M = 20$ order fit.

exactly. This is shown in figure 3.2 where $M = 15$. This mode may fit the data set perfectly, however, it will not generalise well as it had extracted the residual variation of the data as a feature [127]. This problem is known as *over-fitting*. Over-fitting is a common problem in machine learning algorithms, especially for large complex models.

Regularisation

Regularisation is a method used to combat over-fitting. Without it, we would not be able to learn a complex model on data sets of limited size. The oscillations between data points in figure 3.2 (d) is caused by the weights exploding to large values [76]. To prevent this, a penalisation term is included when maximising the log likelihood. The sum of the likelihood and penalisation is known as a loss or cost function

$$E(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) + P(\boldsymbol{\theta}). \tag{3.15}$$

The most commonly used penalisation term is the sum of the squares of the weights

$$P(\boldsymbol{\theta}) = \frac{\lambda}{2\sigma^2}\boldsymbol{\theta}^T\boldsymbol{\theta} \tag{3.16}$$

where λ is the regularisation coefficient, and controls the strength of the regularisation. Parameters such as λ, which are not obtained during the learning processes, are known as *hyperparameters*. Maximising equation 3.15 with respect to $\boldsymbol{\theta}$ gives

$$\boldsymbol{\theta}_{\mathrm{RR}} = (\lambda\mathbf{I} - \boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{t} \tag{3.17}$$

which is known as ridge-regression [128].

Setting the value of λ is important for creating a model which generalises well to new data. In each of the plots in figure 3.3 a polynomial model with $M = 15$ is fitted to the same data set as figure 3.2. In figure 3.3 (a) λ is set too large and the model does not fit the data well. This is known as *under-fitting* In figure 3.3 (b) λ has been reduced significantly and the model does fit the data set well. However, it exhibits large oscillations, which suggest the model is once again over-fitting. Finally, figure 3.3 (c) shows a model where λ has been optimised. This model is expected to generalise well to unseen data. Optimising λ is another problem which falls under model selection. This will be discussed in the following section.

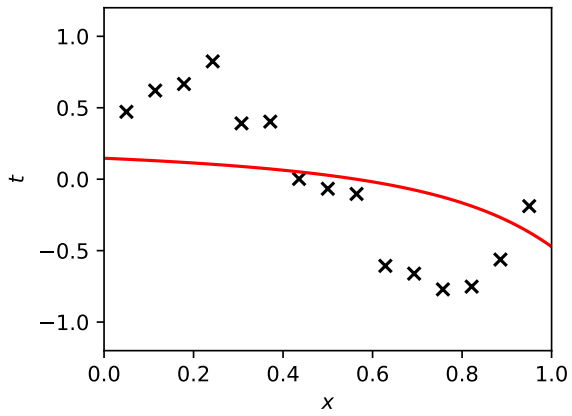
Other regularisation terms can be used. These are often of the form

$$P(\boldsymbol{\theta}) = \frac{\lambda}{2\sigma^2}\sum_{j=1}^{M}|\theta_j|^q \tag{3.18}$$

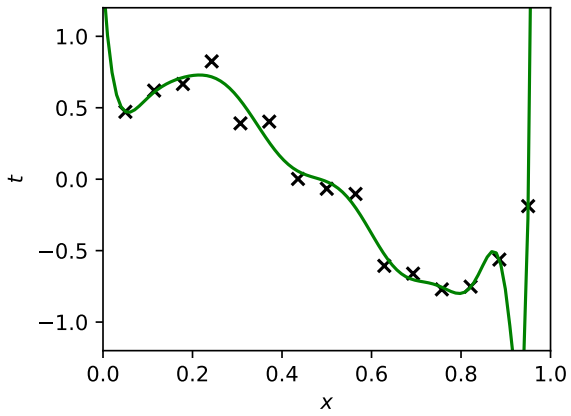
where $q = 2$ gives ridge-regression. When $q = 1$ equation 3.18 is known as the Lasso regression penalty [129]. This penalty has the benefit of driving some of the model weights to zero. This allows us to start from a complex model then learn the most important features, producing a simpler model. However, there is no analytical solution for Lasso regression and the optimisation must be performed numerically.

Model Selection

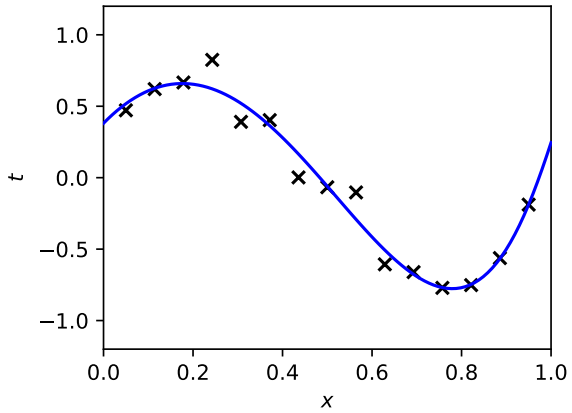
There are aspects of the model that are not learned during the training process. Two examples we have encountered are hyperparameters and the form of $L(\boldsymbol{\theta})$. Optimising these is known as model selection. One of the most common approaches to model selection is cross-validation [130]. In cross-validation, the data set is separated into a training set and a validation set. A range of different models are trained on the training set. The likelihood function is then calculated for each model using the validation set. This is used as a model performance metric as it gives the ability of the model to generalise to unseen data. The model with the maximum likelihood is chosen as the optimum model.



(a) $\lambda = 10$



(b) $\lambda = 10^{-10}$



(c) $\lambda = 10^{-4}$

Figure 3.3: Ridge-regression fits to a synthetic data set generated by $t(x) = 0.6\sin(2\pi\, x) + \mathcal{N}(0, 0.25)$. (a) $\lambda = 10$ and the model is under-fitting. (b) $\lambda = 10^{-10}$ and the model is over-fitting. (c) $\lambda = 10^{-4}$ and the model is optimised

Cross-validation works well for large data sets. However, it does not work as well for small data sets as the partitioning reduces the size of the training set further. A small validation set is also an issue as the performance metric may be poor due to noise. To solve theses issues, k-fold cross-validation is often used. Here, the data set is separated into k subsets. Training is then performed on $k-1$ subsets and validation on the remaining subset. This is then repeated k times, where a different subset is left out each time. The model which performs best on average is selected as the optimum. This provides a robust method for model selection, however, also greatly increases the computational time as k times more models are trained [76].

The are other approaches to model selection, such as information criterion methods that do not require the partitioning of the data set. These include the Akaike information criterion (AIC) [131] and the Bayesian information criterion (BIC) [132]. The AIC and BIC metrics are simple to calculate, however, they tend to favour overly simple models [76]. In section 3.2.3 we will see a Bayesian analysis provides a robust approach to model

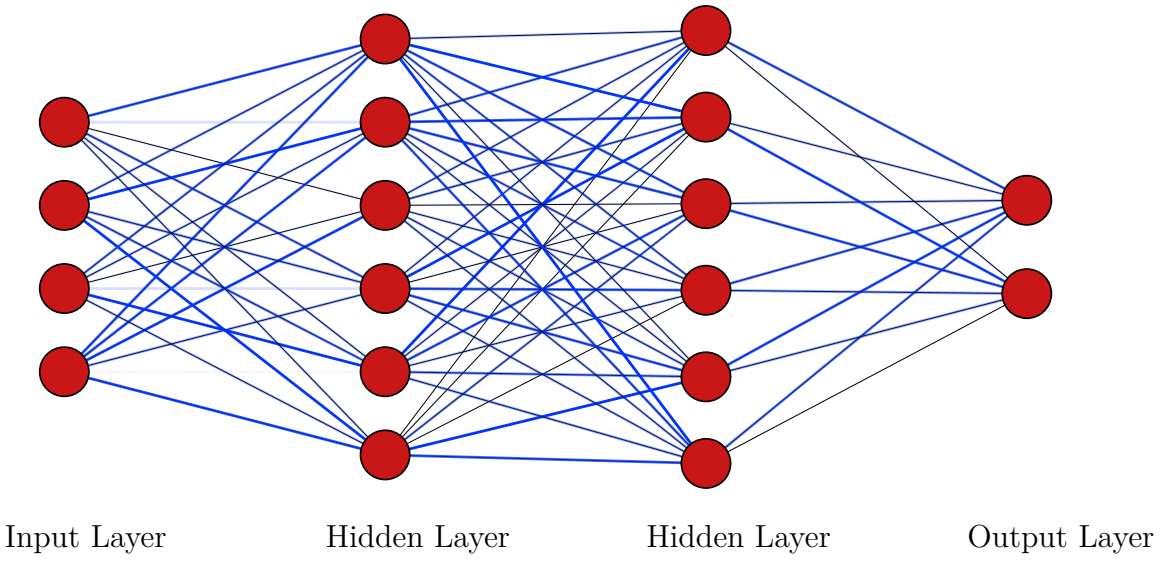


Figure 3.4: Diagram of feedforward neural network with two hidden layers. The transparency of the connections between nodes represents the weights θ_{ij}. Diagram was created with NN-SVG [133].

selection that does not involve partitioning the data.

3.2.2 Neural Network Regression

Linear regression is of limited use as we are required to set the basis functions before training. If we have prior knowledge of the data set features, we can select appropriate basis functions for the data. If we do not have prior knowledge, we must use a large number of basis functions to cover a wide range of possible features. The number of basis sets rises exponentially with the dimension of the data. Therefore, this is not a practical approach for high dimensional data. Neural network regression gets around this problem by using nonlinear parametric basis functions which adapt during the training process.

Artificial neural networks (NNs), are a collection of connected nodes which are arranged into layers. They are loosely based on the biological network of the brain. By encoding $f(\mathbf{x})$ within the network they can be used as a nonparameteric regression method. In a feedforward NN all the layers are fully connected to adjacent layers and the connections do not form loops. A diagram of a feedforward NN is shown in figure 3.4. To make a prediction with a pre-trained NN, the components of $\mathbf{x}$ are passed into the nodes of the input layer. The output of the nodes in the second layer are given by the sum of the inputs multiplied by a weight, θ_{ij}, and passed through a nonlinear activation function $h(\cdot)$. This then continues through the network and $f(\mathbf{x})$ is obtained at the output layer. The transparency of the connections in figure 3.4 represent the value of θ_{ij}.

Neural Network Functional Form

We will now give a more mathematically rigorous description of neural network regression [76, 134]. As we have already mentioned, NN regression gets around the problem of fixed basis functions by allowing them to change during the training process. The basis functions are chosen to have the same form as 3.9, transformed by a nonlinear activation function

$$z_j = h\left(\sum_{i=0}^{D} \theta_{ij}^{(1)}\, x_i\right) \tag{3.19}$$

where $\theta_{ij}^{(1)}$ are the weights where (1) refers to the first set of connections and $i = 0, ..., D$ and $j = 0, ..., M$ are the indices of the nodes in the first and second layer respectively. There are many choices of activation function which change how the NN learns. Model selection can be used to choose the optimum activation function. However, this is computationally expensive, so in this work we have used the sigmoid function

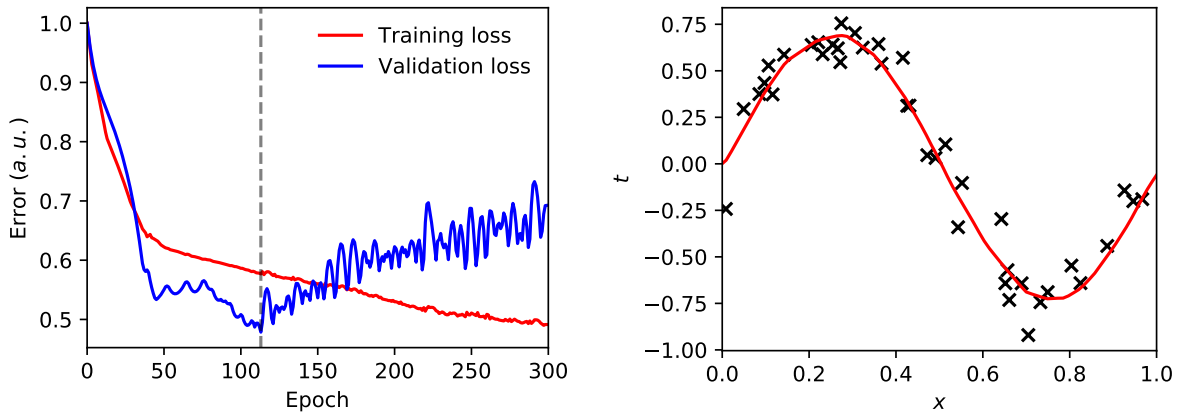
$$h(x) = \frac{1}{1 + e^{-x}} \tag{3.20}$$

which is one of the more commonly used activation functions [134]. For a multidimensional target, with K dimensions, the network function is given by

$$f_k(\mathbf{x}, \boldsymbol{\theta}) = \sum_{j=0}^{M} \theta_{kj}^{(2)}\, h\left(\sum_{i=0}^{D} \theta_{ij}^{(1)}\, x_i\right) \tag{3.21}$$

where $k = 1, ..., K$. To add extra layers to the network, we simply pass equation 3.21 through the activation function and use this as the basis function. Extra layers allow for progressively higher-level features to be extracted. However, even single hidden layer NNs can approximate any continuous function to arbitrary accuracy if the network has a sufficiently large number of hidden nodes [135].

Some of the topics discussed in section 3.2.1 can also be applied to NN regression. Deep (many layers) NNs often have a large number of free parameters, making overfitting a problem. They are therefore better suited to large data sets. However, overfitting can be reduced by adding a regularisation penalty for θ_{ij}. In the deep learning literature this is referred to as weight decay. Model selection is also an important part of designing an accurate NN. NNs have a large number of hyperparameters which are not learned during the training process such as the number of nodes and layers in the network. Cross-validation can be used, however, this can come at a large computational cost, as it involves training many neural networks. To reduce the computational costs, intelligent optimisation algorithms have been designed which make use of Bayesian optimisation [136].



(a) NN fit to data set

(b) Training and validation loss

Figure 3.5: Example of a NN regression. (a) Shows the training and validation loss. The dash grey line shows the iteration at which the training should stop. (b) Shows the data set and NN regression fit.

Training a Neural Network

To optimise the weights, we will use MLE. The log-likelihood is given by

$$\log L(\boldsymbol{\theta}) = \frac{N}{2}\log\left(\frac{1}{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2}\sum_{k=1}^{K}\sum_{n=1}^{N}\{t_{kn} - f_k(\mathbf{x}_n, \boldsymbol{\theta})\}^2 \tag{3.22}$$

where we have once once again assumed the data is Gaussian distributed. $f_k(\mathbf{x}_n, \theta)$ is a complicated function, so there is no closed form solution to $\nabla_\theta \log L(\boldsymbol{\theta}) = 0$. Therefore, a numerical method such as gradient descent is used. This involves randomly initialising $\boldsymbol{\theta}$ and updating it by

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta\nabla L(\boldsymbol{\theta}^{(i)}) \tag{3.23}$$

where η is known as the *learning rate*. Most modern deep learning packages, (*e.g.* TensorFlow [137]) include extensions to gradient descent such as the AdaGrad [138] and Adam [139] algorithms. These offer a reduction in the learning time compared to basic gradient descent. These algorithms require an efficient method of calculating $\nabla L(\boldsymbol{\theta}^{(i)})$. This is carried out using a method called backpropagation. This involves calculating the derivative by repeatedly applying the chain rule and iterating from the end of the network back. A more thorough discussion of the backpropagation algorithm is given by Bishop [76].

Due to the large number of parameters in the neural network, this training algorithm will likely cause overfitting. To prevent this, cross-validation is used during the training process. During the first few interactions of equation 3.23 the validation loss will decrease. However, once the model begins to overfit the validation loss will increase. This is shown in figure 3.5 (a). At the turning point, the model is optimised for generalising to unseen data. Figure 3.5 (b) shows the optimised model fitted to the data set.

3.2.3 Bayesian Regression

Bayesian regression offers a more robust way to assess uncertainty in the model. It also provides us with the ability to compare models without splitting the data into a test and validation set. In reviewing Bayesian regression, we will again discuss the linear regression model given by equation 3.13, however, we will now treat $\boldsymbol{\theta}$ as a random variable. Assuming the data is Gaussian distributed (see equation 3.10), if we assign a Gaussian prior on $\boldsymbol{\theta}$, the posterior will also be Gaussian due to the conjugate prior property. Assuming a zero mean prior, $p(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta} \mid \mathbf{0}, \boldsymbol{\Sigma}_0)$, the posterior is given by $p(\boldsymbol{\theta} \mid \mathbf{t}, \mathbf{X}) = \mathcal{N}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_\mathrm{N}, \boldsymbol{\Sigma}_N)$ where

$$\boldsymbol{\theta}_\mathrm{N} = \sigma^{-2}\boldsymbol{\Sigma}_\mathrm{N}\boldsymbol{\Phi}^T\mathbf{t} \qquad \boldsymbol{\Sigma}_\mathrm{N}^{-1} = \boldsymbol{\Sigma}_0^{-1} + \sigma^{-2}\boldsymbol{\Phi}^T\boldsymbol{\Phi} \tag{3.24}$$

are the posterior mean and covariance respectively. If we let $\boldsymbol{\Sigma}_0 = \alpha\mathbf{I}$, the log of the posterior is

$$\begin{aligned} \log p(\boldsymbol{\theta} \mid \mathbf{t}, \mathbf{X}) &= \frac{N}{2}\log\left(\frac{1}{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2}(\mathbf{t} - \boldsymbol{\theta}^T\boldsymbol{\Phi})^T(\mathbf{t} - \boldsymbol{\theta}^T\boldsymbol{\Phi}) - \frac{\alpha}{2}\boldsymbol{\theta}^T\boldsymbol{\theta} \\ &= \log L(\boldsymbol{\theta}) + P(\boldsymbol{\theta}) \end{aligned} \tag{3.25}$$

where $P(\boldsymbol{\theta})$ is the same as equation 3.16 with $\alpha = \lambda/\sigma^2$. Therefore, the MAP solution for Bayesian regression with a Gaussian prior is equivalent to ridge regression. This shows how the prior acts to regularise our estimate of $\boldsymbol{\theta}$.

The aim of building a model is not just to estimate $\boldsymbol{\theta}$ but also to make predictions, t_*, at unseen data points $\mathbf{x}_*$. To do so, we need to calculate the predictive distribution

$$p(t_* \mid \mathbf{x}_*, \mathbf{X}, \mathbf{t}) = \int p(t_* \mid \mathbf{x}_*, \boldsymbol{\theta})p(\boldsymbol{\theta} \mid \mathbf{t}, \mathbf{X})\mathrm{d}\boldsymbol{\theta}. \tag{3.26}$$

Once again, the mathematical properties of the Gaussian distribution allow this integral to be calculated analytically, giving

$$p(t_* \mid \mathbf{x}_*, \mathbf{X}_*, \mathbf{t}) = \mathcal{N}(t_* \mid \boldsymbol{\theta}_\mathrm{N}^T\phi(\mathbf{x}_*), \sigma^2 + \phi(\mathbf{x}_*)^T\Sigma_\mathrm{N}\phi(\mathbf{x}_*)). \tag{3.27}$$

An example of Bayesian regression is plotted in figure 3.6. The mean function $\mathbb{E}[t] = \boldsymbol{\theta}_\mathrm{N}^T\phi(\mathbf{x}_*)$ is shown as the red line. The shaded blue area shows one standard deviation from the mean function. There is a contribution from both the intrinsic noise in the data set and the uncertainty in the model parameters, which grows rapidly outside the domain of the data set.

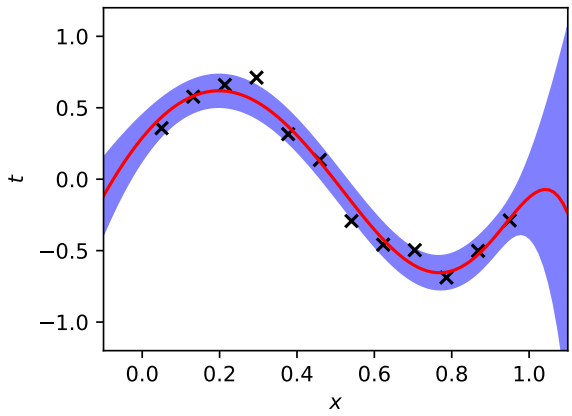


Figure 3.6: A Bayesian regression fitted to synthetic data. The data was generated by the function $t(x) = 0.3 + 0.7x + \mathcal{N}(0, 0.1)$. The red line shows the mean function and the blue shaded area shows the standard deviation from equation 3.27.

Bayesian Model Selection

The problem with cross-validation when used for model selection is that it requires partitioning of the data set. Bayesian model selection offers a robust alternative where none of the data is wasted. To explain Bayesian model selection we will use a hierarchical description, where Bayes' rule is successively applied [79]. At the first level of the hierarchy are the model parameters, $\boldsymbol{\theta}$, at the second level are the hyperparameters, $\boldsymbol{\lambda}$, and at the third and final level are the discrete set of model structures, $\{\mathcal{H}_i\}$. In Bayesian model selection, each of $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$ and $\mathcal{H}$ are treated as random variables.

We can write the posterior over $\boldsymbol{\theta}$ for a given model structure and set of hyperparameters as

$$p(\boldsymbol{\theta} \mid \mathbf{t}, \mathbf{X}, \boldsymbol{\lambda}, \mathcal{H}_i) = \frac{p(\mathbf{t} \mid \mathbf{X}, \boldsymbol{\theta}, \mathcal{H}_i) p(\boldsymbol{\theta} \mid \boldsymbol{\lambda}, \mathcal{H}_i)}{p(\mathbf{t} \mid \mathbf{X}, \boldsymbol{\lambda}, \mathcal{H}_i)} \tag{3.28}$$

where the denominator is

$$p(\mathbf{t} \mid \mathbf{X}, \boldsymbol{\lambda}, \mathcal{H}_i) = \int p(\mathbf{t} \mid \mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \mathcal{H}_i) p(\boldsymbol{\theta} \mid \boldsymbol{\lambda}, \mathcal{H}_i) \mathrm{d}\boldsymbol{\theta}. \tag{3.29}$$

Taking this as the likelihood at the second level and introducing a prior over the hyperparameters, $p(\boldsymbol{\lambda} \mid \mathcal{H}_i)$, the posterior over $\boldsymbol{\lambda}$ is given by

$$p(\boldsymbol{\lambda} \mid \mathbf{t}, \mathbf{X}, \mathcal{H}_i) = \frac{p(\mathbf{t} \mid \mathbf{X}, \boldsymbol{\lambda}, \mathcal{H}_i) p(\boldsymbol{\lambda} \mid \mathcal{H}_i)}{p(\mathbf{t} \mid \mathbf{X}, \mathcal{H}_i)} \tag{3.30}$$

where the denominator is

$$p(\mathbf{t} \mid \mathbf{X}, \mathcal{H}_i) = \int p(\mathbf{t} \mid \mathbf{X}, \mathcal{H}_i) p(\boldsymbol{\lambda} \mid \mathcal{H}_i) \mathrm{d}\boldsymbol{\lambda}. \tag{3.31}$$

We can then repeat this processes for the model structures by introducing a prior $p(\mathcal{H}_i)$

to give

$$p(\mathcal{H}_i \mid \mathbf{t}, \mathbf{X}) = \frac{p(\mathbf{t} \mid \mathbf{X}, \mathcal{H}_i)p(\mathcal{H}_\mathbf{i})}{p(\mathbf{t} \mid \mathbf{X})} \tag{3.32}$$

where the denominator is

$$p(\mathbf{t} \mid \mathbf{X}) = \sum_i p(\mathbf{t} \mid \mathbf{X}, \mathcal{H}_i)p(\mathcal{H}_\mathbf{i}). \tag{3.33}$$

We can combine equations 3.32, 3.30 and 3.28 to produce the predictive distribution

$$p(t_*|\mathbf{x}_*, \mathbf{X}) = \sum_i \int\int p(t_*|\mathbf{x}_*, \mathcal{H}_i)\, p(\boldsymbol{\theta}|\mathbf{t}, \mathbf{X}, \lambda, \mathcal{H}_i)\, p(\lambda|\mathbf{t}, \mathbf{X}, \mathcal{H}_i)\, p(\mathcal{H}_i|\mathbf{t}, \mathbf{X}) \mathrm{d}\boldsymbol{\theta} \mathrm{d}\lambda. \tag{3.34}$$

Simply put, Bayesian model selection considers all possible models and gives a prediction that is the average of each result weighted by their likelihood.

If $\boldsymbol{\theta}$, λ and $\mathcal{H}$ are high dimensional, Bayesian model selection involves many integrals which many not be solvable analytically. A numerical solution is then required, such as MCMC, which is computationally expensive. Therefore, a common approach is to only perform a fully Bayesian analysis at the first level and use maximum likelihood to obtain the optimum values for the hyperparameters and model structures. This is know as type 2 maximum likelihood [140] and is commonly used for training Gaussian process regression models.

3.2.4 Gaussian Process Regression

Linear regression assumes $f(\mathbf{x})$ is a parametric function, with the form decided before inference. However, as we discussed in section 3.2.2, this can be impractical, as to generate a good predictor it requires some prior knowledge of the data set features. Neural network regression bypasses this issue by allowing the basis functions to change during the training process. However, this introduces more model parameters which leads to over fitting for small data sets. Kernel methods offer an alternative solution to the fixed basis function issue. Here we will discuss Gaussian process regression, which is an example of a Bayesian kernel method.

We can arrive at the regression formula for Gaussian process regression by either assuming a distribution over the model parameters or a distribution over the fitting function [79]. We will start with the model parameters distribution view which follows directly on from linear Bayesian regression. Ignoring the intrinsic noise in the data, we can rewrite

the mean function and variance of equation 3.27 as

$$\begin{aligned}\mathbb{E}[f_*] &= \boldsymbol{\theta}_{\mathrm{N}}^T \phi(\mathbf{x}_*) \\ &= \phi(\mathbf{x}_*)^T \boldsymbol{\Sigma}_0 \boldsymbol{\Phi} (\boldsymbol{\Phi}^T \boldsymbol{\Sigma}_0 \boldsymbol{\Phi} + \sigma^2 \mathbf{I})^{-1} \mathbf{t} \\ &= \mathbf{k}(\mathbf{x}_*)(\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{t}\end{aligned} \tag{3.35}$$

and

$$\begin{aligned}\mathbb{V}[f_*] &= \phi(\mathbf{x}_*)^T \boldsymbol{\Sigma}_{\mathrm{N}} \phi(\mathbf{x}_*) \\ &= \phi(\mathbf{x}_*)^T \boldsymbol{\Sigma}_0 \phi(\mathbf{x}_*) - \phi(\mathbf{x}_*)^T \boldsymbol{\Sigma}_0 \boldsymbol{\Phi} (\boldsymbol{\Phi}^T \boldsymbol{\Sigma}_0 \boldsymbol{\Phi} + \sigma^2 \mathbf{I})^{-1} \boldsymbol{\Phi}^T \boldsymbol{\Sigma}_0 \phi(\mathbf{x}_*) \\ &= k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}(\mathbf{x}_*)^T (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{k}(\mathbf{x}_*)\end{aligned} \tag{3.36}$$

respectively, where we have defined $k(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \boldsymbol{\Sigma}_0 \phi(\mathbf{x}')$, $\mathbf{k}(\mathbf{x}) = \phi(\mathbf{x})^T \boldsymbol{\Sigma}_0 \boldsymbol{\Phi}$ and $\mathbf{K} = \boldsymbol{\Phi}^T \boldsymbol{\Sigma}_0 \boldsymbol{\Phi}$. The function $k(\mathbf{x}, \mathbf{x}')$ is known as a kernel or covariance function. This leads to an alternative view of regression, where instead of defining a set of basis functions which in turn defines a kernel, we can directly define the kernel. This is known as the kernel trick [76] and is useful when the basis functions map to a high dimensional space, making their direct calculation expensive. A commonly used kernel is the radial basis function (RBF)

$$k_{\mathrm{RBF}}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\Big(- \frac{(\mathbf{x} - \mathbf{x}')^2}{2\alpha^2} \Big) \tag{3.37}$$

where α and σ are hyperparameters. The Taylor expansion of the RBF kernel is infinite. This implies the corresponding basis function is an infinite dimensional vector. However, because it is never directly calculated, equations 3.35 and 3.36 are still tractable. For an extensive list of kernel functions and their properties, see Rasmussen and Williams [79].

An alternative approach to deriving equations 3.35 and 3.36 is to define a distribution directly over functions $f(\mathbf{x})$ rather than the model parameters $\boldsymbol{\theta}$. To do so requires the use of stochastic processes. A stochastic process is a generalisation of a probability distribution to functions. In the same way that a joint probability distribution can be defined over a set of random variables, indexed by an integer i, a stochastic process is defined over a set of random variables, with a continuous variable index t. Historically, the index t was often time, leading to the name stochastic "process" [141]. In our case the function $f(\mathbf{x})$ is a set of random variables indexed by the continuous variable $\mathbf{x}$. If any finite set $\{f(\mathbf{x}_i)\}$ has a joint Gaussian distribution, this is known as a Gaussian process. A Gaussian process over $f(\mathbf{x})$ is written as

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{3.38}$$

where the mean function and covariance function are given by:

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}[(f(\mathbf{x} - m(\mathbf{x}))(f(\mathbf{x}' - m(\mathbf{x}')]. \end{aligned} \tag{3.39}$$

To demonstrate the function-space view of Gaussian process regression, we will define a Gaussian process prior with zero mean and an RBF kernel, $f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, k_{\text{RBF}}(\mathbf{x}, \mathbf{x}'))$ Figure 3.7 (a) shows 50 sample functions drawn from the prior. We now wish to update the prior using the training data set. The joint probability of the training outputs, $\mathbf{t}$, observed at $\mathbf{X}$ and noise free predictions $\mathbf{f}_*$ observed at $\mathbf{x}_*$ is given by

$$p\begin{pmatrix} \mathbf{t} \\ \mathbf{f}_* \end{pmatrix} = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I} & K(\mathbf{X}, \mathbf{X}_*) \\ K(\mathbf{X}_*, \mathbf{X}) & K(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix}\right). \tag{3.40}$$

Using the property of conditioning on a joint Gaussian distribution (see Bishop page 85 [76]) we can calculate posterior distribution $p(\mathbf{f}_* \mid \mathbf{X}, \mathbf{t}, \mathbf{X}_*)$ from equation 3.40. This is equivalent to applying Bayes' theorem. This gives a Gaussian process with mean and variance

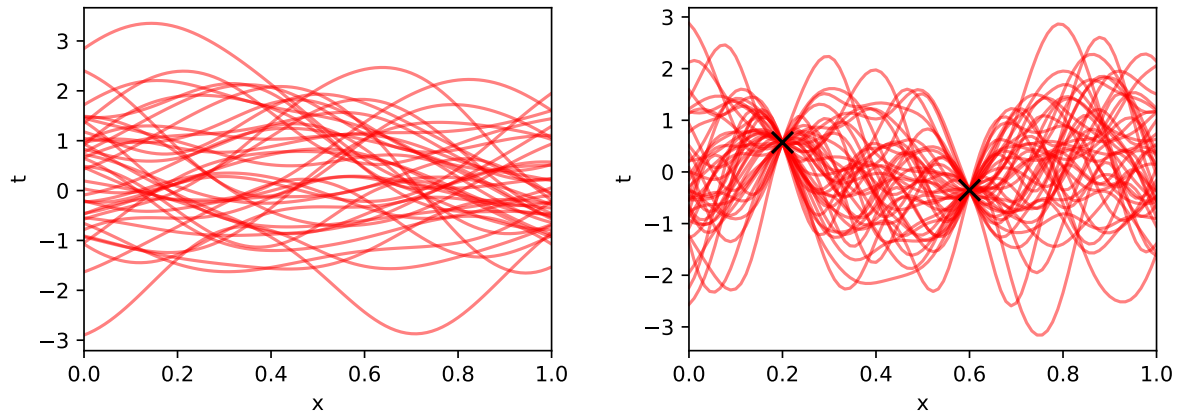
$$\begin{aligned} \mathbb{E}[\mathbf{f}_*] &= K(\mathbf{X}_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I}]^{-1}\mathbf{t} \\ \mathbb{V}[\mathbf{f}_*] &= K(\mathbf{X}_*, \mathbf{X}_*) - K(\mathbf{X}_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma^2\mathbf{I}]^{-1}K(\mathbf{X}, \mathbf{X}_*) \end{aligned} \tag{3.41}$$

respectively. This is the vector equivalent of equations 3.35 and 3.36. Figure 3.7 (b) shows 50 sample functions drawn from the posterior distribution with two data points. We can think of the training process as removing all the functions from the prior which do not pass through the data points.

By defining the regression through a kernel rather than basis functions we have removed the model parameters. However, the kernel includes hyperparameters which must be learned from the data. As Gaussian process regression is a Bayesian approach, we should use Bayesian model selection to set the hyperparameters. However, it is usually too computationally expensive to perform a full Bayesian treatment and a type 2 maximum likelihood approach is used instead. The log marginal likelihood is given by

$$\begin{aligned} \log p(\mathbf{t} \mid \mathbf{X}) &= \log \int p(\mathbf{t} \mid \mathbf{f}, \mathbf{X}) p(\mathbf{f} \mid \mathbf{X}) \mathrm{d}\mathbf{f} \\ &= -\frac{1}{2}\mathbf{t}^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{t} - \frac{1}{2}\log|\mathbf{K} + \sigma^2\mathbf{I}| - \frac{n}{2}\log 2\pi \end{aligned} \tag{3.42}$$

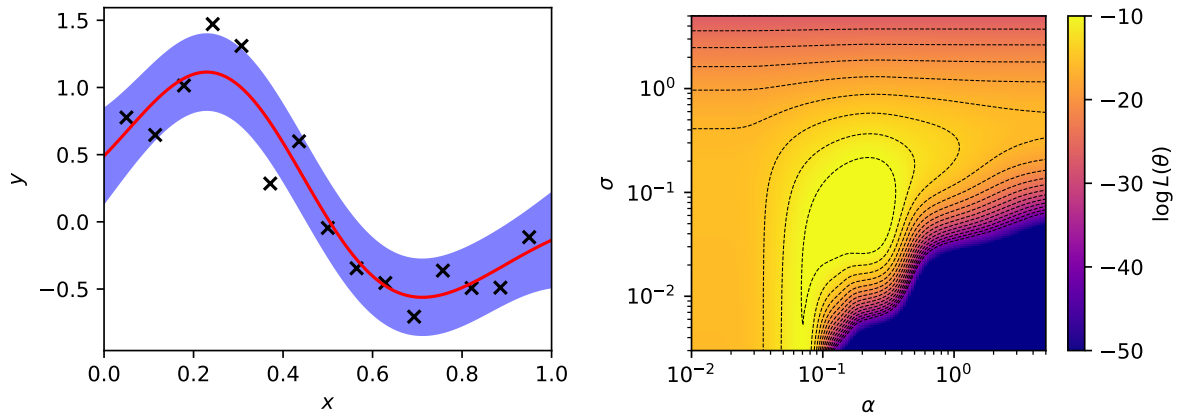
and is maximised w.r.t the kernel hyperparameters. Figure 3.8 (a) shows an example of a Gaussian process regression fitted to a synthetic data set using the RBF kernel. Also shown is a surface plot of $\log p(\mathbf{t} \mid \mathbf{X})$ with varying α and σ. This has a single stationary point giving the optimum values for α and σ. If the log marginal likelihood space was



(a) Gaussian process prior

(b) Gaussian process posterior

Figure 3.7: Gaussian process samples. (a) shows 50 samples drawn from the prior Gaussian process with zero mean and an RBF kernel covariance. (b) shows 50 samples drawn from the posterior Gaussian process.



(a) Gaussian process regression

(b) Log marginal likelihood

Figure 3.8: Gaussian process regression. (a) shows a Gaussian process regression fit to synthetic data. (b) shows the log marginal as a function of α and σ.

multi-modal, the type 2 maximum likelihood approach would not give a good predictor. In this case a full Bayesian treatment is required.

3.3 Monte Carlo Methods

Monte Carlo methods cover a wide range of computational algorithms and are ubiquitous across all areas of physics. The common theme of all Monte Carlo algorithms is the repeated random sampling to obtain a solution. In this work Monte Carlo methods are used both to simulate probabilistic high energy physics systems and solve high dimensional integrals encountered in Bayesian analysis. In both these cases the Monte Carlo algorithm

is used to find the expectation of a function, $f(\mathbf{x})$, given by

$$\mathbb{E}[f] = \int f(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \tag{3.43}$$

where $\mathbf{x}$ is a random variable with an associated probability distribution $p(\mathbf{x})$. If $\mathbf{x}$ has a large number of dimensions, this integral is computationally expensive. However, we can make a Monte Carlo estimate of equation 3.43 given by

$$\mathbb{E}[f] \approx \hat{f} = \frac{1}{N}\sum_{i}^{N} f(\mathbf{x_i}) \tag{3.44}$$

where $\mathbf{x}_i$ is a set of independent random samples of $p(\mathbf{x})$ [142]. The variance on the estimator is

$$\mathbb{V}[\hat{f}] = \frac{\sigma_\mathrm{f}^2}{N} \tag{3.45}$$

where σ_f^2 is the variance of $f(\mathbf{x})$. The error on the estimator decreases as $\frac{1}{\sqrt{N}}$ and is independent of the dimensions of $\mathbf{x}$. Therefore, calculating the Monte Carlo estimate is more efficient than standard numerical integral methods for large dimensional integrals. However, there are caveats to using a Monte Carlo estimate which may reduce this efficiency. Although $\mathbb{V}[\hat{f}]$ is independent of the dimensions of $\mathbf{x}$, generating independent samples of $p(\mathbf{x})$ may not be. Even if the samples can be efficiently generated, $f(\mathbf{x})$ may be small in the region $p(\mathbf{x})$ is large. The computational complexity is therefore shifted to obtaining samples from $p(\mathbf{x})$ in this region. In the remainder of this section we will discuss the sampling and variance reduction techniques used in this work.

3.3.1 Sampling methods

The success of the Monte Carlo method relies on the efficient generation of samples. All the sampling methods discussed here first require a method to sample from the uniform distribution $u \sim U(0,1)$. In this work the Mersenne Twister generator has been used for this purpose [143]. This produces high quality samples with a period of $2^{219937} - 1$. With uniform samples we can apply a more advanced method to sample from a generic distribution. The three methods we will discuss in this section are transform sampling, rejection sampling, and Markov chain Monte Carlo (MCMC) sampling.

Transform sampling

The goal of transform sampling is to find a function which transforms our uniform samples, u, into samples from $p(x)$ (*i.e.* find the function $x = h(u)$). To do this the change of variables formula is used

$$p(x) = p(u)\left|\frac{\mathrm{d}u}{\mathrm{d}x}\right| \tag{3.46}$$

where $p(u) = 1$, and is integrated to give

$$u = F(x) = \int_{-\infty}^{x} p(x')\mathrm{d}x' \tag{3.47}$$

where $F(x)$ is the cumulative distribution function (CDF) of $p(x)$. Samples from $p(x)$ are then given by

$$x = F^{-1}(u). \tag{3.48}$$

This method is relatively simple, however, it is only useful if $F^{-1}(u)$ can be obtained efficiently [76].

Rejection Sampling

Rejection Sampling is a more general sampling method that can still be applied when transform sampling is difficult [144]. It makes use of a proposal distribution, $q(x)$, from which samples can be easily obtained. The first step is to find the scaling constant, given by

$$M = \sup\left\{\frac{p(\mathbf{x})}{q(\mathbf{x})}\right\} \tag{3.49}$$

such that the function $M\,q(\mathbf{x})$ complete encloses $p(\mathbf{x})$. A proposal sample, $\mathbf{x}_0$, is drawn from $q(\mathbf{x})$ and a uniform sample, u_0, is drawn from $U(0,\, M\,q(\mathbf{x}_0))$. The sample $\mathbf{x}_0$ is accepted if $u_0 < p(\mathbf{x}_0)$ otherwise it is rejected and new samples $\mathbf{x}_1$ and u_1 are taken and tested. This continues until a sample is accepted.

The fraction of samples accepted is given by the ratio of the area under $p(\mathbf{x})$ to the area under $Mq(\mathbf{x})$. In general this ratio decreases exponentially with the dimensionality of $\mathbf{x}$. Therefore, using rejection sampling to estimate $\mathbb{E}[f]$ with equation 3.44 offers no benefit over grid based integration methods which also scale exponentially.

Markov Chain Monte Carlo

As we have seen with rejection sampling, it gets exponentially more expensive to generate independent samples as the number of dimensions increases. Markov chain Monte Carlo (MCMC) methods remove this problem by generating non-independent samples. A Markov chain is a stochastic process, describing a sequence of random variables, in which the probability of observing $\mathbf{x}_{i+1}$ only depends on $\mathbf{x}_i$

$$p(\mathbf{x}_{i+1},\ |\ \mathbf{x}_1, \ldots \mathbf{x}_i) = p(\mathbf{x}_{i+1}\ |\ \mathbf{x}_i). \tag{3.50}$$

The chain is fully defined by the probability distribution over the initial variable, $p(\mathbf{x}_0)$ and the transition probabilities between states $T(\mathbf{x}_i, \mathbf{x}_{i+1}) = p(\mathbf{x}_{i+1}\ |\ \mathbf{x}_i)$. The chain is

said to be stationary if $p(\mathbf{x}_{i+1}) = p(\mathbf{x}_i)$ where

$$p(\mathbf{x}_{i+1}) = \sum_{\mathbf{x}_i} T(\mathbf{x}_i, \mathbf{x}_{i+1}) p(\mathbf{x}_i). \tag{3.51}$$

The goal of this sampling method is to define a Markov chain with a stationary distribution given by the distribution which we wish to sample from.

The first algorithm for sampling from a Markov chain was proposed by Metropolis in 1949 [145]. Like rejection sampling, this uses a proposal distribution, $q(\mathbf{x} \,|\, \mathbf{x}_*)$, which can already be sampled from. The proposal distribution gives the probability of obtaining a new state $\mathbf{x}$ and only depends on the previous state $\mathbf{x}_*$. Given some initial state, $\mathbf{x}_0$, the proposal distribution is sampled to give a candidate state, $\mathbf{x}_*$. This state is then accepted with probability

$$A(\mathbf{x}_*, \mathbf{x}_i) = \min\left(1, \frac{p(\mathbf{x}_*)}{p(\mathbf{x}_i)}\right). \tag{3.52}$$

If the state is accepted, we set $\mathbf{x}_{i+1} = \mathbf{x}_*$ otherwise we reject the new state and set $\mathbf{x}_{i+1} = \mathbf{x}_i$.

Unlike rejection sampling, the efficiency of MCMC does not decrease exponentially with the number of dimensions [76]. This makes it a practical method for solving the high dimensional integrals encountered in Bayesian inference. However, there are drawbacks to MCMC methods including the initial samples being drawn from the wrong distribution and the generation of correlated samples. Both these issues increase the Monte Carlo error. The initial sampling problem can be solved by throwing away the start of the chain, a process known as burn-in. Uncorrelated samples can be obtained with an increased computational cost by taking samples which are far apart in the chain and discarding the rest. Extensions to the base Metropolis algorithm have been developed which attempt to reduce the correlation in the chain. These include the Metropolis-Hastings algorithm [146], Gibbs sampling [147] and Hamiltonian Monte Carlo sampling [148]. However, all methods are inefficient if a small number of samples are required.

3.3.2 Variance Reduction

In this work, we are interested in modelling interactions with low probabilities. For example, in a nonlinear Breit-Wheeler experiment, discussed in chapter 5, the probability of a seed electron producing an electron positron pair can be lower than 10^{-8}. If we model N electrons, the number of pairs produced follows a binomial distribution where the probability of each interaction is p. If we want to estimate p, using the crude method given by equation 3.44 the relative error is given by

$$\text{error} = \frac{\sqrt{\mathbb{V}[\hat{p}]}}{\mathbb{E}[\hat{p}]} = \sqrt{\frac{1-p}{Np}} \tag{3.53}$$

where we have used the variance of a binomial distribution as $Np(1-p)$. To get a relative error of 1% would require 10^{12} samples. Variance reduction techniques attempt to reduce the number of samples required for a given error. In this work we have used two types of variance reduction, cross-section biasing and importance sampling.

Cross-Section Biasing

Cross-section biasing artificially increases the probability of a process occurring by multiplying the cross-section by some factor b. In Geant4, we are often modelling discrete processes occurring in volumes with constant material properties. Therefore, the probability of the process occurring over some length l is

$$\begin{aligned} p_{\text{bias}}(x \leq l) &= 1 - e^{-b\,\sigma\rho\,l} \\ &\approx b\,p(x \leq l) \end{aligned} \tag{3.54}$$

where $p(x \leq l)$ is the unbiased probability and the first order Taylor expansion requires $b\,\sigma\rho\,l \ll 1$. Thus, the number samples required to achieve a given error is reduced by factor of b. However, if $b\,\sigma\rho\,l \not\ll 1$ there is no analytical solution linking $p_{\text{bias}}(x \leq l)$ and $p(x \leq l)$. This limits the level of cross-section biasing which can be used.

Importance Sampling

Importance Sampling is a common extension to the crude method of approximating integrals given by equation 3.44 [149]. It is useful for both variance reduction and when $p(\mathbf{x})$ is difficult to sample from. Starting from equation 3.43 we introduce a proposal distribution, $q(\mathbf{x})$ to give

$$\begin{aligned} \int f(\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} &= \int f(\mathbf{x})\frac{p(\mathbf{x})}{q(\mathbf{x})}q(\mathbf{x})\mathrm{d}\mathbf{x} \\ &\approx \frac{1}{N}\sum_{i=1}^{N} f(\mathbf{x}_i)\,w_i \end{aligned} \tag{3.55}$$

where $w_i = \frac{p(\mathbf{x}_i)}{q(\mathbf{x}_i)}$ and the samples are taken from $q(\mathbf{x})$ rather than $p(\mathbf{x})$. The factors w_i are known as the importance weights and correct for the fact we have sampled from the wrong distribution. Therefore, by carefully selecting $q(\mathbf{x})$ we can obtain more samples in areas where $f(\mathbf{x})$ is large, reducing the error in the estimate.

Chapter 4

Modelling Particle-Photon Processes in Geant4

In classical electrodynamics, when two waves of light meet, they simply pass through each other unaffected. The lack of interaction gives rise to a linear theory, evident from Maxwell's equations. However, the prediction of the Breit-Wheeler process in 1934 [18] suggests photons can interact. This implies the theory of quantum electrodynamics must be nonlinear. This has some profound consequences, such as the concept of superposition, at the heart of classical electrodynamics, no longer holding true [150]. In quantum electrodynamics, the two lowest order photon-photon interactions (expanding in α), are the linear Breit-Wheeler process and photon-photon scattering [151, 152]. Both these processes are described as binary interactions as they involve two incoming particles colliding to produce two outgoing particles. In the case of the Breit-Wheeler process, the incoming particles are two photons and the outgoing particles are an electron-positron pair ($\gamma\gamma \to e^+e^-$). For photon-photon scattering all the incoming and outgoing particles are photons ($\gamma\gamma \to \gamma\gamma$). Understanding these processes is not only vital from a theoretical point of view, but it is also of astrophysical importance. They are predicted to play an important role in a range of phenomena throughout the universe [150, 153, 154, 42, 155]. However, from the collision of two real photons, neither photon-photon scattering nor the linear Breit-Wheeler process have been observed in the laboratory.

As shown in figure 4.1, the peak of the Breit-Wheeler cross-section is of the same order of magnitude as Dirac annihilation and Compton scattering, both of which were observed over 80 years ago [97, 156]. However, unlike Dirac annihilation and Compton scattering, the Breit-Wheeler process has a strict centre-of-mass (CM) energy threshold. From the Mandelstam variables (see equation 2.80), the invariant CM energy squared for a two particle system is

$$s = m_1^2 + m_2^2 + 2(E_1E_2 - \overrightarrow{p_1} \cdot \overrightarrow{p_2}). \tag{4.1}$$

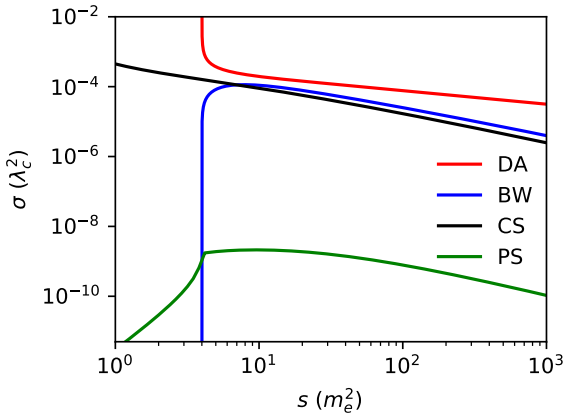


Figure 4.1: Comparison of total cross-section for Dirac annihilation (DA, red), the Breit-Wheeler process (BW, blue), Compton scattering (CS, black) and photon-photon scatter (PS, green).

Here, we have set $c = 1$, which will be used throughout this chapter, along with $\hbar = m_e = 1$. Creating an electron-positron pair at rest requires $s = 4$, which sets this CM threshold. To date, no experiment has interacted two photon sources which are of sufficient energy and density to produces a detectable number of Breit-Wheeler pairs.

In contrast to the Breit-Wheeler process, photon-photon scattering produces two massless particles and exhibits no threshold behaviour. However, the cross-section does decay rapidly below $s = 4$ as $\sigma \propto s^3$. Photon-photon scattering is also a higher order process than the Breit-Wheeler process, and the cross-section is a factor of $\alpha^2 = 5.32 \times 10^{-5}$ smaller. For these reasons, to detect photon-photon scattering we again need high energy and density photon sources, which have not been created in the laboratory.

Despite the difficulties, over the past few years a number of experimental schemes have been proposed to detect the Breit-Wheeler process [157, 158, 159, 160]. All these experiments involve the generation of high energy gamma ray sources to overcome the CM threshold. However, this leads to an experimental environment that is inherently noisy. There are two large sources of noise: the gamma rays interacting directly with single particle detectors; and photons interacting with the experimental setup, producing background electron-positron pairs through the Bethe-Heitler process. For a successful experiment, the ratio of signal Breit-Wheeler pairs to background noise should be maximised. An estimation of the background noise can be obtained by modelling the passage of the high energy particles through the experimental setup. A number of publicly available Monte Carlo particle tracking codes exist for this purpose, including Geant4, Fluka and MCNP [67, 161, 162]. In this work, Geant4 has been used to analyse the background noise of a Breit-Wheeler detection experiment that we will discuss in detail in chapter 6. It would be useful to model both the signal and noise within a single framework, making for easier experimental optimisation. However, the Breit-Wheeler process is not included

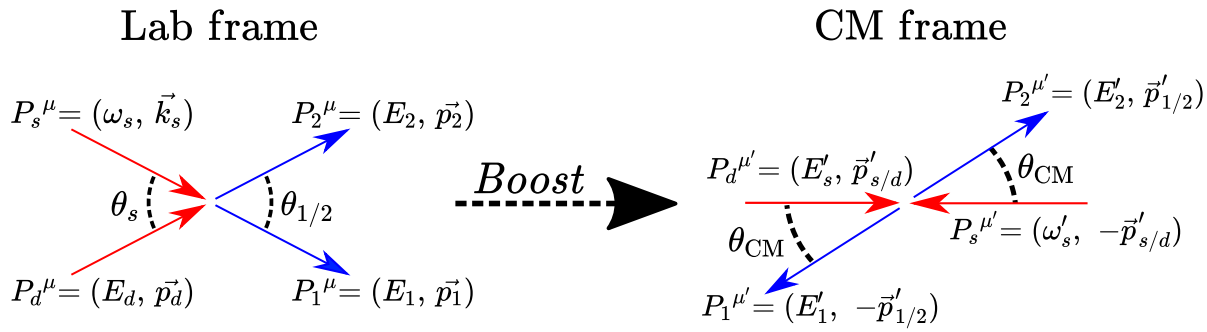


Figure 4.2: Diagram of binary collision in both the laboratory frame and the centre of mass frame.

within the standard Geant4 physics package. Therefore, we have developed a new binary particle-photon interaction package for Geant4, which includes the Breit-Wheeler process.

Geant4 is a Monte Carlo platform for modelling the passage of high energy particles through matter. To include the particle-photon processes of interest here, they must be modelled within this same framework. This has been achieved by treating one of photon sources as a static photon field. High energy particles are sampled and tracked through this field, and this process is referred to as an event. In this work, we will refer to photons from the field as *static* and particles tracked through the field as *dynamic*. Using this method, the temporal evolution of the static photon field is not accounted for. Therefore, this package is designed for modelling experiments with asymmetric sources, in which one is constant in time over the duration of the interaction, to a good approximation.

In this chapter we will start by discussing the development of this physics package, which includes the Breit-Wheeler process, photon-photon scattering and Compton scattering. We will then show how the efficiency of the interaction algorithm can be greatly increased by implementing a Gaussian process regression. Finally, we will demonstrate the capabilities of the package, by studying the prospect of detecting the Breit-Wheeler process and photon-photon scattering in a thermal radiation field.

4.1 Package Development

The base interaction algorithm, which we will discuss here, is the same for all three particle-photon processes. This is because each process consists of a binary collision, which can be represented by the diagram shown in figure 4.2. This diagram, which shows the collision in both the laboratory and CM frames, highlights all the important parameters which are contained within the four momenta of the particles. Here, the subscripts d and s refer to the dynamic particle and static photon respectively, and the subscripts 1 and 2 refer to the two outgoing particles. The primed indices refer to properties in the CM frame.

The interaction algorithm is carried out if a dynamic particle enters a static radiation volume that has been defined within the computational domain. The algorithm has two parts to it. The first part involves calculating the probability that the dynamic particle interacts in the field. If the interaction occurs, the second part of the algorithm is carried out. Here, the properties of the two outgoing particles are calculated. In the following sections, we will review both parts of the algorithm.

4.1.1 Interaction Probability

When a dynamic particle enters a static photon field it can do so at any arbitrary angle in the simulation frame. To simplify the calculation we rotate the frame such that the z-axis is the direction of propagation of the dynamic particle. We will refer to the original and rotated frames as the simulation and dynamic particle frames respectively. The field is fully defined by a spectral photon density per unit solid angle, $n(\omega_s, \theta_s, \phi_s)$. Here, ω_s is the static photon energy and θ_s, ϕ_s are the static photon polar and azimuthal angles defined in the dynamic particle frame. The cross-section, which is proportional to the probability of the interaction occurring, is a function of s. We can use equation 4.1 to calculate s, which, depending on whether the dynamic particle is massive, reduces to

$$s = \begin{cases} 2E_d\omega_s(1-\cos\theta_s) & m_d = 0 \\ m_d^2 + 2E_d\omega_s\left(1-\sqrt{1-\frac{m_d^2}{E_d^2}}\cos\theta_s\right) & m_d \neq 0 \end{cases} \tag{4.2}$$

where E_d is the energy of the dynamic particle and m_d is its mass. For the rest of this chapter, we will only consider massive particles that are electrons or positrons so $m_d = 1$.

The interaction between the dynamic particle beam and the static photon field occurs at a constant average rate making it a Poisson process. Therefore, the probability that the dynamic particle will travel a length, x, is given by the exponential distribution with the following PDF

$$P(x) = \lambda_d^{-1} e^{-x/\lambda_d} \tag{4.3}$$

where λ_d is the mean free path (MFP) of the dynamic particle in the static field. When travelling through a beam of massless particles, the MFP is governed by [163]

$$\frac{1}{\lambda_d} = \int_0^{2\pi} \mathrm{d}\phi_s \int_0^{\pi} \mathrm{d}\theta_s \int_0^{\infty} \mathrm{d}\omega_s \; \Theta(s - s_{\min})\, \sigma(s)\, n(\omega_s, \phi_s, \theta_s)\, (1-\cos\theta_s) \tag{4.4}$$

where $(1-\cos\theta_s)$ accounts for the relative velocity between the particles and $\Theta(s-s_{\min})$ is the Heaviside step function which sets the CM energy threshold. For the case of photon-photon scattering there is no threshold, however, by defining $s_{\min}$ we can speed up the calculation by ignoring highly unlikely interactions. This three dimensional integral is solved by quadrature using Simpson's rule.

After solving equation 4.4, the distance which the dynamic particle travels before interacting is obtained by sampling from the PDF in equation 4.3. This is performed using inverse transform sampling (see section 3.3.1) as the CDF can be inverted analytically. Therefore, the propagation length is given by

$$x = \frac{-\log(1-u)}{\lambda_d} \tag{4.5}$$

where $u \sim \mathcal{U}(0,1)$. If x is longer than the static photon field, the dynamic particle will propagate through unaffected. However, if x is shorter, the dynamic particle will propagate to x, then interact.

4.1.2 Interaction Dynamics

At the interaction point, the dynamic particle is removed and replaced with two new particles, the type of which depends on which process occurred (an electron-positron pair for the Breit-Wheeler process, two photons for photon-photon scattering and a photon and electron or positron for Compton scattering). However, before adding in the new particles, their four momentum in the simulation frame, $P^{\mu}_{1/2}$, must be known. This can easily be found in the CM frame, $P^{\mu'}_{1/2}$ due to conservation laws. However, the CM frame is not yet defined as we do not known which static photon the dynamic particle interacted with. The properties of this photon (ω_s, θ_s and ϕ_s) are obtained by sampling from the integrand of equation 4.4. For an arbitrary static photon field, this is a complex function of three dimensions. Inverse transform sampling would be inefficient, therefore, rejection sampling is used. With both incoming particles defined, equation 4.2 is used to calculate s.

$P^{\mu'}_{1/2}$ has three free parameters, $E'_{1/2}$, $\theta'_{1/2}$ and $\phi'_{1/2}$. We can calculate $E'_{1/2}$ from equation 4.2, which simplifies in the CM frame to

$$E'_{1/2} = \frac{(s + m^2_{1/2} - m^2_{1/2})}{2\sqrt{s}}. \tag{4.6}$$

For both the Breit-Wheeler process and photon-photon scattering, the outgoing particles have the same mass, therefore, they receive the same energy of $E'_{1/2} = \sqrt{s}/2$. For Compton scattering, however, the masses are not equal and the electron/positron receives $E'_1 = (s+1)/2\sqrt{s}$ while the photon receives $E'_2 = (s-1)/2\sqrt{s}$.

Due to conservation of momentum, the outgoing particles scatter along the same axis. The angle between the scatter axis and the interaction axis is given by θ_{CM} and is shown in figure 4.2. θ_{CM} is distributed according to the differential cross-section. We therefore set $\theta'_1 = \theta_{\mathrm{CM}}$ and again use inverse transform sampling to obtain θ_{CM}. The CDF for the

differential cross-section is

$$C(\theta_{\mathrm{CM}}) = \frac{\int_0^{\theta_{\mathrm{CM}}} (\mathrm{d}\sigma/\mathrm{d}\Omega) \sin\theta_{\mathrm{CM}} \, \mathrm{d}\theta_{\mathrm{CM}}}{\int_0^{\pi} (\mathrm{d}\sigma/\mathrm{d}\Omega) \sin\theta_{\mathrm{CM}} \, \mathrm{d}\theta_{\mathrm{CM}}} \tag{4.7}$$

which cannot be inverted analytically. Therefore, equation 4.7 is inverted numerically and tabulated over a range of values of s. θ_{CM} is then interpolated from the table as required. Likewise, we set $\phi_1' = \phi_{\mathrm{CM}}$ which is sampled from the uniform distribution, $\phi_{\mathrm{CM}} \sim \mathcal{U}(0, 2\pi)$. Using conservation of momentum, the scattering angles for particle 2 are then easily obtained through $\theta_2 = \pi - \theta_1$ and $\phi_2 = \phi_1 - \pi$.

Having obtained $P_{1/2}^{\mu'}$ in the CM frame, we now need to transform it into the simulation frame. This involves applying the inverse of the transform which takes the incoming particles to the CM frame. Starting from the dynamic particle frame, a trivial rotation of ϕ_s around the z-axis is performed such that the y-component of the static photon momentum is zero. After doing so, the total four momentum of the system is

$$P_T^{\mu} = P_d^{\mu} + K_i^{\mu} = \begin{pmatrix} E_d + \omega_s \\ \omega_s \sin\theta_s \\ 0 \\ p_d + \omega_s \cos\theta_s \end{pmatrix} \tag{4.8}$$

where p_d is the magnitude of the three momentum of the dynamic particle. If the dynamic particle is a photon, $p_d = E_d$, whereas for an electron or positron, $p_d = \sqrt{E_d^2 - 1}$. The CM frame is obtained by applying a Lorentz boost along a vector in the $x - z$ plane given by $\Gamma^{\mu}{}_{\nu}$. However, rather than boosting along an arbitrary axis, it simpler to break the operation into a rotation around the y-axis and a Lorentz boost along the z-axis

$$P_{\mathrm{CM}}^{\mu} = \Gamma^{\mu}{}_{\nu} \, P_T^{\nu} = R^{\mu}{}_{\nu} \, \Lambda^{\nu}{}_{\rho} \, P_T^{\rho}. \tag{4.9}$$

Here, $R^{\mu}{}_{\nu}$ is the rotation with an associated angle of

$$\tan\phi = \frac{\omega_s \sin\theta_s}{p_d + \omega_s \cos\theta_i} \tag{4.10}$$

and $\Lambda^{\nu}{}_{\rho}$ is the Lorentz boost with a velocity of

$$\beta = \frac{\sqrt{p_d^2 + \omega_s^2 + 2\, p_d\, \omega_s \cos\theta_s}}{E_d + \omega_s}. \tag{4.11}$$

Upon applying the inverse of these transforms to obtain $P_{1/2}^{\mu}$, the two new particles are added to the simulation.

Table 4.1: Static photon energy spectra implemented in the package.

Source Name	$f(\omega_s)$	Parameters and Description
Monoenergetic	$n'\delta(\omega_s - \omega')$	Monoenergetic source with energy ω' and density n'. Used for benchmarking and for approximating a laser.
Grey-body	$\frac{\omega_s^2}{\pi^2}\frac{\epsilon}{e^{\omega_s/T}-1}$	Thermal radiation source of temperature T and emissivity ϵ. Used to model a photon source in thermodynamic equilibrium, such as astrophysics systems or radiation from a hohlraum.
User-defined	Tabulated	Spectrum read from file. Used to model complex sources of radiation, possibly from an atomic simulation or experimentally measured.

4.1.3 Static Photon Field

By treating the photon source as a static field and fully defining it by $n(\omega_s, \theta_s, \phi_s)$, we are ignoring both spatial and temporal gradients. To account for spatial gradients, a static photon field can be built up from smaller sub fields, each with a different $n(\omega_s, \theta_s, \phi_s)$. However, due to the static nature of the Geant4 computational domain, it is not easy to account for temporal gradients. This reduces the types of experiments which can be modelled using this package. However, there are also benefits to using this method including a reduction in the computational complexity. Modelling binary interactions between two beams of particles where both are treated dynamically is an N-body problem with complexity that scales as $\mathcal{O}(N^2)$ where N is the number of particles. In the algorithm presented here, this is reduced to $\mathcal{O}(N)$, however, the integral in equation 4.4 must be carried out for each event and is expensive to compute. In section 4.3 we will discuss methods for speeding this up.

Further constraints are also placed on the static field for computational savings. Defining $n(\epsilon_i, \theta_i, \phi_i)$ requires the allocation of a three dimensional array which can be memory intensive if many sub-fields are used. To reduce this, we assume the static photon's energy does not depend on its angle. This allows us to separate the static photon density into purely angle and energy dependant parts, $n(\epsilon_s, \theta_s, \phi_s) = f(\omega_s)\,\Phi(\theta_s, \phi_s)$. By including multiple ways of defining $f(\omega_s)$ and $\Phi(\theta_s, \phi_s)$, we are able to model a diverse range of interactions. These are summarised in tables 4.1 and 4.2, and any combination can be used to build up the field.

Table 4.2: Static photon angle distributions implemented in the package.

Source Name	$\Phi(\theta_s, \phi_s)$	Description and Parameters
Beam	$\delta(\theta_s - \theta')\delta(\phi_s - \phi')$	Collimated source in the direction defined by θ' and ϕ'. Used for benchmarking and approximating a large focal length laser.
Point source	$\frac{1}{r^2}\delta(\theta_s - \theta')\delta(\phi_s - \phi')$ $\theta' = \cos^{-1}\frac{z}{r}$ $\phi' = \tan^{-1}\frac{y}{x}$	Emission at a distance $r = (x, y, z)$ from a point source. Used when the size of the emission surface is small compared to r.
Isotropic	$\frac{\sin\theta_s}{4\pi}$	Photon distribution is the same at all angles. Used to model a photon source in thermodynamic equilibrium, such as astrophysics systems or radiation from a hohlraum.
User defined	Tabulated	Distribution read from file. Used to model a complex photon source, such as the radiation observed close to a burn-through foil.

4.2 Test Simulations

In this section we will present some benchmark simulations which test the implementation of the algorithm discussed in section 4.1. By choosing a relatively simple setup, calculations for both the interaction probability and the outgoing particle's momentum can be performed directly, working with distributions over parameters. We will then compare these direct calculations to the estimates made by the Monte Carlo algorithm.

To carry out these test simulations, we will use the most simple photon source possible, with density given by

$$n(\omega_s, \theta_s, \phi_s) = n'\delta(\omega_s - \omega')\delta(\theta_s - \theta')\delta(\phi_s - \phi') \tag{4.12}$$

which corresponds to a monoenergetic and beamed source in tables 4.1 and 4.2 respectively. We will also assume the dynamic particle is travelling perpendicular to the static

photons. The momenta of the particles is then given by

$$P_d{}^{\mu} = \begin{pmatrix} E_d \\ 0 \\ 0 \\ p_d \end{pmatrix} \qquad k_s{}^{\mu} = \begin{pmatrix} \omega \\ \omega \\ 0 \\ 0 \end{pmatrix} \tag{4.13}$$

which simplifies equations 4.10 and 4.11 to

$$\tan\phi_R = \frac{\omega_s}{p_d} \qquad \beta = \frac{\sqrt{p_d^2 + \omega_s^2}}{E_d + \omega_s}. \tag{4.14}$$

The final simplification we will make is that the particles scatter in the CM frame with $\phi_{\text{cm}} = 0$. This angle corresponds to a rotation around the Lorentz boost axis and has no effect on the energy of the particles in the simulation frame. Therefore, we can write the momentum of the final state particles in the CM frame as

$$P_1{}^{\mu'} = \begin{pmatrix} E_1' \\ 0 \\ p_{1/2}' \sin\theta_{\text{cm}} \\ p_{1/2}' \cos\theta_{\text{cm}} \end{pmatrix} \qquad P_2{}^{\mu'} = \begin{pmatrix} E_2' \\ 0 \\ -p_{1/2}' \sin\theta_{\text{cm}} \\ -p_{1/2}' \cos\theta_{\text{cm}} \end{pmatrix}. \tag{4.15}$$

4.2.1 Interaction Probability

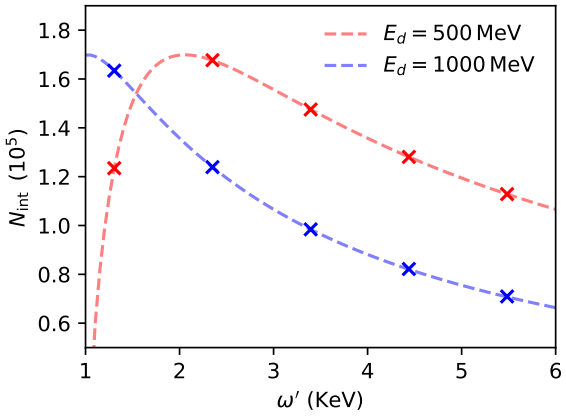
Upon inserting the field density from equation 4.12 into equation 4.4 and using $\theta' = 90°$, the MFP simplifies significantly to the following

$$\frac{1}{\lambda_d} = \begin{cases} 0 & \mathrm{s}' < \mathrm{s}_{\text{min}} \\ n'\sigma(s') & \mathrm{s}' > \mathrm{s}_{\text{min}} \end{cases} \tag{4.16}$$

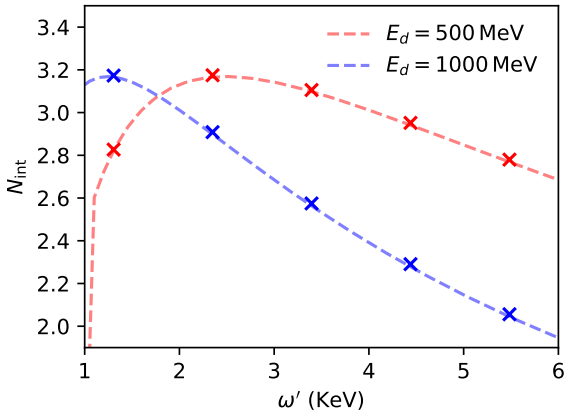
where $s' = 2E_d'\,\omega'$. If the static photon field has a length l_s, and we pass through a beam of N_d dynamic particles, the expected number of interactions is

$$N_{\text{int}} = N_{\text{d}}\,(1 - e^{-x/\lambda_d}). \tag{4.17}$$

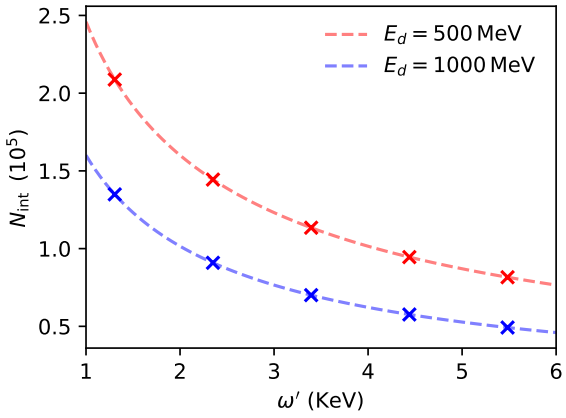
This equation is plotted in figure 4.3 as a function of ω', for each of the three processes. Also plotted are estimates of N_{int} calculated using the algorithm described in the previous section. We can see the Monte Carlo estimates agree well with the analytical calculation.



(a) Breit-Wheeler process



(b) Photon-photon scattering



(c) Compton scattering

Figure 4.3: Number of interactions against static photon energy using analytical calculation (dash line) and the Monte Carlo algorithm (markers). (a) shows the Breit-Wheeler process, (b) shows photon-photon scattering and (c) shows Compton scattering events. Cross-section biasing has been used to reduce errors.

4.2.2 Interaction Dynamics

The second part of the algorithm involves calculating the four momenta of the final state particles in the simulation frame. Here, we will focus on the following two components of the four momentum

$$E_{1/2} = P^0_{1/2} \qquad \theta_{1/2} = \cos^{-1}\left(\frac{P^3_{1/2}}{p_{1/2}}\right) \tag{4.18}$$

which are the simulation frame energy and polar scattering angle respectively. By applying transforms to the particles' momentum distribution in the CM frame, we can directly obtain PDFs for $E_{1/2}$ and $\theta_{1/2}$. These PDFs can also be estimated from the Monte Carlo algorithm and compared.

By assuming $\phi_{\text{cm}} = 0$, the only free parameter of the system in the CM frame is the polar scattering angle θ_{cm}. This has a PDF, $p(\theta_{\text{cm}})$, which is distributed according to

the differential cross-section. From $p(\theta_{\rm cm})$ we can obtain distributions over $x = E_{1/2}, \theta_{1/2}$ through the change of variables formula

$$p(x) = \sum_{\theta_{\rm cm} \in g^{-1}(x)} \frac{p(\theta_{\rm cm})}{|g'(\theta_{\rm cm})|} \tag{4.19}$$

where $x = g(\theta_{\rm cm})$ and the sum accounts for multiple solutions of $g^{-1}(x)$. To solve equation 4.19, we require $E_{1/2}$ and $\theta_{1/2}$ as a functions of $\theta_{\rm cm}$. These are obtained through equation 4.18 where

$$\begin{aligned} {P_{1/2}}^{\mu} &= {R^{\mu}}_{\nu}\, {\Lambda^{\nu}}_{\rho'}\, {P_{1/2}}^{\rho'} \\ &= \begin{pmatrix} \gamma({E_{1/2}}' \pm \beta\, p' \cos\theta_{\rm cm}) \\ \sin\theta_{\rm cm} \\ \gamma \sin\phi_R\, (\beta {E_{1/2}}' \pm p' \sin\theta_{\rm cm}) \\ \gamma \cos\phi_R\, (\beta {E_{1/2}}' \pm p' \cos\theta_{\rm cm}) \end{pmatrix}. \end{aligned} \tag{4.20}$$

Here, $\gamma = (1-\beta^2)^{-1/2}$ is the Lorentz factor and $+/-$ is used for particle $1/2$ respectively. In general, $E_{1/2}, \theta_{1/2} = g(\theta_{\rm cm})$ cannot be inverted analytically, so a numerical inversion is used[1].

When the two outgoing particles have equal mass, the distributions over the simulation frame parameters are identical (*i.e.* $p(E_1) = p(E_2)$ and $p(\theta_1) = p(\theta_2)$). This is the case for both the Breit-Wheeler process and photon-photon scattering. However, for Compton scattering the particles have a different mass and the distributions over the simulation frame parameters are different (*i.e.* $p(E_1) \neq p(E_2)$ and $p(\theta_1) \neq p(\theta_2)$). Therefore, across the three processes, there are four unique distributions over both energy and angle, which are shown in figure 4.4. Also plotted here are estimates of the distributions obtained from the Monte Carlo algorithm given in section 4.1. A good agreement with the analytical calculation is found across all processes.

4.3 Gaussian Process Regression Extension

For a realistic two beam experiment, the source distributions cannot be described by delta functions as used in the section 4.2. In this case, the direct calculation we performed is more inefficient than the Monte Carlo algorithm presented in section 4.1. However, due to the low probability of interaction the algorithm requires a large number of particles to be simulated to reduce the Monte Carlo error. Variance reduction techniques, such as cross-section biasing (see section 3.3.2), can be used to reduce some of the computational

[1]The inversion is only possible analytically for photon-photon scattering due to both the outgoing particles being massless.

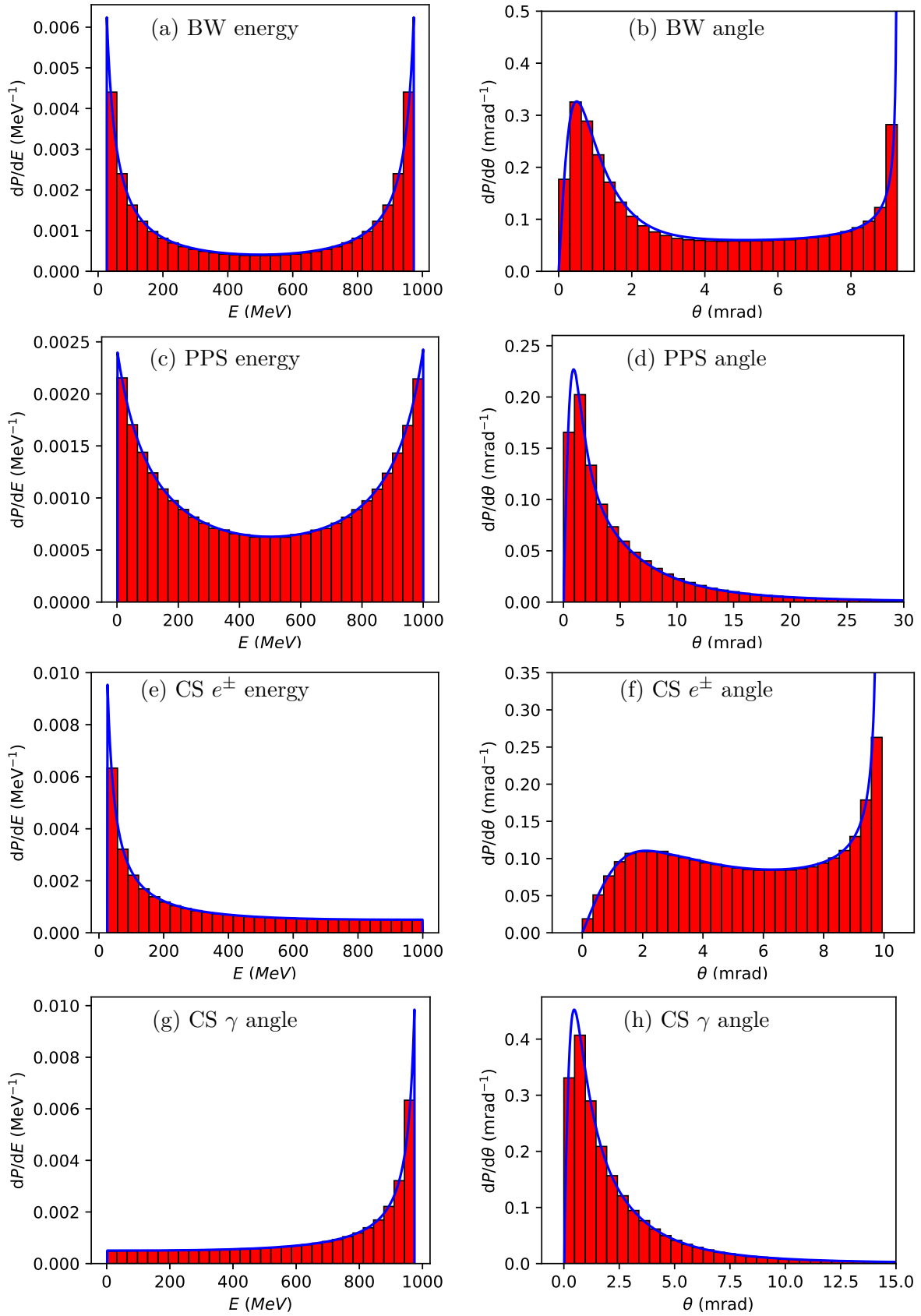


Figure 4.4: Unique distributions of $E_{1/2}$ and $\theta_{1/2}$ for the three processes in the package. Here an $E_d = 1\,\text{GeV}$ dynamic particle beam was interacted at 90° with an $\omega_s = 5\,\text{keV}$ static photon field. In each plot the blue curve shows the direct calculation of equation 4.19 and the red histogram shows the estimate from the algorithm in section 4.1. (a) and (b) shows E_e and θ_e for the Breit-Wheeler process. (c) and (d) shows E_γ and θ_γ for photon-photon scattering. (e) and (f) shows E_e and θ_e for the Compton scattered electron/positron. (g) and (h) shows E_γ and θ_γ for the Compton scattered photon.

costs, although not all. The two expensive parts of the model are calculating λ_d from equation 4.4 and sampling the static photon properties. In most simulations of interest to this work, the sampling is rarely performed as the probability of a single event interacting is much less than 1. However, λ_d must be calculated for each event making it by far the most intensive part of the algorithm.

We can solve this problem using an approximation to λ_d demoted by λ_a, calculated with a cheaper function than equation 4.4. An interpolation table is a common method of approximation and is already used in the algorithm for sampling from the differential cross-section. Here, the differential cross-section table is calculated once and then used in all future simulations. However, a λ_a interpolation table would be simulation specific, as it depends on the photon density $n(\omega_s, \theta_s, \phi_s)$ which changes between simulations. Generating a three dimensional interpolation table before every simulation would likely take more computational resources, not less.

In this work, we have used an alternative approach, where a regression model is used to calculate λ_a which is trained dynamically as the simulation progresses. We have used Gaussian process regression (GPR) as this is a nonparametric method, so no prior knowledge of the functional form of λ_d is required (see section 3.2.4). GPR is also a Bayesian method, allowing us to quantify the uncertainty of our estimate. As we will see, the Bayesian paradigm is an important aspect which allows us to extend the algorithm presented in section 4.1 with a regression model that is both trained and used during runtime. In this section, we will discuss the extension of the algorithm and provide examples showing its benefits.

4.3.1 Gaussian Process Regression Algorithm

Equation 4.4 is an expensive function mapping the dynamic particle's properties, $\mathbf{x} \equiv (E_d, \theta_d, \phi_d)$, to the MFP, $\lambda_d = \lambda_d(\mathbf{x})$. The components of $\mathbf{x}$ are not explicitly stated in equation 4.4, however, it depends on them through the rotation to the dynamic particle frame. Due to the threshold behaviour of some QED processes, λ_d can vary rapidly with small changes in $\mathbf{x}$. Therefore, it is more computationally efficient to learn a function which returns the log of the MFP estimate, *i.e.*

$$\lambda_a = e^{f(\mathbf{x})} \tag{4.21}$$

where $f(\mathbf{x})$ is the GPR model. In using a GPR, we have assumed that $f(\mathbf{x})$ is Gaussian distributed. Therefore, λ_a is log-normal distributed with mean and variance given by

$$\mu_{\lambda_a} = \exp\left(\mu_f - \frac{\sigma_f^2}{2}\right) \qquad \sigma^2_{\lambda_a} = \left[\exp(\sigma_f^2) - 1\right] \exp(2\mu_f + \sigma^2) \tag{4.22}$$

where μ_f and σ_f^2 are the mean and variance of $f(\mathbf{x})$.

To train the GPR a data set, $\mathcal{D} = \{\mathbf{x}_i, \log(\lambda_i)\}$, is required. This is generated by calculating equation 4.4 for a limited number of events. In the units system of Geant4, the dimensions of $\mathbf{x}$ vary over vastly different scales. Therefore, if the regression model is trained directly on $\mathcal{D}$ it would perform poorly. To solve this problem $\mathbf{x}$ is normalised using min-max feature scaling.

We can summarise the GPR extension to the algorithm in the following three stages:

Gaussian Process Algorithm

1. *Data accumulation*: For the first n_t events, λ is obtained by solving equation 4.4. The result is saved, generating $\mathcal{D}$.

2. *Training*: After simulating n_t events the GPR is trained on $\mathcal{D}$ by optimising the hyperparameters.

3. *Acceleration*: For subsequent events, σ_{λ_a} is calculated using equation 4.22. If $\sigma_{\lambda_a} < \sigma_{\max}$ where $\sigma_{\max}$ is a user defined limit, we set $\lambda_d = \mu_{\lambda_a}$. If $\sigma_{\lambda_a} > \sigma_{\max}$ equation 4.4 is solved and the result is appended to $\mathcal{D}$. After another n_t points are added to $\mathcal{D}$ the hyperparameters are again optimised.

During stage 1, apart from saving $\mathcal{D}$, the extended and original algorithms perform the same calculations. Therefore, both calculate events at approximately the same rate. During stage 2, the extended algorithm pauses the calculations of new events while the hyperparameters are optimised. The original algorithm will overtake the extended algorithm slightly at this point. However, during stage 3, this is reversed. As $\mathcal{D}$ grows, σ_{λ_a} will decrease. There is then a higher probability that the GPR is used over equation 4.4, and can be performed much quicker. This then leads to a dramatic increase in the rate at which new events are calculated.

Through the extended algorithm presented here, the acquisition of $\mathcal{D}$ occurs at runtime. This offers several advantages over a pre-trained model where $\mathcal{D}$ is generated before. For example, a pre-trained model requires prior knowledge of how $\mathbf{x}$ is distributed to be effective. If not, generating $\mathcal{D}$ involves unnecessary calculations of highly unlikely values of $\mathbf{x}$. We may also encounter large errors from extrapolation of points outside the boundary of $\mathcal{D}$. Having a prior knowledge of how $\mathbf{x}$ is distributed is difficult if the dynamic particles are not primary but generated through a multi-step process. Finally, using a dynamic data set has the advantage that it automatically minimises the number of times equation 4.4 is calculated for a desired uncertainty.

4.3.2 GPR Algorithm Test

Having discussed the advantages of the extended GPR algorithm, we will now demonstrate this. Here, we will use a simplified Breit-Wheeler detection experiment. By performing the same simulation with both algorithms, we can observe the computational saving provided by the extended one. Here, we will use a dynamic gamma ray source of $N_\gamma = 10^8$ particles, with a divergence of 10 mrad and a Gaussian energy spectrum with mean 1 GeV and standard deviation 0.25 GeV. This will interact with an isotropic black body radiation field

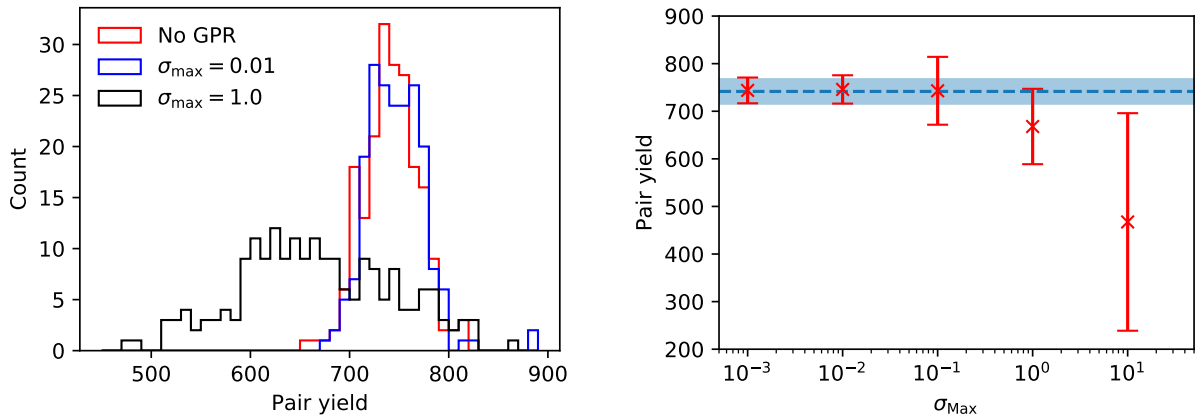
$$n(\omega_s, \theta_s, \phi_s) = \frac{\omega_s^2}{\pi^2} \frac{1}{e^{\omega/T} - 1} \frac{\sin\theta_s}{4\pi} \tag{4.23}$$

of temperature $T = 300\,\text{eV}$ and length $l = 1\,\text{cm}$. However, before we can use the algorithm to its full potential, the two additional parameters $\sigma_{\max}$ and n_t should be optimised.

Optimising Algorithm Parameters

The parameter $\sigma_{\max}$ sets the limit above which the GPR is too uncertain to use. If this is set too high, there will be a large error in our estimate of the number of interactions, N_{BW}. However, if it is too low, equation 4.4 will be performed unnecessarily and there will be little gain in efficiency. The optimum $\sigma_{\max}$ is given by the largest value at which the GPR extended algorithm and the original algorithm produce the same result. However, both these algorithms are statistical and include uncertainty in N_{BW} due to the finite number of events simulated. Therefore, we will run an ensemble of simulations for both methods and compare the distributions over N_{BW}. An example of this comparison is shown in figure 4.5 (a) for two different values of $\sigma_{\max}$. When $\sigma_{\max} = 0.01$ there is little difference between the distributions. However, for $\sigma_{\max} = 1.0$ this is not the case and the GPR algorithm adds significantly more uncertainty to the estimate. This is more clearly demonstrated in figure 4.5 (b) which presents a scan over $\sigma_{\max}$. We can also see the mean of N_{BW} shifts down as $\sigma_{\max}$ is increased. This is due to the exponential transform applied in equation 4.21. This is an asymmetric transform and leads to λ_a being overestimated on average and the mean of N_{BW} decreasing.

The second parameter, n_t, is the number of data points taken between optimising the GPR hyperparameters. Setting this to a large value will have a detrimental effect on the efficiency as the GPR is queried many times without optimised hyperparameters. To obtain the smallest error for each event, we would set $n_t = 1$. However, the hyperparamter optimisation is the most computationally intensive part of GPR, scaling as $\mathcal{O}(N_d^3)$ where N_d is the number of events in $\mathcal{D}$. Apart from the initial few events, it is also unlikely that the addition of a single event will have much effect on the optimum hyperparameters. Therefore, setting n_t too low will also have a detrimental effect on efficiency. Figure 4.6 is a plot of simulation run-time against n_t and demonstrates the trade-off between these



(a) Pair yield histograms

(b) Pair yield against $\sigma_{\max}$

Figure 4.5: Effect of varying $\sigma_{\max}$. (a) compares histograms for the original algorithm, given in red, to the GPR extended algorithm, given in blue ($\sigma_{\max} = 0.01$) and black ($\sigma_{\max} = 1.0$). (b) shows the change in the average and standard deviation of N_{BW} against $\sigma_{\max}$. The blue dashed line and shaded area shows the mean and standard deviation of the original algorithm.

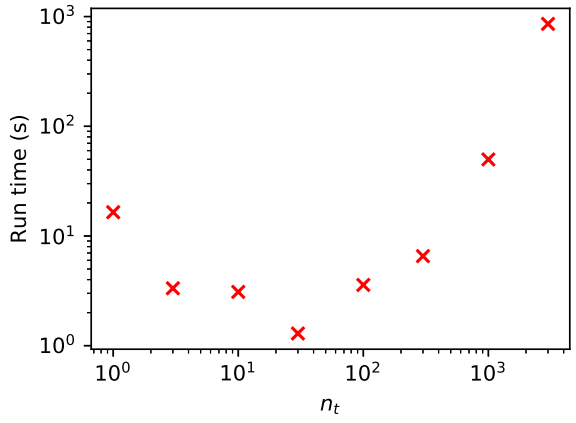
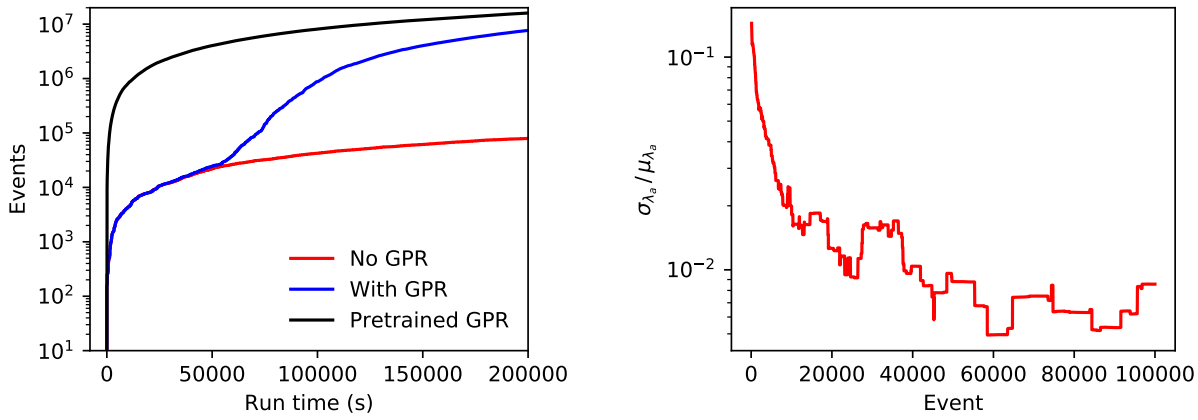


Figure 4.6: Effect of varying the parameter n_t on the simulation run-time

two factors.

With the parameters set to $\sigma_{\max} = 0.01$ and $n_t = 30$, figure 4.7 shows a comparison of the number of events calculated against run-time and a running average of the GPR error. For the first $\approx 10^4$ particles, both algorithms have the same rate of event calculation. We can see that during this period, the average σ_{λ_a} remains above $\sigma_{\max}$, meaning the GPR is rarely used. However, beyond this point the average σ_{λ_a} drops below $\sigma_{\max}$ and the event rate increases dramatically. The average σ_{λ_a} then fluctuates just below $\sigma_{\max}$ as few new data points are added to $\mathcal{D}$. By the end of the simulation, the inclusion of the GPR has increased the rate of event calculation by a factor of ~ 100.

In some circumstances the same static photon field will be used for multiple simulations. For example, when optimising an experimental setup with a fixed photon source.



(a) Events calculated against run-time

(b) Model error against event number.

Figure 4.7: Demonstration of GPR algorithm efficiency. (a) shows the number of events calculated against the simulation run-time for the original algorithm (red), the untrained GPR extended algorithm (blue) and the pre-trained GPR extended algorithm (black). (b) shows a running average, taken over 10^3 events, of σ_{λ_a} for the untrained GPR extended algorithm.

In this case, we can save the GPR model at the end of the first simulation, then use it pre-trained in the follow simulations. The effect this has on the rate of event calculation is also shown in 4.7 (a). We can see this avoids the slows data accumulation stage, providing an increase in efficiency from the start. By still allowing $\mathcal{D}$ to grow dynamically for the pre-trained model, we still maintain the advantages discussed previously.

4.4 QED Experiments With a Thermal Radiation Field

In this final section, we will demonstrate the capabilities of this new physics package by using it to model a Breit-Wheeler and photon-photon scattering detection experiment. The reason for using Geant4 as the framework for this package is that it has an extensive toolkit, which enables full start-to-end simulations to be performed. For example, within a single simulation we can model the generation of the dynamic particle source; the tracking of particles through fields; the interaction between particles and detectors; and the generation of background noise. Although vital for a successful experiment, we will not consider the last two parts here. However, in chapter 6 they will be discussed in detail for a real Breit-Wheeler detection experiment.

Here, we will study the prospect of detecting the Breit-Wheeler process and photon-photon scattering using a thermal radiation field. This setup was originally proposed by Pike *et. al* [157], using a laser-heated hohlraum to produce the thermal radiation field. Using a megajoule class laser, such as the national ignition facility (NIF), a hohlraum radiation temperature of 300 eV can be reached over a length of ≈ 1 cm [164]. To generate

the second sources, they suggested using the bremsstrahlung emission from a beam of LWFA electrons in a high-Z material foil. LWFA are capable of producing $> 2\,\mathrm{GeV}$ beams with charges in excess of $100\,\mathrm{pC}$ [165].

The time-scales of these sources are vastly different, with the radiation from a hohlraum varying over nanoseconds and a LWFA beam varying over femtoseconds. Therefore, we can safely assume the thermal radiation field is static over the duration of the interaction. This makes these sources ideal for demonstrating the package presented here.

Pike [166] has studied in detail how important experimental parameters, such as radiation temperature, electron beam energy, and bremsstrahlung foil thickness, affect the yield of Breit-Wheeler pairs. Therefore, we will not investigate them all here but instead focus on where this package excels. This is the ability to perform experimental optimisation within a fully integrated 3-dimensional simulation. We will assume the experiment is to be performed with a fixed laser system. Therefore, the electron beam and radiation field cannot be easily changed. The only part which can be easily varied is the converter foil. The properties of the bremsstrahlung emission depends on both the thickness of the foil and the material. However, as we will see, these can both be summarised in a single parameter, known as the radiation length.

4.4.1 Bremsstrahlung Converter

As a beam of electrons passes through a material, bremsstrahlung radiation is emitted, due to electrons scattering off the nuclear field of atoms. The high energy gamma rays produced can themselves interact in the nuclear field, producing electron-positron pairs through the Bethe-Heitler process. This causes a cascade as the number of particles increases but their individual energy decreases. This is detrimental for two reasons. The production of electron-positron pairs leads to a large background noise sources, especially for detecting the Breit-Wheeler process. Also, the gamma ray energy may fall below the required CM threshold.

Both bremsstrahlung and the Bethe-Heitler process occur over a similar characteristic length, X_0, known as the radiation length [167]. For $Z < 4$, a good approximation for the radiation length in a single element material is [168]

$$\frac{1}{X_0} = 4\alpha r_e^2 \frac{N_A}{A}\Big\{Z^2[L_{\mathrm{rad}} - f(Z)] + Z L'_{\mathrm{rad}}\Big\} \tag{4.24}$$

where A is the mass number, α is the fine structure constant, r_e is the classical electron radius, N_A is Avogadro's constant and

$$L_{\mathrm{rad}} = \ln(184.15\, Z^{-1/3}); \qquad L'_{\mathrm{rad}} = \ln(1194\, Z^{-2/3}). \tag{4.25}$$

The function $f(Z)$ is known as the Coulomb correction and for $Z < 92$, can be approxi-

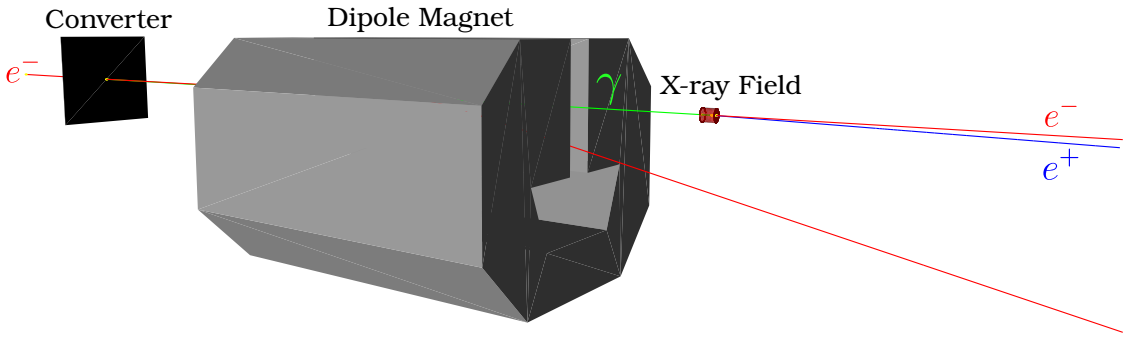


Figure 4.8: Visualisation of a Breit-Wheeler production event. An electron (red line) is incident on a gold foil target. This produces a high energy photon (green line) through bremsstrahlung emission. The electron is swept off axis by a magnetic field and the photon propagates into a thermal radiation field. Here the photon undergoes the Breit-Wheeler process, producing an electron (red line) / positron (blue line) pair.

mated by [169, 167]

$$f(Z) = a^2 \left[(1 + a^2)^{-1} + 0.20206 - 0.0369\, a^2 + 0.0083\, a^4 - 0.002\, a^6 \right] \tag{4.26}$$

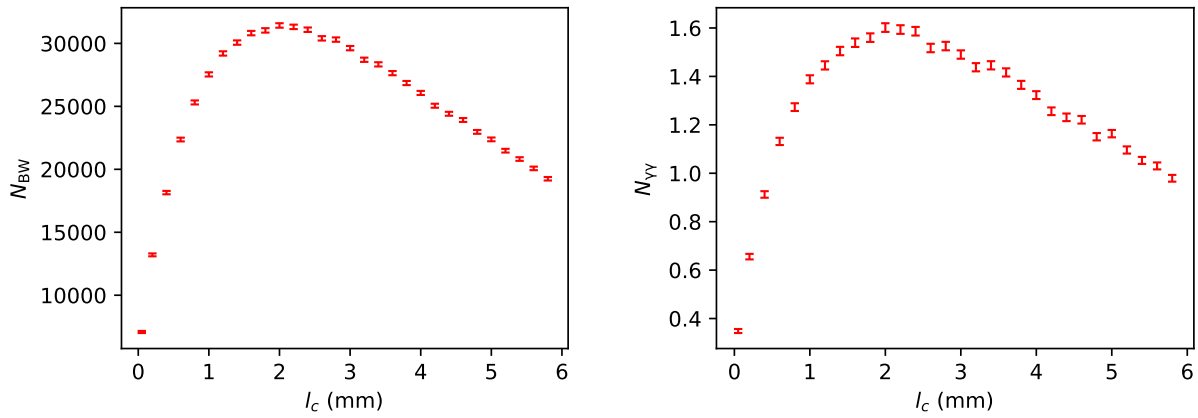
where $a = \alpha Z$.

Through bremsstrahlung, on average an electron's energy will decrease by a factor of $1/e$ over X_0. Similarly, the mean free path of a photon due to the Bethe-Heitler process is $\frac{9}{7}X_0$. Therefore, to suppress cascades, producing the highest energy gamma ray beam possible, the converter foil thickness should be on the order of X_0.

4.4.2 Detecting the Breit-Wheeler Process

To detect the Breit-Wheeler process, it is important that the Bethe-Heitler pairs generated in the converter foil and residual electrons are removed. This can be achieved by placing an on-axis magnet before the radiation field. For electrons up to 2 GeV, a 30 cm magnet with a field strength of $B = 1\,\mathrm{T}$ is sufficient. This is demonstrated in figure 4.8, which shows a single event from a Geant4 simulation. Here, we can see the generation of a bremsstrahlung photon, the tracking of particles through the field, and the production of a Breit-Wheeler pair in the radiation field.

To optimise the converter for the maximum number of Breit-Wheeler pairs, we could vary both the thickness and material. However, the only important parameter is the number of radiation lengths. This is determined by both the thickness and material of



(a) Breit-Wheeler

(b) Photon-photon scattering

Figure 4.9: Geant4 simulations caried out to optimise the converter thickness using 10^9 primary events. (a) shows the Breit-Wheeler pair yield and (b) shows the Photon-Photon scatter yield, both as functions of converter thickness. The errors arise from the Monte Carlo method. To reduce these, cross-section biasing has been used.

the converter. Therefore, we can fix the material as gold and only vary the thickness. The radiation length for gold is $X_0 = 3.344\,\text{mm}$. Figure 4.9 (a) shows the Breit-Wheeler pair yield, N_{BW}, when the converter thickness is varied around this value. We can see the competition between the photon creation and cascades, with an optimum reach at $\approx 2\,\text{mm}$. Here, a large number of pairs are produced (~ 30000 from 10^9 primary electrons). This demonstrates the feasibility of using this scheme to detect the Breit-Wheeler process, provided the background noise is not too significant.

By performing a more detailed simulation, with the thickness set to 2 mm, we can examine the energy and angle distribution of the Breit-Wheeler pairs produced. This is shown in figure 4.10. We can see the Breit-Wheeler pairs have a broad energy spectrum extending up to the energy of the primary electrons. However, the angle distribution is narrow, and sharply peaked around the gamma ray axis. The beaming is due to the gamma ray carrying most of the momentum in the interaction. This is one of the major benefits of this scheme. The Breit-Wheeler pairs can be easily captured and transported off to detectors with a second on-axis magnet. If they were emitted at all angles, the interaction region would have to be surrounded by detectors to capture a large fraction of the pairs.

4.4.3 Detecting Photon-Photon Scattering

We can also perform a similar analysis to study the prospect of using this setup to detect photon-photon scattering. In figure 4.9 (b), a scan over the converter thickness has again been performed, this time investigating the effect on the photon-photon scattering yield. The maximum is again obtained at $\approx 2\,\text{mm}$, so the setup is optimised for both

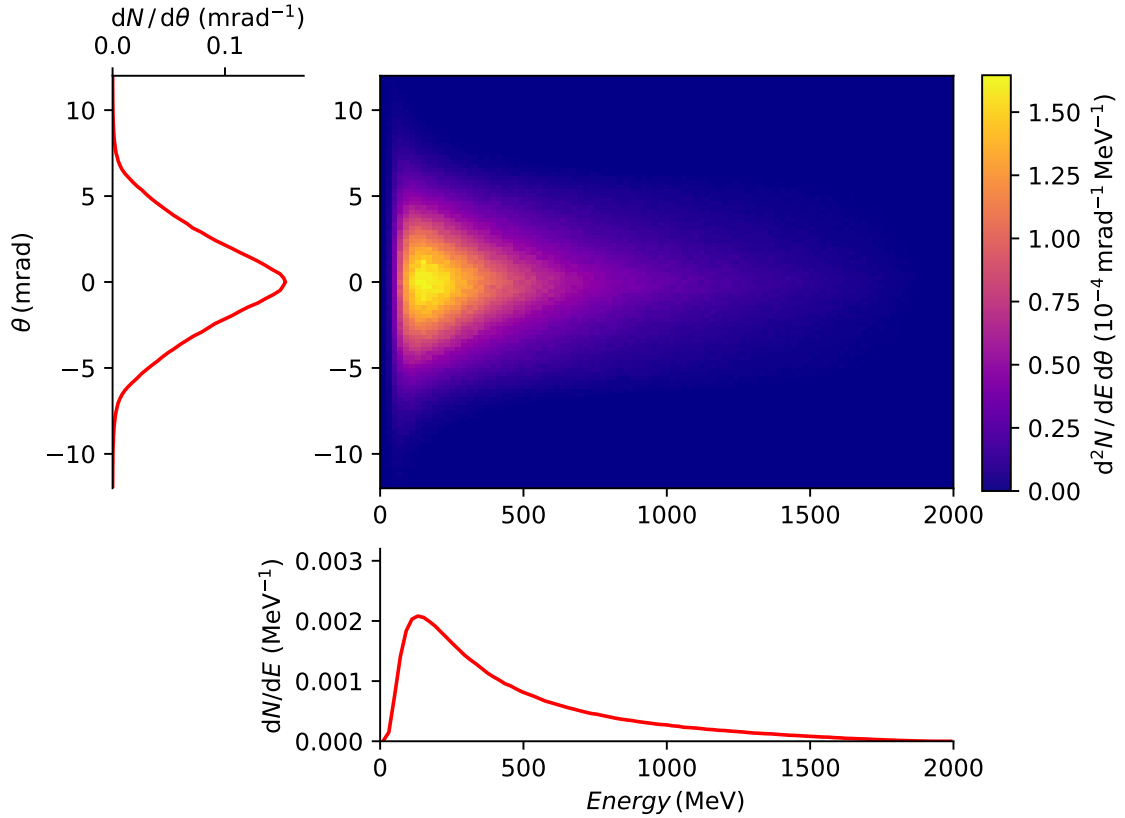


Figure 4.10: Energy and angular distribution of electron-positron pairs, generated through the Breit-Wheeler process. Also shown are the marginalised distributions..

processes. This is to be expected, as the cross-section for both the Breit-Wheeler process and photon-photon scattering are maximised at the same CM energy ($s = 4$). However, the total yield of photon-photon scattering events is significantly reduced compared to Breit-Wheeler pairs. With only ≈ 1 event per shot, when the noise is also considered, this would be a very challenging measurement to make.

The energy and angle distribution from photon-photon scattering with a converter thickness of 2 mm is shown in figure 4.11. We can see the interaction products are again beamed along the gamma ray source axis. Although this is a benefit for detecting the Breit-Wheeler process, it becomes a hindrance for detecting photon-photon scattering. As both the signal and bremsstrahlung background consist of photons, separating them is difficult. The only way to do so is to detect both signal and noise, then look for signatures in the distribution due to photon-photon scattering.

In figures 4.12 (a) and (b) we can see the marginalised energy and angle distributions, for both the bremsstrahlung background and the photon-photon scattering signal. The bremsstrahlung background is many orders of magnitude brighter and covers the same range of energies and angles as the photon-photon signal. Therefore, it is not possible to detect photon-photon scattering with this setup. However, it may be possible to reduce the wings of the bremsstrahlung angle distribution if the beam is collimated before entering

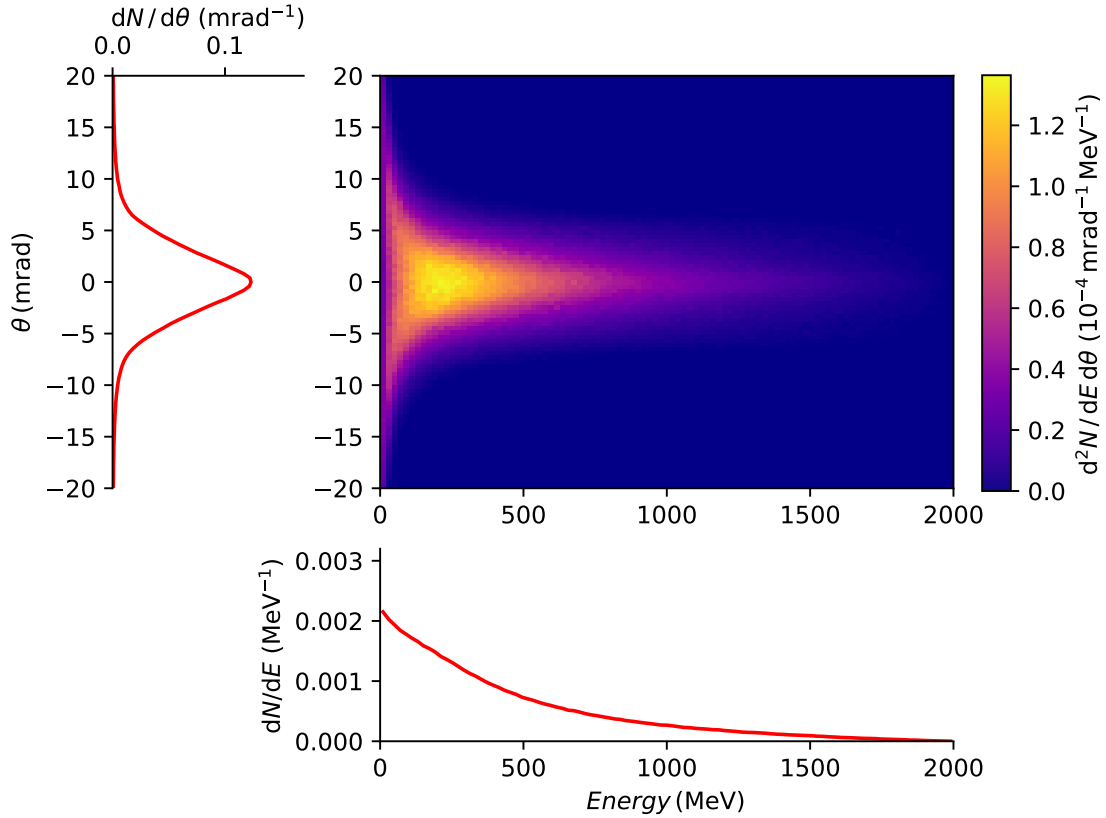
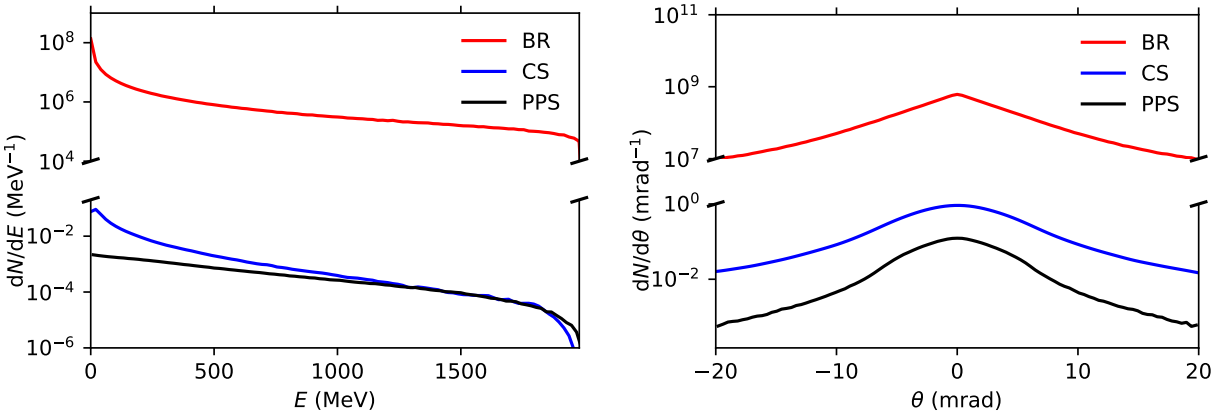


Figure 4.11: Energy and angular distribution of electron-positron pairs, generated through photon-photon scattering. Also shown are the marginalised distributions.

the radiation field.

The probability of photon-photon scattering occurring is proportional to α^4. Therefore, we should also consider the probability of two processes proportional to α^2 happening together. For example, a photon could produce an electron positron pair through the Breit-Wheeler process, then emit a photon through Compton scattering. This interaction will also be proportional to α^4, however, it will depend on the square of the photon density and radiation field length. The energy and angle distributions of photons produced through this two-step interaction are also shown in figures 4.12 (a) and (b). For the parameters used here, the two-step process occurs at a rate ≈ 10 times higher than photon-photon scattering. As these photons are generated at the same location, they cannot be removed with a collimator. The only way to suppress the two-step process, without limiting photon-photon scattering, is to bring the CM energy below the Breit-Wheeler threshold. However, this is difficult if sources with a broad energy spectrum are used.



(a) Breit-Wheeler

(b) Photon-photon scattering

Figure 4.12: Energy and angle distributions of photons produced in Breit-Wheeler process (BW, red), Compton scattering (CS, blue), and photon-photon scattering (PPS, black). (a) shows the energy distributions and and (b) shows the angular distributions.

4.5 Summary

In this chapter, we have presented the development of a new physics package for Geant4 that models particle-photon processes. To integrate the package with Geant4, we treat the photon source as static, which limits the types of experiments which can be modelled. However, in doing so we have reduced the complexity of the problem. It has also led to a scheme in which a GPR can be implemented, leading to a large increase in the algorithm efficiency.

Using Geant4 as the framework for the package enables start-to-end simulations of experiments to be performed. We have demonstrated this by studying the prospect of using the photon-photon collider, proposed by Pike *et. al*, to detect both the Breit-Wheeler process and photon-photon scattering. Using realistic experimental parameters, this scheme can produce a large number of Breit-Wheeler pairs. This is in agreement with the previous study of this setup by Pike [166], suggesting a measurement could be made with only a few shots.

However, there are several challenges that will make using this setup to detect photon-photon scattering extremely difficult. These challenges include: the small value for the cross section; the outgoing particles being the same particle type as the bremsstrahlung source; the scattered photons beaming along the bremsstrahlung source axis; and background noise due to the two-step process involving the Breit-Wheeler process and Compton scattering. The first two will be true for any photon-photon scattering experiment. However, the second two are due to the setup. If two symmetric photon sources are used, there is no dominant axis, making photons more likely to scatter into an angle less dominated by noise. Because the two-step process depends on the square of the field density

and length, by reducing these parameters we can bring the rate below the photon-photon scattering rate. However, this also reduces the number of photon-photon scattering events, which is already small. To suppress the two-step process without this happening, the CM energy of the collision can be brought below the threshold for the Breit-Wheeler process. The cross-section for photon-photon scattering also decays rapidly below this value, so two monoenergetic sources at $\approx 0.5\,\mathrm{MeV}$ should be used. This could be generated using an inverse Compton source, as suggested by Micieli *et. al.* [170, 159].

The processes discussed here are binary, and depend on the photon density in a linear way. At higher photon densities, nonlinear effects can dominate. For the nonlinear Breit-Wheeler process, the extra particles in the interaction reduce the energy required by each one to overcome the CM threshold. This allows low energy but high density photon sources to be used, such as a high-power IR laser. These types of experiments will be studied in the next chapter.

Chapter 5

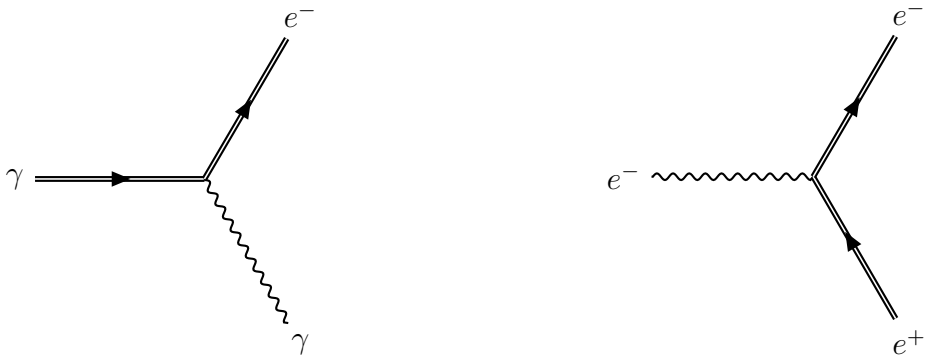
Modelling Strong Field QED Interactions

The classical mechanism through which an accelerating charge emits electromagnetic radiation has been extensively studied and understood since the late 19th century [85]. As the charge radiates, it loses energy at a rate given by the Larmor formula. This loss of energy implies a damping force on the charge, however, this is not accounted for in the Lorentz force. Early attempts to amend this by Abraham [171] and Dirac [92], led to unphysical predictions such as runaway solutions and forces from the future (see section 2.1.7). It was not until the 1970s that Landau and Lifshiz found a solution to this problem, by reducing the order of the Abraham force [83]. The resulting Landau-Lifshiz equation of motion is generally accepted as the correct form for radiation reaction in the classical limit.

In deriving the Landau-Lifshiz equation, it is assumed that the force experienced by the charge from the external field is much larger than the dampening force due to radiation. However, in an ultra-strong electromagnetic field with $\eta \gtrsim 1$ (see equation 2.93), the radiation term can no longer be considered small, making the Landau-Lifshiz equation invalid. To derive a consistent description of radiation reaction, which holds at all fields strengths, requires a quantum theory.

The quantum theory of radiation reaction is an active area of research, requiring experimental evidence to test predictions. However, reaching these conditions in the laboratory is challenging and only a few experiments have been carried out to date. The first to probe the quantum regime was the E144 experiment at SLAC which took place in the 1990s [55, 56]. By interacting a 46.6 GeV electron beam with an $a_0 \approx 0.3$ laser, a nonlinear quantum parameter of $\eta \approx 0.2$ was reached. This provided the first evidence of nonlinear Compton scattering (see figure 5.1 (a)). However, no direct measurements were made to determine the effect of radiation reaction on the electron spectrum.

As a laser with $a_0 < 1$, was used, the experiment operated in the predominantly linear regime, in which nonlinear effects were highly suppressed. However, in two more



(a) Nonlinear Compton scattering (b) Nonlinear Breit-Wheeler process

Figure 5.1: Feynman diagrams for (a) nonlinear Compton scattering and (b) the nonlinear Breit-Wheeler process in a strong background field. The double lines used for the leptons represent the interaction with the strong background field.

recent experiments by Cole *et al.* [57] and Poder *et al.* [58] radiation reaction has been observed in the nonlinear regime with $a_0 > 10$. The measurements obtained during the Cole experiment were found to be more consistent with the fully quantum description of radiation reaction, whereas the Poder measurements were found to be more consistent with the quantum corrected Landau-Lifshitz equation. However, both experiments had large uncertainties in the overlap location due to the micron precision required. In addition to this, only a limited number of interactions were observed, highlighting the need for more experimental data.

A second nonlinear QED interaction of interest in a strong electromagnetic field is the nonlinear Breit-Wheeler process (see figure 5.1 (b)). This causes a photon passing though a field with $\chi \gtrsim 1$ (see equation 2.93), to decay into an electron-positron pair. Although the linear Breit-Wheeler process, which was discussed in chapter 4, has never been directly observed in the laboratory, its nonlinear counterpart has. This was during the E144 experiment at SLAC [56], in which the high energy gamma rays emitted through nonlinear Compton scattering combined with laser photons to produce electron-positron pairs. On average $n \approx 6$ laser photons were absorbed per interaction [172]. As this experiment operated in a regime dominated by linear effects, the interaction was highly unlikely, with only ≈ 100 pairs detected over 22000 shots.

This two-step interaction through which electron-positron pairs were created is an example of a QED cascade [173, 174, 175, 176, 177]. Here, through multiple nonlinear Compton scattering and nonlinear Breit-Wheeler events, a single seed particle can produce many final state particles. Understanding the dynamics of these cascades is important from an astrophysical perspective as they are predicted to occur in a number of extreme environments throughout the universe [35, 150].

Modelling strong field QED cascades requires a number of considerations. A direct QED approach is intractable due to the high level of multiplicity [178]. State of the art

calculations are limited to systems which contain only two additional particles in the final state [179, 172, 107, 108, 180]. As well as this, prolific generation of pairs leads to the depletion of the background field [181]. To account for these effects, the QED-PIC method has been developed [119, 120, 182]. This couples a Monte Carlo emission algorithm with the standard particle-in-cell (PIC) approach. To do so, the cascade is approximated by the product of multiple first order interactions with a classical transition in between. This is referred to as a semi-classical treatment and enables us to model both the generation and dynamics of cascades.

Current laser facilities are only able to operate with a relatively modest nonlinear quantum parameter ($\chi, \eta < 1$) and the extent of a cascade is limited. However, a large number of high energy photons are still generated, leading to a noisy experimental environment (for the same reason discussed in chapter 4). Detailed numerical modelling is then vital for both pre-experiment optimisation and post-experiment analysis. Therefore, we have developed a new strong field QED package for this purpose. This can either be used as a stand-alone piece of software or integrated with Geant4 to perform start-to-end simulations of experiments.

In this chapter we will discuss the development of this package and demonstrate some of its applications. We will start by reviewing the set of equations which the package solves and outline our specific implementation. One of the possible uses of this package is as a forward model for performing post experimental inference. However, the Monte Carlo algorithm used by the package is too slow for this purpose. Therefore, a machine learning emulator which can perform calculations far more quickly, has been developed. In this chapter we will discuss the implementation of this emulator and demonstrate its ability by performing Bayesian inference on mock experimental data.

5.1 Strong Field QED Package Development

The underlying Monte Carlo algorithm, which this package uses, was originally developed by Kirk *et. al* [119] and then adapted to include stochastic effects by Duclous *et.al* [120]. Since its development it has been integrated into a number of PIC codes [183, 184, 185]. This allows the bulk dynamics of an electron-positron plasma, generated through QED cascades, to be modelled. However, current laser facilities are not capable of reaching the regime in which plasma effects are important on the cascade dynamics. As the package discussed here has been developed as a tool for designing such experiments, it does not model the plasma dynamics. By ignoring particle-particle interactions, we can treat a strong field QED experiment, involving a beam of particles moving through a background field, as a system of independent events. This can be parallelised to a higher degree than QED-PIC, reducing the simulation run-time.

In this section we will present an overview of this new strong field QED package,

starting with a discussion of the equations which are solved. We will then review the specifics of the package, including the numerical methods used and the options the package provides. Finally, we will discuss how importance sampling can be used to reduce the computational costs in modelling an experiment.

5.1.1 System of Equations

At the heart of the Monte Carlo algorithm is the quasi-classical treatment of the radiation field. Using the Furry picture [103], we can split the electromagnetic field into a large classical background and small quantum fluctuations (as discussed in section 2.3). The motion of the particles through the field is then treated classically, with point-like QED interactions. Under the quasistatic and weak field approximations (see section 2.3.2), the rate of nonlinear Compton scattering, W_γ, and the nonlinear Breit-Wheeler process, $W_\pm$, are given by equations 2.107 and 2.116 respectively.

This semi-classical treatment of the particle's motion leads to a kinetic equation describing the dynamics of the probability distribution function (PDF) over the particle's phase space coordinates $\mathbf{x} = (\mathbf{r}, \mathbf{p})$. The PDF for the electrons and positrons is denoted by $f_\pm$, for gamma rays by f_γ and their motion governed by [182, 175, 121, 174]

$$\begin{aligned} \frac{\partial f_\pm}{\partial t} + \mathbf{v}\cdot\nabla f_\pm + \mathbf{F}_{\text{ext}}\cdot\nabla_{\text{p}} f_\pm &= \left(\frac{\partial f_\pm}{\partial t}\right)_{\text{QED}} \\ \frac{\partial f_\gamma}{\partial t} + c\mathbf{v}\cdot\nabla f_\gamma &= \left(\frac{\partial f_\gamma}{\partial t}\right)_{\text{QED}} \end{aligned} \tag{5.1}$$

where $\mathbf{F}_{\text{ext}} = e(\mathbf{E}_{\text{ext}} + \mathbf{v}\times\mathbf{B}_{\text{ext}})$ is the external Lorentz force. The left hand side of equation 5.1 describes the classical motion of the particles. This is a null geodesic for photons and the Lorentz motion for electrons and positrons. The terms on the right hand side of equation 5.1 account for particle sources and sinks due to QED effects and, using the jargon of statistical mechanics, can be viewed as a collision operator. The only two QED processes which we will consider here are nonlinear Compton scattering and the nonlinear Breit-Wheeler process. Therefore, we can write the collision operator for the electrons and positrons as [174]

$$\begin{aligned} \left(\frac{\partial f_\pm}{\partial t}\right)_{\text{QED}} = &\int f_\pm(t,\mathbf{x},\mathbf{p}')w_\gamma(\mathbf{p}',\mathbf{p}'-\mathbf{p})\mathrm{d}^3\mathbf{p}' \\ &- f_\pm(t,\mathbf{x},\mathbf{p})\int w_\gamma(\mathbf{p},\mathbf{k})\mathrm{d}^3\mathbf{k} + \int f_\gamma(t,\mathbf{x},\mathbf{k})w_\pm(\mathbf{k},\mathbf{p})\mathrm{d}^3\mathbf{k} \end{aligned} \tag{5.2}$$

where the first term accounts for an electron or positron leaving the state through nonlinear Compton scattering; the second term accounts for the reverse process where an

electron or positron enters the state through nonlinear Compton scattering; and finally, the third term accounts for electrons or positrons entering the state through the nonlinear Breit-Wheeler process. The equivalent collision operator for photons is [174]

$$\left(\frac{\partial f_\gamma}{\partial t}\right)_{\text{QED}} = \int [f_-(t,\mathbf{x},\mathbf{p}) + f_+(t,\mathbf{x},\mathbf{p})] w_\gamma(\mathbf{p},\mathbf{k}) \mathrm{d}^3\mathbf{p} - f_\gamma(t,\mathbf{x},\mathbf{k}) \int w_\pm(\mathbf{k},\mathbf{p}) \mathrm{d}^3\mathbf{p} \tag{5.3}$$

where the first term accounts for the emission of photons from nonlinear Compton scattering and the second term accounts for the photon annihilating through the nonlinear Breit-Wheeler process.

In equations 5.2 and 5.3, $w_\gamma(\mathbf{p},\mathbf{k})$ is the rate at which an electron or positron with momentum $\mathbf{p}$ emits a photon with wavevector $\mathbf{k}$ and is given by [117, 175]

$$w_\gamma(\mathbf{p},\mathbf{k}) = \int_0^{\eta/2} \frac{\mathrm{d}W_\gamma}{\mathrm{d}\chi} \delta\Big(\hbar\mathbf{k} - \frac{2\chi}{\eta}\mathbf{p}\Big) \mathrm{d}\chi \tag{5.4}$$

where $\mathrm{d}W_\gamma/\mathrm{d}\chi$ is given by equation 2.107. The delta function ensures that the photon is emitted in the same direction as the electron or positron[1]. Equivalently, $w_\pm(\mathbf{k},\mathbf{p})$ is the rate at which a photon with wavevector $\mathbf{k}$ annihilates to form an electron and positron with momentum $\mathbf{p}$

$$w_\pm(\mathbf{k},\mathbf{p}) = \int_0^{2\chi} \frac{\mathrm{d}W_\pm}{\mathrm{d}\eta} \delta\Big(\mathbf{p} - \frac{\eta}{2\chi}\hbar\mathbf{k}\Big) \mathrm{d}\eta \tag{5.5}$$

where $\mathrm{d}W_\pm/\mathrm{d}\eta$ is given by equation 2.114.

Equations 5.1 to 5.5 form a system of six-dimensional, coupled, nonlinear integro-differential equations and as a result, approximations are often required to solve them. For example, a study of cascades has been carried out by Sokolov *et al.* [174] and Bulanov *et al.* [177] using a one dimensional plane wave approximation. An alternative approach was suggested by Neitz and Di Piazza [186], where a Fokker-Planck expansion is performed in the limit of the emitted photon energy being much smaller than the electron energy. Niel *et al.* [187] have shown that this approximation is appropriate for $\eta \lesssim 0.25$. However, current laser facilities are capable of operating in regimes of $\eta > 0.25$, if a tightly focused laser is used. Under such conditions neither the one dimensional plane wave or Fokker-Planck approximations are valid. The only option for solving the set of equations is a Monte Carlo treatment. In the following section we will review this Monte Carlo treatment, however, further discussion can be found in refs [175, 120, 182].

[1] In reality, the photons will be emitted into a cone of angle $1/\gamma$, however, for ultra-relativistic particles this is negligible.

5.1.2 Monte Carlo Algorithm

In the Monte Carlo approach, rather than solving equation 5.1 directly, the dynamics of particles sampled from $f_\pm$ and f_γ at time $t = 0$ are solved. Here, we will refer to the sampling and simulation of a seed particle as an event. This event consists of the trajectory of the seed particle and all secondary particles that are created. As we are not considering the action of the particles on the background field, these events can be treated as independent.

In our implementation, the flow of each event is controlled by five base classes. These are: an event generator, particle pusher, field manager, process manager and output manager. The event generator is responsible for sampling the initial phase space coordinates, $\mathbf{x} = (\mathbf{r}, \mathbf{p})$, from $f_\pm$ or f_γ. The particle pusher, field manager and process manager are then invoked at each timestep. Finally the output manager, which is responsible for handling histograms, is called at the end of the event. The particle pusher, field manager and process manager are more involved so will be discuss in detail in this section. In addition to this, we will also discuss how the package integrates with Geant4.

Particle Pusher

Once the initial state of the particle has been sampled, the particle pusher is invoked to solve the equation of motion. If the particle is a photon, it will always follow a simple geodesic. In this case, a first order Euler method is sufficient for solving the equation of motion.

The package offers the choice of two options for the electron and positron equation of motion. If we are interested in the stochastic behaviour of particles, radiation reaction is handled by the process manager and the particle pusher solves the Lorentz force

$$m\dot{\mathbf{v}} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{5.6}$$

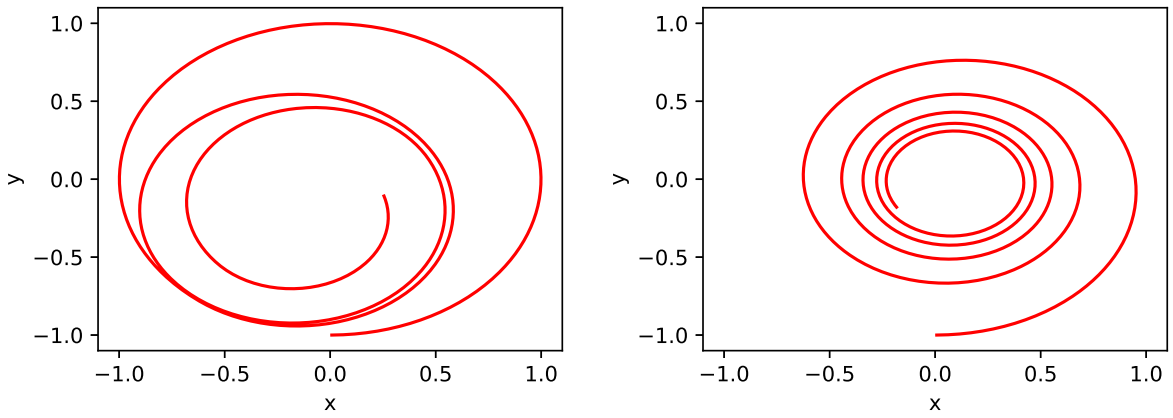
If we are not interested in the stochastic behaviour but the average motion of an electron or positron, the particle pusher solves the quantum corrected Landau-Lifshitz equation of motion (see equation 2.113), which includes the radiation reaction force explicitly

$$m\dot{\mathbf{v}} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - \frac{2}{3}\alpha\eta^2 g(\eta)\hat{\mathbf{v}}. \tag{5.7}$$

where $g(\eta)$ is the Gaunt factor, given by equation 2.112. To avoid performing the integration required to obtain $g(\eta)$, we use the following approximation [188]

$$g(\eta) \approx [1 + 4.8(1 + \eta)\ln(1 + 1.7\eta) + 2.44\eta^2]^{-2/3}. \tag{5.8}$$

Both the Lorentz and quantum corrected Landau-Lifshitz equations of motion are solved



(a) Lorentz

(b) Corrected Landau-Lifshitz

Figure 5.2: Trajectory of an electron in a static magnetic field. In (a) the Lorentz equation of motion has been solved and the radiation reaction force, handheld by the process manager. In (b) the corrected Landau-Lifshitz equation of motion has been solved.

using a 4th order Runge Kutta method.

Figure 5.2 compares the trajectories of electrons in the same static magnetic field but with the two different solvers. In (a) the Lorentz force is solved and radiation reaction is handled by the process manager. The electron exhibits an erratic motion as the radius of curvature changes discretely. However, in (b) the corrected Landau-Lifshitz force is solved and the electron exhibits a smooth path inwards.

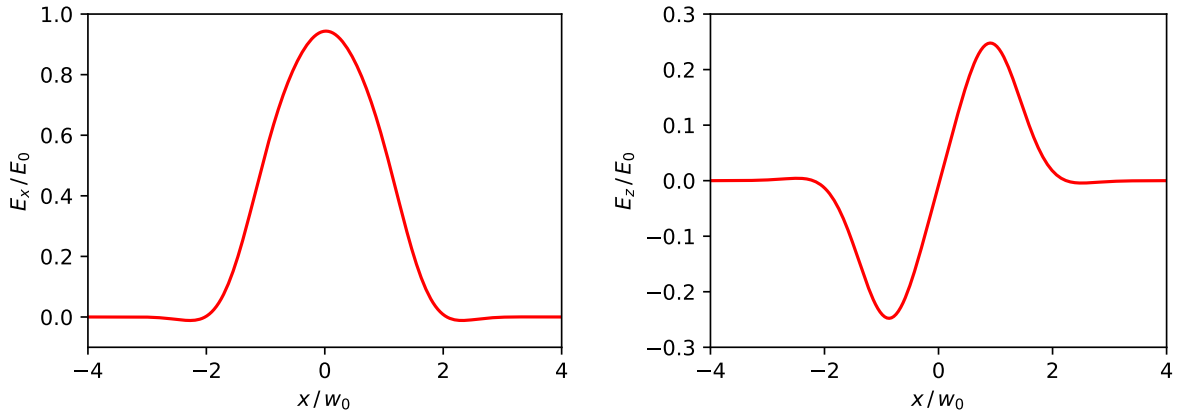
Field Manager

To solve equation 5.6 or 5.7 the electric and magnetic fields are required. In a PIC code these are obtained self consistently by solving Maxwell's equations. However, here we assume the action of the particles has little effect on the background field. Therefore, to reduce computational costs, the fields are provided by a predefined function. The function takes in a position and time coordinate and returns the electric and magnetic field vectors, $\mathbf{E}, \mathbf{B} = f(\mathbf{x}, t)$. We have implemented three methods for defining $f(\mathbf{x}, t)$: a static homogeneous field, a plane wave, and a focusing Gaussian beam. The static homogeneous field simply returns the same field vectors for all position and time coordinate, *i.e.* $\mathbf{E}_0, \mathbf{B}_0 = \mathbf{f}(\mathbf{x}, t)$ where $\mathbf{E}_0, \mathbf{B}_0$ are predefined vectors.

The other two fields are used to model a laser pulse. Therefore, we separate $f(\mathbf{x}, t)$ into an oscillatory function $\mathbf{g}(\mathbf{x}, t)$ and a temporal envelope function, $h(\mathbf{x}, t)$. The envelope, which is defined as a function of the wave phase, $\phi = \mathbf{k} \cdot \mathbf{x} - \omega_0 t$, is given by a Gaussian wave packet

$$h(\phi) = \exp\left(- \frac{\phi^2}{4\omega_0^2 \tau^2} \right). \tag{5.9}$$

Here, τ is the time for the intensity to fall by e^{-1}. For the plane wave, the oscillatory



(a) Transverse field

(b) Longitudinal field

Figure 5.3: Normalised electron field components for Gaussian pulse with $\epsilon = 0.8$. (a) shows the transverse field whereas (b) shows the longitudinal field

function is defined as a simple sinusoidal wave

$$\mathbf{g}(\mathbf{x}, t)_E = \mathbf{E}_0 \sin(\phi) \qquad \mathbf{g}(\mathbf{x}, t)_B = \mathbf{B}_0 \sin(\phi) \tag{5.10}$$

where $\mathbf{B}_0 = -\mathbf{k} \times \mathbf{E}_0/\omega_0$. This function is fast to compute, however, does not include any longitudinal fields that occur when a pulse is tightly focused. Such effects are included by the focusing Gaussian beam. In this case, the fields are given by the solution to the Helmholtz equation resulting from the propagation of an electromagnetic wave in a vacuum. Here, an expansion in powers of the diffraction angle ϵ is used, where terminating at first order corresponds to the paraxial approximation. The diffraction angle is given by $\epsilon = w_0/Z_r$ where w_0 is the waist of the beam at focus and $Z_r = |\mathbf{k}|w_0^2/2$ is the Rayleigh length. Here, we use the field obtained by expanding to fourth order, however, an expansion up to the 11th order can be found in ref. [189]. A plot of the transverse and longitudinal electric fields for $\epsilon = 0.8$ at the focusing plane is shown in figure 5.3.

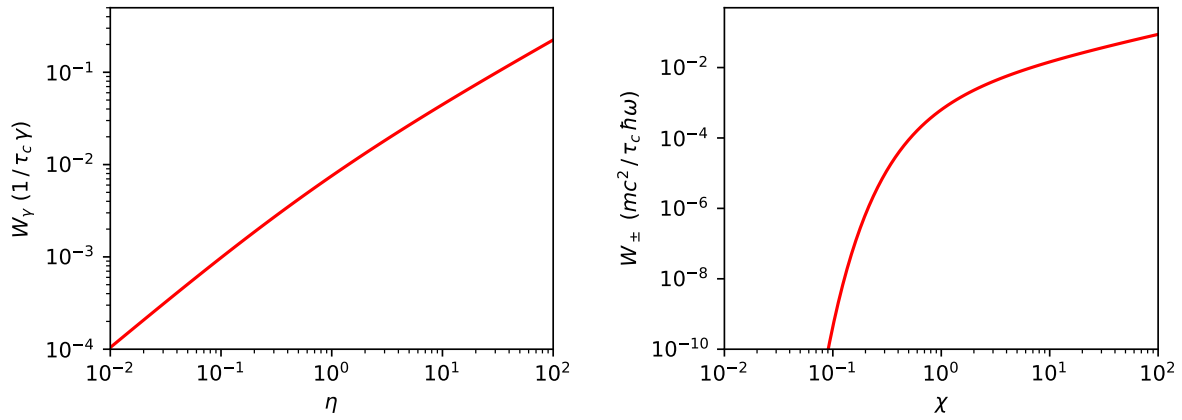
Process Manager

There are two roles performed by the process manager: calculating the distance a particle travels before undergoing an interaction; and calculating the properties of the particles which leave the interaction. To perform the first task, it is useful to introduce the optical depth of a particle

$$\tau(t) = \int_0^t W \mathrm{d}t' \tag{5.11}$$

where W is the rate of a QED interaction, given by

$$W_\gamma = \frac{\sqrt{3}\alpha}{2\pi\tau_c}\frac{\eta}{\gamma}\int_0^{\eta/2} \mathrm{d}\chi\, F(\eta, \chi) \qquad W_\pm = \frac{\alpha}{\tau_c}\frac{mc^2}{\hbar\omega}\chi T(\chi) \tag{5.12}$$



(a) Nonlinear Compton scattering

(b) Nonlinear Breit-Wheeler process

Figure 5.4: Rates for QED processes. (a) shows the nonlinear Compton scattering rate whereas (b) shows the nonlinear Breit-Wheeler process rate

for nonlinear Compton scattering and the nonlinear Breit-Wheeler process respectively, where $F(\eta, \chi)$ is given by equation 2.108 and $T(\chi)$ is given by equation 2.117. Figure 5.4 show plots of W_γ and $W_\pm$. The optical depth can be viewed as the number of mean free paths that the particle has travelled. Therefore, the probability of the interaction occurring at an optical depth of τ is $p(\tau) = e^{-\tau}$.

At the start of each event, and when a new particle is generated, a final optical depth, τ_f is sampled from $p(\tau)$. This uses inverse transform sampling (see section 3.3.1), *i.e.*

$$\tau_f = -\log(1 - u) \tag{5.13}$$

where $u \sim \mathcal{U}(0, 1)$. As the particle travels through the simulation, its current optical depth, τ, is obtained by solving equation 5.11 using a first order Euler method. Once $\tau = \tau_f$, the particle undergoes the interaction and τ is reset.

Having determined that an interaction has occurred, the process manager then moves on to calculate the properties of the outgoing particles. For nonlinear Compton scattering, this involves adding a photon to the event and updating the momentum of the electron or positron, whereas for the nonlinear Breit-Wheeler process the interacting photon is removed from the event and an electron-positron pair is added. Before adding a new particle to the event, its momentum must be calculated. The quantum parameter of the new particle can be obtained by sampling from the process differential rate. This is again carried out using inverse transform sampling. The CDF of the differential rates are given by

$$C(\chi) = \frac{\int_0^{\chi} (\mathrm{d}W_\gamma/\mathrm{d}\chi')\mathrm{d}\chi'}{\int_0^{\eta/2} (\mathrm{d}W_\gamma/\mathrm{d}\chi')\mathrm{d}\chi'} \qquad C(\eta) = \frac{\int_0^{\eta} (\mathrm{d}W_\pm/\mathrm{d}\eta')\mathrm{d}\eta'}{\int_0^{2\chi} (\mathrm{d}W_\pm/\mathrm{d}\eta')\mathrm{d}\eta'} \tag{5.14}$$

for nonlinear Compton scattering and the nonlinear Breit-Wheeler process respectively.

These CDFs cannot be inverted analytically, so it is performed numerically and tabulated. Values for η and χ are interpolated from this table during run-time. If we assume the outgoing particles are emitted parallel to the motion of the incoming particle, we can use equation 2.93 to calculate the outgoing particles' momentum. With this, the new particles are then added to the event and the simulation continues.

Geant4 Event Generator

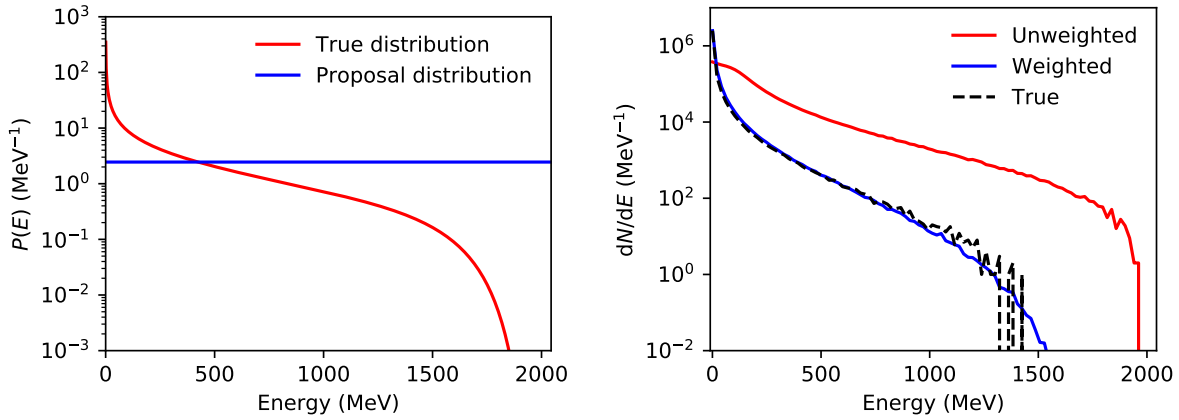
In chapter 4 we discussed one way that particle-photon interactions can be modelled within Geant4. To do so requires the photon source to be modelled as a static field due to the static nature of the Geant4 computational domain. This limits the types of interaction that can be modelled to those in which the two sources vary on vastly different timescales. This is not the case for a strong field QED experiment at a current laser facility in which both sources vary on femtosecond timescales.

To solve this problem, we have used an alternative approach to that used in chapter 4 and developed a strong field QED primary event generator for Geant4. Here, the event is started by the strong field QED algorithm discussed above. Once it has finished, the output is passed to Geant4 as a seed event. Through this scheme, the two pieces of software can be coupled without additional approximations.

5.1.3 Importance Sampling

We can see from figure 5.4 (b), that when $\chi < 1$ the pair production rate is severely suppressed. During an experiment, operating in this regime, we can only expect a few nonlinear Breit-Wheeler interactions to occur per shot. For example, during the E144 experiment a positron was detected approximately once every 200 shots [56]. Using the Monte Carlo algorithm discussed above to make an estimate of the pair production yield of such an experiment, would require simulating a large number of events. Therefore, to reduce the computational costs, we have adapted the algorithm to include importance sampling (see section 3.3.2).

To increase the number of nonlinear Breit-Wheeler interactions that occur in a simulation, we need to increase the number of high energy photons. We can do this by applying importance sampling to the nonlinear Compton scattering emission. This is achieved by sampling the photon's energy from a proposal distribution, $q(E)$, rather than the true distribution, $p(E)$. The proposal distribution is chosen such that the probability of sampling a higher energy photon is increased. This is demonstrated in figure 5.5 (a), where we have chosen a uniform distribution for $q(E)$. This leads to an increase in the number of high energy photons, however, if we also update the electron's momentum using the proposal distribution, it will follow the wrong trajectory. Therefore, we also sample from the true distribution to update the electron's momentum.



(a) Sampled distributions

(b) Total pair emissivity

Figure 5.5: Plots of photon energy distribution emitted by a 2 GeV electron beam travelling through a 5.5×10^5 T magnetic field. (a) shows the true (red) and proposal (blue) distributions that are sampled from during the electron's first emission. (b) shows the final photon energy distribution obtained by sampling from: the proposal distribution (red); the proposal distribution then weighted (blue); and the true distribution (black dash).

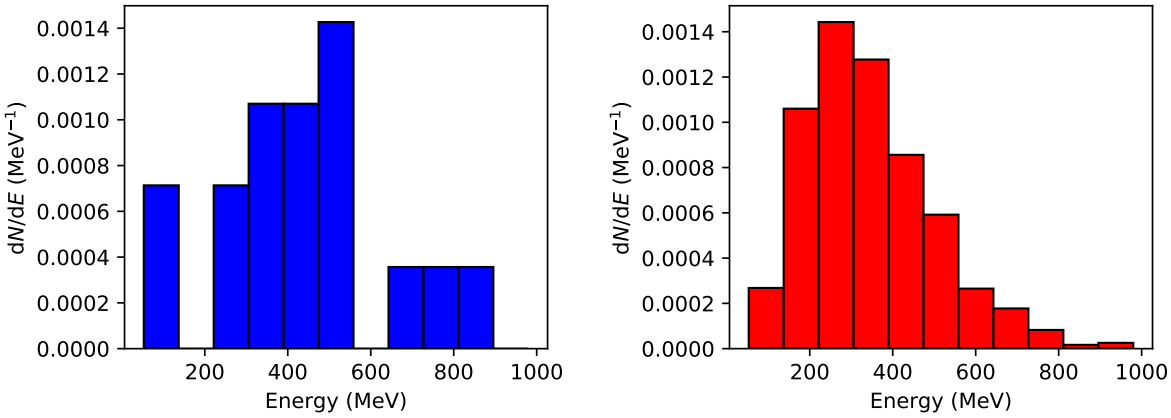
In figure 5.5 (b) we can see the effect the proposal distribution has on the photon spectrum emitted by a beam of electrons with energy 2 GeV, travelling through a magnetic field of strength 5.5×10^5 T, such that $\eta \approx 0.5$. The energy spectrum obtained by sampling from the proposal distribution is not uniform due to the electrons undergoing multiple emission events which reduces their energy. However, we can see there are still far more high energy photons compared to sampling from the true distribution.

Making an estimate based on the proposal distribution alone would give a biased result. This is corrected by assigning a weight to the emitted photons, given by $w = p(E)/q(E)$. Figure 5.5 (b) also shows that accounting for this weight, the spectrum is the same as the that obtained by sampling from the true distribution. These weighted photons are treated in the same way as unweighted photon. However, if the photon decays, the secondary particles will inherit this weight.

The benefit of applying importance sampling on the positron energy spectrum, using the same simulation setup as above, is shown in figure 5.6. Both simulations used the same number of seed particles, however, a more highly resolved estimate is obtained using importance sampling.

5.2 Test Simulations

In this section we will present some benchmark simulations which test the implementation of the algorithm discussed in section 5.1. This will involve both testing the numerical convergence of the algorithm as well as comparisons against analytical theory.



(a) Without importance sampling

(b) With importance sampling

Figure 5.6: Positron energy spectrum obtained from a simulation of 10000 seed electrons with an energy of 2 GeV, travelling through a 5.5×10^5 T magnetic field. In (a) importance sampling has not been used, whereas in (b) importance sampling has been used.

5.2.1 Numerical Convergence

The algorithm discussed in section 5.1 allows particles to move freely in space, but they are discretised in time. Therefore, the time step, Δt, is the only parameter relevant to convergence. In this algorithm, the same value of Δt is used to solve both the particle's equation of motion and probability of emission. However, we would expect these to have different characteristic timescales. For the algorithm to produce reliable results, Δt must be small enough to resolve the shortest timescale dynamics of the two solvers.

For an electron or positron in a static magnetic field, the gyro-period is the shortest timescale that the particle pusher must resolve. Therefore, we will define the particle pusher constraint as

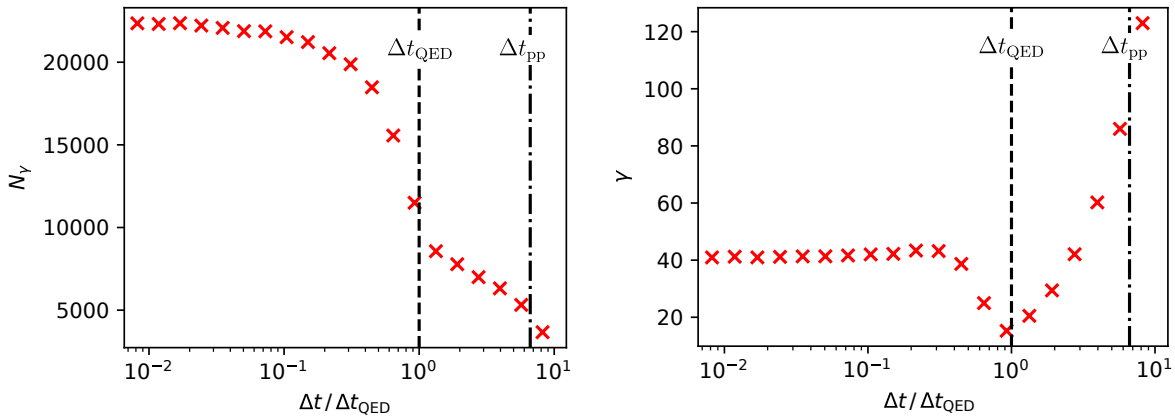
$$\Delta t_{\text{pp}} = \frac{2\pi\gamma m}{e|\mathbf{B}|}. \tag{5.15}$$

As the process manager can only interact once per timestep, the constraint should be set such that the probability of multiple events is low. For an electron or positron in an static magnetic field, we can estimate the probability of N emission events occurring in a time Δt by assuming the emission is a Poisson process[2]

$$p(N) = \frac{(W_\gamma \Delta t)^N e^{-W_\gamma \Delta t}}{N!}. \tag{5.16}$$

To suppress the probability of multiple emissions per timestep, we require that $W_\gamma \Delta t \ll 1$.

[2]This is not strictly true as the reduction in energy of the electron from the emission of the photon will change η and the rate $W_\gamma(\eta)$ is not constant.



(a) Emitted photons

(b) Average γ

Figure 5.7: Effect of increasing the time-step on (a) the number of photons emitted and (b) the average final energy of an electron.

Therefore we can define the process manager constraint as [182]

$$\Delta t_{\mathrm{QED}} = \frac{1}{\max W_\gamma} = \frac{2\pi\,\tau_c}{\sqrt{3}\,\alpha h_0}\frac{B_{\mathrm{crit}}}{|\mathbf{B}|} \tag{5.17}$$

where $h_0 = h(0) = 5.24$ and $B_{\mathrm{crit}} = E_{\mathrm{crit}}/c$ is the Schwinger magnetic field. A similar calculation can be performed for the nonlinear Breit-Wheeler process, however, it will always be larger than the photon emission [182]. The ratio of the two time-step constraints is

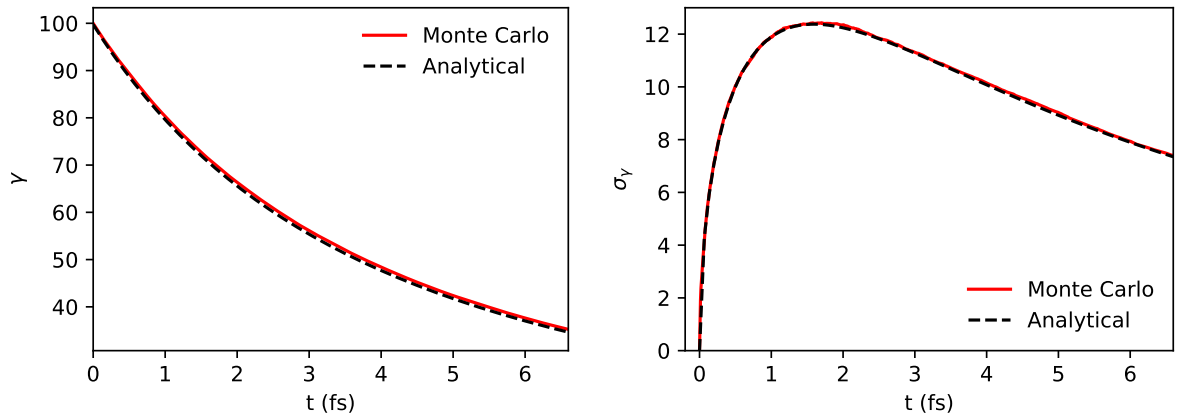
$$\frac{\Delta t_{\mathrm{pp}}}{\Delta t_{\mathrm{QED}}} = \sqrt{3}\alpha h_0 \gamma. \tag{5.18}$$

Therefore, if $\gamma \gtrsim 15$, which is usually the case in a strong field QED experiment, $\Delta t_{\mathrm{QED}} < \Delta t_{\mathrm{pp}}$.

Figure 5.7 shows the effect of varying Δt on a simulation, consisting of a beam of 1000 electrons with $\gamma = 100$, travelling perpendicular to a static magnetic field with $|\mathbf{B}| = 1 \times 10^{-3} B_{\mathrm{crit}}$ for 5 fs. In (a) we can see the total number of photons emitted, which decreases dramatically around $\Delta t_{\mathrm{QED}} = 1$. In (b) we can see the average γ of the electrons at the end of the simulation. Initially, increasing Δt, leads to a slight increase in γ, as the electrons are emitting fewer photons. However, as Δt is increased further the particle pusher is no longer able to resolve the Larmor orbit and introduces large errors.

5.2.2 Analytical Benchmark

With $\Delta t \ll \Delta t_{\mathrm{QED}}$, we will now turn our attention to comparing results from the Monte Carlo algorithm with analytical calculations. This will involve performing separate tests of the nonlinear Compton scattering and nonlinear Breit-Wheeler modules. To simplify the analytical calculations, in both cases the test simulation will consist of a beam of



(a) γ mean

(b) γ s.d.

Figure 5.8: Nonlinear Compton scattering benchmark simulations. (a) shows the average γ of an electron beam travelling through a static magnetic field against time. (b) shows the standard deviation of the same electron beam against time

10000 particles travelling through a static magnetic field of strength $|\mathbf{B}| = 1 \times 10^{-3} B_{\text{crit}}$.

Nonlinear Compton Scattering

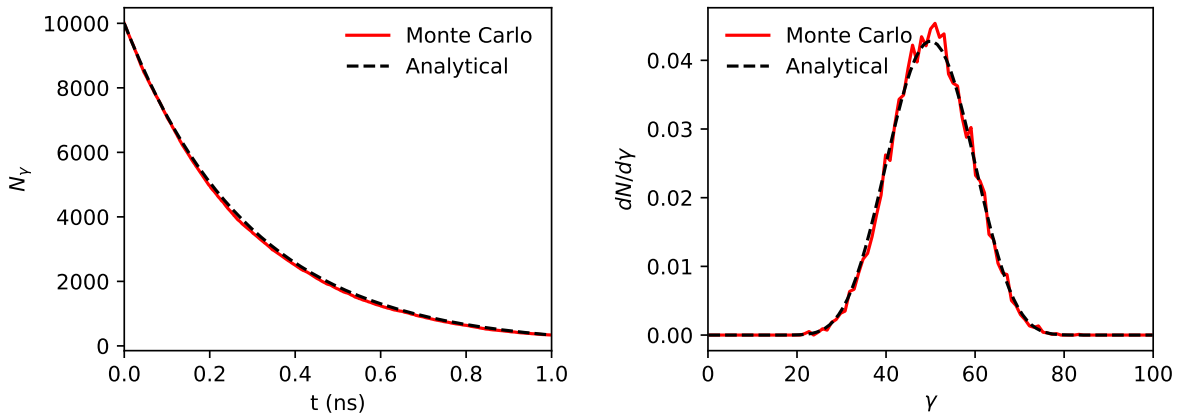
Even with this relatively simple field configuration, equation 5.1 cannot be analytically solved. However, it is possible to analytically calculate the moments of the distribution function. The first moment is the average evolution of the particles' momentum and is given by the corrected Landau-Lifshiz equation of motion (see equation 2.113). The second moment is the evolution of the variance of the particles' momenta. If we assume the momentum distribution is Gaussian at all times with $\sigma \ll \langle\gamma\rangle$, in the limit $\eta \ll 1$, Ridgers *et al.* [188] have shown the variance evolves according to

$$\left(\frac{\mathrm{d}\sigma^2}{\mathrm{d}t}\right) = -2\frac{\langle \Delta\gamma g P_{\text{cl}}\rangle}{m_e c^2} + \frac{\langle S\rangle}{m_e^2 c^4} \tag{5.19}$$

where P_{cl} is the classical radiated power given by equation 2.34 , $\Delta\gamma = \gamma - \langle\gamma\rangle$, and

$$s(\eta) = \frac{55\alpha\gamma c}{24\sqrt{3}\tau_c} m_e^2 c^4 \eta^3 g_2(\eta) \qquad g_2(\eta) = \frac{144}{55\pi\eta^4}\int_0^{\eta/2} \chi F(\eta,\chi)\mathrm{d}\chi. \tag{5.20}$$

Figure 5.8 (a) and (b) show how the first and second moments of an electron beam's distribution function vary with time, starting initially as a delta function with $\gamma = 100$. These calculations have been performed both analytically and using the Monte Carlo estimate and shows good agreement between the methods.



(a) Differential pair emissivity

(b) Total pair emissivity

Figure 5.9: Nonlinear Breit-Wheeler benchmark simulations. (a) shows the number of photons in a beam travelling through a static magnetic field against time. (b) shows the energy distribution of the electrons (or positrons) produced through the nonlinear Breit-Wheeler.

Nonlinear Breit-Wheeler Process

Testing the nonlinear Breit-Wheeler module is a simpler task as the momentum of a photon is not affected by a background field unless it annihilates. The decay of a beam of photons in a static magnetic field is a Poisson process, and the average number of photons left after a time t is $N = N_0 \exp(-t/W_\pm)$, where N_0 is the initial number of photons. This is plotted, along with an estimate from the Monte Carlo algorithm, in figure 5.9 (a) for a beam of photons with energy $E_\gamma = 100\, m_e c^2$.

It is also possible to analytically calculate the energy distribution of the emitted electrons and positrons. Defined in terms of the fraction of energy assigned to the electron (or positron), f, the distribution is given by [37]

$$p(f \mid \chi) = \frac{2 + f(1-f)}{Af(1-f)} K_{2/3}\left[\frac{1}{3\chi f(1-f)}\right] \tag{5.21}$$

where A is a normalisation constant. This is plotted in figure 5.9 (b) along with an estimate from the Monte Carlo algorithm for the same photon beam used above. Once again, a good agreement is found between the analytical and Monte Carlo results.

5.3 Emulating the Monte Carlo Algorithm

Aside from performing start-to-end simulations of experiments, a second application of this package is as a forward model for post experimental, statistical inference. There are two commonly used approaches for statistical inference, maximum likelihood estimation

(MLE) and Bayesian inference (BI). Both these methods require calculating a forward model many times. However, the Monte Carlo algorithm discussed in section 5.1.2 is computationally expensive, making it impractical for this use. To solve this problem, we have developed a machine learning emulator, trained on a data set generated by the Monte Carlo algorithm, which takes far less time to calculate.

In this section we will discuss the development of this emulator. We will start by examining the data set which the emulator is trained on, before discussing the type of emulator which we have used. Finally we will demonstrate the capability of the emulator by using it in BI of mock experimental data.

5.3.1 Emulator Data Set

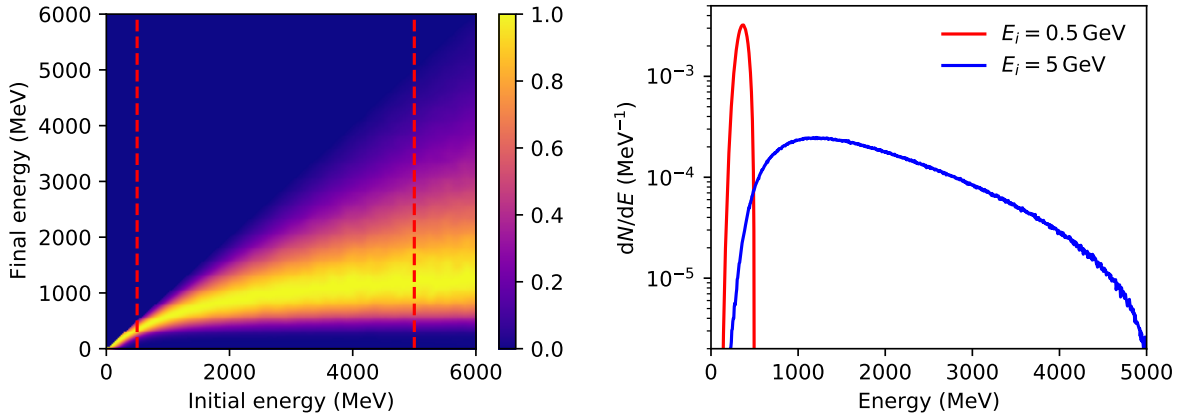
Before applying any machine learning method, it is important to study the data set to determine if the method is appropriate. Here, the data set consists of N input-output pairs $\mathcal{D} = \{\mathbf{x}_i, \mathbf{y}_i \mid i = 1, \ldots, N\}$ where $\mathbf{x}_i$ is the input vector of dimensions D_x and $\mathbf{y}_i$ is the output vector of dimensions D_y. The number of inputs to the Monte Carlo algorithm is larger, however, most can be categorised as either a particle parameter or a field parameter. To simplify the discussion, we will only consider the particle parameters as inputs and keep the field parameter as fixed at the values given in table 5.1. This is due to the current implementation of the Monte Carlo algorithm, making it easier to generate a training data set in which the particle parameters are varied. However, if the algorithm was adapted so that the field parameters could vary each event, then these could also be treated as input parameters.

If we are modelling a radiation reaction experiment, where the nonlinear Breit-Wheeler process rate is highly suppressed ($\chi < 1$), and focus only on the electrons, each event consists of a mapping of the electron's initial phase space coordinates $\mathbf{x}_i \equiv (\mathbf{r}_i, \mathbf{p}_i)$ to its final phase space coordinates $\mathbf{x}_f \equiv (\mathbf{r}_f, \mathbf{p}_f)$. When modelling an experiment, the Monte Carlo algorithm deals with interactions on a micrometre scale. However, the scale of the experiment is metres, so we can treat the particles as being emitted from a point source, ignoring the spatial component of the output, $\mathbf{x}_f = (\mathbf{p}_f)$. Therefore the goal of the machine learning algorithm is to learn the predictive, conditional distribution $p(\mathbf{x}_f \mid \mathbf{x}_i)$ where $D_x = 6$ and $D_y = 3$. In the following discussion, we will use an alternative coordinate system in which the z-component of the particle momentum is replaced with its energy: $\mathbf{x}_i = (\mathbf{r}_i, E_i, p_{ix}, p_{iy})$ and $\mathbf{x}_f = (E_f, p_{fx}, p_{fy})$.

To generate the training data set, we can uniformly sample values of $\mathbf{x}_i$ over a region of interest, then simulate the events using the Monte Carlo algorithm. A plot of the training data set that we will use is shown in figure 5.10 (a) where we have marginalised over all variables other than E_i and E_f. This demonstrates the probabilistic nature of the data set. Therefore, we may think a Bayesian approach, such as a Gaussian process regression,

Table 5.1: Field and particle parameters used in test simulations throughout chapter.

Field Parameters		Particle Parameters	
Field type	focusing Gaussian	Particle type	electron
a_0	30	Energy distribution	Gaussian
Duration	25 fs	Energy mean	2 GeV
Waist	5 μm	Energy s.d.	0.5 GeV
		Duration	30 fs
		Waist	1 μm
		Divergence	1 mrad



(a) Joint distribution

(b) Conditional distribution

Figure 5.10: Sample of the data set used to train the emulator. (a) shows the joint distribution $p(E_i, E_f)$ with red dashed lines corresponding to the conditional distributions $p(E_f \mid E_i)$ shown in (b).

would be a good candidate for the machine learning algorithm. However, implementing a Gaussian process in this case would be challenging. To examine why, we will consider the conditional distribution $p(E_f \mid E_i)$, shown in figure 5.10 (b), obtained by taking a line-out from figure 5.10 (a). The conditional distribution appears Gaussian at $E_i = 0.5\,\text{GeV}$, however, this is not the case at 5 GeV. Therefore, the data set is heteroscedastic, with a non-Gaussian likelihood. It is possible to apply a Gaussian process regression to non-Gaussian data [190], however, it is more computationally expensive, requiring either an MCMC or variational approach. There is a second issue as the training time for a Gaussian process scales poorly for larger data sets ($\mathcal{O}(N^3)$ where N is the size of the data set). Although the Monte Carlo algorithm is too slow to be used as a forward model in statistical inference, it is still possible to generate a larger data set[3]. For these two reasons, the cost of training a Gaussian process emulator would negate any benefits from increased run-time.

[3] $N \approx 10^7$ events can be simulated in $\sim$ 2 hours using a 24 core, Intel Xeon Processor E5-2680 v3 [191].

Instead, we have opted for a neural-network based machine learning method, as the training time scale is favourable ($\mathcal{O}(N)$) for large data sets. The aim of the emulator is to model a predictive distribution with heteroscedastic variance. Therefore, a standard feed-forward network (FFN) is not appropriate. To solve this problem, we have used a mixture density network (MDN), originally developed by Bishop [192]. In the next section, we will demonstrate the issues that arise when using a standard FFN and show how they are solved with a MDN.

5.3.2 Mixture Density Network Emulator

If we parameterise the Monte Carlo algorithm using an FFN, then the emulator function, $\mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta})$, is given by equation 3.21. To learn the network parameters, $\boldsymbol{\theta}$, we can use MLE, which involves defining a likelihood function, that is usually assumed to be Gaussian

$$p(\mathbf{x}_f \mid \mathbf{x}_i, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}_f \mid \mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta}), \boldsymbol{\Sigma}). \tag{5.22}$$

The result of applying a Gaussian likelihood and training a FFN on the data set is demonstrated in figure 5.11 (a). Here we have used a network architecture of two hidden layers with 500 nodes each, a rectified linear unit activation function, and an L2-norm regularisation term applied to the hidden layers. The network was trained to completion using cross-validation and $\boldsymbol{\Sigma}$ is given by the mean of the squared residuals. This network is capable of predicting the mean E_f for a given E_i, however, gives a poor prediction of the data variance. This is highlighted further in figure 5.11 (b) showing the conditional distribution $p(E_f \mid E_i)$ for $E_i = 0.5,\ 5.0\,\text{GeV}$.

The issue with this network is that it is incapable of handling the heteroscedastic variance. To solve this issue, we can assume that the variance is also a function of the input. This is achieved by adding extra nodes to the output of the network, returning both the mean function and the variance, *i.e.* $(\boldsymbol{\mu},\, \boldsymbol{\sigma}) = \mathbf{f}(\mathbf{x}_i, \boldsymbol{\theta})$. The likelihood is then given by

$$p(\mathbf{x}_f \mid \mathbf{x}_i, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}_f \mid \mu(\mathbf{x}_i, \boldsymbol{\theta}),\, \boldsymbol{\Sigma}(\mathbf{x}_i, \boldsymbol{\theta})) \tag{5.23}$$

and the result of applying this to our data set is shown in figure 5.11 (c). Here we have used the same network architecture as before. This model is more flexible and capable of modelling the changing variance. In figure 5.11 (d) we can see plots of the conditional distribution, $p(E_f \mid E_i)$. The network gives a good prediction of $p(E_f \mid E_i)$ at small values of E_i, however, at high values of E_i the prediction is poor. To understand why this is the case, we can study the Fokker-Planck treatment of radiation reaction provided by Niel *et al.* [187]. In a Fokker-Planck treatment, if the initial energy distribution is Gaussian, the final energy distribution will also be Gaussian with a shifted mean and increased variance. Niel *et al.* have shown that this treatment is appropriate in the intermediate quantum

regime, $\eta < 0.25$. This is the same regime under which we would expect $p(E_f \mid E_i)$ to be Gaussian. At larger values of η, higher order moments become important and we cannot use a single Gaussian to describe the distribution.

To account for the higher order moments, we need a more flexible model which can change the functional form of $p(E_f \mid E_i)$ with E_i. This can be achieved using a mixture density network (MDN). Here, we assume the likelihood function is a mixture model, given by the weighted sum of K components. The weighting coefficients, π_k, mean functions, $\boldsymbol{\mu}_k$ and covariances $\boldsymbol{\Sigma}_k$ are all assumed to be functions of the input and are provided by the output of the neural-network, $(\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \mathbf{f}_k(\mathbf{x}_i, \boldsymbol{\theta})$. We are free to choose any set of basis functions for the components, however, the Gaussian family are a common choice as they form a complete set, allowing any distribution to be created. The likelihood is now given by

$$p(\mathbf{x}_i \mid \mathbf{x}_f, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k(\mathbf{x}_i, \boldsymbol{\theta})\, \mathcal{N}(\mathbf{x}_f \mid \boldsymbol{\mu}_k(\mathbf{x}_i, \boldsymbol{\theta}),\, \boldsymbol{\Sigma}_k(\mathbf{x}_i, \boldsymbol{\theta})). \tag{5.24}$$

and the result of fitting a MDN with $K = 10$, using the same network architecture as before, is shown in figure 5.11 (e). This looks similar to the single component network (figure (c)), however, looking at the conditional distribution in figure 5.11 (f) we can see the MDN gives a far better prediction at larger values of E_i.

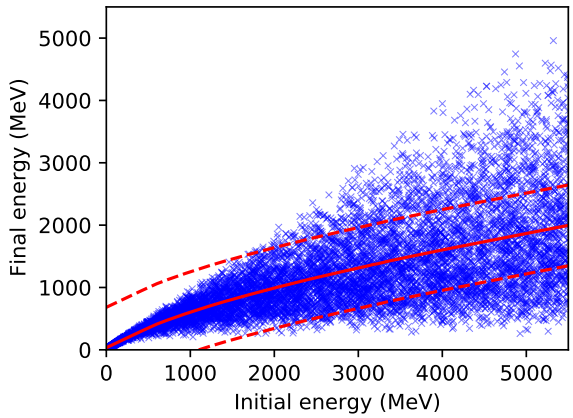
Emulating an Experiment

The MDN emulator provides the ability to quickly calculate the conditional distribution, $p(\mathbf{x}_f|\mathbf{x}_i)$. When modelling an experiment we are usually interested in the marginal output distribution

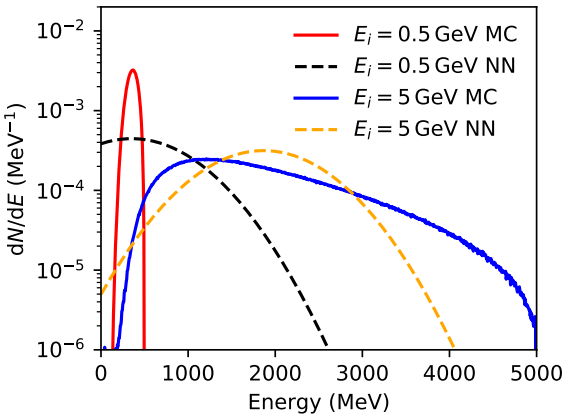
$$p(\mathbf{x}_f) = \int p(\mathbf{x}_f \mid \mathbf{x}_i) p(\mathbf{x}_i) \mathrm{d}\mathbf{x}_i. \tag{5.25}$$

This is a six dimensional integral, so a Monte Carlo estimate is used. We may now wonder if the MDN offers any computational saving over the Monte Carlo algorithm as a Monte Carlo estimate is required to make a useful calculation. However, far fewer samples are required by the MDN compared to the Monte Carlo algorithm.

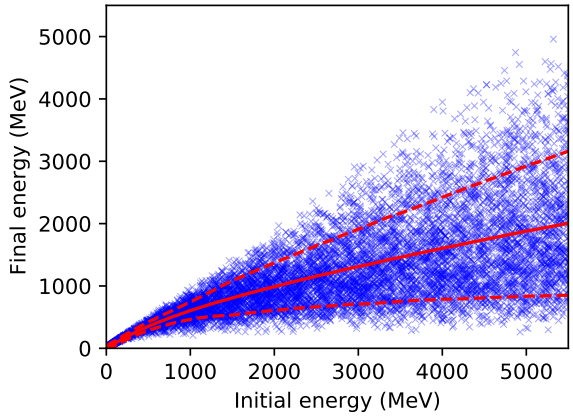
To demonstrate this, we will consider a test simulation consisting of an electron beam interacting with a focusing Gaussian pulse. The simulation parameters are give in table 5.1. Figure 5.12 (a) shows the initial and final energy distribution, obtained by simulating $N_e = 10^6$ events. We will take the final energy distribution to be the ground truth and use it as a convergence test for the two methods. To perform this test, we require a metric that gives a measure of how close a distribution is to the ground truth. Here we will use the Kolmogorov–Smirnov (KS) statistic. Given two probability distributions $p(x)$ and



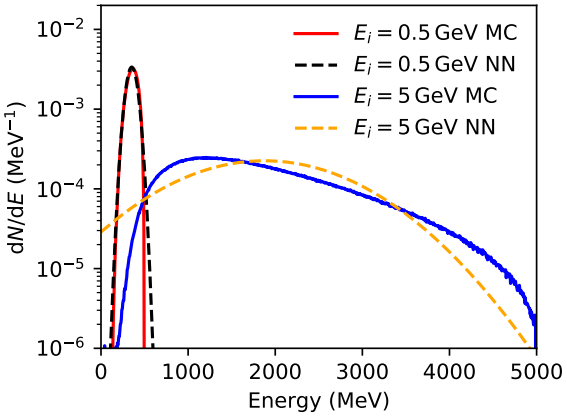
(a) FFN joint distribution



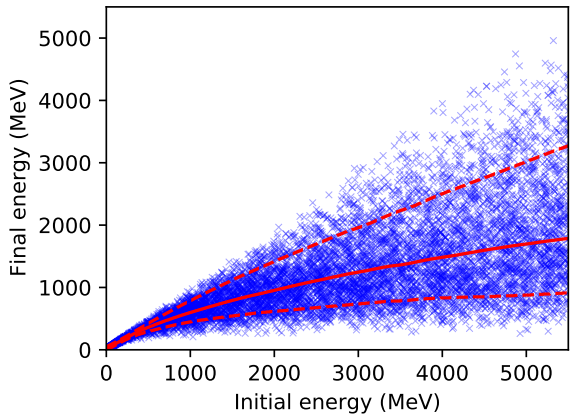
(b) FFN conditional distribution



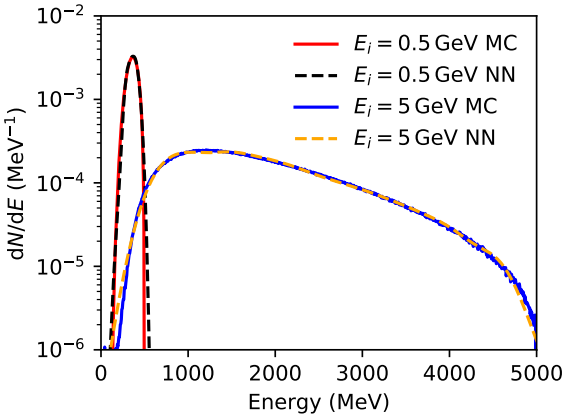
(c) 1 component joint distribution



(d) 1 component conditional distribution

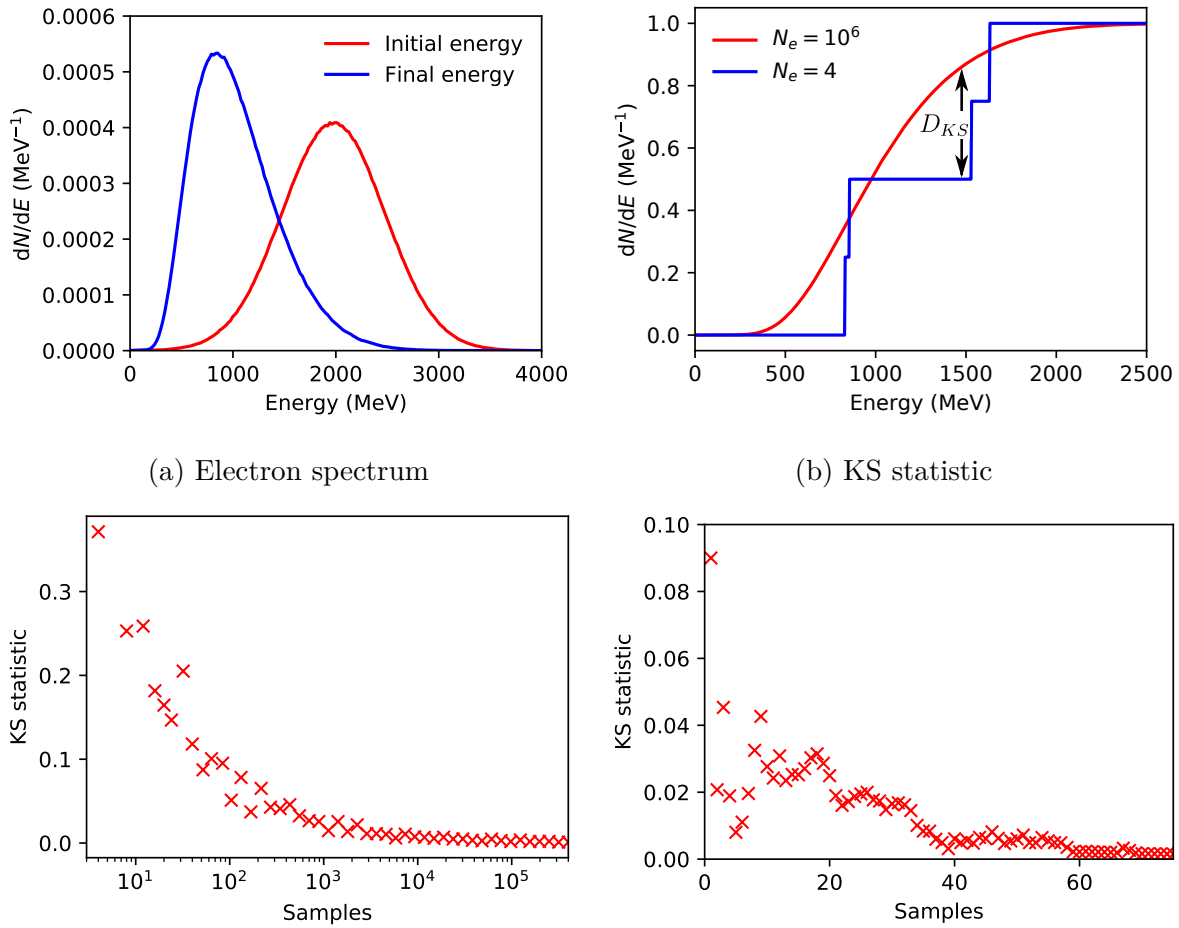


(e) 10 components joint distribution



(f) 10 components conditional distribution

Figure 5.11: Neural network models fitted to a strong field QED data set. (a), (c) and (e) shows the joint distribution $p(E_i, E_f)$ with a mean fit (sold red line) and two-sigma quantile (red dashed lines). (b), (d) and (f) shows the conditional distribution $p(E_f \mid E_i)$ at $E_i = 0.5,\ 5.0\,\text{GeV}$ for both the MDN and Monte Carlo algorithm.



(a) Electron spectrum

(b) KS statistic

(c) Monte Carlo algorithm

(d) MDN

Figure 5.12: Sample convergence tests for both the MDN and Monte Carlo algorithm. (a) shows the initial and final electron energy distribution from a highly resolved simulation. (b) shows an example of the KS statistic for a poorly resolved simulation. (c) shows the convergence of the convergence of the Monte Carlo algorithm estimate. (d) shows the convergence of the MDN estimate.

$q(x)$, the KS statistic is the maximum distance between the CDFs

$$D_{KS} = \sup|P(x) - Q(x)| \tag{5.26}$$

where $P(x)$ and $Q(x)$ are the CDFs of $p(x)$ and $q(x)$ respectively. This is demonstrated in figure 5.12 (b) showing the KS statistic for a Monte Carlo simulation involving only 4 particles.

The effect of varying the number of samples on the Monte Carlo algorithm and MDN estimates are shown in figures 5.12 (c) and (d) respectively. We can see the MDN converges far faster than the Monte Carlo algorithm, requiring a factor of $\sim 10^3$ fewer samples. This is due to each sample of the Monte Carlo algorithm being a point estimate, whereas samples from the MDN provide an estimate of the conditional distribution. Directly

comparing the computational efficiency of the two methods is difficult as we use a graphics processing unit (GPU) for the MDN. However, performing an average over 60 samples using the MDN emulator takes $\approx$ 35 ms on a single NVIDIA TITAN Xp GPU [193], whereas simulating 10^5 particles with the Monte Carlo algorithm takes $\approx$ 72 s on a 24 core, Intel Xeon Processor E5-2680 v3 [191].

Multiple Output Dimensions

So far we have only demonstrated the MDN's ability to emulate a single output dimension. However, equation 5.24 uses a mixture of multivariate Gaussians, enabling the network to emulate the full three dimensional probability distribution. This is demonstrated in figure 5.13 (a), showing a surface plot of the joint distribution $p(E, p_x)$, generated by the MDN with input parameters given by table 5.1. Here we have marginalised over p_y as this is perpendicular to the polarisation of the wave and not affected during the interaction. Shown along side in figure 5.13 (b) is an estimate made by the Monte Carlo algorithm. We can see a good agreement between the two estimates which is highlighted further in figure 5.13 (c) and (d) showing the marginalised distributions $P(E_f)$ and $p(p_{xf})$ respectively.

5.3.3 Bayesian Inference of a Radiation Reaction Experiment

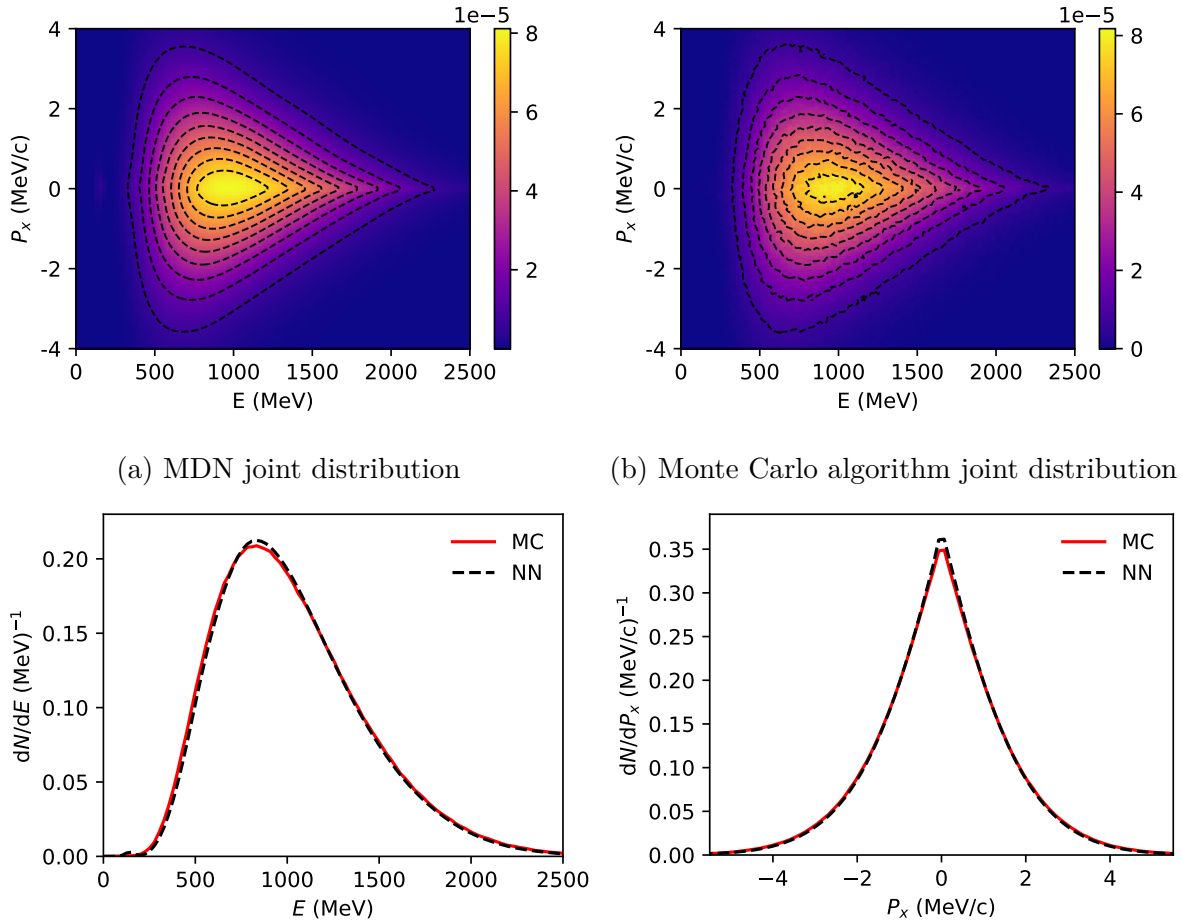
The development of a machine learning emulator, which takes a fraction of the time to compute, has applications in both experimental optimisation and post experimental, statistical inference. In this section we will demonstrate this ability by carrying out statistical inference on a numerical mock experiment.

Statistical inference is an invaluable tool which can be used to estimate unknown experimental parameters, $\mathbf{x}$, and to discriminate between different theoretical models, $\{\mathcal{H}_i\}$. During an experiment, we make a set of observations, $\mathcal{D} = \{y_i | i = 1, \ldots, N\}$, where N is the number of observations made. To perform inference on this data set, requires a forward model of the experiment $y = f(\mathbf{x}, \mathcal{H}_i)$, usually provided by a simulation. The optimum values of $\mathbf{x}$ and $\mathcal{H}_i$ can be inferred by varying the inputs to the forward model until the outputs match the experimental observations. This is equivalent to a maximum likelihood estimate of the parameters.

As discussed in section 3.1.1, issues arise when applying MLE due to obtaining a local maximum. A solution to this is Bayesian inference (BI), where an entire probability distribution over the unknown parameters, $p(\mathbf{x}, \mathcal{H}_i | \mathbf{y})$, is inferred. To carry out BI, Bayes' theorem is applied

$$p(\mathbf{x}, \mathcal{H}_i \mid \mathbf{y}) = \frac{p(y \mid \mathbf{x}, \mathcal{H}_i) p(\mathbf{x}) p(\mathcal{H}_i)}{\sum_{i=1}^{h} \int p(y \mid \mathbf{x}, \mathcal{H}_i) p(\mathbf{x}) p(\mathcal{H}_i) \mathrm{d}\mathbf{x}} \tag{5.27}$$

where h is the number of theories that we are comparing, $p(\mathbf{x})$ and $p(\mathcal{H}_i)$ are priors



(a) MDN joint distribution

(b) Monte Carlo algorithm joint distribution

(c) Marginalised energy distribution

(d) Marginalised x-momentum distribution

Figure 5.13: Multi-dimensional comparison between the MDN and Monte Carlo algorithm. (a) and (b) show the joint energy and x-momentum distributions for the MDN and Monte Carlo algorithm respectively. (c) shows the marginalised energy distributions and (d) shows the marginalised x-momentum distributions.

and $p(\mathbf{y} \mid \mathbf{x}, \mathcal{H}_i)$ is the likelihood. If we make the common assumption that our observations exhibit Gaussian noise with variance σ^2, then the likelihood function is given by $p(y \mid \mathbf{x}, \mathcal{H}_i) = \mathcal{N}(y \mid f(\mathbf{x}, \mathcal{H}_i), \sigma^2)$. Therefore, performing the integral in equation 5.27 requires $f(\mathbf{x}, \mathcal{H}_i)$ to be calculated many times. The Monte Carlo algorithm discussed in section 5.1.2 is too slow to be practically used as a forward model. However, the MDN emulator can be carried out many times per second, making it an ideal tool for performing BI.

Numerical Experiment

A radiation reaction experiment involves overlapping a tightly focused laser pulse with an electron beam to micron precision. Due to the large shot-to-shot jitter in such an experiment there are many unknown interaction parameters. This includes the temporal

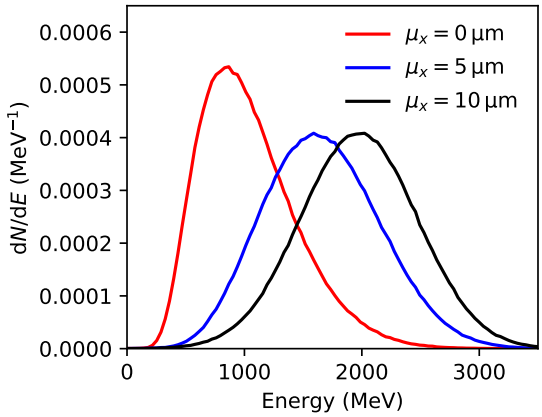


Figure 5.14: Electron energy spectrum after interaction with $\mu_y = 0$ different values of μ_x.

and transverse overlap of the two beams, and the initial electron beam spectrum. To determine these parameters from an experiment requires a diagnostic. However, during a typical radiation reaction experiment, the only diagnostics available are the energy spectra of the gamma rays and electrons after the interaction. Therefore, our goal is to show how the unknown parameters can be inferred from these diagnostics.

As this discussion is for demonstrational purposes, we will analyse a simplified experiment, in which the only unknown parameter is the overlap location of the two beams, $\mathbf{x} = (\mu_x, \mu_y)$. All other interaction parameters will be set using the values in table 5.1. We will also make a further simplification, and assume the experiment has only one diagnostic which is the electron energy spectrum after the interaction, $y = p(E)$. The effect that varying $\mathbf{x} = (\mu_x, \mu_y)$ has on the electron spectrum is demonstrated in figure 5.14. As our mock experimental data, we will take the spectrum generated with an overlap location of $\mathbf{x} = (0, 5\mu\text{m})$.

Bayesian Inference

Solving Bayes' theorem for the posterior distribution over the unknown parameters, $p(\mathbf{x} \mid y)$, requires a prior, $p(\mathbf{x})$ and a likelihood $p(y \mid \mathbf{x})$. It is common to assume the experimental noise is Gaussian, making the likelihood function a Gaussian distribution. However, our data set does not consist of point observations but a set of probability distributions, $y = p(E)$. This is also true for the forward model output, $q(E) = f(\mathbf{x})$. Therefore, we need a modified version of the Gaussian distribution, capable of handling a probability distributions as its input. If we write the Gaussian distribution as

$$\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}) = \frac{1}{\sqrt{2\pi \det(\boldsymbol{\Sigma})}} \exp(-D_M^2/2) \tag{5.28}$$

we can identify the exponent as the square of a distance $D_M = \sqrt{(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma} (\mathbf{x}-\boldsymbol{\mu})}$, known as the Mahalanobis distance. Here, we require an alternative metric that gives the distance between two probability distributions. A commonly used metric for this purpose is the Kullback–Leibler (KL) divergence

$$D_{\mathrm{KL}}(p||q) = \int_0^\infty p(E)\log\left(\frac{p(E)}{q(E)}\right)\mathrm{d}E. \tag{5.29}$$

which can replace D_M in the Gaussian distribution to give the following likelihood function

$$p(y \mid \mathbf{x}) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{D_{\mathrm{KL}}(p||q)^2}{\sigma^2}\right). \tag{5.30}$$

By defining this likelihood, we have introduced an additional unknown parameter σ to account for the random error in the experimental measurements. Although our data set has been generated by a simulation, we would still expect errors due to the Monte Carlo method and the emulator being imperfect. We can treat σ in the same way as the other unknown parameters and learn a posterior distribution over it, $p(\mathbf{x}, \mid y)$, during inference.

Applying an appropriate prior can reduce the size of the space over which we are searching. In this example, if $\sqrt{\mu_x^2+\mu_y^2} \gg w_0$ where w_0 is the laser waist at focus, the electron beam will miss the laser. The energy spectrum will then encode no information about the interaction location and there is no reason to search this space. Therefore, we will set the following uniform prior $p(\mu_{x,y}) = \mathcal{U}(-10\mu\mathrm{m}, 10\mu\mathrm{m})$ for both μ_x and μ_y. The only prior information we have about σ is that it cannot be negative. Therefore, we will also assume a uniform prior with $p(\sigma) = \mathcal{U}(0, 10\mu\mathrm{m})$. If the upper bound is larger than this, the data will be too noisy to make any inference.

Having defined the prior and likelihood, the posterior can be obtained using equation 5.27. However, this involves an expensive three dimensional integral over σ and the components of $\mathbf{x}$. Therefore, rather than performing the integral directly, we can sample from the posterior using a Markov chain Monte Carlo (MCMC) method (see section 3.3.1). The result is a three dimensional probability distribution and a plot of $p(\mu_x, \mu_y \,|\, y)$, obtained by marginalising out σ, is given in figure 5.15. To generate this distribution, 4 chains of 10000 samples were taken. Using a rate of one simulation every 72s (see section 5.3.2), obtaining 40000 samples with the Monte Carlo algorithm would take $\approx$ 800 hours. However, using the MDN the samples were obtained in $\approx$ 20 minutes.

In figure 5.15, we can see a ring of high probability. This appears due to the symmetry of the interaction, as the laser intensity is constant around the central axis at a fixed radius. We may expect the polarisation of the laser, which is in the $\hat{\mathbf{x}}$ direction, to break this symmetry. Electrons which interact at $\mathbf{x} = (5\,\mu\mathrm{m}, 0)$ are driven parallel to the intensity gradient, whereas electrons interacting at $\mathbf{x} = (0, 5\,\mu\mathrm{m})$ are driven perpendicular

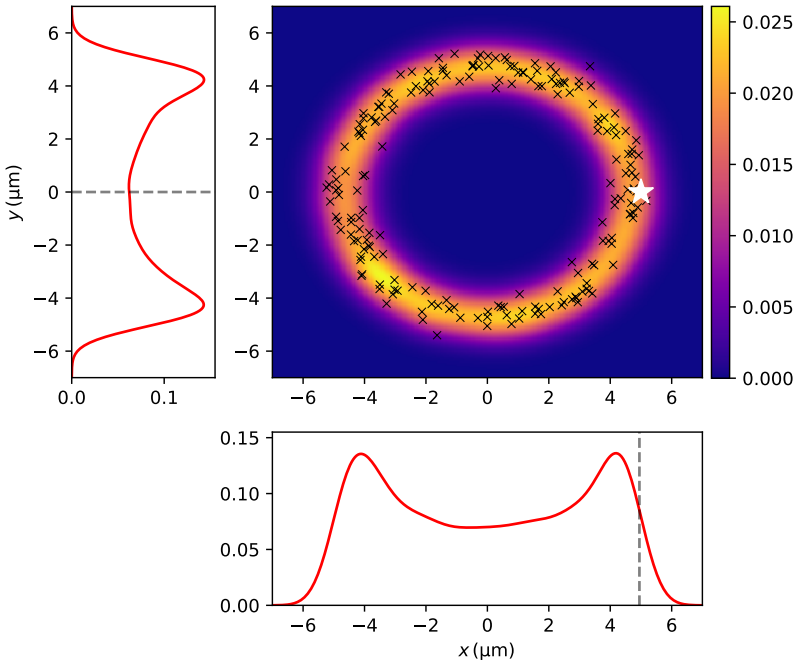


Figure 5.15: Posterior probability distribution over the unknown overlap location of an electron beam and laser pulse. Also shown are the marginalised distributions and a white star corresponding to the true overlap location.

to the intensity gradient. However, as the electrons oscillation amplitude is given by $\lambda_0 a_0/\gamma \ll w_0$ the effect is negligible. We can see the true value of the overlap location, $\mathbf{x} = (5\,\mu\text{m}, 0)$, lies well within the region of high probability.

5.4 Summary

In this chapter, we have presented the development of a new strong field QED package, for the design and analysis of experiments. The same Monte Carlo algorithm used here has been previously integrated into a number of PIC codes. Using the PIC method, the production and bulk dynamics of cascades can be modelled. However, here we have opted for a different approach, developing the package to integrate with the Geant4 framework. This allows for start-to-end simulations of experiments to be carried out, enabling faster optimisation of experimental parameters.

A second application of this package is post experimental, statistical inference. Performing statistical inference requires a forward model of the experiment. We have found that the Monte Carlo algorithm is too slow for this purpose. Therefore, we have developed a machine learning emulator, using a mixture density network, which can be carried out orders of magnitude more quickly. This enables us to perform Bayesian inference on

experimental data, which was demonstrated in section 5.3.3.

When developing the MDN emulator a number of assumptions were made. For example the background field parameters were kept constant. This was due to the implementation of the Monte Carlo algorithm, making it is more difficult to generate the required data set. Future work may involve updating the Monte Carlo algorithm, such that the field parameters are also sampled at the start of each event. This will enable the generation of a suitable data set, that can be used to train an emulator with the field parameters as inputs.

To further simplify the data set for the MDN, secondary particles were ignored. However, during a radiation reaction experiment, the emitted photon spectrum will also be measured. Therefore, it would also be useful to have a photon emulator when performing Bayesian inference. This could be achieved by developing a separate MDN for the different particle types. Including secondaries in the emulator requires additional considerations, as the number of particles per event is no longer constant. To deal with this, an extra node could be added to the network output which returns a distribution over the number of particles.

The discussion of the previous two chapters has been focused on the development, rather than application, of new physics packages. However, this will change in the next chapter as we discuss the design and analysis of a Breit-Wheeler detection experiment.

Chapter 6

Design and Analysis of a Photon-Photon Collider Experiment

A fundamental difference between classical and quantum electrodynamics, is the non-linearity of the electromagnetic field. A consequence of this is the direct interaction of photons, which can occur through a number of difference processes. The two most widely discussed are the Breit-Wheeler process ($\gamma\gamma \rightarrow e^+e^-$), shown in figure 6.1 (a), and photon-photon scattering ($\gamma\gamma \rightarrow \gamma\gamma$), shown in figure 6.1 (b). In section 1.2, we discussed the importance of studying these processes from both a theoretical and astrophysical viewpoint. However, despite their relevance neither has been directly observed in any laboratory experiment to date[1].

The reason that the Breit-Wheeler process has never been observed whereas other similar QED processes have (*e.g.* Compton scattering and Dirac annihilation), is that the Breit-Wheeler cross-section exhibits a relatively high centre of mass (CM) energy threshold. For the collision of two photons, the CM energy squared is given by

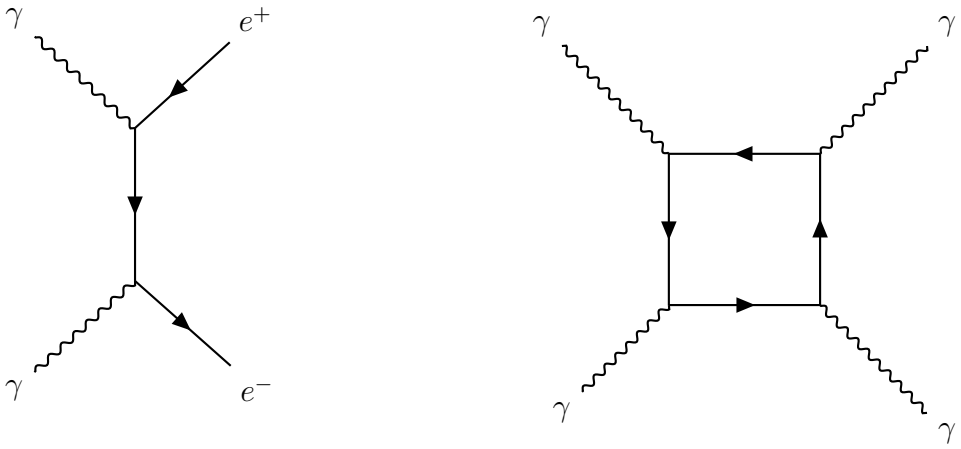
$$s = 2\omega_1\omega_2(1 - \cos\theta) \tag{6.1}$$

where ω_1 and ω_2 are the energy of the two colliding photons, θ is the angle between them, and we have set $c = m_e = \hbar = 1$. To generate an electron-positron pair at rest, the CM energy must be at least twice the rest mass of the electron ($s > 4$). Generating two sources of photons which overcome this threshold and are dense enough to produce a detectable number of pairs is challenging.

This has been apparent since the prediction of the Breit-Wheeler process, with G. Breit and J. A. Wheeler ending their seminal paper with[18]

> "It is also hopeless to try to observe the pair formation in laboratory experiments."

[1]Here we are referring to the linear Breit-Wheeler process and photon-photon scattering with real photons. The nonlinear Breit-Wheeler process was observed at SLAC in 1997 [56] and photon-photon scattering of virtual photons was observed at CERN in 2019 [194].



(a) Breit-Wheeler process (b) Photon-photon scattering

Figure 6.1: Example of lowest order Feynman diagrams which contribute to (a) the Breit-Wheeler process and (b) photon-photon scattering.

However, this dire prediction was made before the advent of the laser, which has lead to an exponential growth in obtainable photon densities over the past few decades. This has lead to a number of authors proposing experimental schemes to make the first observations of the Breit-Wheeler process [157, 158, 159, 160]. What these schemes have in common is the production of high energy photons ($\omega > 1\,\mathrm{MeV}$), yet where they differ is the method of generating them and their expected pair yield.

All photon-photon collider experiments require the generation of two photon sources, which can either be asymmetric or symmetric in energy. An example of an asymmetric setup is proposed by Pike *et al.* (2014) [157]. Here the authors suggest colliding a high energy γ-ray beam ($\omega_1 > 100\,\mathrm{MeV}$) with a thermal x-ray field ($\omega_2 \sim 500\,\mathrm{eV}$). To generate the γ-ray beam, bremsstrahlung emission of a laser wakefield accelerated (LWFA) electron beam interacting with a high Z-material target is used. To generate the x-ray field, the scheme takes advantage of a laser heated hohlraum, capable of producing a dense source of photons [195]. They predict a pair yield of up to 10^5 per shot, however, generating the laser heated hohlraum requires a high energy facility, which have low repetition rates.

If a symmetric collider design is used, the energy requirement of the γ-ray photon is reduced from $\omega_1 > 100\,\mathrm{MeV}$ to $\omega_1 > 0.511\,\mathrm{MeV}$. A number of symmetric collider designs have been reviewed by Ribeyre *et al.* [158]. They found synchrotron emission from a highly energetic electron beam in an intense laser field to be the most efficient, with a predicted yield of 10^4 pairs. A separate investigation by Drebot *et al.* [159] demonstrated the possibility of developing a symmetric collider at a low energy facility. Using two Compton sources, produced by joule-class lasers, they predicted a yield of $\sim 10^{-4}$ pairs per shot. This is far lower than the other schemes, however, opens up the possibility of using a high repetition rate facility. A statistically significant observation of the Breit-Wheeler process can then be made over many shots.

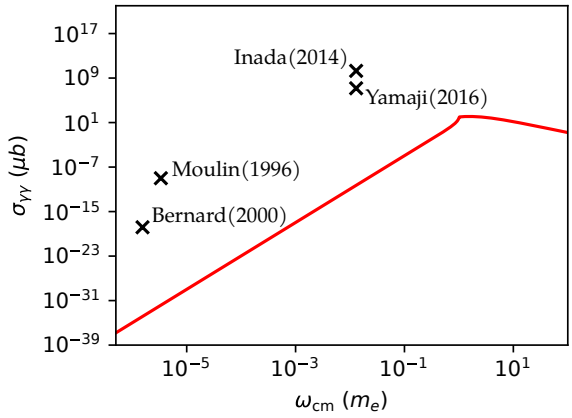


Figure 6.2: Photon-photon scattering cross-section against CM energy. Also shown are bounds applied to the cross-section by previous experiments [200, 201, 202, 203].

These photon collider schemes could also be applied to study photon-photon scattering. However, this introduces additional difficulties which must be considered. Although there is no strict threshold in the photon-photon scattering cross-section, it decays rapidly below the rest mass of the electron. Therefore, high energy photons are still required to prevent the cross-section becoming vanishingly small. Photon-photon scattering is higher order interaction than the Breit-Wheeler process, making the cross-section a factor of $\alpha^2 \approx 5.3 \times 10^{-4}$ times smaller. Finally, as all the incoming and outgoing particles are photons, identifying an interaction event above noise will be extremely challenging.

Due to these challenges, in recent years experimental studies of photon-photon scattering have focused on bounding the cross-section rather than observing the interaction. Finding these upper limits on the cross-section is part of the search for physics beyond the standard model. Deviations from the expected QED value could suggest photons are capable of coupling to unknown weakly interacting particles such as axion-like or minicharged particles [196, 197, 198, 199]. The cross-section limits that have been made by previous studies are shown in figure 6.2. The closest of these to the QED cross-section is that of Bernard *et al.* (2000) [200], however, this is still $\sim$ 18 orders of magnitude higher. This is due to all four experiments operating in a regime where $\omega_{\mathrm{CM}} \ll m_e$. As we can see in figure 6.2, the cross-section is severely suppressed here.

In 2018, motivated by these photon collider proposals, an experimental campaign took place which set out to provide the first laboratory measurements of the Breit-Wheeler process, as well as to make the most stringent bound on the photon-photon scattering cross-section. The experimental setup was based on the asymmetric photon collider proposed by Pike *et al.* (2014) [157]. However, rather than using a low repetition rate system (*e.g.* NIF with $\sim$ 1 shot / day), the Gemini laser at the Central Laser Facility (CLF) was used ($\sim$ 1 shot / 20 s). This significantly reduces the expected pair yield, however, with more shots a better characterisation of the background noise and a better statistical

analysis can be performed.

A key aspect of both the design and analysis of this experiment was detailed computational modelling. Based on rough estimates of the Breit-Wheeler yield, it was expected that the signal-to-noise ratio of the experiment would be low. Therefore, large scale Geant4 simulations, using the photon physics package discussed in chapter 4, were performed to predict the signal-to-noise ratio and optimise the setup. During post experimental analysis, these same simulations provided a method for comparing the obtained measurements to theory.

This computational modelling is the topic of this chapter. However, before beginning, an in depth review of the experiment will be provided. We will then discuss the numerical modelling that was performed to optimise the setup and increase the signal-to-noise ratio. With the optimised setup we will review the prospect of using this scheme to detect the Breit-Wheeler process. Finally, we will discuss the numerical modelling that took place post-experiment. From this we will be able to determine if the Breit-Wheeler process was detactable as well as bound the photon-photon scattering cross-section.

6.1 Experiment Overview

The Gemini laser is a dual beam, 300 TW Ti:Sa system, allowing us to generate and collide two high energy-density photon sources. The experimental setup was based on the scheme by Pike *et al.* (2014) with asymmetrical photon sources [157]. The higher energy source was provided by bremsstrahlung emission of a LWFA electron beam in a high Z-material target, whereas the lower energy source was generated through direct laser heating of a burn-through foil. A diagram of the full experimental setup can be found in figure 6.3. This can be separated into three parts: the γ-ray source; the x-ray source; and the analyser magnets and particle detectors. In this section we will review each of these parts in detail.

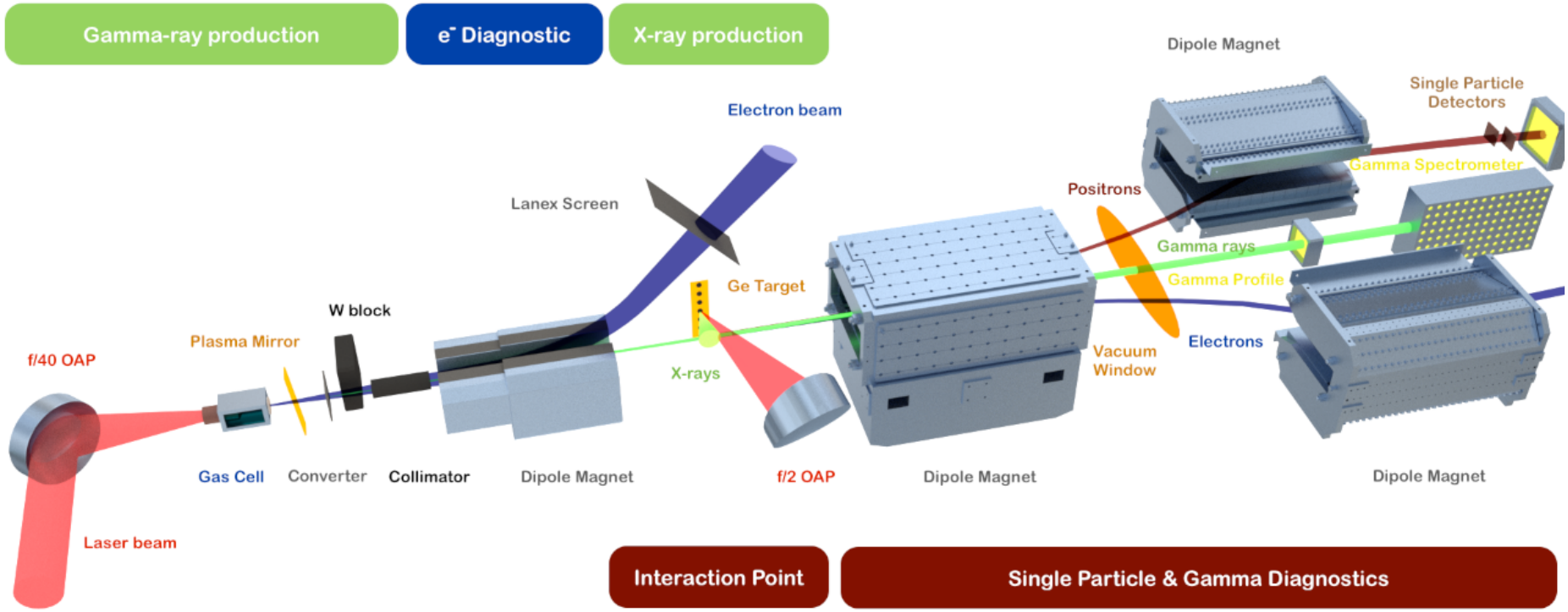


Figure 6.3: Diagram of the experimental setup showing the key components with the chamber and lead shielding removed. Starting from the left, this diagram shows the first laser (red) focused by an off-axis parabola (OAP) into a gas cell target. This produces a beam of electrons (blue) through the LWFA mechanism which propagate into a bismuth converter foil. This generates a beam of γ-rays which are cut in half and collimated with blocks of tungsten. The particles then pass through an on-axis magnet which removes residual electrons and positrons generated in the foil. The γ-ray beam then continues into the x-ray field, generated by the second laser (red) interacting with a germanium target mounted on Kapton tape. Any positrons produce in the interaction are sent through a magnetic chicane to single particle detectors in a well shielded, low noise area of the experiment. This diagram was provided by E. Gerstmayr

6.1.1 X-Ray Source

The x-ray field was generated through the thermal emission of a laser heated 100 nm germanium (Ge) foil. As this solid Ge foil is heated, it turns into a plasma, leading to the emission of intense x-ray radiation predominantly due to M-L band transitions at an energy of $\sim$ 1.5 keV[204]. The laser pulse had a duration of 40 ps fwhm, and a total energy of delivering an energy of 10.72 ± 0.28J. It was focused to an elliptical spot, with major and minor axes of $(217 \pm 6)\,\mu$m $\times$ $(77 \pm 6)\,\mu m$ respectively, delivering an energy of 10.72 ± 0.28 J each shot. The duration of the pulse was 40 ps fwhm, which is roughly the same as the duration of the x-ray emission.

The atomic number of Ge is fairly high ($Z = 32$) and the foil is close to the γ-ray beam axis. This makes the foil a potential noise source, due to γ-rays decaying into electron-positron pairs through the Bethe–Heitler process. Crucially, these would have the same energy range as the expected Breit-Wheeler pairs, and be generated very close to the photon collision point, making it impossible to distinguish this noise source from the signal. Therefore, to reduce this the Ge targets were mounted on a Kapton ($C_{22}H_{10}N_2O_5$) tape with a lower average atomic number, limiting the mass of Ge close to the interaction. A motorised tape-drive was used to change targets between shots.

To diagnose the x-ray field, a pinhole imaging system and crystal spectrometer were used. The pinhole imaging system gave an on-shot measure of both the emission spot size and the target alignment. The spectrometer used a flat, thallium acid phthalate (TlAP) crystal, with a spectral window of $\sim$ 700 eV, centred around $\sim$ 1.6 keV. This spectral window is around the M-L band transitions of Ge.

6.1.2 γ-ray Source

The γ-ray source was generated through bremsstrahlung emission, which first requires a beam of high energy electrons. To produce these electrons, the second laser pulse was focused into a 17.5 mm gas cell filled with helium and a 2% nitrogen dopant. The duration of the pulse was 45 ± 5 fs and the spot size was $(44 \pm 2)\,\mu$m $\times$ $(53 \pm 2)\,\mu$m. This provided an energy on target of 5.51 ± 0.64 J, corresponding to a normalised vector potential $a_0 = 1.13 \pm 0.18$. Through the LWFA mechanism, a beam of high energy electrons was emitted, with a charge of $\sim$ 50 pC.

These electrons were interacted with a 0.5 mm thick bismuth (Bi) foil, acting as a bremsstrahlung converter. This emits a beam of high energy γ-rays (> 100 MeV) with a similar duration to that of the driving laser pulse. However, bremsstrahlung converters also generate a large number of low energy, divergent γ-rays, which are a potential noise source. If these divergent γ-rays were to interact with the Ge foil, or another part of the experiential setup, they would produce background pairs through the Bethe–Heitler process. Therefore, to prevent this, an on-axis beam collimator was used. This consisted

of a 100mm block of tungsten (W) with a 2 mm diameter hole drilled through the centre. This absorbs any γ-ray with a divergence of greater than 10 mrad. To reduce the noise further, a 50 mm tungsten block was placed just off axis, shadowing the Ge foil from the γ-ray beam. Placing a high Z-material close the γ-ray beam axis will itself generate a large number of background Bethe–Heitler pairs. Therefore, a 300 mm on-axis magnet with a field strength of $|\mathbf{B}| = 1\,\mathrm{T}$, was used to remove these before the photon-photon interaction zone.

To diagnose the γ-ray beam, two scintilator detectors were used, each consisting of a stack of caesium iodide (CsI) crystals. The crystals were wrapped in aluminium foil to prevent light leakage and mounted in an aluminium frame. The first detector consisted of 20×20 CsI crystals, each with dimensions $2\,\mathrm{mm} \times 2\,\mathrm{mm} \times 20\,\mathrm{mm}$. This was positioned such that the crystals were stacked transverse to the γ-ray beam axis. This then gives us a measurement of the γ-ray beam profile. The second detector consisted of 33×47 CsI crystals doped with thallium and with dimensions of $5\,\mathrm{mm} \times 5\,\mathrm{mm} \times 50\,\mathrm{mm}$. In this case, the crystals were stacked along the direction of the γ-ray beam axis. As higher energy γ-rays are capable of penetrating deeper into the stack, this provided a measurement of the γ-ray beam spectrum.

6.1.3 Analyser Magnets and Single Particle Detectors

Due to the asymmetry in the energy of the colliding photons, if any Breit-Wheeler pairs are produced, they will be highly beamed along the dominant photon axis. This is a key advantage that a scheme with asymmetrical photon sources has over a scheme with symmetrical sources. With the signal emitted in a known direction, detectors can be situated in a low noise area of the experiment. If the signal was emitted at all angles, the entire interaction zone would have to be surrounded by detectors to maximise the detection efficiency.

To transport the particles to the detectors, a magnetic chicane was used, consisting of an on-axis magnet and two off-axis magnets. The first of these magnets was situated inside the target chamber and had a length of 600mm and a field strength of $|\mathbf{B}| = 0.6\,\mathrm{T}$. To prevent the edge of the γ-ray beam clipping the poles, the magnet had a large gap of 100 mm. This separates the electrons from the positrons, directing them both away from the γ-ray beam axis. The particles then leave the target chamber through a thin Kapton window. When the electrons and positrons pass through the first magnet they are dispersed in energy. Therefore, a set of two off-axis magnets, with a length of 600mm and a field strength of $|\mathbf{B}| = 0.6\,\mathrm{T}$, were used to re-collimate the particles. These particles then continue to a well shielded region at the back of the experimental area. By moving the position of the magnets, the acceptance energy bandwidth of the chicane can be altered. To both maximise the number of Breit-Wheeler pairs and minimise the background noise,

an acceptance band of $220 - 380\,\mathrm{MeV}$ was used.

At the end of the positron arm of the chicane, single particle detectors were situated. No detectors were used on the electron arm. This is because there are more background processes which release electrons (*e.g.* the photoelectric effect), so the expected noise on the positron arm was lower. The single particle detectors consisted of two TimePix3 [205] detectors and a scintillation detector. Each TimePix3 detector consisted of a 256×256 silicon pixel array, spanning an area of $\sim 200\,\mathrm{mm}^2$, with thicknesses of $300\,\mu\mathrm{m}$ and $150\,\mu\mathrm{m}$. These were arranged in layers to enable positron tracking and particle discrimination based on direction. As the pixel array consists of a thin layer of silicon, these detectors were insensitive to background γ-rays. The scintillation detector consisted of a stack of 20×15 CsI crystals with dimensions $5 \times 5 \times 50\,\mathrm{mm}$ and coated with a thin layer of TiO_2. Therefore, the detector spanned an area of $7500\,\mathrm{mm}^2$, far larger than the TimePix3 detectors. However, as the scintillation detector consisted of a larger volume, it is much more sensitive to background γ-rays.

6.2 Pre-Experiment Design and Optimisation

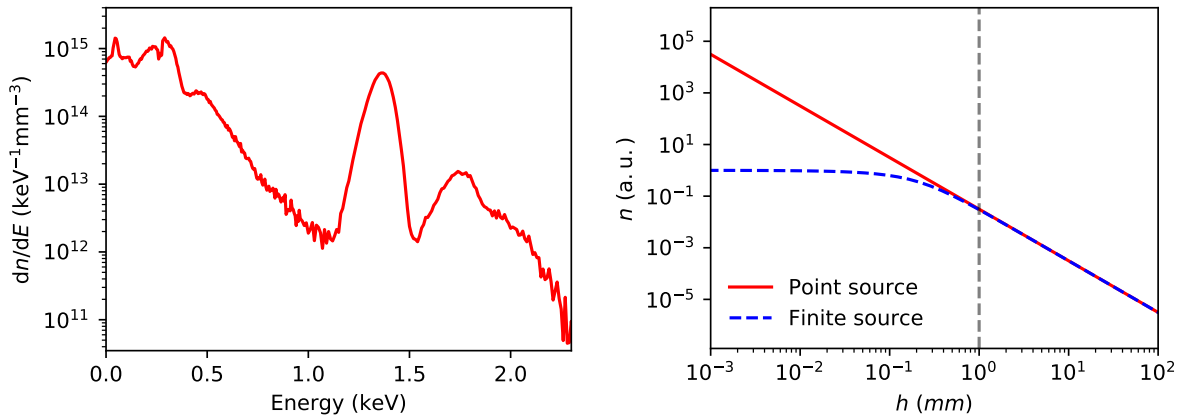
The key factor which determines the success of a photon-photon collider experiment is the signal-to-noise ratio. As we can see from the previous section, this is a complex experiment with many different parts. Therefore, calculating the signal-to-noise ratio is challenging, and we must rely on large scale numerical simulations. To carry out these simulations the Monte Carlo code Geant4 was used, with the addition of the photon physics package discussed in chapter 4.

In this section we will review these large scale numerical simulation. This will involve independently modelling both photon sources, starting with the x-ray field. We will then go on to show how Geant4 simulations can be used to optimise the γ-ray source. Finally, with a fully optimised setup, we will present calculations of the signal-to-noise ratio and discuss the efficacy of the experimental setup.

6.2.1 Modelling the X-Ray Source

As discussed in chapter 4, to use the Geant4 photon physics package we must designate one source as a static field, while the second source is treated dynamical. Due to the difference in the duration of the driving laser pulses, the x-ray source will be treated as static and the γ-ray beam as dynamic. To model the static field, the photon density distribution is separated into angle and energy dependant parts $n(\epsilon, \theta, \phi) = f(\omega)\,\Phi(\theta, \phi)$. These two parts are then set by choosing the most appropriate options from tables 4.1 and 4.2.

To gain a better understanding of the radiation energy spectrum, $f(\omega)$, the laser-



(a) X-ray spectrum

(b) Source density

Figure 6.4: Simulations of x-ray field. (a) shows the x-ray energy spectrum from a radiation-hydrodynamics simulation carried out by J. Morton (AWE). (b) shows the x-ray field density as a function of height above a point (red) or finite size (blue) source. Also shown is the closest approach a γ-ray will make to the source (grey dashed).

solid interaction was modelled using a radiation-hydrodynamics code (performed by John Morton (AWE)). The output of this simulation can be seen in figure 6.4 (a). This shows strong emission in the $1-1.5\,\mathrm{keV}$ window, corresponding to M-L band transitions. Using the tabulated source from table 4.1, we can use this exact spectrum to model $f(\omega)$.

The photon angle distribution close to a finite size emitter is complex and reviews of the subject can be found in refs. [206] and [207]. The only option from table 4.2 that is capable of modelling this is the tabulated source. However, using this source is both computationally expensive and memory intensive, which will limit the number of simulations which can be performed. As we move away from a finite emitter it begins to look more like a point source, becoming a good approximation when the distance from the emitter, h, is much larger than the source size a. The size of the x-ray emitter will be roughly that of the driving laser beam, so we can set $a \approx 250\,\mu\mathrm{m}$. The photon density from an emitter of this size as well as a point source are shown as a function of h in figure 6.4. Also marked on this plot is the closest approach a γ-ray will make to the Ge foil, which is limited due to the tungsten block in the beam path. At this value, both sources are almost identical, suggesting a point source is an appropriate, and faster, alternative to a tabulated source.

Having defined $n(\epsilon, \theta, \phi)$, we will now test if the temporal and spatial assumptions that are made by the photon package (see section 4.1.3) hold for this field. The temporal assumption is that the field remains static during the duration of the interaction, whereas the spatial assumption is that gradients in the field can be modelled using multiple sub-fields. To test these two assumptions, ray tracing simulations will be performed.

Time Variation of the Field

To test the static field approximation, we will assume the duration of the x-ray field, τ_x, and γ-ray beam, τ_γ, are roughly the same as the duration of the driving laser pulses. Therefore, in comparison to τ_x we can let $\tau_\gamma = 0$ to a good approximation. However, this alone is not sufficient to ensure the x-ray field is static. We also require that the time taken for the γ-ray beam to traverse the field is much smaller than τ_x. As the speed of the γ-rays is fixed at c, this time is only a function of h.

To analyse the impact that τ_x and h have, we can perform ray tracing simulations of γ-rays propagating through a time varying field. Examples of the observed density at $h = 1\,\text{mm}$ and $h = 6\,\text{mm}$ are shown in figures 6.5 (a) and (b) respectively. Also included in these plots are the observed density for $\tau_x \to \infty$. At $h = 1\,\text{mm}$ the γ-ray passes through the field quickly, so the duration has little effect. However, this is not the case for $h = 6\,\text{mm}$ which shows a significant difference.

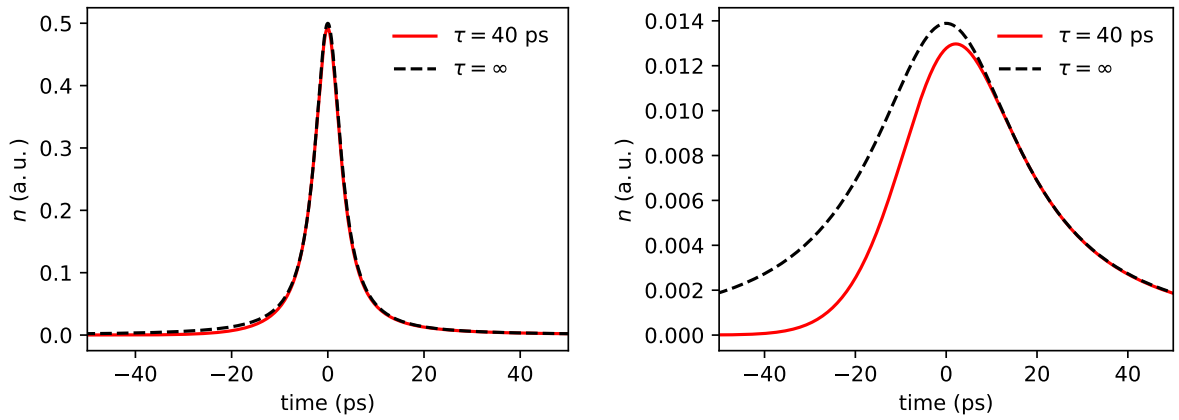
We can further study this temporal effect by plotting the ratio of the line integrated field density at $\tau_x = 40\,\text{ps}$ and $\tau_x = \infty$ as a function of h. This is shown in figure 6.5 (c), and also includes a line at $h = 4\,\text{mm}$ corresponding to the furthest approach of a γ-ray. At this furthest approach, including the field duration only reduces the integrated density by $\sim 15\%$. Therefore, treating the field as static (*i.e.* $\tau_x = \infty$) will only introduce small errors in the estimate.

Spatial Discretization of the X-ray Field

As the computational domain of a Geant4 simulation is comprised of blocks with uniform density, to model spatial gradients the field must be split into multiple sub-fields, each with a different value of $n(\omega, \theta, \phi)$. As a γ-ray passes through the field, the same calculation is performed within each sub-field. Therefore, the computational time will scale linearly with the average number of sub-fields entered. To optimise both efficiency and accuracy, the spatial resolution of the field should be set prior to performing expensive calculations.

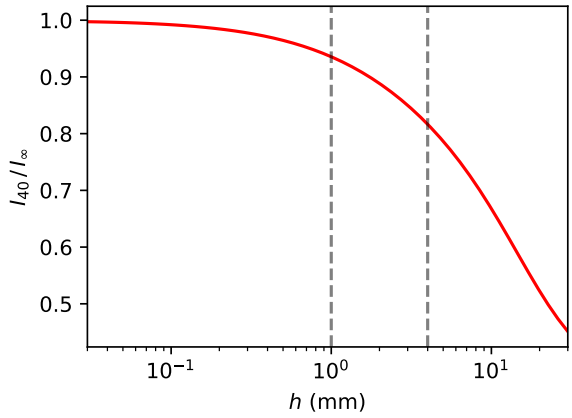
To find the optimum resolution, we can again perform ray tracing simulations. Examples of these simulations are shown in figure 6.6 (a). Here we can see the observed field using low resolution and high resolutions, as well as the true field[2]. Shown alongside in figure 6.6 (b) is the integrated density as a function of resolution, with the dashed line corresponding to the true field value. An optimum resolution is reached at $\sim 15^3$ voxels. Increasing beyond this point offers diminishing returns with increases in computational costs.

[2]As this is a numerical simulation, the "true field" case still requires discretization, however, it is the γ-ray path that is discretized rather than the x-ray field.



(a) $h = 1\,\text{mm}$

(b) $h = 6\,\text{mm}$



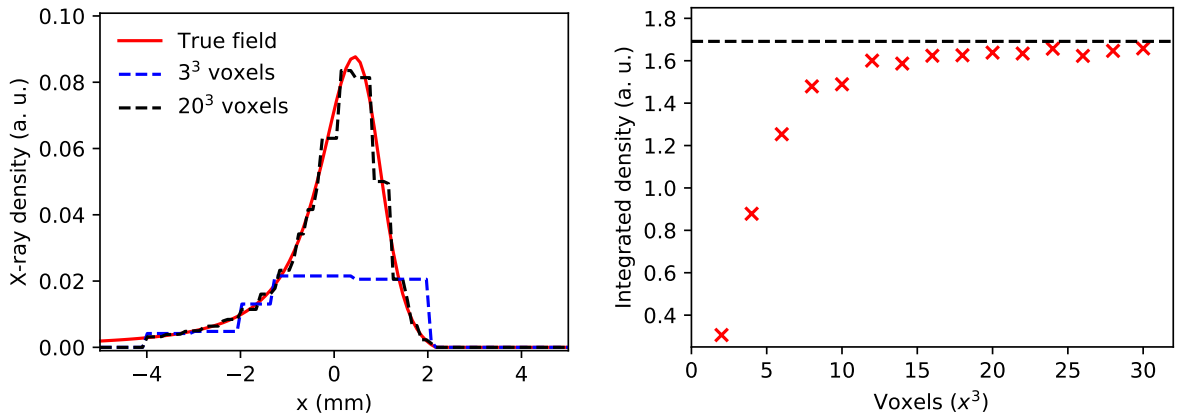
(c) Line integrated density ratio

Figure 6.5: Ray tracing simulations of x-ray density observed by a γ-ray. (a) shows the observed density for a γ-ray travelling 1 mm above the source (b) shows the observed density for a γ-ray travelling 6 mm above the source. (c) shows the line integrated density at $\tau_x = 40\,\text{ps}$ and $\tau_x = \infty$ as a function of h. Also shown are the closest and furthest approach made by a γ-ray (grey dashed).

6.2.2 Optimising the γ-ray Converter

Generating the γ-ray source required a beam of high energy electrons that was provided by a LWFA. If we wanted to perform a full start-to-end simulation of the entire experiment, the LWFA interaction could be modelled using a particle-in-cell (PIC) code. However, this is unnecessary as LWFA experiments have taken place using the Gemini laser on multiple occasions. An example of the electron beam spectra obtained during a recent experiment by Poder *et al.* [165] is shown in figure 6.7.

By firing this beam of electrons into a converter foil, a beam of γ-rays are produced. This conversion occurs through the bremsstrahlung process, due to the electrons scattering off the nuclear field of an atom. The emitted γ-rays can themselves interact with the nuclear field, generating electron-positron pairs through the Bethe-Heitler process. These



(a) Observed x-ray density

(b) Line integrated x-ray density

Figure 6.6: Ray tracing simulations showing the effect of spatial discretization. (a) shows the field observed by a γ-ray with a low resolution (3^3 voxels, blue), a high resolution (20^3 voxels, black) and the true field (red). (b) shows the line integrated density as a function of resolution. Also shown is the true value (black dashed).

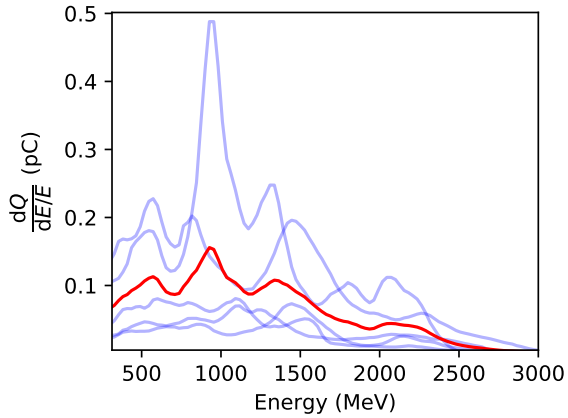
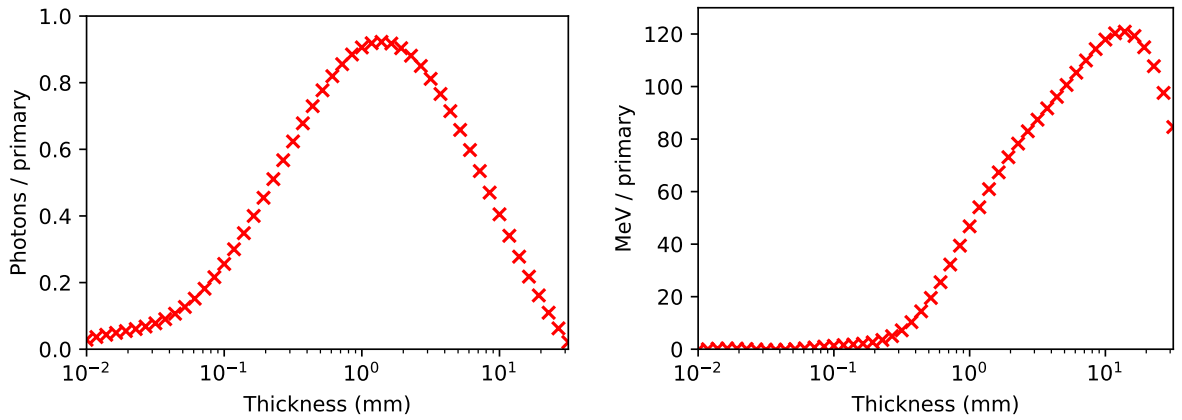


Figure 6.7: Example of high energy electron spectra obtained during a Gemini experiment (blue). Also shown is the average spectrum (red). This figure was adapted from ref. [165].

pairs can undergo further bremsstrahlung emission, setting up a QED cascade. This cascade has the effect of increasing the number of particles, but decreasing their average energy.

By changing the length, l_c, and material, Z, of the converter the extent of the cascade can be controlled to suit the problem at hand. However, as discussed in section 4.4.1, these are both manifestations of a single latent variable, the radiation length X_0. Therefore, when designing the converter, we are free to fix Z and focus on optimising l_c. To reduce the volume of material required, bismuth was selected as it has a high atomic number of $Z = 83$.

If we are simply interested in maximising the signal, the optimum converter would



(a) Maximised metric

(b) Minimised metric

Figure 6.8: Plots of metrics against converter thickness. (a) shows the number of photons with an energy greater than 170 MeV per primary electron. (b) shows the energy emitted with an angle greater than 10 mrad per primary.

produce the maximum number of γ-rays above the threshold for the Breit-Wheeler process (or at the peak of the photon-photon scattering cross-section). Taking the x-ray energy to be $\omega_x \sim 1.4\,\text{keV}$ and using equation 6.1, the threshold γ-ray energy is $\omega_\gamma \sim 170\,\text{MeV}$ for a head on collision. However, for a successful experiment, we are not just interested in maximising the signal, but also minimising the background noise. The most likely particles to generate noise are those emitted with a divergence greater than 10 mrad.

As metrics for optimising the converter thickness, we will use both the number of γ-rays with $\omega_\gamma > 170\,\text{MeV}$ and the total energy of particles with a divergence greater than 10 mrad. The former of these should be maximised whereas the latter should be minimised. In figure 6.8 we can see the result of Geant4 simulations, showing the effect of varying l_c on these two metrics. The number of photons over threshold is maximised at $l_c \approx 2\,\text{mm}$, with ≈ 0.9 γ-rays per primary electron. However, around this thickness the energy in divergent particles starts to increase significantly. Therefore, as a trade off, a converter thickness of $l_c = 0.5\,\text{mm}$ was used in the experiment.

6.2.3 Background Noise Reduction

Using the Gemini laser, electron beams with a charge in excess of 100 pC can be produced [165]. This corresponds to $\sim 10^9$ seed electrons interacting with the bremsstrahlung converter. This then emits an even greater number of particles, which pass through the experimental setup. As we will see in section 6.2.4, the expected number of Breit-Wheeler events per shot is far lower (~ 1/shot). Therefore, it is vital that the experimental setup is optimised such that the number of background particles reaching the detectors does not overwhelm any signal. This is achieved by applying shielding to different parts of the setup.

There are two different types of noise source, photons and leptons (electrons/positrons). Reducing the photons is of secondary importance as they are less likely to interact with the TimePix3 detectors and will produce a different signature if they do. However, it is vital that the lepton noise is reduced as they cannot be distinguished from the signal. To find the optimum shielding configuration that achieves this, Geant4 simulations were performed iteratively for a number of shielding configurations. This was carried out until an acceptable noise level was reached. In this section, rather than reviewing this iterative process in detail, we will identify noise sources and demonstrate how they can be removed.

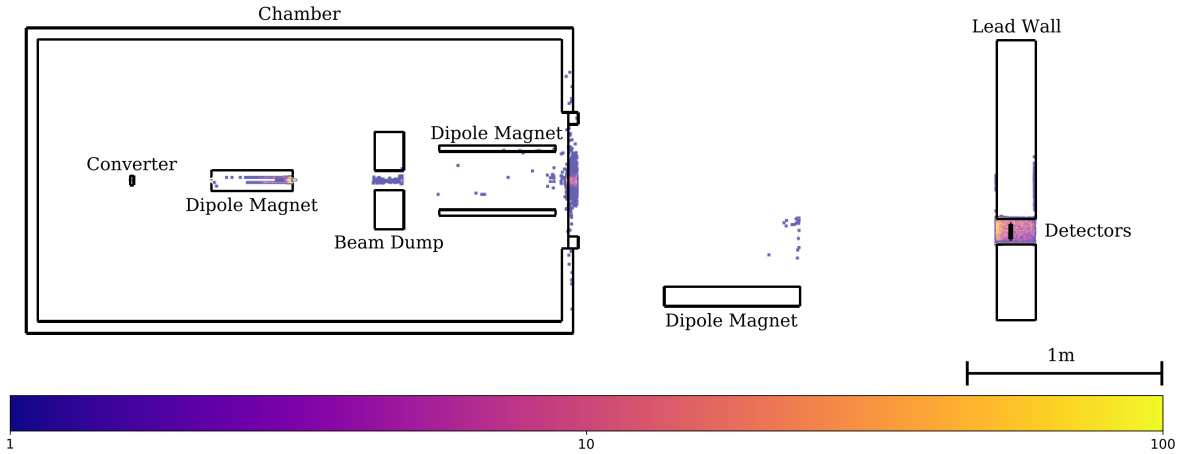
Low Shielding Configuration

As a starting point for the iterative calculations, a low shielding configuration was used. Two cross-sectional diagrams of this configuration are shown in figure 6.9. Starting from the left, visible in these diagram are: the target chamber walls; the γ-ray converter; the residual electron removal magnet; the residual electron beam dump; the first magnet in the chicane; the chamber exit window; the second magnet in the chicane; the target area back wall; and finally the detector. To reduce the background, we need to identify the parts of the setup that are noise sources. This can be achieved by backtracking the particles which pass through the detector location and recording their point of origin. From this, we can generate heat maps of noise sources, which are shown in figures 6.9 (a) and (b) for leptons and photons respectively.

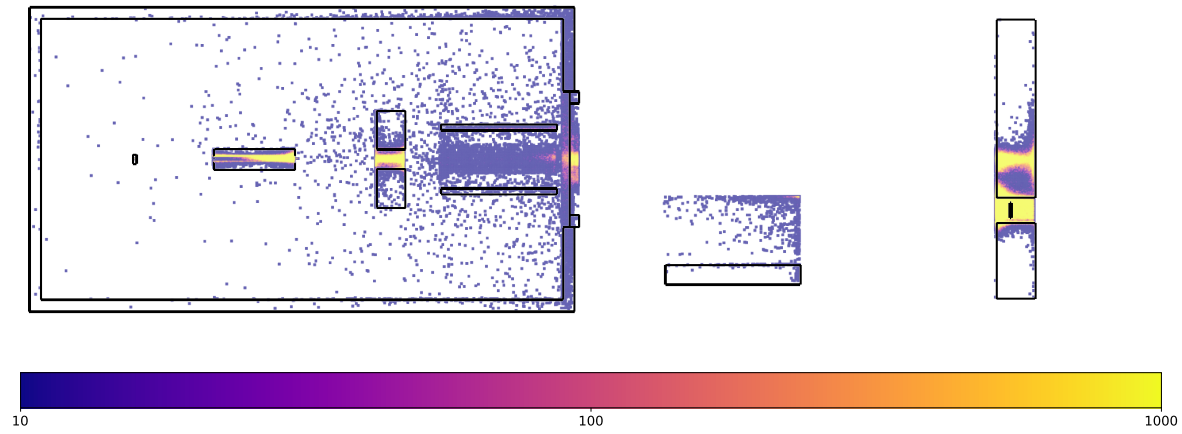
From figure 6.9 (a) and (b) we can identify three design issues, which are the cause of the majority of the background particles. The first is the γ-ray converter, which generates divergent particles which scatter off the experimental setup. This leads to both background leptons and photons which originate inside the target chamber. The second issue is the direct line of sight from scattering location, to the detector. Examples of this include the particles which originate from the target chamber walls and exit window. The final issue is the γ-ray beam dump. It is insufficient in size, leading to particles leaking out into the detector location. We will now discuss the optimised shielding configuration, showing how each of these issues can be addressed.

High Shielding Configuration

In figures 6.10 (a) and (b) we can see cross-sectional diagrams of the optimised shielding configuration with lepton and photon source heat maps respectively. To reduce the noise originating inside the target chamber, an off-axis block and a collimator, both made of tungsten, have been added. The block removes the half of the γ-ray beam that is directed towards the x-ray foil, while the collimator removes the highly divergent particles. By introducing lead walls just outside the target chamber and surrounding the γ-ray beam dump, direct lines of sight from scattering locations have been removed. Finally the γ-ray



(a) Lepton noise source (a.u.)

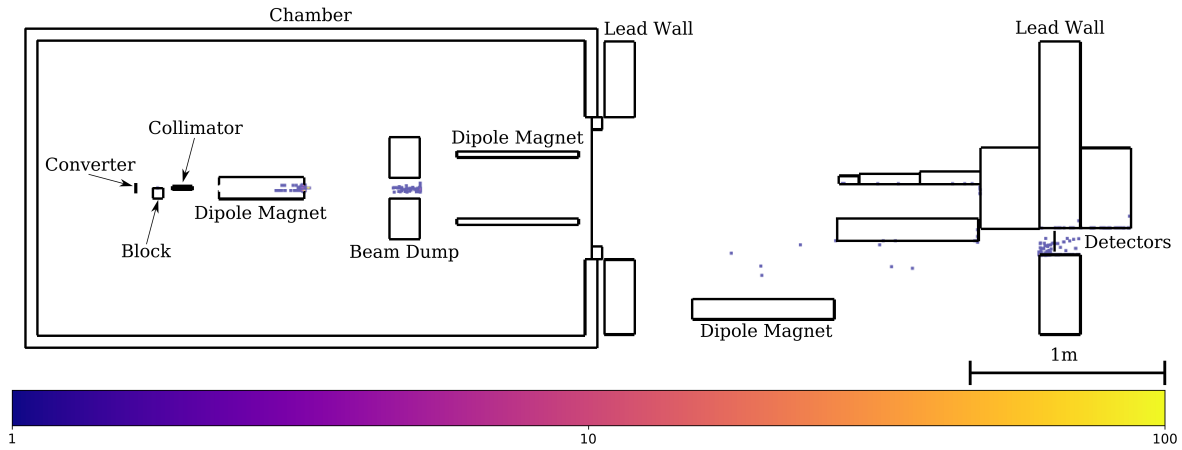


(b) Photon ray noise source (a.u.)

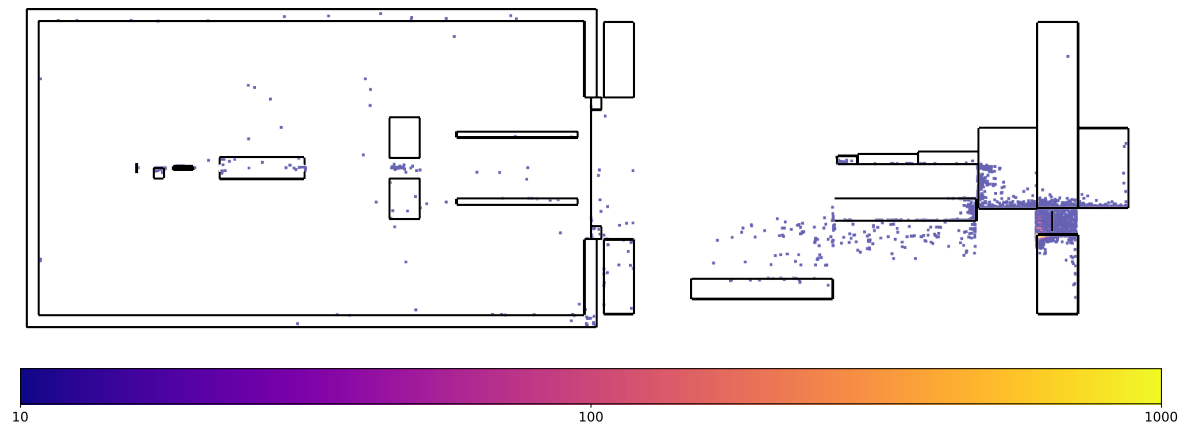
Figure 6.9: Cross-sectional diagrams of experimental setup in low shielding configuration. (a) shows the lepton noise sources and (b) shows the photon noise sources.

beam dump has been significantly reinforced to prevent particles leaking into the detector hole.

Comparing the diagrams in figure 6.9 to those in figure 6.10 we can see that both lepton and photon noise sources have been significantly reduced. There are still some locations of concern, such as the x-ray foil and the residual electron beam dump, however, it was not possible to find a configuration that removed these without having a detrimental impact on the expected signal. The number of particles reaching the detectors, for both shielding configurations, is summarised in table 6.1. If we take the number of primary particles in the electron beam to be 10^9 (corresponding to a charge of $Q = 160\,\mathrm{pC}$), then we can expect a background of ~ 17 leptons per shot.



(a) Lepton noise source (a.u.)



(b) Photon noise source (a.u.)

Figure 6.10: Cross-sectional diagrams of experimental setup in high shielding configuration. (a) shows the lepton noise sources and (b) shows the photon noise sources.

6.2.4 Signal-to-Noise Ratio Calculation

As we have mentioned before, the most important factor which determines whether a photon-photon collider experiment will be successful is the signal-to-noise ratio. Estimating the signal-to-noise ratio clearly involves two parts, a background calculation (which was performed in previous section) and a signal calculation. With the aid of the photon physics package discussed in chapter 4, the signal calculation can also be performed within Geant4, using the same setup as the previous section.

Although the additional shielding reduced the background, blocking part of the γ-ray beam will also reduce the signal. Therefore, signal calculations should be performed for both shielding configurations. The result of these calculations is given in table 6.1.

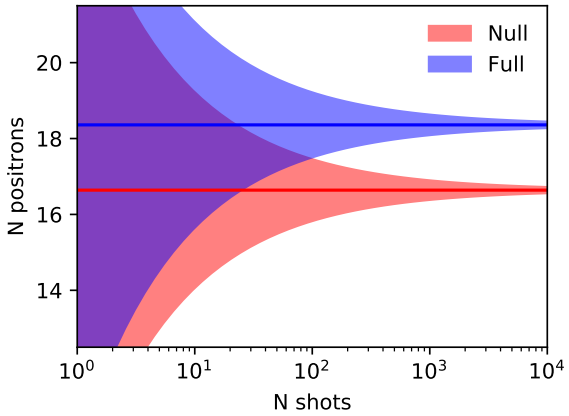


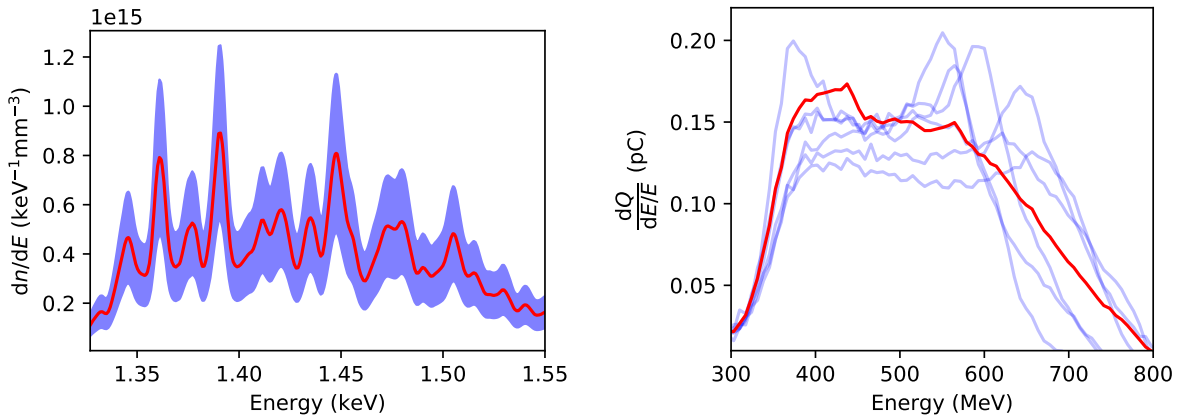
Figure 6.11: Diagram showing the 2σ confidence level of μ_N and μ_F as a function of the number of shots.

The presence of the shielding reduce the Breit-Wheeler signal by $\sim 40\%$ and background by $\sim 92\%$. Therefore, the benefits far outweigh the negatives. However, even with the high shielding configuration the signal-to-noise ratio is still low at ~ 0.1. With such a low value, detecting the Breit-Wheeler process will require a statistical analysis. This involves taking multiple null (without the x-ray field) and full (with the x-ray field) shots. Evidence of the Breit-Wheeler process is then given by a statistically significant difference in the number of particles measured on the null and full shots.

Using the expected values in tables 6.1, we can perform a rough estimate of the number of shots required to make this measurement. Assuming 10^9 primary electrons per shot, the number of particles detected on both the null and full shots are Poisson variables with a means $\mu_N = 16.64$ and $\mu_F = 16.64 + 1.72 = 18.36$ respectively. If we take N null and full shots each, then the standard error on our estimate of the mean is $\sigma = \sqrt{\mu/N}$. In figure 6.11, we can see the 2σ error for both μ_N and μ_F against N. These become significantly different with $N \sim 100$ shots, which is an achievable number over the duration of a Gemini campaign.

Table 6.1: Expected number of background particles and signal Breit-Wheeler pairs reaching the detectors. The error on the background estimates is due to the Monte-Carlo method. For the signal calculations, variance reduction methods were used to lower the error to a negligible level.

	Background leptons (10^{-9} / primary)	Background photons (10^{-7} / primary)	Breit-Wheeler pairs (10^{-9} / primary)
Low shielding	201.5 ± 6.3	651 ± 3	2.86
High shielding	16.6 ± 2.3	6.5 ± 0.4	1.72



(a) X-ray spectrum

(b) Electron spectrum

Figure 6.12: X-ray and electron spectra obtained on the experiment. (a) Shows the average x-ray spectra (red) as well as the uncertainty (blue shade). (b) shows the electron spectra (blue) for each shot as well as their average (red).

6.3 Post-Experiment Analysis

Numerical simulations are not only useful for pre-experiment optimisation, but they are also an invaluable tool for post-experimental analysis. By enabling a comparison between experimental measurements and theory, inference can be performed. In this section we will review the numerical calculations that were carried out in the post-experiment phase. We will start by updating the signal-to-noise ratio calculations of the Breit-Wheeler process, based on the electron and x-ray spectrum obtained during the experiment. From this, we will be able to estimate the likelihood with which the Breit-Wheeler process could be detected. We will then switch our attention to photon-photon scattering, using the experiment to find an upper limit on the cross-section.

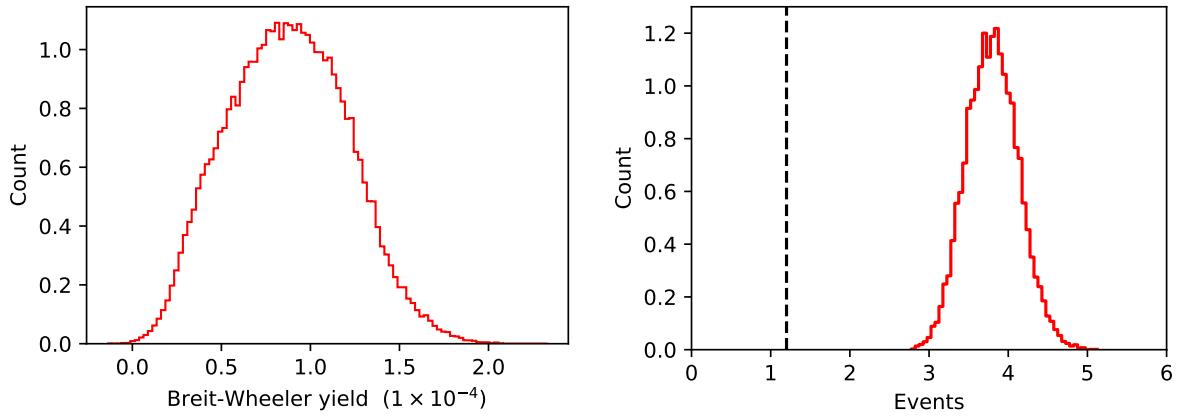
6.3.1 Signal-To-Noise Ratio of the Experiment

When performing the signal-to-noise calculation in section 6.2.4, a number of assumptions were made in regards to the input electron beam and x-ray field. On the experiment, diagnostics were set up to measure both, which we can use to update our signal-to-noise calculations. Figure 6.12 (a) shows the average x-ray spectrum as well as the uncertainty due to shot-to-shot fluctuations. Comparing this with figure 6.4 (a), we can see the measurement on the experiment has far more detail than the simulated spectrum. However, both are the same order of magnitude, so this is unlikely to drastically change the expected Breit-Wheeler yield.

It was not possible to take shot-to-shot measurements of the electron spectrum as the γ-ray converter destroyed the beam. Therefore, the beam path had to be cleared and only a limited number of measurements were taken immediately prior to each data run. The

Table 6.2: Measured variables and their uncertainty which affect the Breit-Wheeler signal-to-noise ratio.

Beam charge (pC)	X-ray density ($10^{13}\,\mathrm{mm}^{-3}$)	Divergence (mrad)	Pointing (mrad)
50 ± 7	9.0 ± 3.6	2.3 ± 0.3	0 ± 0.6



(a) Signal

Figure 6.13: Signal-to-noise ration calculations (a) Shows a distribution over the number of Breit-Wheeler pairs produced per shot. (b) shows the simulated number of background leptons (black-dashed) and experimentally observed background leptons (red).

resulting measurements are shown in figure 6.12 (b), as well as their average. Comparing this with figure 6.7, we can see the electron spectrum is significantly lower in energy than expected. On top of this, table 6.2 shows the average beam charge (corresponding to 3.1×10^8 primary electrons) was also lower than predicted pre-experiment. Therefore, unlike the x-rays, the updated electron beam will affect our signal-to-noise calculation.

Changes to the electron beam will affect both the background noise and Breit-Wheeler signal. Therefore, to update our signal-to-noise estimate, we should repeat both these calculations. All the measured parameters which affect these calculations are summarised in table 6.2. On top of the ones that have already been mentioned are the divergence and pointing of the electron beam. To include the shot-to-shot fluctuations in the signal calculations, we can use the Monte-Carlo method. This involves performing the calculation multiple times with different inputs sampled from table 6.2 assuming a Gaussian distribution on each variable. The result of this is shown in figure 6.13 (a). This updated expected value of $(0.9 \pm 0.3) \times 10^{-4}$ is far lower than the previous estimate of 1.7 pairs per shot. The main contribution to this reduction is the nonlinear scaling of Breit-Wheeler yield with electron beam energy. This is due to both the energy dependence of the bremsstrahlung process and the energy threshold behaviour of the Breit-Wheeler process.

To update the background noise calculation, the same Monte Carlo procedure can be

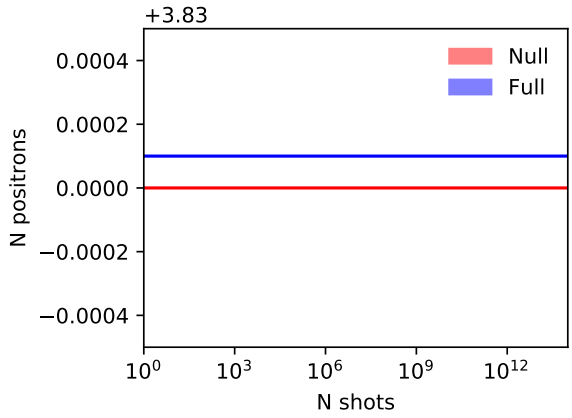


Figure 6.14: Diagram showing the 2σ confidence level of μ_N and μ_F as a function of the number of shots using the measured electron and x-ray spectrum

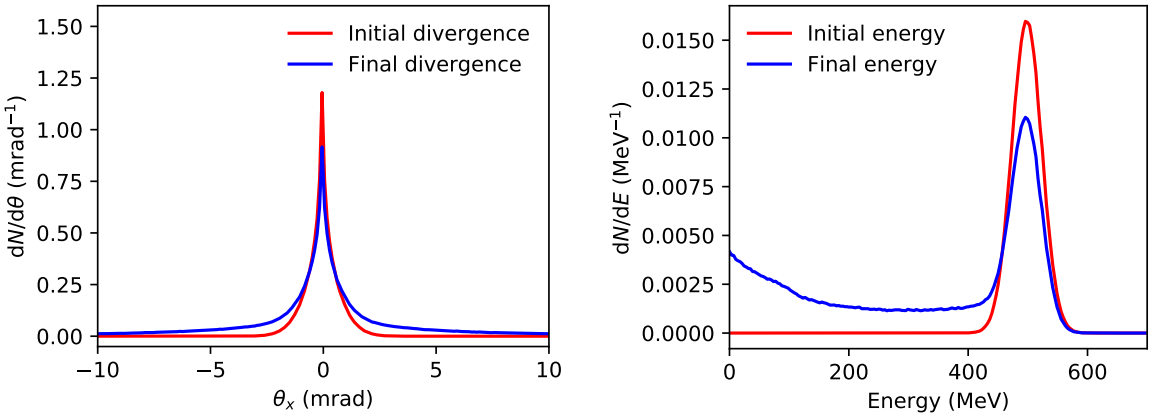
performed. However, as each calculation is computationally expensive (taking $\sim$ 1 day on a 500 core cluster) this is not practical. Instead we will just perform the calculation once using the average values in table 6.2. The resulting number of background positrons per shot is shown as a vertical dashed line in figure 6.13 (b). Also included in this plot is a histogram of the number of positrons observed by the TimePix3 detectors on the null shots of the experiment. The detected background is a factor of $\sim$ 4 times larger than the simulations predict. Possible reasons for this include γ-rays interacting with the TimePix3 detectors causing false positive readings and the simulation setup not being identical to the experimental setup.

Over the duration of the experiment $\sim$ 100 shots were obtained. Given an average detected background of 3.83 and an expected yield of 0.9×10^{-4} positrons per shot, it is highly unlikely that a measurement of the Breit-Wheeler process can be made. However, we can still perform the same analysis used in section 6.2.4 to estimate the number of shots required to make a 2σ confidence level measurement. This is demonstrated in figure 6.14, showing that the null and full shot means only become separated after $\sim 10^{10}$ shots. Using the Gemini shot rate of 0.05 Hz, obtaining this many shots would take $\sim$ 6300 years, far longer than was allocated for the experiment.

6.3.2 Bounding the Photon-Photon Scattering Cross-Section

Based on the analysis of the previous section, it is highly unlikely that the experiment will provide the first evidence of the Breit-Wheeler process. Therefore, the same is also true for photon-photon scattering, as the QED cross-section is several orders of magnitude smaller. However, as we discussed in the introduction to this chapter, the absence of photon-photon scattering is still an interesting result, as it allows us to put a bound on the cross-section.

If prolific photon-photon scattering had occurred, we would expect the presence of the



(a) Divergence

(b) Energy

Figure 6.15: γ-ray beam (a) divergence and (b) energy spectrum before and after interacting with a high density x-ray field. The Breit-Wheeler process was not included in these calculations.

x-ray field to produce two detectable signatures in the γ-ray beam. These are a change in both the divergence and energy spectrum, which are demonstrated in figures 6.15 (a) and (b) respectively. These plots show a γ-ray beams before and after interacting with a high density x-ray field. As this is an asymmetric interaction, on average a scattering event will increase the energy of the x-ray whereas the energy of the γ-ray will decrease. As well as this, the presence of scattering leads to an increase in the divergence of the beam.

As discussed in section 6.1, the experiment used two scintillator based detectors to measure both the profile and spectrum of the γ-ray beam. Therefore, if prolific photon-photon scattering occurred, we should be able to detect both the effects demonstrated in figure 6.15. If not, this suggests a limit on the photon-photon scattering cross-section. From figure 6.15, it appears that the effect on the energy spectrum is more significant than the divergence. Therefore, we expect the γ-ray spectrometer will be a better diagnostic for extracting a bound

In this section we will detail the steps required to go from the signal on the γ-ray spectrometer to a bound on the photon-photon scattering cross-section. This has involved developing a Bayesian spectral retrieval algorithm, based on the maximum likelihood approach by Behm *et al.* [208], which will also be discussed in this section.

Forward Model

The γ-ray spectrometer consisted of a stack of 33×47 CsI crystals doped with thallium and imaged with a 14-bit EMCCD camera. This was positioned inside the γ-ray beam path, with higher energy γ-rays penetrating deeper into the stack than lower energy γ-rays. Extracting the spectrum from this spectrometer is an inverse problem. To solve an

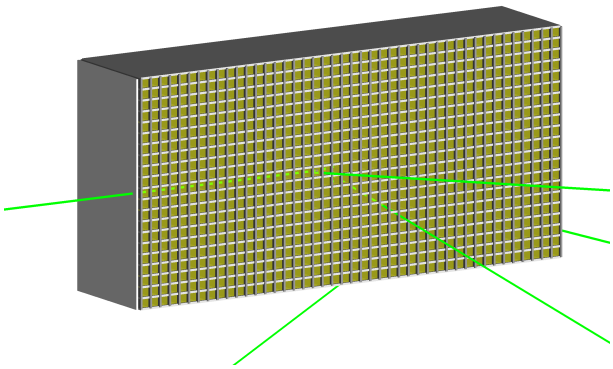


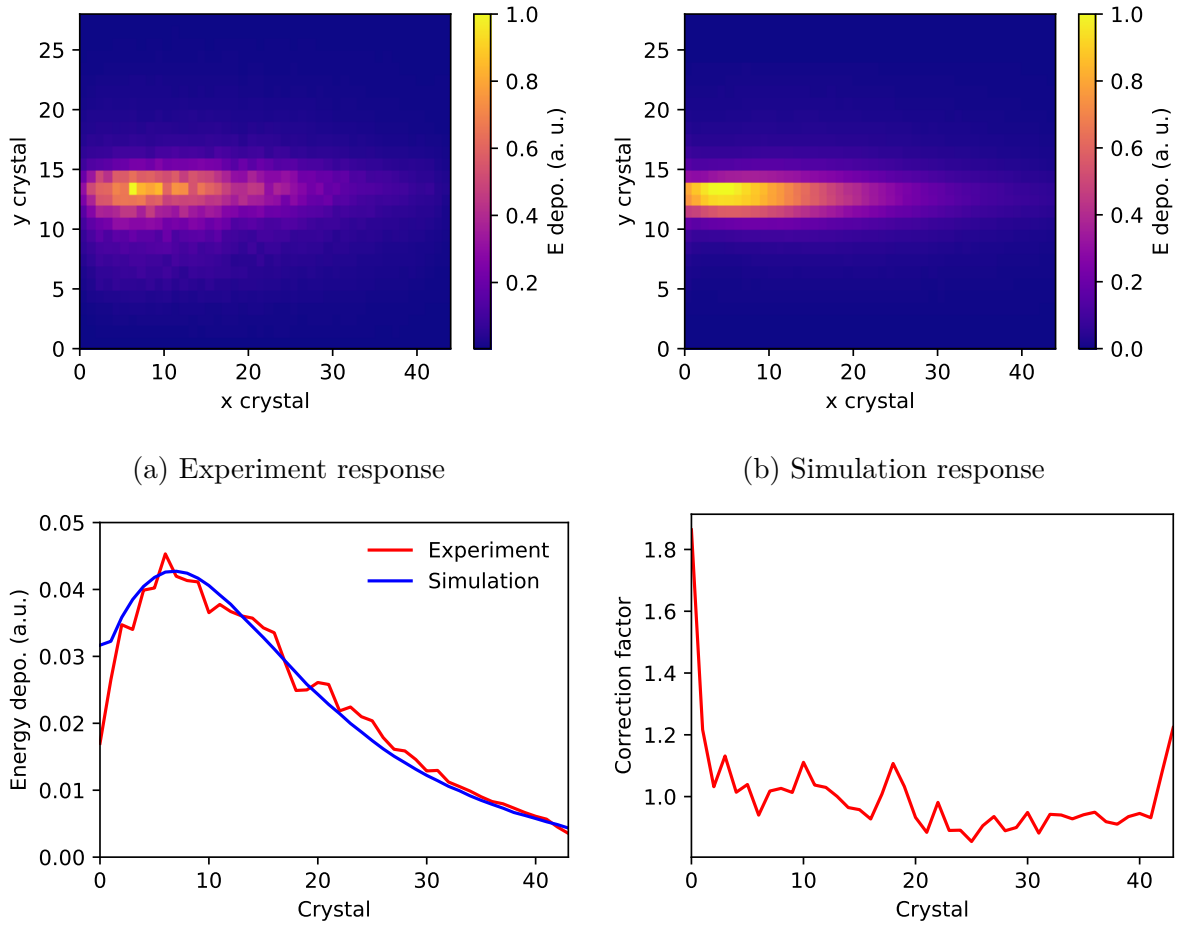
Figure 6.16: Visualisation of a Geant4 simulation consisting of a single high energy γ-ray, interacting with the spectrometer.

inverse problem, we require a forward model of the system, whose inputs can be varied until they agree with the measurements. In this case the input is the γ-ray spectrum and the measurement is the signal on the spectrometer. This type of interaction can be simulated using Geant4, as is demonstrated in figure 6.16.

Before we use Geant4 to build a forward model, we should test its ability to reproduce the experimental measurements. In figure 6.17 (a) we can see the spectrometer response, obtained by averaging over all the shots taken during the experiment[3]. In figure 6.17 (b) we can also see the spectrometer response from a Geant4 simulation, where the average electron spectrum in figure 6.12 has been used. These two images have roughly the same shape, however, the experiment response is both noisier and broader along the y-direction. The spread in the y-direction is likely due to the electron beam pointing fluctuation which is not included in the simulations. To remove this, we can integrate along the y-direction, producing the result seen in figure 6.17 (c). The increased noise on the experimental response is still present and is likely due to a systematic error, such as the crystals being misaligned. This systematic error can also be removed by introducing a correction factor for each crystal row. This correction factor is obtained by dividing the average simulated response by the average experimental response and is shown in figure 6.17 (d).

With the correction factor applied, the Geant4 simulations provide an accurate forward model of the system. However, they are too computationally expensive to be used in a Bayesian analysis due to the number of calls to the forward model. To solve this problem, we will use the same approach as Behm *et al.* [208]. This involves performing a limited number of Geant4 simulations with monoenergetic γ-ray beams of varying energy. From this, we can generate a crystal response function, $\rho_i(E)$, where i is the crystal index. A plot of $\rho_i(E)$ is shown in figure 6.18. From this the crystal response for some arbitrary

[3]This is not the direct readout from the camera. Postprocessing has been applied to find the energy deposited in each crystal. This work was carried out by E. Gerstmayr.



(a) Experiment response

(b) Simulation response

(c) Column integrated response

(d) Correction factor

Figure 6.17: Comparison between experimental and simulated spectrometer response. (a) shows the experimental response, (b) shows the simulated response, (c) shows the experimental and simulated response obtained by summing over the columns and (d) shows a crystal row correction factor obtained by dividing the simulated response by the experimental response.

γ-ray spectrum, $f(E)$, can be quickly calculated with the following integral

$$I_i[f] = C_i \int_0^\infty \rho_i(E) f(E) \mathrm{d}E. \tag{6.2}$$

where C_i is the correction factor given by figure 6.17 (d). This calculation can be performed orders of magnitude faster than a full Geant4 simulation, making it an effective forward model for Bayesian inference.

Bayesian Spectrum Estimation

The γ-ray spectrum, $f(E)$, that we want to estimate is a continuous function. However, in order to effectively apply Bayesian inference, we must find a low dimensional parameterisation of $f(E)$. To do this, we can perform Geant4 simulations of the bremsstrahlung

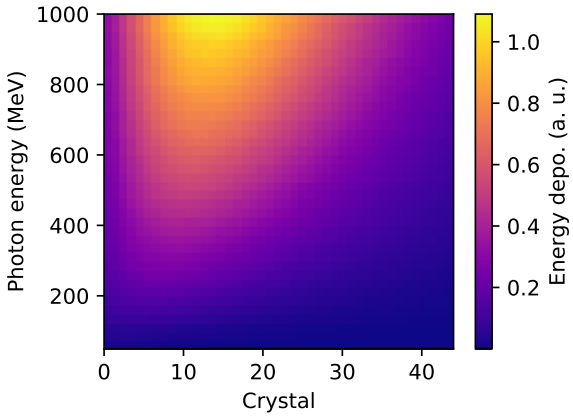


Figure 6.18: Crystal response function, $\rho_i(E)$, obtained by performing Geant4 simulations of monoenergetic γ-ray beams.

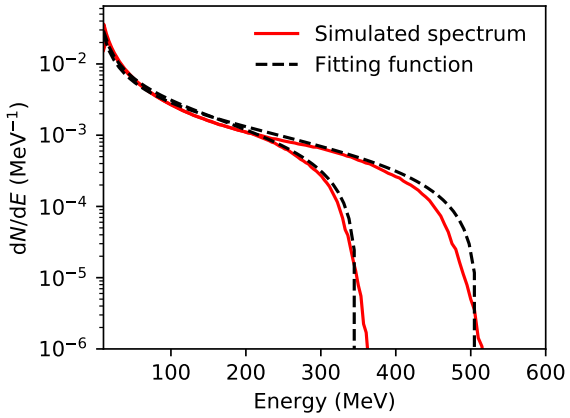


Figure 6.19: γ-ray spectra obtained from Geant4 simulations of the bremsstrahlung converter using the electron spectra in figure 6.12 (b) as inputs (red). Also plotted is the fitting function, equation 6.3, which agrees well with the simulated spectra.

converter, using the electron spectra in figure 6.12 (b), to get a data set of typical γ-ray spectra. From this data set, the following function was found to provide a good approximation to the spectra

$$f(E;\alpha,E_c) = \alpha\left(1 - 0.18\frac{E^2}{E_c^2}\right)E^{-0.94}. \tag{6.3}$$

Here, we have used the same parameters as Behm *et al.* [208], which are the mean energy of the γ-ray beam, E_c, and a scaling factor, α. The ability of equation 6.3 to recreate the simulated spectrum is shown in figure 6.19.

With this parameterisation, we can now obtain the γ-ray spectrum on each shot by applying Bayesian inference to estimate a distributions over $\mathbf{x} = (E_c, \alpha)$. This involves

applying Bayes' theorem

$$p(\mathbf{x} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y} \mid \mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x}}. \tag{6.4}$$

where $\mathbf{y}$ is an observed data point, corresponding to a vector of the crystal responses. In this equation, $p(\mathbf{y} \mid \mathbf{x})$ is the likelihood and $p(\mathbf{x}) = p(E_c)p(\alpha)$ is a prior which we must set.

If we make the assumption that the crystals exhibit random Gaussian noise, σ, we can write the likelihood function as

$$p(\mathbf{y} \mid \mathbf{x}) = \prod_i \frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{\Big(y_i - I_i[f(\mathbf{x})]\Big)^2}{\sigma^2} \right). \tag{6.5}$$

This introduces a new parameter, σ, which should be treated the same way as E_c and α. Given that we have little prior knowledge of E_c, α and σ, other than the fact that they cannot be negative, we will set uniform priors on each with a lower bound of zero. We know the upper bound for E_c cannot be greater than the maximum energy of the electrons ($\sim 800\,\mathrm{MeV}$) so the prior will be $p(E_c) = \mathcal{U}(0, 800\,\mathrm{MeV})$. Through appropriate normalisation of the data set, we can ensure that α is never greater than 10, allowing us to apply the prior $p(\alpha) = \mathcal{U}(0, 10)$. Finally, we will set the prior on σ to be $p(\sigma) = \mathcal{U}(0, 1)$ as if the limit is greater than this, it will be too noisy to make any inference.

With the likelihood and priors set, we can use equation 6.5 to calculate the posterior. Given that the numerator involves a three-dimensional integral, it is most efficiently solved using a Markov chain Monte Carlo (MCMC) method. In figure 6.20 we can see an example of this calculation performed on a random shot. Also shown are the marginal distributions $p(\alpha)$, $p(E_c)$ and $p(\sigma)$.

Bounding the Cross-section

Having developed a robust method for extracting the γ-ray spectrum from the crystal response, we can now test if the presence of the x-ray field has an effect on the γ-ray beam. To do this, we can run the Bayesian spectral retrieval algorithm on each shot of the experiment and compare the distributions over E_c for null and full shots. These two distributions are shown in figure 6.21, which appear similar. We can perform a statistical check to see if this is the case , using the two-sample Kolmogorov–Smirnov (KS) test. This involves calculating KS statistic, given by

$$D = \sup\Big|F_{\mathrm{N}}(E_c) - F_{\mathrm{F}}(E_c)\Big| \tag{6.6}$$

where $F_{\mathrm{N}}(E_c)$ and $F_{\mathrm{F}}(E_c)$ are the cumulative distribution functions for the null and full shots respectively. The null hypothesis of this test is that both null and full data sets have been sampled from the same distribution. This is accepted at the 95% confidence

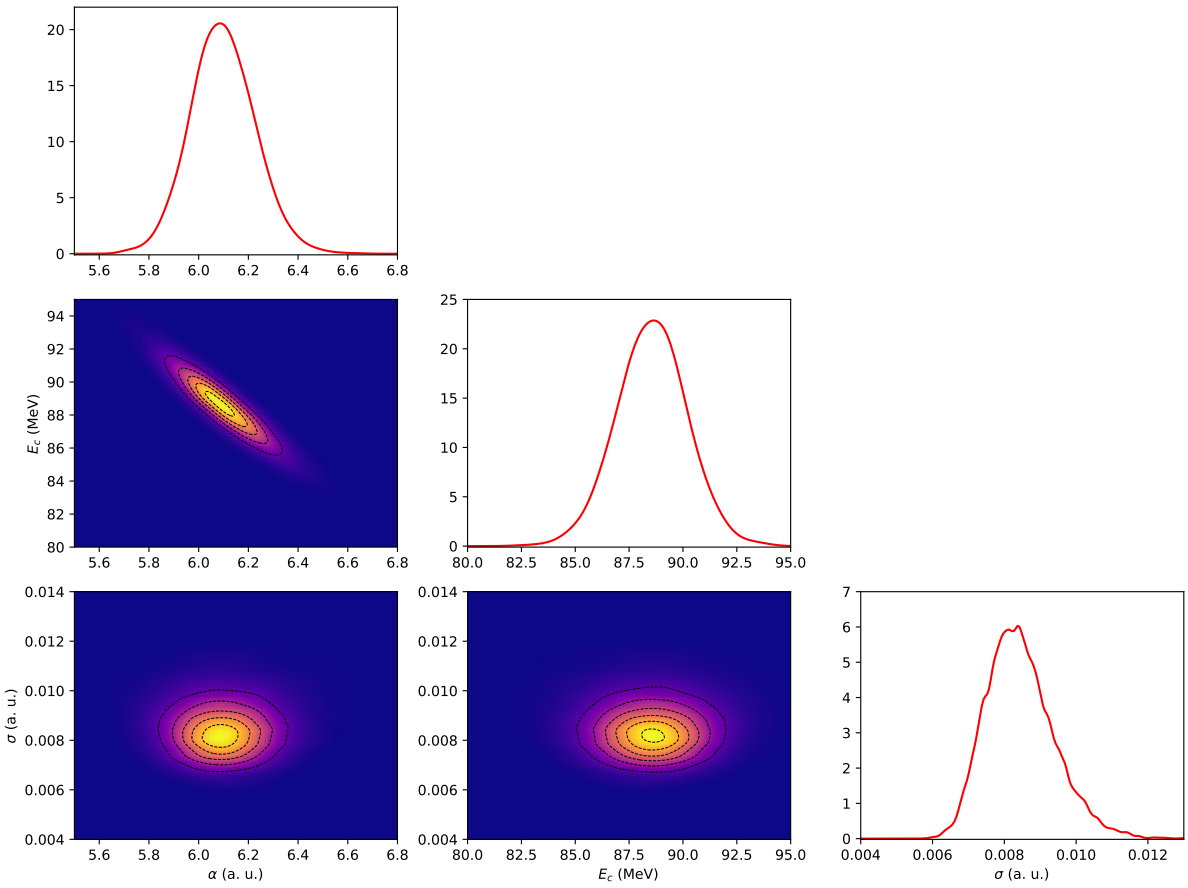
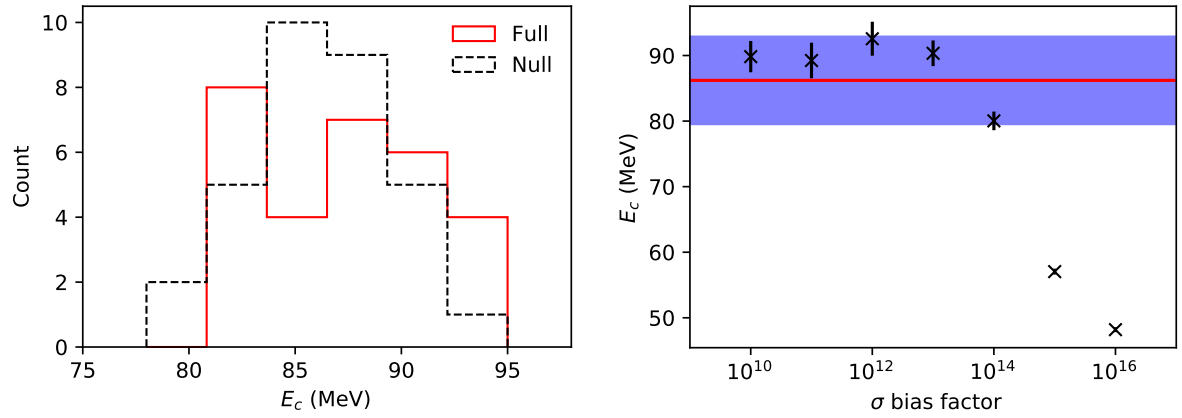


Figure 6.20: Marginal plots of the posterior distribution for a randomly sampled shot.



(a) E_c distribution

(b) E_c against bias factor

Figure 6.21: (a) shows distributions over E_c for both null and full shots. (b) shows the effect of increasing the cross-section bias factor on E_c. Also shown on this plot is the mean E_c from the null shots and a 95% confidence level.

level if $D < 1.358$ [209]. The value obtained for this data set is $D = 0.216$, so we can accept the null hypothesis.

As there is no difference between the two distributions, we can conclude there was

Table 6.3: Comparison of different photon-photon scattering experiments and the limits put on the cross-section.

	ω_{cm} (m_e)	σ (μb)	Bound (μb)	Bound / σ
Moulin (1996)	3.35×10^{-6}	1.6×10^{-34}	1×10^{-10}	6.2×10^{23}
Bernard (2000)	1.55×10^{-6}	1.8×10^{-36}	1.5×10^{-18}	5.5×10^{17}
Inada (2014)	1.29×10^{-2}	2.6×10^{-13}	1.7×10^{10}	6.4×10^{22}
Yamaji (2016)	1.29×10^{-2}	2.6×10^{-13}	1.9×10^{7}	7.2×10^{19}
This work	0.66	0.0049	4.9×10^{12}	1×10^{15}

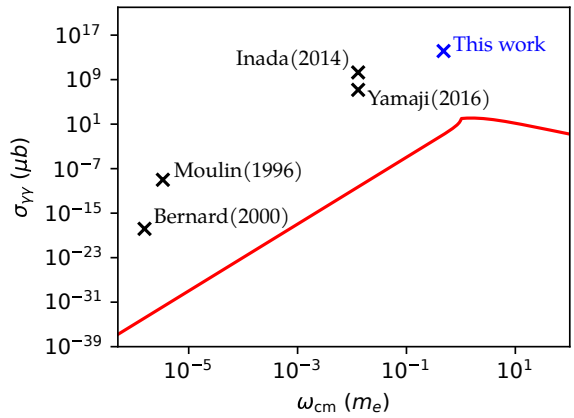


Figure 6.22: Photon-photon scattering cross-section with bounds placed by previous experiment[200, 201, 202, 203] and this work.

not a detectable level of photon-photon scattering. To find how much we would have to bias the cross-section by so that a difference would be detectable, we can perform multiple simulations of the experiment with an increasing cross-section. This is shown in figure 6.21 (b), which also includes the mean and 95% confidence limit of E_c for the null shots. Increasing the bias on the cross-section up to $\sim 10^{13}$ has little effect on E_c. Beyond this point, E_c starts to decrease, however, the first value to be fully outside the 95% confidence level is 10^{15}. This conservative bound corresponds to the lowest limit put on the photon-photon scattering cross-section to date.

Using equation 6.1 and taking the energy of the γ-ray to be $\omega_1 = 86.3$MeV and the energy of the x-ray to be $\omega_2 = 1.4\,$keV, the photon CM frame energy is $\sim 0.66\,m_e$. With this, we can update the plot in figure 6.2 to include this work, as shown in figure 6.22. With a two order of magnitude improvement, this is closest bound to the QED photon-photon scattering cross-section, to our knowledge,.

6.4 Summary

In this chapter we have presented the numerical simulations that were carried out to design, optimise and analyse a photon-photon collider experiment. The majority of these simulations used the Monte-Carlo platform Geant4. With the addition of the photon-physics package developed for this purpose (discussed in chapter 4), both signal and noise calculations could be performed within this framework.

This capability was demonstrated in section 6.2, where full start-to-end simulations of the experiment were performed. This provided a platform to optimise various aspects of the experiment, starting with the photon sources. With these optimised, background noise sources were identified and removed through an iterative process over different shielding configurations. This lead to a significant reduction in the expected levels of background noise. Although this shielding configuration reduced the expected signal, the reduction in background noise was far more significant. With previously achieved electron beam parameters, these simulations suggest that a statistically significant detection of the Breit-Wheeler process can be made at Gemini with ~ 100 shots.

Unfortunately, during the experimental campaign, the LWFA electron beams were far from optimum, resulting in both a lower beam charge and energy spectrum. This was found to have more of a detrimental effect on the Breit-Wheeler signal than the background noise resulting in a far lower signal-to-noise ratio. With an expected signal of $\sim 10^{-4}$ positrons per shot and a detected background of ~ 4 positrons per shot, it is highly unlikely that any observation has been made.

The experiment may not have been successful in its primary goal, however, the data was still useful in enabling the photon-photon scattering cross-section to be bounded. Previous work in this area has focused on high density sources with a lower photon energies. At low energies the cross-section is severely suppressed, resulting in the previous most stringent bound being ~ 20 orders of magnitude above the value predicted by QED [200]. In this experiment the average CM energy was $\omega_{\mathrm{CM}} = 0.66\, m_e$, close to the peak value of the QED cross-section. By looking for shifts in the average energy of the γ-ray spectrum, an upper limit on the cross-section of $\sigma < 4.9 \times 10^{12} \mu$b was made with a 95% confidence limit. We believe this is most stringent bound on the QED cross-section to date and is placed at an astrophysical relevant CM energy, just below the Breit-Wheeler threshold.

Although the method we used here has provided the most stringent bound on the photon-photon scattering cross-section to date, we are aware of its limitations. For the process to actually be detected, the x-ray density would have to increase by many orders of magnitude. This is ultimately a limitation of the asymmetric collider setup, in which the scattered photons are emitted almost parallel to the dominant photon axis. Any measurement then relies on photon-photon scattering causing a bulk change to the γ-ray beam

properties. For these reasons a symmetric collider, with collimated and monoenergetic photon sources, seems like the most appropriate setup for such a measurement.

The analysis presented here shows the Breit-Wheeler process can be detected using the Gemini laser facility. However, this will rely on the facility operating at peak performance for a large number of shots. Therefore, if the experiment is to be carried out again, improvements should be made to increase the signal-to-noise ratio. This will both increase the chance of detecting the Breit-Wheeler process and allow a more stringent bound to be placed on the photon-photon scattering cross-section. One method of increasing the signal is to bring the two photon sources closer together, increasing the densities at the interaction zone. As shown in figure 6.9, the distance between the γ-ray converter and x-ray foil is $\sim 800\,\mathrm{mm}$. This could be reduced by simply removing the space between the converter, the tungsten block and the collimator. It could then be reduced further by using a shorter residual electron magnet. To maintain the same line integrated magnetic field (and and electron dispersion) the field strength of the magnet would have to increase. This could be achieved using a pulsed electromagnet, capable of generating a far higher field strength compared to a permanent magnet. A second method of increasing the signal is reducing the duration of the x-ray driver. With the current setup, a large fraction of the x-rays cannot participate in interactions as they are emitted from the foil well before or after the γ-ray beam passes over. For a fixed laser energy, decreasing the duration reduces the number of unused x-rays and increases the peak field density.

Further optimisation of the setup should also take place to reduce the background noise further. Figure 6.9 shows that two of the main noise sources are the residual electron magnet and the beam dump aperture. Both these noise sources could also be reduced by using a shorter, pulsed electromagnet rather than a permanent magnet. The noise source at the residual electron magnet is due to divergent gamma rays which scatter off the magnet edge. With a shorter magnet the area of potential scattering sites is reduced. The noise generated at the beam dump aperture is due to high energy electrons which experience little deflection in the magnetic field. With a higher strength field these could be sufficiently deflected into the beam dump.

Although we believe these changes will enable the Breit-Wheeler process to be detected using the Gemini laser, it may be more optimal to carry out the experiment at a different facility. Two desirable upgrades are an increase in laser energy or an increase repetition rate. For example, using a higher energy system such as the national ignition facility (NIF), a far higher x-ray density can be obtained. As was shown in section 4.4, this increases the expected Breit-Wheeler yield by many orders of magnitude, enabling a measurement to be taken with only a few shots. An example of a high repetition rate facility, which would be ideal for carrying out this experiment, is the the Extreme Photonics Applications Centre (EPAC) at the CLF. This facility is currently under development and will house a 10 Hz, petawatt class laser (built to replace the Gemini laser). This will

have a similar peak power to Gemini, however, with an increase in repetition rate by a factor of 200, the experiment can be carried out in a fraction of the time.

Chapter 7

Conclusion

QED is the most stringently tested theory in physics to date. However, there still remains many QED processes for which little to no experimental evidence exists. In recent years, the exponential growth of peak obtainable laser intensities has opened up the opportunity to probe some of these processes in the laboratory. However, these experiments often have a low interaction probability, and as a result, suffer from a low signal-to-noise ratio. For this reason, performing detailed numerical simulations of experiments is vital to optimise both the signal-to-noise ratio and probability of a successful measurement. Therefore, the work in this thesis has been concerned with developing the capability to perform these simulations.

As part of this work, we have developed two Monte Carlo QED packages. The first uses the framework of Geant4 and models two-body scattering processes, including the Breit-Wheeler process, photon-photon scattering and Compton scattering. To seamlessly integrate the package with Geant4, the interactions are treated as a beam of particles travelling through a static photon field. This limits the types of photon sources that can be modelled, however, also reduces the computational complexity of calculations. Due to the object-oriented design, extending the package with new processes would be a straightforward task. In future work, we plan to implement processes involving axion-like [199] and minicharged particles [196], to study the potential of detecting beyond standard model physics using a photon-photon collider.

The second package models QED interactions in a strong background field, and includes the nonlinear Breit-Wheeler processes and nonlinear Compton scattering. Although the algorithm we have used is similar to a number of QED-PIC codes, our package offers a number of advantages. At the intensities that are available by lasers today, the production and dynamics of particles is unlikely to have a detectable impact on the background field. Therefore, when modelling these interactions, accounting for the action of the particles on the classical field (*i.e.* solving Maxwell's equations as a PIC code does) is unnecessary and computationally expensive. On top of this, a PIC code is not the correct platform to model many aspects of an experiment, such as particle tracking and detector

interactions. By developing a strong field QED event generator, we have integrated the package with Geant4, enabling many aspects of an experiment to be modelled within a single framework. We believe this package will act as a useful tool for the design and analysis of future experiments. For example, we have already applied it to study the background noise of an upcoming strong field QED experiment at FACET-II, SLAC.

A key aspect of both the packages developed in this work are machine learning algorithms, introduced to increase the efficiency of calculations. For the particle-photon package, a Gaussian process regression was used to replace the expensive mean free path calculation. This is a Bayesian regression method, providing the ability to carry out uncertainty quantification. Using this we have developed an algorithm that minimises the number of expensive calculations, given a predefined error. We believe this algorithm could be applied to other computational physics problems, replacing low fidelity parameterisations or lookup tables of complex physics with an in-line solver and regression. Given the stochastic nature of the strong field QED package, it would have also been desirable to use a Bayesian regression method. However, this was not possible due to the large size of the training data set and the likelihood being non-Gaussian. Therefore, we instead opted for a mixture density network, consisting of a deep feedforward neural network and a Gaussian mixture model. This method has the flexibility to model complex distributions and also scales well to large data sets. Both machine learning methods were found to provide a large increase in efficiency and we believe their use will become more common in computational physics.

In the final part of this work, we presented the design and analysis of a photon-photon collider experiment. Both signal and noise calculations were performed within the Geant4 framework, made possible by the particle-photon package developed for this purpose. During the experiment, the electron beam energies were significantly lower than have previously been obtained using the Gemini laser. Due to the nonlinear scaling of the Breit-Wheeler yield with electron energy, the expected signal-to-noise ratio of the experiment was low at $\sim 2 \times 10^{-5}$. This makes it highly unlikely that the experiment was successful in its primary aim of detecting the Breit-Wheeler process.

However, we were still able to carry out the secondary objective of the experiment, putting a bound on the photon-photon scattering cross-section. Previous experiments that have attempted to do this have operated at a far lower CM energy. At low CM energies, the cross-section is highly suppressed, resulting in a bound that is many orders of magnitude off the predicted QED cross-section. In our experiment, the average CM energy was $\omega_{\mathrm{CM}} \approx 0.66\, m_e$, close to the peak of the cross-section. No signatures of photon-photon scattering were detected, enabling a bound to be placed of $4.9 \times 10^{12}\, \mu$b. Although this is still 15 orders of magnitude off the predicted QED cross-section, it is the most stringent bound placed to date, improving upon previous work by a factor of ~ 1000.

Although present day experiments are still far away from detecting photon-photon scattering, our results suggest that if optimum electron beams are obtained a Breit-Wheeler detection experiment can be carried out using the Gemini laser facility. Therefore, it may soon be possible to do what John Breit and Gregory Wheeler once claimed as "hopeless" and observe pair formation in laboratory experiments.

Bibliography

[1] Carlson, J. B. Lodestone compass: Chinese or olmec primacy? *Science*, 189, 1975.

[2] Prothero, D. R. *The Story of the Earth in 25 Rocks*. 2018.

[3] Priestley, J. *The History and Present State of Electricity*. 1767.

[4] de Coulomb, C.-A. Premier mémoire sur l'électricité et le magnétisme. *Histoire de l'Académie Royale des Sciences*, pages 569–577, 1785.

[5] Mitolo, M. and Araneo, R. A brief history of electromagnetism, 2019.

[6] Faraday, M. V. experimental researches in electricity. *Philosophical Transactions of the Royal Society of London*, 122, 12 1832.

[7] Maxwell, J. C. On physical lines of force. *Philosophical Magazine*, 90, 1861.

[8] Einstein, A. Über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt. *Annalen der Physik*, 322, 1905.

[9] Schrödinger, E. Quantisierung als eigenwertproblem. *Annalen der Physik*, 384, 1926.

[10] Dirac, P. A. M. The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 117, 1928.

[11] Dirac, P. A. M. A theory of electrons and protons. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 126, 1930.

[12] Anderson, C. D. The apparent existence of easily deflectable positives. *Science*, 76, 1932.

[13] H., B. and W., H. On the stopping of fast particles and on the creation of positive electrons. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 146, 1934.

[14] Oppenheimer, J. R. and Plesset, M. S. On the production of the positive electron, 1933.

[15] Sauter, F. Über die bremsstrahlung schneller elektronen. *Annalen der Physik*, 412, 1934.

[16] Racah, G. Sulla nascita degli elettroni positivi. *Il Nuovo Cimento*, 11, 1934.

[17] Landau, L. D. and Lifshitz, E. On the production of electrons and positrons by a collision of two particles. *Phys. Z. Sowjet*, 6:244–undefined, 1934.

[18] Breit, G. and Wheeler, J. A. Collision of two light quanta. *Physical Review*, 46, 1934.

[19] Sauter, F. Über das verhalten eines elektrons im homogenen elektrischen feld nach der relativistischen theorie diracs. *Zeitschrift für Physik*, 69, 1931.

[20] Schwinger, J. On gauge invariance and vacuum polarization. *Physical Review*, 82, 1951.

[21] Euler, H. and Kockel, B. Über die streuung von licht an licht nach der diracschen theorie. *Die Naturwissenschaften*, 23, 1935.

[22] Heisenberg, W. and Euler, H. Folgerungen aus der diracschen theorie des positrons. *Zeitschrift für Physik*, 98, 1936.

[23] Tomonaga, S. On a relativistically invariant formulation of the quantum theory of wave fields. *Progress of Theoretical Physics*, 1, 1946.

[24] Feynman, R. P. Space-time approach to non-relativistic quantum mechanics. *Reviews of Modern Physics*, 20, 1948.

[25] Feynman, R. P. Space-time approach to quantum electrodynamics. *Physical Review*, 76, 1949.

[26] Feynman, R. P. The theory of positrons. *Physical Review*, 76, 1949.

[27] Schwinger, J. Quantum electrodynamics. i. a covariant formulation. *Physical Review*, 74, 1948.

[28] Schwinger, J. Quantum electrodynamics. ii. vacuum polarization and self-energy. *Physical Review*, 75, 1949.

[29] Schwinger, J. Quantum electrodynamics. iii. the electromagnetic properties of the electron-radiative corrections to scattering. *Physical Review*, 76, 1949.

[30] Dyson, F. J. The s matrix in quantum electrodynamics. *Physical Review*, 75, 1949.

[31] Dyson, F. J. The radiation theories of tomonaga, schwinger, and feynman. *Physical Review*, 75, 1949.

[32] Odom, B., Hanneke, D., D'urso, B., and Gabrielse, G. New measurement of the electron magnetic moment using a one-electron quantum cyclotron. *Physical Review Letters*, 97, 2006.

[33] Piran, T. The physics of gamma-ray bursts, 2004.

[34] Uzdensky, D. A. and Rightley, S. Plasma physics of extreme astrophysical environments, 2014.

[35] Harding, A. K. and Lai, D. Physics of strongly magnetized neutron stars. *Reports on Progress in Physics*, 69, 2006.

[36] Blandford, R. D. and Znajek, R. L. Electromagnetic extraction of energy from kerr black holes. *Monthly Notices of the Royal Astronomical Society*, 179, 1977.

[37] Daugherty, J. K. and Harding, A. K. Pair production in superstrong magnetic fields. *The Astrophysical Journal*, 273, 1983.

[38] Daugherty, J. K. and Lerche, I. Theory of pair production in strong electric and magnetic fields and its applicability to pulsars. *Physical Review D*, 14, 1976.

[39] Erber, T. High-energy electromagnetic conversion processes in intense magnetic fields. *Reviews of Modern Physics*, 38, 1966.

[40] Mignani, R. P., Testa, V., Caniulef, D. G., et al. Evidence for vacuum birefringence from the first optical-polarimetry measurement of the isolated neutron star RX j1856.5-3754. *Monthly Notices of the Royal Astronomical Society*, 465, 2017.

[41] Nikishov, A. I. Absorption of high-energy photons in the universe. *Zhur. Eksptl'. i Teoret. Fiz.*, 42, 1961.

[42] Gould, R. J. and Schréder, G. P. Opacity of the universe to high-energy photons. *Physical Review*, 155, 1967.

[43] Aleksić, J., Antonelli, L. A., Antoranz, P., et al. MAGIC observations and multiwavelength properties of the quasar 3C 279 in 2007 and 2009. *Astronomy and Astrophysics*, 530, 2011.

[44] Strickland, D. and Mourou, G. Compression of amplified chirped optical pulses. *Optics Communications*, 56, 1985.

[45] Mourou, G. Nobel lecture: Extreme light physics and application. *Reviews of Modern Physics*, 91, 2019.

[46] Mangles, S. P. D., Murphy, C. D., Najmudin, Z., et al. Monoenergetic beams of relativistic electrons from intense laser-plasma interactions. *Nature*, 431, 2004.

[47] Faure, J., Glinec, Y., Pukhov, A., et al. A laser-plasma accelerator producing monoenergetic electron beams. *Nature*, 431, 2004.

[48] Geddes, C. G. R., Toth, C. S., Tilborg, J. V., et al. A laser – plasma accelerator producing monoenergetic electron beams. *Nature*, 431, 2004.

[49] Gitomer, S. J., Jones, R. D., Begay, F., et al. Fast ions and hot electrons in the laser–plasma interaction. *Physics of Fluids*, 29, 1986.

[50] Fews, A. P., Norreys, P. A., Beg, F. N., et al. Plasma ion emission from high intensity picosecond laser pulse interactions with solid targets. *Physical Review Letters*, 73, 1994.

[51] Beg, F. N., Bell, A. R., Dangor, A. E., et al. A study of picosecond laser-solid interactions up to 1019 w cm-2. *Physics of Plasmas*, 4, 1997.

[52] Kneip, S., McGuffey, C., Martins, J. L., et al. Bright spatially coherent synchrotron x-rays from a table-top source. *Nature Physics*, 6, 2010.

[53] Sarri, G., Corvan, D. J., Schumaker, W., et al. Ultrahigh brilliance multi-MeV -ray beams from nonlinear relativistic thomson scattering. *Physical Review Letters*, 113, 2014.

[54] Sarri, G., Poder, K., Cole, J. M., et al. Generation of neutral and high-density electron-positron pair plasmas in the laboratory. *Nature Communications*, 6, 2015.

[55] Bula, C., McDonald, K. T., Prebys, E. J., et al. Observation of nonlinear effects in compton scattering. *Physical Review Letters*, 76, 1996.

[56] Burke, D. L., Field, R. C., Horton-Smith, G., et al. Positron production in multi-photon light-by-light scattering. *Physical Review Letters*, 79, 1997.

[57] Cole, J. M., Behm, K. T., Gerstmayr, E., et al. Experimental evidence of radiation reaction in the collision of a high-intensity laser pulse with a laser-wakefield accelerated electron beam. *Physical Review X*, 8, 2018.

[58] Poder, K., Tamburini, M., Sarri, G., et al. Experimental signatures of the quantum nature of radiation reaction in the field of an ultraintense laser. *Physical Review X*, 8, 2018.

[59] Hogan, M. Preliminary conceptual design report for the FACET-ii project at SLAC national accelerator laboratory, 2016.

[60] Burkart, F., Aß{}mann, R., Brinkmann, R., et al. LUXE - a QED Experiment at the European XFEL. *10th International Particle Accelerator Conference*, 6 2019.

[61] Abramowicz, H., Altarelli, M., Aßmann, R., et al. Letter of intent for the luxe experiment, 2019.

[62] Eckhardt, R. Stan Ulam, John von Neumann, and the Monte Carlo method. *Los Alamos Science*, 15, 1987.

[63] Ulam, S., Richtmeyer, R. D., and von Neumann, J. Statistical methods in neutron diffusion, 1947.

[64] Metropolis, N. The beginning of the Monte Carlo method. *Los Alamos Science (1987 Special Issue dedicated to Stanislaw Ulam)*, 1987.

[65] U.S. Information Agency. Photograph of world's first computer, the electronic numerical integrator and calculator, series: Master file photographs of U.S. and foreign personalities, world events, and american economic, social, and cultural life, ca. 1953 - ca. 1994, `https://catalog.archives.gov/id/594262`, Date accessed: 28/04/21.

[66] Frank Rosenblatt with the Mark I Perceptron computer in 1960, `https://i.redd.it/bo3xoa8y4m341.jpg`. Date accessed: 28/04/21.

[67] Agostinelli, S., Allison, J., Amako, K., et al. GEANT4 - a simulation toolkit. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506, 2003.

[68] Allison, J., Amako, K., Apostolakis, J., et al. Geant4 developments and applications. *IEEE Transactions on Nuclear Science*, 53, 2006.

[69] Allison, J., Amako, K., Apostolakis, J., et al. Recent developments in Geant4. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 835, 2016.

[70] J. Ferrari, A. Sala, P. F. A. R. FLUKA: A multi-particle transport code (program version 2005). *Cern-2005-010*, 2005.

[71] Böhlen, T. T., Cerutti, F., Chin, M. P., et al. The FLUKA code: Developments and challenges for high energy and medical applications. *Nuclear Data Sheets*, 120, 2014.

[72] Forster, R. A. and Godfrey, T. N. K. MCNP - a general monte carlo code for neutron and photon transport, 2006.

[73] Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3, 1959.

[74] Turing, A. M. Computing machinery and intelligence. *Mind*, 59, 1950.

[75] Rosenblatt, F. The perceptron - a perceiving and recognizing automaton, 1 1957.

[76] Bishop, C. M. *Pattern Recoginiton and Machine Learning*. 2006.

[77] Minsky, M. and Papert, S. Perceptrons: expanded edition. *MIT Press Cambridge MA*, 522, 1969.

[78] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. *Nature*, 323, 1986.

[79] Rasmussen, C. E. and Williams, C. K. I. *Gaussian processes for machine learning. 2006*, volume 38. 2006.

[80] Jackson, J. D. *Classical Electrodynamics, Second Edition*. 1975.

[81] Peskin, M. and Schroeder, D. *An Introduction To Quantum Field Theory*. 1995.

[82] Goldstein, H. Classical mechanics. boston. *Book*, 1980.

[83] Landau, L. D. and Lifshitz, E. M. *The Classical Theory of Fields*. Addison-Wesley Publishing Company, 1971.

[84] Longair, M. S. *High Energy Astrophysics*. 2010.

[85] Larmor, J. Lxiii. on the theory of the magnetic influence on spectra; and on the radiation from moving ions. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 44, 1897.

[86] Liénard, A. *Champ électrique et magnétique produit par une charge électrique concentrée en un point et animée d'un mouvement quelconque*. G. Carré et C. Naud, 1898.

[87] Lau, Y. Y., He, F., Umstadter, D. P., and Kowalczyk, R. Nonlinear thomson scattering: A tutorial. volume 10, 2003.

[88] Sarachik, E. S. and Schappert, G. T. Classical theory of the scattering of intense laser radiation by free electrons. *Physical Review D*, 1, 1970.

[89] Esarey, E., Ride, S. K., and Sprangle, P. Nonlinear Thomson scattering of intense laser pulses from beams and plasmas. *Physical Review E*, 48, 1993.

[90] Olausen, S. A. and Kaspi, V. M. The mcgill magnetar catalog. *Astrophysical Journal, Supplement Series*, 212, 2014.

[91] Abraham, M. Prinzipien der dynamik des elektrons. *Annalen der Physik*, 315, 1902.

[92] Dirac, P. A. M. Classical theory of radiating electrons. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 167, 1938.

[93] Burton, D. A. and Noble, A. On the entropy of radiation reaction. *Physics Letters, Section A: General, Atomic and Solid State Physics*, 378, 2014.

[94] Higuchi, A. and Martin, G. D. Classical and quantum radiation reaction for linear acceleration. volume 35, 2005.

[95] Srednicki, M. *Quantum Field Theory*. Cambridge University Press, 2007.

[96] Jauch, J. M. and Rohrlich, F. *The Theory of Photons and Electrons*. 2 edition, 1980.

[97] Compton, A. H. A quantum theory of the scattering of X-rays by light elements. *Physical Review*, 21, 1923.

[98] Klein, O. and Nishina, T. Über die streuung von strahlung durch freie elektronen nach der neuen relativistischen quantendynamik von dirac. *Zeitschrift für Physik*, 52, 1929.

[99] Tollis, B. D. The scattering of photons by photons. *Il Nuovo Cimento Series 10*, 35, 1965.

[100] Lundström, E., Brodin, G., Lundin, J., et al. Using high-power lasers for detection of elastic photon-photon scattering. *Physical Review Letters*, 96, 2006.

[101] Hanneke, D., Fogwell, S., and Gabrielse, G. New measurement of the electron magnetic moment and the fine structure constant. *Physical Review Letters*, 100, 2008.

[102] MacKenroth, F. and Piazza, A. D. Nonlinear compton scattering in ultrashort laser pulses. *Physical Review A - Atomic, Molecular, and Optical Physics*, 83, 2011.

[103] Furry, W. H. On bound states and scattering in positron theory. *Physical Review*, 81, 1951.

[104] Bohr, N. Über die serienspektra der elemente. *Zeitschrift für Physik*, 2, 1920.

[105] Volkov, D. M. Über eine klasse von lösungen der diracschen gleichung. *Zeitschrift für Physik*, 94, 1935.

[106] Ritus, V. I. Quantum effects of the interaction of elementary particles with an intense electromagnetic field. *Journal of Soviet Laser Research*, 6, 1985.

[107] Ilderton, A. Trident pair production in strong laser pulses. *Physical Review Letters*, 106, 2011.

[108] King, B. and Ruhl, H. Trident pair production in a constant crossed field. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 88, 2013.

[109] Seipt, D. and Kämpfer, B. Two-photon compton process in pulsed intense laser fields. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 85, 2012.

[110] Mackenroth, F. and Piazza, A. D. Nonlinear double compton scattering in the ultrarelativistic quantum regime. *Physical Review Letters*, 110, 2013.

[111] Chistyakov, M. V., Kuznetsov, A. V., and Mikheev, N. V. Photon splitting above the pair creation threshold in a strong magnetic field. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 434, 1998.

[112] Nikishov, A. I. and Ritus, V. I. Quantum processes in the field of a plane electromagnetic wave and in a constant field. *Soviet Physics JETP*, 19, 1964.

[113] Baier, V. N., Katkov, V. M., and Strakhovenko, V. M. An operator approach to quantum electrodynamics in external field: Mass operator. *Zh. Eksp. Teor. Fiz.*, 67:453, 1974.

[114] Brown, L. S. and Kibble, T. W. Interaction of intense laser beams with electrons. *Physical Review*, 133, 1964.

[115] Ivanov, D. S. and Zhigilei, L. V. Combined atomistic-continuum model for simulation of laser interaction with metals: Application in the calculation of melting thresholds in Ni targets of varying thickness. volume 79, 2004.

[116] Harvey, C., Heinzl, T., and Ilderton, A. Signatures of high-intensity Compton scattering. *Physical Review A - Atomic, Molecular, and Optical Physics*, 79, 2009.

[117] Blackburn, T. G. Qed effects in laser-plasma interactions. *Doctoral dissertation, Oxford University, UK*, 2015.

[118] Mackenroth, F., Kumar, N., Neitz, N., and Keitel, C. H. Nonlinear Compton scattering of an ultraintense laser pulse in a plasma. *Physical Review E*, 99, 2019.

[119] Kirk, J. G., Bell, A. R., and Arka, I. Pair production in counter-propagating laser beams. *Plasma Physics and Controlled Fusion*, 51, 2009.

[120] Duclous, R., Kirk, J. G., and Bell, A. R. Monte Carlo calculations of pair production in high-intensity laser-plasma interactions. *Plasma Physics and Controlled Fusion*, 53, 2011.

[121] Shen, C. S. and White, D. Energy straggling and radiation reaction for magnetic bremsstrahlung. *Physical Review Letters*, 28, 1972.

[122] Baier, V. N., Katkov, V. M., and Strakhovenko, V. M. *Electromagnetic Processes at High Energies in Oriented Single Crystals*. 1998.

[123] Goldstein, M. Subjective Bayesian analysis: Principles and practice. *Bayesian Analysis*, 1, 2006.

[124] Wasserman, L. *All of Statistics : A Concise Course in Statistical Inference*. Springer, 2004.

[125] Wishart, J. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A, 1928.

[126] Frome, E. L. The analysis of rates using Poisson regression models. *Biometrics*, 39, 1983.

[127] Burnham, K. P. and Anderson, D. R. *Model Selection and Multimodel Inference*. Springer-Verlag New York, 2 edition, 2004.

[128] Hoerl, A. E. and Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 1970.

[129] Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 1996.

[130] Stone, M. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36, 1974.

[131] Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 1974.

[132] Schwarz, G. Estimating the dimension of a model. *The Annals of Statistics*, 6, 1978.

[133] LeNail, A. Nn-svg: Publication-ready neural network architecture schematics. *Journal of Open Source Software*, 4, 2019.

[134] Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning*. Springer New York Inc., 2001.

[135] Ripley, B. D. *Pattern recognition and neural networks*. 1996.

[136] Kandasamy, K., Neiswanger, W., Schneider, J., et al. Neural architecture search with bayesian optimisation and optimal transport. volume 2018-December, 2018.

[137] Abadi, M., Agarwal, A., Barham, P., et al. Tensorflow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[138] Duchi, J., Hazan, E., and Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12, 2011.

[139] Kingma, D. P. and Ba, J. L. Adam: A method for stochastic optimization. 2015.

[140] Berger, J. O. *Statistical Decision Theory and Bayesian Analysis. Second Edition.* 2 edition, 1985.

[141] Doob, J. L. *Stochastic Processes.* Wiley, 1953.

[142] Brooks, S., Gelman, A., Jones, G. L., and Meng, X. L. *Handbook of Markov Chain Monte Carlo.* 2011.

[143] Matsumoto, M. and Nishimura, T. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. volume 8, 1998.

[144] Robert, C. P. and Casella, G. *Monte Carlo Statistical Methods.* 1999.

[145] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., et al. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21, 1953.

[146] Hastings, W. K. Monte carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 1970.

[147] Geman, S. and Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, 1984.

[148] Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. Hybrid Monte Carlo. *Physics Letters B*, 195, 1987.

[149] Robert, C. and Casella, G. *Introducing Monte Carlo Methods with R.* 2010.

[150] Ruffini, R., Vereshchagin, G., and Xue, S. S. Electron-positron pairs in physics and astrophysics: From heavy nuclei to black holes, 2010.

[151] Karplus, R. and Neuman, M. The scattering of light by light. *Phys. Rev.*, 83:776–784, Aug 1951.

[152] Karplus, R. and Neuman, M. Non-linear interactions between electromagnetic fields. *Physical Review*, 80, 1950.

[153] Kandus, A., Kunze, K. E., and Tsagas, C. G. Primordial magnetogenesis, 2011.

[154] Bonometto, S. and Rees, M. J. On possible observable effects of electron pair production in QSOs. *Monthly Notices of the Royal Astronomical Society*, 152, 1971.

[155] Svensson, R. and Zdziarski, A. Photon-photon scattering of gamma rays at cosmological distances. *The Astrophysical Journal*, 349, 1990.

[156] Anderson, C. D. Cosmic-ray positive and negative electrons. *Physical Review*, 44, 1933.

[157] Pike, O. J., MacKenroth, F., Hill, E. G., and Rose, S. J. A photon-photon collider in a vacuum hohlraum. *Nature Photonics*, 8, 2014.

[158] Ribeyre, X., D'Humières, E., Jansen, O., et al. Pair creation in collision of -ray beams produced with high-intensity lasers. *Physical Review E*, 93, 2016.

[159] Drebot, I., Micieli, D., Milotti, E., et al. Matter from light-light scattering via Breit-Wheeler events produced by two interacting compton sources. *Physical Review Accelerators and Beams*, 20, 2017.

[160] Yu, J. Q., Lu, H. Y., Takahashi, T., et al. Creation of electron-positron pairs in photon-photon collisions driven by 10-PW laser pulses. *Physical Review Letters*, 122, 2019.

[161] Battistoni, G., Boehlen, T., Cerutti, F., et al. Overview of the FLUKA code. *Annals of Nuclear Energy*, 82, 2015.

[162] Briesmeister, J. F. MCNPTM – a general Monte Carlo N-particle transport code. *Los Alamos National Laboratory*, 2000.

[163] Weaver, T. A. Reaction rates in a relativistic plasma. *Physical Review A*, 13, 1976.

[164] Lindl, J. D., Landen, O. L., Edwards, J., et al. Review of the national ignition campaign 2009-2012, 2014.

[165] Poder, K. Characterisation of self-guided laser wakefield accelerators to multi-GeV energies. *Thesis of Imperial College London*, 2016.

[166] Pike, O. Particle interactions in high temperature plasmas. *Imperial College London*, 2015.

[167] Tanabashi, M., Hagiwara, K., Hikasa, K., et al. Review of particle physics, 2018.

[168] Tsai, Y. S. Pair production and bremsstrahlung of charged leptons. *Reviews of Modern Physics*, 46, 1974.

[169] Davies, H., Bethe, H. A., and Maximon, L. C. Theory of bremsstrahlung and pair production. ii. integral cross section for pair production. *Physical Review*, 93, 1954.

[170] Micieli, D., Drebot, I., Bacci, A., et al. Compton sources for the observation of elastic photon-photon scattering events. *Physical Review Accelerators and Beams*, 19, 2016.

[171] Abraham, M. Theorie der elektrizität. zweiter band: Elektromagnetische theorie der strahlung. *Monatshefte für Mathematik und Physik*, 17, 1905.

[172] Hu, H., Müller, C., and Keitel, C. H. Complete QED theory of multiphoton trident pair production in strong laser fields. *Physical Review Letters*, 105, 2010.

[173] Bell, A. R. and Kirk, J. G. Possibility of prolific pair production with high-power lasers. *Physical Review Letters*, 101, 2008.

[174] Sokolov, I. V., Naumova, N. M., Nees, J. A., and Mourou, G. A. Pair creation in QED-strong pulsed laser fields interacting with electron beams. *Physical Review Letters*, 105, 2010.

[175] Elkina, N. V., Fedotov, A. M., Kostyukov, I. Y., et al. QED cascades induced by circularly polarized laser fields. *Physical Review Special Topics - Accelerators and Beams*, 14, 2011.

[176] Piazza, A. D., Müller, C., Hatsagortsyan, K. Z., and Keitel, C. H. Extremely high-intensity laser interactions with fundamental quantum systems. *Reviews of Modern Physics*, 84, 2012.

[177] Bulanov, S. S., Schroeder, C. B., Esarey, E., and Leemans, W. P. Electromagnetic cascade in high-energy electron, positron, and photon interactions with intense laser pulses. *Physical Review A - Atomic, Molecular, and Optical Physics*, 87, 2013.

[178] Blackburn, T. G., Seipt, D., Bulanov, S. S., and Marklund, M. Benchmarking semi-classical approaches to strong-field QED: Nonlinear compton scattering in intense laser pulses. *Physics of Plasmas*, 25, 2018.

[179] King, B. Double compton scattering in a constant crossed field. *Physical Review A - Atomic, Molecular, and Optical Physics*, 91, 2015.

[180] Dinu, V. and Torgrimsson, G. Single and double nonlinear compton scattering. *Physical Review D*, 99, 2019.

[181] Seipt, D., Heinzl, T., Marklund, M., and Bulanov, S. S. Depletion of intense fields. *Physical Review Letters*, 118, 2017.

[182] Ridgers, C. P., Brady, C. S., Kirk, J. G., et al. Dense electron-positron plasmas generated by 10PW lasers in the QED-plasma regime. volume 8780, 2013.

[183] Arber, T. D., Bennett, K., Brady, C. S., et al. Contemporary particle-in-cell approach to laser-plasma modelling, 2015.

[184] Derouillat, J., Beck, A., Pérez, F., et al. Smilei: A collaborative, open-source, multi-purpose particle-in-cell code for plasma simulation. *Computer Physics Communications*, 222, 2018.

[185] Fonseca, R. A., Silva, L. O., Tsung, F. S., et al. Osiris: A three-dimensional, fully relativistic particle in cell code for modeling plasma based accelerators. volume 2331 LNCS, 2002.

[186] Neitz, N. and Piazza, A. D. Stochasticity effects in quantum radiation reaction. *Physical Review Letters*, 111, 2013.

[187] Niel, F., Riconda, C., Amiranoff, F., et al. From quantum to classical modeling of radiation reaction: A focus on stochasticity effects. *Physical Review E*, 97, 2018.

[188] Ridgers, C. P., Blackburn, T. G., Sorbo, D. D., et al. Signatures of quantum effects on radiation reaction in laser-electron-beam collisions. *Journal of Plasma Physics*, 83, 2017.

[189] Salamin, Y. I. Fields of a gaussian beam beyond the paraxial approximation. *Applied Physics B: Lasers and Optics*, 86, 2007.

[190] Rasmussen, C. E. and Nickisch, H. Gaussian processes for machine learning (gpml) toolbox. *Journal of Machine Learning Research*, 11, 2010.

[191] Intel. Intel® Xeon® processor E5-2680 v3 `https://ark.intel.com/content/www/us/en/ark/products/81908/intel-xeon-processor-e5-2680-v3-30m-cache-2-50-ghz.html`, Date accessed: 29/04/21.

[192] Bishop, C. M. Mixture density networks. *Aston University*, 1994.

[193] NVIDIA. NVIDIA TITAN Xp GPU `https://www.nvidia.com/en-us/titan/titan-xp/`, Date accessed: 29/04/21.

[194] Aad, G., Abbott, B., Abbott, D. C., et al. Observation of light-by-light scattering in ultraperipheral Pb+Pb collisions with the ATLAS detector. *Physical Review Letters*, 123, 2019.

[195] Lindl, J. D., Amendt, P., Berger, R. L., et al. The physics basis for ignition using indirect-drive targets on the National Ignition Facility, 2004.

[196] Gies, H., Jaeckel, J., and Ringwald, A. Polarized light propagating in a magnetic field as a probe for millicharged fermions. *Physical Review Letters*, 97, 2006.

[197] Jaeckel, J. and Ringwald, A. The low-energy frontier of particle physics. *Annual Review of Nuclear and Particle Science*, 60, 2010.

[198] Villalba-Chávez, S., Meuren, S., and Müller, C. Minicharged particles search by strong laser pulse-induced vacuum polarization effects. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 763, 2016.

[199] Beyer, K. A., Marocco, G., Bingham, R., and Gregori, G. Axion detection through resonant photon-photon collisions. *Physical Review D*, 101, 2020.

[200] Bernard, D., Moulin, F., Amiranoff, F., et al. Search for stimulated photon-photon scattering in vacuum. *European Physical Journal D*, 10, 2000.

[201] Moulin, F., Bernard, D., and Amiranoff, F. Photon-photon elastic scattering in the visible domain. *Zeitschrift für Physik C: Particles and Fields*, 72, 1996.

[202] Inada, T., Yamaji, T., Adachi, S., et al. Search for photon-photon elastic scattering in the X-ray region. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 732, 2014.

[203] Yamaji, T., Inada, T., Yamazaki, T., et al. An experiment of x-ray photon–photon elastic scattering with a laue-case beam collider. *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics*, 763, 2016.

[204] Bearden, J. A. and Burr, A. F. Reevaluation of X-ray atomic energy levels. *Reviews of Modern Physics*, 39, 1967.

[205] Poikela, T., Plosila, J., Westerlund, T., et al. Timepix3: A 65K channel hybrid pixel readout chip with simultaneous ToA/ToT and sparse readout. *Journal of Instrumentation*, 9, 2014.

[206] Hubbell, J., Bach, R., and Herbold, R. Radiation field from a circular disk source. *Journal of Research of the National Bureau of Standards, Section C: Engineering and Instrumentation*, 65C, 1961.

[207] Smith, J. H. and Storm, M. L. Generalized off-axis distributions from disk sources of radiation. *Journal of Applied Physics*, 25, 1954.

[208] Behm, K. T., Cole, J. M., Joglekar, A. S., et al. A spectrometer for ultrashort gamma-ray pulses with photon energies greater than 10 MeV. *Review of Scientific Instruments*, 89, 2018.

[209] Knuth, D. *The Art of Computer Programming*, volume 40. 1998.