Simulating Fundamental Physics: Numerical Endeavours in Astronomical and Particle Phenomenology

JAMES LUKE MAXWELL

Simulating Fundamental Physics

Numerical Endeavours in Astronomical and Particle Phenomenology

James Maxwell

A Thesis presented for the degree of
Doctor of Philosophy

Institute for Particle Physics Phenomenology
Department of Physics
Durham University
United Kingdom

June 2025

Simulating Fundamental Physics

Numerical Endeavours in Astronomical and Particle Phenomenology

James Maxwell

Submitted for the degree of Doctor of Philosophy

June 2025

Abstract: The fundamental theories of physics, general relativity and quantum field theory allow for the description of highly complex interacting systems. To truly realise the nature of these theories, one must take advantage of the numerical power provided by modern computational hardware. In this thesis, we begin by exploring the possibility, through numerical simulations, that signatures of multi axion phenomenology may be imprinted on the high energy photon spectra of Blazars. We follow this with a presentation of our work on the determination, numerically, of the dynamical friction force that would be experienced by a black hole propagating through a field of scalar particles. Finally, we return from the astrophysical realm, and present a novel code that may allow for the rapid computation of precision matrix elements relevant to future terrestrial collider experience. By accelerating pre-existing implementations of the Laprota algorithm, during the the preliminary benchmarking of our code, we observed a reduction in execution time by an order of magnitude.

Contents

List of Figures

List of Tables

Declaration

The work in this thesis is based on research carried out in the Department of Physics at Durham University. No part of this thesis has been submitted elsewhere for any degree or qualification.

Chapter 2 is based on the collaborative work of ref. [5] undertaken with Dr Francesca Chadha-Day and Dr Jessica Turner who took advisory roles.

Chapters 4 and 5 contain work intended for publication. Chapter 4 was produced in conjunction with Dr Francesca Chadha-Day, Dr Katy Clough and Zipeng Wang with the author of this thesis acting as the lead programmer.

The work presented in chapter 5 is the result of a collaboration with Dr Stephen Jones and Jack Franklin. The GPU based Lagrange interpolator was produced entirely by the author of this thesis with parts of the parser being the result of a collaborative effort with Jack Franklin. Dr Stephen Jones acted as an advisor for this project.

Acknowledgements

Firstly, I would like to thank my supervisor Francesca Chadha-Day for her help and guidance during my PhD. I would also like to thank my housemates Helen, Mia and Rollo and all of the PhD students at the IPPP. I would further like to thank my proof readers, Francesca, Mia, and my parents Ruth and Peter. Finally, I would like to thank my Aunt's cats, Walter, Willow and Milo and my housemate's cat Tibbs for their generous moral support throughout.

Then Ilúvatar said to them: 'Of the theme that I have declared to you, I will now that ye make in harmony together a Great Music'

— from *The Silmarillion* by J. R. R. Tolkien

Dedicated to

Walter and Willow

Preamble

Prior to the advent of computational hardware the solutions to the equations that govern the dynamics of the universe could only be probed through the use of analytical approximations and arduous numerical calculations. Whilst in the classical era of Newtonian mechanics, these techniques proved, in many cases, sufficient to extract verifiable predictions from theoretical laws, in the modern age of quantum mechanics and general relativity their applicability is limited to all but the most simplistic of scenarios. These scenarios often include isolated systems with minimal interactions. For example, the derivations of the analytical Black Hole solutions to Einstein field equations rely on the high degree of symmetry associated with a trivial matter environment. Similarly, in particle physics, purely analytical methods can only access low precision terms in perturbative expansions of the path integral, with those of a higher order necessitating the evaluation of complicated loop integrals.

The preceding two decades have seen momentous developments in computational power. These increases in memory capacity and processing speed mean it is now possible to simulate, to a high degree of precision, complicated and coupled systems involving large scale ranges. In chapter 2 we discuss the propagation of high energy photons from their Blazar source to the Earth in the presence of a non-trivial magnetic field structure and with couplings to string motivated axion-like particles. In chapter 4 we consider the more intensive problem of a Black Hole moving trough a field of scalar particles. The vast difference in length scales associated with this problem necessitate the use of a High Performance Computing (HPC) cluster to achieve a sufficient operational throughput to appropriately resolve important

dynamics. HPC's offer a high degree of parallelism and have been the de facto standard platform for intensive numerical investigations. With the rise of Artificial Intelligence in recent years and the accompanying growth in the capabilities of Graphics Processing Units (GPU's), a new paradigm for the study of numerical physics has emerged. GPU's provide extremely high operational throughputs with a memory structure that offers minimal latency. In chapter 5 we present a novel, proof of concept code that may aid in the computation of high precision matrix elements, potentially capable of accelerating their evaluation by an order of magnitude.

This thesis explores three vastly different aspects of computational physics. In advance of each subject, therefore, we provide a detailed introduction and a specification of the notations and conventions used therein. Note, in particular, that during our discussions on particle physics and general relativity, we use different metric signatures and different choices of natural units, so as best to remain in keeping with the existing literature and the conventions of these fields.

We begin now with an introduction to one of the most well verified theoretical descriptions of the universe, that of the Standard Model of particle physics.

Chapter 1

Particle Physics Introduction

1.1 The Standard Model

The Standard Model (SM) represents the state of the art in our understanding of the behaviour of the infinitesimal. It is a unification of special relativity and quantum mechanics and has allowed for some of the most accurate predictions about our universe. The SM will form the basis for the majority of the work in this thesis. Accordingly, we present an overview of its core facets to give context for what follows.

1.1.1 Gauge Theory

Through over 100 years of experimentation, the local (Gauge) symmetries of the universe have been determined. It is these that form the basis for the SM. Specifically, we require the Lagrangian to be invariant under the following Lie group:

$$\mathrm{SU(3)_c} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y}. \tag{1.1.1}$$

Where, $\mathrm{SU(3)_c}$ is associated with the strong interaction and $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ describes electro-weak processes [6]. The SM Lagrangian can then be constructed by writing down every gauge and Lorentz invariant term that is renormalizable and corresponds

to an observed particle. To illustrate this process, we consider a toy model involving a non-interacting, massive complex scalar φ with U(1) gauge symmetry:

$$\mathcal{L} = (\partial_\mu \varphi)^\dagger (\partial^\mu \varphi) - m^2 \varphi^\dagger \varphi. \tag{1.1.2}$$

Where, under $U(x) \in U(1)$ transformation:

$$\varphi \to U(x)\varphi. \tag{1.1.3}$$

Here, $U(x)$ are elements of the representation labelled by a charge g. In the Lagrangian's current state eq. (1.1.2), the presence of the partial derivatives spoil the gauge symmetry:

$$\partial_\mu (U(x)\varphi(x)) = \varphi(x)\partial_\mu (U(x)) + U(x)\partial_\mu \varphi(x) \tag{1.1.4}$$

$$\neq U(x)\partial_\mu \varphi(x). \tag{1.1.5}$$

To address this issue, we must introduce an additional, so called *gauge field* $A_\mu(x)$, with which we define a gauge covariant derivative:

$$\mathcal{D}_\mu = \partial_\mu - igA_\mu. \tag{1.1.6}$$

We further stipulate that A_μ possesses the following transformation property under $U(1)$:

$$A_\mu \to UA_\mu U^{-1} - \frac{i}{g}(\partial_\mu U)U^{-1}. \tag{1.1.7}$$

Finally, exchanging ∂_μ for $\mathcal{D}_\mu$ in eq. (1.1.2) yields the desired gauge invariance. Interestingly, our theory of a single particle has now become one of two particles. At present, however, A_μ is arbitrary and non-dynamical. As we are free to add additional terms to our Lagrangian, provided they satisfy the symmetries, we seek

to introduce a derivative term for A_μ. This can be constructed in a gauge invariant way using the field strength tensor:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{1.1.8}$$

where it can be shown that the field strength tensor transforms as $F_{\mu\nu} \to U F_{\mu\nu} U^\dagger$. The following contraction can be appended to our Lagrangian, giving:

$$\mathcal{L} = (\mathcal{D}_\mu \varphi)^\dagger (\mathcal{D}^\mu \varphi) - m^2 \varphi^\dagger \varphi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{1.1.9}$$

$$= (\partial_\mu \varphi)^\dagger (\partial^\mu \varphi) - m^2 \varphi^\dagger \varphi + g^2 A_\mu A^\mu \varphi^2 - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \tag{1.1.10}$$

This is a fully gauge invariant Lagrangian that describes the interaction between two dynamical fields.

The true SM Lagrangian can be constructed in a way analogous to this, albeit with some additional complications. In particular, in the above example we considered a very simple gauge symmetry to postpone the discussion of group representations and algebras. We will now address these complexities to motivate the physical Lagrangian.

The SM gauge groups are Lie groups. That is to say that the group operation is differentiable. This means that at a given point in spacetime, an element g of a group G of dimension N can be written as:

$$g = e^{-\theta^a T^a} \qquad\qquad a \in 1, 2, ...N, \tag{1.1.11}$$

where the θ^a are continuous parameters that label the group elements and the T^a are known as the infinitesimal generators of the group:

$$T^a = i \frac{\partial g}{\partial \theta^a}\bigg|_{\theta=0}. \tag{1.1.12}$$

The group generators are related according to the Lie algebra:

$$\left[T^a, T^b\right] = if^{abc}T^c, \tag{1.1.13}$$

where the f^{abc} are known as *structure constants* and $[,]$ is the *Lie bracket*. The Lie algebra of the group can be used to construct the various group representations. A group representation associates each element in the group with an $M \times M$ matrix. The different particles in the SM will transform under different group representations. Specifically, the fermion fields will transform either under the *fundamental* or the *trivial* representations, whereas the gauge fields will transform under the *adjoint* representation.

The trivial representation simply maps every group element to the identity matrix. Fields that are uncharged under a given symmetry will transform under the trivial representation, for example, leptons under $SU(3)_c$. We can express the non-trivial representations in terms of their generators according to eq. (1.1.11). The Fundamental representation is the smallest injective representation. It is conventional to normalise the fundamental generators as follows:

$$\mathrm{Tr}(T_F^a T_F^b) = \frac{\delta^{ab}}{2}. \tag{1.1.14}$$

For $SU(3)$ the generators that satisfy these conditions are known as the Gell-Mann matrices.

The generators of the adjoint representation can be expressed in terms of the structure constants as:

$$(T_A^a)^{bc} = -if^{abc}. \tag{1.1.15}$$

This is because the adjoint representation acts on elements of the lie algebra so its generators must encode the relationship expressed in eq. (1.1.13). Acting on

the basis elements (T^a) of the Lie algebra, the adjoint generators will yield the lie bracket:

$$T^a_A T^b = \left[T^a, T^b\right]. \tag{1.1.16}$$

The action of a group element $g \in G$ on an element of the algebra $h \in \mathfrak{g}$ is that of composition:

$$gh = g \circ h. \tag{1.1.17}$$

The Lie algebra consists of matrices so this composition can be written:

$$g \circ h = ghg^{-1}. \tag{1.1.18}$$

Comparing this to the transformation property of the gauge field A_μ in eq. (1.1.7), we see that the first part of the right hand side is recovered by the requirement that A_μ transforms under the adjoint representation. Interestingly, this also means that the values that the gauge field takes are actually the elements of the Lie algebra. The second term in eq. (1.1.7) arises because the field is defined locally and simply accounts for the derivatives present in eq. (1.1.2).

By requiring that the gauge fields belong to the Lie algebra of the symmetry group, one can ensure appropriate transformation properties when constructing the covariant derivative. It should be noted that the A_μ considered in the above example is not expressed in the general form for a Lie algebra element. This is because $U(1)$ has only a single generator. Generally, a gauge field will be expressed as a superposition of the fundamental generators:

$$A_\mu = A^a_\mu T^a_F. \tag{1.1.19}$$

This leads to the following modification to the field-strength tensor:

$$F_{\mu\nu} = F^a_{\mu\nu} T^a \tag{1.1.20}$$

$$= \partial_\mu A_\nu - \partial_\nu A_\mu - ig \left[A_\mu, A_\nu \right]. \tag{1.1.21}$$

Putting all this together, we can now construct the SM Lagrangian.

1.1.2 Fermion Lagrangian

In the previous example, we considered a scalar field and a vector gauge field. Fermions, however, exist in a different space - that of *spinors*. Spinors were first introduced by Klein in 1897 and encode the property of spin as discovered by the Stern-Gerlach experiment [7]. As with all objects in the SM, spinors must behave in a way such that their physics is invariant under the Lorentz group $SO(1,3)$, i.e. Lorentz scalars formed from spinors should be invariant under the group action. Lorentz 4-vectors naturally satisfy this condition. However, they do not reflect the spin property. Spinors change sign under a spatial rotation by 360^o instead requiring a rotation by 720^o to return to their original state. There are no representations of the rotational subgroup $SO(3) \subset SO(1,3)$ of the Lorentz group that satisfy this property. Spinors instead, transform under the fundamental representation of $SL(2,\mathbb{C})$. $SL(2,\mathbb{C})$ is a double cover (2 to 1 homomorphism) of $SO(1,3)$. One full rotation in the subgroup $SU(2,\mathbb{C}) \subset SL(2,\mathbb{C})$ maps to two full rotations in $SO(3)$, thus encoding the spin property.

To see this more explicitly consider the fundamental representations of $SO(3)$ and $SU(2,\mathbb{C})$. The generators of $SO(3)$ (J^x, J^y, J^z) are given by:

$$J^x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad J^y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \quad J^z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{1.1.22}$$

and obey the Lie algebra:

$$[J_x, J_y] = iJ_z, \quad [J_y, J_z] = iJ_x, \quad [J_z, J_x] = iJ_y. \tag{1.1.23}$$

A generic rotation matrix in 3 dimensions can then be written:

$$A = e^{i\theta^a J^a}. \tag{1.1.24}$$

The generators for $SU(2,\mathbb{C})$ are the Pauli matrices $\{\sigma^a\}, a \in [1,3]$, given by:

$$\sigma^1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1.1.25}$$

obeying:

$$\left[\sigma^i, \sigma^j\right] = i\epsilon_{ijk}\sigma^k. \tag{1.1.26}$$

Where ϵ_{ijk} are the Levi-Civita symbols. A rotation in $SU(2,\mathbb{C})$ can be written:

$$A' = e^{i\theta'^a \sigma^a}. \tag{1.1.27}$$

These are related to rotations in $SO(3)$ by:

$$\theta'^a = 2\theta^a. \tag{1.1.28}$$

Including boosts, a general spinor will transform under the fundamental representation of $SL(2,\mathbb{C})$ and can be represented by a two component complex vector. There is one additional subtlety concerning spinors when modelling fermionic particles. This is chirality (or handedness). Experiments have shown that there are two forms of spinor that behave differently under the standard model gauge symmetries. These are referred to as being *left-handed* or *right-handed* and can be distinguished by the representation of $SU(2,\mathbb{C})$ under which they transform. Specifically, for a rotation, a left-handed spinor will transform under:

$$A'_L = e^{i\theta'^a \sigma^a}. \tag{1.1.29}$$

Whereas a right-handed spinor will transform under the inverse:

$$A'_R = e^{-i\theta'^a \sigma^a}. \tag{1.1.30}$$

These two component complex spinors are known as *Weyl spinors*. Left handed Weyl spinors are said to transform under the $(1/2, 0)$ representation of the Lorentz group, meaning that they transform under the first of the two representations above and trivially under the second. A right handed Weyl spinor is said to transform under the $(0, 1/2)$ representation.

In any massless theory, the left and right handed Weyl spinors will be uncoupled and the Lagrangian could simply be written in terms of separate spinors. For example:

$$\mathcal{L}_{\text{kinetic}} = i\psi_L^\dagger \sigma^\mu \partial_\mu \psi_L + i\psi_R^\dagger \overline{\sigma}^\mu \partial_\mu \psi_R. \tag{1.1.31}$$

Where ψ_L and ψ_R are the left and right handed spinor respectively, $\sigma^\mu = (\mathbb{I}, \vec{\sigma})$ are the Pauli matrices and $\overline{\sigma}^\mu = (\mathbb{I}, -\vec{\sigma})$. The presence of the Pauli matrices in eq. (1.1.31) effectively generalises the Lorentz invariance of ∂_μ to the space of spinors. Fermions, however, are known to be massive and mass terms lead to a coupling between left-handed and right-handed spinors. It is thus useful to introduce a doublet of left and right handed spinors known as a *Dirac spinor*:

$$\Psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}. \tag{1.1.32}$$

The fermion mass term can then be written:

$$\mathcal{L}_{\text{mass}} = -m\overline{\Psi}\Psi \tag{1.1.33}$$

$$= -m\left(\psi_L^\dagger \sigma^0 \psi_R + \psi_R^\dagger \overline{\sigma}^0 \psi_L\right), \tag{1.1.34}$$

where:

$$\overline{\Psi} = \Psi^\dagger \gamma^0. \tag{1.1.35}$$

Here σ^0 and $\overline{\sigma}^0$ effectively encode contractions in left handed and right handed Weyl spinor spaces and γ^0 encodes contraction in the combined Dirac space. These ensure that the mass term is a Lorentz scalar. Eq. (1.1.31) can be rewritten in Dirac spinor space as:

$$\mathcal{L}_{\text{kinetic}} = i\overline{\Psi}\gamma^\mu \partial_\mu \Psi, \tag{1.1.36}$$

where the gamma matrices, which take the role of the Pauli matrices, are given by:

$$\gamma^0 = \begin{pmatrix} 0 & \mathbb{I}_2 \\ \mathbb{I}_2 & 0 \end{pmatrix} \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}, \quad i = 1, 2, 3. \tag{1.1.37}$$

If a Dirac spinor contains only one non-zero Weyl spinor, it is said to have the chirality of that spinor. The chirality of a Dirac spinor can be determined using the chirality operator γ^5:

$$\gamma^5 \begin{pmatrix} \psi_L \\ 0 \end{pmatrix} = -\begin{pmatrix} \psi_L \\ 0 \end{pmatrix}; \tag{1.1.38}$$

$$\gamma^5 \begin{pmatrix} 0 \\ \psi_R \end{pmatrix} = \begin{pmatrix} 0 \\ \psi_R \end{pmatrix}. \tag{1.1.39}$$

Where,

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3. \tag{1.1.40}$$

Dirac spinors can be decomposed into their right-handed and left-handed components using right and left chiral projection operators:

$$\Psi = \Psi_L + \Psi_R = P_L\Psi + P_R\Psi. \tag{1.1.41}$$

Hare Ψ_R and Ψ_L are Dirac spinors and the projection operators are given by:

$$P_L = \frac{1}{2}\left(1 - \gamma^5\right), \quad P_R = \frac{1}{2}\left(1 + \gamma^5\right). \tag{1.1.42}$$

As stated previously, the SM gauge group does not treat fermions with opposing chirality equally. Specifically, only left handed fermions transform non-trivially under $SU(2)_L$. This means that the mass term eq. (1.1.34) is in fact not gauge invariant as it mixes left and right handed components. This conundrum can be resolved with the keystone of the standard model - the *Higgs mechanism*.

1.1.3 The Higgs Mechanism

The Higgs mechanism allows for massive fermions and gauge bosons without violating gauge invariance through a process of spontaneous symmetry breaking. We introduce a complex scalar doublet that transforms in the fundamental representations of $SU(2)_L$ and $U(1)_Y$:

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix}\varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4\end{pmatrix}. \tag{1.1.43}$$

This is the Higgs field, the dynamics of which are given by its Lagrangian:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\Phi)^\dagger(D_\mu\Phi) - V(\Phi), \tag{1.1.44}$$

The covariant derivative D_μ can be written in terms of the vector $SU(2)_L$ gauge fields $W^a_\mu,\ a \in \{1,2,3\}$ and the $U(1)_Y$ gauge field B_μ

$$D_\mu \Phi = \left(\partial_\mu - i\frac{g}{2}\sigma_a \cdot W_\mu^a - i\frac{g'}{2}B_\mu \right) \Phi, \tag{1.1.45}$$

where σ_a are the Pauli matrices that generate $SU(2)_L$. The most general form one can write down for the Higgs potential $V(\Phi)$ is:

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2. \tag{1.1.46}$$

At this stage, μ and λ are free parameters. To produce symmetry breaking, however, we require the Higgs field to posses a non-zero Vacuum Expectation Value (VEV). This occurs when $V(\Phi)$ is minimal for non-zero Φ. To form such a potential, we require $\mu^2 > 0$ and $\lambda > 0$. For this choice of parameters, the potential will be minimised when $\Phi^\dagger \Phi = \frac{\mu^2}{2\lambda}$ and the Higgs VEV is:

$$v = \sqrt{\frac{\mu^2}{2\lambda}}. \tag{1.1.47}$$

The Higgs field has four degrees of freedom, so there are infinitely many states that minimise the potential $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 + \varphi_4^2 = \frac{\mu^2}{2\lambda}$. These states are all equivalent up to a gauge transformation and so we are free arbitrarily to select a gauge in which only a single degree on freedom is non-zero. Conventionally, we choose $\varphi_4 \neq 0$, rewriting it in terms of a new field h with a trivial VEV as:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ h + v \end{pmatrix}. \tag{1.1.48}$$

This is known as the unitary gauge. When the Higgs field is in its vacuum state, the $SU(2)_L \times U(1)_Y$ is spontaneously broken to a single $U(1)_{\mathrm{EM}}$ corresponding to electromagnetism.

The gauge boson mass terms now arise automatically by evaluating the Higgs kinetic term at the VEV ($h = 0$). With the following field definitions:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \mp iW_\mu^2\right), \tag{1.1.49}$$

$$Z_\mu^0 = \frac{1}{\sqrt{g^2+g'^2}}\left(gW_\mu^3 - g'B_\mu\right), \tag{1.1.50}$$

$$A_\mu = \frac{1}{\sqrt{g^2+g'^2}}\left(gW_\mu^3 + g'B_\mu\right), \tag{1.1.51}$$

$$\tag{1.1.52}$$

we obtain:

$$(D_\mu\Phi)^\dagger(D^\mu\Phi)|_{h=0} = \frac{1}{2}\left(\frac{gv}{2}\right)^2\left(|W_\mu^+|^2 + |W_\mu^-|^2\right) + \frac{1}{2}\left(\frac{v\sqrt{\left(g^2+g'^2\right)}}{2}\right)^2 |Z_\mu^0|^2 \tag{1.1.53}$$

$$\equiv \frac{1}{2}m_W^2\left(|W_\mu^+|^2 + |W_\mu^-|^2\right) + \frac{1}{2}m_Z^2|Z_\mu^0|^2. \tag{1.1.54}$$

The $W^\pm$ and Z^0 bosons gain masses $\frac{gv}{2}$ and $\frac{v\sqrt{g^2+g'^2}}{2}$ respectively. Note that there is no mass term for the photon A_μ. This is because the Higgs VEV is still invariant under $U(1)_{\text{EM}}$ after symmetry breaking.

To derive the fermion mass terms, we must introduce a gauge invariant Yukawa coupling between the Higgs and the fermion fields. To achieve this, we must first pair the left-handed fermions into $SU(2)_L$ doublets. Consider as an example the doublet formed from two Dirac spinors representing a down quark (d) and an up quark (u):

$$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}. \tag{1.1.55}$$

Note that only the left-handed projections of the spinor fields are included in the doublet as only they transform non-trivially under $SU(2)_L$. The full Yukawa interaction for these quarks can then be written:

$$\mathcal{L}_Y^{\text{Quarks}} = -y_d \overline{Q}_L \Phi d_R - y_u \overline{Q}_L \tilde{\Phi} u_R + h.c. \tag{1.1.56}$$

Where y_d and y_u are the Yukawa couplings and $h.c.$ is the hermitian conjugate. Each term in eq. (1.1.58) constitutes an inner product in $SU(2)_L$ space between a quark doublet and the Higgs doublet. Once again, substituting the VEV for the Higgs field, one obtains fermionic masses of the form:

$$m_f = \frac{1}{\sqrt{2}} y_f v. \tag{1.1.57}$$

This process also applies to leptons, however, there are no terms involving right-handed neutrinos as these are massless in the SM. For example, for electrons:

$$\mathcal{L}_Y^{\text{Leptons}} = -y_e \overline{L}_L \Phi e_R + h.c. \tag{1.1.58}$$

with doublet:

$$L_L = \begin{pmatrix} e_L \\ \nu_{eL} \end{pmatrix}. \tag{1.1.59}$$

These quark and lepton groupings are known as *generations* (or families). The SM has three generations; (see fig. 1.1 for details). Focusing on a single generation, and combining everything together, the SM Lagrangian can be written:

$$\mathcal{L}_{\text{SM}} = -\frac{1}{4} G^{\alpha}_{\mu\nu} G^{\alpha,\mu\nu} - \frac{1}{4} W^{a}_{\mu\nu} W^{a,\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.1.60}$$

$$+ (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi^\dagger \Phi) \tag{1.1.61}$$

$$+ i\overline{Q}_L \not{D} Q_L + i\overline{Q}_R \not{D} Q_R + i\overline{L}_L \not{D} L_L + i\overline{L}_R \not{D} L_R \tag{1.1.62}$$

$$- (y_d \overline{Q_L} \Phi d_R + y_u \overline{Q}_L \tilde{\Phi} u_R + h.c.) - (y_e \overline{L_L} \Phi e_R + h.c.). \tag{1.1.63}$$

Here, the first line contains the gauge boson kinetic terms. The second line describes the Higgs field and its potential. The Third line describes the fermion kinetic terms and the final line contains the fermion-Higgs Yukawa interactions.

This has all been constructed by writing down every renormalisable term involving the empirical fields that is symmetric under both the SM gauge group and the Lorentz group. Working within this vein, there is one additional term that one could, in principle, write down pertaining to the gluon and electroweak fields. There have, however, been no experimental studies able to detect its existence. The seeming absence of this term is known as the *strong CP problem*.

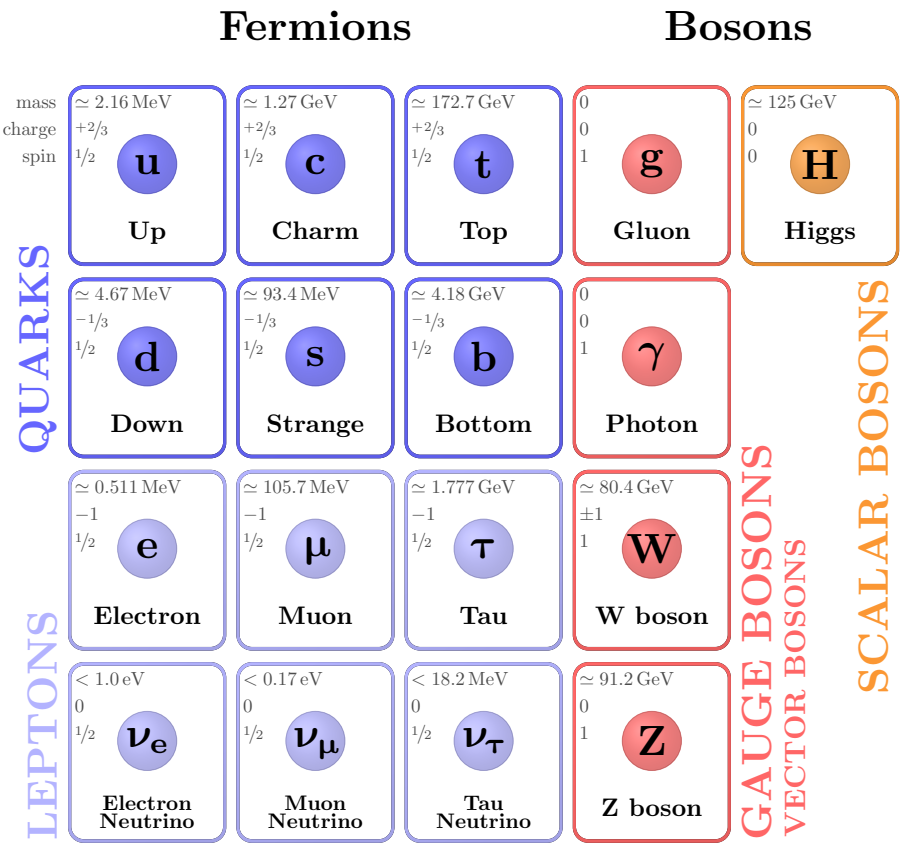


Figure 1.1: The particle content of the SM. The fermions are shown on the left with the three generations separated into individual columns. Mass values were taken from [1].

1.2 The Strong CP Problem

In the theory of Quantum Chromo Dynamics (QCD), the SM symmetries permit the inclusion of a term of the form:

$$\mathcal{L} \supset \theta \frac{g^2}{32\pi^2} G^{\alpha}_{\mu\nu} \tilde{G}^{\alpha,\mu\nu}, \tag{1.2.1}$$

known as the QCD theta term, where $\tilde{G}^{\alpha,\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\sigma\rho}G^{\alpha}_{\sigma\rho}$ is the dual gluon field strength tensor and θ is known as the vacuum angle. This additional Lagrangian term can in fact be written as total derivative:

$$G^{\alpha}_{\mu\nu}\tilde{G}^{\alpha,\mu\nu} = \partial_{\mu}K^{\mu}, \tag{1.2.2}$$

where K^{μ}, known as the *Chern-Simons* current, is given by:

$$K^{\mu} = \epsilon^{\mu\nu\rho\sigma}\left(A^{a}_{\nu}\partial_{\rho}A^{a}_{\sigma} + \frac{g}{3}f^{abc}A^{a}_{\nu}A^{b}_{\rho}A^{c}_{\sigma}\right). \tag{1.2.3}$$

As a consequence of this, one can conclude that the theta term will not affect perturbations about the vacuum. This can be seen by considering the space-time integral in the path integral which, for total derivatives, can be reformulated using Gauss law into a surface integral over the boundary at infinity. The standard boundary conditions then ensure that this integral is null. For this reason, it might seem that the theta term is redundant, and indeed it was neglected in early formulations of QCD. It was later realised, however, that its inclusion gives rise to a number of important non-perturbative effects. These effects arise due to the non-trivial vacuum structure of QCD. That is to say that there exist SM ground states that cannot be rotated into one another through an $SU(3)$ gauge transformation. This non-trivial vacuum results from the topology of the $SU(3)$ manifold, so the theta term is sometimes referred to as being *topological*.

If the universe existed purely in a single vacuum state, then we would not observe any effect from the theta term. Semi-classical solutions to the Equations Of Motion (EOM) have been discovered that take the system from one vacuum state to another. These solutions are known as *instantons* and, importantly, (through the theta term) enable the violation of the combined discrete symmetry of Charge conjugation and Parity (collectively CP symmetry). Charge conjugation symmetry refers to invariance under the exchange of a particle with its anti-particle and parity describes the effect of inverting the space-time directions. Instantons map between states

of differing baryon number, thereby increasing the matter content of the universe relative to anti matter. This equates to a CP symmetry violation. In addition, and of more relevance experimentally, the theta term implies a non-zero electric dipole moment for the neutron. Empirical studies have constrained this value to be very small $\leq 10^{-26}$e cm [8]. This translates to a constraint on the vacuum angle of $|\theta| \leq 10^{-10}$. The vanishing size of this parameter is what is known as the strong CP problem. It is problematic from the point of view of naturalness with it being significantly smaller than the other numbers in the SM Lagrangian.

1.2.1 The QCD Axion

One can attempt to explain the strong CP problem by moving beyond the SM. The most widely accepted, although currently, unproven, resolution to the strong CP problem was put forward by R. D. Peccei and H. Quinn in 1977 [9, 10]. Prior to this seminal work by Peccei and Quinn, it was shown that if at least one quark has no bare mass, then the theta term is irrelevant and can be set to zero through chiral rotation [11]. Unfortunately, this is inconsistent with empirical data and, although some models have been proposed that try to explain the quark masses as being a consequence of renormalization, the magnitude of this effect seems to be insufficient [12]. Motivated by this, and the way quark masses arise through the Higgs mechanism, Peccei and Quinn introduced a new global $U(1)_A$ *axial* symmetry. Axial in this context denotes that the group respects chirality, with elements operating inversely on left-handed ($\psi_L \to \psi_L e^{i\alpha}$) versus right-handed ($\psi_R \to \psi_R e^{-i\alpha}$) spinors. Acting on Dirac spinors, the group action can be expressed succinctly using the γ^5 matrix:

$$\psi \to \psi e^{i\alpha\gamma^5}. \tag{1.2.4}$$

This new $U(1)_A$ symmetry is often referred to as the Peccei-Quinn symmetry $U(1)_{\mathrm{PQ}}$. In addition, the Peccei-Quinn (PQ) mechanism introduces a new scalar field φ_{PQ}

charged under $U(1)_{\rm PQ}$ with the following zero-temperature potential:

$$V(\varphi_{\rm PQ}) = \mu_{\rm PQ}^2 \left(|\varphi_{\rm PQ}|^2 - \frac{f_a^2}{2} \right)^2 . \tag{1.2.5}$$

Where f_a is known as the *decay constant* and $\mu_{\rm PQ}$ is a dimensionless parameter. As we saw with the Higgs mechanism, a potential of this form has a non-zero VEV given by:

$$|\varphi_{\rm PQ}| = \frac{f_a}{\sqrt{2}}. \tag{1.2.6}$$

Expanding around the VEV, we can write:

$$\varphi_{\rm PQ}(x) = \frac{1}{\sqrt{2}} \left(f_a + \rho(x) \right) e^{ia(x)/f_a}. \tag{1.2.7}$$

Here $\rho(x)$ is a massive degree of freedom and $a(x)$ is a massless Goldstone boson. As we are working with a global symmetry, the Goldstone boson doesn't get eaten by the gauge fields to give them mass and instead persists as its own particle known as the *axion*. The massive field on the other hand decouples at low energies and can be neglected to give an effective axion theory. To see this, one need only compute its mass. By expanding the potential in eq. (1.2.5) around the VEV we find $m_\rho \sim \mu_{\rm PQ} f_a$. We will see shortly that the strength of the axion gluon coupling is inversely proportional to f_a. Experimental bounds show that f_a and m_ρ must be very large, thus ρ can be integrated out, giving an effective scalar field:

$$\varphi_{\rm PQ}(x) = \frac{f_a}{\sqrt{2}} e^{ia(x)/f_a}. \tag{1.2.8}$$

From this, we can see that the axion field transforms as $a(x) \to a(x) + \alpha f_a$ under $U(1)_{\rm PQ}$. At present, the axion can couple to fermions through Yukawa couplings as we saw for the Higgs boson. This, on its own, however, does not produce anything that looks like the QCD theta term. Of the other terms we can add to the Lagrangian, one of the simplest would be a current-current interaction of the form:

$$\mathcal{L} \supset J_{\varphi,\mu} J^{\mu}_{\psi} \tag{1.2.9}$$

Where $J_{\varphi,\mu}$ is the Noether current associated with the PQ charged scalar and J^{μ}_{ψ} is that associated with fermions. The axion Noether current associated with the PQ symmetry is given by

$$J^{\mu}_{\varphi} = \frac{\partial \mathcal{L}}{\partial(\partial_{\mu}\varphi_{\mathrm{PQ}})}\delta\varphi_{\mathrm{PQ}} + \frac{\partial \mathcal{L}}{\partial(\partial_{\mu}\varphi^{*}_{\mathrm{PQ}})}\delta\varphi^{*}_{\mathrm{PQ}}. \tag{1.2.10}$$

Where the transformation can be written as $\delta\varphi_{\mathrm{PQ}} = i\alpha\varphi_{\mathrm{PQ}}$. We find that

$$J^{\mu}_{\varphi} = i\varphi^{*}_{\mathrm{PQ}} \overset{\leftrightarrow}{\partial^{\mu}} \varphi_{\mathrm{PQ}}. \tag{1.2.11}$$

Substituting the low energy effective φ_{PQ} eq. (1.2.8) gives:

$$J^{\mu}_{\mathrm{PQ}} \approx f_a \partial^{\mu} a. \tag{1.2.12}$$

We therefore find that our current-current term can be expressed as:

$$J_{\varphi,\mu} J^{\mu}_{\psi} = f_a \partial_{\mu} a J^{\mu}_{\psi}. \tag{1.2.13}$$

Using integration by parts, the divergence can be shifted onto the fermion current, giving:

$$J_{\varphi,\mu} J^{\mu}_{\psi} = -f_a a \partial_{\mu} J^{\mu}_{\psi}. \tag{1.2.14}$$

This kind of derivative coupling is a feature of Goldstone bosons and has important consequences on the form of the interaction. In particular, a classical interpretation of Noether's theorem would suggest that eq. (1.2.14) is identically zero due to the PQ symmetry. In reality, however, regularisation of the one loop triangle diagram leads to a breakdown of the conservation law and the divergence of the fermion current becomes:

$$\partial_\mu J^\mu_\psi = \frac{g_s^2}{32\pi^2} N G^a_{\mu\nu} \tilde{G}^{a\mu\nu}, \tag{1.2.15}$$

where N relates to the PQ charge of the Fermions. This invalidation of the conservation law is known as the *chiral* or *axial anomaly*. Quantum anomalies describe the breakdown of classical symmetry laws when transitioning to a quantum field theory. Unlike classical theories, where the physical path taken through phase space is that which minimises the action, in QFT the system can be thought to take all possible paths, weighted and summed into the path integral. It is this distinction that leads to the problem. Specifically, it is not sufficient that the action is invariant under a transformation; to realise the conservation, the path integral measure must also transform by a trivial Jacobian factor. The current eq. (1.2.15) can be calculated through a long-winded process, the details of which can be found here [13]. The CP violating terms in the Lagrangian are now:

$$\mathcal{L} \supset \theta \frac{g^2}{32\pi^2} G^\alpha_{\mu\nu} \tilde{G}^{\alpha,\mu\nu} + \frac{a}{f_a} \frac{g_s^2}{32\pi^2} N G^a_{\mu\nu} \tilde{G}^{a\mu\nu}. \tag{1.2.16}$$

With this, we are starting to see how the PQ mechanism might account for the unnatural θ-angle. There are still a number of steps remaining, however. The axion field is currently a massless Goldstone mode and we have a dynamical effective θ-angle given by $\theta_{\text{eff}} = \theta + Na/f_a$. We can fix both these issues by considering the axion effective potential. Non-perturbative instantons give rise to a potential of the form:

$$V_{\text{eff}} \sim -\cos\left(\theta + N\frac{a}{f_a}\right). \tag{1.2.17}$$

This potential is minimised at the VEV, given by $\langle a \rangle = \theta f_a / N$. Finally, we can perform a field redefinition, expanding the new axion field around its VEV $a \to \langle a \rangle + a$. This new physical axion gains a mass through the explicit symmetry breaking potential induced by the instanton effects. We can also see that the strong

CP problem is now resolved, with the effective θ-term in the ground state being forced to zero by the axion VEV:

$$\theta_{\text{eff}} = \theta + N\frac{\langle a \rangle}{f_a} = 0. \tag{1.2.18}$$

Without breaking the rest of the SM, the PQ mechanism naturally explains the empirical lack of CP violation and provides a new pseudo-scalar degree of freedom with which to verify the theory. The axion is one of the most well motivated extensions to the SM and since its inception many experiments have searched for its existence. Although as yet no such particle has been discovered, the axion has motivated an entire branch of Beyond the Standard Model (BSM) theories based on a weakly coupled pseudo-scalar. This wider class of particles are said to be axion like and are the focus of our investigations in chapter 2. To avoid confusion, we shall, henceforth, refer to the axion described above as the QCD axion and more general axions as Axion Like Particles or ALPs.

Both the QCD axion and ALPs are also well motivated as possible dark matter (DM) candidates [14]. Through the misalignment mechanism, these fields could evolve into non-relativistic pressure-less clouds of particles, compatible with cold DM. The QCD axion and ALPs may thus constitute a significant fraction, or even the entirety, of the DM in the universe.

1.3 Axion Like Particles

ALPs arise both as a generalisation of the QCD axion and as a consequence of other theories. Generic ALPs do not posses the same relationship between their mass and coupling as the axion. This vastly increases the available parameter space in which to search for their existence. The QCD axion parameter space is difficult to probe experimentally and many new search strategies have begun to focus entirely on generic ALP models.

As well as generating a single generic ALP, some theories also predict the existence of a spectrum of ALPs of differing masses. This is of particular interest with regard to the work in this thesis. String theories regularly give rise to these ALP spectra and are a strong motivation for our work.

1.3.1 The String Axiverse

String theory is a vast subject and not the focus of this thesis. Conveniently, however, ALP modes are present in even the most simplistic string models. In this section, we will discuss bosonic string theory and show the general mechanism through which ALPs emerge. This section is based on ref. [15].

On a classical level, string theory promotes a classical point particle to a string. The path of the particle through space (its world-line) gets promoted to a world-sheet. In D dimensional space, the world-sheet is a $D-2$ dimensional hypersurface. This simple modification leads to a theory that may be able to explain both gravitation and SM physics in a renormalisable way that reduces to QFT and general relativity in the low energy limit. The starting point for bosonic string theory is the Polyakov action:

$$S = -\frac{T}{2}\int d^2\sigma\sqrt{-g}\, g^{\alpha\beta}\partial_\alpha X^\mu \partial_\beta X_\mu. \tag{1.3.1}$$

Here, we define coordinates on the world-sheet $\sigma_\alpha = (\tau, \sigma); \alpha \in \{1, 2\}$, where $\sigma \in [0, 2\pi)$ is the position along the string and τ is related to the time. The world-sheet is embedded in the background Minkowski space (known as the target space) according to $X_\mu(\sigma_\alpha)$, with the induced metric on the world-sheet being given by $g^{\alpha\beta}$ and $\sqrt{-g} = \sqrt{-\det(g_{\mu\nu})}$. The indices α and β refer to world-sheet coordinates and μ and ν refer to the target space. The Polyakov action is classically equivalent to an integral over the area of the world-sheet scaled by the tension in the string T. That is to say that, solving the classical EOMs for $g^{\mu\nu}$ and X^μ and substituting these back into the action gives:

$$S = -T \int d^2\sigma \sqrt{-\gamma}, \tag{1.3.2}$$

where $\gamma_{\mu\nu}$ is the, now non-dynamical, induced metric on the on the world-sheet. This equation, known as the Nambu-Goto (NG) action, and indeed eq. (1.3.1) have been constructed to be reparametrisation invariant with respect to the world-sheet coordinates σ_α. Interestingly, this simple string action is sufficient to generate both gravitons and axions.

Although eq. (1.3.2) and eq. (1.3.1) are classically equivalent, the Polyakov action is easier to work with at a quantum level. The extra degrees of freedom in the Polyakov action are related to gauge symmetries. Of particular importance is the invariance of eq. (1.3.1) under a local scaling of the metric:

$$g_{\alpha\beta} \to \Omega^2(\sigma) g_{\alpha\beta}. \tag{1.3.3}$$

This is known as *Weyl Invariance* and allows the $g_{\alpha\beta}$ to be written as the flat Minkowski metric $\eta_{\alpha\beta}$. Omitting the details, this is achieved by firstly using the σ_α reparametrisation invariance to pick a gauge in which the metric is locally conformally flat, i.e:

$$g_{\alpha\beta} = e^{2\varphi(\sigma_\alpha)} \eta_{\alpha\beta}. \tag{1.3.4}$$

Then using the Weyl invariance to set $\varphi = 0$ [15]. With this gauge choice, the Polyakov action reduces to the simple form:

$$S = -T \int d^2\sigma \, \partial^\alpha X \cdot \partial_\alpha X. \tag{1.3.5}$$

Where $(\cdot)$ indicates contraction through the target-space metric. Promoting the world-sheet metric to a dynamical field in the Polyakov action avoids the square-root present in the NG action. With this simplified action, the theory can now be

quantised. This can be most easily achieved through canonical quantisation. The first step in this procedure is to find the classical solutions to the EOM. Written in terms of *lightcone* coordinates ($\sigma^{\pm} = \tau \pm \sigma$), the EOM reads:

$$\partial_+ \partial_- X^\mu = 0. \tag{1.3.6}$$

In general, the solutions to eq. (1.3.6) will take the form:

$$X^\mu(\sigma_\alpha) = X^\mu_L(\sigma^+) + X^\mu_R(\sigma^-). \tag{1.3.7}$$

The labels L and R refer to left moving and right moving waves respectively. For the closed string sector, the equations can be solved subject to the constraint that the sting is periodic $X^\mu(\tau, \sigma) = X^\mu(\tau, \sigma + 2\pi)$. These solutions are:

$$X^\mu_L(\sigma^+) = \frac{1}{2}x^\mu + \frac{1}{2}\alpha' p^\mu \sigma^+ + i\sqrt{\frac{\alpha'}{2}} \sum_{n\neq 0} \frac{\tilde{\alpha}^\mu_n}{n} e^{-in\sigma^+}. \tag{1.3.8}$$

$$X^\mu_R(\sigma^-) = \frac{1}{2}x^\mu + \frac{1}{2}\alpha' p^\mu \sigma^- + i\sqrt{\frac{\alpha'}{2}} \sum_{n\neq 0} \frac{\alpha^\mu_n}{n} e^{-in\sigma^-}. \tag{1.3.9}$$

Where $\alpha' = 1/(2\pi T)$. In addition to the EOMs, the Euler-Lagrange equations also provide two constraint equations of the form:

$$(\partial_+ X)^2 = (\partial_- X)^2 = 0. \tag{1.3.10}$$

Applying these to the solutions gives:

$$(\partial_- X)^2 = \alpha' \sum_n L_n e^{-in\sigma^-}; \tag{1.3.11}$$

$$(\partial_+ X)^2 = \alpha' \sum_n \tilde{L}_n e^{-in\sigma^+}. \tag{1.3.12}$$

Where,

$$L_n = \frac{1}{2}\sum_m \alpha_{n-m} \cdot \alpha_m; \tag{1.3.13}$$

$$\tilde{L}_n = \frac{1}{2}\sum_m \tilde{\alpha}_{n-m} \cdot \tilde{\alpha}_m. \tag{1.3.14}$$

For $n \in \mathbb{Z}$ and α_0^μ is defined to be:

$$\alpha_0^\mu = \tilde{\alpha}_0^\mu = \sqrt{\frac{\alpha'}{2}} p^\mu. \tag{1.3.15}$$

Here the L_n and $\tilde{L}_n$ can be interpreted as the Fourier decomposition of the constraints eq. (1.3.10), which themselves obey the constraints:

$$L_n = \tilde{L}_n = 0 \quad n \in \mathbb{Z}. \tag{1.3.16}$$

Looking at the $n = 0$ constraints, one finds (using eq. (1.3.15)) that:

$$\frac{\alpha'}{2} p_\mu p^\mu + \sum_{n>0} \alpha_{n-m} \cdot \alpha_m = 0. \tag{1.3.17}$$

Noting that, in Minkowski space, $p_\mu p^\mu$ is equivalent to the effective mass squared $p_\mu p^\mu = -M^2$, we may write:

$$M^2 = \frac{4}{\alpha'} \sum_{n>0} \alpha_{n-m} \cdot \alpha_m. \tag{1.3.18}$$

Crucially, because $\alpha_0^\mu = \tilde{\alpha}_0^\mu$, we can also equate this to the sum of left-moving oscillators:

$$M^2 = \frac{4}{\alpha'} \sum_{n>0} \tilde{\alpha}_{n-m} \cdot \tilde{\alpha}_m. \tag{1.3.19}$$

Having two expressions for the mass allows us to relate the left movers to the right movers. This is known as the *level matching* condition.

To quantise this theory, the string position vector X_μ and its conjugate momenta $\Pi_\mu = \dot{X}_\mu/(2\pi\alpha')$ can be promoted to operators with the canonical commutation relations:

$$[X^\mu(\sigma,\tau), \Pi_\nu(\sigma',\tau)] = i\delta(\sigma-\sigma')\delta^\mu_\nu, \tag{1.3.20}$$

$$[X^\mu(\sigma,\tau), X^\nu(\sigma',\tau)] = [\Pi_\mu(\sigma,\tau), \Pi_\nu(\sigma',\tau)] = 0. \tag{1.3.21}$$

Rewriting these in terms of the Fourier modes gives the following non-zero commutators:

$$[\alpha^\mu_n, \alpha^\nu_m] = [\tilde{\alpha}^\mu_n, \tilde{\alpha}^\nu_m] = n\eta^{\mu\nu}\delta_{n+m,0}. \tag{1.3.22}$$

Drawing parallels with single particle quantum mechanics, we can interpret the α^μ_n and $\tilde{\alpha}^\mu_n$ and annihilation operators and α^μ_{-n} and $\tilde{\alpha}^\mu_{-n}$ are creation operators for right and left moving modes respectively. Upon a generic vacuum string state with momentum eigenvalue p^μ, the annihilation operators act with the following effect:

$$\alpha^\mu_n|0\rangle = 0, \quad \tilde{\alpha}^\mu_n|0\rangle = 0, \quad \text{for } n > 0. \tag{1.3.23}$$

Akin to the classical string picture, a level matching condition also exists for the quantised operators. Some additional complexities arise as a result of dealing with operators instead of numbers. The details of this are not relevant here, although the result is an important one. The quantum mechanical version of the level-matching condition is:

$$M^2 = \frac{4}{\alpha'}\left(N - \frac{D-2}{24}\right) \tag{1.3.24}$$

$$= \frac{4}{\alpha'}\left(\tilde{N} - \frac{D-2}{24}\right). \tag{1.3.25}$$

Where N and $\tilde{N}$ can be thought of as number operators given by:

$$N = \sum_{i=1}^{D-2} \sum_{n>0} \alpha_{-n}^{i} \alpha_{n}^{i}, \tag{1.3.26}$$

$$\tilde{N} = \sum_{i=1}^{D-2} \sum_{n>0} \tilde{\alpha}_{-n}^{i} \tilde{\alpha}_{n}^{i}. \tag{1.3.27}$$

We can now extract the physical fields from the theory. In string theory, the familiar fields of QFT and general relativity present in the form of string excitations. In fact, the first particle that appears comes from the ground state of the system. Substituting $N = 0$ into eq. (1.3.27) generates a particle with a mass squared given:

$$M^2 = -\frac{1}{\alpha} \frac{D-2}{6}. \tag{1.3.28}$$

This particle is knowns as the *tachyon* and its complex mass is one of the ways in which this simple bosonic string theory breaks down. Axion states do still occur even in this over-simplified theory. Axions and gravitons result from the first string harmonic. This can be generated by acting with a creation operator $\tilde{\alpha}_{-1}^{i}$ or α_{-1}^{i}. It might seem then that there are two possible first harmonics, one that is left-moving and one that is right-moving. The level-matching condition, however, requires that there be an equal number of left modes as there are right, hence the first harmonic can be written:

$$\tilde{\alpha}_{-1}^{i} \alpha_{-1}^{i} |0, p\rangle. \tag{1.3.29}$$

The states this generates have a mass of:

$$M^2 = \frac{4}{\alpha'} \left(1 - \frac{D-2}{24}\right). \tag{1.3.30}$$

Due to the symmetries in the Lagrangian, states that are generated by a creation operator must transform as a vector under the Poincaré group ($SO(1, D-1)$). The

state generated by eq. (1.3.29) must therefore transform as a tensor. It can be shown that, for this to be the case, these states must be massless. Consequently, the target space must have dimension $D = 26$. Working with massless states and in the current gauge, the Poincaré symmetry reduces to $SO(24)$ with the state eq. (1.3.29) transforming in the $\mathbf{24} \otimes \mathbf{24}$ (vector $\otimes$ vector) representation. In principle, one could now associate a particle field with this. However $SO(24)$, is not irreducible so this field would not be fundamental. In addition, the indices i and j don't run over the entire target space. $SO(24)$ can be written as the direct sum of three irreducible representations. These correspond to a traceless-symmetric part, an anti-symmetric part and a singlet. This gives three separate fields $G_{\mu\nu}(X)$, $B_{\mu\nu}(X)$ and the singlet $\Phi(X)$. The first and last field are associated with the *graviton* and *diltaton* respectively, but it is the second field $B_{\mu\nu}(X)$ that gives rise to ALPs.

ALP states generally occur through the process of compactification, whereby the 26 dimensional space required for string theory is reconciled with our 4 dimensions of space-time. On the compactified manifold, $D - 4$ of the dimensions are cyclically warped back onto themselves. This is akin to rolling an infinitely long sheet of paper along one direction to form a cylinder. On a macroscopic scale, this cylinder will appear to be a 1 dimensional line. The topology of the compactified manifold can have non-trivial effects on the resultant particle spectrum, but most structures will result in ALPs. After compactification, the $B_{\mu\nu}$ field can be split into spacetime components $B_{\mu\nu}$; $\mu, \nu \in [1, 4]$, mixed components $B_{a\mu}$; $a \in [5, 26]\, \mu \in [1, 4]$ and compactified components B_{ab}; $a, b \in [5, 26]$. It is these compact components that become the ALPs. To produce an ALP, the compactified dimensions must be integrated over. The number of ALPs generated is given by the number of distinct surfaces over which the integration can be performed. B_{ab} are the components of a 2-form field $B = B_{ab} dX^a \wedge dX^b$ which defines a 2 dimensional volume element. The integration surfaces must therefore be 2 dimensional. Roughly, a 2 dimensional closed surface is known as a 2-cycle. As an analogy, one can consider 1-cycles on a Torus. In this case, there are 2 distinct 1-cycles, corresponding to a loop around

the body and a loop around the hole (see 1.2). The number of distinct k-cycles is given by the k^{th} *Bretti number*. The number of ALP modes is given by the second Bretti number, with each ALP mode b_i given by the following integration over a given 2-cycle Σ_i:

$$b_i = \int_{\Sigma_i} B \tag{1.3.31}$$

These scalar fields can be shown to behave as ALPs in effective theories built by substituting the new field spectra back into the string Lagrangian. The exact properties of these ALPs depends on the form of the string Lagrangian and the compactification scheme. In particular, the decay constant is related to the volume ($\mathcal{V}_{\Sigma_i}$) of a given 2-cycle (Σ_i) as follows:

$$f_a^2 \sim \frac{1}{\mathcal{V}_{\Sigma_i}} \tag{1.3.32}$$

Relating f_a to the full compactified volume $\mathcal{V}$, one can write:

$$f_a^2 \sim \frac{M_p^2}{\mathcal{V}^\alpha} \tag{1.3.33}$$

Where M_p is the plank mass and α is a theory dependent parameter [16].

Some theories predict ALP spectra containing over 100 particles with distinct masses [17–21]. Few studies have considered the effect that large numbers of ALP states could have on phenomenology. We explore some novel results in this regard in the following chapter.

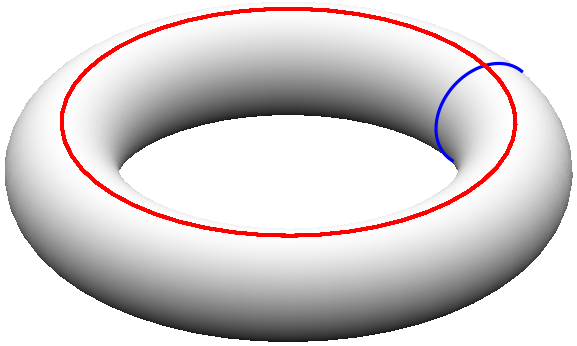


Figure 1.2: The 1-cycles around a torus. The two topologically distinct 1-cycles are shown in blue and red.

Chapter 2

ALP Anarchy

As discussed in chapter 1, ALPs are a well motivated extension to the SM, with the original QCD axion providing a resolution to the strong CP problem and generic ALPs occurring in large numbers in the particle spectra of string theories. In addition, ALPs have also emerged as one of the leading candidate for cold dark matter [22–26]. Unlike the QCD axion, ALPs need not couple to gluons; therefore, their mass and SM couplings are not necessarily related. ALPs are pseudo-scalar fields and are singlets under the SM gauge group. This determines the form of their interactions with the SM fields. For example, the Lagrangian of a single ALP φ interacting with photons and electrons is given by

$$\mathcal{L} \supset -\frac{1}{2}\partial^{\mu}\varphi\partial_{\mu}\varphi - \frac{1}{2}m^{2}\varphi^{2} - g^{\gamma}\varphi\tilde{F}^{\mu\nu}F_{\mu\nu} + \frac{g^{e}}{2m_{e}}\bar{\psi}\gamma^{\mu}\gamma_{5}\psi\partial_{\mu}\varphi\,, \tag{2.0.1}$$

where g^{γ} is the coupling of φ to the electromagnetic field strength tensor, $F^{\mu\nu}$, g^{e} is the dimensionless ALP-electron coupling, m_e is the electron mass and ψ denotes the electron field.

Models containing many ALPs may differ qualitatively in their phenomenology from those containing a single axion or ALP. Recent works have explored many ALP scenarios in a range of contexts [27–44]. The goal of this chapter is to explore the interaction of many ALP models with the SM. As pointed out in Ref. [39], this can

be understood by considering the misalignment between the ALPs' mass basis and the basis in which only one linear combination of ALPs couples to electromagnetism. As shown below, this results in oscillations between the electromagnetically coupled ALP and a number of orthogonal hidden ALP states where the physics at work is analogous to that of neutrino oscillations. This effect has also been studied in the context of Kaluza-Klein axions [45]. This chapter will expand on the results of Ref. [39] by considering several new aspects of these ALP oscillations. In particular, we will introduce ALP anarchy models, similar to the anarchy approach that has been used in neutrino physics to explain the structure of the leptonic mixing matrix. We will show that these many ALP models display dramatically different phenomenology to single ALP models with the same effective couplings. The framework developed here also applies to any system with multiple axion or ALP fields.

This chapter is structured as follows. In section 2.1 we will introduce many ALP models and the oscillation effect. In section 2.2 we will introduce ALP anarchy models. In section 2.3, section 2.4 and section 2.5 we will calculate the phenomenology of ALP anarchy models in the CERN Axion Solar Telescope, in magnetic white dwarfs and in the gamma-ray spectra of distant blazars respectively. Finally, in section 2.6, we will discuss our results further and conclude.

2.1 Axion Oscillations

We will consider a model containing N ALPs, each coupling to both photons and electrons:

$$\mathcal{L} \supset \sum_{i=1}^{N} \left(-\frac{1}{2}\partial^{\mu}\varphi_i \partial_{\mu}\varphi_i - \frac{1}{2}m_i^2\varphi_i^2 - g_i^{\gamma}\varphi_i \tilde{F}^{\mu\nu}F_{\mu\nu} + \frac{g_i^e}{2m_e}\bar{\psi}\gamma^{\mu}\gamma_5\psi\partial_{\mu}\varphi_i \right), \qquad (2.1.1)$$

where φ_i is an ALP field with mass m_i, g_i^{γ} is the coupling of φ_i to the electromagnetic field strength tensor, $F^{\mu\nu}$, g_i^e is the dimensionless ALP-electron coupling, m_e is the

electron mass and ψ denotes the electron field. In this chapter, we will consider only ALP interactions with photons and electrons, but similar considerations would apply to other couplings. We note that eq. (2.1.1) is in the mass and kinetic basis. In this work, we will consider string-motivated ultra-light ALPs, and hence, we will assume that the QCD axion (which may also emerge from string theory) is too heavy or too weakly coupled to contribute to the scenarios considered here. Furthermore, string ALPs with masses heavier than the energy scales considered below will also not contribute. It will be convenient to rotate to the electromagnetic basis, in which only a single ALP (the 'electromagnetic ALP') couples to photons:

$$\mathcal{L} \supset -\sum_i \frac{1}{2}\partial^\mu\varphi_i\partial_\mu\varphi_i - \sum_{i,j}\frac{1}{2}M^\gamma_{ij}\varphi_i\varphi_j + \sum_i \frac{g^e_i}{2m_e}\bar{\psi}\gamma^\mu\gamma_5\psi\partial_\mu\varphi_i - g^\gamma\varphi_1\tilde{F}^{\mu\nu}F_{\mu\nu}\,, \quad (2.1.2)$$

where $g^\gamma = \sqrt{\sum_i g_i^{\gamma 2}}$ is the total effective ALP-photon coupling, we have chosen $\varphi_1 = \sum_i g_i^\gamma\varphi_i/g^\gamma$ as the electromagnetic ALP and φ_i and g^e_i have been appropriately redefined. This basis has the advantage that only φ_1 interacts directly with the photon. The other ALPs, $\varphi_{\{2\ldots N\}}$, are 'hidden' with respect to the electromagnetic interaction.

For example, any ALP production via the electromagnetic interaction *in cases where the mass is irrelevant* will produce the electromagnetic ALP state φ_1. The production rate can, in this case, be calculated simply by considering a single ALP with coupling g^γ to photons. However, this does not necessarily mean we can treat the system as though there is only one ALP with coupling to photons g^γ. As we assume that the ALP mass states differ and that the mass and electromagnetic basis are misaligned, φ_1 may mix with the hidden states, $\varphi_{\{2\ldots N\}}$. This is an analogous effect to neutrino oscillations resulting from misalignment between the neutrinos' interaction and mass eigenbases.

Similarly, we can define the electronic basis in which only one ALP (the 'electronic ALP') couples to electrons:

$$\mathcal{L} \supset -\sum_i \frac{1}{2}\partial^\mu \varphi_i \partial_\mu \varphi_i - \sum_{i,j} \frac{1}{2} M^e_{ij} \varphi_i \varphi_j - \sum_i g^\gamma_i \varphi_i \tilde{F}^{\mu\nu} F_{\mu\nu} + \frac{g^e}{2m_e} \bar{\psi}\gamma^\mu\gamma_5\psi\partial_\mu\varphi_1 \,, \quad (2.1.3)$$

where $g^e = \sqrt{\sum_i g_i^{e2}}$ and we have now chosen $\varphi_1 = \sum_i g^e_i \varphi_i / g^e$ as the electronic ALP and φ_i and g^γ_i have been appropriately redefined. As with the electromagnetic ALP, this basis has the advantage that only φ_1 interacts directly with the electron while the other ALPs $\varphi_{\{2\ldots N\}}$ are 'hidden' with respect to the electron interaction. Note that the electronic and electromagnetic ALP states are generally neither orthogonal nor colinear. Hence, in scenarios where both ALP-photon and ALP-electron interactions are relevant, we must potentially consider three different bases - the mass, the electromagnetic, and the electronic. If other interactions between the ALPs and the SM are relevant, these may introduce further relevant bases.

As shown below, ALP oscillations are phenomenologically significant for many observations, particularly when we hope to detect an ALP that has propagated a large distance. However, ALP search strategies that rely only on the disappearance of SM particles or energy into ALP degrees of freedom, such as stellar cooling bounds on ALPs [46–48], are not significantly affected by ALP oscillations. Therefore, comparison between ALP searches is substantially more complicated in many ALP systems than if we assume only a single axion or ALP.

2.2 Anarchy Models

String theory provides a framework to understand ALP properties, including their mixing matrices that parameterise the misalignment between the interaction and mass bases. Calculating these mixing matrices is a highly non-trivial task [49]. Nonetheless, the ALP photon coupling has been modelled in a range of string axiverse scenarios [21, 50]. Such modelling of the electronic ALP properties from string theory has not been undertaken.

This current lack of knowledge of the mixing matrices, from a first principle string theory calculation, presents a challenge when motivating the choice of ALP mixing parameters and couplings. In this work, we circumvent this issue and remain agnostic to the ALPs' particular ultraviolet physics by considering a large set of randomly sampled mixing matrices. This framework, also known as *anarchy*, has been applied in neutrino physics and refers to the postulate that the neutrino mass matrix has no particular structure but that its elements are randomly chosen $\mathcal{O}(1)$ parameters [51–62]. Randomness in $\mathcal{O}(1)$ coupling constants is expected in sufficiently complicated models or with many fields mixing with each other. While it remains unclear if string theory predicts an anarchy-like mixing pattern, in this work, we use anarchy to explore the general properties of multi-ALP phenomenology and their relation to the number of ALP mass eigenstates.

To implement this anarchical approach and determine the mixing matrices between the hidden and visible ALPs, we assume that the non-diagonal ALP mass matrices, M^γ and M^e, are real and can, therefore, be related to the mass basis states as follows:

$$M^\alpha = U^\alpha D U^{\alpha T} \quad \alpha = e, \gamma \,, \tag{2.2.1}$$

where we assume $U^\alpha \in SO(N)$ and $D = \text{diag}\,(m_1, m_2, \ldots, m_N)$ is a real diagonal matrix with m_i denoting the mass of ALP field φ_i. Following our assumption that the electromagnetic and electronic bases are misaligned, a given model will be described by a set of two rotations U^α - one for each interaction basis. We sample $SO(N)$ such that elements are uniformly distributed over the group manifold to generate these mixing matrices. This can be achieved using the Haar measure, which describes the density of elements in a Lie group. In spherical coordinates, the Haar measure for $SO(N)$ is:

$$dV = \left(\prod_{i\in[1,N-1], j\in[i,N-1]} \sin^{i-1}\theta_{i,j} \right) d\theta_{1,1} \cdots d\theta_{N-1,N-1} \quad , \tag{2.2.2}$$

where there are $N(N-1)/2$ mixing angles, $\left\{\theta_{ij}\right\}_{1\leq i\leq j\leq N-1}$ with $\theta_{1,j} \in [0, 2\pi]$ and $\theta_{i,j} \in [0, \pi]$ for $i, j > 1$. Sampling uniformly in dV yields the desired distribution of the angles that parameterise the mixing matrices U^γ and U^e and we outline our numerical procedure for this task in appendix A. Mixing matrices, although providing a simple means to understand a given parametrisation, contain a large degree of redundancy. In practice, and in what follows, we are more interested in the relationship between the couplings in the mass basis – $\{g_i^\gamma\}$ and $\{g_i^e\}$. Given, for example, the EM coupling in the EM-ALP basis, the mass basis couplings are given by:

$$\begin{pmatrix} g_1^\gamma \\ g_2^\gamma \\ \vdots \end{pmatrix} = U^\gamma \begin{pmatrix} g^\gamma \\ 0 \\ \vdots \end{pmatrix}. \tag{2.2.3}$$

An analogous statement is also true for the electron-ALP couplings. From eq. (2.2.3), it is clear that the top row of U^γ parameterises the mixing of the electromagnetic ALP with the hidden states, which will have observable phenomenological consequences. The remaining $N-1$ rows determine the mixing *between* the hidden states, which is not observationally relevant.

We note that the mixing matrices U parameterise only the misalignment between the electromagnetic, electronic and mass bases and *not* the magnitude of the total effective couplings g^γ and g^e. This allows us to distinguish between the effects of basis misalignment and the effects of simply varying the total coupling, which is also present in the single ALP case. In the following sections, we will explore the phenomenology of ALP anarchy models. In particular, we will compare anarchy models with different numbers of ALPs to single ALP models with the same total effective couplings. We note here that, in a string axiverse scenario, the fundamental parameters are the couplings g_i^γ and g_i^e of the mass eigenstates and the effective couplings of the electromagnetic and electronic ALP states that emerge from these. In this case, the possible values of the effective couplings g^γ and g^e are determined

by the string model.

2.3 The CERN Axion Solar Telescope

In this section, we will outline the production of electromagnetic and electronic solar ALPs in the Sun and their detection by the CERN Axion Solar Telescope (CAST). We will discuss how oscillations, within the framework of an anarchical mixing pattern, influence the number of electromagnetic solar ALP states that reach the CAST detector. Finally, we will explain how we reinterpret the CAST experimental constraint on the single ALP parameter space for our multi-ALP scenario. Of the potential astrophysical sources, the Sun is one of the most accessible to the search for ALPs as the internal dynamics of the Sun are well understood and can be accurately modelled by a weakly coupled plasma. Within this setting, it is possible to find analytical forms for the emitted ALP flux, and these can be used to set competitive bounds on ALP couplings for mass ranges relevant to solar processes. ALPs are produced predominantly through three mechanisms: axio-bremsstrahlung (the Primakoff process) [63, 64], axio-recombination and axio-de-excitation [65–67], and Compton scattering [68–70], see Ref. [71] for a comprehensive overview. Of these processes, shown in fig. 2.1, only the Primakoff process depends on the ALP-photon coupling with the others arising from the ALP-electron interaction. CAST was a helioscope designed to use the Primakoff effect to scatter axions or ALPs produced by the Sun into photons. It consisted of a 9.2 m long evacuated cylinder with a sustained magnetic field oriented transverse to the ALP propagation direction. A bound on g^{γ} and g^{e} can be placed based on the lack of an ALP signal. In this work, we consider how the multi-ALP scenario will alter this bound. This modification occurs as the electromagnetic and electronic ALP states produced in the Sun may oscillate into hidden ALP states which would be undetectable using CAST.

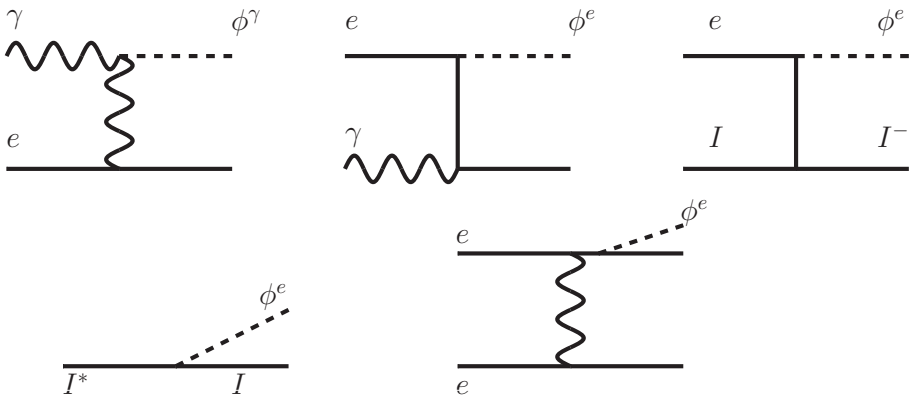


Figure 2.1: Top left image shows the Primakoff process that produces the electromagnetic ALP, and the middle and top right images show the Compton scattering and axio-recombination processes that produce the electronic ALP. The bottom row shows axio-de-excitation and Bremsstrahlung processes that produce the electronic ALP.

2.3.1 Solar ALP emission and detection

If ALPs couple to the electromagnetic field strength tensor or an electronic current, as in eq. (2.1.1), they can be produced in the Sun. The relevant ALP production processes are shown in fig. 2.1, and the dominant ALP production mechanisms are Bremmstrahlung (B), Compton Scattering (C) and the Primakoff process (P). In the well-studied single ALP case, with no oscillation into hidden ALPs, the fluxes at Earth, in units of $\mathrm{m}^{-2}\,\mathrm{year}^{-1}\,\mathrm{keV}^{-1}$, generated by these mechanisms are given by [72]:

$$\left.\frac{\mathrm{d}\Phi_a}{\mathrm{d}\omega}\right|_B = 8.3\times10^{20}\left(\frac{g_e}{10^{-13}}\right)^2 \frac{\omega}{1+0.667\omega^{1.278}} e^{-0.77\omega}\,, \tag{2.3.1}$$

$$\left.\frac{\mathrm{d}\Phi_a}{\mathrm{d}\omega}\right|_C = 4.2\times10^{18}\left(\frac{g_e}{10^{-13}}\right)^2 \omega^{2.987} e^{-0.776\omega}\,, \tag{2.3.2}$$

$$\left.\frac{\mathrm{d}\Phi_a}{\mathrm{d}\omega}\right|_P = 2.0\times10^{18}\left(\frac{g_\gamma}{10^{-12}\mathrm{GeV}^{-1}}\right)^2 \omega^{2.450} e^{-0.829\omega}\,, \tag{2.3.3}$$

where the ALP energy, ω, is in keV. A more detailed study of the solar ALP flux can be found in [73]. Integrating over the energy range ($0.8-6.8$ keV) of the CAST analysis [74], the total fluxes can be found for Φ_B, Φ_C and Φ_P:

$$\frac{\Phi_B}{\mathrm{m}^{-2}\,\mathrm{year}^{-1}} = 4.1 \times 10^{46} g^{e2}\,, \tag{2.3.4}$$

$$\frac{\Phi_C}{\mathrm{m}^{-2}\,\mathrm{year}^{-1}} = 5.2 \times 10^{45} g^{e2}\,, \tag{2.3.5}$$

$$\frac{\Phi_P}{\mathrm{m}^{-2}\,\mathrm{year}^{-1}} = 1.0 \times 10^{43} \left(\frac{g^{\gamma}}{\mathrm{GeV}^{-1}}\right)^2\,. \tag{2.3.6}$$

The CAST experimental constraints assume a single ALP that couples to electromagnetism and electrons. We will now consider the CAST signal from models with multiple ALP mass eigenstates.

2.3.2 Oscillation: The two ALP case

We will expand upon the results of [39] by discussing the basics of ALP oscillations relevant to CAST in a simplified two ALPs scenario where φ_e and φ_γ are linear combinations of only two massive ALP states, φ_1 and φ_2, which have masses m_1 and m_2 respectively. The Lagrangian in the mass basis is

$$\mathcal{L} \supset -\frac{1}{2}\partial^\mu\varphi_1\partial_\mu\varphi_1 - \frac{1}{2}\partial^\mu\varphi_2\partial_\mu\varphi_2 - \frac{1}{2}m_1^2\varphi_1^2 - \frac{1}{2}m_2^2\varphi_2^2 \tag{2.3.7}$$

$$- g_1^\gamma\varphi_1\tilde{F}^{\mu\nu}F_{\mu\nu} - g_2^\gamma\varphi_2\tilde{F}^{\mu\nu}F_{\mu\nu} + \frac{g_1^e}{2m_e}\bar{\psi}\gamma^\mu\gamma_5\psi\partial_\mu\varphi_1 + \frac{g_2^e}{2m_e}\bar{\psi}\gamma^\mu\gamma_5\psi\partial_\mu\varphi_2\,. \tag{2.3.8}$$

We can rotate to a basis where a single ALP field, the electromagnetic ALP, couples to electromagnetism while the other hidden electromagnetic ALP field does not. These are given by

$$\varphi_\gamma = \frac{g_1^\gamma\varphi_1 + g_2^\gamma\varphi_2}{\sqrt{g_1^{\gamma 2} + g_2^{\gamma 2}}}\,, \quad \varphi_{\gamma_h} = \frac{g_2^\gamma\varphi_1 - g_1^\gamma\varphi_2}{\sqrt{g_1^{\gamma 2} + g_2^{\gamma 2}}}\,, \tag{2.3.9}$$

respectively. Analogously, the electronic ALP and the hidden electronic ALP are the following orthogonal combinations:

$$\varphi_e = \frac{g_1^e \varphi_1 + g_2^e \varphi_2}{\sqrt{g_1^{e2} + g_2^{e2}}}\,, \quad \varphi_{e_h} = \frac{g_2^e \varphi_1 - g_1^e \varphi_2}{\sqrt{g_1^{e2} + g_2^{e2}}}\,, \tag{2.3.10}$$

respectively. Note that the electromagnetic and electron ALPs φ_γ and φ_e are generally neither colinear nor orthogonal. Therefore, φ_{γ_h} *will* in general have some non-zero coupling to electrons and φ_{e_h} *will* in general have some non-zero coupling to photons. The following unitary rotation matrices relate the mass basis and the electromagnetic and electron bases:

$$\begin{aligned} \begin{pmatrix} \varphi_\gamma \\ \varphi_{h_\gamma} \end{pmatrix} &= \begin{pmatrix} \cos(\theta^\gamma) & \sin(\theta^\gamma) \\ -\sin(\theta^\gamma) & \cos(\theta^\gamma) \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}, \\ \begin{pmatrix} \varphi_e \\ \varphi_{h_e} \end{pmatrix} &= \begin{pmatrix} \cos(\theta^e) & \sin(\theta^e) \\ -\sin(\theta^e) & \cos(\theta^e) \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}, \end{aligned} \tag{2.3.11}$$

with

$$\cos(\theta^\alpha) = \left(\frac{g_1^\alpha}{\sqrt{g_1^{\alpha 2} + g_2^{\alpha 2}}} \right), \quad \sin(\theta^\alpha) = \left(\frac{g_2^\alpha}{\sqrt{g_1^{\alpha 2} + g_2^{\alpha 2}}} \right), \tag{2.3.12}$$

where $\alpha = \gamma, e$.

We will assume that all couplings are real, and their values are determined using the anarchical approach outlined in section 2.2. We will also assume that m_1 and m_2 are much less than other relevant energy scales. In particular, we will assume $m_1, m_2 < 10^{-2}\,\text{eV}$, corresponding to CAST bounds for evacuated magnet bores. In this mass range, the ALP masses are also much lower than their production energy in the Sun and can be treated in the relativistic limit. This means their effects on ALP production may be neglected, so axio-recombination, axio-de-excitation and Compton scattering produce the state φ_e while Primakoff production produces the state φ_γ. Furthermore, in this mass range, the CAST sensitivity is independent of mass, and therefore CAST will detect the state φ_γ as a single signal.

As CAST aims to detect electromagnetic ALPs, φ_γ, we are interested in the probability that a solar electronic or electromagnetic ALP oscillates to an electromagnetic

ALP after travelling a distance L. We can calculate these probabilities using the fact that the mass eigenstates propagate as $|\varphi_i(L)\rangle = \mathrm{e}^{-i\frac{m_i^2 L}{2\omega}} |\varphi_i(0)\rangle$, where L is the distance travelled. The former is given by

$$P(\varphi_e \to \varphi_\gamma) \equiv P_{e\to\gamma} = \frac{1}{2}\left(1 + \cos(2\theta^e)\cos(2\theta^\gamma) + \sin(2\theta^e)\sin(2\theta^\gamma)\cos\left(\frac{\Delta m^2 L}{2\omega}\right)\right), \tag{2.3.13}$$

where $\Delta m^2 = m_2^2 - m_1^2$ is the mass squared splitting between the ALP mass states. Rewriting eq. (2.3.13) in terms of couplings, g^α, yields

$$P_{e\to\gamma} = \frac{\left(g_1^{e2} g_1^{\gamma 2} + g_2^{e2} g_2^{\gamma 2}\right)}{\left(g_1^{e2} + g_2^{e2}\right)\left(g_1^{\gamma 2} + g_2^{\gamma 2}\right)}\left(1 + \frac{2 g_1^e g_1^\gamma g_2^e g_2^\gamma}{\left(g_1^{e2} g_1^{\gamma 2} + g_2^{e2} g_2^{\gamma 2}\right)}\cos\left(\frac{\Delta m^2 L}{2\omega}\right)\right). \tag{2.3.14}$$

In addition to the electronic ALPs produced in the Sun oscillating to electromagnetic ALPs when they reach CAST, we must consider the survival probability of the solar electromagnetic ALP which is the usual two-state survival probability familiar from neutrino physics:

$$P(\varphi_\gamma \to \varphi_\gamma) \equiv P_{\gamma\to\gamma} = 1 - \sin^2(2\theta^\gamma)\sin^2\left(\frac{\Delta m^2 L}{4\omega}\right) \tag{2.3.15}$$

$$= 1 - 4\frac{g_1^{\gamma 2} g_2^{\gamma 2}}{\left(g_1^{\gamma 2} + g_2^{\gamma 2}\right)^2}\sin^2\left(\frac{\Delta m^2 L}{4\omega}\right). \tag{2.3.16}$$

Several simplifications can be made for solar ALP oscillations. Firstly, the matter potential induced by the solar electron background is negligibly small and therefore does not affect the electron ALP propagation through the Sun. This can be estimated from the fact that the potential experienced by electron neutrinos from electrons in the Sun is $V \approx G_F N_e \sim 10^{-12}\,\mathrm{eV}$ where G_F is Fermi's constant and N_e is the number density of electrons in the Sun's core. In contrast, the potential experienced by the electronic ALP (with coupling $\frac{g^e}{2m_e} = 10^{-11}\,\mathrm{GeV}^{-1}$) is $V \approx \left(\frac{g^e}{2m_e}\right)^2 N_e \sim 10^{-29}\mathrm{eV}$.

Likewise, the potential induced by the Sun's magnetic field is negligible, so vacuum oscillation between the ALP states will be applied. Secondly, for Sun-Earth distances and keV solar ALP with $\Delta m^2 > 10^{-12}\,\mathrm{eV}^2$, the oscillation probability of eq. (2.3.13) and eq. (2.3.14) will be averaged when integrated over the CAST energy range yielding:

$$P_{e\to\gamma} = \frac{\left(g_1^{e2} g_1^{\gamma 2} + g_2^{e2} g_2^{\gamma 2}\right)}{\left(g_1^{e2} + g_2^{e2}\right)\left(g_1^{\gamma 2} + g_2^{\gamma 2}\right)}\,, \quad P_{\gamma\to\gamma} = \frac{g_1^{\gamma 4} + g_2^{\gamma 4}}{\left(g_1^{\gamma 2} + g_2^{\gamma 2}\right)^2}\,. \tag{2.3.17}$$

2.3.3 Oscillation: The many ALP case

We now turn to the case where many ALP mass eigenstates couple to electrons and photons:

$$\mathcal{L} \supset \sum_i^N \left(\frac{1}{2}\partial^\mu \varphi_i \partial_\mu \varphi_i - \frac{1}{2} m_i^2 \varphi_{i^2} - g_i^\gamma \varphi_i \tilde{F}^{\mu\nu} F_{\mu\nu} + g_i^e \bar{\psi}\gamma^\mu \gamma_5 \psi \partial_\mu \varphi_i\right)\,, \tag{2.3.18}$$

such that the electromagnetic and electronic ALPs produced in the Sun are linear combinations of the mass states:

$$\varphi_\gamma = \frac{\sum_i g_i^\gamma \varphi_i}{\sqrt{\sum_i g_i^{\gamma 2}}}\,, \quad \varphi_e = \frac{\sum_i g_i^e \varphi_i}{\sqrt{\sum_i g_i^{e2}}}\,, \tag{2.3.19}$$

where we again assume that all ALP masses considered are $m_i < 10^{-2}\,\mathrm{eV}$. Any mass eigenstates with $m_i > 10^{-2}\,\mathrm{eV}$ would not contribute to the signal considered here as they would not produce a signal in CAST with an evacuated bore. Again, we will assume that $\Delta m^2 > 10^{-12}\,\mathrm{eV}^2$ so that the oscillation probabilities average when integrated over the CAST energy range. Under these conditions, it can be shown that the electromagnetic ALP survival probability is

$$P_{\gamma\to\gamma} = \frac{1}{N}\left(1 + \frac{N^2 \mathrm{VAR}\left(\{g_i^{\gamma 2}\}\right)}{g^{\gamma 4}}\right) = \frac{\sum_i^N g_i^{\gamma 4}}{\left(\sum_i^N g_i^{\gamma 2}\right)^2}\,, \tag{2.3.20}$$

where N is the number of ALP mass eigenstates, $g^\gamma = \sqrt{\sum {g_i^\gamma}^2}$ is the coupling of the electromagnetic ALP to photons, and $\mathrm{VAR}(\{{g_i^\gamma}^2\})$ is the variance of the ALP-photon couplings squared in the mass basis which is given by

$$\mathrm{VAR}\left(\{{g_i^\gamma}^2\}\right) = \frac{\sum_i^N \left[{g_i^\gamma}^2 - \frac{{g^\gamma}^2}{N}\right]^2}{N}\,. \tag{2.3.21}$$

Likewise, the probability that an electronic ALP oscillates to an electromagnetic ALP is

$$P_{e\to\gamma} = \frac{1}{N}\left(1 + \frac{N^2\mathrm{COVAR}\left(\{{g_i^\gamma}^2\},\{{g_i^e}^2\}\right)}{{g^\gamma}^2 {g^e}^2}\right) = \frac{\sum_i^N {g_i^e}^2 {g_i^\gamma}^2}{\sum_i^N {g_i^e}^2 \sum_i^N {g_i^\gamma}^2}\,, \tag{2.3.22}$$

where $g^e = \sqrt{\sum {g_i^e}^2}$ is the coupling of the electronic ALP to electrons, and the covariance of the ALP-photon and ALP-electron couplings squared is given by

$$\mathrm{COVAR}\left(\{{g_i^\gamma}^2\},\{{g_i^e}^2\}\right) = \frac{\sum_i^N \left[{g_i^\gamma}^2 - \frac{{g^\gamma}^2}{N}\right]\left[{g_i^e}^2 - \frac{{g^e}^2}{N}\right]}{N}\,. \tag{2.3.23}$$

2.3.4 Reinterpretation of CAST results

The non-observation of an excess number of photons allows CAST to place constraints on the (g^γ, g^e) parameter space, which bounds the solar ALP fluxes on Earth. In the multi-ALP case, the fluxes of the electromagnetic and electronic ALP on Earth, namely the oscillated fluxes, are, respectively:

$$\Phi_\gamma^{\mathrm{osc}} = P_{\gamma\to\gamma}\Phi_\gamma + P_{e\to\gamma}\Phi_e\,, \tag{2.3.24}$$

$$\Phi_e^{\mathrm{osc}} = P_{\gamma\to e}\Phi_\gamma + P_{e\to e}\Phi_e\,, \tag{2.3.25}$$

where $\Phi_\gamma = \Phi_P$ and $\Phi_e = \Phi_B + \Phi_C$ and $P_{a\to b}$ is the probability of species a oscillating into species b during propagation as derived in section 2.3.3. Since CAST makes

use of a strong magnetic field to convert ALPs, the relevant flux to consider for the recasting from the single to the multi-ALP scenario is $\Phi_\gamma^{\rm osc}$. CAST has placed bounds on $(g^\gamma)^2$ as a function of ALP mass and on g^γ as a function of g^e, for small ALP mass m_a, for a single ALP. Here we determine an upper bound on g^γ as a function of g^e for multi-ALP scenarios with $m_a \leq 10^{-2}\,\text{eV}$ and $\Delta m^2 > 10^{-12}\,\text{eV}^2$ for all relevant ALP mass eigenstates; for which the bound becomes mass independent [74]. We do this by considering the maximum flux compatible with the non-detection of ALPs by CAST. Given the original single ALP bound on the ALP-electromagnetic coupling, which we denote as $g^\gamma_{N=1}(g^e)$, (shown by the black line in fig. 2.2), the maximum flux of electromagnetic ALPs

$$\Phi^{\rm Max}(g^e, g^\gamma) = \frac{\left(g^\gamma_{N=1}(g^e)\right)^2}{g^{\gamma\,2}} \Phi^{N=1}(g^e, g^\gamma_{N=1}(g^e))\,, \tag{2.3.26}$$

where $\Phi^{N=1} = \Phi_P + \Phi_B + \Phi_C$ is the flux at the detector in the single ALP case. With $\Phi_\gamma^{\rm osc}(g^e, g^\gamma)$ in hand, we now perform a grid scan over the coupling parameter space, (g^e, g^γ), to determine the new bound[1]. For a given point in the ALP anarchy parameter space to be allowed by existing CAST data, the total EM-ALP flux at the detector given in eq. (2.3.24) must be less than $\Phi^{\rm Max}$.

To determine the bound on the total effective couplings in the ALP anarchy scenario, it is therefore necessary to compute $P_{\gamma\to\gamma}$ and $P_{e\to\gamma}$. From section 2.3.3, these are given by eq. (2.3.20) and eq. (2.3.22), respectively, and can be written in terms of the relationship between the individual ALP couplings (g_i^e and g_i^γ). It is this relationship that encodes the mixing between the various ALP states. We consider 10^4 realisations of $\{g_i^\gamma\}$ and $\{g_i^e\}$ for each overall coupling pair (g^γ, g^e). For each point in (g^γ, g^e) space, we determine the proportion of viable realisations such that $\Phi_\gamma^{\rm osc} < \Phi^{\rm Max}(g^e, g^\gamma)$. These results are shown in fig. 2.2; for a detailed outline of this numerical procedure, see appendix B. The black line indicates the original $N = 1$ bound ($\varphi_e \equiv \varphi_\gamma$) where in the mass independent region $\log_{10}\left(g^\gamma\,[\text{GeV}^{-1}]\right) \lesssim -10$.

[1] To accelerate this procedure we developed a C++ package with python bindings, allowing for optimised parallel computation through the use of `OpenMP`.

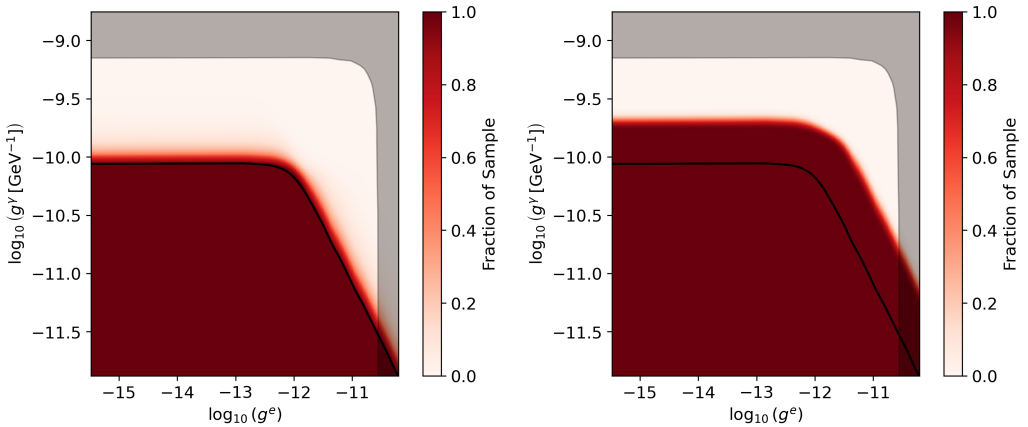


Figure 2.2: The fraction of realisations consistent with non-detection as a function of coupling (g^e and g^γ) shown for two different values of N – Left: $N = 2$; Right: $N = 30$. The grey region indicates the excluded region from solar neutrinos [2].

The boundary between regions where 0 and 100% of realisations are viable can be interpreted as the bound on g^γ as a function of g^e in the ALP anarchy scenario.

From fig. 2.2, it can be seen that increasing the number of ALPs decreases the competitiveness of the bound. The left plot of fig. 2.2 shows the bound with $N = 2$ ALP fields, and we observe that the effect of an additional state is marginal; however, for $N = 30$, we observe that the bound on g^γ is relaxed by almost half an order of magnitude with $\log_{10}(g^\gamma\,[\mathrm{GeV}^{-1}]) \lesssim -9.6$. As the number of hidden states increases, the oscillated flux of electromagnetic ALPs on Earth decreases since they can oscillate into the hidden ALP states that are not detectable by CAST. Hence, the effective coupling g^γ can increase to compensate for this decrease in the detectable flux.

To quantify this relationship, we consider ALP multiplicities $N \in [2, 30]$. For each N in this set, we determine the value of g^γ, $g^\gamma_{50}(N)$ for which 50% of mixing realisations satisfying $\Phi^{\mathrm{osc}}_\gamma < \Phi^{\mathrm{Max}}(g^e, g^\gamma)$ at a point in the horizontal region of (fig. 2.2). We fix $g^e = 10^{-15}$. For this low ALP-electron coupling, the production processes for φ_e, namely bremsstrahlung and Compton scattering, are ineffective, and the production of φ_γ in the Sun dominates. $g^\gamma_{50}(N)$ can be interpreted as an approximate bound on g^γ in the ALP anarchy scenario. This result is shown in fig. 2.3, which we fit the following function to:

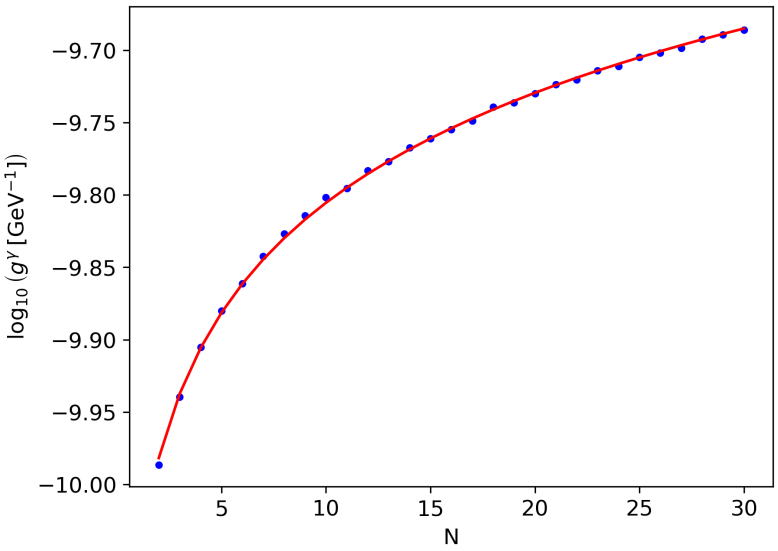


Figure 2.3: Bounds on photon ALP coupling (g^γ) as a function of number of ALPs (N) for $g^e = 10^{-15}$. The scatter plot depicts the minimum upper bound consistent with 50% of mixing matrix realisations. The fit to that is shown as a line plot.

$$g^\gamma = e^{m \log N - c}, \tag{2.3.27}$$

where we find $m = 0.25$ and $c = -23$. This fitting function is also shown in fig. 2.3.

We can understand the dependence of the bound on N - $g^\gamma_{50}(N) \propto N^{1/4}$. We first note that when electromagnetic ALP production dominates, CAST is sensitive to $g^{\gamma 4}$, as the ALP must be produced and detected via this coupling to photons. In the many ALP scenario, the bound placed on $g^{\gamma 4}$ is weakened in proportion to the electromagnetic ALP survival probability $P_{\gamma\to\gamma}$ given in eq. (2.3.20). For large N, the variance term becomes negligible and we have $P_{\gamma\to\gamma} \sim \frac{1}{N}$. We therefore obtain $g^\gamma_{50}(N) \propto N^{1/4}$, as found numerically.

We have seen that if the ALP-photon and ALP-electron interactions are an effect from multiple ALP mass eigenstates, the CAST bounds on the total effective couplings may be somewhat reduced. We have calculated this reduction in the ALP anarchy scenario for mass differences $\Delta m^2 > 10^{-12}\,\mathrm{eV}^2$. As shown in [39], for $\Delta m^2 < 10^{-14}\,\mathrm{eV}^2$, there is no significant oscillation into hidden states. For intermediate mass differences,

there is a non-trivial oscillation structure depending on ω.

2.4 Magnetic white dwarfs

Having considered the constraints from the CAST experiment, we now examine how the limits on the single ALP electromagnetic and electronic coupling from observations of magnetic white dwarfs (MWDs) can be applied to our multi-ALP scenario. MWDs produce $\omega \sim$ keV energy ALPs via axio-bremsstrahlung, which can convert to X-rays in the magnetosphere surrounding the MWD. Subsequently, searches for observable X-ray signals provide one of the most stringent constraints on the ALP parameter space, see e.g. Ref. [3]. More specifically, in the single ALP scenario, where $N = 1$ and $\varphi_e \equiv \varphi_\gamma$, the (g^e, g^γ) parameter space is constrained from the non-observation of astrophysical X-ray emission from the MWD RE J0317-853 by the Suzaku telescope [75]. The flux of ALP-induced X-ray photons on Earth, in the low ALP mass regime, is approximately proportional to $\Phi_{\text{X-ray}} \propto (g^e g^\gamma)^2$ as the ALP luminosity is proportional to ${g^e}^2$ while the probability the ALP transitions to X-ray photons is proportional to ${g^\gamma}^2$. The non-observation of excess X-rays provides an upper bound on $\Phi_{\text{X-ray}}$ and hence a corresponding bound on (g^e, g^γ).

In the multi-ALP case, since the radius of the MWD is relatively small (less than one percent of the Sun's radius [76]), the oscillations between the electronic ALP and hidden states do not have time to develop, assuming $\Delta m^2 \lesssim 4R = 10^{-10}\,\text{eV}^2$, where R is the MWD's radius. Moreover, we assume that all ALP masses considered are $m_i < 10^{-2}\,\text{eV}$. To perform a simple recast of the single to the multi-ALP case, we compute the conversion probability ($P_{e\to\gamma}$) and scale the $N = 1$ bound on $g^e g^\gamma$, denoted as $b_{N=1}$, as follows:

$$b_{N>1} = b_{N=1}/P_{e\to\gamma} \tag{2.4.1}$$

where $b_{N>1}$ is the new bound on $g^e g^\gamma$ assuming the existence of N ALP states.

A decrease in the conversion probability will allow for a greater possible coupling strength. In the MWD setting, in the case where oscillations do not have time to develop, $P_{e\to\gamma}$ is given by the scalar product of φ_e and φ_γ:

$$P_{e\to\gamma} = \left| \frac{\sum_i g_i^e g_i^\gamma}{g^e g^\gamma} \right|^2 \tag{2.4.2}$$

Note that this is a different limit to that considered in section 2.3 as the distance propagated is much lower. The resulting bound as a function of number of axions is shown in fig. 2.4 for values: $L = 4 \times 10^{-3}$ Solar Radii [76], $E = 5\,\text{keV}$ and masses distributed logarithmically in $[10^{-9}, 10^{-6}]\,\text{eV}$. The conversion probability, accounting for the presence of hidden ALPs (in the simple two-ALP case, see eq. (2.3.13)), is given by

$$P_{e\to\gamma} = \cos^2(\theta^e - \theta^\gamma)\,, \tag{2.4.3}$$

which describes the projection of the electronic ALP state onto the EM ALP state. In the limit $\theta^e = \theta^\gamma$, naturally, the probability $P_{e\to\gamma} = 1$ because the electron and EM ALP states are completely aligned, and the $N = 1$ case is recovered. If these angles are very different, then the probability $P_{e\to\gamma}$ can be significantly smaller and make the bound proportionally weaker.

In fig. 2.4, we show how the limit on $|g^\gamma g^e|$, in the low ALP mass regime, from Ref. [3] changes as the number of hidden states is increased. We observe that as the number of hidden states is increased, the total allowed coupling strength of the ALP to EM and electrons increases. This occurs because increasing the number of hidden states leads to an increase in the typical orthogonality between the electromagnetic and electronic ALP states, decreasing the chance that an ALP produced in an electron interaction will convert into a photon. We find that the bound decreases almost three orders of magnitude as the number of ALP mass states increases from 2 to 18.

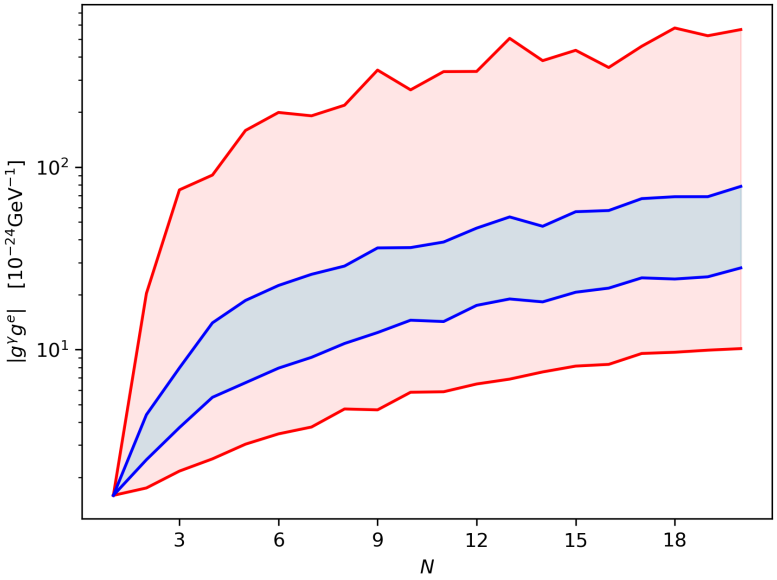


Figure 2.4: The bounds on $g_\gamma g_e$ as a function of N for the low mass ($m_a < 10^{-6}$eV), low propagation limit of the MWD. The bounds were computed for a set 1000 coupling pairs (g^e and g^γ). The central 90% of bounds lie within the red band, with the central third being encompassed by the blue band. The one ALP bound was recast from Ref. [3].

2.5 Very high-energy blazars

In this final section, we examine how our multi-ALP scenario can be constrained by the observation of very high-energy gamma rays. We begin by outlining the simulation of the propagation of these high-energy photons emitted by blazars toward Earth, considering the possibility of photon mixing with multiple ALP states. We detail the density matrix equations and the model of the magnetic fields used in the simulation. Additionally, we discuss the selection of blazars and how these sources can be utilised to constrain the electromagnetic coupling as a function of the number of ALPs in our multi-ALP scenario.

Blazars produce a large flux of Very High Energy (VHE) photons. These TeV scale photons can scatter off the isotropic extragalactic background light (EBL) as they propagate to Earth, producing positron-electron pairs. The probability with which this scattering occurs increases with energy. As such, we expect significant attenuation of high-energy photons travelling through intergalactic space. Telescope

observations suggest that the Universe may be more transparent than expected to VHE photons [77, 78], although the evidence for this effect is not conclusive [79, 80]. Several phenomenological studies have introduced ALPs to explain this discrepancy [81–84]. VHE photons can oscillate into ALPs in the magnetic field of the blazar or the intergalactic medium and thus travel unimpeded through the Universe. These oscillations may conspire for appropriate masses and couplings, such that the ALPs reconvert into photons in the Milky Way, allowing for their detection on Earth. In this case, the measured flux of VHE photons on Earth will be amplified, accounting for the observations. In this section, we will explore this effect in the context of our multi-ALP model. Following and extending the analysis of Ref. [84], we consider an arbitrarily mass-mixed set of ALP states. We note that the ALP-electron coupling does not play a significant role in this process, and therefore will not be considered in this section.

2.5.1 Simulation

In the following subsections, we describe the approach by which we simulate the propagation of VHE photons from their blazar source to Earth. We consider both photon-EBL scattering and ALP-photon mixing. As the ALP mass is relevant for this effect, we work in the mass basis where each ALP couples separately to the photon. We again determine these couplings anarchically. Photon-EBL scattering is dissipative (VHE photons can scatter off EBL photons, creating positron-electron pairs) and introduces non-unitarity into the evolution. To account for this effect, we use a density matrix formalism.

Our simulation can be broken down into three spatial regions: propagation through the galaxy cluster hosting the blazar; propagation through the intergalactic medium (IGM); and propagation through the Milky Way (MW). The principal difference between each region is the form of the magnetic field present. Photon to ALP conversion occurs most readily in the galaxy cluster where the VHE-photon density

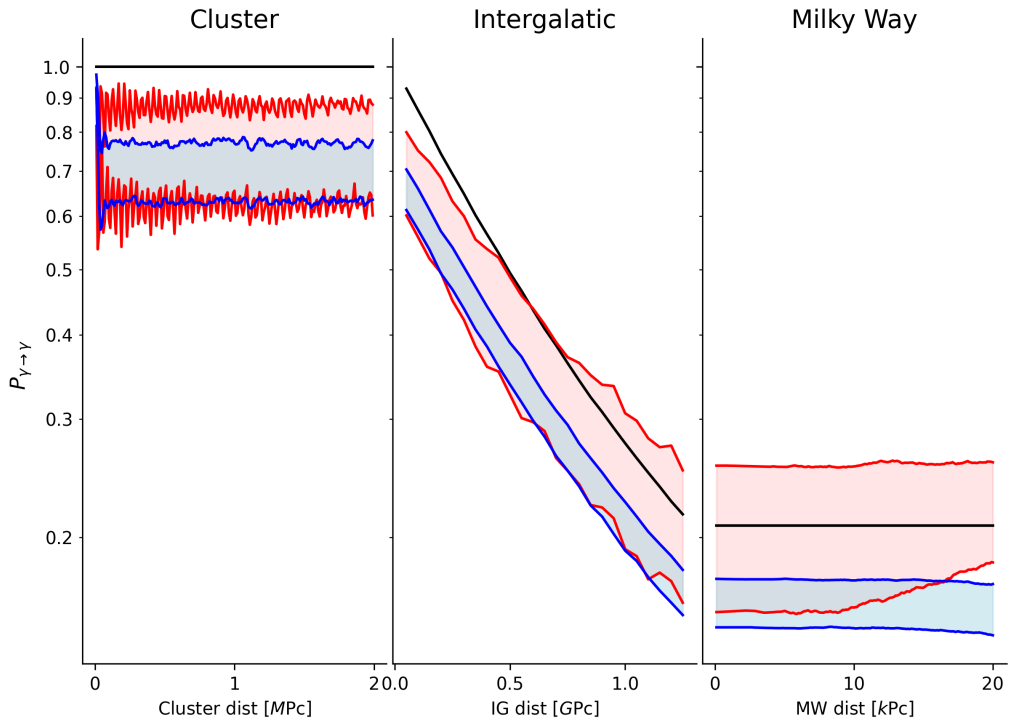


Figure 2.5: Photon survival probability against propagation distance for a photon energy of 400 GeV produced by 1ES0414+009. The zero ALP case is shown in black, with the central third of samples shown in red and blue for the 1 ALP and 20 ALP cases respectively.

and magnetic fields are large. As the VHE photon/ALP beam propagates out of the cluster, the relatively strong cluster fields give way to a much weaker intergalactic field, largely suppressing any ALP-photon processes. The vast IGM is instead responsible for the attenuation of photons by EBL scattering. Finally, once the blazar jet reaches the Milky Way, we expect to see ALP-photon back-conversion induced by the strong galactic magnetic fields.

The effect of each propagation region is depicted in fig. 2.5 which shows the photon survival probability of a photon with $E = 400\,\text{GeV}$ as a function of the distance of propagation from the blazar. The solid black line indicates the scenario without ALPs ($N = 0$), while the red and blue regions correspond to the scenarios of $N = 1$ and $N = 20$ ALP states respectively. In the no-ALP scenario, the survival probability remains unity until propagation through the IGM where, due to EBL scattering, it decreases to ~ 0.2. For the cases that include ALPs, a number of model unknowns

associated with oscillation lead us to consider a large set possible $P_{\gamma\to\gamma}$ – each element generated with a unique B-field structure and, for $N > 1$, an anarchical choice of $\{g_i^\gamma\}$. The central third of these $\{P_{\gamma\to\gamma}\}$ are indicated by the filled regions in the figure. In both the one and 20 ALP cases, we see the effect of photon to ALP conversion in the galaxy cluster. In the one ALP case, we see a significant increase in the photon survival probability as the beam travels through the Milky Way, corresponding to the reconversion of ALPs to photons. However, this increase is not present in the 20 ALP case. This is because the electromagnetic ALP produced in the galaxy cluster is very likely to oscillate into a hidden ALP before it reaches the Milky Way. Due to oscillation into hidden ALPs, if the ALP-photon coupling is spread over 20 ALPs, we always expect a *decrease* in the blazar luminosity rather than the increase seen in the single ALP case.

The galaxy cluster

The simulation begins with a jet of pure photons at the site of the blazar. These photons are unpolarised, hence we take our initial state to be the following the density matrix:

$$\rho_\gamma = 1/2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots \end{bmatrix}, \tag{2.5.1}$$

where the first two diagonal elements correspond to the photons' two polarisations, and the remaining diagonal elements correspond to the ALP mass eigenstates; off-diagonal elements are associated with a superposition of states. To facilitate comparison between the multi-ALP and single ALP effects, we follow [84] in our description of the magnetic fields. We take the cluster field to have a domain-like structure with a radially dependent magnitude given by:

$$B^C(r) = B_0^C(1 + (r/r_{\rm core})^2)^{-\eta}\,, \tag{2.5.2}$$

where, the core radius is $r_{\rm core} = 200\,{\rm kpc}$ and the central field strength is $B_0^C = 10\,\mu{\rm G}$ and $\eta = 0.5$. Within a given magnetic domain, the field is assumed to be constant. The direction of the magnetic field is randomised in each domain. The coherence of the magnetic field, therefore, depends on the domain length, which is taken to be $\Delta L_{\rm c} = 10\,{\rm kpc}$. The total radius of the cluster is 2 Mpc.

Galaxy clusters are home to a large population of charged particles that will give an effective mass to the photon. To accurately describe these, we consider the electron density given by:

$$n_{\rm el}(r) = n_{\rm el}^0(1 + r/r_{\rm core})^{-1}\,, \tag{2.5.3}$$

where $n_{el}^0 = 10^{-2}{\rm cm}^{-3}$. With this description of the cluster, we evolve the state matrix of eq. (2.5.1) by constructing an evolution operator (G) for each domain:

$$G = e^{iH\Delta L}\,, \quad H = \begin{bmatrix} \Delta^{\rm pl} + 2\Delta^{\rm QED} & 0 & (\Delta_x^{\varphi\gamma})_1 & (\Delta_x^{\varphi\gamma})_2 & (\Delta_x^{\varphi\gamma})_3 \\ 0 & \Delta^{\rm pl} + \frac{7}{2}\Delta^{\rm QED} & (\Delta_y^{\varphi\gamma})_1 & (\Delta_y^{\varphi\gamma})_2 & (\Delta_y^{\varphi\gamma})_3 \\ (\Delta_x^{\varphi\gamma})_1 & (\Delta_y^{\varphi\gamma})_1 & \Delta_1^{\varphi} & 0 & 0 \\ (\Delta_x^{\varphi\gamma})_2 & (\Delta_y^{\varphi\gamma})_2 & 0 & \Delta_2^{\varphi} & 0 \\ (\Delta_x^{\varphi\gamma})_3 & (\Delta_y^{\varphi\gamma})_3 & 0 & 0 & \ddots \end{bmatrix}, \tag{2.5.4}$$

where H denotes the propagation Hamiltonian with the following parameters:

$$\Delta^{\rm pl} = \frac{-1.1\times10^{-10}{\rm GeV}}{E\times10^{-3}}\frac{n_e}{10^{-3}{\rm cm}^3} \tag{2.5.5}$$

$$\Delta^{\rm QED} = 4.1\times\frac{10^{-6}{\rm GeV}}{E\times10^{-3}}\frac{B_x^2 + B_y^2}{\mu{\rm G}^2} \tag{2.5.6}$$

$$\Delta_i^{\varphi} = \frac{-7.8\times10^{-3}{\rm GeV}}{E\times10^{-3}}\left(\frac{m_i}{10^{-8}{\rm GeV}}\right)^2 \tag{2.5.7}$$

$$(\Delta_x^{\varphi\gamma})_i = 7.6 \times 10^{-2} \frac{g_i^\gamma \text{GeV}}{5 \times 10^{-11}} \frac{B_y}{\mu\text{G}} \tag{2.5.8}$$

$$(\Delta_y^{a\gamma})_i = 7.6 \times 10^{-2} \frac{g_i^\gamma \text{GeV}}{5 \times 10^{-11}} \frac{B_x}{\mu\text{G}} \,. \tag{2.5.9}$$

Here, E is the photon/ALP energy (in GeV), n_e is the electron density and B_x and B_y are the x and y components of the magnetic field strength (in μG) [85]. The mass of the i^{th} ALP state, denoted m_i, is taken to be distributed logarithmically within $\left[10^{-8}, 10^{-5}\right]$ eV. The various components of H can be interpreted as follows: Δ_i^φ are the ALP mass terms; Δ^{pl} is a photon effective mass term, induced by thermal effects in the electron plasma; $(\Delta_x^{\varphi\gamma})_i$ and $(\Delta_y^{\varphi\gamma})_i$ are the ALP-photon couplings for each photon polarisation; and, Δ^{QED} implements vacuum polarisation effects. As we are working in the mass basis, each ALP is independent of all other ALP states so H is diagonal in the ALP sector. We construct a new evolution operator, G_i, for each domain using the central values of B and n_{el}. The state of the system after propagation through the cluster is then given by:

$$\rho_{\text{out}}^C = \tilde{G}^C \rho_\gamma \left(\tilde{G}^C\right)^\dagger \,, \tag{2.5.10}$$

$$\tilde{G}^C = \prod_i G_i^C \,, \tag{2.5.11}$$

where i runs over each domain in the cluster.

Intergalactic space

The intergalactic magnetic field is modelled with a similar domain-like structure to that of the AGN cluster, albeit with a significantly lower overall strength and a much larger domain length $\Delta L^{IG} = 50\,\text{Mpc}$. The intergalactic medium has a luminosity redshift (z) dependent field strength given by:

$$B^{IG}(z) = B_0^{IG}(1+z)^2 \,, \tag{2.5.12}$$

where $B_0^{IG} = 1\,\mathrm{nG}$. The intergalactic electron density is approximated with a constant value of $n_{\mathrm{el}}^{IG} = 10^{-7}\,\mathrm{cm}^{-3}$. To account for VHE photon scattering off the EBL we introduce a non-unitary decay matrix (D) for each domain:

$$D(\tau) = \begin{bmatrix} \exp(-\tau/2) & 0 & 0 \\ 0 & \exp(-\tau/2) & 0 \\ 0 & 0 & \ddots \end{bmatrix}, \tag{2.5.13}$$

where τ is the optical depth associated with propagation through a given domain. In general, τ depends on the photon energy and the redshift of the domain. We use the EBL opacity model given in Ref. [86]. After computing $D(\tau_i)$ for each domain, the ALP-photon state is propagated as follows:

$$\rho_{\mathrm{out}}^{IG} = \tilde{G}^{IG} \rho_{\mathrm{out}}^{C} \left(\tilde{G}^{IG}\right)^{\dagger}, \tag{2.5.14}$$

$$\tilde{G}^{IG} = \prod_i D(\tau_i) G_i^{IG}, \tag{2.5.15}$$

where G_i^{IG} is calculated as before using eq. (2.5.4) with i running over each domain in the IGM.

The Milky Way

Compared to the previous two field scenarios, the Milky Way field has a significantly more complex structure that comprises a halo field that surrounds a spatially compact disk field. Following Ref. [87], we describe the disk by a series of logarithmic spirals and the halo field by a superposition of piece-wise functions extending relatively far above and below the galactic plane, see fig. 2.6. It is this extensive halo field that is responsible for the majority of ALP-Photon conversion. A detailed description of the field model used is provided in [87]. We note that in our work, a small alteration has been made to correct the function describing boundaries between the consecutive log-spiral regions in the disk:

$$r_{\rm bound} = r_j e^{(\theta-\pi)\cot\theta}\,, \qquad \theta = \frac{\pi}{180}(90-\alpha_{\rm open})\,, \tag{2.5.16}$$

where $\alpha_{\rm open} = 11.5^o$ is the opening angle of the log-spiral and r_j are the radii at which each spiral boundary crosses the negative x-axis. Finally, we use a constant electron density of $0.1\,{\rm cm}^{-3}$ to construct the MW evolution operator. The final state of the system is then given by the product over domains i:

$$\rho_{\rm out}^{MW} = \tilde{G}^{MW}\rho_{\rm out}^{IG}\left(\tilde{G}^{MW}\right)^\dagger\,, \tag{2.5.17}$$

$$\tilde{G}^{MW} = \prod_i G_i^{MW}\,, \tag{2.5.18}$$

where again we discretise our computation of the evolution operator (G_i) - in this case, using an approximate-coherence length of 100 pc. We begin the Milky Way propagation stage when the ALP/photon reaches a radial distance of 20 kpc from the galactic centre. The probability that an emitted photon is detected as a photon on Earth is found by projecting the final state ($\rho_{\rm out}^{MW}$) onto the two possible photon states:

$$P_{\gamma\to\gamma} = {\rm Tr}\left(2\rho_\gamma \mathcal{T}_{\rm out}^{MW}\right) \tag{2.5.19}$$

$$= {\rm Tr}\left(2\rho_\gamma\left[\tilde{G}^{MW}\tilde{G}^{IG}\tilde{G}^{C}\right]\rho_\gamma\left[\tilde{G}^{MW}\tilde{G}^{IG}\tilde{G}^{C}\right]^\dagger\right)\,. \tag{2.5.20}$$

2.5.2 Code

To achieve optimal efficiency, the code[1] for the simulation was written in C++ and relies heavily on the `Eigen` matrix library [88] which allows for parallelised execution of matrix operations. The software was written with a focus on usability. For example, it allows different sections of the simulation to be activated and deactivated

[1]The code is available in the following repository: `https://gitlab.dur.scotgrid.ac.uk/James_Maxwell/blazarphotonaxion`.

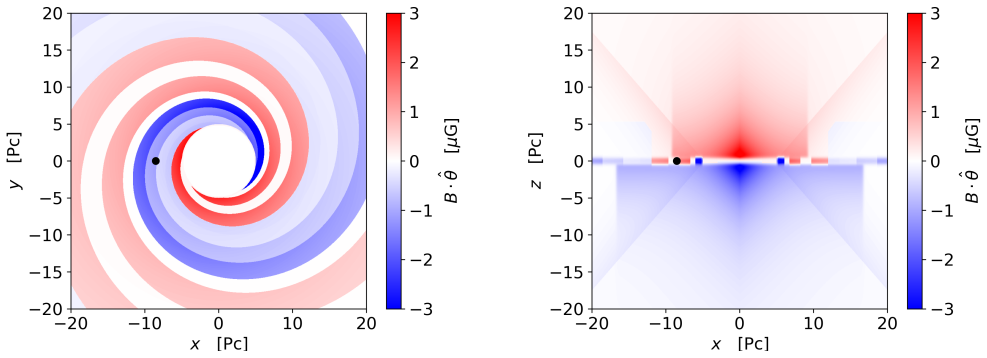


Figure 2.6: The magnitude of the Milky Way field in two perpendicular planes through the galactic centre. Left: cross-section at $z = 0$ looking upon the galactic disk; Right: cross-section perpendicular to disk at $y = 0$. The sign of the field indicates the orientation of the azimuthal component: positive - anticlockwise (looking down negative z); negative - clockwise. The location of the earth, $(-8.5, 0, 0)$kpc, is marked by the black circle.

without modification to the code. This promotes its use for similar but distinct problems that may not require the simulation of all propagation domains. Indeed, at the time of writing this thesis, the software is currently being used in this way. The code was verified by considering the single ALP scenario discussed in ref. [89].

2.5.3 Results

Due to observational limitations, the field orientations in each domain of the galaxy cluster and intergalactic medium are unknown and small changes thereto could have a large effect on the final value of $P_{\gamma\to\gamma}$. It is thus necessary to simulate with a large set of different field directions. $P_{\gamma\to\gamma}$ now represents a stochastic quantity that we use to marginalise over the unknown field states.

As with the rest of this work, we also encounter an ambiguity in the choice of mixing parameters ($\{g_i^\gamma\}$), which we again account for using the ALP anarchy model described above. Each simulation run has a unique set of coupling parameters and magnetic field orientations. The final results described below were obtained using a sample of $N_{\text{samp}} = 1000$ simulation runs and a set of 6 Blazar sources. These sources

j	Source	Experiment	Fit Function
1	Mkn 501	HEGRA	LP
2	1ES0414+009	H.E.S.S	PL
3	1ES0229-200	H.E.S.S.	PL
4	Mkn421	H.E.S.S.	LP
5	1ES1101-232	H.E.S.S	PL
6	1ES0347-121	H.E.S.S	PL

Table 2.1: List of VHE gamma-ray sources used in our analysis and their corresponding fit functions; Log Parabola (LP) or Power Law (PL), see appendix C.

(listed in table 2.1) were selected based on the availability of their data and on their inclusion in other works on this subject [84]. We plan to perform a more detailed follow-up study using a larger sample of sources and more recent data.

For a given choice of blazar source and grid point (N, g^γ) in model parameter space, the result of our simulation is a set of $N_{\rm samp}$ survival probability spectra $P_{\gamma\to\gamma}(E)$, where E is the photon energy. To marginalise over these, we use the students p-statistic (p_t), following [84]; the calculation of p_t is described in appendix C. For each grid point, we obtain a distribution of $N_{\rm samp}$ p-statistics. We collate these data by determining the value of p_t corresponding to the field configuration resulting in a better agreement between the model and corrected spectra than 95% of other configurations. We denote this as p_{95}. In this way, we arrive at a two-dimensional parameter scan for each source that preserves more of the underlying p_t distribution than could be achieved, for example, by simply averaging the p_t over $N_{\rm samp}$. This method facilitates comparison with Ref. [84]. Note that this method results in relatively conservative bounds on ALP scenarios as we are choosing a B field configuration that agrees rather well with the data.

Finally, to construct an overall parameter scan we determine p_{95} on the union of the sample data for each source. The resulting grid scan is shown as a heat map in fig. 2.7. Lower values of $-\log_{10}(p_{95})$ correspond to the model better reproducing the observations. We see a significant improvement over the SM with the addition of a single ALP state, this corroborates the findings of Ref. [84]; the introduction of ALP

states can alleviate tensions in standard intergalactic opacity models. Interestingly, and novel to this work, we also observe that increasing N provides a worse fit to the VHE data. As the number of ALP states increases, so too does the number of hidden states. Working in the interaction basis, states in models with many ALPs will exist less frequently as the interaction state. We might expect, therefore, a suppression in the back conversion probability. This can be seen explicitly in fig. 2.5, where in the 20 ALP case, we do not observe any increase in $P_{\gamma\to\gamma}$ in the MW. This effect can be very significant, in this case, completely negating the opacity decrease caused by reduced EBL scattering; $P_{\gamma\to\gamma}$ is significantly lower in the 20 ALP case than it is for no ALPs (black line in the fig. 2.5). It should be noted that the results depicted in fig. 2.5, having been computed for a single source and energy, may not be representative of other sources at different locations. In contrast, fig. 2.7 reflects our statistical analysis using multiple blazars and energies, where the single ALP state shows the greatest agreement with observed spectra.

In addition to favouring fewer ALP states, we also see a lower degree of tension at greater ALP couplings. The lower regions of fig. 2.7 behave as no ALP models, and there is no significant reduction in EBL scattering.

2.6 Discussion

In this chapter, we have explored the phenomenology of ALP anarchy models comprising many axion-like particles whose masses and SM couplings are related by random matrices. String compactifications typically generate many ALPs; therefore, the phenomenology of many ALP systems is an important direction in studying physics beyond the SM. The ALP anarchy scenario provides a benchmark for this phenomenology.

We have shown that a key feature of many ALP phenomenology is oscillations between the ALP states that couple to the SM and hidden ALP states that do not. A given ALP model can be characterised by the total ALP photon and ALP electron couplings $g^{\gamma 2} = \sum g_i^{\gamma 2}$ and $g^{e2} = \sum g_i^{e2}$. However, oscillation into hidden

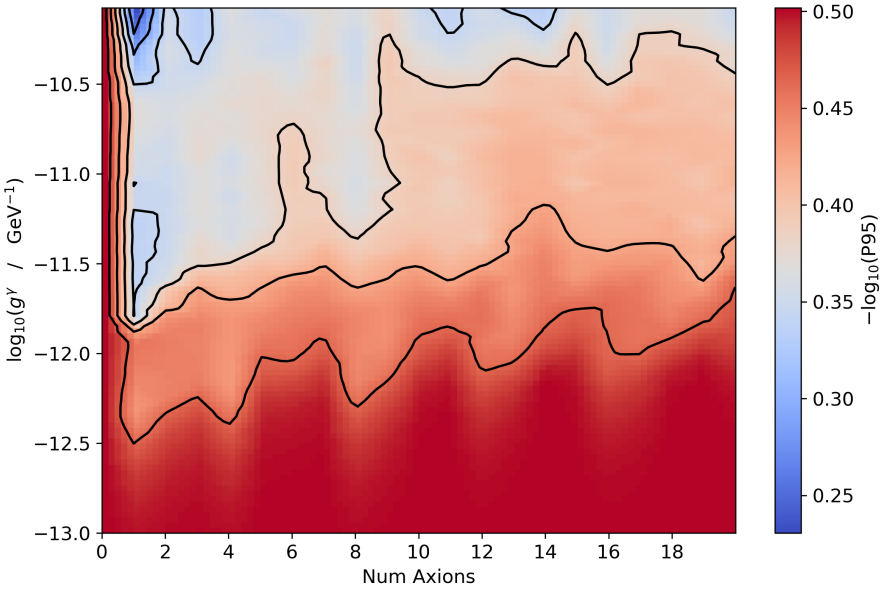


Figure 2.7: A heat map showing p_{95} computed using 6 sources and $N_{\text{samp}} = 1000$ samples per source. Lower values (higher p_{95}) indicate better agreement between the model and the data.

ALP states can significantly reduce the signal in some ALP searches, such as CAST and IAXO. Other ALP search strategies sensitive only to photon disappearance into ALP degrees of freedom, such as arguments from stellar cooling, will be largely insensitive to the number of ALP fields for a given total ALP-photon coupling. Still, other search strategies such as ADMX [90] and other Dark Matter haloscopes rely on a mass resonance and would, therefore, be sensitive to each ALP mass eigenstate individually rather than to the total effective ALP couplings. If the total ALP-photon coupling and dark matter density is shared over many ALP states, the expected signal in haloscope experiments for a given mass would be correspondingly reduced.

As discussed in section section 2.5, ALPs have been proposed as a solution to increase the observed transparency of the Universe to very high energy photons. However, we found that for many ALP systems, the photon survival probability instead decreases due to the presence of ALPs - the opposite effect to that observed for a single ALP. We therefore conclude that many ALP phenomenology leads to a number of new effects not captured by consideration of a single ALP field.

Chapter 3

General Relativity Introduction

Up until now, everything we have considered has taken place in flat Minkowski space-time. Minkowski space is appropriate in regimes of weak gravitation, such as we experience on the Earth. Much of the remaining content of this thesis, however, concerns the search for signatures of ALPs and other scalar particles around massive astrophysical objects. The extreme energies and gravitational and electromagnetic fields generated by these structures make them ideally suited to the search for weakly coupled particles. To fully describe processes taking place around these objects, we must take account of the curvature of space-time using General Relativity (GR). In the following section, we overview the core features of GR working up to the black-hole solutions to Einstein's field equations. Following this, we discuss the process by which these highly non-linear systems can be probed numerically through the use of high performance computers (HPCs).

Throughout this chapter and the next, we use geometric units such that the speed of light is $c = 1$ and Newton's gravitational constant is $G = 1$. When referring to space-time coordinates, we use the Greek alphabet $\mu, \nu, ... \in [0, 3]$ and when referring to purely spatial coordinates we use the Latin alphabet $i, j, ... \in [1, 3]$. Additionally, and contrary to the preceding chapters, we use the mostly positive metric signature $\eta_{\mu\nu} = \text{diag}(-1, +1, +1, +1)$.

As we have seen in the previous chapters, particles exert forces on one another

through the propagation of gauge fields. One force, however, is notably absent from the SM Lagrangian, this being gravitation. Attempts to reconcile gravity within the framework of the particle theory have as yet, been unyielding. Particle theories concerning gravity and its proposed graviton have proved to be non-renormalisable and unable to generate useful predictions. Gravity can by understood, not with a quantum field theory, but rather a classical field theory. This theory was developed throughout the 20^{th} century by appealing to two core tenets. These being *general covariance* and the *principle of equivalence*. The former is the requirement that the laws of physics remain invariant under a change of observer and the latter reflects the observation that objects accelerate under gravity at a mass independent rate.

3.1 Equivalence principle

An early form of the equivalence principle was deduced by Galileo, and later refined by Newton. This form, known as *weak equivalence* is the statement that gravitational mass and inertial mass are the same. That is to say that the acceleration due to gravity is mass independent, contrary to all other forces of nature.

This distinguishing observation of the gravitational force was later developed by Einstein, giving the *strong equivalence* principle. The strong equivalence principle describes the notion that an observer in a free falling box will experience weightlessness in the same way as an observer adrift in empty space. More technically, the laws of physics in a free-falling reference frame are locally equivalent to those in the absence of a gravitational field. The effect of gravitation is only apparent when considering an extended system. Returning to the free falling box example, consider the case that an observer drops two test masses (that we assume cannot gravitate towards one another) from different locations. If the box is falling in the Earth's, spherical, gravitational field, then the masses would begin to converge. On the other hand, if the box and its occupants are falling in a uniform field, then the masses will remain in the same position, see fig. 3.1.

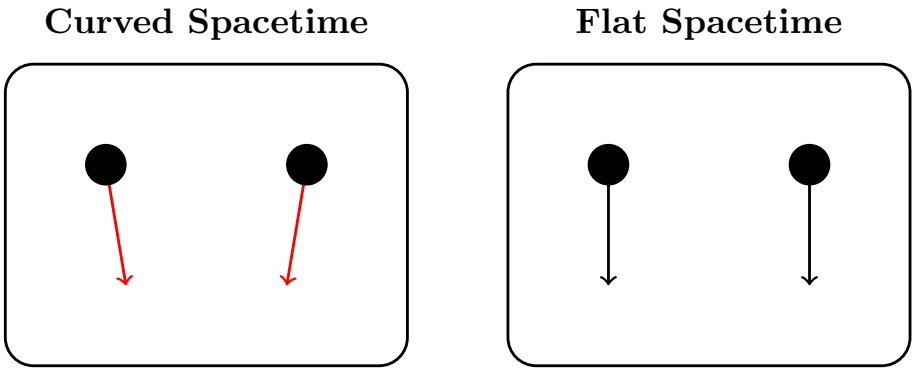


Figure 3.1: An illustration of how the curvature of space time can be observed in free fall, provided the system has a finite size. The balls falling in the curved spacetime (left) converge due to the spatial variation in the field. In the absence of any field, or variations in the field, the balls remain in the same location relative to one another (right).

A form of the strong equivalence principle also holds for a scenario complimentary to that of a free-falling box, in which the box is now at rest on the Earth's surface. An observer within this box will now experience a force from the base of the box. Equivalence implies that this force, and indeed all the laws of nature, will be locally equivalent to those within an appropriately accelerating box in free space. It was this result that led Einstein to infer that gravitation should be described by a geometrical theory. Acceleration describes changes in space and time and Einstein developed his theory by considering gravity as a warping of a space-time surface in the presence of massive objects. This warping then gives rise to gravitational acceleration. John Wheeler summarised this succinctly in Ref. [91, pg.235]:

'Spacetime tells matter how to move; matter tells spacetime how to curve.'

3.1.1 Differential Geometry

Having identified GR as a geometric theory, Einstein required an appropriate mathematical framework, through which to express this notion. For this, he turned to differential geometry, a formalism that extends the concepts of calculus to geometrical, smooth surfaces known as manifolds.

Returning to gravity, we begin with the question of what does it mean for space to curve? To answer this, it is necessary to introduce a concept of length to the manifold. The size of an infinitesimal line element is given by the, aptly named, metric:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \tag{3.1.1}$$

where $g_{\mu\nu}$ is known as the metric tensor. $g_{\mu\nu}$ is symmetric and contains information about the shape of the manifold. For flat Euclidean space $\mathbb{R}^3$, $g_{\mu\nu}$ is simply the identity matrix and the metric becomes a statement of Pythagoras' theorem:

$$ds^2 = dx^2 + dy^2 + dz^2. \tag{3.1.2}$$

Consider now the less trivial example of a 2-sphere S^2 in polar coordinates (θ, ϕ). In this case, the metric becomes:

$$ds^2 = R^2 \left(d\theta^2 + \sin^2 \theta \, d\phi^2 \right). \tag{3.1.3}$$

The metric in-fact contains all of the information regarding the shape and curvature of the manifold. It follows that to successfully realise Einstein's geometric theory of gravity, we are required to determine the metric tensor $g_{\mu\nu}(x_\mu)$ throughout the manifold.

Before discussing the field equations that govern the behavior of $g_{\mu\nu}$, it is instructive to explore another important property of the metric, namely, it provides a natural way to define an inner product between vectors. As a brief aside, when we refer to vectors in differential geometry, we are really referring to tangent vectors. Tangent vectors simply live in the space tangential to the manifold (M) at a given point (p), known as the tangent space $T_P M$. A tangent vector can be defined in terms of a smooth curve $\gamma(s)$ as:

$$V^\mu = \frac{d\gamma^\mu}{ds}\Big|_{s=0}. \tag{3.1.4}$$

In local coordinates, this vector can be written:

$$V = V^\mu \partial_\mu, \tag{3.1.5}$$

where $\partial_\mu = \frac{\partial}{\partial x^\mu}$. It is often useful to define a vector as a function of the position $V(x_\mu)$ on the manifold. This is known as a vector field and is an element of the disjoint union of all tangent spaces, known as the tangent bundle $V(x_\mu) \in TM$. In local coordinates, one may write:

$$V = V^\nu(x_\mu)\partial_\nu. \tag{3.1.6}$$

We see that in differential geometry, vectors and vector fields correspond to differential operators. A related concept is that of the dual vector space T_p^*M, an element of which (in local coordinates) can be written:

$$V^* = V_\mu dx^\mu, \tag{3.1.7}$$

where dx^μ form a basis of one-forms and the lower index of V_μ indicates that the vector belongs to T_p^*M. V^* is dual in the sense that, when it is operated on by a vector will yield a scalar:

$$V^\mu \partial_\mu V_\mu dx^\mu = V^\mu V_\mu. \tag{3.1.8}$$

Elements of the dual space are known as covectors and elements of the tangent space are known as contravectors. They are related to one another through the metric:

$$V_\mu = g_{\mu\nu} V^\mu. \tag{3.1.9}$$

In this way, the metric naturally provides a definition for an inner product:

$$(V, U) = V^\mu g_{\mu\nu} U^\nu. \tag{3.1.10}$$

3.2 Geodesics

As identified previously, the metric contains the curvature information of the manifold and thus will dictate the motion of bodies under gravitation. We now describe how this motion can be determined. In the absence of gravity, spacetime will be flat and a test mass will not experience any acceleration, hence:

$$\frac{d^2 x^\mu}{d\lambda^2} = 0, \tag{3.2.1}$$

where λ is an affine parameter that labels points in the trajectory $x^\mu(\lambda)$. If the test mass encounters a non-trivial gravitational field, then eq. (3.2.1) must be augmented with an additional *curvature* term, giving:

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma} \frac{dx^\rho}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0, \tag{3.2.2}$$

where $\Gamma^\mu_{\rho\sigma}$ are known as *Christoffel symbols* and encode spatial variations in the metric. This equation is known as the *geodesic* equation, and in the absence of any back reaction, can be solved to determine the path of any free-falling object in a gravitational field. An example of a test mass free-falling into a 2d black hole (BH) geometry is depicted in fig. 3.2 using a Flamm paraboloid[1].

The Christoffel symbols that dictate the geodesic path are given by:

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2} g^{\lambda\rho} \left(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu} \right), \tag{3.2.3}$$

where $g^{\mu\nu}$ is the inverse metric. The Christoffel symbols describe how the tangent spaces change from one point to another. It should be noted that the presence of

[1] Flamm's embedding encodes the non-Euclidean 2d geometry of the Schwarzchild metric as a Euclidean surface in 3d by equating the metrics $\left(1 - \frac{2M}{r}\right)^{-1} dr^2 = dr^2 + dz^2$.

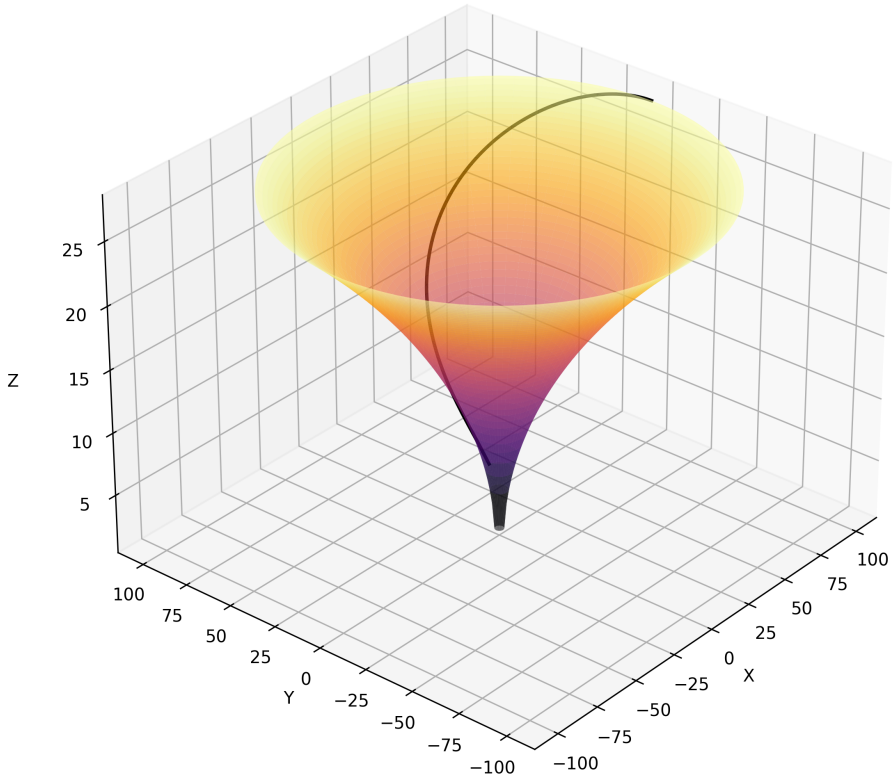


Figure 3.2: A geodesic (black line) for a test mass free-falling on a 2d Schwarzchild geometry. To visualise the curvature, the manifold is depicted in 3d through Flamm's embedding. The geometry corresponds to a BH of mass $M = 1$ and the test mass was initialised with a velocity of $0.99c$ in the angular direction.

partial derivatives in the eq. (3.2.3) implies that the Christoffel symbols are not tensors. Under coordinate transformation, these symbols will gain a correction term containing additional derivative operations.

In GR, geodesics can be classified according to their 4-velocity k^μ. Geodesics corresponding to the motion of massive objects are said to be *time-like*, and satisfy the normalisation condition $g_{\mu\nu}k^\mu k^\nu < 0$. For massless particles, trajectory geodesics will satisfy $g_{\mu\nu}k^\mu k^\nu = 0$ and are said to be *null* or *light-like*. Finally, geodesics for which $g_{\mu\nu}k^\mu k^\nu > 0$ are said to be *space-like*.

3.2.1 Covariant derivative

The Christoffel symbols provide a way to relate vectors in one tangent space to those in an adjacent space. This allows one to define a tensor derivative that encapsulates the structure of space-time, known as the *covariant derivative*. Acting on a scalar function ($f(x_\mu)$), the covariant derivative simply reduces to partial differentiation:

$$\nabla_\mu f = \partial_\mu f. \tag{3.2.4}$$

This reflects the fact that scalars are unaffected by the structure of the manifold. Acting on a vector field ($V = V^\mu e_\mu$), however, the covariant derivative takes the form:

$$\nabla_\mu V^\nu = e_\mu V^\nu + \Gamma^\nu_{\mu\lambda} V^\lambda, \tag{3.2.5}$$

where e_μ are the basis vectors. The relationship between the covariant derivative and the partial derivative in this case is most apparent in the coordinate basis ($e_\mu = \partial_\mu$), for which:

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\lambda} V^\lambda. \tag{3.2.6}$$

Hence, in flat space, $\nabla_\mu V^\nu = \partial_\mu V^\nu$.

The covariant derivative is related to the concept of parallel transport. Parallel transport describes the act of continuously translating a vector from one point on the manifold to another such that it remains invariant up to changes due to the curvature. A vector V^μ is said to be parallel transported along a curve with tangent vector u^μ if it satisfies the following relation:

$$u^\nu \nabla_\nu V^\mu = 0. \tag{3.2.7}$$

3.2.2 Curvature

The Christoffel symbols, which embody the variation of the metric over the manifold, have already proved invaluable. Unfortunately, however, they do not transform covariantly under a change of basis. This is at odds with our desire to construct a basis independent theory of gravity. Through specific combinations of these symbols it is possible to construct new objects that do transform appropriately. These objects can then be used to build a differential equation for the structure of the metric.

The first such object we will consider is the *Riemann tensor*, given by:

$$R^\rho_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\mu\sigma}. \tag{3.2.8}$$

Unlike the Christoffel symbols, the Riemann Tensor can be transformed simply by contracting it with an appropriate Jacobian tensor. To develop some intuition about the Riemann tensor, consider the parallel transport of a tangent vector on a manifold. If the vector V^ρ is transported along the sides of a geodesic triangle (see fig. 3.3), then in general, the resulting vector will differ from the original. For an infinitesimal triangle (side length dx^μ), and to first order, the difference between the two vectors will be given by:

$$\delta V^\rho \approx R^\rho{}_{\sigma\mu\nu} V^\sigma dx^\mu dx^\nu. \tag{3.2.9}$$

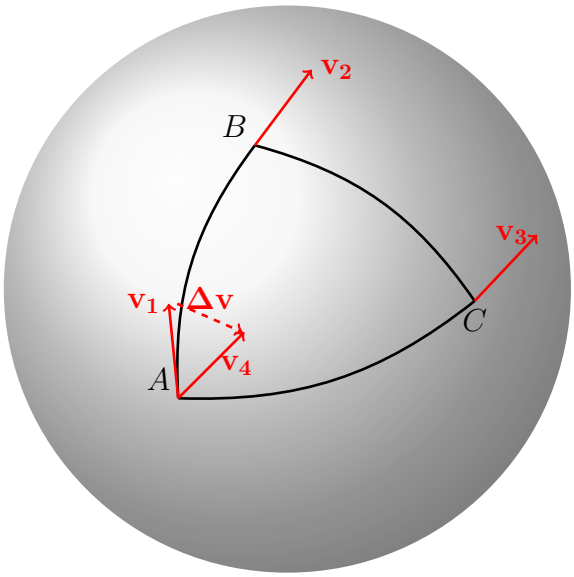


Figure 3.3: Depiction of Riemann curvature. The change in the vector V_1 after parallel transporting it along the path of the geodesic triangle (ABC) is given by ΔV. For an infinitesimal triangle, $\Delta V = \delta V$.

For flat space the Riemann tensor is zero, reflecting the fact that a vector transported round a triangle will not change.

We can construct another covariant quantity by further contracting the indices of the Riemann Tensor. This object, known as the *Ricci tensor*, is defined:

$$R_{\mu\nu} = R^{\lambda}{}_{\mu\lambda\nu}. \tag{3.2.10}$$

The Ricci tensor describes the congruence of geodesics, i.e. whether adjacent geodesic curves are converging or diverging. To visualise this, consider a small circle (known as a geodesic ball) on a two-sphere S^2. If the circle is allowed to free-fall along its perimeter geodesics, then its volume will change, see fig. 3.4. This change can be approximately expressed through the following differential equation:

$$\frac{d^2V}{d\tau^2} \approx -R_{\mu\nu}u^\mu u^\nu V, \tag{3.2.11}$$

where V is the original volume, u^μ is the geodesic tangent vector at the centre of the circle and τ is the geodesic affine parameter. Note that on a general mani-

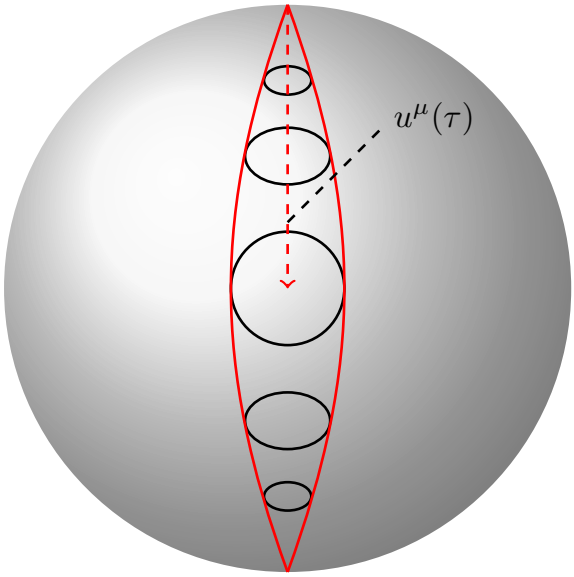


Figure 3.4: The change in area of a geodesic ball flowing according to its perimeter geodesics on S^2. The solid red lines indicate the tangential perimeter geodesics. The central geodesic tangent u^μ follows the dashed line. The change in the area of the ball is described by the Ricci tensor.

fold, a geodesic ball will not maintain its shape in free-fall. To fully describe the volume change, one must account for vortices and shear in the geodesic congruence. Eq. (3.2.11) is a special case of the *Raychaudhuri equation* which takes these effects into account; more information on this can be found in Ref. [92].

Finally, we introduce the *Ricci scalar*, defined simply as the trace over the Ricci Tensor:

$$R = g^{\mu\nu} R_{\mu\nu}. \tag{3.2.12}$$

The Ricci scalar condenses all the curvature information into a single number at each point on the manifold. If $R > 0$, the curvature is said to be positive and geodesics will tend to converge along their paths. This is the case with a sphere. For negative curvature ($R < 0$) geodesics diverge. Hyperbolic planes such as saddles have negative curvature.

3.2.3 Einstein's Field Equations

Having identified the metric as the core structure encapsulating the effects of gravity, we now require a way for matter to influence $g_{\mu\nu}$. This is achieved in the form of Einstein's Field Equations (EFEs):

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{3.2.13}$$

where $T_{\mu\nu}$ is known as the *stress energy tensor* and encodes the distribution of energy and mass in the universe and $G_{\mu\nu}$ is the *Einstein tensor* that contains the curvature information:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}. \tag{3.2.14}$$

In this subsection, we will follow in Einstein's footsteps and derive his field equations relating matter to curvature. The starting point in this endeavour is to write down an appropriate action. As the action corresponds to an integral over space-time, we require an appropriate definition of the infinitesimal volume. Our manifold is endowed with a metric, so we can use the natural volume form given by $\sqrt{-g}$. We are now free to choose terms to add to the integrand. When deriving the SM Lagrangian, we simply wrote down every term allowed under the symmetry group that corresponded to an observed particle. We will use a similar approach here. To ensure covariance, we may only write down scalar quantities. The simplest scalar that encapsulates the curvature is R; we are thus motivated to write down the action:

$$S = \frac{1}{16\pi G} \int R\sqrt{-g}\, d^4x. \tag{3.2.15}$$

Applying the principle of least action to eq. (3.2.15), one obtains the vacuum field equations:

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 0. \tag{3.2.16}$$

To include the effects of matter, we simply add the matter Lagrangian into the action:

$$S = \frac{1}{16\pi G} \int (R + \mathcal{L}_{\text{matter}}) \sqrt{-g}\, d^4x. \tag{3.2.17}$$

As an example, for a scalar field φ with potential $V(\varphi)$, $\mathcal{L}_{\text{matter}}$ might take the form:

$$\mathcal{L}_{\text{matter}} = \frac{1}{2} g^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi - V(\varphi). \tag{3.2.18}$$

To obtain the non-trivial EFEs, we can vary eq. (3.2.17) with respect to $g^{\mu\nu}$. We begin by defining the stress energy tensor to be the variation in the matter action S_{matter}:

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta S_{\text{matter}}}{\delta g^{\mu\nu}}, \tag{3.2.19}$$

where:

$$S_{\text{matter}} = \frac{1}{16\pi G} \int \mathcal{L}_{\text{matter}} \sqrt{-g}\, d^4x. \tag{3.2.20}$$

Finally, by combining equations eq. (3.2.19) and eq. (3.2.16) we derive equation eq. (3.2.13), which, in expanded form and taking $c = 1$, reads:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{3.2.21}$$

The left-hand side of these equations describes the curvature of space-time and the right-hand side describes the mass and energy content of the universe. As we saw previously, the Ricci tensor and scalar are defined in terms of the Riemann tensor, which itself is comprised of second order partial derivatives of the metric. EFEs thus constitute a set of 16 second order partial differential equations. For a generic system, solving these equations analytically is intractable. In recent years, the development of HPCs has allowed these complicated problems to be studied numerically. A large

portion of the work in this thesis concerns the use of these methods to study scalar fields in the presence of BHs. We give an introduction to these numerical relativity (NR) techniques in section 3.3. In the following subsection, we will study some of the analytical solutions to EFEs.

3.2.4 Black Holes

BHs constitute some of the simplest solutions to EFEs. Of these, the Schwarzchild solution, describing a non-rotating, uncharged BH is the most approachable. As a starting point, we assume the universe to be devoid of mass, hence $T^{\mu\nu} = 0$ and EFEs reduce to the vacuum equations eq. (3.2.16). Working in spherical polar coordinates (t, r, θ, ϕ), it can be shown that the following metric solves these equations:

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \frac{dr^2}{1 - \frac{2GM}{r}} + r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2\right). \tag{3.2.22}$$

This is the Schwarzchild metric for a BH of mass M. The coordinate basis used here to express the metric is associated asymptotic inertial observers (stationary observers at an infinite distance BH). From this perspective, the metric has a singularity at $r = 2GM$. This value, known as the *Schwarzchild radius* subtends a spherical *event horizon* from within which no matter or light can escape. Within the event horizon, all time-like and light-like geodesics will point inwards, see fig. 3.5. It can also be seen that as the Schwarzchild radius is approached, the temporal metric component will tends to zero. From the perspective of inertial asymptotic observers, therefore, no time will appear to pass at the event horizon. If such an observer were to watch a ball fall into the BH, they would see it slow down to a stop as it approached this point. This is seemingly at odds with the classical notion of gravitation. In fact, the singularity giving rise to this phenomenon is known as a coordinate singularity and is only present in certain coordinate systems. If instead, one were to consider an inward moving frame, following the path of a geodesic, then this singularity would disappear. Coordinate systems for which time propagation does not asymptote to

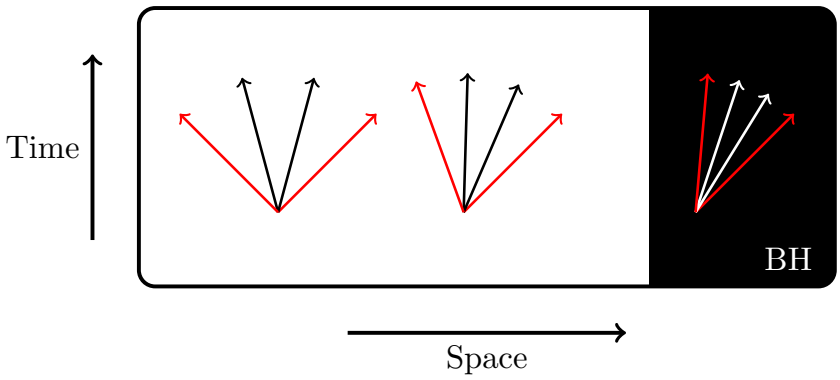


Figure 3.5: Possible geodesics in and around a BH. The light-like geodesics are shown in red, with the time like geodesics depicted in black/white. Outside the event horizon objects can move in an outwards spatial direction. Within the Schwarzchild radius, however, even light-like paths point inwards towards the centre.

zero are said to be horizon-penetrating. We will see later how these are necessary to appropriately resolve dynamics close to BHs.

In addition to the singularity at $r = 2GM$, eq. (3.2.22) also has a singularity at $r = 0$. This is, however, not a coordinate singularity and is present in all coordinate systems. To see this, we can consider the basis independent Ricci Scalar, which, for the Schwarzchild geometry, is given by:

$$R = \frac{4GM}{r^3} - \frac{8G^2M^2}{r^4}. \tag{3.2.23}$$

At the centre of the BH, $R \to \infty$, indicating that there is infinite curvature. All time-like and light-like geodesics will end at this point and GR is unable to predict what will happen here.

The Schwarzchild metric is not the only BH geometry that arises from the vacuum field equations. In general, physical BHs will possess a non-zero spin and charge and can be described by the *Kerr-Newman* metric [93]. Additionally, the matter environment around physical BHs will be non-trivial and for a thorough treatment of their physics, one must turn to numerical simulations.

3.3 Numerical Relativity

Physical systems involving the interplay between matter and BHs will generally not posses a high degree of symmetry. It is therefore difficult to solve the field equations that describe these systems analytically. This issue is compounded by the fact that the matter distribution and thus metric will evolve in time. Solving EFEs as well as the matter evolution equations across both space and time, whilst including back reactions effects between these is a task that can only be undertaken numerically. In this subsection, we discuss how this can be achieved by evolving an initial spatial state temporally using methods of finite difference.

EFEs treat space and time on an equal footing. This is problematic when it comes to predicting the temporal evolution of a system. Expressed in the form eq. (3.2.21), EFEs constitute a set of 10 independent, non-linear Partial Differential Equations (PDEs) relating changes in space, time, mass and energy to one another. To evolve a system numerically, we wish to rephrase these equations in the form of an initial value (or Cauchy) problem:

$$\partial_t \mathbf{u} = \mathbf{F}\left(\mathbf{u}, \partial_i \mathbf{u}, \partial_i \partial_j \mathbf{u}\right). \tag{3.3.1}$$

Where $\boldsymbol{u}$ are evolutions variables and $\boldsymbol{F}$ is a set of functions of $\boldsymbol{u}$ and their spatial derivatives. Expressing the EFEs in this way, the time dependence of the metric can be determined purely from its spatial structure and energy distribution at a given time. Numerical integration techniques can then be used to solve for the metric at later times.

To morph the field equations into the form of eq. (3.3.1), a number of key challenges must be addressed. The most fundamental of these is in the bifurcation of space and time. Known as $3+1$ decomposition, space-time is foliated into a series of space-like submanifold laminae, each labelled by a different time-like coordinate t. There are many different ways to perform this decomposition, but perhaps the most simple is the *Arnowitt–Deser–Misner* (ADM) formulation. We will see later some

more sophisticated approaches. In ADM decomposition, one defines the line element associated with the laminae:

$$d\sigma^2 = \gamma_{ij} dx^i dx^j \tag{3.3.2}$$

and expresses the full 4-metric as:

$$ds^2 = -\alpha^2 dt^2 + \gamma_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt), \tag{3.3.3}$$

where λ_{ij} is referred to as the *induced metric*, β^i, the *shift*, and α, the *lapse*. β^i encodes the change in spatial coordinates between consecutive laminae and α is an expression of change in proper time ($d\tau = \alpha dt$) between laminae separated by an infinitesimal dt. Explicitly, the metric tensor and its inverse are given by:

$$g_{\mu\nu} = \begin{pmatrix} -\alpha^2 + \beta^k \beta_k & \beta_i \\ \beta_j & \gamma_{ij} \end{pmatrix} \quad , \quad g^{\mu\nu} = \begin{pmatrix} -\frac{1}{\alpha^2} & \frac{\beta^i}{\alpha^2} \\ \frac{\beta^j}{\alpha^2} & \gamma^{ij} - \frac{\beta^i \beta^j}{\alpha^2} \end{pmatrix}. \tag{3.3.4}$$

The covariant and contravarient spatial vectors are related to one another by the spatial metric. For example the covariant shift is $\beta_i = \gamma_{ij}\beta^j$. The contravariant spatial metric γ^{ij} is simply the matrix inverse of the induced metric tensor (γ_{ij}).

EFEs are most easily expressed in terms of curvature quantities (R and $R_{\mu\nu}$). In the ADM formalism, we can try to do the same thing, although we'll need some new definition of curvature that can be written in $3+1$ form. Given the time-like normal vector (n^μ) to the laminae,

$$n^\mu = \frac{1}{\alpha}\left(\partial_t^\mu - \beta^i \partial_i^\mu\right), \tag{3.3.5}$$

the extrinsic curvature is defined:

$$K_{ij} = -\frac{1}{2}(\mathcal{L}_{\vec{n}} \gamma_{ij}) \tag{3.3.6}$$

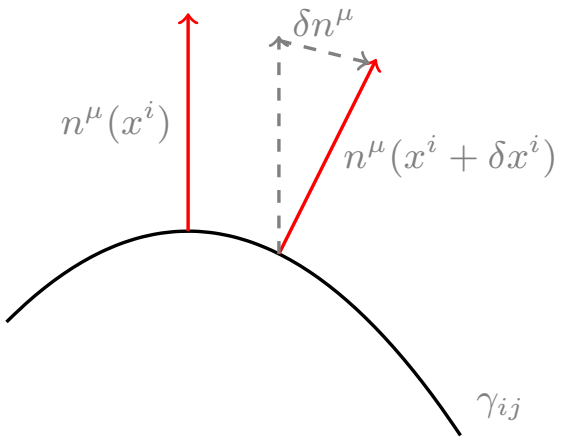


Figure 3.6: An embedding of a 1d hyper-surface within a 2d base space. The surface has an induced metric $\gamma_{ij},\ i,j \in \{0\}$ and extrinsic normal vector field $n^\mu(x^i),\ \mu \in \{0,1\},\ i \in \{0\}$. The extrinsic curvature is related to the change in the surface normal δn^μ. This curvature is extrinsic in the sense that it describes how a higher dimensional object varies over the surface. R simply describes the change of the metric and so is in some sense intrinsic to the submanifold, whereas K encodes the embedding and, in this case, effectively relates temporal components to spatial ones.

$$= g_i^\mu g_j^\nu \nabla_\mu n_\nu, \tag{3.3.7}$$

where $\mathcal{L}_{\vec{n}}$ is the Lie derivative in the $\vec{n}$ direction. The extrinsic curvature describes the rate of change of the normal n^μ as it moves across the space-like sub-manifold, see fig. 3.6.

We can now break down EFE's into a set of equations of the form eq. (3.3.1) and a number of elliptical constraint equations. The time evolution of the induced metric is given by:

$$\partial_t \gamma_{ij} = -2\alpha K_{ij} + \nabla_i \beta_j + \nabla_j \beta_i. \tag{3.3.8}$$

and the evolution of the extrinsic curvature is:

$$\partial_t K_{ij} = \beta^k \partial_k K_{ij} + K_{ki} \partial_j \beta^k + K_{kj} \partial_i \beta^k$$

$$
\begin{aligned}
&- \nabla_i \nabla_j \alpha \\
&+ \alpha \left({}^{(3)}R_{ij} + K K_{ij} - 2K_{ik}K^k_{\ j} \right) \\
&+ 4\pi\alpha(\gamma_{ij}(S-\rho) - 2S_{ij}).
\end{aligned} \tag{3.3.9}
$$

Here ${}^{(3)}R_{ij}$ refers to the Ricci tensor calculated on the induced metric, not the spatial components of the 4-tensor. Similarly, ∇_i is the covariant derivative on the sub-manifold. K is the trace of the extrinsic curvature and the quantities, S, ρ and S_{ij} are projections of the stress energy tensor ($T^{\mu\nu}$) in the direction of $\vec{n}$:

$$\rho = n^\alpha n^\beta T_{\alpha\beta}, \tag{3.3.10}$$

$$S_i = -n^\alpha T_{\alpha i}, \tag{3.3.11}$$

$$S_{ij} = T_{ij}, \tag{3.3.12}$$

$$S = \gamma^{ij} S_{ij}. \tag{3.3.13}$$

They are the quantities that would be measured by normal observers to the spatial slices [94]. For example ρ is the energy density and S_i is the momentum density.

From EFEs, we can also write down a Hamiltonian constraint, given by:

$$H = R + K^2 - K_{ij}K^{ij} - 16\pi\rho \tag{3.3.14}$$

and a momentum constraint, given by:

$$M_i = \nabla^j(\gamma_{ij}K - K_{ij}) - 8\pi S_i \tag{3.3.15}$$

We now have our 10 EFEs in ADM form; 4 constraint equations and 6 evolution equations (the symmetric components of k_{ij} and γ_{ij}). The constraint equations don't give any information on the evolution of the system, they are instead a restriction on the choice of metric and mass distribution. For instance, in their absence, we would be free to write down a scenario in which a large mass was situated in a completely

flat spacetime. This, of course, would be non-physical. The constraint equations are particularly important when specifying the initial data for a NR simulation. We will see later that it is a non-trivial task to determine a suitable starting metric that is both consistent with the stress-energy and appropriate for the types of problems one might wish to solve. For now though, we will discuss the implementation of time evolution in NR codes.

3.3.1 Time Integration

The work presented in the following chapter builds on the code base of `GRChombo` [95]. The focus of this subsection will be on the time integration schemes used by this software.

Each time evolution step can be decomposed into two parts; an update to the background, followed by an update to the matter distribution. The update to the background involves the integration of eq. (3.3.8) and eq. (3.3.9). The right hand sides of these equations can be found using the metric, the stress-energy, and numerical derivatives thereof. The time integration can then be completed using the 4^{th}-order Runge-Kutta algorithm (RK4). This algorithm uses 4 numerical evaluations of the derivatives to ensure good agreement with the analytic solution.

With the background updated, the evolution of the matter terms can now be computed. To achieve this, it is necessary it introduce additional equations of motion. EFEs describe the evolution of the background geometry but not that of the matter. The matter evolution can be determined using a classical Hamiltonian approach, albeit with a curved background. Returning to the example from section 3.2.3 of a scalar field, minimally coupled to gravity:

$$S = \int d^4x\sqrt{-g}\left(-\frac{1}{2}g^{\mu\nu}\nabla_\mu\varphi\nabla_\nu\varphi - V(\varphi)\right). \tag{3.3.16}$$

We can find the Hamiltonian by contracting the stress-energy tensor $T_{\mu\nu}$ with the normal n^μ to the space-like hyper-surfaces;

$$H_\varphi = n^\mu n^\nu T_{\mu\nu} \tag{3.3.17}$$

For a scalar field, the stress-energy in covariant form can be written [96]:

$$T^{\mu\nu} = \nabla^\mu\varphi\nabla^\nu\varphi - g^{\mu\nu}\left(\frac{1}{2}g^{\alpha\beta}\nabla_\alpha\varphi\nabla_\beta\varphi + 2V(\varphi)\right). \tag{3.3.18}$$

Contracting with n^μ and rewriting in terms of ADM variables, one finds:

$$H_\varphi = \alpha\sqrt{\gamma}\left(\frac{1}{2}\frac{\pi_\varphi^2}{\gamma} + \frac{1}{2}\gamma^{ij}\partial_i\varphi\partial_j\varphi + V(\varphi)\right) + \beta^i\pi_\varphi\partial_i\varphi, \tag{3.3.19}$$

where π_φ is the conjugate momentum defined by:

$$\pi_\varphi = \frac{\delta\mathcal{L}}{\delta(\partial_t\varphi)} = \frac{1}{\alpha}\left(\partial_t\varphi - \beta^i\partial_i\varphi\right). \tag{3.3.20}$$

Hamilton equations of motion are given by the following variational derivatives:

$$\partial_t\pi_\varphi = -\frac{\delta H_\varphi}{\delta\varphi} \tag{3.3.21}$$

$$= \beta^i\partial_i\pi_M + \gamma^{ij}\left(\alpha\partial_j\partial_i\varphi + \partial_j\varphi\,\partial_i\alpha\right) + \alpha K\pi_M - \gamma^{ij}\Gamma^k_{ij}\partial_k\varphi + \frac{dV}{d\varphi}. \tag{3.3.22}$$

and

$$\partial_t\varphi = \frac{\delta H_\varphi}{\delta\pi_\varphi} \tag{3.3.23}$$

$$= \alpha\pi_\varphi + \beta^i\partial_i\varphi. \tag{3.3.24}$$

These equations are in the form of Cauchy problems and in `GRChombo` are solved, once again, using the RK4 algorithm. In summary, there are 6 dynamical equations and 4 constraint equations for the metric and, in the case of a scalar field, 1 additional equation of motion (EOM) for the matter. In general, one would have a matter EOM for each degree of freedom as well as additional matter constraint equations.

Under ideal circumstances, where numerical integration was completely accurate, the ADM constraint equations would always be satisfied. In reality, however, numerical errors can compound leading to instabilities that can result in significant constraint violation. As with all numerical simulations, the easiest way to reduce the error is to increase the resolution. In some cases, however, a better choice of numerical scheme is required. In NR, this corresponds to selecting a different $3+1$ decomposition.

3.3.2 Adaptive Mesh Refinement

There are two options when it comes to increasing the resolution. Decrease the time step or increase the number of spatial points. `GRChombo` uses a mesh-based system, where space is divided into a grid of cubes. The spatial resolution could be improved by simply reducing the global grid spacing. This approach would come at a large cost, however, both in terms of execution time (as more boxes must be iterated through) and in terms of the memory required to store the data. In regions of low curvature, the additional information afforded by the higher resolution mesh would be redundant. A common strategy to maximize both efficiency and accuracy is with Adaptive Mesh Refinement (AMR) where the mesh is iteratively sub-divided in regions of interest (high curvature) and left sparse in flat regions.

AMR can be performed on the fly during the simulation. This can be achieved through a number of different algorithms. `GRChombo` inherits its AMR scheme from the fluid dynamics code `Chombo` [97], which itself is based on the *Berger-Rigoutsos* algorithm [98]. This algorithm works by first estimating the error associated with each grid element, then determining a set of cuboidal boxes to encompass the regions of large error. The cells in these regions are then subdivided and the process is repeated until the error tolerance is met or the user specified maximum refinement level is reached. See fig. 3.7 for an illustration of AMR.

Once the mesh has been subdivided into refinement levels, the simulation proceeds, beginning with the update step on the coarsest level. The grid cells belonging

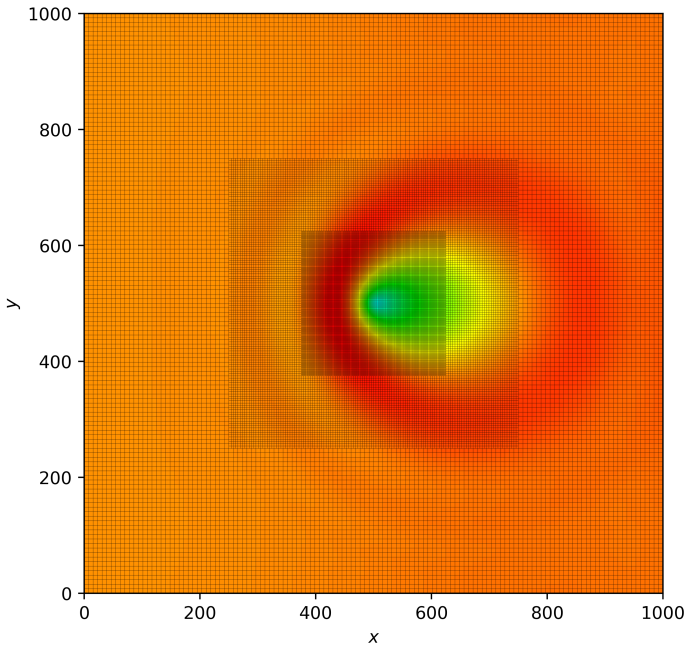


Figure 3.7: Example of AMR in an x-y plane surrounding a BH. Three levels of refinement are depicted centred around a boosted BH (the colour is associated with the real part of a scalar field that surrounds the BH). Each successive refinement level has twice the resolution of the preceding level in each dimension.

to higher refinement levels are then updated through the repeated application of progressively smaller time evolution steps. To maintain consistent evolution at the boundaries between refinement levels, the data coarse boundary cells are interpolated to populate so called *ghost cells* that are used during the evolution of the finer levels. Once temporal synchronized is achieved between the different refinement levels, any coarse cells that are overlapped by fine cells have their data overwritten. This processes is repeated iteratively through the successively finer levels until all cells are updated.

AMR allows for efficient partitioning of both spatial and temporal resolution. It allows large simulations to be performed without sacrificing accuracy around regions of interest. This is particularly useful for studying BH phenomena, where, to avoid the use of non-physical boundaries, the simulation volume should be large enough to approximate Minkowski space at its peripheries. Without AMR a simulation of this kind would be computationally unattainable.

3.3.3 Beyond ADM

As mentioned previously, sometimes, simply increasing the resolution isn't enough to ensure numerical convergence. To achieve convergence, the evolution equations must be *well-posed*. That is, they must have a unique solution that varies continuously with the initial conditions. Secondly, when simulate physics, one must further stipulate that disturbances in the solution propagate in a wavelike manner (i.e. at a finite speed). Well-posed PDEs of this nature are said to be *hyperbolic*. In the absence of the constraints, the ADM evolution equations are not truly hyperbolic as some parameters may increase in an unbounded way [96]. Unfortunately, numerical errors in the metric evolution can lead to constraint violation, thereby invalidating the use of the ADM scheme. Numerical GR codes use modified *strongly hyperbolic* versions of the ADM equations that are well posed even in the absence of the energy and momentum constraints. Common choices are *BSSN* and *CCZ4*. For more information on these topics, please refer to Ref. [99].

3.3.4 Initial Condition

We now turn our attention back to the constraint equations. Unlike the hyperbolic, evolution equations, these are elliptical in nature and do not produce wave-like solutions with finite speeds of information transfer. This, however, is not as pathological as it first appears. The constraint equations do not contain any time derivatives so their propagation speed is not a speed in the canonical sense and we are not in violation of special relativity. These elliptical constraint equations do present other challenges. Before one can run a numerical simulation, the state from which the system should begin evolving must be chosen. Thanks to the constraints, this is a non-trivial task. In some cases, such as when simulating low density gases around massive compact objects like BHs, it is possible to neglect the back reaction of the low density matter and simply use the metric of the massive object. In other cases, the results of previous numerical simulation can be used as a starting point. In more

difficult cases, specialised codes can be used to solve the elliptical equations. One such code, produced by the `GRChombo` collaboration is `GRTrensa` [100].

3.3.5 Fixed Background

In the remainder of this thesis, we will only be concerned with the evolution of matter on a fixed metric background. This simplifies the situation considerably from the general case discussed above. In particular, since we don't evolve the field equations, we are at no risk of violating the constraints and we are free to work in ADM coordinates. In addition, we will be working in a regime where the back reaction can be ignored. Consequently, we can simply fix the metric to an analytical solution to the vacuum Einstein equations and freely choose the initial conditions for the matter fields. This removes two of the key difficulties in numerical GR. We will still make use of AMR and matter evolution, but take analytical values for the metric quantities.

Chapter 4

Dynamical Friction

New light scalar and pseudo-scalar particles appear in many well-motivated extensions of the SM. As discussed in section 1.2 the QCD axion is an overwhelmingly popular solution to the strong CP problem and as outlined in section 1.3 ALPs occur quite generically in string compactification. Such (pseudo) scalars must be very weakly coupled to the SM to have evaded detection. Additionally, their SM couplings are rather model dependent. It is therefore well motivated to search for these new degrees of freedom via their gravitational interactions, for example by understanding their behaviour in the vicinity of BHs.

BHs are expected to readily encounter bosonic clouds, whether this be as a result of accretion or through more hypothetical processes such as *superradiance* around spinning BHs [101, 102]. Whilst the effects of superradiance may be detectable through instability collapse and Bosenovae [103], where the bosonic cloud becomes unstably dense, more generic bosonic clouds cannot be detected in this way. It may, however, be possible to infer the presence of such clouds by the force they exert on their source BH. As a massive object moves through a system of smaller gravitating bodies, it will accrete and sling-shot them, forming an over-density to its rear. This will exert a force, known as Dynamical Friction (DF), on the moving body as the surrounding masses are accelerated. DF was first described by Chandrasekhar in 1943 [104] for non-relativistic stars moving under Newtonian gravity. Later

works extended this principle to describe motion through fluids [105–107]. More recently, numerical simulations have been used to determine relativistic corrections to DF [4, 108]. These corrections have proved to be significant in certain scenarios such as during Extreme Mass Ratio Inspirals (EMRIs) [109].

4.1 Background

Chandrasekha's Newtonian description of DF pertains to a high mass star moving through a field of stochastically distributed lower mass stars. Chandrasekha described the DF force experienced by a star of mass M, moving with velocity v through a star field of density ρ as follows [104]:

$$\mathcal{F}_{\text{Chandra}} = 4\pi\rho\frac{M^2}{v^2}\ln\left(\frac{b_{\max}}{b_{\min}}\right). \tag{4.1.1}$$

Here, the logarithmic term, known as the Coulomb logarithm, accounts for the range of possible impact parameters $[b_{\min}, b_{\max}]$. The lower parameter ($b_{\min}$) corresponds to the capture impact of the focal star $b_{\min} \approx M/v^2$. The maximum impact parameter is related to the size of the gravitational wake. If the star were to enter a homogeneous gas cloud, initially its wake would be negligible, hence $b_{\max}$ would initially be zero. Over time, $b_{\max}$ would grow, in conjunction the DF force.

Chandrasekha's result for the DF can be thought of as the high mass (μ) limit of the following expression for the DF generated by a cloud of scalar particles [110]:

$$\mathcal{F}_{d,\text{nonrel}} = 4\pi\rho\left(\frac{M}{v}\right)^2\left\{\ln(2kr) - 1 - \text{Re}\left[\Psi(1+i\beta)\right]\right\}, \tag{4.1.2}$$

where Ψ is the digamma function, $k = \mu v$ and β is given by:

$$\beta = \frac{b_{\min}}{\lambda_{\text{dB}}}, \tag{4.1.3}$$

with $\lambda_{\rm dB} = 1/k$ being the de Broglie wavelength. As the impact parameter becomes large compared to $\lambda_{\rm dB}$, eq. (4.1.2) will tend towards eq. (4.1.1). This occurs for high M and μ and low velocities.

Although this field based description of DF correctly accounts for wave-like effects, both of these Newtonian approaches underestimate the force at relativistic speeds. In the relativistic regime, the effect of time dilation increases the gravitational deflection, thereby increasing the net momentum transfer. In this regime and in the limit of weak scattering (where the minimum impact parameter $b_{\rm min} \gg M$), eq. (4.1.2) must be augmented by the following multiplicative factor [111]:

$$\mathcal{F}_{d,\rm rel} = \gamma^2(1+v^2)^2\ \mathcal{F}_{d,\rm nonrel}. \tag{4.1.4}$$

Note that in the ultra-relativistic limit where $v = 1$, corresponding to photon propagation, the term $(1+v^2) \to 2$. This is associated with the photon gravitational deflection angle:

$$\theta_{\rm photon} = 2\theta_{\rm matter}. \tag{4.1.5}$$

Where $\theta_{\rm matter}$ is the slingshot angle for non-relativistic matter around a massive object in the weak scattering limit.

As well as the relativistic correction, an additional multiplicative factor must be included to account for the case where the boost velocity exceeds the speed of sound (c_s) of the matter field [112–114]. This is because oscillations in the field induced by the boosted mass cannot propagate outside the Mach cone described by the angle (θ_M) to the negative of the propagation vector:

$$\theta_M = \arcsin\frac{c_s}{v}. \tag{4.1.6}$$

This leads to an anisotropic accumulation of density and pressure around the cone. This can be accounted for in the DF with the introduction of the following pressure

factor that effectively replaces $\rho \to \rho + p$ in eq. (4.1.2):

$$\mathcal{F}_d = \frac{\rho + p}{\rho} \mathcal{F}_{d,\mathrm{rel}}. \tag{4.1.7}$$

Where p is the isotropic pressure in the absence of any perturbations induced by the mass M.

In 2021, numeral GR simulations were used by the authors of Ref. [4] to verify the relativistic and supersonic DF expression given by:

$$\mathcal{F}_{\mathrm{d}} = \mathcal{F}_{\mathrm{d,\ nonrel}} \times \gamma^2 (1 + v^2)^2 \times \frac{\rho + p}{\rho} \tag{4.1.8}$$

They considered a BH moving relative to complex scalar field of mass μ described by:

$$\varphi(x, t) = \varphi_0 e^{i(\omega t - kx)}, \tag{4.1.9}$$

where $k = \mu v$ and $\omega = \mu^2 + k^2$. From this, they determined a pressure correction of the form:

$$\frac{\rho + p}{\rho} = 1 + \kappa \frac{v^2}{1 + v^2} \tag{4.1.10}$$

where κ accounts for the distinction between more wave-like (small μ) and particle-like (large μ) fields. With $\kappa \to 1$ for wave-like fields and $\kappa \to 0$ for particle-like fields.

The numerical evolution was achieved using the AMR code `GRDZhadzha` [4]. This is is a branch of `GRChombo` [97] that is designed for fixed background simulations. As discussed in the previous chapter, the use of a fixed background greatly simplifies the process of extracting the forces acting on the BH and reduces the simulation run time. It does, however, mean that all back reactions on the metric are neglected. For this to be a valid approximation, the matter energy density must be sufficiently

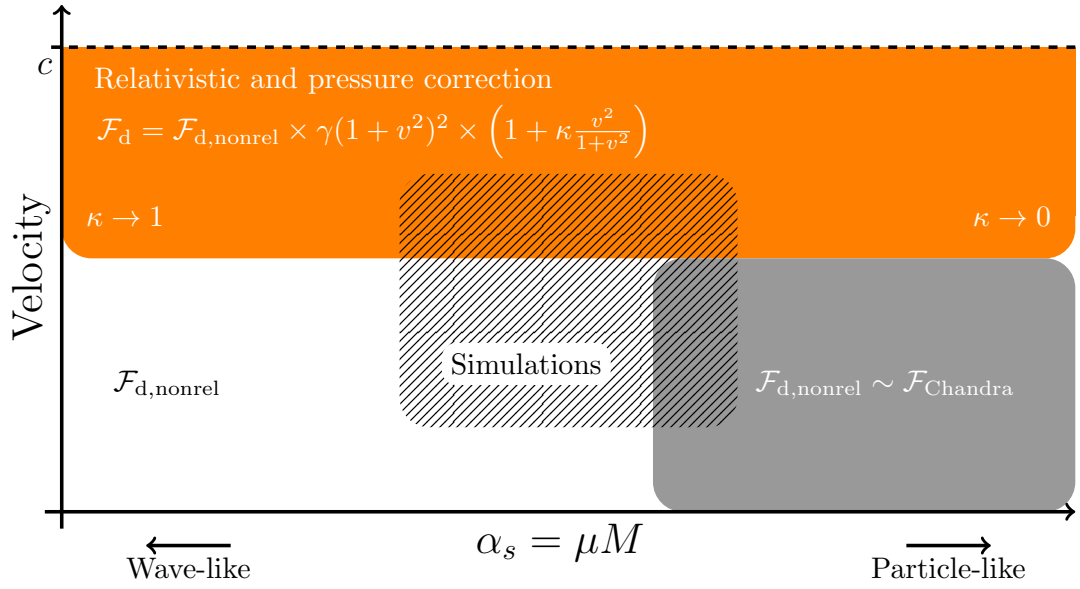


Figure 4.1: A cartoon depicting the different regions of validity for the various approximation to DF. Simulations allow for accurate investigations around the limits of validity of analytical approaches. This is a reproduction of fig. 4 in [4].

small $\rho \ll 1$. In the case of superradiance mentioned above, matter densities are not expected to exceed $\rho < 10^{-5}$. The agreement found between the fixed background numerical result and eq. (4.1.8) further indicates the accuracy of this approximation.

In addition to verifying the historical analytical results, these types of simulations open up new regions of the mass-velocity plane that were inaccessible to purely analytical techniques. The cartoon in fig. 4.1 illustrates this point.

The simulations carried out by the authors of Ref. [4] were performed using a fixed Schwarzchild metric written in *isotropic* coordinates. Isotropic coordinates are associated with stationary, asymptotic observers in the far field. As a result, when written in this form, the metric approaches that of Minkowski space at large distances from the BH. Consequently, the DF as measured by isotropic observers can be associated directly with the Newtonian understanding of the force. Unfortunately, however, the isotropic metric possesses a coordinate singularity at the event horizon. This leads to reduced temporal evolution close to the BH and decreases the accuracy

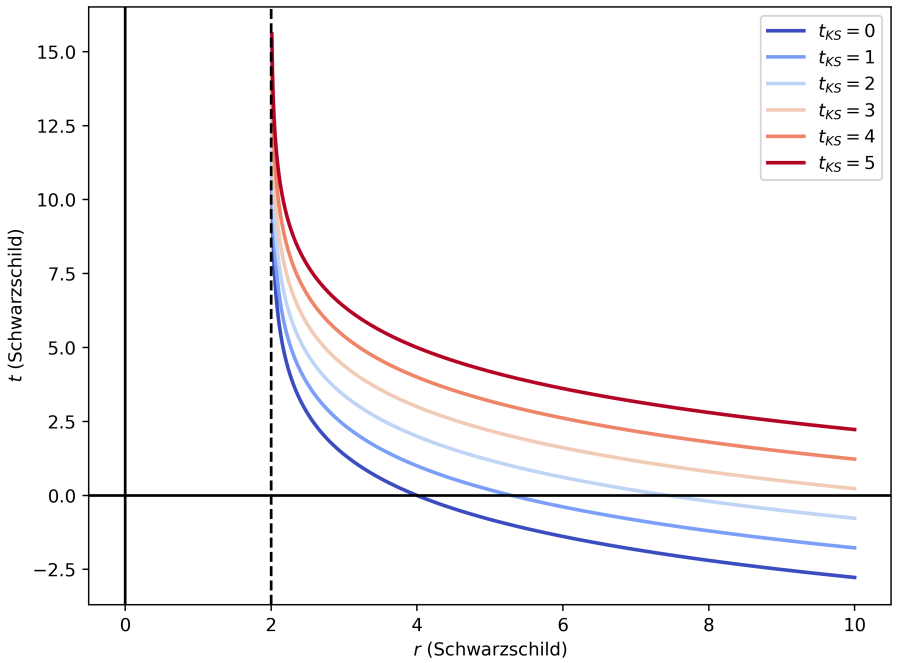


Figure 4.2: Constant KS time curves in Schrawzchild coordinates. A mass of $M = 1$ was used to produce this plot, giving a Schwarzchild radius of $r = 2$. Notice that, at the event horizon (dashed line), any finite change KS time relates to an infinite passage of Schwarzchild time. Note also that the constant time curves in KS become asymptotically parallel to those in Schwarzchild coordinates at large radii.

of the simulation as well as its computational efficiency. Previous works sought to address these issues by excising a central sphere around the BH and only evolving the regions exterior to this [4, 108]. While this approach is valid in simple situations, in cases where there are non-trivial dynamics within or close to the excised region, potentially important details will be unresolved. To access these proximal regions numerically, it is necessary to use a horizon-penetrating gauge. In this chapter, we make use of the *Kerr-Schild* (KS) basis. This coordinate system corresponds to observers that are infalling along null geodesics and therefore exhibits the necessary horizon-penetration. Fig. 4.2 depicts the difference in temporal evolution between the (non-horizon-penetrating) Schwarzchild and KS bases as a function of radius.

In what follows, we present a method for calculating the isotropic DF force using KS coordinates. The process of transforming data between the two gauges is not

entirely intuitive and we provide a detailed overview of this in section 4.3. Before tackling the augmented computation, however, we begin with a description of the process purely within isotropic coordinates.

4.2 Isotropic Formulation

In this section we describe the procedure used to determine the DF corresponding to inertial, asymptotic observers that is induced by a scalar field on a boosted BH using a fixed isotropic background. The information in this section is based on the following works [4, 108].

4.2.1 Metric

The natural choice of inertial observers would be those associated with the standard Schwarzchild coordinates. Unfortunately, however, the use of these coordinates leads to analytically unwieldy expressions in the $3+1$ decomposition of the metric. It is much easier to consider the observers for whom the spatial metric appears locally isotropic. This is the case for observers infalling along time-like geodesics. At spatial infinity, these observers will asymptotically agree with their Schwarzchild counterparts. If a numerical simulation is carried out over a sufficiently large spatial volume, then the frictional forces can be obtained using the isotropic gauge.

In general, and as discussed in chapter 3, the choice of space-time foliation scheme ($3+1$ decomposition) is central to success of numerical time integration. In the case of a fixed background, however, it is sufficient to use the simple ADM decomposition. In isotropic coordinates, the ADM decomposition of the Schwarzchild metric can be written:

$$ds^2 = -\alpha^2 dt^2 + \gamma_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt), \tag{4.2.1}$$

where,

$$\alpha^2 = \left(\frac{1-\frac{M}{2r}}{1+\frac{M}{2r}}\right)^2, \tag{4.2.2}$$

$$\beta^i = 0, \tag{4.2.3}$$

$$\gamma_{xx} = \gamma_{yy} = \gamma_{zz} = \left(1+\frac{M}{2r}\right)^4. \tag{4.2.4}$$

To experience a non-zero DF force, the BH must be boosted with respect to the surrounding scalar field. There are of course two canonical perspectives from which to view this boost. One could hold the BH in place and initialise the scalar field with a homogeneous initial velocity, or, instead, boost the BH relative to a non-moving field. The former may seem more straight-forward, however, it would lead to complicated boundary conditions. It is, in fact, more straightforward to apply the boost directly to the BH. This introduces another issue, namely that the BH will eventually drift outside any finite simulation region. To lock the BH in place, the Lorentz boost can be combined with a Galilean shift in the opposite direction. Relative to these boosted-shifted coordinates (t, x, y, z), the un-boosted and un-shifted coordinates $(t_{\rm ub}, x_{\rm ub}, y_{\rm ub}, z_{\rm ub})$ can be written:

$$t_{\rm ub} = (t/\gamma - \gamma v x); \quad x_{\rm ub} = \gamma x; \quad y_{\rm ub} = y; \quad z_{\rm ub} = z. \tag{4.2.5}$$

The boosted, shifted isotropic ADM variables can be written:

$$\alpha^2 = \frac{AB}{\gamma^2(B-Av^2)}, \qquad \beta^i = \delta^i_x A v, \tag{4.2.6}$$

$$\gamma_{xx} = \gamma^2(B-Av^2), \qquad \gamma_{yy} = \gamma_{zz} = B. \tag{4.2.7}$$

and A and B are given by:

$$A = \left(\frac{1-\frac{M}{2\bar{r}}}{1+\frac{M}{2\bar{r}}}\right)^2; \tag{4.2.8}$$

$$B = \left(1 + \frac{M}{2\bar{r}}\right)^4, \tag{4.2.9}$$

where $\bar{r}^2 = \gamma^2 x^2 + y^2 + z^2$. In the above expressions, γ refers to the Lorentz factor:

$$\gamma = \frac{1}{\sqrt{1 - v^2}}. \tag{4.2.10}$$

In general, one might expect that boosting the metric in this way would have an effect on the measured DF. Conveniently, it can be shown that no such correction to the DF force is required as momenta remain invariant under the boost and shift transformations [4]:

$$\frac{dP_x}{dt} = \frac{dP_{x_{\rm ub}}}{dt_{\rm ub}}, \tag{4.2.11}$$

where P_x and $P_{x_{\rm ub}}$ are the momenta in the boosted and un-boosted frames respectively.

4.2.2 Scalar Field

Scalar fields are the simplest to implement numerically and study analytically. In addition, their existence around BH is well motivated with ALPs expected to undergo superradiant amplification upon a Kerr type background. Real scalar fields lack a phase degree of freedom, and so will propagate via mass dependent oscillations in their energy density:

$$\varphi_{\rm Re} = \varphi_0 \cos(\mu t). \tag{4.2.12}$$

This can be problematic when evolving the field numerically, leading to increased errors. Complex scalar fields, however, do not suffer this issue and yet retain most of the simplicity afforded by their real counterparts. Indeed, in the absence of any metric back-reaction, the real and imaginary parts of the field will decouple from one another [4]. The action of a minimally coupled complex scalar field can be written:

$$S = \int d^4x\sqrt{-g}\left[\frac{1}{2}g^{\mu\nu}\nabla_\mu\varphi^*\nabla_\nu\varphi - V(|\varphi|)\right]. \tag{4.2.13}$$

where the potential is simply $V = \frac{1}{2}\mu^2|\varphi|^2$. The evolution of this field is given by eq. (3.3.22) and eq. (3.3.24).

The stress-energy is crucial to determination of the DF and diagnostic quantities described below. For the action given by eq. (4.2.13), this tensor can be expressed as:

$$T_{\mu\nu} = \frac{1}{2}\left[(\nabla_\mu\varphi\nabla_\nu\varphi^* + \nabla_\nu\varphi\nabla_\mu\varphi^*) - g_{\mu\nu}\nabla_\lambda\varphi\nabla^\lambda\varphi^* - g_{\mu\nu}V(|\varphi|)\right]. \tag{4.2.14}$$

The stress-energy can be written in ADM form by its projections onto the normals to the spatial slices using equations eq. (3.3.10) to eq. (3.3.13).

With the above specification of static, boosted metric, the choice of initial conditions is somewhat arbitrary - Working in the limit of zero back-reaction, the elliptical constraint equations can be neglected, and since the BH already encodes the boost, the only requirement on the initial field configuration is that it is non-moving. The simplest non-trivial field configuration is that with a spatially constant, non-zero amplitude φ, given by:

$$\mathrm{Re}\left[\varphi(t=0,r)\right] = \varphi_0; \tag{4.2.15}$$

$$\mathrm{Im}\left[\varphi(t=0,r)\right] = 0; \tag{4.2.16}$$

$$\mathrm{Re}\left[\partial_t\varphi(t=0,r)\right] = 0; \tag{4.2.17}$$

$$\mathrm{Im}\left[\partial_t\varphi(t=0,r)\right] = \mu\varphi_0. \tag{4.2.18}$$

Similarly, selecting appropriate boundary conditions is also straightforward in this regime. All of the interesting dynamics will be fixed around the centre of the simulation region, so it is appropriate to simply set homogeneous boundary conditions at a large distance from the BH. AMR is required to provide sufficient resolution

near the event-horizon while also allowing the simulation domain to reach sufficiently into the far-field.

4.2.3 Diagnostics and Friction

DF is a force, and as such, can be determined by considering the rate of change of momentum of the BH. Working in the fixed background limit, changes in the BH momentum are neglected by the simulation. To extract an effective force, it is necessary to consider changes in the dynamical momentum of the scalar field. EFEs contain a set of 4 elliptical constraint equations pertaining to energy and momentum conservation. In the limit of zero back reaction, energy conservation holds, however, as we discuss below, momentum conservation is broken. This is a consequence of the structure of the metric and can be taken into account with the introduction of a momentum source term related to the exchange of momentum with the background. This adapted momentum constraint and indeed the standard energy conservation law constitute important diagnostic tools in numerical GR. As we describe in section 4.4, many simulation parameters such as grid spacing and derivative fidelity can lead to instabilities that render the results invalid. Verifying the Hamiltonian and momentum constraints is necessary to ensure the accuracy of a given simulation. In this section, we present a derivation of, firstly, the more straightforward energy constraint, and, later, the momentum constraint. Finally, we identify the terms in the momentum constraint that are associated with DF.

One can define a general current J^μ in GR by projecting the Stress-Energy tensor onto a particular direction ζ^μ:

$$J^\mu = \zeta^\nu T^\mu_\nu. \tag{4.2.19}$$

The energy current ($J^\mu_{\rm E}$) is associated with the projection onto the time direction ∂_t, hence $\zeta^\mu_{\rm E} = (1, 0, 0, 0)$. This choice of $\zeta^\mu_{\rm E}$ is in fact a Killing vector of the metric, i.e.:

$$\mathcal{L}_{\zeta_{\rm E}} g_{\mu\nu} = \nabla_\mu \zeta_{{\rm E}\nu} + \nabla_\nu \zeta_{{\rm E}\mu} = 0, \tag{4.2.20}$$

where $\mathcal{L}_{\zeta_{\rm E}} g_{\mu\nu}$ is the Lie derivative of the metric along the direction of $\zeta^\mu_{\rm E}$. This shows that the metric is unchanged under flows along this direction. This choice of projection vector therefore encodes a continuous symmetry of the system. Consequently, Noether's theorem implies conservation of the energy current:

$$\nabla_\mu J^\mu_{\rm E} = \nabla_\mu (T^{\mu\nu} \zeta_{{\rm E}\nu}), \tag{4.2.21}$$

$$= 0. \tag{4.2.22}$$

where the second line arises from the Killing vector relation and the constraint laws of EFEs, which imply that:

$$\nabla_\mu T^{\mu\nu} = 0. \tag{4.2.23}$$

Integrating over the divergence of the current gives:

$$\int d^4x \sqrt{-g}\, \nabla_\mu J^\mu_{\rm E} = 0. \tag{4.2.24}$$

Working in ADM decomposition, we wish to rewrite eq. (4.2.24) in terms of integrals over the individual spatial slices. To do so, we define the energy *charge* $Q_{\rm E}$ associated with the conserved current ($J^\mu_{\rm E}$) as its projection onto the normal n^μ to a given 3d spatial hypersurface Σ:

$$Q_{\rm E} = n_\mu J^\mu_{\rm E}. \tag{4.2.25}$$

For the spatial surface Σ, the unit normal is in the time direction and can be written:

$$n^\mu = \left(\frac{1}{\alpha}, -\frac{\beta^i}{\alpha} \right) \tag{4.2.26}$$

With a dual vector given by $n_\mu = \alpha(1,0,0,0)$. We can re-express eq. (4.2.24) as a relationship between the temporal variation in the net energy $Q_{\rm E}$ within the volume Σ and the total flux across its boundary $\partial\Sigma$ [94]:

$$\frac{\partial}{\partial t}\int_\Sigma d^3x\,\sqrt{\gamma}\,Q_{\rm E} = -\alpha\int_\Sigma d^3x\,\sqrt{\gamma}\,\nabla_i J^i_{\rm E}. \tag{4.2.27}$$

This relationship can be obtained by separating the spatial and temporal components of the covariant derivative:

$$\nabla_\mu J^\mu_{\rm E} = \partial_t J^0_{\rm E} + \Gamma^\mu_{\mu 0} J^0_{\rm E} + \nabla_i J^i_{\rm E} = 0. \tag{4.2.28}$$

and noting that the temporal Christoffel symbols can be written as a time derivative of the volume form:

$$\Gamma^\mu_{\mu 0} = \frac{1}{\sqrt{g}}\partial_t\sqrt{g}. \tag{4.2.29}$$

We conclude due to the static nature of the background, that $\Gamma^\mu_{\mu 0} = 0$. Hence:

$$\frac{\partial}{\partial t}\int_\Sigma d^3x\,\sqrt{\gamma}\,J^0_{\rm E} = -\int_\Sigma d^3x\,\sqrt{\gamma}\,\nabla_i J^i_{\rm E}. \tag{4.2.30}$$

Finally, substituting $J^0_{\rm E} = Q_{\rm E}/\alpha$ yields eq. (4.2.27). We can now apply Gauss's law to this equation to obtain:

$$\frac{\partial}{\partial t}\int_\Sigma d^3x\,\sqrt{\gamma}\,Q_{\rm E} = -\int_{\partial\Sigma} d^2x\,\sqrt{\sigma}\,F_{\rm E}. \tag{4.2.31}$$

Where $\partial\Sigma$ is the 2d spatial hypersurface with induced metric σ that bounds the 3d spatial volume Σ and $F_{\rm E}$ is the flux through this surface, given by:

$$F_{\rm E} = \alpha N_i J^i_{\rm E}. \tag{4.2.32}$$

Here N^i is the 3d vector describing the normal to $\partial\Sigma$. We will take $N_i = (r = 1, \theta = 0, \phi = 0)$, i.e. pointing radially outwards, thereby defining $\partial\Sigma$ to be a spatial 2-sphere. It will be useful in the following section on KS coordinates to have covariant expression for the flux. To obtain this, we can extend our definition of $N_i \to N_\mu$ with the following normalisation condition $N_\mu N^\mu = 1$. This implies that $N_0 = \beta^i N_i$, hence, we derive the covariant force term:

$$F_{\rm E} = \alpha N_\mu J^\mu_{\rm E} - \alpha N_0 J^0_{\rm E} \tag{4.2.33}$$

$$= \alpha N_\mu J^\mu_{\rm E} - \beta^\mu N_\mu n_\nu J^\nu_{\rm E} \tag{4.2.34}$$

$$= N_\mu(\alpha J^\mu_{\rm E} - \beta^\mu n_\nu J^\nu_{\rm E}). \tag{4.2.35}$$

Where we assign $\beta^0 = 0$. We have now arrived at eq. (4.2.31) that relates an integral over the energy density $Q_{\rm E}$ contained within a spherical volume to the flux through its boundary. By computing these values at regular intervals, the convergence of the simulation can be verified. It is in this sense that we use this conservation law as a diagnostic.

A similar calculation can be carried out to obtain the momentum flux, with the momentum in a particular spatial direction being associated with the projection of the stress-energy onto that direction. The difficulty in this case is that due to the spatial variation of metric, the current associated with momentum $J^\mu_{\rm P} = \zeta^\nu_{\rm P} T^\mu_\nu$ is not conserved. Put another way, any spatial momentum projection vector, e.g. $\zeta^\mu_{\rm P} = (0, x = 1, y = 0, z = 0)$, will not satisfy Killing's equation, and as a consequence, Noether's theorem implies the existence of a source term $S_{\rm P}$, such that:

$$\int d^4x\,\sqrt{-g}\,\nabla_\mu J^\mu_{\rm P} = \int d^4x\,\sqrt{-g}\,T^\mu_{\ \nu}\nabla_\mu \tilde{\zeta}^\nu_{\rm P}. \tag{4.2.36}$$

Which, repeating the above procedure, leads to:

$$\partial_t \int_\Sigma d^3x\,\sqrt{\gamma}\,Q_{\rm P} = -\int_{\partial\Sigma} d^2x\,\sqrt{\sigma}\,F_{\rm P} + \int_\Sigma d^3x\,\sqrt{\gamma}\,S_{\rm P}, \tag{4.2.37}$$

where:

$$S_{\mathrm{P}} = \alpha T^{\mu}_{\ \nu} \nabla_{\mu} \zeta^{\nu}_{\mathrm{P}}. \tag{4.2.38}$$

This equation relates the rate of change of momentum within the volume Σ to the flux across its boundary combined with the momentum generated in its bulk due to the lack of spatial translational symmetry in the metric. There are a number of subtleties regarding the integrals presented in this subsection. Firstly, it should be noted that at each point in the domain of the integrals, a different frame of reference is used and the quantities within the integrand are those that would be measured by an observer within this frame. Each of these frames corresponds to isotropic observers located at a specific location. At an infinite distance from the BH, all observers will be stationary and will agree on the values of the integrands and the integrals will effectively be carried out within a single frame. By performing the flux integration at sufficiently large radii, one can recover the change in energy and momentum expected by an inertial asymptotic observer.

The second, more problematic feature of these integrals is that we have taken Σ to fully encompass the BH. To perform the simulation, one would thus be required to compute the matter evolution in the presence of metric singularities. In isotropic coordinates, this issue is two-fold as a the system exhibits not only the true origin singularity, but also a coordinate singularity at the event-horizon. To account for this, a central region of Σ must be excised. The boundary surface $\partial\Sigma$ now becomes a pair of concentric disconnected spheres, see fig. 4.3. No evolution will be preformed within the inner excision region. Ideally this buffer zone would be as small as possible to reduce the loss of information. If the excision boundary is too close to the BH, however, a significant degree of noise can be introduced [4].

We now consider how the momentum constraint eq. (4.2.37) relates to the force on the BH. From a Newtonian perspective, DF can be computed by summing the gravitational force generated by the surrounding matter. For a continuous fluid, this

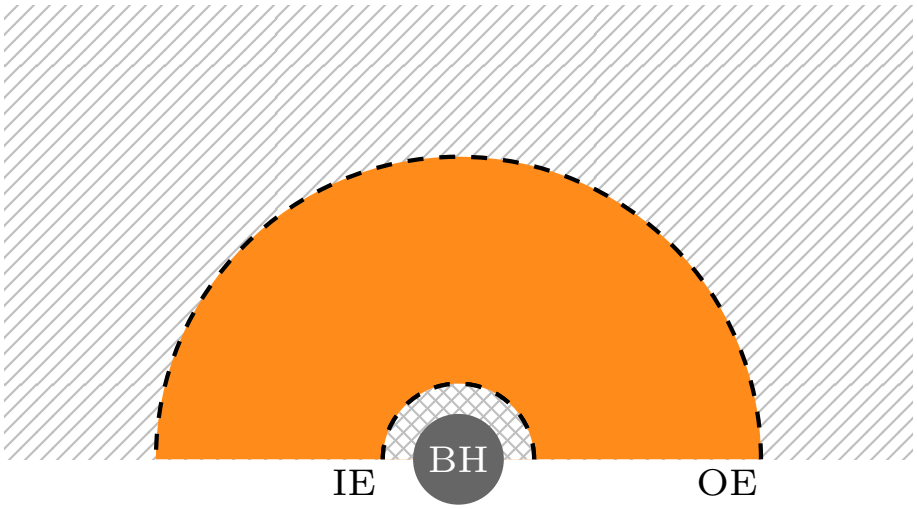


Figure 4.3: An illustration of the inner (IE) and outer (OE) excision regions. Evolution is carried out in all but the cross-hatched region within the inner excision. The separation between the inner excision and the BH prevents numerical instabilities. The computation of diagnostic quantities and frictional forces is performed only in the orange region between the excision radii.

force, projected onto the x^i direction, will be given by:

$$\mathcal{F}^i_{\text{d, nonrel}} = \int \frac{GM\rho}{r^2}\frac{x^i}{r}dV. \tag{4.2.39}$$

In GR, obtaining an integral for the DF force is less-trivial as gravity is encoded in the curvature of spacetime. In a full relativistic scenario, the matter field will generate its own perturbation to the curvature and it is this modification that leads to the relativistic version of DF. Working in the limit of negligible back reaction, this perturbation to the background metric is encoded in the non-conservation of momentum from the context of the scalar field. This is equivalent to how the momentum of a ball thrown in the air on earth is not conserved, because it is transferred to the earth and neglected. From this, one can conclude that the total relativistic force on the BH is given by:

$$\mathcal{F} = \int_\Sigma d^3x\,\sqrt{\gamma}\,S_{\text{P}}. \tag{4.2.40}$$

Where, in this equation, Σ is the entire volume of 3d space. It is common to make the

distinction between the force due to DF and that which is induced by the accretion of matter into the BH. Taking the Newtonian result as inspiration, the DF should arise from the integral over the field that is exterior to the BH. The accretion term is then associated with the momentum flux into the event-horizon. Breaking the force up in this way, we can write:

$$\mathcal{F} = \int_{\Sigma - \Sigma_{\rm BH}} d^3x \, \sqrt{\gamma} \, S_{\rm P} + \int_{\partial \Sigma_{\rm BH}} d^2x \, \sqrt{\sigma} \, F_{\rm P}, \tag{4.2.41}$$

where $\Sigma_{\rm BH}$ is the volume encompassed by the event-horizon. Here, the right most integral, known as the *Bondi* term [115] describes accretion, whilst the left most integral over the source describes the DF. In simulations, $\Sigma_{\rm BH}$ can be approximated by the inner excision sphere.

The distinction made between the accretion and friction terms in eq. (4.2.41) seems very intuitive, and indeed, within a specific gauge, this makes sense. It is easy to see, however, that under a change of observer, these terms can mix. Take for example a horizon-penetrating, compared with a non-penetrating, gauge. The accretion flux in the non-penetrating gauge will be zero, whilst in general this will not be the case in the penetrating gauge.

Note that in a simulation where the full integration volume is finite, the total rate of change of momentum eq. (4.2.37) also includes a flux integral over the outer boundary. This corresponds to the momentum of the field entering the simulation region. During the early stages of the evolution from the initial conditions, the net momentum of the field within the simulation domain may evolve somewhat. Eventually, however, the system should settle to a steady state where the total momentum within the domain is constant. At this point, the net momentum flux will exactly cancel the total source momentum. Once this point is reached, the rate of change of the integrated momentum flux can be interpreted as the total force on the BH. Additionally, because this flux integral is carried out over a sphere of large radius, this force will be approximately that corresponding to inertial asymptotic

observers, as desired. Overall then, we write the net force (DF plus Bondi term) acting on the BH as:

$$\mathcal{F} = -\int_{\partial\Sigma} d^2x\,\sqrt{\sigma}\,F_{\rm P} + \int_{\partial\Sigma_{\rm BH}} d^2x\,\sqrt{\sigma}\,F_{\rm P}. \tag{4.2.42}$$

Which is valid provided that the simulation has reached a state in which:

$$\partial_t \int_{\Sigma} d^3x\,\sqrt{\gamma}\,Q_{\rm P} = 0. \tag{4.2.43}$$

4.3 Kerr-Schild Formulation

We now move on to discuss our novel approach to computing the DF force using the horizon-penetrating KS gauge. We begin in much the same was as for the isotropic case. A number of corrections and modifications must be made along the way to account for the difference in observers, however. We start with a description of the ADM Schwarzchild metric in the KS basis, following this, we discuss the changes made to the scalar field initial conditions. Finally, we derive an augmented expression for the DF.

4.3.1 Metric

KS coordinates were first introduced to describe the rotating Kerr metric. In this work we take advantage of their manifest horizon-penetration. The KS basis is constructed by considering observers that are infalling along null geodesics. These correspond to massless particles and therefore, do not asymptote to stationary observers at spatial infinity. This feature significantly complicates the evaluation of the forces that may act upon the BH, and requires several transformations and additional correction terms to yield agreement with the isotropic case.

A KS observer will posses a null velocity vector, ℓ^μ, satisfying $g_{\mu\nu}\ell^\mu\ell^\nu = 0$ and will follow a geodesic of the form:

$$\ell^\mu \nabla_\mu \ell^\nu = 0. \tag{4.3.1}$$

At an infinite distance from the BH, $g_{\mu\nu} \to \eta_{\mu\nu}$. Expanding around this flat space, the KS form of a metric is given by:

$$g_{\mu\nu} = \eta_{\mu\nu} + H\ell_\mu \ell_\nu, \tag{4.3.2}$$

where ℓ^μ is chosen such that it is null with respect to both $\eta_{\mu\nu}$ and $g_{\mu\nu}$, and H is a function of the spatial components. For the Schwarzchild metric, the infalling null geodesic is:

$$\ell_\mu = (-1, x/r, y/r, z/r). \tag{4.3.3}$$

and H is given by:

$$H = \frac{2M}{r}. \tag{4.3.4}$$

As in the isotropic case above, we again boost and shift the metric according to eq. (4.2.5). The boosted and shifted KS metric can be written in the form eq. (4.2.1) with the ADM variables being given by:

$$\alpha = \sqrt{\beta^i \gamma_{ij} \beta^j + (1 - v^2) - 2H_b \ell_{b,0} \ell_{b,0}}; \tag{4.3.5}$$

$$\beta^i = \gamma^{ij}(-v\delta_{0j} + 2H_b \ell_{b,j} \ell_{b,0}); \tag{4.3.6}$$

$$\gamma_{ij} = \delta_{ij} + 2H_b \ell_{b,i} \ell_{b,j}, \tag{4.3.7}$$

where $H_b = M/r_b$ and $\ell_b = (-1, x_b/r_b, y/r_b, z/r_b)$ are the boosted forms of eq. (4.3.4) and eq. (4.3.3), for which:

$$x_b = \gamma x; \tag{4.3.8}$$

$$r_b = \sqrt{\gamma^2 x^2 + y^2 + z^2}. \tag{4.3.9}$$

4.3.2 Scalar field

As in the isotropic case, we again perform the simulations using a complex scalar field. We do modify the initial conditions, however. A constant time slice in isotropic coordinates will not correspond to a constant time slice in KS. We account for this by computing the isotropic time coordinate at each point in the spatial KS simulation volume and evolving the isotropic initial conditions as plane waves up to this time. Explicitly, this is:

$$\mathrm{Re}\left[\varphi(t=0,r)\right] = \varphi\cos\left(\mu t_{\mathrm{iso}}(r_{\mathrm{KS},\mu})\right); \tag{4.3.10}$$

$$\mathrm{Im}\left[\varphi(t=0,r)\right] = -\varphi\sin\left(\mu t_{\mathrm{iso}}(r_{\mathrm{KS},\mu})\right); \tag{4.3.11}$$

$$\mathrm{Re}\left[\partial_t\varphi(t=0,r)\right] = -\mu\varphi\sin\left(\mu t_{\mathrm{iso}}(r_{\mathrm{KS},\mu})\right); \tag{4.3.12}$$

$$\mathrm{Im}\left[\partial_t\varphi(t=0,r)\right] = \mu\varphi\cos\left(\mu t_{\mathrm{iso}}(r_{\mathrm{KS},\mu})\right), \tag{4.3.13}$$

where $t_{\mathrm{iso}}(r_{\mathrm{KS},\mu})$ is the isotropic time coordinate written as a function of the spatial KS coordinates $r_{\mathrm{KS},\mu}$. Whilst this correction isn't strictly necessary, we expect it to improve the rate of convergence between the two approaches before the system has settled into a steady state.

4.3.3 Diagnostics and Friction

This subsection concerns the calculation of Boosted Isotropic (BI) quantities using a Boosted Kerr-Schild (BKS) basis. To avoid confusion, we denote BI coordinates with a bar and BKS without. For example, the $v^{\bar{\mu}}$ signifies a vector expressed in BI coordinates and v^{μ} is the same vector expressed in BKS coordinates. We further denote quantities pertaining to the BI basis with a bar, so, for example, the BI spatial unit normal would be $\bar{n}_{\bar{\mu}}$ expressed in BI coordinates and $\bar{n}_{\mu}$ in BKS coordinates.

As identified in the discussion for isotropic observers, the DF is an observer dependent quantity. In GR, classical Newtonian descriptions of DF are associated with asymptotic, inertial observers. Whilst BKS coordinates are asymptotically flat, they are not inertial. Although we could directly apply the framework of section 4.2 in BKS coordinates, the resulting force would not match the isotropic result. This issue arises because, in general, the temporal and spatial directions will not agree between two arbitrary gauges. This means that one would expect that $\zeta \neq \bar{\zeta}$ for $\zeta^\mu = \bar{\zeta}^\mu$ and similarly for the surface normals n^μ. The canonical approach to solving this problem would be to exploit covariance to write the BI $\bar{\zeta}$ and $\bar{n}$ in BKS coordinates. We will begin with this strategy, and later highlight a number of issues with this naive approach.

Starting with the straightforward energy diagnostic, we note that the left hand side of eq. (4.2.31) is simply an integral over the scalar quantity:

$$\bar{Q}_{\mathrm{E}} = \bar{n}_{\bar{\mu}} \bar{J}^{\bar{\mu}}_{\mathrm{E}}. \tag{4.3.14}$$

where the current is given by:

$$\bar{J}^{\bar{\mu}}_{\mathrm{E}} = \bar{\zeta}^{\bar{\nu}}_{\mathrm{E}} T^{\bar{\mu}}_{\bar{\nu}}. \tag{4.3.15}$$

To compute $\bar{Q}_{\mathrm{E}}$ using a simulation performed in BKS coordinates, each of the covariant, BI quantities must be transformed into the BKS basis $x^{\bar{\mu}} \to x^\mu$. We can define the Jacobian for this transformation $\mathcal{J}^\mu_{\bar{\nu}}$ as follows:

$$\mathcal{J}^\mu_{\bar{\nu}} = \frac{\partial x^\mu}{\partial x^{\bar{\nu}}}. \tag{4.3.16}$$

Working with non-dynamical backgrounds, $\mathcal{J}^\mu_{\bar{\nu}}$ can be found simply via analytical pre-computation. The simulation, when performed in BKS coordinates can generate the stress-energy in BKS form, and although its coordinate representation $T^{\mu\nu}$ will differ from that in BI coordinates $\bar{T}^{\bar{\mu}\bar{\nu}}$, the tensors themselves will approximately

match for any given space-time event $\bar{T}(x^{\bar{\mu}}) = T(x^{\mu})$. To find $\bar{Q}_{\mathrm{E}}$ then, we need only transform $\bar{n}_{\bar{\mu}}$ and $\bar{\zeta}^{\bar{\mu}}_{\mathrm{E}}$ by contraction with $\mathcal{J}^{\mu}_{\bar{\nu}}$.

Returning to eq. (4.2.31), we must now perform a 3d volume integral over the spatial hyper-surface $\bar{\Sigma}$. This is where issues start to arise. Assuming $\bar{\Sigma}$ is spherical in BI space, then in general it will be aspherical in BKS coordinates and the temporal and spatial components may even mix; see fig. 4.4. In addition, the volume forms $\sqrt{\bar{\gamma}} \neq \sqrt{\gamma}$ will also differ. If this were a volume element on the whole space, i.e. $\sqrt{-g}$, then accounting for this change would be straightforward, with the transformed element being given by:

$$\sqrt{-g} = |\det \mathcal{J}^{\mu}_{\bar{\nu}}| \sqrt{-\bar{g}}. \tag{4.3.17}$$

Instead of integrating over a non-trivial re-parametrisation of $\bar{\Sigma}(x^{\mu})$ written in the BKS basis and transforming the non-covariant volume element, we opt to perform a spatially spherical integral in BKS coordinates using the BKS induced volume element $\sqrt{\gamma}$. This integral corresponds to the BI energy contained within a non-spherical BI surface. We thus obtain the following expression for the augmented left hand side of eq. (4.2.31):

$$\frac{\partial}{\partial t} \int_{\Sigma} d^3x \, \sqrt{\gamma} \, Q_{\mathrm{E}} \to \frac{\partial}{\partial t} \int_{\Sigma} d^3x \, \sqrt{\gamma} \, \bar{Q}_{\mathrm{E}}. \tag{4.3.18}$$

This is easier to compute numerically, as the BKS quantities are accessible to the simulation and a spherical integration scheme already exists within the code base of `GRDZhadzha`. We will later show how correction terms can be introduced to resolve the disagreement caused by using different integration volumes.

We now consider the right hand side of eq. (4.2.31), where the situation becomes yet murkier. To relate energy and flux, we once again deconstruct the covariant derivative to obtain eq. (4.2.30). In this case, however, it is less straight forward to relate $\bar{J}^0_{\mathrm{E}}$ to $\bar{Q}_{\mathrm{E}}$ because in the BKS coordinates, $\bar{n}_{\mu}$ will no longer simply be given by $(\bar{\alpha}, 0, 0, 0)$. In order to reproduce the conservation law, we are forced to replace

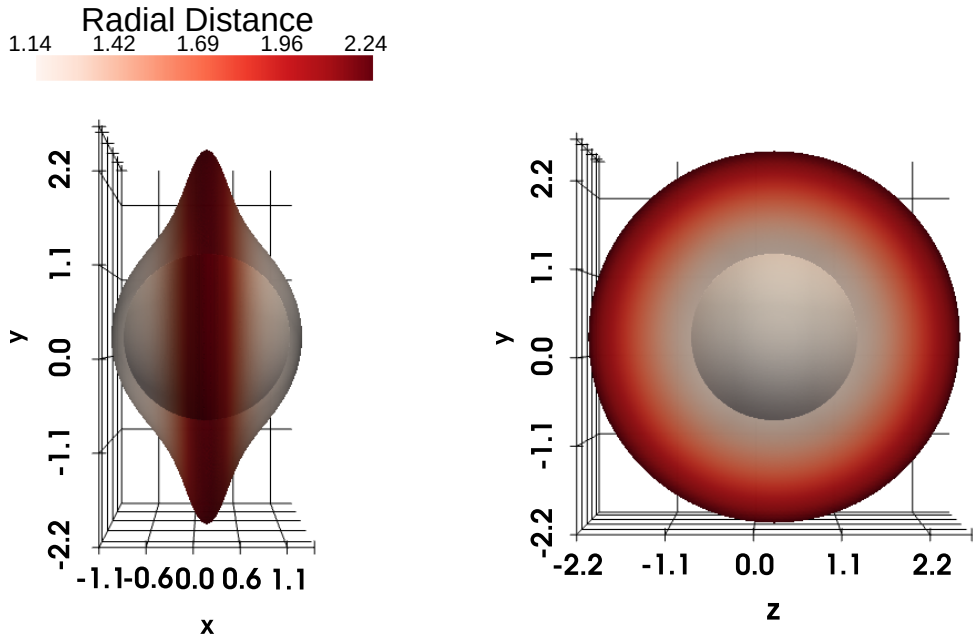


Figure 4.4: A BI sphere of radius $\bar{r} = 1$ as viewed by a BKS observer with boost velocity $v = 0.99$ in the x direction. The colour indicates the radial distance from the centre. The original sphere, i.e. a radius $r = 1$ BKS sphere is shown within the morphed BI sphere. At $v = 0$, both spheres agree. The left plot shows the x-y projection, and the right plot shows the z-y projection.

$\bar{Q}$ with a projection of the current onto the BKS spatial normal $\bar{Q}_{\rm E} \to n_\mu \bar{J}^\mu_{\rm E}$; the difference between the normal vectors $\bar{n}_\mu$ and n_μ is depicted in fig. 4.5. We can then write:

$$\frac{\partial}{\partial t}\int_\Sigma d^3x\,\sqrt{\gamma}\,n_\mu \bar{J}^\mu_{\rm E} = -\int_{\partial\Sigma} d^2x\,\sqrt{\sigma}\,N_\mu(\alpha \bar{J}^\mu_{\rm E} - \beta^\mu n_\nu \bar{J}^\nu_{\rm E}). \tag{4.3.19}$$

This equation is no longer related directly to the energy measured by a BI observer or a BKS observer but can provide an important diagnostic. One could simply use eq. (4.2.31) in the BKS basis to compute the BKS energy and use this to verify the numerical stability. By using the hybrid expression of eq. (4.3.19) it is also possible to verify the accuracy of the transformations.

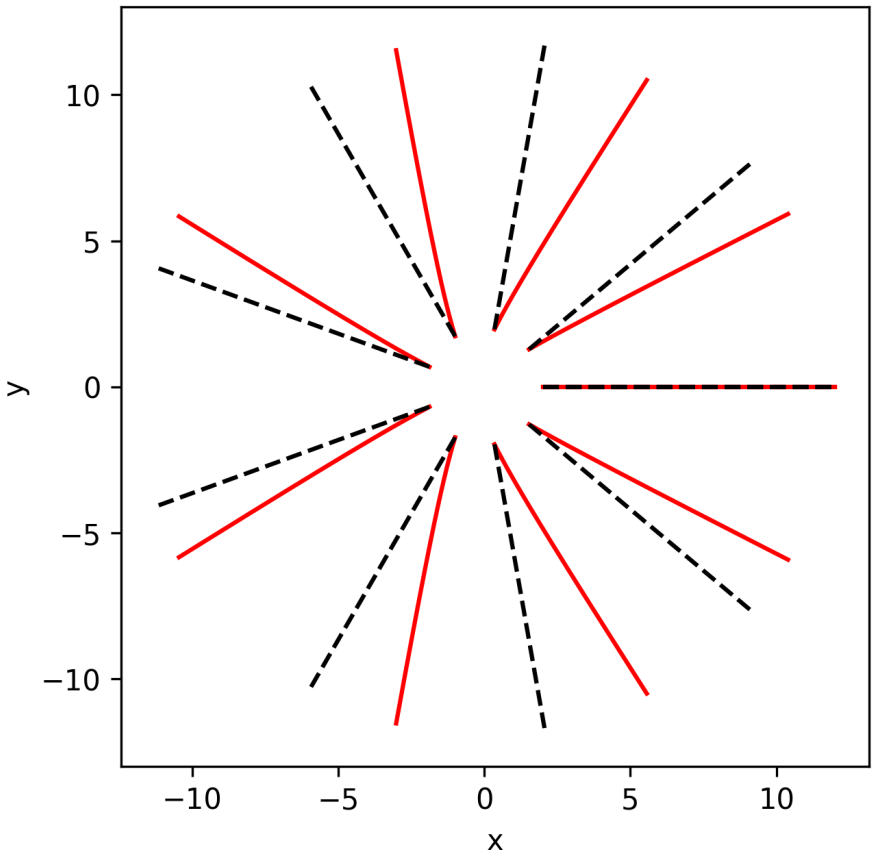


Figure 4.5: Lines in BKS coordinates projected onto the $x-y$ plane tracing the spatial normal directions in BKS, N_μ (dashed black), and BI, $\bar{N}_\mu$ (red). A boost velocity of $v = 0.9$ in the x direction and BH mass of $M = 1.0$ were used to produce this figure. Note that the spatial components of the normal directions agree for zero boost velocity.

We now turn to the momentum continuity equation. The momentum current in the BI x direction is found by projecting the stress-energy onto $\bar{\zeta}^{\bar{\mu}}_{\rm P} = (0, x = 1, y = 0, z = 0)$:

$$\bar{J}^{\mu}_{\rm P} = \bar{\zeta}^{\nu}_{\rm P} T^{\mu}_{\nu}. \tag{4.3.20}$$

This introduces an additional source term as before. Overall, performing the integrations in BKS space, we obtain the quasi BI momentum diagnostic:

$$\frac{\partial}{\partial t}\int_\Sigma d^3x\,\sqrt{\gamma}\,n_\mu \bar{J}^\mu_{\rm P} = -\int_{\partial\Sigma} d^2x\,\sqrt{\sigma}\,N_\mu(\alpha\bar{J}^\mu_{\rm P} - \beta^\mu n_\nu \bar{J}^\nu_{\rm P}) + \int_\Sigma d^3x\,\sqrt{\gamma}\,\alpha T^\mu{}_\nu \nabla_\mu \bar{\zeta}^\nu_{\rm P}. \tag{4.3.21}$$

This expression is in fact general for any choice of stress-energy projection direction ζ. Whilst we could use eq. (4.3.21) for identifying convergence issues, it is not directly related to the DF force. In the previous section, we related the source term to the change in the ADM momentum of the BH through eq. (4.2.41) and showed that once a steady state is reached, the DF can be expressed purely in terms of the momentum flux. In this case, the steady state can be identified with the condition:

$$\frac{\partial}{\partial t}\int_\Sigma d^3x\,\sqrt{\gamma}\,\bar{Q}_{\rm P} = 0, \tag{4.3.22}$$

Once eq. (4.3.22) is satisfied, the BI force acting on the BH in the x direction will be given by:

$$\bar{\mathcal{F}}^{\bar{x}} = \int_{\partial\Sigma} d^2x\,\sqrt{\sigma}\,\bar{N}_\mu(\bar{\alpha}\bar{J}^\mu_{\rm P} - \bar{\beta}^\mu \bar{n}_\nu \bar{J}^\nu_{\rm P}). \tag{4.3.23}$$

We now define a correction vector of the form:

$$b_\mu = (N_\mu\alpha - N_\nu\beta^\nu n_\mu) - (\bar{N}_\mu\bar{\alpha} - \bar{N}_\nu\bar{\beta}^\nu\bar{n}_\mu). \tag{4.3.24}$$

Finally, using eq. (4.3.21) we can write:

$$\bar{\mathcal{F}}^{\bar{x}} + \int_{\partial\Sigma} d^2x\,\sqrt{\sigma}\,(b_\mu \bar{J}^\mu_{\rm P}) = \tag{4.3.25}$$

$$\int_{\partial\Sigma} d^2x\,\sqrt{\sigma}\,N_\mu(\alpha\bar{J}^\mu_{\rm P} - \beta^\mu n_\nu \bar{J}^\nu_{\rm P}) = \int_\Sigma d^3x\,\sqrt{\gamma}\,\alpha T^\mu{}_\nu \nabla_\mu \bar{\zeta}^\nu_{\rm P} \tag{4.3.26}$$

$$+ \frac{\partial}{\partial t}\int_\Sigma d^3x\,\sqrt{\gamma}\,n_\mu \bar{J}^\mu_{\rm P}, \tag{4.3.27}$$

which, once the stationary state is reached and the final term on the last line reduces to zero, will give to the following expression for the BI force:

$$\bar{\mathcal{F}}^{\bar{x}} = \int_{\Sigma} d^3x \, \sqrt{\gamma} \, \alpha T^{\mu}_{\ \nu} \nabla_{\mu} \bar{\zeta}^{\nu}_{\mathrm{p}} - \int_{\partial\Sigma} d^2x \, \sqrt{\sigma} \, (b_{\mu} \bar{J}^{\mu}_{\mathrm{P}}). \tag{4.3.28}$$

We see that the force is given by a source term combined with a correction that accounts for the difference between the temporal and spatial projection directions as well as the disagreement in ADM decomposition between BI and BKS coordinates.

We could once again try to make the distinction between the force due to accretion and that due to DF. However, because $\bar{\mathcal{F}}^{\bar{x}}_{\mathrm{d}}$ corresponds to a non-horizon-penetrating gauge, the observed net flux over the horizon will be null and $\bar{\mathcal{F}}^{\bar{x}}_{\mathrm{d}}$ is identical to the DF force.

4.4 Numerical Framework

In this section we describe our numerical implementation of the above framework for computing the BI DF force using BKS coordinates. We use the code `GRDZhadzha` as the basis for our simulations [116]. `GRDZhadzha` inherits its AMR and integration classes from `GRChombo`, but provides analytic expressions for the metric quantities instead of solving the EFEs. As mentioned previously, it was originally developed to compute the DF force using simulations performed in isotropic coordinates. In this work, we have developed an additional metric background class for working in BKS coordinates and provide a modified diagnostic class for verifying the continuity equations eq. (4.3.19) and eq. (4.3.21).

4.4.1 Diagnostic Implementation

We perform the numerical evolution of the scalar field entirely within BKS coordinates. The diagnostics and friction are not required to preform this evolution so we only compute these quantities periodically, with an interval of 100 coarse evolution steps. At each diagnostic step, we calculate the left and right hand sides of

eq. (4.3.19) and eq. (4.3.21), modifying Σ to exclude the inner excision. Explicitly, we can write these quasi energy and momentum constraints as:

$$\frac{\partial}{\partial t}\int_{\Sigma} d^3x\,\sqrt{\gamma}\,n_\mu \bar{J}^\mu_{\mathrm{E}} = -\int_{\partial\Sigma_{\mathrm{out}}} d^2x\,\sqrt{\sigma}\,\alpha N_\mu \bar{J}^\mu_{\mathrm{E}} + \int_{\partial\Sigma_{\mathrm{in}}} d^2x\,\sqrt{\sigma}\,\alpha N_\mu \bar{J}^\mu_{\mathrm{E}}; \tag{4.4.1}$$

and,

$$\begin{aligned}\frac{\partial}{\partial t}\int_{\Sigma} d^3x\,\sqrt{\gamma}\,n_\mu \bar{J}^\mu_{\mathrm{P}} = -\int_{\partial\Sigma_{\mathrm{out}}} d^2x\,\sqrt{\sigma}\,\alpha N_\mu \bar{J}^\mu_{\mathrm{P}} + \int_{\partial\Sigma_{\mathrm{in}}} d^2x\,\sqrt{\sigma}\,\alpha N_\mu \bar{J}^\mu_{\mathrm{P}} + \\ \int_{\Sigma} d^3x\,\sqrt{\gamma}\,\alpha T^\mu{}_\nu \nabla_\mu \bar{\zeta}^\nu_{\mathrm{P}},\end{aligned} \tag{4.4.2}$$

respectively, where $\partial\Sigma_{\mathrm{in}}$ and $\partial\Sigma_{\mathrm{out}}$ refer to the inner and outer boundaries of Σ. During a given time evolution step, the simulation will update the ADM projections (ρ, S_i, S_{ij} and S) of the stress-energy tensor (see section 3.3). The first step in computing the diagnostics is to rewrite the ADM variables as covariant quantities. The full 4d stress-energy tensor can be written:

$$T^{\mu\nu} = \rho n^\mu n^\nu + S^\mu n^\nu + S^\nu n^\mu + S^{\mu\nu} \tag{4.4.3}$$

In addition to $T^{\mu\nu}$, we also require the covariant metric in order to compute the contractions and the covariant derivatives present in eq. (4.4.1) and eq. (4.4.2). Whilst we could reconstruct the metric and its derivatives directly in terms of the α, β and $\gamma_{\mu\nu}$, the computation and implementation of the Christoffel symbols could be a significant source of bugs[1]. We thus opted to simply determine the 4-metric from analytic expressions and determine its derivatives and associated Christoffel symbols numerically at run time. This approach is not ideal as it involves multiple computations of the same metric quantities; once in the $3+1$ decomposition, and multiple times in covariant form. At the time of writing this thesis, we are currently in the process of implementing a fully ADM based formulation that we will later verify

[1]The analytical expressions for the Christoffel symbols are unwieldy and difficult to implement accurately.

against our covariant approach. Presently, we determine the Christoffel symbols according to:

$$\Gamma^{\lambda}_{\mu\nu} = \frac{1}{2} g^{\lambda\sigma} \left(\partial_{\mu} g_{\nu\sigma} + \partial_{\nu} g_{\mu\sigma} - \partial_{\sigma} g_{\mu\nu} \right), \tag{4.4.4}$$

with the metric derivatives being computed using finite difference:

$$\partial_{\mu} g^{\nu\sigma} \approx \frac{g^{\nu\sigma}(x^{\rho} + dx^{\rho}) - g^{\nu\sigma}(x^{\rho} - dx^{\rho})}{2dx^{\rho}}. \tag{4.4.5}$$

One could, in principle store the full 4-metric as a variable on the grid and perform the differentiation using the difference in $g^{\mu\nu}$ between adjacent cells. This would, however, incur a significant memory overhead in storing the 10 independent metric components and increase the number of memory copy calls required. To avoid this, we instead compute $g^{\mu\nu}$ twice for each grid cell to compute eq. (4.4.5). Using this method, we are also free to independently specify the fidelity of the derivative step dx^{μ}. During testing, we found that values of $dx^{\mu} < 10^{-}6$ lead to significant numerical errors due to noise, whilst values significantly larger than this fail to resolve the derivatives to a sufficient accuracy.

With the BKS quantities expressed in covariant form, it remains for us to determine the covariant energy and momentum projection operators $\bar{\zeta}^{\mu}_{\mathrm{E}}$ and $\bar{\zeta}^{\mu}_{\mathrm{P}}$ that define the currents $\bar{J}^{\mu}_{\mathrm{E}}$ and $\bar{J}^{\mu}_{\mathrm{P}}$. We achieve this via contraction with the BKS to BI Jacobian ($\mathcal{J}^{\mu}_{\bar{\nu}}$), which can be written analytically in terms of the following three partial Jacobians:

$$(\mathcal{J}_{\mathrm{Boost}})^{\mu}_{\nu} = \begin{pmatrix} \frac{1}{\gamma} & -\gamma v & 0 & 0 \\ 0 & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \tag{4.4.6}$$

$$(\mathcal{J}_{\rm sph})^{\mu}_{\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{x}{\sqrt{x^2+y^2+z^2}} & \frac{xz}{\sqrt{\frac{x^2+y^2}{x^2+y^2+z^2}}\sqrt{x^2+y^2+z^2}} & -y \\ 0 & \frac{y}{\sqrt{x^2+y^2+z^2}} & \frac{yz}{\sqrt{\frac{x^2+y^2}{x^2+y^2+z^2}}\sqrt{x^2+y^2+z^2}} & x \\ 0 & \frac{z}{\sqrt{x^2+y^2+z^2}} & \frac{-x^2-y^2}{\sqrt{\frac{x^2+y^2}{x^2+y^2+z^2}}\sqrt{x^2+y^2+z^2}} & 0 \end{pmatrix}; \tag{4.4.7}$$

$$(\mathcal{J}_{\rm I\,\to KS})^{\mu}_{\nu} = \begin{pmatrix} 1 & -\frac{2m(m+2r_{\rm Iso})}{r_{\rm Iso}(m-2r_{\rm Iso})} & 0 & 0 \\ 0 & 1-\frac{m^2}{4r_{\rm Iso}^2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{4.4.8}$$

where, acting on contravectors e.g. ζ^{μ}, $(\mathcal{J}_{\rm Boost})^{\mu}_{\nu}$ transforms from an un-boosted frame to one that is boosted and shifted. $(\mathcal{J}_{\rm sph})^{\mu}_{\nu}$ maps from a polar basis to a Cartesian one and $(\mathcal{J}_{\rm I\,\to KS})^{\mu}_{\nu}$ transforms from un-boosted polar isotropic coordinates to un-boosted KS coordinates and is parametrised in terms of the former. Overall the full transformation from BKS to BI can be written as the following matrix product:

$$\mathcal{J}^{\mu}_{\bar{\nu}} = (\mathcal{J}^{-1}_{\rm Boost})^{\mu}_{\delta}(\mathcal{J}_{\rm sph}(x_{\rm KS})^{\delta}_{\kappa}(\mathcal{J}_{\rm I\,\to KS}(x_{\rm I}))^{\kappa}_{\iota}(\mathcal{J}^{-1}_{\rm sph}(x_{\rm I}))^{\iota}_{\rho}(\mathcal{J}^{-1}_{\rm Boost})^{\rho}_{\sigma}(\mathcal{J}_{\rm sph}(\bar{x}))^{\sigma}_{\bar{\nu}}, \tag{4.4.9}$$

where $x_{\rm I}$ and $x_{\rm KS}$ are un-boosted isotropic and KS coordinates respectively.

With these quantities in hand, we are now able to evaluate the diagnostic spatial and volume integrals to assess the validity of a given simulation.

4.4.2 Setup and Parameters

We describe now, with regard to their justification in terms of numerical stability and comparability, our chosen simulation parameters and integration schemes.

Using our modified version of `GRDZhadzha`, we evolve the system on an adaptively refined mesh-grid. We use the pre-existing integration scheme to solve the scalar

equations of motion eq. (3.3.24) and eq. (3.3.22). In this approach, the PDEs are reduced to ODEs using the *method of lines*, whereby the spatial partial derivatives are computed using finite difference on the grid and time integration is performed using the RK4 algorithm. Following [4], we make use of the quadrant symmetry of the problem and perform the simulation within a cuboidal spatial domain of volume $\Delta x \times \Delta y \times \Delta z = L \times L/2 \times L/2$, with the BH centred on $(\Delta x/2, 0, 0)$. The diagnostics and frictional forces are then multiplied by a factor of 4 to obtain those which will be measured over the full volume. For comparative reasons, we select the full domain length L according to the values used by the authors of [4]. We discretise the domain onto a coarse grid of dimensions $128 \times 64 \times 64$ and specify a maximum of 7 $(2:1)$ refinement iterations. Between each spatial sub-division, we additionally reduce the time step by a factor of 10. This improves efficiency, allowing larger time steps to be used by coarser levels. The coarsest time step is chosen according to the condition $dt_{\text{coarse}} \leq T/32$, where T is the characteristic oscillation period of the scalar field, $T = 2\pi/\mu$ [4]. The discrete nature of this grid based approach allows for optimised load distribution through the *Message Passing interface*[1] (MPI) [117]. Using a single 64 core, 2.6GHz CPU, a simulation run up to a time of 5000 coarse steps takes $\mathcal{O}(1\text{h})$ to complete.

As outlined previously, the convergence of the diagnostic quantities can be affected by the choice of the inner excision radius. This relationship is depicted in fig. 4.6. Sufficient numerical stability is observed for radii greater than 3.5 and a value of $r_{\text{in}} = 4.083$ was selected to produce the data described below.

4.5 Preliminary Results

At present our code is able to perform a full evolution within the BKS basis and compute the quasi BI diagnostic quantities that constitute eq. (4.4.1) and eq. (4.4.2)

[1] MPI is a framework for distributed memory, parallel computation and is used extensively in high performance computing clusters.

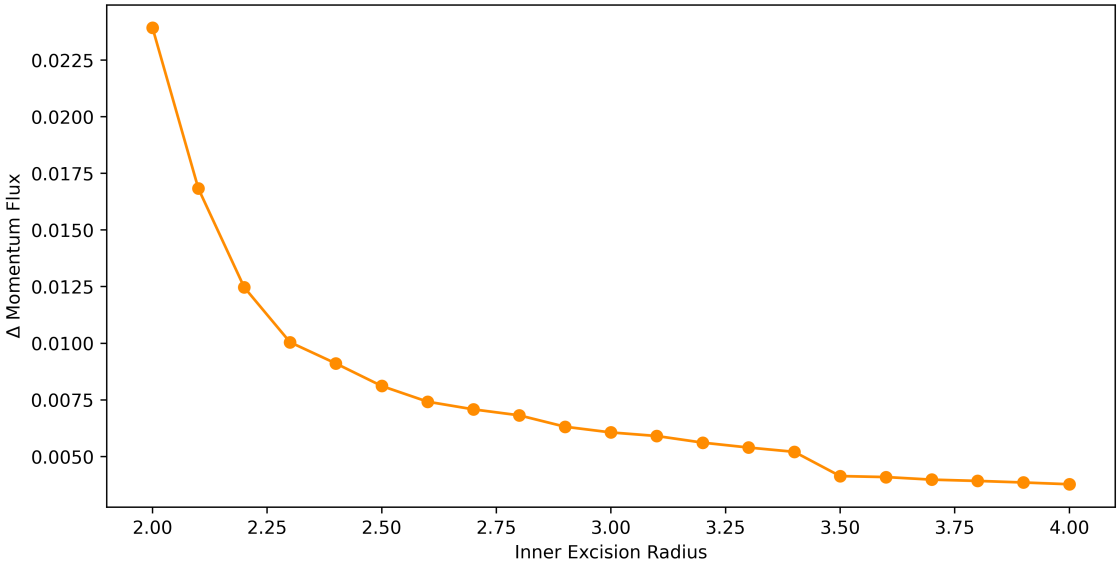


Figure 4.6: Fractional discrepancy in between the left and right hand sides of eq. (4.4.2) against inner excision radius. For radii less than 3.5 the flux through the inner excision cannot be resolved numerically and is neglected in the computation of the momentum discrepancy. The effect of this can be seen in the discontinuity at this radius. The data used to generate this figure were generated using $v = 0.2$ and $\alpha = 0.05$ and correspond to an evolution time of $t = 163.2$.

and we are currently implementing the full BI DF calculation. To verify the self-consistency of our code, we performed two simulation runs up to a time $t = 5000$ with a BH of mass $M = 1$ and scalar field with amplitude $\alpha = 0.05$; for a scalar field of mass $\mu = 1 \times 10^{-12}$eV, this corresponds to a BH with a physical mass of $M_{\rm phys} = 6.6 M_\odot$ (where $M_\odot$ is the mass of the sun). In the first run, we took the boost velocity to be $v = 0$, and in the second, this was chosen to be $v = 0.2$. The simulations were carried out on within a box of length $L = 4096$ with 128 coarse grid cells per dimension and 7 grid refinement levels. The effect of the boost on the scalar field can be seen in its energy density distribution. The energy density in the x-y plane is depicted in fig. 4.7 and fig. 4.8 for zero velocity and boosted cases respectively[1]. In the zero velocity case, the energy density can be seen to build up around the BH as matter is accreted, whereas in the boosted case, an over density in the scalar field accumulates behind the BH. The extent of this wake increases over time, eventually reaching the outer excision radius. Once the wake size saturates, the internal energy and momentum densities will enter a steady state. At this point the total energy within the excision region will saturate to a constant value. We plot the total energy in the right hand plot of fig. 4.9. At late times, the energy approaches a constant value.

The total energy depicted in fig. 4.9 has been calculated in two ways. These correspond to the left-hand side (solid lines) and right-hand side (dashed lines) of eq. (4.4.1) after time integration. We also show the flux, directly, in the left plot of eq. (4.4.1). The agreement between the dashed and solid lines for both $v = 0$ and $v = 0.2$, demonstrates self consistency in our implementation of the framework discussed in section 4.3.3. We also demonstrate momentum conservation according to eq. (4.4.2), both in the x (fig. 4.10) and radial (fig. 4.11) directions.

[1]Energy density and diagnostic figures were produced using excision radii of $r_{\rm in} = 4.083$ and $r_{\rm out} = 900$ and a simulation volume of $4096 \times 2048 \times 2048$. The lower half of the density plots were not the result of direct simulation but were inferred from the upper half using the $y \to -y$ symmetry of the problem.

We conclude that our implementation is able to consistently evolve a system entirely within a BKS gauge whilst computing BI observables. We additionally note that with our chosen simulation parameters, the system is well posed, with the diagnostics not displaying any significant deviation, even at late times. On the back of this, we are now working to implement the DF calculation of eq. (4.3.28).

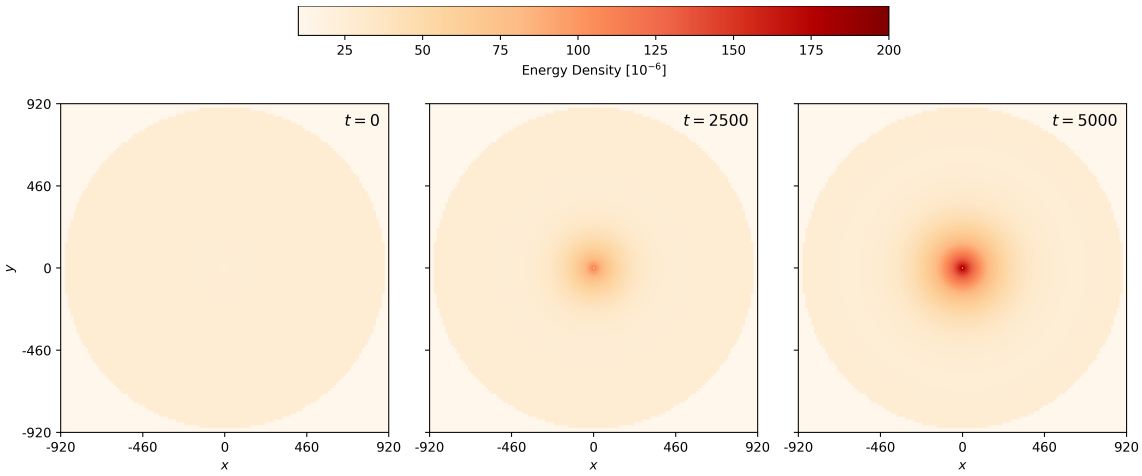


Figure 4.7: The evolution of the energy density in the x-y plane around a non-moving BH. The BH is centred at $(x = 0, y = 0)$, and the visible circle marks the outer excision radius. Over time, the energy density grows around the BH as matter is accreted into its vicinity.

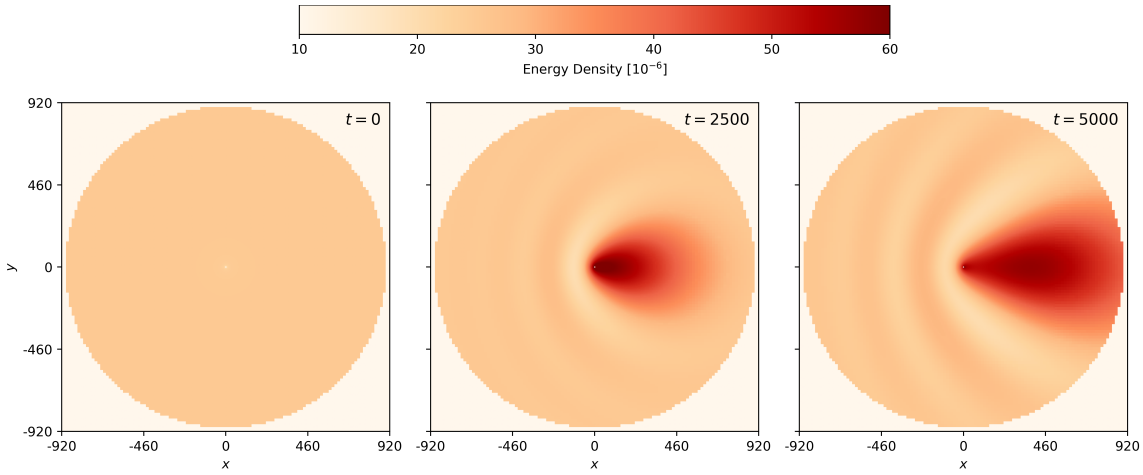


Figure 4.8: The evolution of the energy density in the x-y plane around BH centred at $(x = 0, y = 0)$ and boosted in the negative x direction. The over-density wake grows over time until the influence of the BH reaches the outer excision ($r_{out} = 900$). At late times, the energy density begins to enter a steady state within the finite excision region.

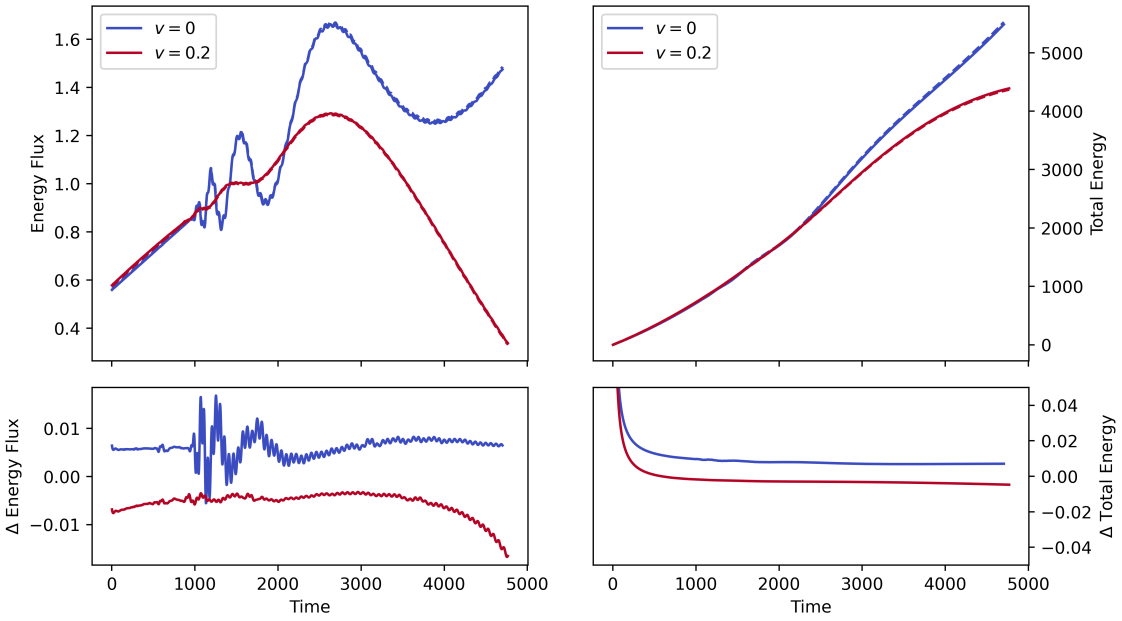


Figure 4.9: Energy flux conservation (left) and total energy conservation (right) for $v = 0$ (blue) and $v = 0.2$ (red). The solid and dashed lines correspond to the left and right hand sides of eq. (4.4.1) respectively, with these being integrated over t in the right plot. The fractional difference between the dashed and solid lines is shown below. Agreement between these lines indicates convergence of the simulations. The decrease in energy flux at $t = 3000$ occurs as the wake reaches the outer excision radius leading to an increase in flux out of this boundary.

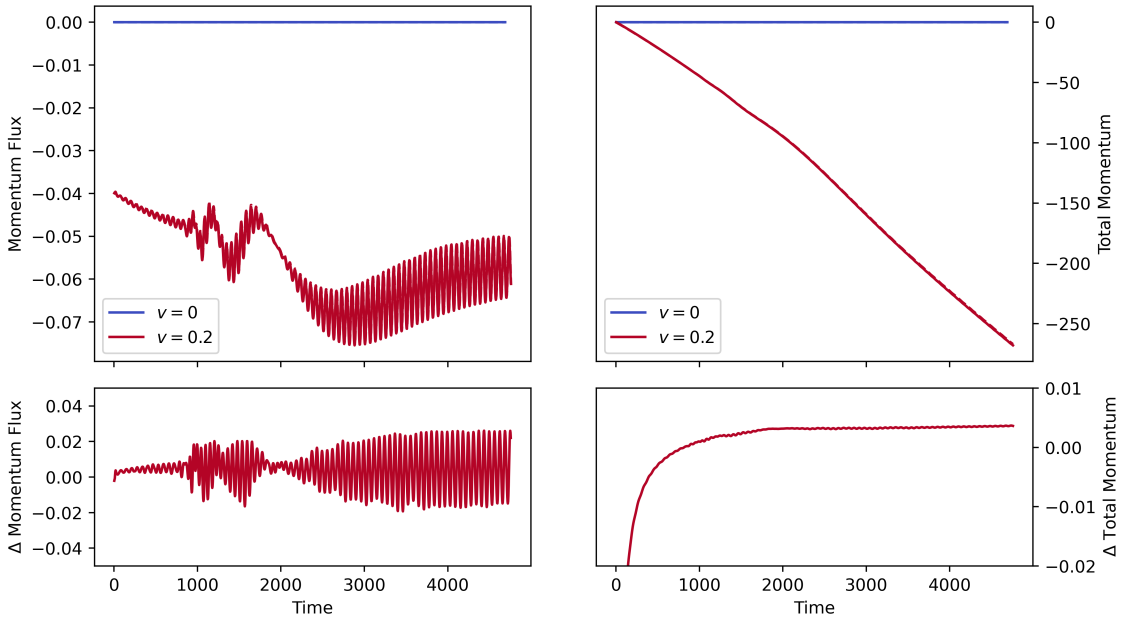


Figure 4.10: Conservation of momentum flux in the x direction (left) and total momentum (right) for $v = 0$ (blue) and $v = 0.2$ (red). The solid and dashed lines correspond to the left and right hand sides of eq. (4.4.2) respectively, with these being integrated over t in the right plot. As in fig. 4.9, the fractional discrepancy between dashed and solid lines is shown below and indicates a high degree of momentum conservation. In the $v = 0$ case, the flux is negligible as the ingoing flux in the x direction cancels on either side of $x = 0$. In the boosted case the momentum grows as it forms a wake moving in the direction of $-x$.

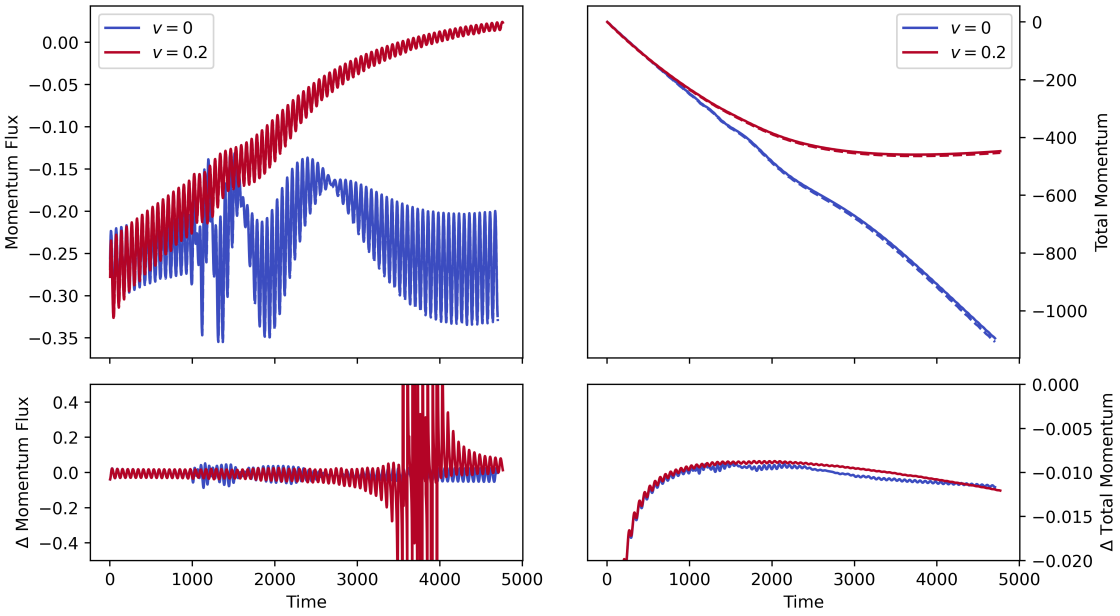


Figure 4.11: Radial momentum conservation. As in fig. 4.10, flux is depicted in the left plot with total momentum on the right. The solid and dashed lines again correspond to the left and right hand sides of eq. (4.4.2) with their fractional disagreement depicted below. In this case, however, the projection vector $\bar{\zeta}^{\mu}_{\mathrm{P}} = (t = 0, r = 1, \theta = 0, \phi = 0)$ has been selected to point in the outwards radial direction. In the un-boosted case, the radial momentum grows linearly as matter accelerates towards the BH isotropically. In the boosted case, however, the momentum growth decreases at late times as the wake of fast moving particles begins to exit through the excision boundary.

4.6 Outlook

In this chapter, we have demonstrated the viability, with respect to energy and momentum conservation, of our horizon-penetrating implementation of scalar field evolution around BHs. We are currently refactoring our code to be more in keeping with the existing code base of `GRDZhadzha` and to minimise redundant metric and stress-energy calculations. Once this is complete, we expect to be able to reproduce the DF presented in [4]. The use of horizon-penetration will effectively result in greater temporal resolution around the horizon. The temporal resolution is inversely

proportional to the grid refinement level, therefore convergence of the diagnostics can be achieved with lower grid resolutions than with non-penetrating gauges. This will reduce the overall execution time, potentially allowing for the investigation of more complicated problems.

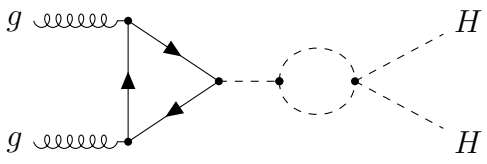
A further benefit of the KS gauge is its applicability to rotating Kerr backgrounds. One of the core motivations for the study of bosonic scalars around BHs is superradiance. Superradiance is fundamentally an effect of rotation. In realistic scenarios, therefore, we expect phenomena such as frame dragging to have an important affect on the observed DF force. At present we take the spin of the BH to be zero, however, with minimal modifications to the metric, this could be included in the simulation background.

Overall, this work may give new insights into the the behaviour of BHs within their surroundings, and potentially aid in the calculation of gravitational wave signals generated by inspiraling BHs. This may allowing for the identification of new fundamental fields that have thus far evaded detection.

Chapter 5

Precision Calculations

The Large Hadron Collider (LHC) located at CERN is currently the foremost terrestrial particle detection experiment. Since its creation, in addition to the successful search for the Higgs Boson completed in 2012, a key science driver for the LHC has been the search for physics Beyond the SM (BSM) [118–125]. To correctly identify an effect as belonging to new physics, it is necessary to determine theoretically the phenomenological implementations of the SM to a high degree of accuracy. With the LHC now undergoing significant upgrades that promise an increase in peak luminosity from $1 \times 10^{34}\text{cm}^{-2}\text{s}^{-2}$ to $7.18 \times 10^{34}\text{cm}^{-2}\text{s}^{-2}$ [126], the precision requirements on theoretical calculations are now greater than ever. For example, many theories of BSM physics predict a non-standard Higgs boson self coupling [127]. To resolve this against the analytical uncertainty associated with perturbative physics, it is necessary to compute cross-sections such as that of 2 Gluon to 2 Higgs ($gg \to HH$) scattering at up to Next to Next to Leading Order (N^2LO) [128]. This involves the evaluation Feynman diagrams containing many loops and legs, such as the following:



(5.0.1)

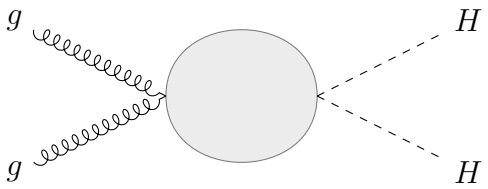
The computation of these high-order corrections is extremely computationally expensive. The most advanced techniques involve identifying *families* of integrals and relating these to a set of *master integrals* whose evaluation is more tractable; this approach is known as the Laporta algorithm [129]. One of the core bottlenecks with this method is in the decomposition of an integral into the master basis. This is usually is achieved through the use of Integration By Parts (IBP) relations and Gaussian elimination. It is this elimination step wherein the problem lies. Sophisticated modern techniques aim to reduce the time complexity of this problem by decreasing the required number of Gaussian elimination procedures through the use of multivariate rational function interpolation. At present, one of the leading tool-chain for this consists of the following triad of software: `Kira` [130, 131], `Ratracer` [132] and `FireFly` [133, 134]. We will explore the use case of these codes in the following subsections, but note that `FireFly` is used to perform the interpolation step.

`FireFly`, whilst benefiting from multithreading, is fundamentally a CPU bound application. The advent of modern GPU hardware offers the possibility of an increase in operational throughput of the order of 100 times that of high end CPUs. In the following chapter, we discuss how this can be utilised to good effect in the context of precision phenomenology. We present a novel, proof of concept code that is capable of performing multivariate polynomial interpolation entirely on a GPU. We further demonstrate that its efficiency surpasses that of `FireFly` for many real world polynomials. Before addressing this, however, we provide an overview of the Laporta algorithm.

5.1 Laporta Algorithm

The computation of any Feynman diagrams with greater complexity than tree level will involve the evaluation of loop integrals. The Laporta algorithm provides a systematic procedure with which one may deconstruct and simplify these integrals to the point where they are numerically soluble. In this section, we overview the core

facets of this algorithm and identify the motivation for the work in this chapter. To aid in this discussion, we find it useful to consider a specific example of $gg \to HH$ scattering:



(5.1.1)

Computing this element up to $\mathrm{N^2LO}$ will involve the evaluation of over 100 integrals, please see ref. [128] for a detailed analysis. These integrals will, in general, possess a rank-2 tensor structure associated with the 2 external gluon legs. The first step in their evaluation is to extract this tensor dependence through the process of *Tensor Reduction*.

5.1.1 Tensor Reduction

To simplify the integrals, we wish to extract their tensor components and write them, instead, as a superposition of tensor basis elements. For the example of $g^a_\mu(p_1)g^b_\nu(p_2) \to H(-p_3)H(-p_4)$ scattering, a general matrix element can be written [128]:

$$M_{ab} = \varepsilon_{1,\mu}\varepsilon_{2,\nu}M^{\mu\nu}_{ab} = \varepsilon_{1,\mu}\varepsilon_{2,\nu}\delta_{ab}\left(F_1 T_1^{\mu\nu} + F_2 T_2^{\mu\nu}\right), \tag{5.1.2}$$

where $\varepsilon_{i,\mu}$ are gluon polarisations and $\{a, b\}$ are colour indices. Here the F_i are known as *form factors* and contain the integral structure and the $T_i^{\mu\nu}$ form the tensor basis and are functions of the external momenta. The tensor basis vectors can be constructed as combinations of the momentum vectors and the metric tensor, for example:

$$T_1^{\mu\nu} = \eta^{\mu\nu} - \frac{p_2^\mu p_1^\nu}{p_1 \cdot p_2}; \tag{5.1.3}$$

$$T_2^{\mu\nu} = \eta^{\mu\nu} + \frac{1}{(p_T)^2(p_1 \cdot p_2)} \left[m_H^2 p_2^\mu p_1^\nu - 2(p_1 \cdot p_3) p_2^\mu p_3^\nu - 2(p_2 \cdot p_3) p_3^\mu p_1^\nu + 2(p_1 \cdot p_2) p_3^\mu p_3^\nu \right], \tag{5.1.4}$$

where m_H is the Higgs mass and:

$$p_T^2 = \frac{(p_2 + p_3)^2 (p_1 + p_3)^2 - m_H^4}{(p_1 + p_2)^2}. \tag{5.1.5}$$

For any given choice of tensor basis, one may determine the form factors by projecting the matrix element onto linear combinations ($P_{I,ab}^{\mu\nu}$) of $T_1^{\mu\nu}$ and $T_2^{\mu\nu}$, i.e.:

$$F_i = P_{i,ab}^{\mu\nu} M_{\mu\nu}^{ab}. \tag{5.1.6}$$

Each form factor generated in this way will comprise a sum of scalar integrals. These loop integrals will typically take the form:

$$I = \int \left(\prod_j \frac{d^d \ell^j}{(2\pi)^d} \right) \frac{N(\{\ell^j\})}{D_1 D_2 \cdots D_n}, \tag{5.1.7}$$

where $N(\{\ell^j\})$ is a function of the loop momenta $\{\ell^j\}$, the D_i are propagator factors and d is the space-time dimension. Note that d is left as a variable here to account for divergences through *dimensional regularisation* and *renormalisation* [6]. The number of propagator factors will depend on the number of internal lines associated with a given loop. For example, a Higgs boson loop with internal momentum ℓ containing two vertices with external momenta p would be associated with two propagators of the form:

$$D_1 = \ell^2 - m_H^2; \tag{5.1.8}$$

$$D_2 = (\ell + p)^2 - m_H^2. \tag{5.1.9}$$

These same propagators may also arise from diagrams with alternative loop structures. Integrals with the same propagator structure are said to belong to a *family*. It is possible to rewrite integrals within a given family as a linear superposition of other members of that family. This is achieved through Integration By Parts (IBP) relations and allows a large set of complicated integrals to be computed from a subset of simpler ones.

5.1.2 Integration By Parts Relations

IBP relations arise as a consequence of Gauss's theorem within dimensional regularisation, which implies that integrals of total derivatives vanish. Crucially, this means that:

$$\int d^D\ell \, \frac{\partial}{\partial \ell^\mu} \left(v^\mu \frac{1}{D_1^{a_1} D_2^{a_2} \dots D_n^{a_n}} \right) = 0, \tag{5.1.10}$$

where v^μ can refer to either loop or external momenta and α_i are powers of the propagators. Expanding the derivative in eq. (5.1.10) and integrating by parts will yield expressions of the form:

$$I_{ij\dots k} = \sum_l c_l I_{i'(l)j'(l)\dots k'(l)}. \tag{5.1.11}$$

Where the integral subscripts $\{i, j, k, i', j', k'\}$ denote the exponents of the propagators pertaining to the family and the c_l are expansion coefficients. Note that the numerator present in eq. (5.1.7) and those which arise during the IBP procedure can be expressed as negative powers of the denominators. In this way, any given loop integral can be written in the form $I_{ij\dots k}$, which itself can be expressed as a superposition of simpler integrals. The integral elements that constitute a minimum spanning set of the family, are known as master integrals. Once all members of the family are expressed in this basis, the master integrals can be evaluated using numerical methods such as sector decomposition. This final evaluation step is beyond the scope of this work. Please see Ref. [135] for more information.

5.1.3 Gaussian Elimination

The first step in reduction to the master basis ($\{M_n\} \subset \{I_n\}$) is to express each Feynman integral as a linear combination of all family members ($\{I_n\}$). In the language of linear algebra, the IBP relations can be written:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} I_1 = M_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \tag{5.1.12}$$

where, for brevity we drop the propagator subscripts and simply denote each integral using a single number. In this example, each of the 3 rows corresponds to a unique IBP relation, and I_1 is assumed to be the master integral in terms of which I_2 and I_3 may be written. In this way, the problem has been reduced to a matrix equation, solvable via the standard techniques of linear algebra. To achieve the reduction to the masters, one need only to perform sufficient Gaussian elimination steps. Whilst this procedure does have very poor time complexity scaling ($\mathcal{O}(N^3)$, for a matrix of size N), its execution time and memory requirements even for the large $\sim 100 \times 100$ matrices that we expect to encounter will be feasibly small. The problem arises due to the fact that the coefficients a_{ij} are complicated analytical function of the external momenta. Modern software such as `Kira` can perform algebraic Gaussian elimination. For small problems this is a simple task. For modern problems, however, hardware limitations make this process highly non-trivial. Specifically, when working purely algebraically, the expressions resulting from Elementary Row Operations (EROs) swell rapidly and can quickly exceed the available computational memory. For large problems, it is instead necessary to perform the elimination steps numerically. This is done by substituting test momenta into the algebraic parameters.

This process necessarily leads to a large amount of information loss. This can be circumvented by using interpolation to build an analytical expression from a set of numerical results. This approach avoids memory swell but requires a large number of elimination procedures, in effect trading the memory bottleneck for a temporal one.

In modern versions of `Kira`, this issue can be addressed using `Ratracer`. `Ratracer`, deriving its name from rational tracer, traces the ERO steps taken by `Kira` for a single numerical substitution and records them to a trace file. This prevents the need for repeated Gaussian eliminations, with the trace file effectively encoding the master integral coefficients, albeit in an very unsimplified form.

The traces generated by `Ratracer` take the form of rational functions and can be simplified using interpolation. Interpolation schemes assume that their target (or black-box) function (a trace in this case) takes a specific form and use repeated numerical evaluations to determine the coefficients through which this can be parametrised. It is because interpolation reconstructs into a chosen form e.g. that of a canonical polynomial:

$$f(x) = a_0 + a_1 x + ... + a_N x^N, \tag{5.1.13}$$

that the result will automatically be in its most simplified state. This approach can be vastly more memory efficient and faster than algebraic simplification algorithms.

The current state of the art tool for this purpose is `FireFly`. `FireFly` is a multi-threaded CPU application that supports MPI. In spite of this, it represents one of the largest bottlenecks in the computation of high order Feynman integrals. During profiling we found that, for physically relevant scenarios, the numerical evaluations of the black-box functions were responsible for the majority of the interpolation time. These black-box evaluations, however, can be computed entirely in parallel and so motivate the use of GPU hardware.

5.2 GPU Interpolation

As identified above, it is the evaluation of the black-box probes that consumes the majority of the execution time for a given interpolation. By refactoring this aspect of the Laporta algorithm for execution on a GPU, one would naively expect a significant

time saving. The most straightforward approach to accelerating `FireFly` would be to simply move the black-box evaluations onto the GPU, whilst continuing to use `FireFly`'s pre-existing CPU based interpolator. The inter-device communication calls that would be required for this are significantly slower than access speeds of both the host main memory and the GPU global memory and would, therefore, incur an unnecessarily large overhead. We instead choose to perform the full interpolation procedure on the GPU.

In the remainder of this chapter we describe the algorithmic choices and optimisation schemes with which our GPU based interpolator is constructed; for an overview of the key hardware differences between GPU and CPU devices please see appendix D. We begin with a description of the core procedures below, saving the details of the implementation until the following section. We note that at present our code does not support direct rational function reconstruction as is required for use with `Kira`. It instead acts as a proof of concept, able to solve the precursor problem of multivariate polynomial interpolation. For the remainder of this work, therefore, we consider the simplified problem in which the `Ratracer` trace files contain only multivariate polynomials.

5.2.1 Finite Fields

When working with numerics, the limitations of the finite precision of computational data types are ever present. Whilst high precision floating point types such as doubles can represent extremely large and small numbers, the possibility of information loss due to their finite mantissa invalidates their use in high precision interpolation. Similarly, integer types also lack the required range to express the interpolation coefficients. A common solution to this issue it to use finite (or *Galois*) fields. A Galois field (GF) labelled by a prime P is a set of positive integers $\mathrm{GF}(P) = \mathbb{Z}$ mod P. The use of such fields prevents the risk of integer overflow while maintaining the precision that would be lost with floating point types. Until recently NVIDIA

GPUs did not provide hardware support for modular arithmetic and, although software emulation of these operations is available, this would incur a significant overhead.

The size of the Galois field also plays a significant role in the efficiency of the interpolation. `Firefly` uses a 64 bit unsigned integer type with $P < 2^{64}$. For function coefficients larger than 2^{64}, the results from multiple interpolation over different fields must be combined to reproduce the full coefficient. This can be achieved using the *Chinese Remainder Theorem* (CRT). Current NVIDIA GPUs do provide support for 64 bit unsigned integers, albeit with less hardware optimization when compared with the 32 bit variant. Performing the CRT on 32 bit types required twice the number of evaluations when compared to their 64 bit counterparts. In this work we opt to use the 32 bit type for the interpolations and perform the CRT reconstruction on the CPU using an arbitrary precision type provided by the *GNU Multiple Precision arithmetic library* (`GMP`).

5.2.2 Parsing

We now have a way to represent numbers on the GPU, but still require a method by which to represent the trace files in a machine readable format. The trace files produced by `Ratracer` are often in excess of 50GB in size for modern problems. This may constitute a significant fraction of the available GPU memory, leaving insufficient memory for the interpolation process. To alleviate this issue, we begin by manually subdividing the input trace file, processing only a single trace $<$ 1GB per run. This is in contrast to `FireFly` that can handle the full system simultaneously.

The human readable trace functions generated by `Ratracer` are not readily comprehensible to the GPU. Ideally, one would represent each trace in a binary format that could be executed directly. To this end, we have investigated the use of Just In Time (JIT) compilation. The `CUDA` compiler (nvcc) is capable of performing significant optimisations such as removing redundant operations and its usage allows

for minimal execution times and overall memory footprints. For large polynomials, however, we find that the overhead due to the compiler runtime exceeds any possible time savings afforded by this strategy. We instead choose to parse each trace as a string and use a FOR loop to iterate through each set of operands and operations.

To reduce the size of this string, we first re-express the chosen trace function in *Reverse Polish Notation* (RPN), where sets of operands are written preceding their operation. This removes the need for parentheses. For example consider the following simple expression:

$x(x+3)$	Infix;
$x3+x\times x$	RPN.

We store the RPN expression associated with a given trace as a character array with each character (or token) representing either an operator or an operand. During each trace evaluation, the variable symbols are substituted for numerical values using a lookup.

The standard method for parsing an expression in RPN relies on the use of a *stack* type memory structure to store intermediate values. Stacks follow a first-in-last-out storage paradigm wherein the length of the stack may grow or shrink dynamically to accommodate variations in number of elements. `CUDA`, unfortunately, does not support device side dynamic memory allocation as would be required for this. It is possible, however, to predetermine the maximum stack depth required to parse an RPN expression. This can be achieved using the algorithm 1.

Using this, a statically sized stack can be constructed by allocating a block of GPU global memory of size `max_depth` $\cdot$ `n_threads` $\cdot$ SIZE_OF(`Field Type`), where the value of `n_threads` determines the number of black-box probes that may be computed simultaneously. For sufficiently large polynomials, the available global

Algorithm 1 Compute the Maximum Stack Depth of an Expression

```
1: Input: A vector of strings representing the expression, expression
2: Output: The maximum stack depth, max_depth
3: stack_depth ← 0
4: max_depth ← 0
5: for all token in expression do
6:     if IS_OPERATOR(token) then
7:         stack_depth ← stack_depth − 1
8:     else
9:         stack_depth ← stack_depth + 1
10:    max_depth ← MAX(max_depth, stack_depth)
11: return max_depth
```

memory will place a bound on this quantity. If the thread count is significantly reduced, however, the device utilization will suffer and performance will be impaired. One method to reduce the necessary stack depth would be to traverse the RPN expression to construct a tree-like structure. In this way, independent branches of the tree could be identified and evaluated individually. For an illustration of this see fig. 5.1. In this proof of concept work we do not implement this strategy, and instead identify it as possible addition for a future update.

5.2.3 Univariate Polynomial Interpolation

The simplest polynomials are those involving only a single variable. The interpolation of such a univariate polynomial of rank N involves the determination of a series of coefficients $\{a_i\}; i \leq N$:

$$P(x) = a_0 + a_1 x + ... + a_N x^N, \tag{5.2.1}$$

which satisfies $\{P(x_i) = y_i\}; i \leq N$, where, in our case, the x_i are randomly generated input probe values and the y_i are the results of corresponding trace function evaluations. Interpolation can be rephrased as a linear algebra problem of the form:

$$V \cdot \mathbf{a} = \mathbf{y}, \tag{5.2.2}$$

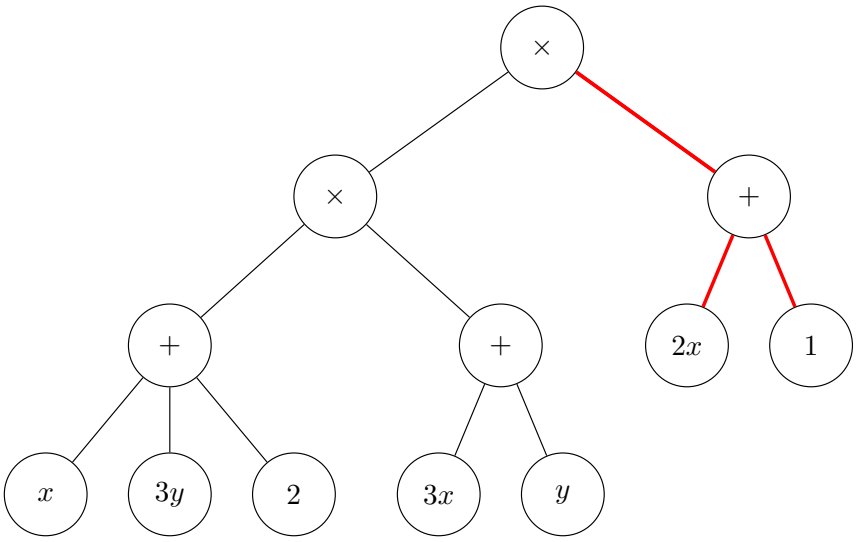


Figure 5.1: A binary tree in RPN for the expression $f(x,y) = ((x+3y+2)(3x+y))(2x+1)$. Representing a function in this way allows independent branches (such as is identified in red) to be identified and computed independently.

where V is known as the Vandermond matrix and can be written:

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^N \\ 1 & x_2 & x_2^2 & \cdots & x_2^N \\ 1 & x_3 & x_3^2 & \cdots & x_3^N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^N \end{bmatrix},$$

and the vectors $\mathbf{a}$ and $\mathbf{y}$ are given by:

$$\mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_N \end{bmatrix}.$$

Written in this form, the coefficients can be determined by simply inverting the Vandermond matrix. Whilst this seems trivial, and indeed is for low rank polynomials,

the poor time complexity scaling of inversion ($\mathcal{O}(N^3)$, for a matrix of size $N \times N$) means this direct approach is not feasible for large matrices.

Since the 17$^{\text{th}}$ century, many interpolation algorithms with greater efficiency than matrix inversion have been developed. Some of the earliest forms were developed by Newton and Lagrange in 1669 and 1795 respectively. Newton's interpolation is a sequential algorithm and involves the reconstruction of each coefficient in turn using the values of the previous coefficients. In contrast, in Lagrange interpolation, each coefficient can be computed independently. Both of these algorithms require $\mathcal{O}(N^2)$ operations to interpolate a polynomial of rank N. In the case of Lagrange interpolation, however, parallel implementations can allow for significantly reduced execution times.

The univariate Lagrange interpolation algorithm is as follows. Given a set of black-box probes $\{x_i\}; i \in [0, N]$ and their corresponding trace function evaluations $P(x_i)$, it is possible to reconstruct any rank N polynomial $P(x)$ as a superposition of rank N Lagrange polynomials $l_n(x)$:

$$P(x) = \sum_{n=0}^{N} \tilde{a}_n l_n(x), \tag{5.2.3}$$

where, the n^{th} Lagrange polynomial can be written:

$$l_n(x) = \frac{(x - x_0)}{(x_n - x_0)} \cdots \frac{(x - x_{n-1})}{(x_n - x_{n-1})} \frac{(x - x_{n+1})}{(x_n - x_{n+1})} \cdots \frac{(x - x_N)}{(x_n - x_N)} \tag{5.2.4}$$

$$= \prod_{\substack{j \in [0,N] \\ j \neq n}} \frac{x - x_j}{x_n - x_j}, \tag{5.2.5}$$

and the expansion coefficients $\tilde{a}_n$ are simply given by:

$$\tilde{a}_n = P(x_n). \tag{5.2.6}$$

Each term in eq. (5.2.3) can be computed independently from every other term, and

if one desired to express the interpolation result in the form eq. (5.2.3), then the whole procedure could be achieved in $\mathcal{O}(N)$ time with optimal parallelism.

5.2.4 Multivariate Polynomial Interpolation

In general, the traces produced by `Ratracer` will be functions of multiple variables associated with the various momenta of the problem. The interpolation of such multivariate polynomials can be achieved through repeated application of univariate algorithms. Consider a two dimensional polynomial of rank (N, M):

$$P(x, y) = \sum_{n<(N-1)} \sum_{m<(M-1)} a_{nm} x^n y^m. \tag{5.2.7}$$

This function can be reconstructed by first performing M univariate interpolations over x, using a different value for y for each interpolation. Each of these polynomials will require a set of N probe values (x_i) to compute. The result of this first stage will be a set of M independent one dimensional polynomials in x. The second dimension can then be reconstructed by performing N independent interpolations over y using the coefficients of the interpolations over x in place of the black-box evaluations used in the previous stage. Overall, the number of black-box probes required for this interpolation would be $N \cdot M$. In this simple 2d case this procedure can be viewed using matrices in place of the vectors in the univariate example above. The input to the system will be the set of random black-box probes $(\{(x_i, y_j)\} i \in [0, N-1], j \in [0, M-1])$. From these, a codomain matrix Y_x corresponding to $\mathbf{y}$ in eq. (5.2.2) can be determined:

$$Y_x = \begin{bmatrix} P(x_0, y_0) & P(x_1, y_0) & P(x_2, y_0) & \cdots & P(x_{N-1}, y_0) \\ P(x_0, y_1) & P(x_1, y_1) & P(x_2, y_1) & \cdots & P(x_{N-1}, y_1) \\ P(x_0, y_2) & P(x_1, y_2) & P(x_2, y_2) & \cdots & P(x_{N-1}, y_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P(x_0, y_{M-1}) & P(x_1, y_{M-1}) & P(x_2, y_{M-1}) & \cdots & P(x_{N-1}, y_{M-1}) \end{bmatrix}, \tag{5.2.8}$$

Beginning with the interpolation over the x direction, each row of eq. (5.2.8) corresponds to an individual interpolation problem. The solution to these will be a set of coefficients $a_i(y_j)$ that parametrise polynomials ($P_{y_j}(x)$) in x that are labelled by the y_j:

$$P(x, y_i) = P_{y_j}(x) = \sum_i a_i(y_j) x^i. \tag{5.2.9}$$

To prepare for the interpolation over y, we form a new codomain matrix Y_y from the coefficients $a_i(y_j)$:

$$Y_y = \begin{bmatrix} a_0(y_0) & a_1(y_0) & a_2(y_0) & \cdots & a_{N-1}(y_0) \\ a_0(y_1) & a_1(y_1) & a_2(y_1) & \cdots & a_{N-1}(y_1) \\ a_0(y_2) & a_1(y_2) & a_2(y_2) & \cdots & a_{N-1}(y_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_0(y_{N-1}) & a_1(y_{N-1}) & a_2(y_{N-1}) & \cdots & a_{N-1}(y_{N-1}) \end{bmatrix}, \tag{5.2.10}$$

such that,

$$Y_y \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^{N-1} \end{bmatrix} = \begin{bmatrix} P(x, y_0) \\ P(x, y_1) \\ P(x, y_2) \\ \vdots \\ P(x, y_{N-1}) \end{bmatrix}. \tag{5.2.11}$$

The interpolations over the y direction can be performed analogously to those over the x. To achieve this, the interpolation must be carried out along the columns instead of the rows. The result of this will be a matrix of pure coefficients a_{ij}, satisfying the relation:

$$\begin{bmatrix} 1 \\ y \\ y^2 \\ \vdots \\ y^{N-1} \end{bmatrix}^T \begin{bmatrix} a_{00} & a_{10} & a_{20} & \cdots & a_{(N-1)0} \\ a_{01} & a_{11} & a_{21} & \cdots & a_{(N-1)1} \\ a_{02} & a_{12} & a_{22} & \cdots & a_{(N-1)2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{0(M-1)} & a_{1(M-1)} & a_{2(M-1)} & \cdots & a_{(N-1)(M-1)} \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^{N-1} \end{bmatrix} = P(x, y). \tag{5.2.12}$$

Note that in this example the results of the interpolation are the sets of canonical coefficients to the individual monomial terms $x^m y^n$. This is not necessary and in fact most interpolation schemes are based around expanding polynomials in terms of basis function that simplify the procedure. For example, in the $1d$ Lagrange interpolation algorithm outlined above, reconstruction is carried out in terms of Lagrange polynomials.

Modern multivariate interpolation schemes such as the *Zipple* algorithm [136] (as used by `FireFly`) utilise sophisticated techniques to achieve time complexities as low as $\mathcal{O}(N \log(N))$. Zipple's algorithm is particularly well suited to the reconstruction of sparse multivariate polynomials as it does not assume the polynomials to contain non-zero coefficients for every possible term. This reduces the total number of coefficients that must be calculated. Unfortunately, Zipple's algorithm is not readily parallelisable and so cannot take full advantage of the high operational throughput of GPU's. In fact, the $\mathcal{O}(N^2)$ Lagrange algorithm is a better choice, and it is this that we implement in this work.

To extend the 1d Lagrange algorithm to N dimensions, we must iteratively interpolate over the coefficients generated by previous interpolation steps. Whilst this could be preformed directly using the Lagrange coefficients $\bar{a}_n$, the result of the full interpolation would be a sum of products of Lagrange polynomials, e.g:

$$P(x, y) = \sum_{ij} (\bar{a}_x)_i (\bar{a}_y)_j l_i(x) l_j(y). \tag{5.2.13}$$

As we are interested in determining a fully simplified expression, in which like terms have been cancelled against one another, we must re-express the result of each interpolation stage in canonical form before proceeding with the next stage.

To rewrite eq. (5.2.5) in canonical form, both the numerator and denominator must be expanded. The denominator consists of pure numbers and can be efficiently computed in parallel using multiplicative reduction, where the product is subdivided and processed across multiple threads. We provide the details of our implementation of this in section 5.3.2.

The numerator, on the other hand, consists of a product of rank 1 polynomials. In direct coefficient space, its expansion through repeated polynomial multiplication is equivalent to an iterative convolution. Convolution is a costly operation with a time complexity of at best $O(N^2)$, where N is the degree of the polynomial. Lagrange polynomials, however, are written as a product of many small polynomials so the overall complexity will be greater as multiple convolution operations would be required. Some early performance gains can by achieved by making use of the parallel GPU architecture to perform multiple convolutions at the same time. At the first level, there are N rank 1 polynomials of the form:

$$P_i(x) = (x_n - x_i) \tag{5.2.14}$$

A set of $(N/2)$ new polynomials of rank 2 can be found by convolving pairs of these. This process can be iterated a total of $\log_2 N$ times to determine the full polynomial of rank N. This divide and conquer approach is significantly faster than performing each convolution sequentially. The early iterations involve a large number of small convolutions that can be computed in parallel. At later iterations, however, a small number of large convolutions is required. As each convolution must be calculated on a single thread the gains found on the first levels will be lost in the later levels. By transforming the system to frequency (or reciprocal) space, the convolution can be performed as element wise multiplication, taking $\mathcal{O}(N)$ time. For complex numbers,

this can be done using a Fourier Transform.

The utility of the FT is entirely made possible by the high efficiency of the *Fast Fourier Transform* (FFT) algorithm. The standard FT algorithm has the same time complexity as convolution and so will not lead to any efficiency gains. By placing the additional restriction that the degree of the polynomial in each dimension be power of 2, one may exchange the direct FT for the FFT. The FFT uses a method of division and conquest to achieve a time complexity of $\mathcal{O}(N\log(N))$. For polynomials of length $N = 100$, as are commonly encountered in precision calculations, the FFT approach will be approximately 15 times faster than regular convolution.

As discussed previously, to prevent integer overflow, we interpolate over Galois fields. These finite fields are incompatible with the standard FFT and we must instead use a variant known as the *Number Theoretic Transformation* (NTT). The use of the NTT in this way places restrictions on the choice of prime (P) over which the underlying Galois field is defined. We discuss this in detail in section 5.3.5, but note that P must be chosen such that it possesses appropriate primitive roots. The process for determining suitable primes can be very time consuming as it requiring trial and error, but can be performed as a precompensation step.

5.2.5 Chinese Remainder Theorem

As discussed earlier, the result of the above interpolation algorithm may only match the true polynomial up to modulation by some large prime number. If the coefficients of the original polynomial are less than this prime and positive, then the result of the interpolation will match the true polynomial. If this is not the case, further work is required to truly reconstruct the polynomial. In the case that coefficients are negative and less than half the value of the prime, their true value can be determined by simply subtracting the prime from the reconstructed value e.g.:

$$-3 \cong 4 \bmod 7; \tag{5.2.15}$$

$$-3 = 4 - 7. \tag{5.2.16}$$

If, however, the true value of a given coefficient is larger than the prime, then the determination of its value will be less straightforward and require the use of the CRT. This theorem states, given a set of congruences

$$\{c_i = C \bmod p_i\} \subset \mathbb{N}, \tag{5.2.17}$$

where $C \in \mathbb{N}$ and the $\{p_i\} \in \mathbb{N}$ (known as moduli) are (at least) pairwise coprime to one another, that the following expression may be uniquely determined:

$$C \bmod P = C \bmod \prod_i p_i. \tag{5.2.18}$$

This can be achieved using algorithm 2.

Algorithm 2 CRT

Input: Arrays of `moduli` and `a`
Output: The true coefficient `A`
`A` $\leftarrow 0$
`M` $\leftarrow 1$
Initialize an array `Ms` of size `nModuli`
Initialize an array `Ys` of size `nModuli`
for $i = 0$ **to** `nModuli` $- 1$ **do**
    `M` $\leftarrow$ `M` $\times$ `moduli`$[i]$
for $i = 0$ **to** `nModuli` $- 1$ **do**
    `Ms`$[i] = \frac{\texttt{M}}{\texttt{moduli}[i]}$
    `Ys`$[i]$ = modInverse(`Ms`$[i]$, `moduli`$[i]$)
for $i = 0$ **to** `nModuli` $- 1$ **do**
    `partialMult` $\leftarrow$ modMultiply(`a`$[i]$, `Ms`$[i]$, `M`)
    `partialMult` $\leftarrow$ modMultiply(`partialMult`, `Ys`$[i]$, `M`)
    `A` $\leftarrow$ modAdd(`A`, `partialMult`, `M`)
return `A`

For a given coefficient, each interpolation over a different prime field (p_i) can be interpreted as an individual a_i. If the true coefficient A is less than the product over

the primes ($A < P$), then the true coefficients can be completely reconstructed by the CRT. In reality, it is necessary for the absolute value of the true coefficient to be less than half the product of primes ($C < P/2$) to allow for the accounting of negative numbers.

The coefficients encountered in precision calculations are frequently as large as $10^{100} \approx 2^{332}$ so may require in excess of ten 32 bit prime fields to fully reconstruct. The maximum unsigned integer type with hardware support on modern GPUs has a depth of only 64 bits. It is for this reason that we perform the final CRT calculation on the CPU using the `GMP` library.

5.3 Numerical Framework

In this section, we describe the details of our implementation of the above interpolation strategy. We begin by discussing Galois fields and parallel reduction algorithms and towards the end of this section we provide a detailed background on the FT and NTT due to their central role in efficient Lagrange interpolation.

5.3.1 Modular Arithmetic

At the heart of polynomial reconstruction is the finite field over which the polynomial is defined. As such, it is crucial that the mathematical operations from which polynomials are comprised are implemented efficiently. Working with unsigned integers, the main hindrance in this regard it the possibility of integer overflow and underflow. The following subsection details the algorithms used to account for this issue while still ensuring a high level of efficiency. As explained previously, we work with Galois fields using modular arithmetic to ensure that the result of any mathematical operation will not exceed the value of field defining prime, P.

Although the result of any modular operation will necessarily be less than P, intermediate results need not be. Unsigned integer overflow can be thought of as a

modulation operation; modulo the integer size, for example 2^{32}. The modulo operator is non commutative, so this presents an issue for the storage of intermediate expressions. For example, the addition of two large, 32 bit values $a, b > 2^{31}$ will overflow, and the following sequence of operations will yield an incorrect result for their sum modulo some $P < 2^{32}$:

$$\text{MODSUM}(a, b, P) = ((a + b) \mod 32) \mod P \tag{5.3.1}$$

$$\neq (a + b) \mod P. \tag{5.3.2}$$

To avoid the overflow operation (the modulation by 32 in eq. (5.3.1)), we stipulate that $P < 2^{31}$.

With addition accounted for, the next issue is subtraction. Here, there is the added difficulty that we are working with unsigned types that cannot represent negative numbers directly. To account for this we interpret the upper half of the prime field as corresponding to negative numbers, i.e.:

$$-a \cong (P - a) < 2^{32}, \tag{5.3.3}$$

for $2^{31} > a > 0$ and $|a| < 2^{31}$. During the reconstruction process, both positive and negative numbers may be represented by this region of the field. This is not an issue and is a consequence of the cyclicity of the Galois field. Negativity is only relevant when the congruencies are combined to give the full coefficients $A_i \in \mathbb{Z}$. We define the modular subtraction between the numbers $a, b \in GF(P)$ as follows:

$$\text{MODSUB}(a, b) = (a - b + P) \mod P. \tag{5.3.4}$$

Here, the result of the first subtraction may lead to underflow although this will be immediately corrected with an overflow from the addition of P.

Unlike addition and subtraction, multiplications can overflow and underflow in more

challenging ways. The multiplication of two 32 bit unsigned integers (a and b) will yield a result with a size of up to 64 bits. Modular multiplication could be implemented straightforwardly by either casting to 64 bits or restricting the input size to $a, b < 2^{16}$. In our present implementation, we rely on casting, although, this is not ideal due to the reduced hardware support for 64 bit types. We are currently in the process of updating to a more sophisticated approach. Modular multiplication is a key step in many cryptographic algorithms. As a result, several highly efficient fixed type algorithms have been developed. The most notable of these is Montgomery multiplication, given by algorithm 3.

Algorithm 3 Montgomery Multiplication Algorithm

Input: Integers a, b, modulus P, and $R = 2^k$ with $\gcd(P, R) = 1$
Output: $(a \times b) \mod P$
Ensure: $T = (a \cdot b \cdot R^{-1}) \mod P$
Precompute P' such that $R \cdot P' \equiv -1 \pmod{P}$
$T \leftarrow a \cdot b$
$m \leftarrow (T \cdot P') \mod R$
$u \leftarrow (T + m \cdot P)/R$
if $u \geq P$ **then**
    $u \leftarrow u - P$
return u

During our preliminary testing, we found Montgomery's method to be $\sim 41\%$ faster than performing the calculation in 64 bits[1]. For the large number of multiplication operations expected to be present in `Ratracer` trace files, this could lead to a significant time saving.

Polynomials involve exponentiation, sometimes by large numbers. Direct exponentiation requires repeated multiplication steps to be performed. For the large exponents this is very inefficient. In this work we perform exponentiation by squaring. For example, 3^9 can be written $3 \times ((3^2)^2)^2$, requiring 4 multiplication operations instead of 9. In practice, a bit-shift operation can be used to keep track of the exponent,

[1]Tests were performed using an $80GB$ NVIDIA A100 GPU. For each test a set of 10^7 multiplications were computed, with each test being repeated $10,000$ times to account for variations in GPU warm-up time and thread scheduling.

further increasing the efficiency.

The final arithmetic operation required for polynomial interpolation is modular division. This is the least trivial to implement as it is defined as inverse modular multiplication. Division can be performed using the extended Euclidean method, algorithm 4.

Algorithm 4 Extended Euclidean Algorithm for Modular Inverse

Input: Integer a, modulus p
Output: $a^{-1} \mod p$ (modular inverse of a modulo p)
Ensure: Compute t_1 such that $(a \cdot t_1) \equiv 1 \pmod{p}$
$r_1 \leftarrow p$, $r_2 \leftarrow a \mod p$
$t_1 \leftarrow 0$, $t_2 \leftarrow 1$
while $r_2 \neq 0$ **do**
    $q \leftarrow \lfloor r_1 / r_2 \rfloor$
    $r_{\text{tmp}} \leftarrow r_2$
    $r_2 \leftarrow r_1 - q \cdot r_2$
    $r_1 \leftarrow r_{\text{tmp}}$
    $t_{\text{tmp}} \leftarrow t_2$
    $t_2 \leftarrow t_1 - q \cdot t_2$
    $t_1 \leftarrow t_{\text{tmp}}$
return $(t_1 + p) \mod p$

5.3.2 Parallel Reduction

Devices with a high degree of parallelism can combine large sets of numbers efficiently using reduction algorithms. Numerous implementations of these exist in efficient libraries such as NVIDIA's `Thrust` [137]. These pre-existing libraries rely on the data being formatted in a specific way. To map the data generated during interpolation into the desired from would involve duplicating the memory footprint and carrying out a large number of memory copy operations. To optimise both execution time and memory usage, we have developed a custom reduction method that minimises data copying. Our implementation makes use of the block, warp, thread hierarchy of NVIDIA GPUs and their associated memory access patterns. To describe this,

we will use the example of addition, where the problem we wish to solve is given by the sum:

$$S = \sum_{i=0}^{N} a_i. \tag{5.3.5}$$

We begin by discussing the simpler case in which N is less than the thread capacity of a block. In this case, each coefficient a_i can be loaded into the high speed shared memory of a warp within the block. Each warp contains $32 = 2^5$ threads and will store the value of up to 32 different a_i. The first stage of the reduction involves combining pairs of (a_i, a_{i+1}) within each warp. This can be done using the ___SHFL_DOWN_SYNC() `CUDA` call that allows for efficient memory read/write operations without race conditions or blocking. The second stage involves repeating this procedure, this time, using a step size of 2 instead of 1. The full procedure continues in the same way and is depicted in fig. 5.2. After completing this, each warp will contain a single number within the first element of its shared memory. The warps will then be synchronised and the result of the sum in each warp will be copied into the the shared memory of the first warp. The reduction process can then be repeated for a second time to complete the reduction. Performing the reduction entirely without the use of global memory ensures minimal run times. If a given sum consists of more elements than the thread capacity of a block then a single warp is used to compute the sum over the excess elements before carrying out the above procedure.

5.3.3 Discrete Fourier Transform

The FT is crucial to the simplification of Lagrange polynomials. For the finite set of coefficients that define a polynomial, the relevant form of the FT is the Discrete Fourier Transform (DFT). The DFT allows one to express a discrete signal $\{x[n]; n \in [0, N-1]\}$ in the reciprocal basis $\{X[k]; k \in [0, N-1]\}$, given by:

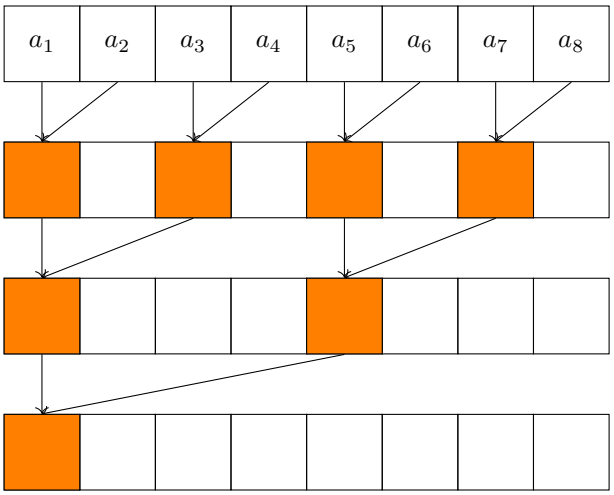


Figure 5.2: Intra warp reduction procedure for a hypothetical warp with 8 threads. The orange boxes indicate the read and write location within the array at each stage. The arrows indicate addition operations. The result of the reduction will be stored in the first element of the array.

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i\frac{2\pi}{N}kn}, \quad k = 0, 1, \ldots, N-1. \tag{5.3.6}$$

In this basis, each element $X[k]$ corresponds to an amplitude of an oscillatory function of frequency $\frac{2\pi}{N}k$. For this reason, the reciprocal basis is often referred to as the frequency domain. Consider now, a polynomial with complex coefficients:

$$P(x) = a_0 + a_1 x + ... + a_{N-1} x^{N-1}. \tag{5.3.7}$$

This polynomial can be represented by the sequence $\{a_0, ..., a_{N-1}\}$. Given a second polynomial ($Q(x)$) represented by $\{b_0, ..., b_{N-1}\}$, the product of P and Q, can be written:

$$P(x) \cdot Q(x) = \{a_0 \cdot b_0, a_0 \cdot b_1 + a_1 \cdot b_0, ..., a_{N-1} \cdot b_{N-1}\} \tag{5.3.8}$$

$$= \{a_0, ..., a_{N-1}\} * \{b_0, ..., b_{N-1}\}. \tag{5.3.9}$$

where, $*$ represents discrete convolution. Convolution is acieved with a double loop

over $[0, N-1]$, yielding $\mathcal{O}(N^2)$ time complexity. The convolution theorem states that:

$$\mathrm{DFT}\{x[n] * h[n]\} = \mathrm{DFT}\{x[n]\} \cdot \mathrm{DFT}\{h[n]\}. \tag{5.3.10}$$

This implies that convolution can be computed by a DFT followed by element-wise multiplication followed by an inverse DFT. Element-wise multiplication scales linearly in N. Unfortunately, the DFT algorithm has quadratic complexity, so this approach appears to only trade one bottleneck for another. Luckily, and as alluded to earlier, there exists a class of FFT algorithms that can perform a full DFT in only $O(N \log_2(N))$ time. This is achieved by repeatedly subdividing the problem into increasingly small DFTs.

5.3.4 Fast Fourier Transform

There are many variations of FFT, but the most well known is the *radix* 2 Decimation In Time (DIT) Cooley-Turkey algorithm [138]. The radix refers to the way in which eq. (5.3.6) is subdivided. Being radix 2, the DFT is separated into two sums at each stage:

$$X[k] = \sum_{n=0}^{N/2-1} x[2n] e^{-i\frac{2\pi}{N}k(2n)} + \sum_{n=0}^{N/2-1} x[2n+1] e^{-i\frac{2\pi}{N}k(2n+1)}, \quad k = 0, 1, \ldots, \frac{N}{2} - 1. \tag{5.3.11}$$

This can be rewritten in terms of two DFTs with period $N/2$ by extracting a factor of $e^{-\frac{2\pi i}{N}k}$ from the second sum:

$$X[k] = \sum_{n=0}^{N/2-1} x[2n] e^{-i\frac{2\pi}{N/2}kn} + e^{-\frac{2\pi i}{N}k} \sum_{n=0}^{N/2-1} x[2n+1] e^{-i\frac{2\pi}{N/2}kn}, \quad k = 0, 1, \ldots, \frac{N}{2} - 1. \tag{5.3.12}$$

Here, the left sum corresponds to a DFT over the even indexed inputs, and the right corresponds to a DFT over the odd indexed inputs. The preceding exponential is

known as a twiddle factor. The key to the FFT's efficiency is that, written in this way, the periodic nature of the complex exponential can be exploited to express the $X[k + N/2]$; $k < N/2$ in terms of quantities already known from the computation of the $X[k]$; $k < N/2$:

$$X[k + N/2] = \sum_{n=0}^{N/2-1} x[2n] e^{-i\frac{2\pi}{N/2}(k+N/2)n} + e^{-\frac{2\pi i}{N}(k+N/2)} \sum_{n=0}^{N/2-1} x[2n+1] e^{-i\frac{2\pi}{N/2}(k+N/2)n} \tag{5.3.13}$$

$$= \sum_{n=0}^{N/2-1} x[2n] e^{-i\frac{2\pi}{N/2}kn} e^{-2\pi ni} + e^{-\frac{2\pi i}{N}k} e^{-\pi i} \sum_{n=0}^{N/2-1} x[2n+1] e^{-i\frac{2\pi}{N/2}kn} e^{-2\pi ni} \tag{5.3.14}$$

$$= \sum_{n=0}^{N/2-1} x[2n] e^{-i\frac{2\pi}{N/2}kn} - e^{-\frac{2\pi i}{N}k} \sum_{n=0}^{N/2-1} x[2n+1] e^{-i\frac{2\pi}{N/2}kn} \tag{5.3.15}$$

$$= E_k - e^{-\frac{2\pi i}{N}k} O_k \tag{5.3.16}$$

where E_k and O_k refer to the k^{th} even and odd DFTs respectively. In this way, the total number of required DFTs is halved. By repeating this process until each DFT is computed on a set of two inputs, the time complexity can be further reduced. The final DFTs are order $\mathcal{O}(1)$ and the initial DFT for each of the N outputs is subdivided $\log_2(N)$ times, so the overall complexity is $\mathcal{O}(N \log_2(N))$. An illustration of the signal flow by which multiple DFTs are combined to give the full FFT is depicted in fig. 5.3.

The above description of the FFT operates on a complex field. In this work, however, polynomials are reconstructed over Galois fields. Galois fields, like the complex exponential, are cyclic, and are thus compatible with a modified version of the FFT known as a Number Theoretic Transformation (NTT).

5.3.5 Number Theoretic Transformation

The NTT is achieved by replacing the exponentials in the complex DFT with appropriate primitive roots of unity in the Galois field $\text{GF}(P) = \mathbb{Z} \mod P$:

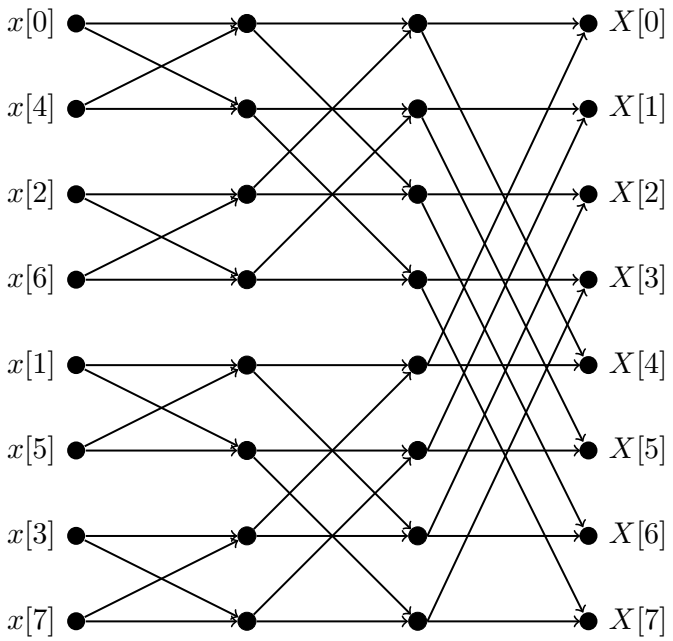


Figure 5.3: The signal flow used to compute the FFT. The input array is shown on the left with the result of the FFT on the right. The pairings over which each individual DFT is performed are indicated by arrows. Owing to their appearance, these pairings are often referred to as butterfly operations.

$$e^{-i\frac{2\pi}{N}kn} \to \omega_N^{kn} \mod P, \tag{5.3.17}$$

where, ω_N is the N^{th} root of unity:

$$\omega_N^N \to 1 \mod P, \tag{5.3.18}$$

and $\omega_N^0, \omega_N^1, ..., \omega_N^{N-2}$ are all distinct.

With this adaptation, the number theoretic DFT can be written:

$$X[k] = \sum_{n=0}^{N-1} x[n]\omega_N^{kn}, \quad k = 0, 1, \ldots, N-1. \tag{5.3.19}$$

To perform an NTT using the FFT algorithm, where the DFTs are iteratively subdivided, it is necessary to find M^{th} primitive roots for all $M \in \{2^1, 2^2, 2^3, ..., N\}$. The main difficulty in going from a complex DTT to an NTT is in the determination

of prime fields for which these exit. In general, a prime field will not give rise to an n^{th} primitive root, so a scan of large primes $P \approx 2^{32}$ is required. If a field has an n^{th} root, the following condition will be satisfied:

$$(P-1) \mod n = 0. \tag{5.3.20}$$

This is equivalent to saying that there exists a cyclic subgroup of GF(P) of order n. The generators of the subgroup are the primitive roots and can be found using algorithm 5. This approach assumes that n is a power of 2, a condition that is also required for the NTT to be efficient.

Algorithm 5 Determine Primitive Root Modulo p

1: **Input:** Prime number p and integer n dividing $p-1$
2: **Output:** A primitive root g of order n modulo p
3: Generate a random value $x \in \mathbb{Z}_p$
4: Compute $g \leftarrow x^{(p-1)/n} \mod p$
5: **if** $g^{n/2} \not\equiv 1 \pmod{p}$ **then**
6: **return** g (a primitive root)
7: **else**
8: Repeat with a new random x

The first step for the evaluation of polynomial multiplications is to precompute a set of primes, whose finite fields have multiplicative subgroups of order $2^1, 2^2, ..., 2^{13}$ along with their $2^i th$ primitive roots[1]. With these in hand, it is possible to perform Fourier transforms on polynomials of length $2^1, 2^2, ..., 2^{13}$, thereby allowing the complete expansion of the Lagrange polynomials. It should be noted that in the computation of the primes according to eq. (5.3.20), it is only necessary to check this criterion for the largest power of 2, i.e. $n = 2^{13}$ as all lower powers will implicitly satisfy the condition.

[1]Larger orders could in principle be used, however, we find that memory constraints restrict the maximum polynomial size to be at most 2^{13}.

5.4 Results and Discussion

To verify the effectiveness of our code, we have benchmarked it against `FireFly`, using both a synthetic univariate polynomial test as well as a real-world multivariate master coefficient. For the polynomial test, reconstruction was performed over functions of the form:

$$f(x) = \sum_{i=1}^{N} ix^i. \tag{5.4.1}$$

By varying N, the time scaling of the two codes was determined. This is shown in fig. 5.4. Note that for these runs, the parser was not used in either program. For small polynomials, `FireFly` was found to be around an order of magnitude faster than our code. Additionally, our code had no dependence on the polynomial degree up to approximately $N = 10^3$. These effects can be attributed to the overhead associated with the initialisation of the GPU; this involves memory allocation and driver loading. For our test system[1], this start up time was found to be ~ 0.4s.

From fig. 5.4, it can be seen that the execution time for `FireFly` has a relatively poor scaling with N, and although `FireFly` is faster for small polynomials, it falls short of our code for polynomials of degree $N > 100$. For the largest polynomials tested, we found our code to be 274 times faster, with `FireFly` taking 16.8min to complete, compared to 3.7s for our code.

Although the gains afforded by our interpolation scheme grow with increasing problem complexity, there is a limit to the size of polynomials that can be interpolated. This is set by the memory constraints of the system. The reconstruction process has a minimum memory requirement given by:

$$\text{Mem}_{\text{min}} = (4N_{\text{vars}}N_{\text{terms}}^2) \text{ bytes}, \tag{5.4.2}$$

where N_{vars} is the number of variables and N_{terms} is the maximum degree of the

[1] 80GB NVIDIA A100 GPU and 16 core and 2.1GHz Intel(R) Xeon(R) Silver 4216 CPU.

polynomial and must be a power of 2. For 80GB of available ram, this equates to a univariate polynomial of degree $\sim 141 \times 10^3$. The requirement that $N_{\text{terms}} \in 2^n;\ n \in \mathbb{N}$ is necessary to allow the NTT to be performed. Note that if the degree of the polynomial is not close to a power of 2, then an unnecessarily large number of black-box probes will be computed to pad the NTT array to the appropriate size. We do not account for this issue in this work, however a simple strategy to reduce the GPU workload would simply be to pad the NTT array using a single probe value for all additional elements. This wouldn't affect the interpolation result whilst still allowing the NTT to be performed.

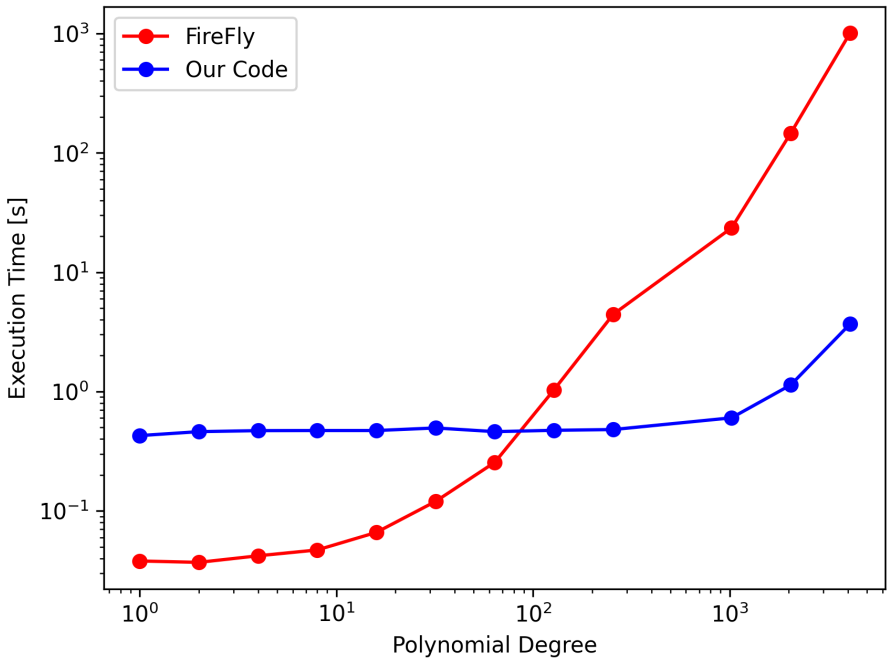


Figure 5.4: Interpolation times for 1d polynomials of varying degree for `FireFly` (red) and our code (blue). Our code outperforms `FireFly` for polynomials of degree ~ 100. The polynomials used to produce this figure were selected with coefficients that could be reconstructed using a single prime field.

For the real-world test, we reconstructed a single master coefficient generated for the process $gg \to hh$ at 2 loop. The chosen coefficient was a rational function of 2 momentum variables (s and t). Before performing the interpolation, we used a separate `Mathematica` script [139] to multiply out the denominator. The resulting

polynomial had a degree of 75 in s and 81 in t and each of the polynomial coefficients were $\sim 10^{94}$ in size, requiring ~ 5 independent 64 bit prime fields and ~ 10 independent 32 bit fields to reconstruct. The results of this test are broken down into the parsing time and execution time and are summarised in table 5.1. In spite of requiring more prime fields, our code outperformed `FireFly` in terms of the total execution time by approximately an order of magnitude.

We found the parsing time for our code to constitute a significant fraction of the total run time, 25% for the real world example in table 5.1. This is in contrast to `FireFly`, where parsing time takes less than 1% of the total. We have not dedicated much time to the optimisation of our parser as it a CPU bound process and not the focus of this work. In future versions of our code, however, efficient algorithms such as the *shunting yard parser* used by `FireFly` could be implemented to further reduce execution times.

From table 5.1, it is clear that the reconstruction of rational functions is a more intensive procedure than that of polynomials, with `FireFly` taking over twice as long to process the rational function compared to the expanded polynomial. There are many ways to perform rational interpolation, but perhaps the most suited to our current approach is to interpolate the numerator and denominator as individual polynomials. If GPU acceleration yields the same time savings as observed for polynomials, this technique could lead to a significant increase in the accessible precision of QFT calculations.

Overall, we suggest the use of a hybrid approach between CPU and GPU based interpolation. The minimal start up overhead of CPUs makes them ideal for handling small coefficients, whilst the huge operational throughput of GPUs can greatly accelerate the processing of larger coefficients. This approach would also circumvent the issues caused by the memory limitations of current GPUs by allowing the parallel batch processing of small coefficients on the CPU. Furthermore, the overhead of initialising the GPU could be diluted by continuously processing coefficients in a sequential manner such that the main process remains active on the GPU throughout.

This will prevent multiple spool up operations.

	Our Code (Pol)	`FireFly` (Pol)	`FireFly` (Rat)
Parsing Time [s]	0.15	0.03	0.04
Interpolation Time [s]	0.44	5.66	12.30
Total Time [s]	0.59	5.69	12.34

Table 5.1: Parsing and interpolation times for a master integral coefficient generated for the $gg \to hh$ at two loop. This coefficient was a 2d function of the s and t Mandelstam variables. As, at present, our code is only able to handle polynomial interpolation, we preprocessed this coefficient by multiplying out the denominator using an external program. The timings associated with the interpolation of this polynomial are indicated by (Pol). For completeness, we also present the interpolations time for `FireFly` to solve the full rational problem (Rat).

5.5 Outlook

We have demonstrated the utility of modern GPUs in precision calculations in QFT. We see that, in sufficiently large, real world scenarios, our code can perform interpolation over a time span that is an order of magnitude lower than state of the art CPU based codes. At N^2LO, the evaluation of the integrals associated with a single scattering process can take a period of weeks to complete. The vast majority of this time is associated with the interpolation procedure, specifically in the evaluation of black-box probes. If a similar level of GPU acceleration as demonstrated for polynomial interpolation can be obtained for the more complicated process of rational interpolation, then computation times could be reduced from weeks to days. This would remove a barrier to a new frontier of precision physics. With the rapid development of new GPU hardware in the pursuit of Artificial Intelligence (AI) with ever increasing memory capacities and core counts, the time to transition old code to modern GPU hardware is now!

Chapter 6

Conclusions

In this thesis, we have discussed the use of modern computational hardware in solving a wide variety problems in fundamental physics. Our investigations have focussed on the extremes of energy and curvature, where analytical techniques may struggle to resolve important dynamics. The enormous operational throughputs afforded by the highly parallel architectures or modern HPC clusters and GPU devices allow for high accuracy modelling of physical processes. We have demonstrated this both for scenarios where extreme precision is required as well as in cases where, due to unknowns in the theory, large samples of numerical data are required to draw representative conclusions.

Motivated by string theory and the strong CP problem, in chapter 2 we investigated the possibility that ALP-photon interconversion may be responsible for the perceived excess in high energy Blazar spectra. String theories generically predict large numbers of ALPs. Whilst numerous studies have investigated the effect of a single ALP state on Blazar spectra, none, prior to our work, had investigated a truly string motivated model. The main difficulty in this regard arises due to the uncertain nature of the ALP mass and coupling distributions. We addressed this issue through an anarchical framework, marginalising over many samples of the masses and couplings as well as the magnetic field structures of intergalactic space. Through the use of temporal discretisation, we simulated the propagation of photons from their Blazar

source to the earth to determine their survival probabilities. Fitting these to the observed Blazar spectra, our results suggest that if ALPs are responsible for the excess, then models involving fewer ALPs are favoured. In future studies, where the details of the ALP spectra (couplings and masses) can be extracted from their progenitor string theories, investigation of Blazar spectra may provide an insight into the viability of a given theory.

Continuing with our focus on the physics of scalar fields around BHs, in chapter 4 we presented a work in progress fixed background numerical GR code for the determination of the DF force on BHs. As in chapter 2, this involves the time integration of the field equations, albeit this time using a classical field and a non-trivial metric. Novel to our code, we express this metric in the horizon-penetrating, KS coordinates. This allows for increased accuracy around the horizon, but complicates the extraction of inertial observables. During preliminary testing, our code successfully met the convergence criteria, satisfying the Isotropic Hamiltonian and momentum constraints even at late times. We are currently in the process of implementing the DF calculation as well as refactoring the code to achieve improvements in efficiency. Once fully implemented, our software may aid in the calculation of gravitation wave signals emitted by binary systems as their BHs spiral though accreted matter. In addition, the use of the KS gauge greatly simplifies the metric form for Kerr type BHs. With the addition of spin, fixed background simulations in KS coordinates may allow for detailed analyses of superradiant amplification in binary systems where the orbit of the companion BH may be perturbed by the energy loss channel generated by dynamical friction with superradiant cloud. In addition, the growth of the cloud itself may be affected by the orbit of the secondary BH.

In chapter 5, we moved away from the extreme energy and curvature environments of BHs, focussing instead on the high precision calculations required to interpret the results of terrestrial experiments. With the ever increasing energies reached by particle colliders such as the LHC, the demand for high order calculations in perturbative QFT is growing. The determination of the resulting loop integrals

requires the use of sophisticated computational techniques such as the Laprota algorithm. Focussing on one core bottleneck in this procedure, we demonstrated the utility of GPU acceleration in the interpolation of multivariate polynomials. By selecting algorithms that best make use of the parallelism provided by GPU hardware, we achieved a reduction in execution time by an order of magnitude for sufficiently large real world problems. For its use in the Laporta algorithm, our code must be extended to work with rational functions. This can be achieved through repeated polynomial interpolation steps. A large proportion of the efficiency gains seen in our current proof of concept work would likely be maintained in a full, rational code. Improvements in speed of this magnitude would reduce the evaluation time for complicated Feynman diagrams from weeks to hours, making the computation of very high precision matrix elements feasible.

Throughout this thesis, we have used the enormous parallelism of modern devices to probe theories in a way that would have been impossible only decades prior. With the recent development of artificial intelligence, GPUs and other specialised devices with highly parallel architectures have undergone significant development. We are currently on the verge of a revolution in numerical physics. As the memory capacity of these devices grows, and as specialised parallel algorithms are developed, the ability to probe complex interacting systems will only increase. No physical system exists truly in isolation. Numerical simulations, therefore, constitute one of the most effective methods through which to probe the theoretical interplay between the vast array of physics that governs the behaviour of our universe.

Appendix A

Statistical approach to producing mixing matrices using the Haar measure

Here we can outline how the mixing matrices are inductively produced. Following section 2.2, one can generate a sample of $N-1$ mixing angles, $\{\theta_{ij}\}$, in spherical polar coordinates. As a simplifying assumption, we take the mixing matrices, $U \in SO(N)$. Since $SO(N)$ is parametrised by $N(N-1)/2$ mixing angles, $\left\{\theta_{ij}\right\}_{1\leq i\leq j\leq N-1}$, the mixing matrix for $SO(2)$ can be parametrised by a single angle, θ_{11}:

$$U_2 = \begin{bmatrix} \cos\theta_{11} & \sin\theta_{11} \\ -\sin\theta_{11} & \cos\theta_{11} \end{bmatrix} . \tag{A.0.1}$$

The Haar measure of this matrix is given by eq. (2.2.2) and, explicitly, is

$$dV = d\theta_{11} \, , \tag{A.0.2}$$

which informs us we sample over θ_{11} uniformly in $[0, 2\pi]$. Generalising to higher dimensions, the structure of the matrices becomes much less trivial. The most

straightforward approach to writing down a higher $SO(N)$ matrix is as a product of matrices:

$$U_N = U'_N S_N(U_{N-1}) \tag{A.0.3}$$

where $S_N(U_{N-1})$, given by:

$$S_N(U_{N-1}) = \begin{bmatrix} U_{N-1} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{A.0.4}$$

where $\mathbf{0}$ is a $(N-1)$ dimensional zero vector and $S_N(U_{N-1})$ represents rotations in the hyperplane orthogonal to the N^{th} dimension. U'_N is the N-dimensional analogue of U_2, describing rotations that involve the N^{th} coordinate:

$$U'_{N1,1} = \cos\left(\theta_{1,N-1}\right) \tag{A.0.5}$$

$$U'_{N1,i+1} = \cos\left(\theta_{i+1,N-1}\right) \prod_{m=1}^{i} \sin\left(\theta_{m,N-1}\right) \quad \forall \quad i \in [1, N-2] \tag{A.0.6}$$

$$U'_{N1,N} = \prod_{m=1}^{N-1} \sin\left(\theta_{m,N-1}\right) \tag{A.0.7}$$

$$U'_{Ni,j} = 0 \quad \forall \quad j < i-1 \;\; \& \;\; i > 1 \tag{A.0.8}$$

$$U'_{Ni,j} = -\sin\left(\theta_{j,N-1}\right) \quad \forall \quad j = i-1 \;\; \& \;\; i > 1 \tag{A.0.9}$$

$$U'_{Ni+1,i+1} = \cos\left(\theta_{i,N-1}\right)\cos\left(\theta_{i+1,N-1}\right) \quad \forall \quad i \in [1, N-2] \tag{A.0.10}$$

$$U'_{NN,N} = \cos\left(\theta_{N-1,N-1}\right) \tag{A.0.11}$$

Explicitly, the $SO(3)$ mixing matrix has the following form:

$$U_3 = \begin{bmatrix} \cos\theta_{12} & \sin\theta_{12}\cos\theta_{22} & \sin\theta_{12}\sin\theta_{22} \\ -\sin\theta_{12} & \cos\theta_{12}\cos\theta_{22} & \cos\theta_{12}\sin\theta_{22} \\ 0 & -\sin\theta_{22} & \cos\theta_{22} \end{bmatrix} \begin{bmatrix} \cos\theta_{11} & \sin\theta_{11} & 0 \\ -\sin\theta_{11} & \cos\theta_{11} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{A.0.12}$$

Hence we produce the symbolic form of the N-dimensional mixing matrices inductively. We then sample the angles according to the Haar measure as shown in eq. (2.2.2), and we can visualise these samples as populating an N-dimensional sphere uniformly.

Appendix B

Producing CAST bounds

This section provides a detailed overview of the approach used to recast the CAST bounds in section 2.3.4. The original CAST bound can be generalised by considering the detector's sensitivity. From any point $(g^{\gamma}_{N=1}, g^{e}_{N=1})$ on the original bound, the black line shown in fig. 2.2, we can extract a proxy for the maximum signal ($\sigma^{\max}$) that is consistent with non-detection:

$$\sigma_{\max} = \Phi_{\text{tot}}(g^{e}_{N=1}, g^{\gamma}_{N=1})(g^{\gamma}_{N=1})^2\,, \tag{B.0.1}$$

where Φ_{tot} is the total single ALP flux at the detector; computed using eq. (2.3.4), eq. (2.3.5) and eq. (2.3.6):

$$\Phi_{\text{tot}}(g^{e}_{N=1}, g^{\gamma}_{N=1}) = \Phi_P(g^{\gamma}_{N=1}) + \Phi_B(g^{e}_{N=1}) + \Phi_C(g^{e}_{N=1})\,. \tag{B.0.2}$$

For any point in the coupling parameter space, the total flux at the detector must, therefore, be less than:

$$\Phi_{\max}(g^{\gamma}) = \sigma_{\max}/{g^{\gamma}}^2 \tag{B.0.3}$$

We note that to obtain the total flux at the detector we integrated over the differential fluxes of eq. (2.3.1), eq. (2.3.2) and eq. (2.3.3) using a lower integration boundary

of $\omega = 1\,\text{keV}$ (corresponding to the detector threshold) and an upper bound of $\omega = 10\,\text{keV}$.

The recast bounds in fig. 2.2 are shown as heat maps. The weight assigned to each set of couplings (g^e, g^γ), corresponds to the proportion of our sample of survival and conversion probabilities, $\{P_{\gamma\to\gamma}\}$ and $\{P_{e\to\gamma}\}$, that are consistent with eq. (B.0.3). That is to say that, for a given (g^e, g^γ) and probability $P^l_{e\to\gamma} \in \{P_{e\to\gamma}\}$, we find the proportion (η) of $\{P_{\gamma\to\gamma}\}$ that satisfy:

$$\Phi_{\text{max}}(g^\gamma) \geq \Phi_P(g^\gamma)P_{\gamma\to\gamma} + (\Phi_B(g^e) + \Phi_C(g^e))P_{e\to\gamma} \tag{B.0.4}$$

At this point, η could be found by Monte Carlo sampling $\{P_{\gamma\to\gamma}\}$ and determining the number of elements that satisfy eq. (B.0.4). This approach is, however, very computationally expensive. A significant saving can be achieved by instead considering the critical value (P_{Crit}) of $P_{\gamma\to\gamma}$ that achieves equality in eq. (B.0.4):

$$P_{\text{Crit}}(g^e, g^\gamma, P^l_{e\to\gamma}) = \frac{1}{\Phi_P(g^\gamma)}\left[\Phi_{\text{max}}(g^\gamma) - (\Phi_B(g^e) + \Phi_C(g^e))P^l_{e\to\gamma}\right], \tag{B.0.5}$$

η is now given by the cumulative sum - evaluated at P_{Crit} - of the histogram of $\{P_{\gamma\to\gamma}\}$; $\eta = C(P_{\text{Crit}})$. An example of this procedure is depicted in fig. B.1. Repeating this process for every $P^l_{e\to\gamma} \in \{P_{e\to\gamma}\}$, the total weight associated with a coupling pair is given by:

$$W(g^e, g^\gamma) = \frac{1}{N_{ae}}\sum_l C(P_{\text{Crit}}(g^e, g^\gamma, P^l_{e\to\gamma}))\,, \tag{B.0.6}$$

where N_{ae} is the size of $\{P_{e\to\gamma}\}$ - taken here to be $N_{ae} = N_{a\gamma} = 10^4$. $W(g^e, g^\gamma)$ is equivalent to the fraction of $\{P_{\gamma\to\gamma}\}$ and $\{P_{e\to\gamma}\}$ that satisfy eq. (B.0.4) and will saturate to 1 for sufficiently low couplings.

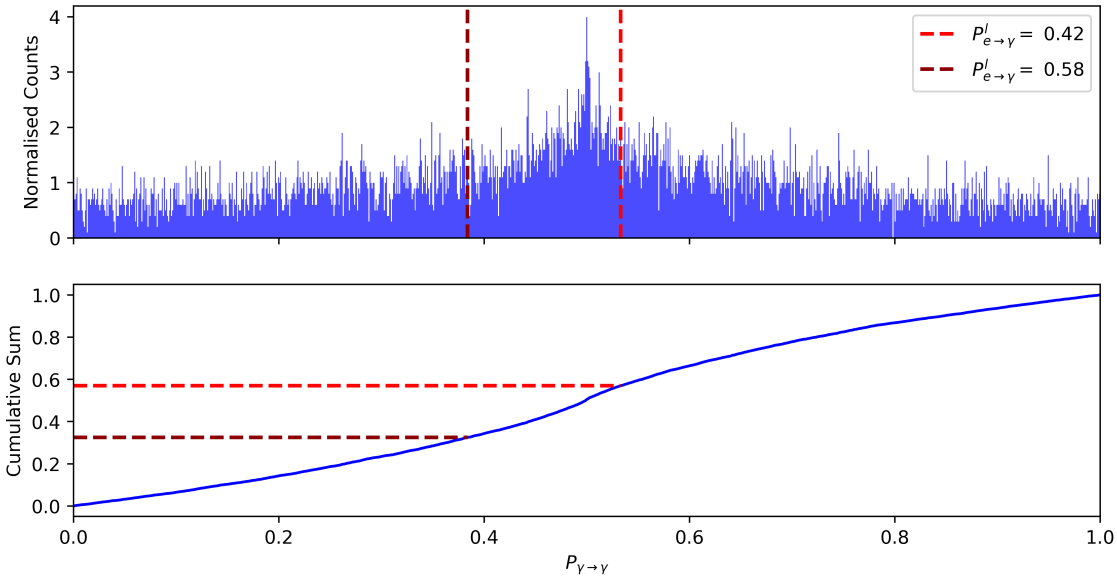


Figure B.1: An example of a photon survival histogram for a 2 ALP scenario with couplings $g_{ae} = 1.14 \times 10^{-12}$GeV and $g_{a\gamma} = 8.02 \times 10^{-11}$GeV. The critical values (P_{Crit}) for two distinct values of $P^l_{e\to\gamma}$ are indicated by the dashed lines. The cumulative sum of the histogram is shown below with the weights (η) being given cumulative sum at each critical value (these are indicated by the horizontal dashed lines).

Appendix C

Fit quality for VHE spectra and statistical approach to multi-ALP parameter space

To assess the fit of our multi-ALP model to the VHE Blazar spectra, we employ the same statistical methods as Refs. [84, 89]. In particular, we use the observed spectra, $\Phi_i^{\rm obs}$, shown in table 2.1.

We note that $\Phi_i^{\rm obs}$ contains data on the blazar source plus all ambient backgrounds. We compute the photon survival probability for a given point in the theory parameter space:

$$\left\langle P_{\gamma\to\gamma}\right\rangle_i (g^\gamma, g^e) = \frac{1}{\Delta E_i}\int_{\Delta E_i} {\rm d}E P_{\gamma\to\gamma}(E)\,, \tag{C.0.1}$$

where E is the photon energy, and this probability can be computed on a bin-by-bin basis (see fig. C.1). The absorption-corrected flux is given by

$$\Phi_i(g^\gamma, g^e) = \left\langle P_{\gamma\to\gamma}\right\rangle_i^{-1} (g^\gamma, g^e)\Phi_i^{\rm obs}\,, \tag{C.0.2}$$

where again, we have kept the theory parameter dependence explicit. For a given point in the model parameter space, (g^γ, g^e), we are tasked with understanding its

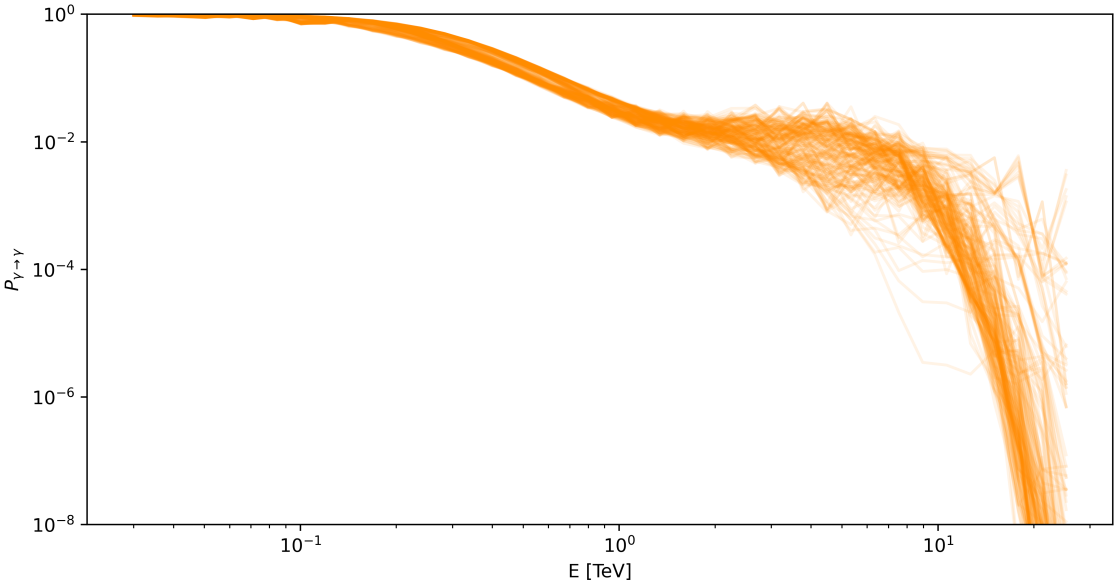


Figure C.1: An example distribution of photon survival probabilities as a function of energy. These probabilities correspond to the source 1ES0414+009.

compatibility with the observed spectra. For this purpose, we must quantify the observed flux in the absence of new physics and do so by fitting $\Phi_i(g^\gamma, g^e)$ to a power law:

$$f(E) = \begin{cases} N_0 \left(E/E_0\right)^{-\Gamma}, & p_{\text{fit}} \geqslant 0.05 \\ N_0 \left(E/E_0\right)^{-(\Gamma+\beta_c \ln(E/E_0))}, & \text{otherwise} \end{cases} \tag{C.0.3}$$

where p_{fit} denotes the fit probability. The first power law contains three parameters: N_0, E_0, and Γ while in the second (which is a parabola in log-log space), there are four fit parameters, N_0, E_0, β_c and Γ. The power law function is used unless the baseline (no ALP) fit probability (as calculated by Ref. [84]) is less than 0.05. An example of a corrected spectrum fitted with a power law is shown in fig. C.2

We note that, when computing the fit parameters, we perform a Chi-Squared analysis, taking the spectral uncertainties to be statistical. For each energy bin (E_i), the fit residual is calculated:

$$\chi_i = \frac{\Phi_i - f\left(E_i\right)}{\sigma_i}, \tag{C.0.4}$$

where σ_i denotes the statistical measurement uncertainty at 68% confidence level.

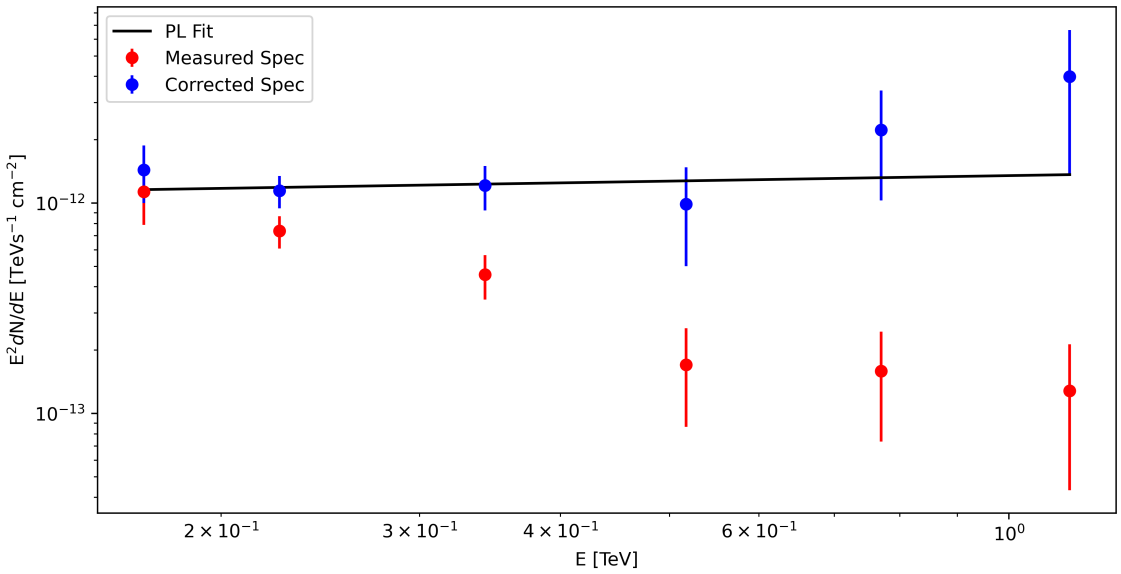


Figure C.2: An example of the measured (red) and corrected (blue) spectrum for 1ES0414+009 for a single realisation of magnetic field structure for a single ALP. The power law fit to the corrected spectrum is indicated by the black line.

The value of χ^2 can now easily be obtained by summing the residuals over all energy bins and squaring.

Assuming that $P_{\gamma\to\gamma}$ accurately predicts the Universe's opacity to VHE γ rays, we would expect the residuals in the optically thick regime to follow a Gaussian distribution with a mean of zero. To test this hypothesis, we employ the t-test, where we calculate the test statistic, t, as follows:

$$t = \frac{\bar{\chi}}{\sqrt{\sigma_\chi / N_\chi}} \,. \tag{C.0.5}$$

where $\bar{\chi}$ denotes the mean and σ_χ the variance of the residual distribution, which comprises a total of N_χ data points. The p_t test statistic we use to quantify the fit of our model to the observed spectra can be obtained using a standard t-test procedure.

Appendix D

GPU Structure

NVIDIA GPUs, as used in this work, are programmed using the `CUDA` API [140]. `CUDA` allows general purpose computational work to be offloaded to a GPU. It introduces a number of software level abstractions to help facilitate optimal parallel execution. In this section, we will briefly describe these abstractions and their relation to the true GPU hardware.

A function that is to be executed in parallel is known as a kernel. Kernels run on the GPU and can be launched from the CPU (Host) or GPU (Device). When a kernel is called, its execution is divided into groups of threads known as warps. Each warp contains 32 threads, where each thread corresponds to a single execution of the kernel. Importantly, every thread in a warp will execute each kernel instruction at the same time. This is known as lockstep execution and can lead to large bottlenecks when working with conditional logic where a small number of threads may execute code in one branch while all others are paused until the next branch is reached. This is known as warp divergence.

As well as warps, `CUDA` introduces two additional thread groupings, thread blocks and grids. A thread block can contain up to 1028 threads and a grid can contain an arbitrary number of blocks.

At the hardware level, GPUs consist of a number of parallel compute units known as Streaming Multiprocessors (SMs). Each SM contains a set of execution cores

and a small amount of high speed shared memory. Additionally, GPUs have a large amount (≈ 80 GB) of slower, global memory that can be accessed by all SMs.

Each thread block will be executed on a single SM and the grid of blocks will be assigned to saturate the SMs. This structure allows for efficient memory transfer between threads. All threads in a warp execute in lockstep so there are no race conditions[1] when accessing shared memory. For inter-warp communication within the same block, it is necessary to ensure thread level synchronization. This can be achieved through the `CUDA` API. Finally, for inter-block communication, global memory must be used.

It should be noted that in reality, SMs execute multiple warps simultaneously. A warp scheduler is used to assign work to the SM cores. To ensure maximal throughput, it is necessary to provide each SM with more than a single warp to prevent gaps in execution.

[1] Race conditions occur when multiple threads attempt to read from, or write to, the same section of memory simultaneously and can lead to undefined behaviour.

Bibliography

[1] P. D. Group, R. L. Workman, V. D. Burkert, V. Crede, E. Klempt, U. Thoma et al., *Review of particle physics*, *Progress of Theoretical and Experimental Physics* **2022** (08, 2022) 083C01, [https://academic.oup.com/ptep/article-pdf/2022/8/083C01/49175539/ptac097.

[2] P. Gondolo and G. G. Raffelt, *Solar neutrino limit on axions and keV-mass bosons*, *Phys. Rev. D* **79** (2009) 107301, [0807.2926].

[3] C. Dessert, A. J. Long and B. R. Safdi, *X-ray Signatures of Axion Conversion in Magnetic White Dwarf Stars*, *Phys. Rev. Lett.* **123** (2019) 061104, [1903.05088].

[4] D. Traykova, K. Clough, T. Helfer, E. Berti, P. G. Ferreira and L. Hui, *Dynamical friction from scalar dark matter in the relativistic regime*, *Physical Review D* **104** (Nov., 2021) .

[5] F. Chadha-Day, J. Maxwell and J. Turner, *Alp anarchy*, *Journal of Cosmology and Astroparticle Physics* **2024** (Sept., 2024) 056.

[6] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995, 10.1201/9780429503559.

[7] W. Gerlach and O. Stern, *Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld*, *Zeitschrift fur Physik* **9** (Dec., 1922) 349–352.

[8] A. Borisevich, F. Gleyzer and V. Gavrilenko, *Feedforward inversion control of dc/dc dual-bridge series resonant converter in buck and boost modes*, 2021.

[9] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, *Phys. Rev. Lett.* **38** (1977) 1440–1443.

[10] F. Wilczek, *Problem of Strong P and T Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978) 279–282.

[11] H. Quinn, *The cp puzzle in the strong interactions*, 2001.

[12] H. Leutwyler, *The ratios of the light quark masses*, *Physics Letters B* **378** (June, 1996) 313–318.

[13] A. Bilal, *Lectures on anomalies*, 2008.

[14] C. B. Adams, N. Aggarwal, A. Agrawal, R. Balafendiev, C. Bartram, M. Baryakhtar et al., *Axion dark matter*, 2023.

[15] D. Tong, *Lectures on string theory*, 2012.

[16] M. Cicoli, M. D. Goodsell and A. Ringwald, *The type iib string axiverse and its low-energy phenomenology*, *Journal of High Energy Physics* **2012** (Oct., 2012) .

[17] P. Svrcek and E. Witten, *Axions In String Theory*, *JHEP* **06** (2006) 051, [hep-th/0605206].

[18] M. Cicoli, M. Goodsell and A. Ringwald, *The type IIB string axiverse and its low-energy phenomenology*, *JHEP* **10** (2012) 146, [1206.0819].

[19] I. Broeckel, M. Cicoli, A. Maharana, K. Singh and K. Sinha, *Moduli stabilisation and the statistics of axion physics in the landscape*, *JHEP* **08** (2021) 059, [2105.02889].

[20] M. Demirtas, N. Gendler, C. Long, L. McAllister and J. Moritz, *PQ Axiverse*, 2112.04503.

[21] N. Gendler, D. J. E. Marsh, L. McAllister and J. Moritz, *Glimmers from the Axiverse*, 2309.13145.

[22] J. Preskill, M. B. Wise and F. Wilczek, *Cosmology of the Invisible Axion*, *Phys. Lett. B* **120** (1983) 127–132.

[23] L. F. Abbott and P. Sikivie, *A Cosmological Bound on the Invisible Axion*, *Phys. Lett. B* **120** (1983) 133–136.

[24] M. Dine and W. Fischler, *The Not So Harmless Axion*, *Phys. Lett. B* **120** (1983) 137–141.

[25] D. J. E. Marsh, *Axion Cosmology*, *Phys. Rept.* **643** (2016) 1–79, [1510.07633].

[26] F. Chadha-Day, J. Ellis and D. J. E. Marsh, *Axion dark matter: What is it and why now?*, *Sci. Adv.* **8** (2022) abj3618, [2105.01406].

[27] M. J. Stott and D. J. E. Marsh, *Black hole spin constraints on the mass spectrum and number of axionlike fields*, *Phys. Rev. D* **98** (2018) 083006, [1805.02016].

[28] V. M. Mehta, M. Demirtas, C. Long, D. J. E. Marsh, L. McAllister and M. J. Stott, *Superradiance in string theory*, *JCAP* **07** (2021) 033, [2103.06812].

[29] M. J. Stott, D. J. E. Marsh, C. Pongkitivanichkul, L. C. Price and B. S. Acharya, *Spectrum of the axion dark sector*, *Phys. Rev. D* **96** (2017) 083510, [1706.03236].

[30] M. Reig, *The stochastic axiverse*, *JHEP* **09** (2021) 207, [2104.09923].

[31] J. E. Kim, H. P. Nilles and M. Peloso, *Completing natural inflation*, *JCAP* **01** (2005) 005, [hep-ph/0409138].

[32] S. Dimopoulos, S. Kachru, J. McGreevy and J. G. Wacker, *N-flation*, *JCAP* **08** (2008) 003, [hep-th/0507205].

[33] P. Agrawal, J. Fan, M. Reece and L.-T. Wang, *Experimental Targets for Photon Couplings of the QCD Axion*, *JHEP* **02** (2018) 006, [1709.06085].

[34] N. Kitajima and F. Takahashi, *Resonant conversions of QCD axions into hidden axions and suppressed isocurvature perturbations*, *JCAP* **01** (2015) 032, [1411.2011].

[35] I. Obata, *Implications of the cosmic birefringence measurement for the axion dark matter search*, *JCAP* **09** (2022) 062, [2108.02150].

[36] T. Higaki, N. Kitajima and F. Takahashi, *Hidden axion dark matter decaying through mixing with QCD axion and the 3.5 keV X-ray line*, *JCAP* **12** (2014) 004, [1408.3936].

[37] N. Glennon, N. Musoke and C. Prescod-Weinstein, *Simulations of multifield ultralight axionlike dark matter*, *Phys. Rev. D* **107** (2023) 063520, [2302.04302].

[38] B. Gavela, P. Quílez and M. Ramos, *Multiple QCD axion*, 2305.15465.

[39] F. Chadha-Day, *Axion-like particle oscillations*, *JCAP* **01** (2022) 013, [2107.12813].

[40] Z. Chen, A. Kobakhidze, C. A. J. O'Hare, Z. S. C. Picker and G. Pierobon, *Phenomenology of the companion-axion model: photon couplings*, *Eur. Phys. J. C* **82** (2022) 940, [2109.12920].

[41] D. Cyncynates, T. Giurgica-Tiron, O. Simon and J. O. Thompson, *Resonant nonlinear pairs in the axiverse and their late-time direct and astrophysical signatures*, *Phys. Rev. D* **105** (2022) 055005, [2109.09755].

[42] S. Gasparotto and E. I. Sfakianakis, *Cosmic Birefringence from the Axiverse*, 2306.16355.

[43] D. Cyncynates, O. Simon, J. O. Thompson and Z. J. Weiner, *Nonperturbative structure in coupled axion sectors and implications for direct detection*, *Phys. Rev. D* **106** (2022) 083503, [2208.05501].

[44] D. Cyncynates and J. O. Thompson, *Heavy QCD axion dark matter from avoided level crossing*, *Phys. Rev. D* **108** (2023) L091703, [2306.04678].

[45] K. R. Dienes, E. Dudas and T. Gherghetta, *Invisible axions and large radius compactifications*, *Phys. Rev. D* **62** (2000) 105023, [hep-ph/9912455].

[46] A. Ayala, I. Domínguez, M. Giannotti, A. Mirizzi and O. Straniero, *Revisiting the bound on axion-photon coupling from Globular Clusters*, *Phys. Rev. Lett.* **113** (2014) 191302, [1406.6053].

[47] G. G. Raffelt, *Axion Constraints From White Dwarf Cooling Times*, *Phys. Lett. B* **166** (1986) 402–406.

[48] S. I. Blinnikov and N. V. Dunina-Barkovskaya, *The cooling of hot white dwarfs: a theory with non-standard weak interactions, and a comparison with observations*, *Mon. Not. Roy. Astron. Soc.* **266** (1994) 289–304.

[49] J. Halverson and F. Ruehle, *Computational Complexity of Vacua and Near-Vacua in Field and String Theory*, *Phys. Rev. D* **99** (2019) 046015, [1809.08279].

[50] J. Halverson, C. Long, B. Nelson and G. Salinas, *Towards string theory expectations for photon couplings to axionlike particles*, *Phys. Rev. D* **100** (2019) 106010, [1909.05257].

[51] L. J. Hall, H. Murayama and N. Weiner, *Neutrino mass anarchy*, *Phys. Rev. Lett.* **84** (2000) 2572–2575, [hep-ph/9911341].

[52] N. Haba and H. Murayama, *Anarchy and hierarchy*, *Phys. Rev. D* **63** (2001) 053010, [hep-ph/0009174].

[53] A. de Gouvea and H. Murayama, *Statistical test of anarchy*, *Phys. Lett. B* **573** (2003) 94–100, [hep-ph/0301050].

[54] A. de Gouvea and H. Murayama, *Neutrino Mixing Anarchy: Alive and Kicking*, *Phys. Lett. B* **747** (2015) 479–483, [1204.1249].

[55] J. R. Espinosa, *Anarchy in the neutrino sector?*, hep-ph/0306019.

[56] J. Heeck, *Seesaw parametrization for n right-handed neutrinos*, *Phys. Rev. D* **86** (2012) 093023, [1207.5521].

[57] Y. Bai and G. Torroba, *Large N (=3) Neutrinos and Random Matrix Theory*, *JHEP* **12** (2012) 026, [1210.2394].

[58] X. Lu and H. Murayama, *Neutrino Mass Anarchy and the Universe*, *JHEP* **08** (2014) 101, [1405.0547].

[59] J.-F. Fortin, N. Giasson and L. Marleau, *Probability density function for neutrino masses and mixings*, *Phys. Rev. D* **94** (2016) 115004, [1609.08581].

[60] J.-F. Fortin, N. Giasson and L. Marleau, *Anarchy and Neutrino Physics*, *JHEP* **04** (2017) 131, [1702.07273].

[61] J.-F. Fortin, N. Giasson and L. Marleau, *Probability Density Functions for CP-Violating Rephasing Invariants*, *Nucl. Phys. B* **930** (2018) 384–398, [1801.10165].

[62] J.-F. Fortin, N. Giasson, L. Marleau and J. Pelletier-Dumont, *Mellin transform approach to rephasing invariants*, *Phys. Rev. D* **102** (2020) 036001, [2004.14284].

[63] G. G. Raffelt, *ASTROPHYSICAL AXION BOUNDS DIMINISHED BY SCREENING EFFECTS*, *Phys. Rev. D* **33** (1986) 897.

[64] L. M. Krauss, J. E. Moody and F. Wilczek, *A STELLAR ENERGY LOSS MECHANISM INVOLVING AXIONS*, *Phys. Lett. B* **144** (1984) 391–394.

[65] S. Dimopoulos, G. D. Starkman and B. W. Lynn, *Atomic Enhancements in the Detection of Weakly Interacting Particles*, *Phys. Lett. B* **168** (1986) 145–150.

[66] S. Dimopoulos, J. A. Frieman, B. W. Lynn and G. D. Starkman, *Axiorecombination: A New Mechanism for Stellar Axion Production*, *Phys. Lett. B* **179** (1986) 223–227.

[67] M. Pospelov, A. Ritz and M. B. Voloshin, *Bosonic super-WIMPs as keV-scale dark matter*, *Phys. Rev. D* **78** (2008) 115012, [0807.3279].

[68] K. O. Mikaelian, *Astrophysical Implications of New Light Higgs Bosons*, *Phys. Rev. D* **18** (1978) 3605.

[69] M. Fukugita, S. Watamura and M. Yoshimura, *Light Pseudoscalar Particle and Stellar Energy Loss*, *Phys. Rev. Lett.* **48** (1982) 1522.

[70] M. Fukugita, S. Watamura and M. Yoshimura, *Astrophysical Constraints on a New Light Axion and Other Weakly Interacting Particles*, *Phys. Rev. D* **26** (1982) 1840.

[71] J. Redondo, *Solar axion flux from the axion-electron coupling*, *JCAP* **12** (2013) 008, [1310.0823].

[72] I. G. Irastorza et al., *Towards a new generation axion helioscope*, *JCAP* **06** (2011) 013, [1103.5334].

[73] S. Hoof, J. Jaeckel and L. J. Thormaehlen, *Quantifying uncertainties in the solar axion flux and their impact on determining axion model parameters*, *JCAP* **09** (2021) 006, [2101.08789].

[74] K. Barth et al., *CAST constraints on the axion-electron coupling*, *JCAP* **05** (2013) 010, [1302.6283].

[75] A. Harayama, Y. Terada, M. Ishida, T. Hayashi, A. Bamba and M. S. Tashiro, *Search for non-thermal emissions from an isolated magnetic white dwarf, euve j0317 855, with suzaku*, *Publications of the Astronomical Society of Japan* **65** (2013) 73.

[76] B. Kulebi, S. Jordan, E. Nelan, U. Bastian and M. Altmann, *Constraints on the origin of the massive, hot, and rapidly rotating magnetic white dwarf RE J 0317-853 from an HST parallax measurement*, *Astron. Astrophys.* **524** (2010) A36, [1007.4978].

[77] G. I. Rubtsov and S. V. Troitsky, *Breaks in gamma-ray spectra of distant blazars and transparency of the Universe*, *Soviet Journal of Experimental and Theoretical Physics Letters* **100** (Nov., 2014) 355–359, [1406.0239].

[78] K. Kohri and H. Kodama, *Axion-like particles and recent observations of the cosmic infrared background radiation*, *Physical Review D* **96** (sep, 2017) .

[79] J. Biteau and D. A. Williams, *THE EXTRAGALACTIC BACKGROUND LIGHT, THE HUBBLE CONSTANT, AND ANOMALIES: CONCLUSIONS FROM 20 YEARS OF TeV GAMMA-RAY OBSERVATIONS*, *The Astrophysical Journal* **812** (oct, 2015) 60.

[80] A. Domínguez and M. Ajello, *Spectral analysis of Fermi-LAT blazars above 50 GeV*, *Astrophys. J. Lett.* **813** (2015) L34, [1510.07913].

[81] A. De Angelis, G. Galanti and M. Roncadelli, *Relevance of axion-like particles for very-high-energy astrophysics*, *Phys. Rev. D* **84** (2011) 105030, [1106.1132].

[82] M. Simet, D. Hooper and P. D. Serpico, *Milky way as a kiloparsec-scale axionscope*, *Physical Review D* **77** (mar, 2008) .

[83] M. A. Sanchez-Conde, D. Paneque, E. Bloom, F. Prada and A. Dominguez, *Hints of the existence of Axion-Like-Particles from the gamma-ray spectra of cosmological sources*, *Phys. Rev. D* **79** (2009) 123511, [0905.3270].

[84] M. Meyer, D. Horns and M. Raue, *First lower limits on the photon-axion-like particle coupling from very high energy gamma-ray observations*, *Physical Review D* **87** (feb, 2013) .

[85] E. Masaki, A. Aoki and J. Soda, *Photon-axion conversion, magnetic field configuration, and polarization of photons*, *Physical Review D* **96** (aug, 2017) .

[86] A. Franceschini, G. Rodighiero and M. Vaccari, *Extragalactic optical-infrared background radiation, its time evolution and the cosmic photon-photon opacity*, *Astronomy & Astrophysics* **487** (jun, 2008) 837–852.

[87] R. Jansson and G. R. Farrar, *A NEW MODEL OF THE GALACTIC MAGNETIC FIELD*, *The Astrophysical Journal* **757** (aug, 2012) 14.

[88] G. Guennebaud, B. Jacob et al., "Eigen v3." http://eigen.tuxfamily.org, 2010.

[89] M. Meyer, *The Opacity of the Universe for High and Very High Energy γ-Rays*, Ph.D. thesis, Hamburg U., 2013.

[90] ADMX collaboration, C. Boutan et al., *Piezoelectrically Tuned Multimode Cavity Search for Axion Dark Matter*, *Phys. Rev. Lett.* **121** (2018) 261302, [1901.00920].

[91] J. A. Wheeler and K. Ford, *Geons, black holes, and quantum foam: A life in physics*. 1998.

[92] M. O. Tahim, R. R. Landim and C. A. S. Almeida, *Spacetime as a deformable solid*, 2007.

[93] E. T. Newman, R. Couch, K. Chinnapared, A. Exton, A. Prakash and R. Torrence, *Metric of a Rotating, Charged Mass*, *J. Math. Phys.* **6** (1965) 918–919.

[94] K. Clough, *Continuity equations for general matter: applications in numerical relativity*, *Classical and Quantum Gravity* **38** (July, 2021) 167001.

[95] T. Andrade, L. A. Salo, J. C. Aurrekoetxea, J. Bamber, K. Clough, R. Croft et al., *Grchombo: An adaptable numerical relativity code for fundamental physics*, *Journal of Open Source Software* **6** (2021) 3703.

[96] K. Clough, *Scalar fields in numerical general relativity*, 2017.

[97] D. Martin, M. Adams, P. Colella, D. Graves, H. Johansen, J. Johnson et al., *Chombo Design Document*, .

[98] M. Berger and I. Rigoutsos, *An algorithm for point clustering and grid generation*, *IEEE Transactions on Systems, Man, and Cybernetics* **21** (1991) 1278–1286.

[99] L. I. Petrich, S. L. Shapiro, R. F. Stark and S. A. Teukolsky, *Accretion onto a Moving Black Hole: A Fully Relativistic Treatment*, **336** (Jan., 1989) 313.

[100] J. C. Aurrekoetxea, S. E. Brady, L. Aresté-Saló, J. Bamber, L. Chung-Jukko, K. Clough et al., *Grtresna: An open-source code to solve the initial data constraints in numerical relativity*, 2025.

[101] R. Brito, V. Cardoso and P. Pani, *Superradiance: New Frontiers in Black Hole Physics*. Springer International Publishing, 2020, 10.1007/978-3-030-46622-0.

[102] R. H. Dicke, *Coherence in spontaneous radiation processes*, *Phys. Rev.* **93** (Jan, 1954) 99–110.

[103] K. Clough, T. Helfer, H. Witek and E. Berti, *Ghost Instabilities in Self-Interacting Vector Fields: The Problem with Proca Fields*, *Phys. Rev. Lett.* **129** (2022) 151102, [2204.10868].

[104] S. Chandrasekhar, *Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction.*, **97** (Mar., 1943) 255.

[105] M. A. Ruderman and E. A. Spiegel, *Galactic Wakes*, **165** (Apr., 1971) 1.

[106] Y. Rephaeli and E. E. Salpeter, *Flow past a massive object and the gravitational drag*, **240** (Aug., 1980) 20–24.

[107] E. C. Ostriker, *Dynamical friction in a gaseous medium*, *The Astrophysical Journal* **513** (Mar., 1999) 252–258.

[108] D. Traykova, R. Vicente, K. Clough, T. Helfer, E. Berti, P. G. Ferreira et al., *Relativistic drag forces on black holes from scalar dark matter clouds of all sizes*, *Physical Review D* **108** (Dec., 2023) .

[109] N. Speeney, A. Antonelli, V. Baibhav and E. Berti, *Impact of relativistic corrections on the detectability of dark-matter spikes with gravitational waves*, *Physical Review D* **106** (Aug., 2022) .

[110] L. Hui, J. P. Ostriker, S. Tremaine and E. Witten, *Ultralight scalars as cosmological dark matter*, *Physical Review D* **95** (Feb., 2017) .

[111] D. Syer, *Relativistic dynamical friction in the weak scattering limit*, *Monthly Notices of the Royal Astronomical Society* **270** (Sept., 1994) 205–208.

[112] E. C. Ostriker, *Dynamical friction in a gaseous medium*, *The Astrophysical Journal* **513** (Mar., 1999) 252–258.

[113] M. A. Ruderman and E. A. Spiegel, *Galactic Wakes*, **165** (Apr., 1971) 1.

[114] Y. Rephaeli and E. E. Salpeter, *Flow past a massive object and the gravitational drag*, **240** (Aug., 1980) 20–24.

[115] H. Bondi, *On spherically symmetrical accretion*, **112** (Jan., 1952) 195.

[116] J. C. Aurrekoetxea, J. Bamber, S. E. Brady, K. Clough, T. Helfer, J. Marsden et al., *Grdzhadzha: A code for evolving relativistic matter on analytic metric backgrounds*, 2023.

[117] Message Passing Interface Forum, *MPI: A Message-Passing Interface Standard Version 4.0*, June, 2021.

[118] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim et al., *Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc*, *Physics Letters B* **716** (Sept., 2012) 1–29.

[119] S. Chatrchyan, V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo et al., *Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc*, *Physics Letters B* **716** (Sept., 2012) 30–61.

[120] P. Athron, C. Balázs, T. Bringmann, A. Buckley, M. Chrząszcz, J. Conrad et al., *A global fit of the mssm with gambit*, *The European Physical Journal C* **77** (Dec., 2017) .

[121] ATLAS collaboration, G. Aad et al., *ATLAS Run 2 searches for electroweak production of supersymmetric particles interpreted within the pMSSM*, *JHEP* **05** (2024) 106, [2402.01392].

[122] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, S. H. Abidi et al., *Search for supersymmetry in final states with charm jets and missing transverse momentum in 13 tev pp collisions with the atlas detector*, *Journal of High Energy Physics* **2018** (Sept., 2018) .

[123] ATLAS collaboration, G. Aad et al., *Search for a heavy Higgs boson decaying into a Z boson and another heavy Higgs boson in the $\ell\ell bb$ and $\ell\ell WW$ final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) 396, [2011.05639].

[124] ATLAS Collaboration collaboration, M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, S. H. Abidi et al., *Search for low-mass dijet resonances using trigger-level jets with the atlas detector in pp collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. Lett.* **121** (Aug, 2018) 081801.

[125] L. Masetti, *Atlas results and prospects with focus on beyond the standard model*, *Nuclear and Particle Physics Proceedings* **303-305** (2018) 43–48.

[126] *High-Luminosity Large Hadron Collider (HL-LHC) : Preliminary Design Report*, .

[127] J. de Blas et al., *Higgs Boson Studies at Future Particle Colliders*, *JHEP* **01** (2020) 139, [1905.03764].

[128] G. Heinrich, S. Jones, M. Kerner, T. Stone and A. Vestner, *Electroweak corrections to Higgs boson pair production: the top-Yukawa and self-coupling contributions*, *JHEP* **11** (2024) 040, [2407.04653].

[129] Laporta*International Journal of Modern Physics A* **15** (2000) 5087.

[130] J. Klappert, F. Lange, P. Maierhöfer and J. Usovitsch, *Integral reduction with kira 2.0 and finite field methods*, *Computer Physics Communications* **266** (Sept., 2021) 108024.

[131] P. Maierhöfer, J. Usovitsch and P. Uwer, *Kira—a feynman integral reduction program*, *Computer Physics Communications* **230** (Sept., 2018) 99–112.

[132] V. Magerya, *Rational tracer: a tool for faster rational function reconstruction*, 2022.

[133] J. Klappert and F. Lange, *Reconstructing rational functions with firefly*, *Computer Physics Communications* **247** (Feb., 2020) 106951.

[134] J. Klappert, S. Y. Klein and F. Lange, *Interpolation of dense and sparse rational functions and other improvements in firefly*, *Computer Physics Communications* **264** (July, 2021) 107968.

[135] G. HEINRICH, *Sector decomposition*, *International Journal of Modern Physics A* **23** (Apr., 2008) 1457–1486.

[136] R. Zippel, *Interpolating polynomials from their values*, *Journal of Symbolic Computation* **9** (1990) 375–403.

[137] N. Bell and J. Hoberock, *Thrust: A productivity-oriented library for cuda*, 2012, https://api.semanticscholar.org/CorpusID:60775233.

[138] J. W. Cooley and J. W. Tukey, *An algorithm for the machine calculation of complex fourier series*, *Mathematics of Computation* **19** (1965) 297–301.

[139] W. R. Inc., "Mathematica, Version 14.2."

[140] J. Nickolls, I. Buck, M. Garland and K. Skadron, *Scalable Parallel Programming with CUDA*, *Queue* **6** (2008) 40–53.