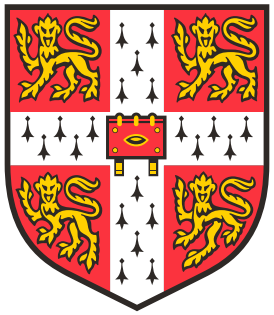
MOTION IN A SCALAR FIELD



LEONG KHIM WONG

Department of Applied Mathematics and Theoretical Physics
Trinity College, University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

July 2020

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other university or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other university or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the Mathematics Degree Committee.

Leong Khim Wong

MOTION IN A SCALAR FIELD

Leong Khim Wong

Abstract

Light scalar fields emerge as a generic prediction in physics beyond the Standard Model. For example, they arise as new degrees of freedom in modified gravity, as Kaluza–Klein modes from extra compactified dimensions, and as Nambu–Goldstone bosons from spontaneously broken symmetries. Far from being just objects of theoretical interest, these scalar fields could also play crucial roles in resolving some of the most important open problems, such as the nature of dark matter and dark energy. Given this ubiquity in modern theoretical physics and their potentially far-reaching implications, efforts to detect or otherwise rule out these hypothetical scalars have burgeoned into a global enterprise in recent years. This thesis contributes to this ongoing effort by updating our understanding of how light scalar fields influence the dynamics of moving bodies, focusing on two novel scenarios.

We begin by reanalysing the motion of electrons in laboratory experiments designed to deliver high-precision measurements of the fine-structure constant. The vacuum chambers employed in these setups make them ideal testing grounds for a class of scalar–tensor theories that screen the effects of their scalar mode based on the ambient density. If unscreened, the scalar exerts an attractive "fifth" force on the electron and, moreover, transforms the vacuum cavity into a dielectric medium due to its interactions with electromagnetic fields. Because these effects introduce different amounts of systematic bias into each experiment, good agreement between independent measurements of the fine-structure constant can be used to establish meaningful constraints on the parameter spaces of these models.

In the second part of this thesis, we turn to investigate how ambient scalar fields influence the motion of binary black holes. Even though the models we consider are subject to no-hair theorems, the interplay between absorption at the horizons and momentum transfer in the bulk of the spacetime still gives rise to interesting phenomenology. We show that this interaction causes a fraction of the ambient field to be ejected from the system as scalar radiation, while the black holes themselves are seen to feel the effects of an emergent fifth force. Moreover, if the ambient field corotates with the binary, it can extract energy from the orbital motion and grow exponentially through a process akin to superradiance. Although these effects turn out to be highly suppressed in the regime amenable to analytic methods, the novel techniques developed herein lay the groundwork for future studies of these complex gravitational systems.

Preface

The material presented in this thesis is based on five publications that are the result of work undertaken at the University of Cambridge between October 2016 and March 2020:

[1] L. K. Wong and A.-C. Davis, One-electron atoms in screened modified gravity, Phys. Rev. D **95**, 104010 (2017).

[2] P. Brax, A.-C. Davis, B. Elder, and L. K. Wong, Constraining screened fifth forces with the electron magnetic moment, Phys. Rev. D **97**, 084050 (2018).

[3] L. K. Wong, A.-C. Davis, and R. Gregory, Effective field theory for black holes with induced scalar charges, Phys. Rev. D **100**, 024010 (2019).

[4] L. K. Wong, Superradiant scattering by a black hole binary, Phys. Rev. D **100**, 044051 (2019).

[5] L. K. Wong, Evolution of diffuse scalar clouds around binary black holes, Phys. Rev. D **101**, 124049 (2020).

References [1] and [2] form the basis of the first half of this thesis, with Ref. [1] making up all of Chapter 3 and some of Chapter 5, while Ref. [2] makes up all of Chapter 4 and most of Chapter 5. As for the second half, the results in Ref. [3] are spread across Chapters 6 to 8, whereas Chapter 9 reproduces the work published in Refs. [4] and [5].

It must be said that none of these results would have been possible without the advice, guidance, and support I have received from my supervisor, Anne-Christine Davis, to whom I owe a tremendous amount of gratitude. Thanks is also due to my collaborators, Philippe Brax, Benjamin Elder, and Ruth Gregory, who — along with Anne — have helped me achieve this milestone in my career. Benjamin is also owed another expression of gratitude for performing the numerical calculations described in Chapter 4.

Within the University of Cambridge, I wish to thank John Barrow, Harvey Reall, Jorge Santos, and Ulrich Sperhake for the invariably enlightening discussions that have impacted my research in one way or another. Further afield, I am also pleased to be able to thank Cliff Burgess, Clare Burrage,

Vitor Cardoso, Sam Dolan, Joe Keir, Eugene Lim, David Marsh, Ira Rothstein, and Andrew Tolley for similarly helpful discussions. For their moral support and camaraderie, my friends in Pavilion B — Amelia, Bogdan, Ed, Josh, Muntazir, Nakarin, Oliver, Sam, and Theo — also deserve a mention, as does Amanda Stagg; our wonderful group administrator.

At this stage, I would be remiss not to thank my dear friends Jish, Sara, and Shivy, who are scattered across the globe, and my friends closer to home at the Cambridge University Malaysian Society, for never failing to offer a welcome respite from the trials of life as a graduate student. Finally, I am indebted to my parents, Koon Tatt and Cheng Woei, and my siblings, Mei and Shaun, for their unconditional love and unwavering support.

The work in this thesis was funded primarily by a Cambridge International Scholarship from the Cambridge Commonwealth, European and International Trust. In addition, I received funding in the form a Research Scholarship and multiple Rouse Ball travel grants from Trinity College, Cambridge, and was also supported by a research studentship award (Ref. S52/064/19) from the Cambridge Philosophical Society. As a member of the department, I have benefited from the following STFC Consolidated Grants: No. ST/L000385/1, No. ST/L000636/1, No. ST/P000673/1, and No. ST/P000681/1.

Contents

Chapter 1
Introduction

Fundamental physics has proven to be both tremendously successful and also gravely incomplete. On the one hand, it offers an incredibly detailed picture of how this Universe works, accounting for everything from the properties of subatomic particles like quarks to the processes around astrophysical giants like quasars. In the laboratory, high-precision experiments refined over the course of decades have now validated the accuracy of QED to the level of about one part per billion [6], while the nascent rise of gravitational-wave astronomy has opened a new window into the highly dynamical, strong-field regime of gravity; further establishing the efficacy of Einstein's general theory of relativity in the process [7–10].

These successes notwithstanding, there are still compelling reasons to believe that general relativity and the Standard Model are not the end of the story. When confronted with observations on cosmological scales, for instance, these cornerstones of modern physics suggest that *we* — by which I mean all matter composed of the Standard Model particles — make up only a meagre 5% of the energy density today [11]. A further 26% is comprised of what is known as dark matter: an extra component that cannot be seen directly, but whose gravitational effects are crucial for seeding the distribution of large-scale structure and sustaining the rotational velocities of galaxies. The final remaining 69%, meanwhile, is dominated by what is presently called dark energy, and is responsible for driving the late-time accelerated expansion of this Universe. While the precise nature and physical origin of this dark sector remain unknown, there is little doubt today that the solution must lie beyond the Standard Model.

At the other end of the spectrum, our understanding of physics is also lacking on the smallest of scales. General relativity is nonrenormalisable when treated as a quantum field theory [12]; hence, the proliferation of ultraviolet (UV) divergences can be controlled only if Einstein's theory is just the leading term in a derivative expansion involving all possible local interactions [13, 14]. Within this

effective field theory (EFT) framework, quantum corrections at low energies and small curvatures are well understood, but little can be said about physics at very high energies. This presents a problem since such high-energy scales are relevant at least during the first few moments in our Universe's history, and also within the interiors of black holes. All of this points to the need for a UV completion to general relativity, but the search for a theory of quantum gravity remains arguably the biggest open problem in theoretical physics.

Interestingly, while the origin of the dark sector and the nature of quantum gravity appear to be two, fundamentally distinct mysteries, involving different scales separated by many orders of magnitude, their potential solutions could be intimately related. Many candidate theories that seek to extend general relativity and the Standard Model — either in the UV or in the infrared (IR) — inevitably introduce new, light scalar degrees of freedom. This is good news because the existence of a new, light particle in the Universe is a falsifiable prediction. It is then no surprise that efforts to detect or otherwise rule out these hypothetical scalars have flourished into a burgeoning enterprise in recent decades.

The goal of this thesis is to support this endeavour by updating our understanding of how light scalar fields influence the dynamics of moving bodies. We will be interested in two broad classes of models that have the potential to exhibit novel effects: Scalar fields that arise in alternative theories of gravity will be the subject of the first half of this thesis, whereas in the second half, we will focus on the impact of scalar fields that emerge after compactification in string theory. Some, though not all, of these scalars have the potential to serve as either dark matter or dark energy candidates. To set the stage, let us begin with a broad overview of these different models.

1.1 Scalars from modified gravity

All cosmological observations to date are well described by the ΛCDM model,[1] which contends that the late-time acceleration of this Universe is driven by a cosmological constant. While this minimalism is certainly appealing, questions about the nature of dark energy cannot be addressed without alternative models that can serve as a basis for comparison. Moreover, because the relatively large uncertainties in the data do not preclude such a possibility at present [20, 21], it is interesting to explore models in which dark energy has a dynamical component.[2] The simplest models, which go by the name

[1] One notable exception is a 4.4σ tension between local measurements of the Hubble constant and the value inferred from fitting ΛCDM to Planck data [15–19]. It remains unclear if this discrepancy is due to unknown systematics or is a sign of new physics; hence, we will not dwell too much on this issue here.

[2] Another motivation for contemplating alternatives to ΛCDM is the cosmological constant problem [22]: This quantity receives radiative corrections from matter fields, whose contributions to the vacuum energy density $\rho_{\rm vac}$ scale with the fourth power of their mass. Thus, even if the electron were the heaviest particle

of quintessence [23–25], are reminiscent of inflation in that late-time acceleration is attributed to a minimally coupled scalar field slow-rolling down its potential. More recent attempts at explaining dark energy, meanwhile, have taken a less conventional approach. In the last two decades, there has been a surge in development of models that account for late-time acceleration by introducing IR modifications to the laws of gravity.

On the surface, theories that modify gravity can be classified according to how they violate Lovelock's theorem [26–28]. (See Ref. [29] for a comprehensive review.) Roughly speaking, it states that general relativity is the only four-dimensional theory of gravity derivable from an action principle that is local, diffeomorphism invariant, built from only the metric, and has second-order equations of motion. Thus, in addition to establishing the remarkable uniqueness of Einstein's theory, Lovelock's theorem implies that an alternative theory of gravity follows from relaxing any one or more of its assumptions. To remain somewhat conservative, it is common to retain locality and the use of an action principle; hence, we are left with four options: (1) introduce new fields, (2) alter the number of spacetime dimensions, (3) break diffeomorphism invariance, and/or (4) allow for higher-order field equations.

While these four options are inherently distinct, they turn out to be phenomenologically equivalent [30]. Case in point, models based on string theory invariably need to be compactified down to four dimensions to be observationally viable, at which point the extra dimensions emerge in the low-energy EFT as Kaluza–Klein modes. Thus, from a phenomenological standpoint, adding extra dimensions is no different from introducing new fields. In a similar vein, theories that explicitly break diffeomorphism invariance can be reformulated as a diffeomorphism-invariant theory with an extended field content via the Stueckelberg mechanism [31]. (An example is massive gravity [32, 33].) Theories with higher-order field equations, on the other hand, are typically pathological because they contain Ostrogradsky instabilities [34, 35], but those that manage to evade these issues turn out to be equivalent to second-order theories with additional fields. (An example being $f(R)$ gravity [36].) The takeaway message should now be clear: At the phenomenological level, modifying general relativity essentially

---

in the Universe, its contribution $\sim m_e^4 = (511\ \text{keV})^4$ is already 33 orders of magnitude greater than the observed value $\rho_{\text{vac}} \approx (2.4\ \text{meV})^4$, indicating a severe fine-tuning problem. Unfortunately, the theories that we consider in Part I of this thesis do nothing to address this cosmological constant problem directly, but instead assume that an eventual solution will ultimately set $\rho_{\text{vac}} = 0$. They then aim to provide an alternative mechanism for driving late-time acceleration. Such a two-pronged approach could still be compelling were it not for the fact that these alternatives are also plagued by their own fine-tuning issues. Thus, at least from a naturalness standpoint, these models offer little advantage over simply tolerating a fine-tuned cosmological constant. What they do offer, however, is a useful *phenomenological* framework for exploring what kinds of interesting effects can arise when new, dynamical degrees of freedom are present. This will be our key motivation for continuing to study these kinds of theories, and there will be more to say in Section 1.1.2 and also at the end of Chapter 4.1.

amounts to introducing new fields. The simplest and best studied models are so-called *scalar–tensor theories*, which extend general relativity by introducing a single, new scalar degree of freedom.

The prototypical example of a scalar–tensor theory emerged from the halls of Princeton University in 1961 as an attempt to better incorporate Mach's principle [37] into general relativity. The Brans–Dicke theory [38, 39], as it is known today, introduces a massless scalar field that couples to matter and, consequently, mediates an additional gravitational-like force. The strength of this "fifth" force — so named as to distinguish it from the other four known fundamental forces of nature[3] — is set by a single free parameter in the theory, which local tests of gravity strongly constrain [52, 53]. Most notably, the Lunar Laser Ranging experiment [54] and redshift measurements of radio waves transmitted to and from the Cassini spacecraft [55] constrain deviations from the quasistatic, weak-field limit of general relativity to be less than about one part in 10^4 within the Solar System. (See Refs. [56–58] for comprehensive reviews on model-independent tests of gravity).

These stringent experimental bounds now present an interesting obstacle that a viable dark energy candidate must overcome. Because the scalar has to be cosmologically light if its evolution is to drive late-time acceleration, it is effectively massless on Solar System scales and is therefore subject to these constraints from local tests of gravity. Of course, one way of skirting these bounds is to simply posit a scalar that is decoupled from matter and so does not mediate a fifth force, as is done in quintessence scenarios, but such an arrangement appears to be unnatural in the context of the renormalisation group. One typically expects the scalar to couple to all of the other fields in the theory unless a symmetry prohibiting such interactions is present. This intuition is supported by string theory, wherein at least one scalar field — the dilaton — is known to have a Brans–Dicke-like kinetic term with a coupling constant of approximately gravitational strength [59–61].

Modern scalar–tensor theories therefore circumvent these constraints by instead relying on self-interactions, which can lead to a phenomenology that is markedly different from the Brans–Dicke case. A nonlinear effect known as *screening* is a generic feature in many of these models that allows the scalar to adjust the strength and/or range of its fifth force dynamically based on the local environment. In so doing, the scalar is able to suppress its effects in and around dense objects, like the Earth and

[3] It is interesting to note that the term "fifth force" first appeared in a *New York Times* article [40] reporting on a completely different proposal for a new fundamental force in nature. In 1986, Fischbach and his collaborators [41] suggested that data from the Eötvös experiment [42], anomalies in neutral kaon experiments [43, 44], and geophysical measurements of Newton's gravitational constant [45, 46] were together hinting at the existence of a new massive vector boson that coupled to either baryon number or hypercharge. Naturally, their proposal incited an onslaught of new experiments that were designed to independently search for this novel interaction, but all bar one reported no significant departure from Newtonian gravity. Within a decade, the experimental evidence was insurmountable: Fischbach et al.'s fifth force was dead. See Refs. [47–51] for reviews on this period in history.

the Sun, thereby satisfying all of the aforementioned tests of gravity, while simultaneously remaining light on cosmological scales, where it can couple to matter with gravitational strength. If this were the end of it, such models would be virtually indistinguishable from ΛCDM on small scales, but as it turns out, screening mechanisms also trigger an array of novel signatures in both laboratory and astrophysical settings. Consequently, the study of screening has inspired an increasing number of new experiments that test the laws of gravity in innovative ways. Seeing as no significant departure from general relativity has been detected to date, data from these experiments currently place strict limits on the models that remain observationally viable. The literature on this subject is vast and is expertly reviewed in Refs. [62–68]. Below, I briefly highlight just a few key predictions that will be important in later chapters.

1.1.1 Chameleon-like theories

While the specific details of how screening operates can vary considerably from model to model, it is useful to group different scalar–tensor theories into broad categories based on similarities in their phenomenology. The first of three categories we will discuss is comprised of theories that utilise nonderivative self-interactions to suppress the effects of its scalar mode in high-density environments. Such theories are said to be chameleon-like. This category includes Khoury and Weltman's eponymous chameleon model [69, 70], Hinterbichler and Khoury's symmetron model [71, 72], and Brax et al.'s environmentally-dependent dilaton model [73].

Three novel effects specific to these theories are described below, but first, it is necessary to introduce some important terminology. Consider embedding some object of finite size R within a background environment that varies on a length scale $\gg R$. This body is said to possess a *scalar charge* Q if the scalar-field profile around it is of the form $\phi \sim \Phi + Qe^{-m_{\rm eff}r}/(4\pi r)$ when viewed from a distance $r \gg R$. The field in the absence of this body is here denoted by Φ, while $m_{\rm eff}$ is the scalar's local effective mass in this environment. It follows that, analogous to the electrostatic case, the fifth force between two such objects has magnitude $Q_1 Q_2 e^{-m_{\rm eff}r}/(4\pi r^2)$ when both objects are travelling nonrelativistically and are separated by a distance $r \gg \max(R_1, R_2)$. Moreover, if the background field Φ is spatially inhomogeneous, it too exerts a fifth force on these bodies; given by $Q\nabla\Phi$ in the nonrelativistic limit. (The relativistic version of this equation can be found in Chapter 2.)

Equivalence principle violations Because these charges do not scale linearly with mass, they lead to violations of the strong equivalence principle, which are expected to be most pronounced in systems like dwarf galaxies [74]. To elaborate, small test particles like atoms or molecules, which are not heavy

enough to drive the scalar into its nonlinear regime, do have scalar charges proportional to their masses; but, in contrast, the scalar charge of a macroscopic body like a star is sourced only by a thin shell of mass on the body's surface, as the high density of matter in its interior suppresses any effect from the scalar in this region. Of course, the precise thickness of this shell depends on both the model parameters and its internal composition [70, 72, 75–78], but suffice it to say, this "thin-shell effect" renders the strength of the fifth force between macroscopic bodies negligible in comparison to their Newtonian gravitational interaction. On the other hand, the fifth force between microscopic particles can remain comparable to Newton's inverse-square law in low-density environments.

This distinction causes the gaseous and dark matter components of a galaxy to exhibit kinematics that differ substantially from those of its stellar components. In Ref. [79], six low-surface-brightness galaxies were used to constrain chameleon-like theories on the basis that their gaseous and stellar rotation curves are consistent with the predictions of Newtonian gravity. More recent analyses, meanwhile, used a much larger sample of galaxies to constrain two other novel signatures associated with violations of the strong equivalence principle: warping of stellar disks [80] and a displacement between the centroids of the galaxy's stellar and gaseous components [81]. Quite unexpectedly, these studies find a departure from the predictions of Newtonian gravity at the level of $\sim 7\sigma$ when the data is fit to a chameleon-like model. Tantalising for sure, but it is not yet possible to claim that new physics has been discovered, given that the impact of baryonic physics on galaxy formation is still not fully understood. Future high-resolution, hydrodynamical simulations of galaxy formation are therefore expected to play an important role in assessing which of these potential explanations is most likely. (Further studies that examine the impact of a chameleon-like scalar on galactic dynamics can be found in Refs. [82–87].)

Stellar processes In the interior of a star, the presence of a nontrivial scalar-field profile alters the effective strength of gravity and therefore affects the rate at which different processes occur. Cepheid variable stars, for instance, have been shown to pulsate at a higher frequency in chameleon-like theories than in Newtonian gravity [75, 88]; resulting in important implications for the local distance ladder. If such effects were active but unaccounted for, this would bias the distance inferred from using their period–luminosity relation. In contrast, the luminosities of stars at the tip of the red-giant branch (TRGB) are set by nuclear physics, and so are largely independent of gravity [89]; hence, comparing the inferred distances to a host galaxy using both Cepheid variables and TRGB stars provides a way of probing the effects of chameleon-like theories [90].

Vacuum bubbles In the laboratory, a chameleon-like scalar develops a bubble-like profile within the confines of a vacuum chamber, resulting in a weak but distinctive force on the cavity walls. Casimir

force and torsion balance experiments, which were originally designed to search for Yukawa-like interactions on submillimetre scales, are also sensitive to this bubble-induced force and provide some of the tightest constraints on these models [91–95]. In addition, vacuum bubbles can also be probed by tracking their influence on the motion of test particles, as is done in atom interferometry [96–99] and neutron bouncing experiments [100–103].

1.1.2 Vainshtein-like theories

A Vainshtein-like theory exploits derivative self-interactions to suppress the effects of its scalar mode. It is interesting to note that although theories of this kind are typically much harder to solve, they appeared earlier in the literature than those of the previous category. In fact, the idea of a screening mechanism originated with Vainshtein [104], who argued for its existence as a resolution to the vDVZ discontinuity in massive gravity [105–107]. (vDVZ being short for van Dam, Veltman, and Zakharov.) The Vainshtein mechanism subsequently reemerged in a braneworld-type model of gravity by Dvali, Gabadadze, and Porrati [108], and has since been shown to arise in a large class of scalar–tensor theories (so-called "galileons" and their generalisations) as well [109–116].

These theories turn out to screen so efficiently that they lead to very few novel signatures on small scales when their parameters are such that they can account for late-time acceleration [66]. Nevertheless, their appeal stems from the fact that their solutions admit self-accelerating cosmologies. In other words, they can give rise to late-time acceleration purely through the dynamics of the scalar field and without the need for some kind of cosmological-constant-like contribution. However, the recent advent of gravitational-wave astronomy has facilitated an unprecedented culling of these models.

Speed of gravitational waves The cosmological solutions of many Vainshtein-like theories predict that the graviton's helicity-2 modes propagate at a different speed from that of light. Consequently, the neutron star merger event GW170817–GRB 170817A [117–119], which established that gravitational waves travel at the speed of light to within a few parts in 10^{15}, has decisively ruled out many of these models as the driving force behind late-time acceleration [120–126].

By all accounts, this is tremendous progress. As alternative models continue to be deemed unviable, we gain an increasingly refined understanding of late-time acceleration and its physical origin. This latest achievement also nicely demonstrates a key advantage of testing gravity on multiple scales. Even if a theory is consistent with large-scale cosmological observations, its incompatibility with a smaller-scale experiment offers sufficient grounds to rule it out. (Of course, a certain amount of care is needed with this kind of logic, especially when drawing conclusions based on EFTs with low cutoffs [127].)

Having said all of this, it must be emphasised that such constraints do not necessarily spell the end for Vainshtein-like theories. While it would certainly be nice and convenient if the scalar fields we studied served some useful purpose, like explaining dark energy, nature need not conform to our wishes. Even if it cannot give rise to self-acceleration, a Vainshtein-like theory could still be active on smaller scales. In fact, this has been the attitude towards chameleon-like theories for some time now. It was recognised very early on that these theories could not self-accelerate while also remaining compatible with observational constraints [128, 129]; hence, they can only give rise to late-time acceleration through a cosmological-constant-like contribution. This makes them no better for cosmology than ΛCDM, but the chameleon mechanism and its close relatives continue to be of interest because they offer novel ways to hide light scalars. An open possibility is that new physics could be lurking in regimes we have not yet thought to look. When combined with the fact that light scalar fields appear to be a generic feature in fundamental theories like string theory, experiments that test the laws of gravity can be seen to offer the prospect of probing new physics associated with very high energy scales, and may provide clues as to the nature of quantum gravity.

1.1.3 Theories with strong-field deviations

In this spirit, it is interesting to also consider a different class of models that have been constructed not for cosmological purposes, but to introduce deviations from general relativity in the strong-field regime.[4]

Scalarised neutron stars The Damour–Esposito-Farèse (DEF) model [136–138] is a scalar–tensor theory whose scalar mode is dormant in the weak-field regime, thus guaranteeing no deviation from general relativity in the laboratory and Solar System. However, the scalar is subject to a linear tachyonic instability that triggers novel effects in regions of extreme energy density, such as in the interiors of neutron stars. As a result, neutron stars above a certain mass threshold do not resemble those in general relativity, but develop a scalar charge — in the same sense as in Section 1.1.1 — through a process called spontaneous scalarisation [136].[5] (See also Refs. [145–151] for recent developments on a related phenomenon known as dynamical scalarisation.) When part of a binary system, the motion of this scalarised neutron star leads to the emission of scalar radiation, in addition to gravitational waves.

[4]Chameleon- and Vainshtein-like theories have also been studied in this regime, although screening strongly suppresses any deviations from general relativity [130–135]. That being said, there are still some open questions about how the Vainshtein mechanism operates in binary systems [134].

[5]The tachyonic instability that triggers spontaneous scalarisation is also activated during the early Universe, resulting in a cosmological evolution that deviates from ΛCDM and leads to a present-day vacuum expectation value of the scalar that violates Solar System constraints [139–142]. Fortunately, several ways of preventing this catastrophic behaviour without relying on fine-tuned initial conditions have recently been proposed [143, 144].

If both components of the binary are scalarised, the scalar field also mediates a fifth force between them [137, 138, 152]. Observations of pulsars in binary systems currently set the most stringent bounds on the DEF model [153].

Hairy black holes Another class of models that lead to deviations from general relativity in the strong-field regime is the Einstein–dilaton–Gauss–Bonnet (EdGB) theory and its generalisations [154–162]. These models are interesting because of their novel black hole solutions.

It is today well known that black holes in general relativity are remarkably simple objects, characterised only by their mass and spin. This statement is encoded mathematically in a set of uniqueness theorems [163, 164], which establish the Kerr metric [165] as the only black hole solution to the four-dimensional vacuum Einstein field equations that is stationary, analytic, and asymptotically flat.[6] It was later realised that this unwavering simplicity extends also to more general field theories, including the Brans–Dicke and DEF models, all chameleon-like theories, and most Vainshtein-like theories [174–183]. In these scenarios, black holes are still described by the Kerr metric, while the scalar field is consigned to being everywhere a constant.

The EdGB model circumvents these "no-hair" theorems by coupling the scalar not to matter but to a quadratic curvature invariant. (See Refs. [164, 184, 185] for other black hole solutions that evade the no-hair theorems.) As a result, black holes in these models can possess scalar charges, which — not unlike how neutron stars behave in the DEF model — lead to scalar radiation and the mediation of a fifth force in binary systems [186–188]. At present, the strongest constraints on EdGB theory are set by observations of low-mass X-ray binaries [189] and gravitational-wave data [190–192].

1.2 Scalars from string compactification

Thus far, we have seen how the existence of a light scalar field, as might be required to make sense of dark energy, can be reconciled with the stringent constraints from local tests of gravity by way of a screening mechanism. Yet another interesting possibility, which I have already alluded to, is to furnish the scalar with a symmetry that prohibits the mediation of fifth forces altogether. A pseudoscalar field called the *axion*, which arises as a Nambu–Goldstone boson from the spontaneous breaking of a global $\mathrm{U}(1)$ symmetry, is the quintessential example of this scenario. The axion was introduced in 1977 as a solution to the strong *CP* problem [193–199], although particles with similar properties — now also called axions — have since been shown to emerge in large numbers from string

[6] The Einstein–Maxwell equations admit black hole solutions with an electromagnetic charge [166–169], but such objects are widely considered to have little astrophysical relevance [170–173].

compactifications [200–204]. Nonperturbative effects subsequently bestow each of these particles with a mass (along with higher-order self-interaction terms), whose value can vary wildly depending on the specifics of the construction.

Despite the fact that axions do not mediate fifth forces,[7] they can still give rise to novel signatures through other sorts of interactions with the Standard Model, or through their gravitational effects. As it turns out, axions that occupy different mass ranges are most effectively probed by different kinds of phenomena. Comprehensive reviews on axions and their phenomenology are presented in Refs. [207–218]. In what follows, we shall discuss four key effects that will be relevant in later chapters.

Axion–photon conversion Axions that solve the strong *CP* problem, or "QCD axions" for short, are known to couple to photons via the term[8] $\mathcal{L} \supset -g_{a\gamma\gamma}\phi F_{\mu\nu}{}^{\star}F^{\mu\nu}/4$ (where ${}^{\star}F^{\mu\nu} := \epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}/2$) [219, 220], which allows for axion–photon oscillations in the presence of external electric or magnetic fields. Direct-detection experiments that exploit this phenomenon include "light shining through wall" (LSW) experiments, helioscopes, and haloscopes. All three utilise strong magnetic fields to convert a flux of axions into photons that can more readily be detected, but each is designed to work with a different source. As its name would suggest, an LSW experiment, like GammeV (Gamma to milli-eV particle search) or OSQAR (Optical Search for QED Vacuum Birefringence, Axions and Photon Regeneration), points a laser beam at a wall in the presence of a strong magnetic field in an attempt to detect photons emerging on the other side [221–230]. In contrast, helioscopes like CAST (CERN Axion Solar Telescope) use a dipole magnet aimed at the Sun to search for axions that are ostensibly produced in its interior via the Primakoff effect [231–237]. Finally, haloscopes like ADMX (Axion Dark Matter eXperiment) have been designed to convert axions into microwave photons, assuming they make up at least a fraction of the local dark matter density [238–243]. Seeing as no axion has been detected to date, data from these experiments establish upper bounds on $|g_{a\gamma\gamma}|$ as a function of the axion's mass.

The most universal of these constraints is due to CAST, which bounds $|g_{a\gamma\gamma}| < 6.6 \times 10^{-11}\ \text{GeV}^{-1}$ for any axion lighter than 0.02 eV [237], although in a small mass window around 3×10^{-6} eV, ADMX is able to push this down to $|g_{a\gamma\gamma}| \lesssim 3 \times 10^{-16}\ \text{GeV}^{-1}$ [243]. This constraint is strong enough to rule out two benchmark models — the Kim–Shifman–Vainshtein–Zakharov (KSVZ) axion [196, 197] and the Dine–Fischler–Srednicki–Zhitnitsky (DFSZ) axion [198, 199] — in this mass range, assuming they constitute 100% of the dark matter density. (For reference, both models predict an axion–photon

[7] To be more precise, a U(1) symmetry still allows for a derivative coupling of axions to matter, which does give rise to a long-range interaction between spin-polarised sources [205, 206]. However, axions do not mediate a fifth force between *macroscopic* objects.

[8] If A is a sum of terms and B is one of them, we convey this mathematically by writing $A \supset B$.

coupling that scales like $|g_{a\gamma\gamma}| \sim 7 \times 10^{-16}\ (\mu/3 \times 10^{-6}\ \text{eV})\ \text{GeV}^{-1}$ at the order-of-magnitude level, where μ is the axion mass.) For more details on the latest constraints, see Ref. [218].

It is interesting to note that many of these axion experiments have also been successful in constraining chameleon-like scalars, which couple to photons via an analogous interaction of the form $\sim \phi F_{\mu\nu} F^{\mu\nu}$ [244, 245]. (The difference stems from the fact that chameleons and their cousins are scalars rather than pseudoscalars.) CAST operates in essentially the same way when searching for chameleons [246, 247], but LSW experiments need to be modified since a chameleon's screening mechanism forces it to accumulate in the laser cavity rather than travel through the walls. However, if the laser is subsequently turned off while the magnetic field is left on, the chameleon reconverts into photons; producing an "afterglow" [248, 249]. A modified LSW experiment called GammeV–CHASE (with CHASE being short for CHameleon Afterglow SEarch) [250–253] has been designed to probe this effect, and the same principle is also used by ADMX to constrain chameleons [254].

Fuzzy dark matter While it is remarkable that the QCD axion can account for dark matter if its mass lies somewhere in the range of $10^{-6} - 10^{-3}$ eV [209], it is no less interesting to contemplate a more elaborate scenario in which one axion solves the strong *CP* problem while another is left to be responsible for dark matter. In fact, this kind of setup would seem to be more natural in the context of string theory, which predicts that an abundance of these light particles should exist with masses spanning a vast range from $10^{-33} - 10^{8}$ eV [202]. At the lower end of this spectrum, an axion with a mass $\mu \sim 10^{-22} - 10^{-21}$ eV has become one of the most popular dark matter candidates today [255–261].

Although it couples much more weakly to the Standard Model than does the QCD axion (and so is not accessible to present-day direct-detection experiments[9]), this kind of ultralight particle — often dubbed *fuzzy dark matter* — can still be probed by virtue of its gravitational effects. Specifically, because extremely large occupation numbers are needed to compensate for its incredibly low mass, the collective dynamics of these particles is well described by a classical Klein–Gordon field (of mass μ) whose de Broglie wavelength is on the order of kiloparsecs [215, 259]. Dark matter halos in this model therefore exhibit macroscopic wave-like properties on scales up to this de Broglie wavelength, while being essentially indistinguishable from cold dark matter (CDM) on larger scales. This property makes fuzzy dark matter (FDM) a compelling alternative to the standard CDM paradigm, which has been plagued by several small-scale crises; namely, the core–cusp, missing satellites, and "too big to fail" problems (see Ref. [262] for a review). While it must be acknowledged that these crises

[9] Its interactions with the Standard Model are suppressed by the scale at which symmetry breaking occurs, and so one expects $|g_{a\gamma\gamma}| \lesssim 10^{-16}\ \text{GeV}^{-1}$ for an axion of stringy origin. This is well below what current axion–photon conversion experiments can achieve for masses $\sim 10^{-22}$ eV.

could simply be due to a poor understanding of baryonic feedback that is now improving alongside advancements in numerical simulations [263–272], other studies suggest that FDM could be favoured by observations [273–280].

In any case, an FDM model must not "oversolve" the problem. Because structure formation is suppressed on scales smaller than its de Broglie wavelength, an axion that is too light would reduce the abundance of low-mass halos, decrease the number of galaxies that form at high redshifts, and quench the matter power spectrum at high momenta by too much. Taking these considerations (and others) into account, cosmological observations constrain the mass of an FDM axion to be $\gtrsim 10^{-21}$ eV, assuming it constitutes 100% of the observed dark matter density [215, 281–288]. Similar constraints come from studying the impact of de Broglie-scale fluctuations in the dark matter halo on galactic dynamics [259, 289–291], with the exception of Ref. [292], which derives a more stringent lower bound of $\mu \gtrsim 10^{-19}$ eV. However, the authors of this study state that the limited validity of their approximations could mean that FDM models are still allowed in some parts of the range $10^{-21} - 10^{-20}$ eV.

In addition to these de Broglie-scale effects, FDM halos also exhibit much more rapid Compton-scale fluctuations, which oscillate every ~ 3 months $\times (\mu/10^{-22}\ \text{eV})^{-1}$. When coarse-grained over many periods, these fluctuations average out so as to be irrelevant for most galactic or cosmological studies, but they could potentially be probed by systems evolving on comparable timescales. Pulsar-timing arrays [293–295] and the orbits of pulsar binaries [296, 297] have both been shown to be sensitive to the metric perturbations sourced by these Compton fluctuations, although the limits they impose are presently uncompetitive with those obtained from constraining effects on de Broglie scales. (See Ref. [298] for a summary of constraints on fuzzy dark matter.)

Induced black hole hair Speaking of compact objects, black holes have become prime observational targets when it comes to searching for new physics [299–302]. Recall that while no-hair theorems impose certain restrictions on how a minimally coupled scalar field like the axion can behave around a black hole, these theorems are predicated on a number of crucial assumptions that, if violated, would allow for more interesting effects. When a black hole is embedded in an FDM halo, for instance, oscillations of the scalar field and its absorption by the black hole violate the assumption of stationarity. Accordingly, these processes lead to the formation of a distinctive scalar-field profile around the black hole [303–306], which — for the case of supermassive black holes like the one at the centre of our own Galaxy — can be probed by the orbits of S2-like stars [307, 308].

It is now worth emphasising that this effect is not unique to fuzzy dark matter or to axions. In fact, the idea of dressing a black hole with scalar hair originated with Jacobson [309], who considered the

case of a small black hole embedded in an inflating Universe. (See also Refs. [310–312] for more recent studies.) Boundary conditions require that the inflaton field asymptote to a time-dependent but spatially homogeneous profile at large distances, whereas absorption at the future event horizon forces the inflaton to develop a radial dependence around the black hole. After coarse-graining over distances much greater than the size of this hole, but restricting attention to distances much smaller than the cosmological horizon, one finds a scalar-field profile of the form $\phi \sim \Phi - A\dot{\Phi}/(4\pi r)$ [309,310], where $\Phi \equiv \Phi(t)$ is the spatially homogeneous part of the inflaton field, $\dot{\Phi}$ is its first derivative, and A is the area of the black hole (located at the origin in these coordinates). By analogy with the scalar–tensor theories of the previous section, it is instructive to regard the quantity $Q := -A\dot{\Phi}$ appearing in the numerator of the second term as the *induced* scalar charge of the black hole.

These conclusions can be extended to the late-time Universe as well. Specifically, if dark energy is driven by a slow-rolling scalar field (an axion of mass $\mu \sim 10^{-33}$ eV is a natural dark energy candidate of this kind [215]), then its similarities with inflation automatically imply that black holes in the present epoch should all possess small, albeit nonzero, scalar charges. Going one step further, Horbatsch and Burgess argued that these charges would prompt a binary system of black holes to radiate a fraction of its energy and momentum into scalar waves [313]. As this process would lead to a more rapid decay of the orbital period, observations of the supermassive binary black hole at the centre of quasar OJ 287, whose inspiral is consistent with the predictions of the vacuum Einstein field equations to within an uncertainty of 6% [314–316], impose the constraint $\sqrt{4\pi G}|\dot{\Phi}| \lesssim 7 \times 10^{-7}\ \mathrm{s}^{-1}$ on the local time evolution of any rolling scalar [313]. While this bound is by no means spectacular (one expects a quintessence-type scalar to satisfy $\sqrt{4\pi G}|\dot{\Phi}| \ll H_0$, where $H_0 = 2 \times 10^{-18}\ \mathrm{s}^{-1}$ is the Hubble rate), that black holes are sensitive to this value at all is an interesting observation. Note that the black holes observed by LIGO (the Laser Interferometer Gravitational-Wave Observatory) have also been used to constrain this effect [190], but the bound obtained is much looser since $Q \propto A$.

Black hole superradiance Even more dramatic phenomenology is possible around black holes if a minimally coupled, light scalar field — axion or otherwise — carries a substantial amount of angular momentum. As early as the 1970s, Starobinsky, Teukolsky, and others demonstrated that a coherent bosonic field will undergo amplification as it scatters off a rotating black hole, provided its angular phase velocity is less than that of the horizon [317–320]. In many ways, this phenomenon can be regarded as the wave analogue of the Penrose process [321, 322], although such a comparison obscures some important distinctions between the two cases. In particular, unlike the Penrose process, the amplification

of waves by a black hole does not rely on the existence of an ergosphere, but operates on the basis of a much more generic principle known as superradiance.

At its most fundamental level, superradiance is driven by a *dissipative* interaction that facilitates the exchange of energy between an incident wave and a rotating body. Amplification occurs whenever the wave satisfies the "superradiance condition" $0 < \omega < m\Omega$, where ω is its angular frequency, m is its azimuthal number with respect to the body's axis of rotation, and Ω is the body's angular velocity. By virtue of it being so generic, one might expect to find superradiant phenomena arising in a broad range of contexts and, indeed, this is the case. In his seminal papers on the subject [323, 324], Zel'dovich describes the amplification of electromagnetic waves scattering off a conducting cylinder, while more recently, Torres et al. [325] reported the first laboratory observation of superradiance in water waves scattered by a draining vortex. In astrophysics, the superradiant scattering of light bosonic fields by rotating stars has also recently been studied [326–328].

Having said all of this, black holes occupy a special position in the superradiance literature because of the role that gravity plays. While a single scattering event increases the intensity of an incident scalar wave by at most 0.4% (this number improves to 4% for electromagnetic waves and 138% for gravitational waves) [329], a bosonic field with a small but nonzero mass μ can condense into a gravitationally bound "cloud" (similar to how electrons form clouds around protons in hydrogen atoms) that constantly interacts with the black hole's horizon. In this bound configuration, continual superradiant amplification triggers an instability that leads to exponential growth [330–352]. (See Ref. [329] for a comprehensive review.) To feed this growth, the cloud extracts energy and angular momentum from its host until the black hole can no longer rotate rapidly enough to satisfy the superradiance condition ($\omega \approx \mu$ for bound states), at which point the instability shuts off. It is worth emphasising that this process does not rely on the boson being produced in large abundances during the early Universe; in principle, even a small quantum fluctuation can trigger this instability [339].

This will be true for bosons of any mass, but superradiance proceeds most efficiently when the field's Compton wavelength is comparable to the size of the black hole ($\sim GM$). At this resonance, the minimum value of the e-folding time for this instability is $\min(t_{\rm inst}) \sim 29\ {\rm s}\ (M/M_\odot)$ in the case of a scalar cloud [333], whereas $\min(t_{\rm inst}) \sim 12\ {\rm ms}\ (M/M_\odot)$ for a vector cloud [335]. (The corresponding value for a spin-2 field is not presently known.) The fact that these timescales are so short, especially by astronomical standards, can be exploited as a means of indirect detection. Since a cloud formed by this instability will spin down a black hole within a timescale set by $t_{\rm inst}$, the existence of a boson of mass μ can be inferred if black holes of mass $M \sim (G\mu)^{-1}$ are all observed to be slowly rotating. This is known as a gap in the Regge plane [336, 347, 353–356]. Conversely, the observation of a rapidly rotating

black hole rules out the existence of ultralight bosons with a mass around $\mu \sim (GM)^{-1}$. Because the value of $t_{\rm inst}$ sharply increases on either side of the resonance, especially for scalar fields, observations of rapidly rotating supermassive black holes ($M \sim 10^7 - 10^8\ M_\odot$) rule out axions and similar particles with masses in the narrow range $\mu \sim 10^{-19} - 10^{-18}$ eV [329]. Similar conclusions hold also for stellar-mass black holes (for which $\mu \sim 10^{-13} - 10^{-12}$ eV would be excluded), although spin measurements are less reliable in this case [329].

Several other key features of these superradiant clouds are worth discussing. For one, its rotation around the black hole's spin axis and its Compton-scale oscillations make it a continuous source of monochromatic gravitational waves [357–360].[10] This kind of signal may be resolvable if the host black hole is within our own Galaxy; otherwise, it could be detected as part of the stochastic gravitational-wave background. Speaking of gravitational waves, another promising way to search for these superradiant clouds is through their impact on the evolution of black hole binaries. These clouds could potentially reveal themselves by introducing resonances and other secular effects into the orbits of extreme mass-ratio inspirals [361–364]; prime targets for the future, space-based gravitational-wave detector LISA (the Laser Interferometer Space Antenna) [365]. Comparable-mass binaries are just as interesting, as resonant transitions of the cloud have been shown to lead to a large dephasing of the gravitational-wave signal during the early inspiral [366–369]. Finally, going beyond gravitational waves, recent imaging of the supermassive black hole in M87 by the Event Horizon Telescope [370] has inspired investigations into how black hole shadows are affected by superradiant clouds; see Refs. [371–373].

1.3 Objectives

Having presented a broad overview of the current literature, we are now in a position to identify key directions in which progress can be made. I stated earlier that the aim of this thesis is to update our understanding of how light scalar fields influence the dynamics of moving bodies. We will do so mainly by example; focusing on a number of specific systems in two distinct regimes.

1.3.1 Electrons in laboratory experiments

We begin in the laboratory, where small-scale experiments performed in high vacuum have established themselves as ideal testing grounds for chameleon-like theories. While the most stringent constraints are naturally set by dedicated experiments optimised to be sensitive to small deviations from Newtonian gravity, the fact that chameleon-like scalars couple universally to matter makes it interesting to ask:

[10] This is true for *real* fields; clouds comprised of a complex field need not emit gravitational waves [349, 350].

Could an existing experiment that was not designed to search for this kind of particle nevertheless be repurposed to provide meaningful constraints? Of course, the more precise the experiment, the better the constraint is likely to be; hence, the most promising candidates a priori are a small number of so-called "precision tests of QED." These experiments have been refined over the course of decades to deliver measurements of the fine-structure constant α that are today in excess of 10 significant figures. Given that the data from these high-precision tests are readily available, it is interesting and also cost-effective to explore what kind of constraints they impose on chameleon-like theories. Moreover, by using these existing experiments to rule out models that are in conflict with known physics, we can better identify where in parameter space future, dedicated searches should target their attention.

In keeping with our central theme, Part I of this thesis presents a reanalysis of two precision tests of QED that both operate by probing the motion of an electron. After reviewing some of the more technical aspects of chameleon-like theories in Chapter 2, we begin our analysis in Chapter 3 with an investigation into how a chameleon-like scalar affects the gross, fine, and hyperfine structure of the hydrogen atom. These calculations provide a good introduction to some of the key ideas that will feature repeatedly in this thesis, although much of the phenomenology unique to chameleon-like theories will be absent; the primary reason being that protons and electrons are not heavy enough to drive the scalar into its fully nonlinear regime. We address this shortcoming by considering a second precision test of QED in Chapter 4. An experiment known as the one-electron quantum cyclotron measures the magnetic moment g_e of an electron by suspending it in a cylindrical Penning trap, into which microwave photons are injected. These photons lead to resonant excitations of the electron, whose Landau levels encode information about the value of g_e, which in turn contains information about the value of α. Chapter 4 reanalyses this experiment under the assumption that a chameleon-like scalar is active in the cavity, where we would expect it to form a vacuum bubble.

These two experiments offer plenty of insight into how chameleon-like scalars alter the properties and dynamics of charged, elementary particles like the electron, but they do not contain sufficient information for us to constrain these models when analysed individually, since these are the experiments used to measure the fine-structure constant in the first place. Fortunately, there is a simple fix: We can simply combine data from both experiments. The fact that each one provides a determination of α that is in good agreement with the other sets a limit on the size of the effects that a chameleon-like scalar can introduce. To conclude Part I of this thesis, Chapter 5 uses this approach to establish new constraints on two archetypal models: the chameleon and the symmetron.

1.3.2 Black holes in astrophysical environments

In Part II of this thesis, we turn our attention to much larger astrophysical scales, where even a minimally coupled scalar field can exhibit some truly remarkable phenomenology around a black hole. We have already seen in the previous section that the literature on this subject is extensive, but apart from a few notable exceptions, most of what is known today is about how light scalar fields evolve around a *single* black hole at rest. Questions about how one interacts with a *binary* black hole, on the other hand, remain mostly open, and so our goal in this second half will be to take a number of key steps in this direction.

To be sure, the problem is a difficult one. It has taken decades of concerted development to arrive at our current understanding of how binary black holes evolve and coalesce when isolated in empty space. Now allowing for the presence of an additional field on top of this spacetime invariably complicates the problem by introducing a myriad of new scales, which will almost certainly trigger an onslaught of new effects. Nevertheless, progress can still be made by restricting our attention to certain limits. I stated earlier, for example, that when it is part of a binary system, resonant transitions of the superradiant cloud around a black hole have been shown to lead to a large dephasing of the gravitational-wave signal [366–369]. These effects are under analytic control only during the early part of the inspiral, when the size of the cloud is much smaller than the orbital separation a of the binary. Less is known about field configurations whose characteristic length scale λ is comparable to a, although recent numerical simulations have shown that when a scalar Gaussian pulse of this size scatters off a binary black hole, its evolution is marked by quasinormal modes that are associated with the ringdowns of both the individual black holes and the binary as a whole [374].

So as to complement these studies, we will focus on yet another regime in this thesis, which I will call the "long-wavelength limit." This is the limit in which $\lambda \gg a$, and is exactly the regime in which Horbatsch and Burgess first showed that a binary black hole radiates scalar waves when embedded in a quintessence-type scalar field [313]. Note that this effect is not unique to rolling scalars, however, as outgoing scalar radiation was also subsequently observed in numerical relativity simulations of binary black holes embedded in Gaussian scalar-field profiles and spatial scalar-field gradients [375, 376]. In this light, it becomes natural to ask: How generic is this phenomenon? Extrapolating from these studies leads us to expect that a black hole should gain a scalar charge and, consequently, radiate scalar waves when it is moving through *any* ambient scalar-field profile, but if this is really the case, then it would be nice to have a general framework for studying this phenomenon, which would correctly reproduce the results from the above studies as limiting cases.

Ideally, such a framework should also provide us with a more reliable method for performing calculations. At the moment, the flux of scalar radiation can be computed only when the binary is

surrounded by a scalar field that varies at most linearly with space and time. In this case, the black holes' scalar charges are effectively constant; hence, the calculation proceeds entirely by analogy with neutron stars in the DEF model. New "technology" will have to be developed, however, for ambient scalar fields that exhibit a more complicated spacetime dependence. Such a generalisation is vital as it will allow us to consider a wider range of scenarios of astrophysical interest. For instance, given that black holes have already been shown to possess nontrivial scalar hair when at rest within a fuzzy dark matter halo, a natural next step would be to investigate how a binary system of black holes behaves when embedded in this environment.

Scalar radiation aside, it will also be interesting to explore if an ambient field gives rise to any other effects that can influence the binary's inspiral. In fact, our physical intuition can already be used to make a number of educated guesses. Given that a black hole is known to be subject to both dynamical friction and a drag force due to accretion when it travels through some extended distribution of matter, we should expect that the same is true also when it moves through an ambient scalar field. What would be especially interesting from a theoretical point of view is if all of these different effects — dynamical friction, accretion, scalar radiation, etc. — can be derived from a single, unified framework.

With all of this in mind, we will push forward by constructing an effective field theory for this problem à la Goldberger and Rothstein [377–380]. The assumed hierarchy between the characteristic length and timescales of the black holes and those of the ambient scalar facilitates an effective description of the former in terms of worldlines furnished with composite operators that capture finite-size effects. The key benefit of this description is the ability to disentangle questions about the long-distance, IR physics we are interested in — such as the trajectories of the black holes and the flux of outgoing radiation — from the short-distance, UV physics transpiring near the horizons. Information about the latter is accessible to distant observers, like ourselves, through the way it impacts the black hole's multipolar structure. Mathematically, this is encoded in the EFT by Wilsonian coefficients, whose values can be determined by matching to a more UV-complete theory that is able to resolve details on the scale of the black holes' horizons. For those of astrophysical size, general relativity will suffice as our UV completion.

The construction of this EFT plays out across two chapters. In Chapter 6, we begin by solving for the general behaviour of a light scalar field around the horizon of a Kerr black hole. The results from this calculation are then matched onto a point-particle effective action for the black hole, which we construct in Chapter 7. The construction itself is one of the main novelties in this thesis: By using the in–in formalism to integrate out the composite operators localised on its worldline, we obtain the first, fully relativistic framework that can systematically account for both conservative and dissipative effects

associated with the black hole's horizon. By virtue of their universality, the Wilsonian coefficients that appear in this effective action, once matched, can be applied to study a wide array of scenarios.

Case in point, in Chapter 8 we use this framework to examine how a binary black hole evolves when embedded in a fuzzy dark matter halo. As desired, this EFT provides a single, unified framework for performing calculations: scalar radiation, dynamical friction, accretion, and a number of other effects to be discussed are all seen to arise from appropriate interaction terms which, as in any perturbative theory, can be organised neatly according to Feynman diagrams. Also in this chapter, we will explore how well observations of the quasar OJ 287 constrain the local density of fuzzy dark matter.

Having studied how a binary evolves within a much more massive halo, in Chapter 9 we then turn our attention to the opposite limit by examining how diffuse configurations of a scalar field evolve around a much more massive binary black hole. In so doing, we will be able to address a longstanding question about this kind of system: Can binary black holes amplify long-wavelength fields via superradiance? The natural expectation is: yes. In the same way that its absorptive horizon is what facilitates energy and angular momentum exchange between a single rotating black hole and an external bosonic field, one expects that absorption across the horizons will again be responsible for triggering the onset of superradiance around a binary black hole. The key difference now is that the predominant source of rotational energy is the binary's orbital motion, rather than the spins of its constituents; hence, amplification ought to occur even when the individual black holes are not spinning. These expectations will be substantiated by analytic calculations in Chapter 9, where we will use our EFT approach to compute (among other things) the amplification factor for incident scalar waves and the growth rate for scalar clouds that are gravitationally bound to the binary as a whole.

We will then conclude in Chapter 10 by summarising all of the original results presented in this thesis.

Conventions

Natural units are adopted throughout, such that the reduced Planck constant $\hbar$ and the speed of light c are always both equal to 1. The vacuum permittivity ε_0 and, consequently, the vacuum permeability μ_0 are also set equal to 1. Rather unusually, two different definitions of the reduced Planck mass will be used: Part I adopts the more conventional definition $M_{\rm Pl} \equiv (8\pi G)^{-1/2}$, whereas Part II switches to a lowercase Planck mass given by $m_{\rm Pl} \equiv (32\pi G)^{-1/2}$ to be consistent with the relevant literature; see, e.g., Refs. [377–380]. Our metric signature is mostly plus; i.e., the Minkowski metric $\eta_{\mu\nu} = \text{diag}(-1, +1, +1, +1)$, and conventions for the curvature tensors follow those of Misner,

Thorne, and Wheeler [381]. When used to label directions in spacetime, Greek indices $\{\mu, \nu, \dots\}$ run from 0 to 3, whereas Latin indices $\{i, j, \dots\}$ run from 1 to 3.

Angled brackets serve multiple purposes in this thesis: In Chapters 3, 4, and 7, the notation $\langle \hat{O} \rangle$ is used to denote the expectation value of some quantum mechanical operator $\hat{O}$. In Chapters 8 and 9, where the systems are fully classical, the notation $\langle f \rangle$ instead refers to the time average of some quantity f. As for angled brackets around indices, these always denote the symmetric, trace-free projection of the corresponding tensor; e.g., $T^{\langle ij \rangle} = T^{(ij)} - \delta^{ij}(T^{k\ell}\delta_{k\ell}/3)$.

All other conventions will be defined explicitly when first used in the text.

PART I

Electrons

Chapter 2
Chameleon-like theories

Part I of this thesis presents a reanalysis of two precision tests of QED under the assumption that a chameleon-like scalar is active in the experiment. Its impact on the spectral lines of hydrogen are explored in Chapter 3, while the systematic effects it introduces into measurements of the electron's magnetic moment will be examined in Chapter 4. We will then combine results from both analyses in Chapter 5 to establish new constraints on these models. Before any of this, however, we must discuss all of the different ingredients that make up a chameleon-like theory.

Within an effective field theory (EFT) framework, the action for this sort of theory can be constructed from the bottom up by simply including all possible local interactions permitted by the symmetries we wish to impose.[1] To that end, let us consider a theory in four dimensions whose gravitational sector is comprised of a metric $g_{\mu\nu}$ and a real scalar field ϕ. We will insist that this theory is diffeomorphism invariant, in which case our action should be assembled from products of these fields, derivatives of the scalar, and the curvature tensors only. For simplicity, we shall further restrict our attention to terms in the action containing at most two derivatives, and will also assume that our theory satisfies the weak equivalence principle. These provisions lead to the Bergmann–Wagoner class of scalar–tensor theories [384, 385], defined by the action

$$S = \int \mathrm{d}^4x \sqrt{-g}\,[\phi R - Z(\phi) g^{\mu\nu} \partial_\mu\phi\,\partial_\nu\phi - U(\phi)] + S_m[g]. \tag{2.1}$$

The dynamics of ϕ are determined by two functions, $Z(\phi)$ and $U(\phi)$, which we are free to specify. The Brans–Dicke theory, for instance, is a line in this space of models corresponding to the specific choice $U(\phi) = 0$ and $Z(\phi) = \omega/\phi$, where ω is some constant [38, 39].

[1] While some attempts have been made to embed chameleon-like theories into a UV completion [382, 383], treating them as effective field theories currently remains the most fruitful approach.

Enforcing the weak equivalence principle means that the Standard Model fields, which are encapsulated in the matter action S_m, are coupled only to the metric $g_{\mu\nu}$. Prima facie, this might seem to contradict our discussion earlier in Chapter 1, where I stated that these scalar fields couple to matter and so mediate a fifth force. There is of course no such contradiction, because the Ricci scalar R can always be replaced by the trace of the energy–momentum tensor at the level of the field equations. Alternatively, the scalar's coupling to matter can be made more manifest by performing a field redefinition. Transformations of this kind leave the physical content of a theory intact [386–388],[2] but can be useful for generating equivalent formulations that are more convenient to work with.

2.1 Conformal transformations and the fifth force

The theory in (2.1) is said to be written in the Jordan frame. To make this explicit, let us affix the subscript J onto the fields, such that the action now reads

$$S = \int \mathrm{d}^4x\,\sqrt{-g_{\mathrm{J}}}\,[\phi_{\mathrm{J}}R_{\mathrm{J}} - Z(\phi_{\mathrm{J}})g_{\mathrm{J}}^{\mu\nu}\partial_\mu\phi_{\mathrm{J}}\,\partial_\nu\phi_{\mathrm{J}} - U(\phi_{\mathrm{J}})] + S_m[g_{\mathrm{J}}]. \tag{2.2}$$

The defining feature of this frame is that matter is minimally coupled to the metric g_{J}. A new frame with metric g_{E} can then be obtained through the conformal transformation

$$g_{\mathrm{J}\,\mu\nu}(x) = \Omega^2(x)\,g_{\mathrm{E}\,\mu\nu}(x). \tag{2.3}$$

Any choice of conformal factor $\Omega(x)$ will define a valid frame, but the action takes on the especially simple form

$$S = \int \mathrm{d}^4x\,\sqrt{-g_{\mathrm{E}}}\left(\frac{M_{\mathrm{Pl}}^2}{2}R_{\mathrm{E}} - \frac{1}{2}g_{\mathrm{E}}^{\mu\nu}\partial_\mu\phi_{\mathrm{E}}\,\partial_\nu\phi_{\mathrm{E}} - V(\phi_{\mathrm{E}})\right) + S_m[\Omega^2(\phi_{\mathrm{E}})g_{\mathrm{E}}] \tag{2.4}$$

when we make the particular choice $\Omega(x) = M_{\mathrm{Pl}}\,\phi_{\mathrm{J}}^{-1/2}(x) > 0$; having also defined a new scalar ϕ_{E} such that $\mathrm{d}\phi_{\mathrm{E}} = M_{\mathrm{Pl}}\sqrt{3 + 2\phi_{\mathrm{J}}Z(\phi_{\mathrm{J}})}\,\mathrm{d}\log\phi_{\mathrm{J}}$. In performing this transformation, use has been made of the identity [389]

$$\Omega^2 R_{\mathrm{J}} \equiv R_{\mathrm{E}} - 6g_{\mathrm{E}}^{\mu\nu}\nabla_\mu\nabla_\nu\log\Omega - 6g_{\mathrm{E}}^{\mu\nu}(\partial_\mu\log\Omega)(\partial_\nu\log\Omega), \tag{2.5}$$

where ∇ is the covariant derivative (with Levi–Civita connection) defined with respect to g_{E}.

The form of (2.4) should make the utility of these field redefinitions apparent. The new fields $(g_{\mathrm{E}}, \phi_{\mathrm{E}})$ are governed by much simpler kinetic terms. Indeed, the scalar ϕ_{E} is now simply a (generalised) Klein–

[2] For the same reason, generalising the coupling $\phi R \to f(\phi)R$ does not introduce anything new into the theory.

Gordon field with potential $V \equiv \Omega^4 U$ that is minimally coupled to the metric $g_{\rm E}$, which in turn is governed by nothing more than the Einstein–Hilbert action. For this reason, this frame is known as the Einstein frame. The price we pay for simpler kinetic terms is the introduction of nonminimal couplings between the Standard Model fields and the scalar $\phi_{\rm E}$, whose ramifications can be seen by deriving the equations of motion descending from this new Einstein-frame action.

Moving forward, we shall dispense with the subscript E's when referring to Einstein-frame quantities, but Jordan-frame quantities will continue to be identified with subscript J's. Extremising (2.4) then leads to the field equations

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{1}{M_{\rm Pl}^2}(T_{\mu\nu} + T_{\mu\nu}^{(\phi)}), \tag{2.6a}$$

$$\nabla^\mu \nabla_\mu \phi - V'(\phi) = -\frac{1}{\sqrt{-g}} \frac{\delta S_m}{\delta \phi}, \tag{2.6b}$$

where $V'(\phi) \equiv \mathrm{d}V/\mathrm{d}\phi$. The first of these resembles the usual Einstein equations in general relativity, and similarly states that all forms of energy act as a source of spacetime curvature. On the rhs, it proves useful to isolate the scalar's contribution to the total energy–momentum tensor, which is given by

$$T_{\mu\nu}^{(\phi)} = \partial_\mu \phi \, \partial_\nu \phi - \left(\frac{1}{2} g^{\rho\sigma} \partial_\rho \phi \, \partial_\sigma \phi + V(\phi) \right) g_{\mu\nu}. \tag{2.7}$$

The remaining contribution from the matter sector is defined by

$$T_{\mu\nu} := -\frac{2}{\sqrt{-g}} \frac{\delta S_m}{\delta g^{\mu\nu}}. \tag{2.8}$$

The second field equation in (2.6b) further states that matter can directly source the scalar ϕ, owing to the nonminimal coupling introduced by our field redefinition.

What about the fifth force? To determine the effect of this scalar field on the matter sector, we may begin by noting that the usual Bianchi identities hold in the Einstein frame; hence, acting on (2.6a) with ∇_μ returns $\nabla_\mu (T^{\mu\nu} + T^{(\phi)\mu\nu}) = 0$. After also using (2.7), we find

$$\nabla_\mu T^{\mu\nu} = T \, \nabla^\nu \log \Omega(\phi), \tag{2.9}$$

where $T \equiv T^{\mu\nu} g_{\mu\nu}$. We learn from this equation that the energy–momentum tensor for matter is not conserved in the Einstein frame when gradients of ϕ are present. This nonconservation, which is due to the term on the rhs, is exactly the fifth force that a scalar field exerts onto matter particles.

The same conclusion may also be drawn by thinking about how matter behaves in the Jordan frame. Diffeomorphism invariance guarantees that its energy–momentum tensor in this frame is covariantly conserved; i.e., $(\nabla_{\rm J})_\mu T_{\rm J}^{\mu\nu} = 0$, where $\nabla_{\rm J}$ is the covariant derivative defined with respect to $g_{\rm J}$ and

$$T_{{\rm J}\,\mu\nu} := -\frac{2}{\sqrt{-g_{\rm J}}}\frac{\delta S_m}{\delta g_{\rm J}^{\mu\nu}}. \tag{2.10}$$

By comparing the two definitions in (2.8) and (2.10) and using the chain rule, we obtain the relation

$$T_{\mu\nu} = \Omega^2\, T_{{\rm J}\,\mu\nu}, \tag{2.11}$$

or equivalently, $T^{\mu\nu} = \Omega^6 T_{\rm J}^{\mu\nu}$. The result in (2.9) can then be reproduced by acting on this latter equation with ∇_μ and using the fact that the Levi–Civita connections for these two frames are related by the identity [389]

$$(\Gamma_{\rm J})^\rho{}_{\mu\nu} \equiv \Gamma^\rho{}_{\mu\nu} + 2\delta^\rho{}_{(\mu}\partial_{\nu)}\log\Omega - g_{\mu\nu}g^{\rho\sigma}\partial_\sigma\log\Omega. \tag{2.12}$$

The physical meaning behind these calculations is simple: Since the energy–momentum tensor is conserved in the Jordan frame, massive test particles travel along timelike geodesics of $g_{\rm J}$. Under a conformal transformation, these same test particles cannot also be travelling along timelike geodesics of g; hence, the departure from geodesic motion in the Einstein frame manifests as a fifth force. (Massless test particles, on the other hand, do not experience a fifth force because the null geodesics of a metric are conformally invariant. We will meet this conformal invariance again when we discuss gauge bosons later.)

2.1.1 Perfect fluids

Further insight into the nature of this fifth force can be gleaned by specialising to several concrete examples of matter species. Let us begin by considering a perfect, pressureless fluid. In the Jordan frame, this fluid has energy density $\rho_{\rm J}$ and flows along timelike curves with 4-velocity $u_{\rm J}^\mu$; hence, $T_{\rm J}^{\mu\nu} = \rho_{\rm J} u_{\rm J}^\mu u_{\rm J}^\nu$. There must exist analogous quantities in the Einstein frame, so let us posit that the energy–momentum tensor in this frame has the form $T^{\mu\nu} = \rho u^\mu u^\nu$.

To preserve the unit norm of the 4-velocities in both frames, we will demand that

$$g_{{\rm J}\,\mu\nu}u_{\rm J}^\mu u_{\rm J}^\nu = g_{\mu\nu}u^\mu u^\nu = -1, \tag{2.13}$$

which automatically implies that $u^\mu = \Omega\, u_{\rm J}^\mu$. Further requiring that $T^{\mu\nu}$ and $T_{\rm J}^{\mu\nu}$ are linked according

to (2.11) enforces the relation $\rho = \Omega^4 \rho_{\rm J}$. Substituting these quantities into (2.9) then yields

$$u^\nu \nabla_\mu (\rho u^\mu) + \rho (u^\alpha \nabla_\alpha u^\nu) + \rho\, \partial^\nu \log \Omega = 0. \tag{2.14}$$

To see that the scalar exerts a fifth force on this fluid, project this equation onto the spatial hypersurface orthogonal to u^μ to obtain

$$a^\mu = -(g^{\mu\nu} + u^\mu u^\nu) \partial_\nu \log \Omega, \tag{2.15}$$

where $a^\mu := u^\nu \nabla_\nu u^\mu$ is the acceleration of the worldline tangent to u^μ. We may think of this fluid as being comprised of a large number of test particles, each of mass m; thus, multiplying (2.15) by m gives us the equation of motion

$$m a^\mu = Q (g^{\mu\nu} + u^\mu u^\nu) \partial_\nu \phi \tag{2.16}$$

that each test particle should obey. The term on the rhs is the fifth force, whose strength is determined by the particle's *scalar charge* $Q := -m\, \mathrm{d} \log \Omega / \mathrm{d}\phi$. By analogy with electrostatics,[3] the fluid as a whole therefore possesses a charge density $(Q/m)\rho$, which acts as a source for the scalar field. Indeed, using the chain rule to compute the rhs of (2.6b), one finds

$$-\frac{1}{\sqrt{-g}} \frac{\delta S_m}{\delta \phi} = \frac{\mathrm{d} \log \Omega}{\mathrm{d}\phi} \rho = -\left(\frac{Q}{m} \right) \rho. \tag{2.17}$$

It is worth remarking that this energy density ρ generally turns out to be inconvenient to work with, however, as it is not conserved. To see this, simply contract (2.14) with u_ν to find

$$\nabla_\mu (\rho u^\mu) - \rho u^\mu \partial_\mu \log \Omega = \Omega \nabla_\mu (\Omega^{-1} \rho u^\mu) = 0. \tag{2.18}$$

Interestingly, we see that while ρ is not conserved along the fluid flow lines, there exists a quantity $\rho_c := \Omega^{-1} \rho$ that is. It should be stressed that this quantity is a purely mathematical construct with no physical interpretation — Jordan-frame observers (like us) comoving with the fluid will measure a density $\rho_{\rm J}$, whereas would-be Einstein-frame observers will measure a density ρ. Nevertheless, ρ_c is still useful as the presence of conserved quantities makes solving equations of motion much simpler.

[3] The analogy is not perfect, however, because unlike electric charges, scalar charges are not associated with a conserved current. This terminology is conventional, nonetheless.

2.1.2 Gauge bosons and electromagnetism

Vector gauge bosons are noteworthy because the classical Yang–Mills action is conformally invariant in four dimensions. While this would naively suggest no direct coupling to the scalar, this conformal symmetry turns out to be anomalous; hence, interactions of the form $\sim \phi \,\mathrm{tr}(F_{\mu\nu}F^{\mu\nu})$ are inevitably generated under renormalisation group flow [244, 245, 390]. In the case of gluons, this interaction causes the strength of the fifth force to exhibit some composition dependence when acting on baryonic matter [390], although such effects are typically small compared to the composition-independent part of the fifth force [58], whose strength is set by the conformal factor $\Omega(x)$. For this reason, we will neglect the scalar's coupling to gluons throughout this thesis.

On the other hand, it would be unwise to neglect the coupling of the scalar to photons, given that the systems to be analysed in later chapters invariably involve strong electric and magnetic fields. In such scenarios, this coupling could potentially play an important role. Without appealing to any particular UV completion, we may parametrise this interaction agnostically by introducing the arbitrary function $\varepsilon(\phi)$,[4] such that the action for the electromagnetic sector reads

$$S_{\mathrm{em}} = -\frac{1}{4}\int \mathrm{d}^4x\sqrt{-g}\,\varepsilon(\phi)F_{\mu\nu}F^{\mu\nu}. \tag{2.19}$$

Extremising this action then leads to a set of modified Maxwell equations:

$$\nabla_\mu(\varepsilon F^{\mu\nu}) = -J^\nu, \qquad J^\mu = \frac{1}{\sqrt{-g}}\frac{\delta S_m}{\delta A_\mu}. \tag{2.20}$$

Acting on the first of these with ∇_ν and using the antisymmetry property of the field strength tensor $F_{\mu\nu} := 2\partial_{[\mu}A_{\nu]}$ shows that J^μ is a conserved current. This is to be expected, since the coupling preserves the $\mathrm{U}(1)$ symmetry of the theory. To understand its physical implications, it is instructive to compare (2.20) with the set of Maxwell equations in matter [391]. In doing so, we may deduce that this interaction causes the vacuum to behave like a dielectric medium, whose relative permittivity and relative permeability are given by ε and ε^{-1}, respectively. Of course, the presence of this coupling also means that electromagnetic fields will act as a source for the scalar. Indeed, it is straightforward to show that

$$-\frac{1}{\sqrt{-g}}\frac{\delta S_{\mathrm{em}}}{\delta\phi} = \frac{1}{4}\frac{\mathrm{d}\varepsilon}{\mathrm{d}\phi}F_{\mu\nu}F^{\mu\nu} =: \frac{\mathrm{d}\varepsilon}{\mathrm{d}\phi}\rho_{\mathrm{em}}, \tag{2.21}$$

where the energy density ρ_{em} evaluates to $(\mathbf{B}^2 - \mathbf{E}^2)/2$ on flat space.

[4] Doing so breaks our earlier assumption of the weak equivalence principle, but only for photons.

2.1.3 Electrons

To conclude this section, we should also discuss how the scalar couples to elementary particles like the electron. Recall that, by assumption, the Standard Model is minimally coupled to the Jordan-frame metric $g_{\rm J}$; hence, the electron obeys the standard Dirac equation in this frame. For an electron of mass m_e and charge $-e$, the corresponding action is

$$S_\psi = \frac{1}{2}\int \mathrm{d}^4x\sqrt{-g_{\rm J}}\,\overline{\psi}_{\rm J}[i\,(\not{e}_{\rm J})^\mu(D_{\rm J})_\mu - m_e]\psi_{\rm J} + \text{c.c.}, \tag{2.22}$$

where the Jordan-frame vierbeins $(\mathrm{e}_{\rm J})^\mu_{\hat\alpha}$ have spacetime indices μ and local Lorentz indices $\hat\alpha$. The Dirac matrices satisfy $\{\gamma^{\hat\alpha},\gamma^{\hat\beta}\} = -2\eta^{\hat\alpha\hat\beta}$, the Lorentz generators are $S^{\hat\alpha\hat\beta} := (i/4)[\gamma^{\hat\alpha},\gamma^{\hat\beta}]$, and standard Feynman slash notation is being used; i.e., $(\not{e}_{\rm J})^\mu \equiv \gamma^{\hat\alpha}(\mathrm{e}_{\rm J})^\mu_{\hat\alpha}$. Lastly, the covariant derivative is given by $(D_{\rm J})_\mu := \partial_\mu + (\varpi_{\rm J})_\mu + ieA_\mu$ with spin connection

$$(\varpi_{\rm J})_\mu = -\frac{i}{2}S_{\hat\alpha\hat\beta}\left[(\nabla_{\rm J})_\mu(\mathrm{e}_{\rm J})^{\hat\alpha}_\nu\right]g_{\rm J}^{\nu\lambda}(\mathrm{e}_{\rm J})^{\hat\beta}_\lambda. \tag{2.23}$$

Under the conformal transformation in (2.3), this action reads

$$S_\psi = \frac{1}{2}\int \mathrm{d}^4x\sqrt{-g}\,\Omega^3\overline{\psi}_{\rm J}\left(i\not{e}^\mu D_\mu - \Omega m_e + \frac{3}{2}i\not{e}^\mu\partial_\mu\log\Omega\right)\psi_{\rm J} + \text{c.c.}, \tag{2.24}$$

where the Einstein-frame vierbeins $\mathrm{e}^\mu_{\hat\alpha} = \Omega^{-1}(\mathrm{e}_{\rm J})^\mu_{\hat\alpha}$, while D_μ is defined just like $(D_{\rm J})_\mu$ except without the subscript J's. At this stage, it is useful to define a new spinor $\psi = \Omega^{-3/2}\psi_{\rm J}$ whose kinetic term is then canonical in the Einstein frame:

$$S_\psi = \frac{1}{2}\int \mathrm{d}^4x\sqrt{-g}\,\overline{\psi}(i\not{e}^\mu D_\mu - \Omega m_e)\psi + \text{c.c.} \tag{2.25}$$

As a bonus, this redefinition also exactly cancels the extra $\not{e}^\mu\partial_\mu\log\Omega$ term in (2.24). Just as we saw with energy densities in Section 2.1.1, redefining matter fields in the Einstein frame helps with making the calculations easier. A conserved quantity can also be constructed from ψ. In this case, it is the familiar Noether current $J^\mu = -e\overline{\psi}\not{e}^\mu\psi$ related to the usual global U(1) symmetry.

We can now read off the modified Dirac equation in the Einstein frame from (2.25), obtaining

$$(i\not{e}^\mu D_\mu - \Omega m_e)\psi = 0. \tag{2.26}$$

Evidently, the effect of the conformal factor $\Omega(\phi)$ is to endow the electron with an effective mass that depends on the local value of ϕ.

2.2 Screening mechanisms

So far in this chapter, we have seen how to define a general class of scalar–tensor theories that are nonminimally coupled to the Standard Model in the Einstein frame. One selects a particular model from this class by specifying three free functions: the conformal coupling to matter $\Omega(\phi)$, the conformal coupling to electromagnetism $\varepsilon(\phi)$, and the self-interaction potential $V(\phi)$. Only certain choices of these functions lead to viable screening mechanisms that can circumvent the stringent constraints set by local tests of gravity. In this section, we will discuss two archetypal models of screening: the chameleon and the symmetron.

2.2.1 Chameleon models

In keeping with the principles of effective field theory, the space of coupling functions can be parametrised by expanding $\Omega(\phi)$ and $\varepsilon(\phi)$ in powers of ϕ. Well below the cutoff for this theory, it will suffice to truncate to leading order, in which case

$$\Omega(\phi) = 1 + \frac{\phi}{M_c} + \mathcal{O}(\phi^2), \qquad \varepsilon(\phi) = 1 + \frac{\phi}{M_\gamma} + \mathcal{O}(\phi^2). \tag{2.27}$$

A convenient choice of normalisation [$\Omega(0) = \varepsilon(0) = 1$] has been made such that the matter sector is described by the usual Standard Model Lagrangian when the scalar field is absent. Within the EFT's regime of validity [i.e., when $\phi \ll \min(M_c, M_\gamma)$], these coupling functions always remain close to unity, thus it proves more convenient to speak in terms of the dimensionless coupling strengths

$$\beta_m(\phi) := M_{\rm Pl}\frac{\mathrm{d}\log\Omega}{\mathrm{d}\phi}, \qquad \beta_\gamma(\phi) := M_{\rm Pl}\frac{\mathrm{d}\log\varepsilon}{\mathrm{d}\phi} \tag{2.28}$$

when discussing the scalar's interactions with matter and electromagnetism. For the chameleon, one has that $\beta_m = M_{\rm Pl}/M_c$ and $\beta_\gamma = M_{\rm Pl}/M_\gamma$.

The self-interaction potential $V(\phi)$, in contrast, is assumed to arise from nonperturbative effects [128] and so should not be expanded as a power series. Indeed, treating it nonperturbatively is essential for the realisation of the screening mechanism. The prototypical choice is an inverse power-law potential of the Ratra–Peebles form [23, 70, 392],[5]

$$V(\phi) = \frac{\Lambda^{4+n}}{\phi^n} \qquad (n > 0). \tag{2.29}$$

[5]Note, however, that inverse power laws are somewhat unnatural in an EFT, and so (2.29) should really be viewed as just a *fiducial* model for runaway potentials — chosen for its analytic tractability. Another point worth noting is that the chameleon mechanism can also be realised with non-runaway, positive power-law potentials; i.e., $V(\phi) \propto \phi^{2s}$ with integer values of $s \geq 2$ [393], but such models will not be explored here.

Following the invention of this model, it was quickly observed that torsion balance experiments constrained the model parameter $\Lambda \lesssim 2.4\,\text{meV}$ [70] for matter couplings of gravitational strength ($M_c \sim M_{\text{Pl}}$). That an Earth-bound experiment imposed a bound on Λ close to the dark energy scale was considered of great interest, especially since the chameleon was designed primarily as a string-inspired alternative for late-time acceleration. However, the bare potential in (2.29) with $\Lambda = 2.4\,\text{meV}$ cannot drive an accelerating Universe on its own while remaining consistent with cosmological observations [128, 129]. This can be remedied by choosing a different potential of the form $V(\phi) = \Lambda^4 W(\Lambda^n/\phi^n)$, where W is some dimensionless function. Whenever $\Lambda \ll \phi$, this new potential can be Taylor expanded to recover (2.29) along with a cosmological-constant-like contribution of the right value: $V(\phi) \approx \Lambda^4 + \Lambda^{4+n}/\phi^n$. The result is a theory that is identical to ΛCDM on large scales, but which predicts new phenomenology on smaller scales due to the dynamical nature of the scalar field. For this reason, most studies have fixed $\Lambda = 2.4\,\text{meV}$ when constraining the chameleon, although in principle all of parameter space is open to exploration. We will adopt this latter, more open-minded approach in this thesis. Note that whether or not the constant Λ^4 term is present in the potential will be irrelevant for our purposes, since its effect is negligible on small scales.

With these definitions, the equation of motion for the chameleon in (2.6b) can be rewritten as

$$\nabla^\mu \nabla_\mu \phi = V'_{\text{eff}}(\phi) \tag{2.30}$$

in terms of an effective potential given by

$$V_{\text{eff}}(\phi) = \frac{\Lambda^{4+n}}{\phi^n} + \left(\frac{\rho}{M_c} + \frac{\rho_{\text{em}}}{M_\gamma}\right)\phi. \tag{2.31}$$

The above expression assumes the presence of an ambient matter density ρ and electromagnetic fields with energy density ρ_{em}. As we must always have $\Omega(\phi) \approx 1$ and $\varepsilon(\phi) \approx 1$ in the EFT's regime of validity, the distinction between the different definitions of densities (ρ, ρ_{J}, and ρ_c) highlighted in Section 2.1.1 becomes irrelevant.

To see how screening arises in this model, consider placing a point particle of mass m at the origin and surrounding it by the uniform ambient densities ρ and ρ_{em}. In the absence of this particle, the scalar field relaxes to the constant value ϕ_{min} that minimises its effective potential, $V'_{\text{eff}}(\phi_{\text{min}}) = 0$. Now assuming that this particle is light enough (a more precise condition will be derived in the next section) that it only sources a weak perturbation $\varphi = \phi - \phi_{\text{min}}$, a good approximation for φ can be obtained by linearising (2.30) and solving the resulting equation of motion,

$$[\nabla^\mu \nabla_\mu - m^2_{\text{eff}}(\phi_{\text{min}})]\,\varphi = \frac{\beta_m m}{M_{\text{Pl}}}\delta^{(3)}(\mathbf{x}), \tag{2.32}$$

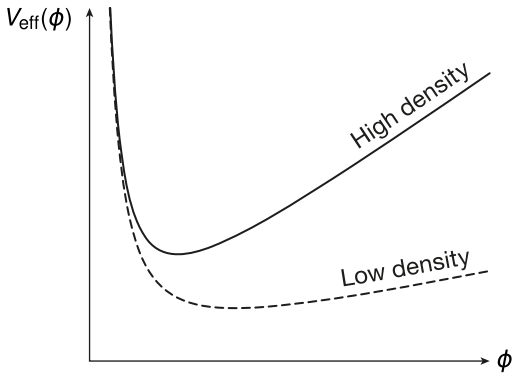


Figure 2.1 Effective potential for the chameleon model. The position and curvature about the minimum varies according to the ambient density, thus allowing the scalar to dynamically alter its effective mass and suppress the range of its fifth force based on its local environment.

where the effective mass for φ is given by $m^2_{\rm eff}(\phi) = V''_{\rm eff}(\phi)$. On flat space, the particular solution to this equation is the familiar Yukawa potential

$$\varphi = -\frac{\beta_m\, m}{M_{\rm Pl}} \frac{e^{-m_{\rm eff}(\phi_{\rm min})r}}{4\pi r}. \tag{2.33}$$

Crucially, observe that the range of this potential depends on the value of $m_{\rm eff}$, which reads

$$m^2_{\rm eff}(\phi_{\rm min}) = n(n+1)\Lambda^{4+n}\left[\frac{1}{n\Lambda^{4+n}}\left(\frac{\rho}{M_c} + \frac{\rho_{\rm em}}{M_\gamma}\right)\right]^{(2+n)/(1+n)} \tag{2.34}$$

after substituting in the expression for $\phi_{\rm min}$. Evidently, the chameleon mechanism operates by allowing the scalar to vary its effective mass based on the ambient energy density. The scalar mediates a long-range fifth force in low-density environments, but this force becomes increasingly short ranged as the ambient density increases. This is illustrated in Figure 2.1.

2.2.2 Symmetron models

Symmetron models, in contrast, screen themselves by relying on a $\mathbb{Z}_2$ symmetry that is spontaneously broken by changes in the ambient density. They have a Higgs-like, double-well potential

$$V(\phi) = -\frac{1}{2}\mu^2\phi^2 + \frac{\lambda}{4}\phi^4, \tag{2.35}$$

which drives the field to a nonzero expectation value $\phi \to \pm\mu/\sqrt{\lambda}$ when in pure vacuum. As with the chameleon, this model cannot drive late-time acceleration on its own unless a cosmological-constant-like term is added to the potential [72]. Also like the chameleon, the symmetron's coupling functions

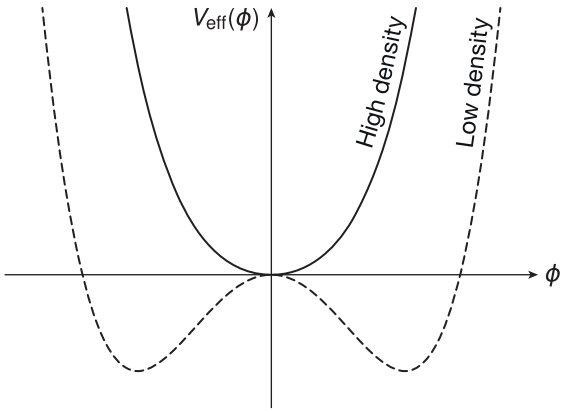


Figure 2.2 Effective potential for the symmetron model. In low-density environments, the presence of two, degenerate local minima in the effective potential forces the scalar to spontaneously break its $\mathbb{Z}_2$ symmetry and acquire a nonzero expectation value. Meanwhile, in high-density environments, the scalar is driven to the single global minimum at $\phi = 0$ and linearly decouples from the Standard Model.

can be expanded as power series that we may truncate to leading order provided $\phi \ll \min(M_s, M_\gamma)$:

$$\Omega(\phi) = 1 + \frac{\phi^2}{2M_s^2} + \mathcal{O}(\phi^4), \qquad \varepsilon(\phi) = 1 + \frac{\phi^2}{2M_\gamma^2} + \mathcal{O}(\phi^4). \tag{2.36}$$

With these definitions, the symmetron can be seen to obey the same equation of motion as in (2.30), albeit with an effective potential given by

$$V_{\rm eff}(\phi) = \frac{1}{2}\mu^2 \left(\frac{\rho}{\mu^2 M_s^2} + \frac{\rho_{\rm em}}{\mu^2 M_\gamma^2} - 1 \right) \phi^2 + \frac{\lambda}{4}\phi^4. \tag{2.37}$$

At sufficiently low densities ($\rho M_s^{-2} + \rho_{\rm em} M_\gamma^{-2} < \mu^2$), the coefficient of the quadratic term above becomes negative, and so causes the symmetron to relax to a nonzero expectation value. In this configuration, the field is said to be in its *broken phase*. Conversely, the field is driven back to zero and the $\mathbb{Z}_2$ symmetry restored when sufficiently large densities render this coefficient positive; see also Figure 2.2. The field is said to be in its *unbroken phase* in this case.

Repeating the same arguments as for the chameleon, it is straightforward to show that a point particle of mass m sources a perturbation to the symmetron field given by the same formula [cf. (2.33)]

$$\varphi = -\frac{\beta_m(\phi_{\rm min})\, m}{M_{\rm Pl}} \frac{e^{-m_{\rm eff}(\phi_{\rm min})r}}{4\pi r}. \tag{2.38}$$

As the symmetron's coupling strengths are field dependent,

$$\frac{\beta_m(\phi)}{M_{\rm Pl}} = \frac{\phi}{M_s^2}, \qquad \frac{\beta_\gamma(\phi)}{M_{\rm Pl}} = \frac{\phi}{M_\gamma^2}, \tag{2.39}$$

and because the minimum of the effective potential depends on the ambient density,

$$\phi_{\rm min} = \begin{cases} \pm\dfrac{\mu}{\sqrt{\lambda}}\left(1 - \dfrac{\rho}{\mu^2 M_s^2} - \dfrac{\rho_{\rm em}}{\mu^2 M_\gamma^2}\right)^{1/2} & \text{(broken phase)} \\ 0 & \text{(unbroken phase),} \end{cases} \tag{2.40}$$

the symmetron linearly decouples from the Standard Model in dense environments. In this region, point particles do not source linear perturbations to the symmetron field and are said to be screened.

Although not essential to how this screening mechanism operates, it is worth noting that the range of the symmetron's fifth force is also limited by its effective mass. At the minimum of the effective potential, it can be shown that

$$m_{\rm eff}^2(\phi_{\rm min}) = \mu^2 \left| \frac{\rho}{\mu^2 M_s^2} + \frac{\rho_{\rm em}}{\mu^2 M_\gamma^2} - 1 \right| \times \begin{cases} 2 & \text{(broken phase)} \\ 1 & \text{(unbroken phase).} \end{cases} \tag{2.41}$$

2.3 Thin-shell effect

Our discussion in the preceding section reveals how a scalar screens the effects of its fifth force by dynamically varying either its effective mass or its coupling strengths to the Standard Model. This description suffices for understanding how chameleon-like scalars interact with small elementary particles, but fails to capture a much richer phenomenology that emerges on macroscopic scales as a result of nonlinearities becoming important. The remainder of this chapter will describe two inherently nonlinear phenomena: violations of the strong equivalence principle via the "thin-shell effect" are discussed here, whereas the formation of vacuum bubbles will be the subject of Section 2.4.

Consider a sphere of radius R and constant density ρ_0 surrounded by a uniform medium with ambient density ρ_∞. Our goal is to solve the equation of motion in (2.16) for the scalar-field profile sourced by this density contrast. (For simplicity, we shall assume that no electromagnetic fields are present in what follows, although including them at a later stage is straightforward.) Although an exact solution is only accessible numerically, it turns out that a surprisingly good analytic approximation can be achieved by dividing the problem into multiple regions, linearising (2.16) in each region, and sewing the resulting solutions together by imposing matching conditions at the boundaries [70, 72, 76, 77, 394, 395]. This is essentially an application of the method of matched asymptotic expansions.

Approximate solution In the simplest case, we divide the problem into two regions. At radii exterior to the sphere ($r \geq R$), we perturb (2.16) about the value ϕ_∞ that minimises the effective potential in

this region. The approximate equation of motion then reads

$$[\nabla^\mu \nabla_\mu - m_\infty^2](\phi - \phi_\infty) = 0, \tag{2.42}$$

where $m_\infty^2 \equiv m_{\rm eff}^2(\phi_\infty)$. Let us further assume that the densities ρ_0 and ρ_∞ are not too large that the underlying spacetime can be taken to be flat, just as before. The most general static, spherically symmetric solution to the above equation is then

$$\phi(r) = \phi_\infty + Q\frac{e^{-m_\infty(r-R)}}{4\pi r} \qquad (r \geq R), \tag{2.43}$$

where Q is an aptly named integration constant to be determined and the boundary condition $\phi \to \phi_\infty$ as $r \to \infty$ has been imposed. This condition reflects the fact that the scalar field would relax to the constant value ϕ_∞ everywhere if the sphere were absent.

In the sphere's interior ($r \leq R$), we instead perturb (2.16) about an as-of-yet unknown central field value ϕ_0. This value need not minimise the effective potential in this region; hence, the corresponding linearised equation of motion is

$$[\nabla^\mu \nabla_\mu - m_0^2](\phi - \phi_0) = V'_{\rm eff,0}, \tag{2.44}$$

where the first derivative of the effective potential

$$V'_{\rm eff}(\phi) = V'(\phi) + \frac{\beta_m(\phi)\rho}{M_{\rm Pl}}, \tag{2.45}$$

and I write $V'_{\rm eff,0} \equiv V'_{\rm eff}(\phi_0)$ as shorthand. Imposing regular boundary conditions ($\mathrm{d}\phi/\mathrm{d}r = 0$) at the origin leads to the solution

$$\phi(r) = \phi_0 - \frac{V'_{\rm eff,0}}{m_0^2} + c_0\frac{\sinh m_0 r}{4\pi r} \qquad (r \leq R), \tag{2.46}$$

where c_0 is another integration constant.

The solutions in these two regions can now be sewn together by requiring that ϕ and its gradient are both continuous across the boundary at $r = R$. These conditions are satisfied when

$$\frac{c_0}{4\pi} = \frac{(\phi_\infty - \phi_0 + V'_{\rm eff,0}/m_0^2)(1 + m_\infty R)}{m_0\cosh(m_0 R) + m_\infty \sinh(m_0 R)}, \tag{2.47}$$

$$\frac{Q}{4\pi} = -\frac{(\phi_\infty - \phi_0 + V'_{\rm eff,0}/m_0^2)[m_0 R - \tanh(m_0 R)]}{m_0 + m_\infty \tanh(m_0 R)}. \tag{2.48}$$

Finally, the central field value ϕ_0 is determined by demanding the self-consistency condition $\phi(0) = \phi_0$, which yields the implicit equation

$$\frac{(\phi_\infty - \phi_0 + V'_{\text{eff},0}/m_0^2)(1 + m_\infty R)}{\cosh(m_0 R) + (m_\infty/m_0)\sinh(m_0 R)} = \frac{V'_{\text{eff},0}}{m_0^2}. \tag{2.49}$$

Limiting behaviour To make sense of these equations, it helps to consider their limiting behaviour when the dimensionless combination $m_0 R$ is either very small or very large. When $m_0 R \ll 1$, the self-consistency condition in (2.49) reduces to $\phi_0 \sim \phi_\infty + \mathcal{O}(m_0^2 R^2)$; indicating that the presence of the sphere only weakly perturbs the scalar field. In this limit, the nonlinearities in the bare potential $V(\phi)$ are negligible; hence, the effective potential in the interior is dominated by the matter-coupling term, $V'_{\text{eff},0} \sim \beta_m(\phi_\infty)\rho_0/M_{\text{Pl}}$. Further noting that $m_\infty^2 < m_0^2$ when $\rho_\infty < \rho_0$, (2.48) reduces to $Q \sim -V'_{\text{eff},0} 4\pi R^3/3$. This can also be written as

$$Q \sim -\frac{\beta_m(\phi_\infty) M}{M_{\text{Pl}}} \qquad (m_0 R \ll 1) \tag{2.50}$$

in terms of the sphere's total mass M ($= 4\pi\rho_0 R^3/3$). Comparing this result with the discussion around (2.16) reveals that the quantity Q is simply the total scalar charge of this macroscopic object. Moreover, note that this result is identical to what we would obtain from the linearised approximation in Section 2.2, which we now learn is valid provided $m_0 R \ll 1$.

Now consider the opposite limit, $m_0 R \gg 1$. In this regime, (2.49) reduces to $V'_{\text{eff},0} \sim 0$; indicitating that the central field value ϕ_0 is minimising the effective potential in the sphere's interior. Moreover, $c_0 \sim 0$ in this limit; hence, from (2.46) we see that the scalar field is essentially frozen at $\phi(r) \sim \phi_0$ for $r < R$. Still, this has to connect smoothly to the solution in (2.43) for $r > R$, and so it must be that only a thin shell on the sphere's surface is responsible for sourcing the scalar-field profile in the exterior, as illustrated in Figure 2.3. The total charge of this shell is given by (2.48), which simplifies to $Q \sim -4\pi R(\phi_\infty - \phi_0)$ when making the reasonable assumption that this sphere is much denser than its environment ($\rho_0 \gg \rho_\infty$), such that $m_0^2 \gg m_\infty^2$. It is instructive to rewrite this as

$$Q \sim -\frac{\beta_m(\phi_\infty) M}{M_{\text{Pl}}} \left(\frac{3\delta R}{R}\right) \qquad (m_0 R \gg 1), \tag{2.51}$$

where the quantity δR can be interpreted as the thickness of the shell. Explicitly, we have that

$$\frac{\delta R}{R} = \frac{M_{\text{Pl}}(\phi_\infty - \phi_0)}{\beta_m(\phi_\infty)\rho_0 R^2}. \tag{2.52}$$

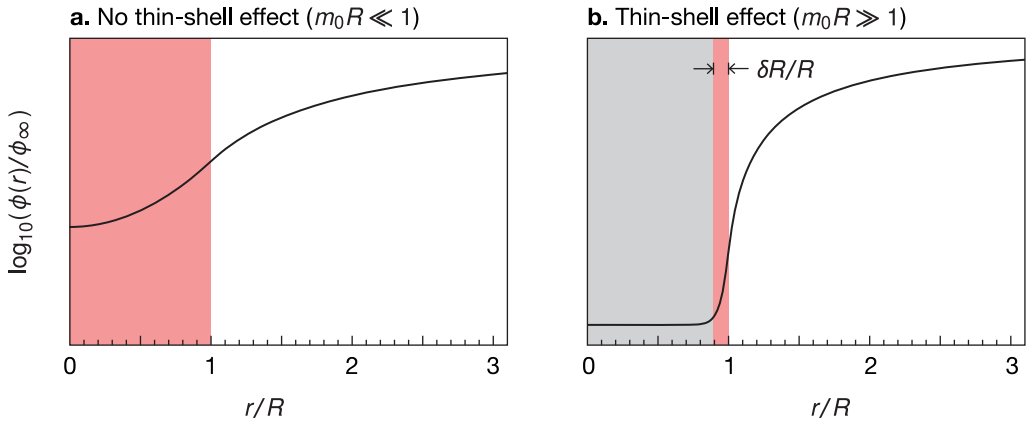


Figure 2.3 Scalar-field profile sourced by a uniform-density sphere of radius R for two illustrative values of the effective mass m_0. When $m_0R \ll 1$, the field is sourced by the entire body of the sphere, but only a thin shell of width δR contributes when $m_0R \gg 1$. The pink shaded regions indicate where in the sphere the scalar varies appreciably, while the grey shaded region indicates where the field is frozen at the minimum of its local effective potential. The approximate solutions in (2.43) and (2.46) were used to draw the black solid curves, but tick marks along the two vertical axes have been omitted as the precise value of $\phi(r)$ is not essential here.

It is now worth verifying that this quantity is always small. Using (2.45) and the fact that $V'_{\rm eff,0} \sim 0$, we obtain $\rho_0/M_{\rm Pl} \sim -V'(\phi_0)/\beta_m(\phi_0)$, which can be used to eliminate ρ_0 from (2.52). We then eliminate $V'(\phi_0)$ by making use of the scaling relation $|V'(\phi_0)/\phi_0| \sim V''(\phi_0) \sim m_0^2$, which is valid for both chameleon and symmetron models up to an $\mathcal{O}(1)$ constant. Also noting that $\phi_\infty \gg \phi_0$ when $\rho_0 \gg \rho_\infty$, we arrive at the expression

$$\frac{\delta R}{R} \sim \left(\frac{\beta_m(\phi_0)}{\beta_m(\phi_\infty)}\frac{\phi_\infty}{\phi_0}\right)\frac{1}{m_0^2R^2}. \tag{2.53}$$

For symmetrons, the ratios in parentheses together evaluate to unity, thus we indeed have that $\delta R \ll R$ whenever $m_0R \gg 1$. For chameleons, this is not always the case and we uncover the more stringent condition $m_0R \gg \sqrt{\phi_\infty/\phi_0}$ for the onset of thin-shell screening.

Having done the math, let us turn to the underlying physical picture. We see that when the scalar has a local Compton wavelength $\sim m_0^{-1}$ in the interior that is much smaller than the object's characteristic size R (i.e., $m_0R \gg 1$), it will settle at the minimum of its effective potential in this region, in which case, thin-shell screening occurs. If instead $m_0R \ll 1$, the scalar has insufficient room to reach this minimum value, thus allowing the full mass of the object to contribute to its scalar charge Q.

The attractive fifth force between this macroscopic sphere and a point particle of mass m situated at a distance $r \gg R$ is given by [78]

$$|F| = \frac{Q_{\rm sph}Q_{\rm pp}}{4\pi r^2}e^{-m_\infty r}, \tag{2.54}$$

where the sphere's charge $Q_{\rm sph}$ is given either by (2.50) or (2.51), while the point particle has charge $Q_{\rm pp} = -\beta_m(\phi_\infty)m/M_{\rm Pl}$. Because the value of $Q_{\rm sph}$ strongly depends on the sphere's internal constitution, this thin-shell effect leads to a violation of the strong equivalence principle. As we discussed in Chapter 1, dwarf galaxies are prime targets for probing this phenomenon [74, 79–81].

2.4 Vacuum bubbles

The formalism developed in the previous section can now be applied to study a related phenomenon wherein a nontrivial scalar-field profile, called a vacuum bubble, develops inside a cavity. A precise understanding of this phenomenon is important for the analysis of laboratory experiments performed in high vacuum, since the presence of a bubble can lead to a fifth force being exerted on small particles within the vacuum chamber. Indeed, the results in this section will play an important role in our reanalysis of the one-electron quantum cyclotron experiment in Chapter 4.

2.4.1 Approximate solutions via matched asymptotic expansions

As a rudimentary model for a vacuum chamber, consider a spherical cavity of radius R and density ρ_0 surrounded by walls of uniform density ρ_∞. Immediately, it should be apparent that the calculations in the previous section can be adapted to this scenario, with the only difference being that we now have $\rho_\infty \gg \rho_0$. In the exterior region ($r \geq R$), the scalar field settles at the minimum ϕ_∞ of its effective potential within a distance, or *skin depth*, $\sim m_\infty^{-1}$ from the boundary at $r = R$. While this toy model assumes that the cavity is surrounded by walls of infinite extent, we can be assured that the scalar field will reach this minimum inside a wall of finite size as long as this wall is much thicker than m_∞^{-1}. (To give an example of the typical scales involved in laboratory experiments, note that we will require $m_\infty^{-1} \ll 0.1$ mm in Chapter 4.) When this condition is satisfied, the behaviour of the scalar in the interior of this vacuum chamber is essentially decoupled from whatever is transpiring beyond its walls.

In this interior region ($r \leq R$), the field climbs to a maximum value ϕ_0 at the centre, which is determined by the self-consistency condition in (2.49). Two reasonable assumptions make it possible to obtain an approximate, analytic solution to this implicit equation:

(1) *Pure-vacuum approximation*: We will assume that the density ρ_0 inside the cavity is negligible, such that the effective potential can be replaced by the bare potential, $V'_{\rm eff,0} \sim V'(\phi_0)$.

(2) *Zero-skin-depth approximation*: We will further assume that the density in the walls is sufficiently large that $m_\infty R \gg 1$ and $m_\infty^2 \gg m_0^2$. In this regime, $\phi_0 - \phi_\infty \sim \phi_0$ and the skin depth $\sim m_\infty^{-1}$ of the exterior solution approaches zero.

Chameleon As a result of these approximations, (2.49) reduces to

$$\frac{\sinh(m_0 R)}{m_0 R} = n + 2 \tag{2.55}$$

for the case of the chameleon model. This simple equation reveals that a chameleon field, when inside a vacuum chamber, adjusts its local Compton wavelength to be comparable to the size of the cavity, $m_0 R \sim \mathcal{O}(1)$ [70]. Using the definition of its effective mass and writing $\mathrm{sinhc}(z) := \sinh(z)/z$ for brevity, (2.55) can be inverted to yield

$$\phi_0 = \left(\frac{n(n+1)\Lambda^{4+n} R^2}{[\mathrm{sinhc}^{-1}(n+2)]^2} \right)^{1/(2+n)} . \tag{2.56}$$

Symmetron For the symmetron, making the pure-vacuum approximation is unnecessary as the bare and effective potentials have the same functional form, differing only in the values of their coefficients. After defining $\mu_0^2 = \mu^2 - \rho_0 M_s^{-2}$, (2.49) reduces to

$$\left(1 + \frac{\mu_0^2 - \lambda\phi_0^2}{m_0^2} [1 - \mathrm{sinhc}(m_0 R)] \right) \phi_0 = 0, \tag{2.57}$$

where the effective mass m_0 is given by $m_0^2 = 3\lambda\phi_0^2 - \mu_0^2$. To simplify this equation further, it will be convenient to normalise ϕ_0 by the maximum possible value it can attain, thus let $\hat{\phi}_0 := \phi_0/(\mu_0/\sqrt{\lambda})$. Assuming that symmetry breaking occurs inside the cavity (i.e., let $\mu_0^2 > 0$), (2.57) simplifies to

$$\frac{\hat{\phi}_0}{3\hat{\phi}_0^2 - 1} \left[2\hat{\phi}_0^2 - (1 - \hat{\phi}_0^2)\, \mathrm{sinhc}\left(\mu_0 R \sqrt{3\hat{\phi}_0^2 - 1} \right) \right] = 0. \tag{2.58}$$

This is still an implicit equation for $\hat{\phi}_0$, but it is one that is straightforward to solve numerically.

To gain physical insight into this equation, consider first its limiting behaviour. The sinhc function diverges in the limit $\mu_0 R \to \infty$, thus we must have that $\hat{\phi}_0^2 \to 1$ for a consistent solution. This squares with our intuition, as we would expect the symmetron to easily minimise its effective potential in the cavity when its Compton wavelength $m_0^{-1} \ll R$. For smaller values of $\mu_0 R$, (2.58) starts to admit multiple roots, but the physical solution can be singled out by choosing the branch that is smoothly connected to the point $\hat{\phi}_0^2(m_0 R \to \infty) = 1$. The value of $\hat{\phi}_0^2$ along this physical branch decreases until it reaches $\hat{\phi}_0^2 = 1/3$ at some threshold value of $\mu_0 R$, below which $\hat{\phi}_0 = 0$ is the only physical solution. This reveals that for a given value of μ_0, a vacuum cavity must be larger than a certain size if it is to

accommodate a symmetron bubble. To summarise, the central field value is given by

$$\hat{\phi}_0^2(\mu_0 R) = \begin{cases} \text{Largest root of (2.58)} & (1/3 < \hat{\phi}_0^2 \leq 1) \\ 0 & \text{(otherwise).} \end{cases} \tag{2.59}$$

The sharp jump between 0 and $1/3$ is an artefact of performing a matched asymptotic expansion; an exact solution is expected to smooth over this discontinuity. Speaking of which, ...

2.4.2 An exact solution for the symmetron

It turns out that choosing a different shape for the vacuum cavity will allow us to obtain an exact, analytic solution for the symmetron field. Specifically, we shall now turn our attention to a plane-parallel cavity in the region $z \in [-\ell, \ell]$ that is surrounded by walls on either side extending to infinity. As before, let ρ_0 be the uniform density in the cavity and ρ_∞ the density in the walls.

Interior solution The problem is translationally invariant along the x and y directions; hence, the symmetron's equation of motion inside the cavity is simply

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} = -\mu_0^2\phi + \lambda\phi^3 \qquad (|z| \leq \ell), \tag{2.60}$$

where recall $\mu_0^2 = \mu^2 - \rho_0 M_s^{-2} > 0$. Once again, it proves convenient to normalise this scalar field by the maximum possible value it can attain, thus let $\hat{\phi} = \phi/(\mu_0/\sqrt{\lambda})$, which can take values in the range $[-1, 1]$. Integrating (2.60) up once then yields

$$\frac{1}{\mu_0^2}\left(\frac{\mathrm{d}\hat{\phi}}{\mathrm{d}z}\right)^2 = -(\hat{\phi}^2 - \hat{\phi}_0^2) + \frac{1}{2}(\hat{\phi}^4 - \hat{\phi}_0^4), \tag{2.61}$$

having chosen boundary conditions such that $\mathrm{d}\hat{\phi}/\mathrm{d}z = 0$ at the origin with $\hat{\phi}(0) = \hat{\phi}_0$. As before, the central field value $\hat{\phi}_0$ is an as-of-yet unknown integration constant to be determined. The above equation can be integrated once more to obtain [2, 92, 396]

$$-\mu_0 z\frac{\hat{\phi}_0}{\sqrt{2v^2}} = F\left(\sin^{-1}\left(\hat{\phi}(z)/\hat{\phi}_0\right), v\right) - K(v), \tag{2.62}$$

where $v^2 := \hat{\phi}_0^2/(2 - \hat{\phi}_0^2)$ and the positive branch $\hat{\phi}_0 > 0$ has been chosen without loss of generality. This implicit solution is expressed in terms of the elliptic integrals of the first kind, defined such that

$$F(u, v) := \int_0^u \frac{\mathrm{d}\vartheta}{\sqrt{1 - v^2\sin^2\vartheta}} \tag{2.63}$$

and $K(v) := F(\pi/2, v)$. To invert (2.62), we make use of the Jacobi elliptic functions

$$\mathrm{sn}(u, v) := \sin F^{-1}(u, v), \tag{2.64a}$$

$$\mathrm{cn}(u, v) := \cos F^{-1}(u, v), \tag{2.64b}$$

$$\mathrm{dn}(u, v) := \sqrt{1 - v^2 \mathrm{sn}^2(u, v)}, \tag{2.64c}$$

and the identity [397]

$$\mathrm{sn}(u + K(v), v) = \frac{\mathrm{cn}(u, v)}{\mathrm{dn}(u, v)} =: \mathrm{cd}(u, v) \tag{2.65}$$

to obtain the final solution [2, 396]

$$\hat{\phi}(z) = \hat{\phi}_0 \,\mathrm{cd}\left(\mu_0 z \frac{\hat{\phi}_0}{\sqrt{2v^2}}, v\right) \qquad (|z| \leq \ell). \tag{2.66}$$

Self-consistency condition The symmetron field in the walls is governed by the equation

$$\frac{1}{\mu_0^2}\left(\frac{\mathrm{d}\hat{\phi}}{\mathrm{d}z}\right)^2 = \left(\frac{\mu_\infty}{\mu_0}\right)^2 \hat{\phi}^2 + \frac{1}{2}\hat{\phi}^4 \qquad (|z| \geq \ell), \tag{2.67}$$

where $\mu_\infty^2 = \rho_\infty M_s^{-2} - \mu^2$ must be positive if its $\mathbb{Z}_2$ symmetry is to be restored in this region. Accordingly, boundary conditions have been chosen such that $\hat{\phi} \to 0$ as $|z| \to \infty$. If we were so inclined, (2.67) can now be solved in a similar fashion to (2.61). The resulting integration constant is then determined by requiring continuity of $\hat{\phi}$ at the boundaries $|z| = \ell$ and a self-consistency equation for $\hat{\phi}_0$ follows from also demanding continuity of the first derivatives; see Ref. [396]. That being said, realistic vacuum chambers typically satisfy the zero-skin-depth approximation, in which case these steps are rendered superfluous. In this limit, the symmetron rapidly plunges to zero once inside the walls, thus we may simply take $\hat{\phi}(\ell) = 0$ as our self-consistency condition [2]. This is an implicit equation for $\hat{\phi}_0$ that is easily solved numerically.

In Figure 2.4, $|\hat{\phi}(\ell)|$ is drawn as a function of $\hat{\phi}_0$ for different values of $\mu_0 \ell$. Above a certain threshold, notice that the equation $\hat{\phi}(\ell) = 0$ begins to admit multiple roots, but unlike the approximate solution in (2.58), each of these roots corresponds to a physical solution. The smaller roots generate field configurations containing domain walls, as shown in the right panel of Figure 2.4. The lesson here is that we should view a symmetron bubble as a solitonic object with a minimum size on the order of μ_0^{-1}. When the size of the cavity is too small, there is insufficient room to accommodate a bubble, and thus the field remains at $\hat{\phi} = 0$ everywhere. Above a threshold value of $\mu_0 \ell \sim \mathcal{O}(1)$, however, the

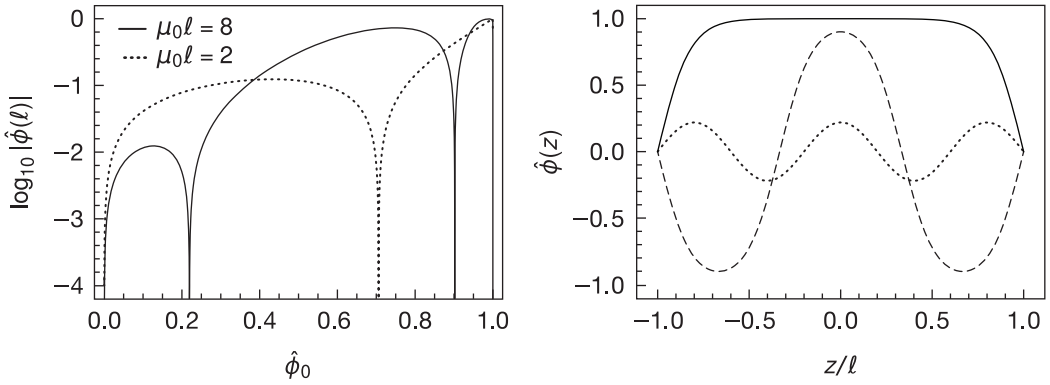


Figure 2.4 The central field value $\hat{\phi}_0$ of a symmetron bubble in a plane-parallel cavity of width 2ℓ is determined by finding the zero(s) of the function $\hat{\phi}(\ell)$, shown on the left for two illustrative values of $\mu_0\ell$. In the right panel, field profiles $\hat{\phi}(z)$ with corresponding central field values $\hat{\phi}_0 \in \{0.22, 0.90, 1.00\}$ for $\mu_0\ell = 8$ are shown.

cavity is finally large enough that a bubble can be accommodated, and so the field starts to grow. For even greater values of ℓ, the symmetron's Compton wavelength is now much smaller than the size of the cavity; hence, it is possible to fit multiple bubbles separated by domain walls.

Only the even-parity solutions are shown in Figure 2.4, but the symmetron will also admit odd-parity solutions, which one can obtain by solving (2.60) with different boundary conditions; namely, $\hat{\phi}(0) = 0$ and $\mathrm{d}\hat{\phi}/\mathrm{d}z \neq 0$ at the origin. (See Ref. [396] for more details.) In an experimental setup, though, it is natural to expect that the symmetron will occupy the state of lowest energy, corresponding to the solution with no domain walls. This is given by the largest value of $\hat{\phi}_0$ that solves $\hat{\phi}(\ell) = 0$, which is drawn as a function of $\mu_0\ell$ in Figure 2.5.

Exact vs approximate methods To conclude this chapter, it will be interesting to assess how well the approximate solution in Section 2.4.1 fares compared to this exact result. It is necessary that we compare like with like; hence, we should use

$$\frac{\hat{\phi}_0}{3\hat{\phi}_0^2 - 1}\left[2\hat{\phi}_0^2 - (1-\hat{\phi}_0^2)\cosh\left(\mu_0\ell\sqrt{3\hat{\phi}_0^2 - 1}\right)\right] = 0 \tag{2.68}$$

in place of (2.58) as the relevant self-consistency equation, which one obtains after repeating the analysis for a plane-parallel cavity. The largest root of (2.68) as a function of $\mu_0\ell$ is also drawn in Figure 2.5. First notice that this curve does not exhibit the same jump from $\hat{\phi}_0^2 = 0$ to $1/3$ as in the spherical case, which can only be attributed to the cosh function being more well behaved than the sinhc function. More importantly, we see from Figure 2.5 that the approximate and exact results are in good agreement. The two methods give essentially identical results close to the end points $\hat{\phi}_0 = 0$ and 1, which correspond to the limiting cases $\mu_0\ell \to 0$ and $\mu_0\ell \to \infty$, respectively. It is only at intermediate values, for which

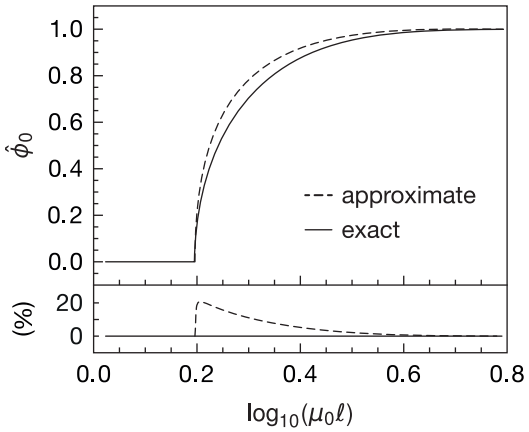


Figure 2.5 Central field value $\hat{\phi}_0$ of a symmetron bubble in a plane-parallel cavity as a function of $\mu_0 \ell$. The bottom panel displays the percentage error between the exact and approximate solutions.

nonlinearities are significant, that the approximate solution does poorly. Even so, it is clear that this rudimentary approximation is able to capture the correct qualitative behaviour of the symmetron, and can therefore be trusted when used to guide our physical intuition for understanding these models.

Chapter 3
The hydrogen atom

In the third volume of his *Lectures on Physics* [398], Feynman asserts that the most dramatic success in the history of quantum mechanics was the understanding of atomic spectra. If so, then most of this success must surely be ascribed to the unique simplicity of the hydrogen atom, whose particular properties allow us to perform calculations at a level of precision that would be unthinkable for most other systems in nature. As we now begin our investigation into the impact of light scalar fields on moving bodies, this simplicity will be a welcome way to ease us into the task.

Our model for atomic hydrogen will be constructed in typical fashion: Its nucleus is taken to be a point particle affixed to the origin of our coordinate system, while the electron in orbit around it will be treated quantum mechanically. This nucleus acts as a classical source of electromagnetic fields and will also give rise to a chameleon-like scalar, if one is active. For gravitational-strength couplings $\beta_m, \beta_\gamma \sim \mathcal{O}(1)$, we would expect the effects from this scalar to be comparable in size to those of the gravitational force between the proton and electron; which is to say, not much, given that the latter is suppressed by some 40 orders of magnitude relative to the Coulomb interaction. (For this reason, we will not include the Newtonian gravitational potential in any of our calculations.) On the other hand, a scalar that couples to matter more strongly than gravity has the potential to lead to interesting effects.

Indeed, Brax and Burrage [399] had already shown almost a decade ago that for matter couplings with strength $\beta_m \gtrsim 10^{15}$, there would be an observable lowering of the energy levels in atomic hydrogen due to the scalar mediating an attractive fifth force between the proton and electron. Their study did not fully exhaust the physical implications of the scalar–photon coupling, however, and we will demonstrate in Section 3.1 that this interaction leads to a kind of vacuum polarisation effect already at the classical level. The next two sections extend the findings of Ref. [399] even further. Section 3.2 examines the impact of embedding this atom in unscreened astrophysical environments, where large scalar-field

gradients may be present, while Section 3.3 takes the nuclear magnetic field into account to study the influence of a chameleon-like scalar on hyperfine structure. In Section 3.4, we then take the first step towards constraining these effects by discussing how they would systematically bias the value of the fine-structure constant inferred from measurements of hydrogen lines if unaccounted for.

3.1 Atomic spectral lines

3.1.1 Fields sourced by a point nucleus

In the interest of generality, we shall suppose that the nucleus has mass m_N and electric charge $+Ze$. The latter gives rise to the electric field

$$\mathbf{E} = \frac{Ze}{4\pi r^2}\,\mathbf{e}_r, \tag{3.1}$$

where $\mathbf{e}_r$ is the unit vector in the radial direction. The presence of this nucleus sources a small perturbation $\varphi = \phi - \phi_0$ to the ambient scalar-field profile ϕ_0, the latter of which may be assumed to be approximately constant over atomic length scales. In the previous chapter, we learnt that point particles like this nucleus are not heavy enough to trigger any nonlinear effects; hence, the perturbation φ is well described by the linearised version of (2.30), which reads

$$(\Box - m_0^2)\varphi = \frac{1}{M_{\rm Pl}}\left(\beta_m m_N \delta^3(\mathbf{x}) - \frac{1}{2}\beta_\gamma \mathbf{E}^2\right). \tag{3.2}$$

On the lhs, $\Box \equiv \eta^{\mu\nu}\partial_\mu\partial_\nu$ is the wave operator on flat space, while $m_0 := m_{\rm eff}(\phi_0) \equiv \sqrt{V''_{\rm eff}(\phi_0)}$ is the scalar's local effective mass, which varies based on the ambient environment. On the rhs, the couplings $\beta_m \equiv \beta_m(\phi_0)$ and $\beta_\gamma \equiv \beta_\gamma(\phi_0)$ are also set by this ambient field value. For notational convenience, we will suppress this explicit dependence on ϕ_0 in what follows.

While it is possible to solve (3.2) as is in terms of exponentials and the exponential integral function, we can make one further simplification by recognising that the scalar's local Compton wavelength $\sim m_0^{-1}$ is typically much larger than the Bohr radius of the atom. (For a concrete example, recall that if this atom were located in the vacuum chamber of some laboratory experiment, then we previously saw in Chapter 2.4 that the chameleon will adjust its local Compton wavelength to be on the order of the size of the cavity.) Consequently, the mass term in (3.2) can be neglected when solving for the behaviour of $\varphi(r)$ in the vicinity of the nucleus. In this region, (3.2) reduces to the Poisson equation, whose particular solution is easily found to be

$$\varphi(r) = -\frac{1}{4\pi M_{\rm Pl}}\left(\frac{\beta_m m_N}{r} + \frac{\beta_\gamma Z^2 e^2}{16\pi r^2}\right). \tag{3.3}$$

This scalar-field profile goes on to exert an attractive fifth force on the electron, which leads to systematic shifts in its energy levels. Additionally, due to its coupling $\varepsilon(\phi)$ to the photon, this scalar-field profile will also induce small corrections to the electric field sourced by the nucleus, which results in further systematic shifts that arise at the same order in G (recall that $G \equiv 1/8\pi M_{\rm Pl}^2$). To see this, perturb $A_\mu \to A_\mu + \delta A_\mu$ in the vacuum Maxwell equations [cf. (2.20)] to obtain

$$\Box\, \delta A^\mu = \frac{\beta_\gamma}{M_{\rm Pl}} F^{\mu\nu}\partial_\nu \phi. \tag{3.4}$$

This first-order equation is written in Lorenz gauge ($\partial_\mu \delta A^\mu = 0$) and the field strength tensor $F_{\mu\nu}$ on the rhs is to be evaluated using the zeroth-order electric field in (3.1).[1] Physically, this equation is telling us that a nontrivial scalar-field profile generates secondary charges and currents that go on to source corrections to the bare electromagnetic fields. Or, put another way, the presence of the scalar field polarises the vacuum (already at the classical level) and causes it to behave like a dielectric medium. For the field profiles in (3.1) and (3.3), the solution to (3.4) is

$$\delta A_\mu = -\delta^0_\mu G\left(\beta_m\beta_\gamma \frac{Zem_N}{4\pi r^2} + \beta_\gamma^2 \frac{Z^3e^3}{96\pi^2 r^3}\right). \tag{3.5}$$

3.1.2 Hamiltonian

The electromagnetic and scalar-field profiles we just derived interact with the electron according to the modified Dirac equation in (2.26). For any spacetime on which a coordinate chart $x^\mu = (t, \mathbf{x})$ can be erected such that spatial hypersurfaces correspond to surfaces of constant t, multiplying (2.26) by $(g^{00})^{-1}\not{e}^0$ gives us the Schrödinger-like equation $i\,\partial_t\psi = H\psi$ with Hamiltonian

$$H = (-g^{00})^{-1}\not{e}^0[\Omega(\phi)m_e + \not{e}^i(\mathbf{p}_i + e\mathbf{A}_i) - i\not{e}^\mu \varpi_\mu] + eA_0, \tag{3.6}$$

where $\mathbf{p}_j := -i\,\partial_j$ is the momentum operator. After taking the flat-space limit and using the expressions in (3.1), (3.3), and (3.5), the resulting Hamiltonian can be written as a sum of two parts: $H = H_0 + H_S$.

The unperturbed Hamiltonian, which can be solved exactly, is

$$H_0 = \gamma^0 m_e + \Sigma^i \mathbf{p}_i - \frac{Z\alpha}{r}. \tag{3.7}$$

Its eigenvalues and eigenfunctions are reviewed in Appendix 3.A. Above, the elementary charge e has been rewritten in terms of the fine-structure constant $\alpha \equiv e^2/(4\pi)$, and $\Sigma^i := \gamma^0\gamma^i$.

[1] In our conventions, the components of the field strength tensor are related to the electric and magnetic fields by the equations $F^{0i} = E^i$ and $F_{ij} = \epsilon_{ijk}B^k$.

The remaining terms in H, meanwhile, all involve the scalar field,

$$H_S = \left(\frac{\beta_m m_e}{M_{\rm Pl}}\right)\gamma^0\varphi + e\,\delta A_0, \tag{3.8}$$

and will be treated as small perturbations. Written out explicitly, we have that

$$H_S = -G\left(\gamma^0\beta_m^2\frac{2m_e m_N}{r} + \gamma^0\beta_m\beta_\gamma\frac{Z^2\alpha m_e}{2r^2} + \beta_m\beta_\gamma\frac{Z\alpha m_N}{r^2} + \beta_\gamma^2\frac{Z^3\alpha^2}{6r^3}\right). \tag{3.9}$$

The first two terms account for the energy shift due to the scalar exerting a fifth force on the electron, while the remaining two are the result of the scalar-field profile φ polarising the vacuum and generating corrections to the electric field. It is worth highlighting the relative sizes of the second and third terms, both of which are proportional to $\beta_m\beta_\gamma$. Because the third term is enhanced by a factor of $\sim 2m_N/m_e$ (~ 4000 in the case of hydrogen) relative to the second, we see that the effect of vacuum polarisation can be just as important as the fifth force in scenarios where electromagnetic fields are present.

3.1.3 Energy shifts

The amount by which H_S shifts each of the electron's energy levels can be calculated in the usual way. At first order in perturbation theory, its expectation value with respect to a given eigenstate ψ of H_0 is

$$\langle H_S\rangle(\psi) := \int \mathrm{d}^3\mathbf{x}\;\psi^\dagger H_S\,\psi. \tag{3.10}$$

To gain a sense for the typical size of each term in H_S, let us evaluate their expectation values with respect to the hydrogen ground state. One should find

$$-2G\beta_m^2 m_e m_N\langle\gamma^0 r^{-1}\rangle(1S_{1/2}) = -(2.4\times10^{-38}\ \text{eV})\,\beta_m^2, \tag{3.11a}$$

$$-\frac{1}{2}G\beta_m\beta_\gamma Z^2\alpha m_e\langle\gamma^0 r^{-2}\rangle(1S_{1/2}) = -(3.5\times10^{-46}\ \text{eV})\,\beta_m\beta_\gamma, \tag{3.11b}$$

$$-G\beta_m\beta_\gamma Z\alpha m_N\langle r^{-2}\rangle(1S_{1/2}) = -(1.3\times10^{-42}\ \text{eV})\,\beta_m\beta_\gamma, \tag{3.11c}$$

after using the closed-form expressions for the radial expectation values found in Ref. [400], which are reproduced in Appendix 3.A for ease of reference.

Caution must be exercised with the last term in (3.9), however. For $S_{1/2}$ and $P_{1/2}$ states, whose wavefunctions are nonzero at the origin, a $1/r^3$ potential is steep enough that it leads to a divergent

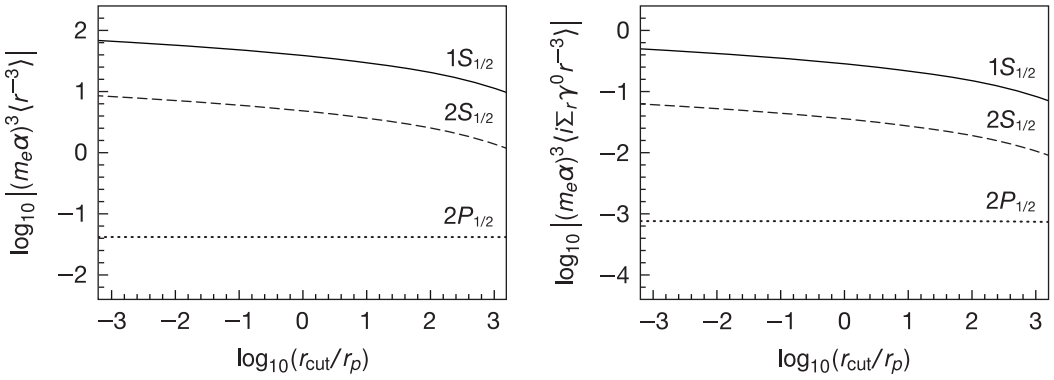


Figure 3.1 Regularised expectation values for a few of the lowest-lying levels in atomic hydrogen.

expectation value. This UV divergence is unphysical, of course, and is signalling the breakdown of the point-particle approximation for the nucleus. In general, this kind of singular behaviour can be dealt with systematically by introducing a convenient regulator and a renormalisation scheme [401]. That said, we will here adopt a simpler, albeit more ad hoc approach that should nevertheless provide a correct order-of-magnitude estimate. The logic is as follows: If we were to UV-complete our model for the nucleus, we should expect the Coulomb potential $eA_0 = -Z\alpha/r$ to be replaced by $eA_0 = -Z\alpha F(\mathbf{x})$ at radii smaller than the nuclear charge radius r_N, where $F(\mathbf{x})$ is some form factor associated with the charge distribution. As a crude example, a uniformly charged sphere will have $F(r) = (2r_N)^{-1}(3 - r^2/r_N^2)$, although the precise form for $F(\mathbf{x})$ will not be essential to our discussion. The key point is that, because r_N is much smaller than the Bohr radius $a_0 \sim (m_e\alpha)^{-1}$, the effect of this form factor on the expectation value $\langle r^{-3}\rangle$ is practically equivalent, as a first approximation, to introducing a hard cutoff.

Therefore, whenever necessary, we shall regularise the expectation value of some operator $\hat{O}$ by introducing a cutoff at the radius $r_{\rm cut}$; i.e.,

$$\langle\hat{O}\rangle(\psi)_{r_{\rm cut}} := \int \mathrm{d}^2\Omega \int_{r_{\rm cut}}^{\infty} r^2 \mathrm{d}r\, \psi^\dagger \hat{O}\psi. \tag{3.12}$$

The precise value of this quantity will naturally depend on our particular choice of cutoff, but Figure 3.1 shows that the result will be of the same order of magnitude for any reasonable choice close to r_N. Taking the cutoff to be equal to the proton charge radius $r_p = 0.85$ fm [402] in the case of atomic hydrogen, our estimate for the ground-state expectation value of the fourth term in H_S is

$$-\frac{1}{6}G\beta_\gamma^2 Z^3\alpha^2\langle r^{-3}\rangle(1S_{1/2})_{r_p} \sim -(1.2\times 10^{-49}\ \mathrm{eV})\,\beta_\gamma^2. \tag{3.11d}$$

3.2 Large background gradients

Thus far, we have assumed that the spacetime around the atomic nucleus is flat and that the ambient scalar-field profile is simply some constant ϕ_0. These assumptions are typical of vacuum chambers in laboratory experiments, but may not be representative of unscreened astrophysical environments, where large scalar-field gradients could be present. Given that spectroscopic surveys of neutral hydrogen constitute an important observable in astrophysics and cosmology, it is important that any systematic effect due to a chameleon-like scalar be understood.

In such circumstances, the nucleus of an atom travels along some worldline Γ with tangent vector u^μ on a background spacetime (g, ϕ). Our goal is to determine the additional corrections to the electron's energy levels engendered by these background fields. Ideally, we would want to evaluate these contributions as perturbations to the same Hamiltonian H_0 in (3.7). Now that the spacetime is curved, care must be taken to choose appropriate coordinates that still have interpretations of physical length and time, at least in the convex normal neighbourhood of Γ. These are the *Fermi normal coordinates* $(t, \mathbf{x})$, whose coordinate time t corresponds to the proper time along Γ and whose spatial origin corresponds to points along Γ [381, 403]. For each point (i.e., for a given time t), the coordinates $\mathbf{x}^i$ measure proper distances along three spatial geodesics whose tangents are orthonormal to one another and to the 4-velocity u^μ.

In this coordinate chart, the background fields can be Taylor expanded in powers of $\mathbf{x}$ to read [403]

$$g_{\mu\nu} = \eta_{\mu\nu} - 2(\mathbf{a}\cdot\mathbf{x})\,\delta^0_\mu\delta^0_\nu + \mathcal{O}(\mathbf{x}^2), \tag{3.13a}$$

$$\phi = \phi_0 + (\nabla\phi_0)\cdot\mathbf{x} + \mathcal{O}(\mathbf{x}^2). \tag{3.13b}$$

Terms of quadratic order and higher will be neglected, as their effects are parametrically suppressed by powers of the small ratio a_0/λ, where λ denotes the characteristic length scale over which the background varies.[2] Assuming no other external forces are acting on the atom, the acceleration $\mathbf{a}$ of its worldline is determined by the fifth force exerted by the scalar gradient $\nabla\phi_0$. Specifically, evaluating (2.15) in Fermi normal coordinates gives us $\mathbf{a} = (\beta_m/M_{\rm Pl})\nabla\phi_0$. In principle, the components of this acceleration vector can still vary as a function of the coordinate time t, but any such variation will typically proceed on timescales much greater than that of the atom. For this reason, we can safely make the adiabatic approximation and take these coefficients to be effectively constant when evaluating expectation values.

[2]The $\mathcal{O}(\mathbf{x}^2)$ terms in the metric lead to tidal deformations, whose effect on the energy levels of atomic hydrogen have been explored in Refs. [404–406].

Substituting the background metric in (3.13a) into the Hamiltonian in (3.6) requires expressions for the vierbeins $\mathrm{e}^{\mu}_{\hat{\alpha}}$ and the spin connection ϖ_μ to first order in $\mathbf{a}$. Because of their orthonormality, it is straightforward to read off from

$$\eta_{\hat{\alpha}\hat{\beta}}\,\mathrm{e}^{\hat{\alpha}}_{\mu}\mathrm{e}^{\hat{\beta}}_{\nu} = g_{\mu\nu} \tag{3.14}$$

that $\mathrm{e}^{\hat{\alpha}}_{\mu} = \delta^{\hat{\alpha}}_{\mu} + (\mathbf{a}\cdot\mathbf{x})\,\delta^{0}_{\mu}\delta^{\hat{\alpha}}_{0}$ and $\mathrm{e}^{\mu}_{\hat{\alpha}} = \delta^{\mu}_{\hat{\alpha}} - (\mathbf{a}\cdot\mathbf{x})\,\delta^{\mu}_{0}\delta^{0}_{\hat{\alpha}}$. Now using the definition of ϖ_μ in (2.23) and noting that the only connection coefficients that are nonzero at this order are $\Gamma^{0}{}_{0i} = \delta_{ij}\Gamma^{j}{}_{00} = \mathbf{a}_i$, we obtain $\varpi_0 = \Sigma^i\mathbf{a}_i/2$ and $\varpi_i = 0$.

Vierbeins and spin connections aside, the acceleration of the worldline and the background scalar gradient also contribute to the Hamiltonian via the vacuum polarisation effect described earlier. Substituting (3.13) into (3.4) and keeping terms that are first order in $\nabla\phi_0$ only, the corrections they induce onto the Coulomb potential of the nucleus are given by[3]

$$\delta A_\mu \supset -\delta^{0}_{\mu}\,(\beta_m - \beta_\gamma)\,(\mathbf{x}\cdot\nabla\phi_0)\frac{Ze}{8\pi M_{\mathrm{Pl}} r}, \tag{3.15}$$

having made the identification $\mathbf{a} = (\beta_m/M_{\mathrm{Pl}})\nabla\phi_0$. Putting everything together, a background scalar gradient is seen to generate the following additional terms in the Hamiltonian:

$$H_S \supset \frac{\beta_m}{M_{\mathrm{Pl}}}(\nabla\phi_0)_i\left(\frac{i}{2}\Sigma^i - \mathbf{x}^i\,\Sigma^j\mathbf{p}_j + \mathbf{x}^i\frac{Z\alpha}{2r}\right) + \frac{\beta_\gamma}{M_{\mathrm{Pl}}}\frac{Z\alpha}{2r}\,(\mathbf{x}\cdot\nabla\phi_0). \tag{3.16}$$

To make sense of these terms, let us align our coordinate system such that $\nabla\phi_0$ points along the z axis. By inspection, we see that the remaining operators have an odd parity overall; hence, the terms in (3.16) shift the electron's energy levels in a way that is completely analogous to the Stark effect. If ψ and ψ' are two degenerate states with nonvanishing matrix element $\langle\psi\,|H_S|\,\psi'\rangle$, the presence of a background scalar gradient $\nabla\phi_0$ breaks this degeneracy. It must be noted, however, that this can occur only if $\nabla\phi_0$ is sufficiently large. Otherwise, such degeneracies would already be lifted by radiative corrections — commonly referred to as the "Lamb shift" [402, 407–410], in which case any response to $\nabla\phi_0$ is nullified at first order in perturbation theory.

This makes it natural to ask: How large must the background gradient be that its effect is not washed out by radiative corrections? Consider the $2P_{1/2} - 2S_{1/2}$ transition, whose Lamb shift is on the order of 10^{-6} eV, as a concrete example. Observing the impact of a background scalar gradient therefore

[3] In astrophysical environments, there may also be large-scale background electromagnetic fields that contribute to the field strength tensor on the rhs of (3.4), but these shall be neglected in our discussion as they are always subdominant to the Coulomb potential in the vicinity of the nucleus.

requires a value of $\nabla\phi_0$ such that $|H_S| \gg 10^{-6}$ eV. Taking $\Sigma^i \sim Z\alpha$ and $r \sim \mathbf{x}^i \sim 1/\mathbf{p}^i \sim a_0$ to obtain an order-of-magnitude estimate [404], we find that we must have

$$M_{\rm Pl}^{-1} \max(\beta_m, \beta_\gamma) Z |\nabla\phi| \gg 10^{19}\, \kappa_\oplus, \tag{3.17}$$

where $\kappa_\oplus = 9.81\ \mathrm{m\,s^{-2}}$ is the surface gravity on Earth. Even neutron stars, thought to be one of the most extreme environments in our Universe, have surface gravities of "only" $\sim 10^{11}\,\kappa_\oplus$, thus we learn that atomic hydrogen and other simple atoms are effectively blind to background scalar-field gradients.

3.3 Hyperfine splitting and the 21 cm line

In addition to an electric charge, atomic nuclei also have magnetic dipole moments, whose interactions with a chameleon-like scalar are the subject of this section. For a nucleus with spin operator $\mathbf{I}$, this dipole moment sources a magnetic field whose corresponding vector potential in the Lorenz gauge is

$$\mathbf{A}_i = \frac{g_N e}{2m_p} \frac{\epsilon_{ijk} \mathbf{I}^j \mathbf{x}^k}{4\pi r^3}, \tag{3.18}$$

where g_N is a dimensionless constant often called the gyromagnetic ratio. (The proton has $g_N \approx 5.59$).

The interaction between this field and the electron's angular momentum leads to hyperfine splitting in the atomic spectral lines, as governed by the Hamiltonian

$$H_{\rm hfs} = \frac{g_N \alpha}{2m_p} \frac{\epsilon_{ijk} \Sigma^i \mathbf{I}^j \mathbf{x}^k}{r^3}, \tag{3.19}$$

which one treats as a small perturbation to H_0. An eigenstate $|n, \ell, j, F\rangle$ of the Hamiltonian $H_0 + H_{\rm hfs}$ is specified by four quantum numbers: n, ℓ, and j denote the principal quantum number, orbital angular momentum, and total angular momentum of the electron, respectively, while F is the eigenvalue corresponding to the total angular momentum of the system: $\langle \mathbf{F}^2 \rangle = F(F+1)$ with $\mathbf{F} = \mathbf{I} + \mathbf{j}$. It can then be shown that the expectation value of this hyperfine-structure Hamiltonian is [411, 412]

$$\langle H_{\rm hfs} \rangle = -\frac{g_N \alpha}{m_p} \langle i \Sigma_r \gamma^0 r^{-2} \rangle \frac{k}{4k^2 - 1} [F(F+1) - I(I+1) - j(j+1)], \tag{3.20}$$

where $k = (-1)^{j-\ell+1/2}(j + 1/2)$ is the angular-momentum-parity quantum number and $\Sigma_r := \hat{\mathbf{x}}_i \Sigma^i$ is the projection of Σ^i along the radial direction.

In the present context, it is interesting to ask what additional effects arise when the interactions between a chameleon-like scalar and this magnetic field are taken into account. As in Section 3.1, two distinct effects can be identified:

(1) *Fifth force*: The nuclear magnetic field sources an additional contribution to the scalar perturbation φ, which exerts a fifth force on the electron and leads to further shifts in its energy levels.

(2) *Vacuum polarisation*: The magnetic field also interacts with φ to generate secondary currents that go on to a source a perturbation $\delta\mathbf{A}$. If the atom's worldline is accelerating and/or if there is a background gradient $\nabla\phi_0$ present, these will lead to additional contributions to $\delta\mathbf{A}$ as well.

Simple order-of-magnitude estimates reveal that the second interaction is likely to be more pronounced; hence, let us concentrate on this latter effect in what follows. For simplicity, we will also neglect the subleading contributions of the bare electromagnetic fields to φ and will focus only on the leading term due to the nuclear mass m_N. Substituting (3.13) and the first term of (3.3) into (3.4), the perturbation to the vector potential is found to be

$$\delta\mathbf{A}_i = \frac{g_N}{2m_p}\frac{e}{4\pi}\bigg(\beta_m\beta_\gamma\frac{Gm_N}{2}\frac{\epsilon_{ijk}\mathbf{I}^j\mathbf{x}^k}{r^4} - \delta_{\ell(m}\epsilon_{k)ij}\frac{\beta_m}{M_{\rm Pl}}\frac{\mathbf{I}^j\mathbf{x}^k\mathbf{x}^\ell(\nabla\phi_0)^m}{r^3} - \delta_{\ell[m}\epsilon_{i]jk}\frac{\beta_\gamma}{M_{\rm Pl}}\frac{\mathbf{I}^j\mathbf{x}^k\mathbf{x}^\ell(\nabla\phi_0)^m}{r^3}\bigg). \tag{3.21}$$

The way this is written might suggest that the second and third terms above are enhanced by one power of $M_{\rm Pl}$ relative to the first, but it is worth noting that there is no such enhancement, as the background field ϕ_0 is itself proportional to $\beta_m/M_{\rm Pl}$ (or $\beta_\gamma/M_{\rm Pl}$) if it is to be a valid solution to the equation of motion in (2.30). All three terms in (3.21) therefore appear at the same order in G ($\equiv 1/8\pi M_{\rm Pl}^2$).

Now contracting this equation with $e\Sigma^i$ returns its contribution to the Hamiltonian H_S. To start with, notice that the angular operator in the first line of (3.21) is identical to that of the hyperfine-structure Hamiltonian in (3.19), thus the expectation value of this first term is

$$\langle H_S\rangle \supset -\frac{g_N\alpha}{2m_p}G\beta_m\beta_\gamma m_N\langle i\Sigma_r\gamma^0 r^{-3}\rangle\frac{k}{4k^2-1}[F(F+1)-I(I+1)-j(j+1)]. \tag{3.22}$$

The hyperfine transition of primary interest is the hydrogen 21 cm line, which involves the $1S_{1/2}^{F=0}$ and $1S_{1/2}^{F=1}$ states. The radial expectation values $\langle i\Sigma_r\gamma^0 r^{-3}\rangle$ for both of these states diverge just like $\langle r^{-3}\rangle$ did in Section 3.1, and so we regularise by imposing a cutoff at the proton charge radius r_p. Having done so, numerical integration shows that this term in H_S alters the magnitude of the 21 cm transition frequency by the relative amount $\sim 2\times 10^{-43}\,\beta_m\beta_\gamma$.

As for the remaining two terms in (3.21), their contribution to the Hamiltonian is

$$H_S \supset -\frac{g_N\alpha}{2m_pM_{\rm Pl}}\bigg(\beta_m\delta_{\ell(m}\epsilon_{k)ij}+\beta_\gamma\delta_{\ell[m}\epsilon_{i]jk}\bigg)\frac{\Sigma^i\mathbf{I}^j\mathbf{x}^k\mathbf{x}^\ell(\nabla\phi_0)^m}{r^3}, \tag{3.23}$$

but these have no effect on the 21 cm line. To see this, note that after fixing the direction of $\nabla\phi_0$ (say, along the z axis), the remaining operators $\Sigma^i \mathbf{x}^k \mathbf{x}^\ell$ that act on the electron's wavefunction have odd parity overall. In general, we expect such operators to lift degeneracies between states, but since the states involved in the 21 cm transition all have identical electron quantum numbers, the relevant matrix elements for this part of the Hamiltonian all vanish. Thus, as was the case with fine structure in the previous section, we see here that the hyperfine structure of atomic hydrogen is also insensitive to the presence of background scalar-field gradients (at first order in G).

3.4 Measuring the fine-structure constant

Let us now discuss the possibility of using these effects to impose constraints on chameleon-like theories. At first order in perturbation theory, we have already seen that background scalar-field gradients have essentially no influence on the energy levels of atomic hydrogen. As for the 21 cm line, laboratory measurements of this transition frequency have achieved a remarkable precision of less than one part per trillion [410], but larger theoretical uncertainties associated with modelling the proton's finite-size effects make this line ill-suited for probing or constraining new physics. Accordingly, it stands to reason that the gross and fine-structure corrections computed in Section 3.1 are our best option.

In principle, using these shifts in the energy levels to detect or constrain a chameleon-like scalar is straightforward. Suppose that a given transition, say the Lyman-alpha line, is measured to have an angular frequency $2\pi\nu(1S_{1/2} - 2S_{1/2})$. We should then compare this value with the theoretical prediction obtained by evaluating $\langle H\rangle(2S_{1/2}) - \langle H\rangle(1S_{1/2})$. If both numbers agree to within experimental and theoretical uncertainties when $\beta_m = \beta_\gamma = 0$, no scalar has been detected. This agreement can then be translated into an upper bound on the coupling parameters, since ones that are too large would lead to an intolerable discrepancy.

There is, however, an important subtlety that we must contend with. A value for the fine-structure constant α is needed to make the theoretical prediction, but the recommended value by CODATA (the Committee on Data for Science and Technology) cannot be used for this purpose because it is a weighted average of several measurements [402], which includes values of α inferred from measurements of atomic hydrogen. To elaborate, these are obtained by equating the experimentally measured values of transition frequencies to their corresponding theoretical predictions from QED and solving for α. By construction, these inferred values already assume no potential contribution from new physics. To avoid this pitfall, we must use an independently measured value of α, but we should expect that whatever independent experiment we choose will also be subject to systematic corrections from a chameleon-like

scalar, which should be taken into account. The question now stands: How do we constrain these scalar field theories in a consistent manner?

The approach we will take is as follows. Consider, in the simplest case, two independent experiments that can be used to measure the fine-structure constant, and call them X_1 and X_2. Assuming no new physics beyond QED, these experiments will produce the inferred values $\overline{\alpha}(X_1)$ and $\overline{\alpha}(X_2)$, respectively. However, if a chameleon-like scalar is actually active in the experiments, then these values are biased. The true value of the fine-structure constant is obtained by subtracting the systematic bias introduced from not taking the scalar into account: $\alpha = \overline{\alpha}(X_i) - \delta\alpha(X_i)$ $(i = 1, 2)$. Because these shifts $\delta\alpha(X_i)$ will typically be different in size for each experiment, a significant discrepancy between $\overline{\alpha}(X_1)$ and $\overline{\alpha}(X_2)$ would be a tantalising hint of new physics. Conversely, if these two values are found to agree with one another within combined error bars of size Δ, then any effect from the scalar is currently too small to probe; hence, a constraint can be obtained by requiring that $|\delta\alpha(X_1) - \delta\alpha(X_2)| < \Delta$.

In this section, we will determine the systematic shift $\delta\alpha(\mathrm{H})$ that is needed to correct the value $\overline{\alpha}(\mathrm{H})$ inferred from spectral lines of atomic hydrogen. In later chapters, this result will be combined with the appropriate shift $\delta\alpha$ from another experiment to impose constraints on chameleon-like theories.

To make full use of the fact that measurements of hydrogen lines have relative uncertainties that are less than $\sim 10^{-10}$, we must be able to perform theoretical calculations at a similar level of precision. In addition to H_0 and H_S, two other sets of terms are needed to obtain a precise value for $\overline{\alpha}(\mathrm{H})$. In descending order of magnitude (and neglecting hyperfine structure), the full Hamiltonian for atomic hydrogen is given by

$$H = H_0 + H_R + H_L + H_S. \tag{3.24}$$

The recoil Hamiltonian H_R is a perturbative series truncated at $\mathcal{O}(\alpha^4)$ that accounts for the motion of the proton about the atom's centre of mass, while the higher-order terms from this series, as well as radiative corrections and finite-size effects of the nucleus, are encoded in the Lamb shift H_L. (See Refs. [402, 407–410] for further details.)

In practice, when extracting a value for the fine-structure constant, linear combinations of transition frequencies are used to circumvent the large theoretical uncertainties in H_L associated with modelling nuclear finite-size effects, which strongly affect the $S_{1/2}$ states [407]. By equating this linear combination of measured frequencies to our theoretical prediction,

$$N 2\pi\nu(\psi_1 - \psi_1') - 2\pi\nu(\psi_2 - \psi_2') = N\left[\langle H\rangle(\psi_1') - \langle H\rangle(\psi_1)\right] - \left[\langle H\rangle(\psi_2') - \langle H\rangle(\psi_2)\right], \tag{3.25}$$

we obtain an implicit equation for α in terms of the integer N, which we are free to specify, and five

parameters: the electron mass m_e, the proton mass m_p, Newton's gravitational constant G, and the scalar couplings β_m and β_γ. Of course, the last three of these parameters appear only as part of the scalar's contribution H_S to the Hamiltonian.

For simplicity, let us represent this implicit equation as $f(\alpha; \Pi, G) = 0$, where Π is shorthand for the other parameters. In the limit $G \to 0$, which is a proxy for turning off the contribution from H_S, the solution to this equation gives us the inferred value $\overline{\alpha}$ under the assumption that there is no new physics beyond QED. The value we will use in this thesis is [413]

$$[\overline{\alpha}(\mathrm{H})]^{-1} = 137.035\,998\,995\,(85). \tag{3.26}$$

If a chameleon-like scalar exists, the true value of the fine-structure constant is actually $\alpha = \overline{\alpha} - \delta\alpha$, where the systematic shift can be obtained from the requirement that $f(\overline{\alpha} - \delta\alpha; \Pi, G) = 0$. To first order in G, the solution is

$$\delta\alpha = \frac{G[\partial f(\overline{\alpha}; \Pi, G)/\partial G]_{G=0}}{[\partial f(\alpha; \Pi, 0)/\partial\alpha]_{\alpha=\overline{\alpha}}}. \tag{3.27}$$

The numerator depends only on H_S, whereas the denominator is a power series in $\overline{\alpha}$ that receives contributions from H_0, H_R, and H_L. As a first approximation, however, it will suffice to truncate this series to leading order in $\overline{\alpha}$, in which case the expectation value $\langle H_0 - m_e \rangle$ reduces to the Rydberg formula $-m_e\overline{\alpha}^2/(2n^2)$, while the effects of H_R and H_L can be neglected entirely. In this limit, the systematic shift due to the scalar is

$$\delta\alpha(\mathrm{H}) = 2G\beta_m^2 m_e m_p + \mathcal{C}_2 G\beta_m\beta_\gamma m_e m_p\overline{\alpha}^2 + \frac{1}{6}\mathcal{C}_3 G\beta_\gamma^2 m_e^2\overline{\alpha}^4. \tag{3.28}$$

Notice that the first term is independent of the choice of spectral lines, whereas the remaining two are not. Their dependence on the quantum numbers of the relevant states are encapsulated in two dimensionless coefficients, which are both given by

$$\mathcal{C}_q = \frac{N[\langle r^{-q}\rangle(\psi_1) - \langle r^{-q}\rangle(\psi_1')] - [\langle r^{-q}\rangle(\psi_2) - \langle r^{-q}\rangle(\psi_2')]}{(m_e\overline{\alpha})^q[N(n_1n_1')^{-2}(n_1'^2 - n_1^2) - (n_2n_2')^{-2}(n_2'^2 - n_2^2)]}. \tag{3.29}$$

For $\mathcal{C}_2$, a subleading effect suppressed by one power of m_e/m_p, which originates from the second term in (3.9), has been neglected.

Given that the value for $\overline{\alpha}(\mathrm{H})$ in (3.26) is a weighted average derived from multiple linear combinations of spectral lines, we should also use an average value for the coefficients $\mathcal{C}_2$ and $\mathcal{C}_3$. Their numerical values for different transitions are presented in Table 3.1, and recall that $\langle r^{-3}\rangle$ is regularised according to (3.12) with $r_{\mathrm{cut}} = r_p$ whenever necessary. As measurements of the different transition frequencies have approximately the same level of uncertainty [402, 407], it will suffice for our purposes

Table 3.1 Numerical values for the coefficients $\mathcal{C}_2$ and $\mathcal{C}_3$ for a number of hydrogen lines. At this level of precision, there is no need for a distinction between a $P_{1/2}$ and $P_{3/2}$ state, or between a $D_{3/2}$ and $D_{5/2}$ state.

Transition frequencies	$\mathcal{C}_2$	$\mathcal{C}_3$	Transition frequencies	$\mathcal{C}_2$	$\mathcal{C}_3$
$\nu(2S-8S)-\nu(2S-2P)$	0.34	−0.14	$7\nu(2S-8S)-\nu(1S-2S)$	−0.031	−0.85
$\nu(2S-8D)-\nu(2S-2P)$	0.35	0.18	$7\nu(2S-8D)-\nu(1S-2S)$	−0.0061	−0.25
$\nu(2S-12D)-\nu(2S-2P)$	0.34	0.17	$7\nu(2S-12D)-\nu(1S-2S)$	−0.0016	−0.24

to calculate the unweighted mean. The average value of $\delta\alpha(\mathrm{H})$ is then

$$\delta\alpha(\mathrm{H}) = G m_e m_p \left[2\beta_m^2 + 0.17\,\beta_m\beta_\gamma\bar{\alpha}^2 - 0.032\,(m_e/m_p)\beta_\gamma^2\,\bar{\alpha}^4\right]. \tag{3.30}$$

Note that our choice of averaging procedure, and also our regularisation prescription, will only affect the bounds on β_γ; the bound on β_m is robust. These constraints will be presented in Chapter 5.

As a last remark, it is worth acknowledging that the analysis above also requires experimentally measured values for the electron and proton rest masses. The most accurate determination of the former derives from combining measurements of its relative atomic mass $A_r(e)$ [402], the relative atomic mass of rubidium $A_r(\mathrm{Rb})$ [414, 415], and the quotient h/m_{Rb} (where h is the Planck constant) measured from recoil experiments [416–418].[4] The mass of the proton can then be deduced after also measuring the proton–electron mass ratio [402]. In principle, the experiments used to obtain these quantities could also be subject to systematic corrections from a chameleon-like scalar, although a detailed analysis of each one is beyond the scope of this thesis. As we will see in Chapter 5, the constraints in a given region of parameter space are typically driven by just one experiment at a time, thus it stands to reason that neglecting possible systematics from these ancillary experiments should still give us a conservative bound.

Appendix 3.A Relativistic hydrogen wavefunctions

The Dirac–Coulomb Hamiltonian H_0 describes an electron bound to the electrostatic potential sourced by a point charge $+Ze$ that is fixed at the origin,

$$H_0 = \gamma^0 m_e + \Sigma^i \mathbf{p}_i - \frac{Z\alpha}{r}. \tag{3.A.1}$$

The solutions to the eigenvalue problem $H_0\psi = E\psi$ are reviewed in this appendix.

[4] The uncertainty in measuring h/m_{Rb} dominates the standard error for $\bar{\alpha}(\mathrm{H})$ in (3.26).

Eigenstates The problem is solved by working in the Dirac representation, in which

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{and} \quad \Sigma^i = \begin{pmatrix} 0 & \sigma^i \\ \sigma^i & 0 \end{pmatrix}, \tag{3.A.2}$$

with σ^i denoting the usual Pauli matrices. The equations are then separable in spherical coordinates, thus we can make the ansatz

$$\psi = \begin{pmatrix} R_1(r) Y_{k\ell m}(\hat{\mathbf{x}}) \\ i R_2(r) Y_{-k,\ell m}(\hat{\mathbf{x}}) \end{pmatrix}, \tag{3.A.3}$$

where $R_1(r)$ and $R_2(r)$ are two radial functions to be determined, and $Y_{k\ell m}(\hat{\mathbf{x}})$ are the spinor spherical harmonics. These are related to the usual (scalar) spherical harmonics $Y_{\ell m}(\hat{\mathbf{x}})$ according to the appropriate set of Clebsch–Gordan coefficients [412, 419],

$$\begin{aligned} Y_{k\ell m}(\hat{\mathbf{x}}) = & \sqrt{\frac{\ell - m\,\mathrm{sgn}(k) + 1/2}{2\ell + 1}}\, Y_{\ell, m-1/2}(\hat{\mathbf{x}}) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ & - \mathrm{sgn}(k) \sqrt{\frac{\ell + m\,\mathrm{sgn}(k) + 1/2}{2\ell + 1}}\, Y_{\ell, m+1/2}(\hat{\mathbf{x}}) \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \end{aligned} \tag{3.A.4}$$

and are specified by three quantum numbers (j, ℓ, m): The integer $\ell = 0, 1, 2, \ldots$ determines the electron's orbital angular momentum, the half-integer $j = \ell \pm 1/2$ is its total angular momentum, and the magnetic quantum number m is the projection of j along the z axis. From j and ℓ, it is convenient to define the angular-momentum-parity quantum number $k = (-1)^{j-\ell+1/2}(j + 1/2)$.

The eigenvalues for this system are

$$E_{nj} = m_e \left(1 + \frac{Z^2\alpha^2}{\left(n - j - 1/2 + \sqrt{(j+1/2)^2 - Z^2\alpha^2} \right)^2} \right)^{-1/2}, \tag{3.A.5}$$

where the integer $n = 1, 2, \ldots$ is the principal quantum number. Finally, it will be convenient to define

$$\mathcal{E} = E_{nj}/m_e, \quad \nu = \sqrt{k^2 - Z^2\alpha^2}, \quad b = \sqrt{1 - \mathcal{E}^2}, \quad \text{and} \quad n_r = n - (j + 1/2), \tag{3.A.6}$$

as these allow us to express the radial eigenfunctions compactly as [400]

$$\begin{pmatrix} R_1(r) \\ R_2(r) \end{pmatrix} = \frac{b^2 m_e^{3/2}}{\nu} \sqrt{\frac{(\mathcal{E}k - \nu) n_r!}{Z\alpha(k - \nu)\Gamma(n_r + 2\nu)}}\, \rho^{\nu-1} e^{-\rho/2} \begin{pmatrix} \mathfrak{R}_1^+ & \mathfrak{R}_1^- \\ \mathfrak{R}_2^+ & \mathfrak{R}_2^- \end{pmatrix} \begin{pmatrix} \rho L_{n_r-1}^{2\nu+1}(\rho) \\ L_{n_r}^{2\nu-1}(\rho) \end{pmatrix} \tag{3.A.7}$$

in terms of the associated Laguerre polynomials [397], having also defined $\rho = 2bm_e r$ and

$$\mathfrak{R}_1^+ = \frac{Z\alpha b}{\mathcal{E}k - \nu}, \qquad \mathfrak{R}_1^- = k - \nu, \qquad \mathfrak{R}_2^+ = \frac{b(k-\nu)}{\mathcal{E}k - \nu}, \qquad \mathfrak{R}_2^- = Z\alpha. \tag{3.A.8}$$

Radial expectation values Three types of integrals are commonly needed when evaluating expectation values of spherically symmetric operators. The options are [420]

$$\langle r^q \rangle = \int_0^\infty \mathrm{d}r\, r^{q+2}[R_1^2(r) + R_2^2(r)], \tag{3.A.9a}$$

$$\langle \gamma^0 r^q \rangle = \int_0^\infty \mathrm{d}r\, r^{q+2}[R_1^2(r) - R_2^2(r)], \tag{3.A.9b}$$

$$\langle i\,\Sigma_r \gamma^0 r^q \rangle = -2\int_0^\infty \mathrm{d}r\, r^{q+2} R_1(r) R_2(r), \tag{3.A.9c}$$

where $\Sigma_r := \hat{\mathbf{x}}_i\,\Sigma^i$ is the projection of Σ^i along the radial direction. For $q \in \{-3, -2, \ldots, 2\}$, analytic expressions can be found in Ref. [400]. Those that are of use in the main text are reproduced below:

$$\langle \gamma^0 r^{-1} \rangle = \frac{m_e b^2}{Z\alpha}, \tag{3.A.10}$$

$$\langle \gamma^0 r^{-2} \rangle = \frac{2b^3 m_e^2 (2\nu^2 - \mathcal{E}k)}{Z\alpha\nu(4\nu^2 - 1)}, \tag{3.A.11}$$

$$\langle r^{-2} \rangle = \frac{2b^3 m_e^2 k(2\mathcal{E}k - 1)}{Z\alpha\nu(4\nu^2 - 1)}, \tag{3.A.12}$$

$$\langle r^{-3} \rangle = 2(bm_e)^3 \frac{3\mathcal{E}^2k^2 - 3\mathcal{E}k - \nu^2 + 1}{\nu(\nu^2 - 1)(4\nu^2 - 1)}. \tag{3.A.13}$$

It is vital to keep in mind that these expressions are not always valid. As both radial functions can be represented as series solutions [420],

$$R_1(r) = r^{\nu-1}\sum_{i=0}^{\infty} R_1^{(i)} r^i, \qquad R_2(r) = r^{\nu-1}\sum_{i=0}^{\infty} R_2^{(i)} r^i, \tag{3.A.14}$$

substituting these into (3.A.9) reveals that the integrals diverge whenever $2\nu + q \leq -1$, due to the singular behaviour of the integrand near $r = 0$. Equivalently, since E_{nj} is real only when ν is real, they diverge when $q \leq -2(j+1)$. Thus, as I pointed out in the main text, expectation values for $S_{1/2}$ and $P_{1/2}$ states diverge when $q \leq -3$.

Chapter 4
The electron's magnetic moment

Away from the limelight of multibillion-dollar projects like the Large Hadron Collider sits a small, tabletop experiment responsible for delivering one of the most precise measurements in scientific history. In 2008, Hanneke and collaborators published the results of their latest experiment, which trapped single electrons in a vacuum chamber for months at a time to measure the particle's magnetic dipole moment g_e to within an uncertainty of less than one part per trillion [421, 422].

In this chapter, we will reanalyse this incredible experiment under the assumption that a chameleon-like scalar is active in the vacuum cavity. Two types of effects arise:

(1) *Quantum corrections*: The scalar generates additional loop corrections to the QED vertex function, which increase the intrinsic value of g_e. Due to its screening mechanism, changes in the scalar's effective mass and/or its couplings to the Standard Model cause the magnitudes of these corrections to vary dynamically as a function of the local environment.

(2) *Cavity shift*: Nonlinear self-interactions drive the scalar to form a bubble profile within the vacuum chamber of the experiment, which results in a fifth force being exerted on the electron. Because the electron's magnetic moment is determined indirectly from measurements of its eigenfrequencies in the cavity, perturbations to its energy levels brought about by this fifth force introduce a systematic bias into the measured value of g_e unless accounted for.

These effects are described in further detail in Sections 4.1 and 4.2, respectively. Then, in Section 4.3, we will discuss how a measurement of g_e can be translated into an experimental determination of the fine-structure constant α; focusing on the systematic corrections needed to account for the presence of a chameleon-like scalar.

4.1 Quantum corrections

Let us begin with some definitions. The magnetic dipole moment of the electron,

$$\boldsymbol{\mu} = -\mathrm{g}_e \mu_B \mathbf{S}, \tag{4.1}$$

is defined in terms of its spin operator $\mathbf{S}$, the Bohr magneton $\mu_B \equiv e/(2m_e)$, and the gyromagnetic factor g_e. This last quantity is also often simply called the magnetic moment, since it alone determines the size of $\boldsymbol{\mu}$; the magnitudes of the other two quantities are fixed by definition.

The Dirac equation predicts that an electron should have $\mathrm{g}_e = 2$ exactly, but quantum fluctuations lead to a slight increase in this value. The difference between the true and tree-level values is called the anomalous magnetic moment,

$$a_e := (\mathrm{g}_e - 2)/2. \tag{4.2}$$

This quantity can be read off from the renormalised QED vertex function, which — for any Lorentz invariant, U(1)-gauge invariant, and *CP*-symmetric theory — has the form [423]

$$\Gamma^\mu(p, p') = F_1(q^2)\gamma^\mu + F_2(q^2)\left(\frac{-iS^{\mu\nu}q_\nu}{m_e}\right) \tag{4.3}$$

when the external electrons are on shell. (Recall we defined $S^{\mu\nu} = (i/4)[\gamma^\mu, \gamma^\nu]$ in Chapter 2.) The 4-vectors p and p' denote the momenta for the ingoing and outgoing electron, respectively, the photon has ingoing momentum $q = p' - p$, and the functions F_1 and F_2 are known as the electric and magnetic form factors. As the constant electric part $F_1(0)$ leads to a simple rescaling of the Lagrangian parameter e, we are always free to set $F_1(0) = 1$ exactly by working in an on-shell renormalisation scheme, in which case $-e$ is exactly the electron charge. Moreover, as the vertex function Γ^μ has been normalised such that it reduces to the Dirac matrix γ^μ at tree level, its constant magnetic part is exactly the anomalous magnetic moment, $F_2(0) = a_e$.

In one of the early crowning achievements of quantum field theory, Schwinger famously showed in

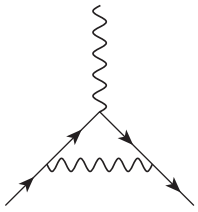


Figure 4.1 The one-loop Feynman diagram that provides the leading contribution to the electron's anomalous magnetic moment. The electron is represented by a solid line while the photon is drawn as a wavy line.

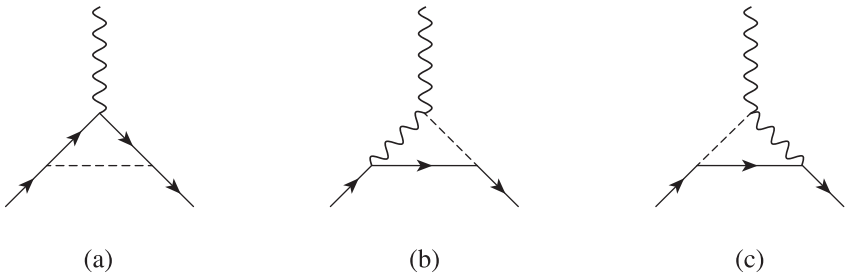


Figure 4.2 One-loop corrections to the electron's magnetic moment from a chameleon-like scalar (dashed line).

1948 that the electron's anomalous magnetic moment $a_e = \alpha/(2\pi)$ at leading, one-loop order in pure QED [424, 425]. This came at a time before the use of Feynman diagrams was prevalent, but in modern parlance, Schwinger's result can be shown to arise from the single Feynman diagram in Figure 4.1. Figure 4.2, meanwhile, shows that three additional Feynman diagrams would also contribute to a_e at the one-loop level if a chameleon-like scalar were present in this Universe.

To compute these extra corrections, consider small quantum fluctuations $\delta\phi = \phi - \langle\phi\rangle$ about the classical field profile $\langle\phi\rangle$ in the vacuum cavity where g_e is to be measured. As the electron remains very close to the centre of this cavity (see Section 4.2.4), it will suffice to take $\langle\phi\rangle \approx \phi_0$, where the constant ϕ_0 is the value of the classical field profile at this position. The local effective mass m_0 for $\delta\phi$ is then given by evaluating $m_0^2 := m_{\rm eff}^2(\phi_0) \equiv V''_{\rm eff}(\phi_0)$ at the centre of this cavity. As for the scalar's interactions with the Standard Model, the two terms that are relevant at one loop are

$$\mathcal{L} \supset -\left(\frac{\beta_m}{M_{\rm Pl}}\right) m_e\, \bar{\psi}\psi\delta\phi - \frac{1}{4}\left(\frac{\beta_\gamma}{M_{\rm Pl}}\right)\delta\phi\, F_{\mu\nu}F^{\mu\nu}. \tag{4.4}$$

These come from linearising the interactions in (2.25) and (2.19), respectively. The dimensionless coupling strengths $\beta_m \equiv \beta_m(\phi_0)$ and $\beta_\gamma \equiv \beta_\gamma(\phi_0)$ are also to be evaluated at the cavity's centre, although in what follows, we will suppress this explicit dependence on ϕ_0 for the sake of readability.

Given these interactions, the first diagram in Figure 4.2(a) can be shown to lead to the contribution

$$a_e \supset 2\beta_m^2\left(\frac{m_e}{4\pi M_{\rm Pl}}\right)^2 I_1(m_0/m_e), \tag{4.5}$$

where $I_1(z)$ is an integral over Feynman parameters. A closed-form expression is derived towards the end of Appendix 4.A, although it will often suffice to replace this integral by the constant value $I_1(0) = 3/2$ since $m_0 \ll m_e$ in most of parameter space. The remaining two diagrams in Figure 4.2, meanwhile, result in UV-divergent contributions to the magnetic form factor. After regularising by

working in $4-\epsilon$ dimensions, they lead to the contribution

$$a_e \supset 4\beta_m\beta_\gamma\left(\frac{m_e}{4\pi M_{\rm Pl}}\right)^2\left[\frac{1}{\epsilon}+\log\left(\frac{\widetilde{\mu}}{m_e}\right)+I_2(m_0/m_e)\right], \tag{4.6}$$

where $\widetilde{\mu}$ is the renormalisation point (or "sliding scale") and $I_2(z)$ is another integral over Feynman parameters, also normalised such that $I_2(0)=3/2$. It must be stated that the diagrams in Figure 4.2 commonly arise in many different contexts (see, e.g., Refs. [426–430]), but for the sake of completeness, a derivation of (4.5) and (4.6) is presented in Appendix 4.A.

Two comments about these results are necessary. First, a comprehensive survey of all of the one-loop diagrams in this theory will reveal that the scalar's potential $V(\phi)$ receives large quantum corrections that, among other things, force the scalar to become much heavier than what the classical theory would predict [431]. This is the usual hierarchy problem. While several attempts have been made at constructing more radiatively stable models [432,433], the issue remains mostly open. For now, we will have to simply tolerate a certain level of fine tuning in order to explore the interesting, albeit technically unnatural phenomenology of screening.

On to the second comment: Because the interaction between the scalar and the photon is due to a nonrenormalisable operator, we naturally find that the UV divergence in (4.6) cannot be absorbed into any of the existing parameters we have defined. This difficulty is unsurprising and is resolved by recognising that the second term in (4.4) is not the only dimension-five operator in the Lagrangian. Whatever the UV completion is, the interaction term

$$\mathcal{L}\supset -c_5\,\overline{\psi}\,S^{\mu\nu}F_{\mu\nu}\,\psi \tag{4.7}$$

must also be present in the low-energy EFT, where c_5 is a Wilsonian coefficient with dimensions of length. One therefore removes the UV divergence in (4.6) by absorbing it into a renormalisation of c_5. Without loss of generality, let us define a dimensionless coefficient a_5 such that $c_5=\mu_B a_5$ to simplify the final result. In the on-shell scheme, the total (theoretical) value for the anomalous magnetic moment of the electron is then

$$a_e^{\rm th}=a_e^{\rm SM}+a_5+\left(\frac{Gm_e^2}{\pi}\right)\left[\beta_m^2 I_1(m_0/m_e)+2\beta_m\beta_\gamma I_2(m_0/m_e)\right], \tag{4.8}$$

where the first term is the contribution from the Standard Model and I include a superscript "th" to distinguish this theoretical prediction from the experimentally measured value in the next section.

As the value of a_5 is a priori unknown, this quantity should be regarded as an additional parameter in the theory to be constrained. That being said, we will suppose hereafter that the UV completion for this

theory is such that a_5 is always much smaller than the other terms in (4.8); thus allowing us to restrict our attention to the region of parameter space where $a_5 \approx 0$. To be clear, there is no particularly strong theoretical motivation to do this — it is purely a simplifying assumption that helps reduce the number of dimensions in parameter space down to something more manageable.

4.2 Cavity shift

Quantum corrections aside, a chameleon-like scalar also introduces systematic effects into the measured value of g_e. In the current state of the art [421, 422], this quantity is determined experimentally by using strong electromagnetic fields to confine a single electron to the centre of a vacuum cavity, called a Penning trap, into which microwave photons are injected. By varying the energy of these photons and looking for resonant transitions, one is able to accurately measure the electron's eigenfrequencies in this trap, which encode information about its magnetic moment.

As we learnt at the end of Chapter 2, a chameleon-like scalar is predisposed to forming a bubble-like profile inside such a cavity. If one is present, then the fifth force it exerts on the electron will affect the values of its eigenfrequencies, and so biases the measurement of g_e unless taken into account. In this section, we will calculate the systematic shift generated by this vacuum bubble.

4.2.1 Cylindrical Penning traps

In Hanneke et al.'s experiment, which they dub the one-electron quantum cyclotron, a single electron is held at the centre of a cylindrical Penning trap by the quadratic electrostatic potential [422]

$$V = \frac{V_0}{2d^2}\left(\frac{r^2}{2} - z^2\right); \tag{4.9}$$

having chosen cylindrical coordinates (r, ϑ, z) whose origin coincides with the cavity's centre. The potential difference is maintained at a value of $V_0 = 101.4$ V and the characteristic length scale

$$d = \left(r_0^2/4 + z_0^2/2\right)^{1/2} \approx 3.5 \text{ mm} \tag{4.10}$$

is defined in terms of the cavity's radius r_0 and half-height z_0. The values of all experimental parameters pertinent to our discussion are curated in Table 4.1. Alongside this electrostatic potential, the uniform, axial magnetic field

$$\mathbf{B} = B_0 \mathbf{e}_z \tag{4.11}$$

is also established within the cavity to split the degeneracy of the electron's spin states.

Table 4.1 Experimental parameters and eigenfrequencies in the cylindrical Penning trap, reproduced from Hanneke et al. [421, 422]. Up to small differences, the theoretical frequencies $(\omega_+, \omega_0, \omega_z)$ are approximately related to their experimentally measured counterparts by $\omega_+ \approx \omega_0 \approx 2\pi\nu_c$ and $\omega_z \approx 2\pi\nu_z$.

Parameter		Value	Parameter		Value
Magnetic field	B_0	5.36 T	Cyclotron frequency	ν_c	150 GHz
Potential difference	V_0	101.4 V	Anomaly frequency	ν_a	174 MHz
Cavity radius	r_0	4.5 mm	Axial frequency	ν_z	200 MHz
Cavity height	$2z_0$	7.7 mm	Magnetron frequency	$\omega_-/(2\pi)$	133 kHz

These electromagnetic fields act as a source for a chameleon-like scalar, whose profile inside the Penning trap is determined by solving the equation of motion [cf. (2.30)]

$$\nabla^2\phi = V'_{\text{eff}}(\phi) = \begin{cases} V'(\phi) + \dfrac{\beta_m(\phi)\rho_0}{M_{\text{Pl}}} + \dfrac{\beta_m(\phi)\rho_{\text{em}}}{M_{\text{Pl}}} & \text{(in the cavity: } r < r_0, |z| < z_0\text{)} \\ V'(\phi) + \dfrac{\beta_m(\phi)\rho_\infty}{M_{\text{Pl}}} & \text{(in the surrounding walls).} \end{cases} \tag{4.12}$$

In addition to the electromagnetic energy density $\rho_{\text{em}} = (\mathbf{B}^2 - \mathbf{E}^2)/2$, the rhs of this equation also requires as input estimates for the densities of matter in the interior and in the surrounding walls. These are denoted by ρ_0 and ρ_∞, respectively. While no direct measurement of the density of gas in the cavity has been made, an estimate from a similar trap design places an upper bound on the number density of atoms at 100 cm^{-3} [422, 434]. Assuming the current experiment satisfies the same bound and taking the average mass of a molecule to be that of nitrogen, we estimate $\rho_0 \lesssim 5 \times 10^{-18}\ \text{kg}\,\text{m}^{-3}$. On the other hand, the trap electrodes and vacuum container surrounding the cavity are composed primarily of silver, quartz, titanium, and molybdenum [422], which have typical densities $\rho_\infty \gtrsim 3 \times 10^3\ \text{kg}\,\text{m}^{-3}$.

Because the cavity's radius and height are comparable in size, the solution to (4.12) must strictly be a function of two variables, $\phi \equiv \phi(r, z)$. This makes the problem too difficult to solve analytically, even with approximate methods, thus numerical techniques are needed to obtain an accurate solution. Having said that, it turns out that we can continue to make a substantial amount of analytic progress for the time being as the experiment is cooled to an extremely low temperature $T \sim 100$ mK. The electron remains very close to the centre of the cavity as a result (a more precise statement will be made in Section 4.2.4); hence, whatever the scalar-field profile is, the motion of the electron will only be sensitive to the first few terms in its Taylor expansion about the origin:

$$\phi \simeq \phi_0 + \phi_{rr}\frac{r^2}{2r_0^2} + \phi_{zz}\frac{z^2}{2z_0^2}. \tag{4.13}$$

Reflection symmetry in all three spatial directions guarantees that this expansion contains only even powers of r and z. The Taylor coefficients ϕ_0, ϕ_{rr}, and ϕ_{zz} are left to be determined numerically, and we will do so in Section 4.3, but for now, it will suffice to work with the generic field profile in (4.13).

4.2.2 Hamiltonian

The experimental parameters are such that the electron at the centre of this cavity is nonrelativistic. In this limit, its Hamiltonian — whose general form is given by (3.6) — reduces to the familiar one appearing in the Schrödinger equation. We shall split this Hamiltonian into two parts: $H = H_0 + \delta H$.

The eigenstates and eigenfrequencies of the unperturbed Hamiltonian

$$H_0 = \frac{(\mathbf{p} + e\mathbf{A})^2}{2m_e} - eV + \mathrm{g}_e \mu_B \mathbf{B} \cdot \mathbf{S} \tag{4.14}$$

can be determined exactly, whereas the remainder,

$$\begin{aligned} \delta H = {} & \frac{\beta_m m_e}{M_{\mathrm{Pl}}} \left(\phi_{rr} \frac{r^2}{2r_0^2} + \phi_{zz} \frac{z^2}{2z_0^2} \right) \\ & - e\delta V + 2\mu_B (\mathbf{p} + e\mathbf{A}) \cdot \delta\mathbf{A} + \mathrm{g}_e \mu_B \mathbf{S} \cdot (\nabla \times \delta\mathbf{A}), \end{aligned} \tag{4.15}$$

will be treated as a small perturbation. This part of the Hamiltonian captures the leading effects from a chameleon-like scalar. The term proportional to β_m in the first line represents the fifth force acting on the electron, while the remaining terms in the second line are the result of the scalar's bubble profile transforming the vacuum into a dielectric medium. By substituting (4.9), (4.11), and (4.13) into the rhs of the modified Maxwell equation in (3.4), this classical vacuum-polarisation effect can be shown to induce the following first-order corrections to the electromagnetic fields:

$$\delta V = \frac{V_0}{2d^2} \frac{\beta_\gamma}{M_{\mathrm{Pl}}} \left(\phi_{rr} \frac{r^4}{16 r_0^2} - \phi_{zz} \frac{z^4}{6 z_0^2} \right), \tag{4.16a}$$

$$\delta\mathbf{A} = B_0 \phi_{rr} \frac{\beta_\gamma}{M_{\mathrm{Pl}}} \frac{r^2}{8 r_0^2} (\mathbf{x} \times \mathbf{e}_z). \tag{4.16b}$$

4.2.3 Unperturbed eigenstates

The unperturbed Hamiltonian in (4.14) can be further decomposed into three, mutually commuting parts. Let $H_0 = H_r + H_z + H_s$, where the radial, axial, and spin Hamiltonians are, respectively,

$$H_r = \frac{1}{2m_e} (\pi_x^2 + \pi_y^2) - \frac{1}{4} m_e \omega_z^2 r^2, \tag{4.17a}$$

$$H_z = \frac{1}{2m_e}\pi_z^2 + \frac{1}{2}m_e\omega_z^2 z^2, \tag{4.17b}$$

$$H_s = \frac{g_e}{2}\omega_0 S_z. \tag{4.17c}$$

These Hamiltonians are written in terms of the mechanical momentum $\boldsymbol{\pi} = \mathbf{p} + e\mathbf{A}$, the (bare) cyclotron frequency $\omega_0 = eB_0/m_e$, and the axial frequency $\omega_z = (eV_0/m_e d^2)^{1/2}$.

From the form of H_z, it is easy to spot that the electron will undergo simple harmonic motion with frequency ω_z in the axial direction. Making the familiar transformation

$$z = \frac{1}{\sqrt{2m_e\omega_z}}(a_z + a_z^\dagger), \qquad \pi_z = -i\sqrt{\frac{m_e\omega_z}{2}}(a_z - a_z^\dagger) \tag{4.18}$$

allows us to express $H_z = \omega_z(a_z^\dagger a_z + 1/2)$ in terms of creation and annihilation operators. It turns out that the radial Hamiltonian H_r can also be diagonalised into a sum of two decoupled harmonic oscillators. To see this, first define two more frequencies $\omega_\pm$ via

$$2\omega_\pm := \omega_0 \pm (\omega_0^2 - 2\omega_z^2)^{1/2} \tag{4.19}$$

and denote their difference by $\Delta\omega := \omega_+ - \omega_-$. Then, making the transformation [435]

$$\begin{aligned} x &= \frac{i(a_c - a_c^\dagger + a_m - a_m^\dagger)}{\sqrt{2m_e\Delta\omega}}, \qquad \pi_x = \sqrt{\frac{m_e}{2\Delta\omega}}\left[\omega_+(a_c + a_c^\dagger) - \omega_-(a_m + a_m^\dagger)\right], \\ y &= \frac{-(a_c + a_c^\dagger - a_m - a_m^\dagger)}{\sqrt{2m_e\Delta\omega}}, \qquad \pi_y = i\sqrt{\frac{m_e}{2\Delta\omega}}\left[\omega_+(a_c - a_c^\dagger) + \omega_-(a_m - a_m^\dagger)\right] \end{aligned} \tag{4.20}$$

ultimately leaves us with

$$H_0 = \omega_+\left(a_c^\dagger a_c + \frac{1}{2}\right) + \omega_z\left(a_z^\dagger a_z + \frac{1}{2}\right) - \omega_-\left(a_m^\dagger a_m + \frac{1}{2}\right) + \frac{g_e}{2}\omega_0 S_z. \tag{4.21}$$

An eigenstate $|n_c, n_z, n_m, m_s\rangle$ of this Hamiltonian is thus specified by four quantum numbers. Three of these correspond to the occupation numbers $n_i = \langle a_i^\dagger a_i\rangle = 0, 1, 2, \ldots$ of the harmonic oscillators, while the fourth is the electron's spin state $m_s = \pm 1/2$.

Physically, the oscillators with frequencies ω_+, ω_z, and ω_- correspond to cyclotron, axial, and magnetron motion, respectively. A cartoon illustrating these different contributions to the electron's overall motion in the Penning trap is shown in Figure 4.3. (Further details can be found in the review by Brown and Gabrielse [435].) That ω_+ is slightly larger than the bare cyclotron frequency ω_0 is due

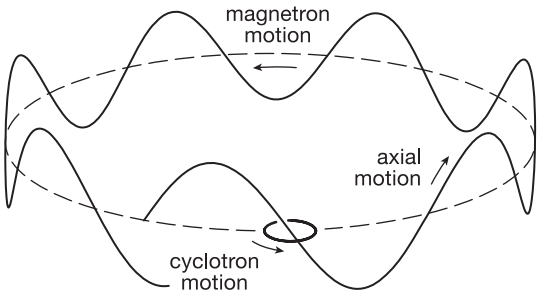


Figure 4.3 Illustration of an electron's orbit in a Penning trap. Magnetron motion corresponds to the electron travelling slowly along the large dashed circle. Combining this circular orbit with the oscillatory motion in the axial direction produces the wavy line known as the guiding-centre motion. Superimposed on this trajectory is the rapid cyclotron motion around the small circle.

to the confining effect of the electrostatic potential. Note also that the minus sign appearing in front of ω_- in (4.21) signifies that magnetron motion is unstable and unbounded from below.

4.2.4 Axial and magnetron motion

For the particular choice of parameters in this experiment (see Table 4.1), we have that

$$\omega_- \ll \omega_z \ll \omega_+. \tag{4.22}$$

This hierarchy ensures that both the axial and magnetron motions are semiclassical. When measurements used to infer the value of a_e are being made, the axial motion is in thermal equilibrium with the detection amplifier circuit at a temperature $T_z \sim 230\,\mathrm{mK}$ [422]. The average axial quantum number is thus given by $n_z \sim k_B T_z/\omega_z \sim 24$.

Similarly, the magnetron motion thermalises with a temperature $T_m \sim -(\omega_-/\omega_z)T_z$, assuming maximum axial sideband cooling [422, 435]. This relation sets the axial and magnetron quantum numbers equal to one another: $n_m \sim n_z \sim 24$. The negative temperature here is again representing the fact that magnetron motion is unstable. Nevertheless, its decay time is on the order of billions of years [422, 435]; hence, this state is metastable on the timescale of the experiment.

These estimates justify truncating the scalar-field profile to quadratic order in (4.13). For $n_c \sim \mathcal{O}(1)$, the expectation values

$$\left\langle \frac{r^2}{r_0^2} \right\rangle = \frac{2(n_c + n_m + 1)}{m_e \Delta\omega r_0^2} \sim 10^{-10}, \qquad \left\langle \frac{z^2}{z_0^2} \right\rangle = \frac{n_z + 1/2}{m_e \omega_z z_0^2} \sim 10^{-7} \tag{4.23}$$

demonstrate that the spread of the electron wavefunction does indeed remain very close to the origin.

4.2.5 Frequency shifts

In what follows, the measured value of an angular frequency will be denoted by $2\pi\nu$ so as to distinguish it from its theoretical counterpart ω. The electron's magnetic moment g_e is determined experimentally by measuring three of these transition frequencies:

(1) The cyclotron frequency ν_c is measured by resonantly exciting the transition $(n_c, m_s) = (0, 1/2) \to (1, 1/2)$ at fixed n_z and n_m. Taking the difference in expectation values $\langle H \rangle$ for these two states leads to the equation

$$2\pi\nu_c = \omega_+ - 3\pi\delta_{\text{rel}} + \delta\omega_c, \tag{4.24a}$$

where $\delta\omega_c$ follows from $\langle \delta H \rangle$ and represents the systematic shift generated by the chameleon-like scalar. Explicitly, one finds

$$\delta\omega_c = \frac{\phi_{rr}}{M_{\text{Pl}} r_0^2}\left[\frac{\beta_m}{\Delta\omega} - \frac{\beta_\gamma\omega_0}{2m_e\Delta\omega}\left(\frac{g_e}{2} + (2n_m+3)\frac{\omega_+}{\Delta\omega}\right)\right.$$
$$\left. + (n_m+1)\frac{\beta_\gamma\omega_z^2}{2m_e\Delta\omega^2} - (2n_m+1)\frac{\beta_\gamma\omega_0\omega_-}{2m_e\Delta\omega^2}\right]. \tag{4.24b}$$

Additionally, a small extra term δ_{rel} ($\approx 10^{-9}\nu_c$) has been inserted by hand into (4.24a) to account for the leading relativistic correction, which is large enough to be relevant at the level of precision achieved by this experiment [422, 435].

(2) The anomaly frequency ν_a corresponds to the transition $(n_c, m_s) = (1, -1/2) \to (0, 1/2)$ at fixed n_z and n_m. One finds

$$2\pi\nu_a = \frac{g_e}{2}\omega_0 - \omega_+ + \delta\omega_a, \tag{4.24c}$$

where the systematic shift from the scalar is

$$\delta\omega_a = -\frac{\phi_{rr}}{M_{\text{Pl}} r_0^2}\left[\frac{\beta_m}{\Delta\omega} + (2n_m+3)\frac{\beta_\gamma\omega_0}{2m_e\Delta\omega}\left(\frac{g_e}{2} - \frac{\omega_+}{\Delta\omega}\right)\right.$$
$$\left. + (n_m+1)\frac{\beta_\gamma\omega_z^2}{2m_e\Delta\omega^2} - (2n_m+1)\frac{\beta_\gamma\omega_0\omega_-}{2m_e\Delta\omega^2}\right]. \tag{4.24d}$$

(3) The axial frequency ν_z is measured from resonant transitions that satisfy $|\Delta n_z| = 1$, with all other quantum numbers fixed:

$$2\pi\nu_z = \omega_z + \delta\omega_z. \tag{4.24e}$$

While the result does not change significantly, for definiteness we shall define $\delta\omega_z$ as the average shift for the two transitions $n_z \to n_z \pm 1$. In this case, we have that

$$\delta\omega_z = \frac{\phi_{zz}}{M_{\rm Pl} z_0^2}\left(\frac{\beta_m}{2\omega_z} - (2n_z+1)\frac{\beta_\gamma}{8m_e}\right). \tag{4.24f}$$

The results in (4.24) form a set of three simultaneous equations that relate g_e to the measured frequencies $\nu_i = (\nu_c, \nu_a, \nu_z)$ and the theoretical parameters $\omega_i = (\omega_0, \omega_z, \omega_+, \omega_-)$. While some of these parameters can be calculated by using the values in Table 4.1, those quantities are not known to the same level of precision as the measured frequencies ν_i. Thus, a precise determination of the electron's magnetic moment is obtained by eliminating all instances of ω_i from (4.24) to obtain an expression for g_e that depends only on ν_i. Performing this elimination requires two additional, independent equations. These are supplied by the definitions of $\omega_\pm$ in (4.19), which can be rearranged to read[1]

$$\omega_0 = \omega_+ + \omega_-, \qquad \omega_- = \omega_z^2/(2\omega_0). \tag{4.25}$$

Because the frequency shifts $\delta\omega_i$ exhibit a highly nonlinear dependence on the parameters ω_i, it is easiest to solve this set of equations in two stages. First, we obtain a zeroth-order solution by neglecting the effects of the scalar field. Also making use of the hierarchy in (4.22) to discard terms that are irrelevant at the level of experimental precision, the result is [421,422]

$$\overline{a}_e^{\rm exp} = (\overline{g}_e^{\rm exp}/2) - 1 = \frac{\nu_a - \nu_z^2/(2\nu_c)}{\nu_c + 3\delta_{\rm rel}/2 + \nu_z^2/(2\nu_c)}. \tag{4.26}$$

The superscript "exp" is used to distinguish this result from the theoretical prediction in (4.8) and the overline signifies that this is the zeroth-order solution. We now take the frequency shifts $\delta\omega_i$ into account by making the replacement $\nu_i \to \nu_i - \delta\omega_i/(2\pi)$ in the expression above. Expanded to first order in $\delta\omega_i$ and judiciously using the hierarchy in (4.22) once again, the final result is

$$a_e^{\rm exp} = \overline{a}_e^{\rm exp} - \delta a_e^{\rm exp}, \tag{4.27}$$

where the systematic "cavity shift" due to the vacuum bubble in the cavity is given by

$$\delta a_e^{\rm exp} = -\frac{\beta_m}{M_{\rm Pl}(2\pi\nu_c)^2}\left(\frac{\phi_{rr}}{r_0^2} + \frac{\phi_{zz}}{2z_0^2}\right) + \frac{\beta_\gamma}{M_{\rm Pl}(2\pi\nu_c)^2}\left(\frac{\pi\nu_a}{m_e}\frac{\phi_{rr}}{r_0^2} + \frac{(2n_z+1)\pi\nu_z}{4m_e}\frac{\phi_{zz}}{z_0^2}\right). \tag{4.28}$$

[1] These relations are exact for an ideal Penning trap, but are also approximately true for a real trap with small imperfections due to the hierarchy in (4.22) and an invariance theorem [436].

This result can be simplified further by using the equation of motion in (2.30) to replace the second derivatives ϕ_{rr} and ϕ_{zz} with the effective potential. After also defining $\xi = (\phi_{zz} z_0^{-2})/(\phi_{rr} r_0^{-2})$ and using the value $n_z \sim 24$, (4.28) can be rewritten as

$$\delta a_e^{\rm exp} = \left[-\frac{\beta_m}{2M_{\rm Pl}(2\pi\nu_c)^2} + \frac{\beta_\gamma}{M_{\rm Pl}(2\pi\nu_c)^2}\left(\frac{\pi\nu_a}{m_e}\frac{1}{2+\xi} + \frac{49\pi\nu_z}{4m_e}\frac{\xi}{2+\xi}\right)\right] V'_{\rm eff}(\phi_0). \tag{4.29}$$

At this stage, it is worth remarking that the terms proportional to β_γ, which arise from the classical polarisation of the vacuum engendered by the scalar field, are strongly suppressed by factors of $\nu_a/m_e \sim \nu_z/m_e \sim 10^{-13}$ relative to the β_m term. Further noting that ξ must be positive definite as ϕ_0 is a maximum point, it is easy to check that these terms are always negligible within the EFT's regime of validity, and so do not lead to any meaningful constraint on β_γ. This conclusion may seem somewhat surprising at first glance, as one might imagine that the large magnetic field in the cavity would be useful for such a purpose. On the contrary, it turns out to offer little advantage because of the particular combination of frequencies that have to be measured in this experiment. The two terms involving $\delta\mathbf{A}$ in (4.15) approximately cancel out in $\delta a_e^{\rm exp}$, thus the leading effect that survives is due to δV, which is much smaller. For this reason, we shall omit the β_γ terms in (4.29) henceforth. Moving forward, the expression we will use for the cavity shift is

$$\delta a_e^{\rm exp} = -\frac{\beta_m V'_{\rm eff}(\phi_0)}{2M_{\rm Pl}(2\pi\nu_c)^2}. \tag{4.30}$$

4.3 Measuring the fine-structure constant, again

Let us now pick up from where we left off in Chapter 3.4 and return to the issue of constraining these models. From our earlier discussion, recall that the presence of a chameleon-like scalar — if not taken into account — biases the value of the fine-structure constant inferred from a given experiment. If $\overline{\alpha}(\mathrm{H})$ is the biased value inferred from the spectral lines of atomic hydrogen, for instance, then the true value of α is obtained by subtracting the bias $\delta\alpha(\mathrm{H})$; i.e., $\alpha = \overline{\alpha}(\mathrm{H}) - \delta\alpha(\mathrm{H})$. The same logic also applies in this chapter. If the effects of the scalar are not accounted for, then the value $\overline{\alpha}(\mathrm{g}_e)$ determined from a measurement of g_e is similarly biased by an amount $\delta\alpha(\mathrm{g}_e)$. In Chapter 5, we will use the fact that $\overline{\alpha}(\mathrm{H})$ and $\overline{\alpha}(\mathrm{g}_e)$ are in good agreement to establish an upper bound on $|\delta\alpha(\mathrm{g}_e) - \delta\alpha(\mathrm{H})|$, which can then be recast into constraints on the model parameters. Before we are able to do so, however, we must first obtain an expression for $\delta\alpha(\mathrm{g}_e)$. That is the goal of this section.

To translate a measurement of the electron's magnetic moment into a determination of the fine-structure constant, we should equate the measured value in (4.27) to the theoretical prediction in (4.8)

and solve for α. Because a_e can be measured to within a relative uncertainty of less than one part per billion [421], obtaining a precise value for α requires that we are able to calculate $a_e^{\rm th}$ to a similar level of precision. Fortunately, seven decades after Schwinger's seminal contribution, heroic calculations by Aoyama and collaborators have made this a possibility [413].

Schematically, the Standard Model prediction for the electron's anomalous magnetic moment is

$$a_e^{\rm SM} = \sum_{n=1}^{N} C_n (\alpha/\pi)^n + a_e^{\rm ew} + a_e^{\rm had}, \tag{4.31}$$

where the first term is the asymptotic series from QED, calculations for which have now been completed up to $N = 5$ loops [413, 437, 438]. (Note from our earlier discussion that $C_1 = 1/2$.) Also relevant at the requisite level of precision are smaller contributions from the electroweak and hadronic sectors, which are encoded in the remaining two terms; see Ref. [428] for further details. By substituting this into (4.8) and equating the result to (4.27), we obtain an implicit equation for α in terms of the lepton masses, several other Standard Model quantities, and the parameters $(G, \beta_m, \beta_\gamma)$. Of course, these last three parameters appear only as part of the scalar's contribution.

Let us represent this implicit equation as $f(\alpha; \Pi, G) = 0$, where Π is shorthand for all of the other parameters. (The astute reader should recognise the parallels with our earlier discussion in Chapter 3.4.) In the limit $G \to 0$, which is a proxy for turning off the scalar's effects, the inferred value for the fine-structure constant obtained from this experiment is [437]

$$[\overline{\alpha}(\mathrm{g}_e)]^{-1} = 137.035\,999\,1491\,(331). \tag{4.32}$$

However, this value is biased if a chameleon-like scalar is active in the experiment. The true value of the fine-structure constant $\alpha = \overline{\alpha} - \delta\alpha$ is obtained by subtracting this bias, which we determine by requiring that $f(\overline{\alpha} - \delta\alpha; \Pi, G) = 0$. To first order in G, the solution is

$$\delta\alpha = \frac{G[\partial f(\overline{\alpha}; \Pi, G)/\partial G]_{G=0}}{[\partial f(\alpha; \Pi, 0)/\partial\alpha]_{\alpha=\overline{\alpha}}}. \tag{4.33}$$

The numerator is the sum of the scalar's contributions to both $a_e^{\rm th}$ and $a_e^{\rm exp}$, while the denominator is simply the first derivative of $a_e^{\rm SM}$ with respect to α. Expanded to leading order, the systematic shift that a chameleon-like scalar introduces into this experiment is found to be

$$\delta\alpha(\mathrm{g}_e) = 2Gm_e^2\left[\beta_m^2 I_1(m_0/m_e) + 2\beta_m\beta_\gamma I_2(m_0/m_e)\right] - \frac{\beta_m V'_{\rm eff}(\phi_0)}{4\pi M_{\rm Pl}\nu_c^2}. \tag{4.34}$$

Observe that the central field value ϕ_0 of the scalar in the Penning trap enters into the above equation via the coupling strengths $\beta_i \equiv \beta_i(\phi_0)$, the effective mass $m_0 \equiv m_{\rm eff}(\phi_0)$, and the first derivative of the effective potential $V'_{\rm eff}(\phi_0)$. While the calculations in this chapter have so far been valid for any chameleon-like theory, we must specialise to specific models if we are to determine the value of ϕ_0. To conclude this chapter, we will consider two archetypal models: the chameleon and the symmetron.

4.3.1 Chameleon profile in the vacuum cavity

I mentioned earlier that an accurate solution to (4.12) is only accessible numerically because the cavity radius and height are of comparable size, which renders the problem strictly two-dimensional. While this is true, relying solely on numerical methods is computationally very expensive, as the calculation must be repeated every time we pick a new point in parameter space. Fortunately, for this particular problem, it is possible to expedite the process, as we are only interested in the value of the scalar field at the centre of the cavity. The trick is to deform the analytic, one-dimensional solutions in Chapter 2.4 by introducing a number of free parameters, whose job is to account for differences associated with the two-dimensional nature of the cylindrical Penning trap. The resulting formulae can then be calibrated to match a small set of numerical calculations. As we will see, this semi-empirical approach provides solutions for the central field value ϕ_0 that agree with the results from numerical calculations to within 1%.

For the chameleon, our starting point is the approximate solution [cf. (2.55)]

$$\phi_0 = \left(\frac{n(n+1)\Lambda^{4+n}R^2}{[\mathrm{sinhc}^{-1}(n+2)]^2}\right)^{1/(2+n)} \tag{4.35}$$

for the central field value in a spherical cavity of radius R. Recall that this solution is contingent on two assumptions: First, we assumed that the walls surrounding the cavity are sufficiently dense that the field immediately minimises its effective potential once it penetrates the boundary. This is called the zero-skin-depth approximation. Second, we also assumed that the interior of the cavity is a perfect vacuum. We will continue to use both of these approximations in what follows, and will discuss their validity shortly.

The key insight needed to extend this result to higher dimensions is to notice that were we to repeat the calculations in Chapter 2.4 for different one-dimensional geometries, we would obtain solutions that are almost identical to (4.35). The only difference is that R would be replaced by the relevant length scale in the problem, while a function more appropriate to the geometry would appear in place of $\mathrm{sinhc}(z)$. For instance, in a plane-parallel cavity of width 2ℓ, the correct solution follows after simply replacing $R \to \ell$ and $\mathrm{sinhc}(z) \to \cosh(z)$ in (4.35) [2]. Similarly, in the case of an infinitely-long

cylindrical cavity, R should be substituted for the cylindrical radius while $\mathrm{sinhc}(z)$ must be replaced by the modified Bessel function $J_0(z)$ [394].

The fact that the solutions for these different geometries are related by simple replacements is not a coincidence. As we already discussed in Chapter 2.4, (4.35) can be interpreted as saying that a chameleon field always adjusts its local Compton wavelength to be on the order of the size of the cavity to which it is confined. If we trust this intuitive picture, then we might expect that the solution for the central field value in any convex cavity ought to take the form

$$\phi_0 \sim \left(\frac{n(n+1)\Lambda^{4+n} L^2}{[f^{-1}(n+2)]^2} \right)^{1/(2+n)}, \tag{4.36}$$

where L is some relevant length scale and $f(z)$ is some function appropriate to the geometry. Although we do not have a systematic method for determining this function, it can always be parametrised in terms of its Padé approximant [439]:

$$\frac{1}{[f^{-1}(n+2)]^2} = \frac{\sum_{i=0}^{p} \delta_i n^i}{1 + \sum_{j=1}^{q} f_j n^j}. \tag{4.37}$$

The integers (p, q) specify the order of this approximation, while the coefficients δ_i and f_j are free parameters to be calibrated. To prevent introducing too many degrees of freedom, we should keep the order as low as possible. It turns out that a Padé approximant with $(p, q) = (1, 1)$ and $\delta_1 = f_1 = 1$, with only δ_0 left unspecified, will be sufficient for our purposes. Thus, our analytic estimate for the central field value in the cylindrical Penning trap is

$$\phi_0 \sim \left[n(n+\delta_0)\Lambda^{4+n} L^2 \right]^{1/(2+n)}. \tag{4.38}$$

As shown in Figure 4.4, the best-fit[2] values $L = 1.40$ mm and $\delta_0 = 2.78$ lead to an analytic estimate for ϕ_0 that agrees with the results from full numerical calculations to within 1%. This excellent agreement guarantees that the former can be used reliably when we turn to establish constraints on the chameleon parameter space in the next chapter.

Let us also say a few words about the numerical method and the validity of our approximations. The full, nonlinear chameleon profile in the cylindrical Penning trap is determined by integrating (4.12) through successive under-relaxation using the Gauss–Seidel scheme [440] for 12 values of $n \in (0, 13)$ at fixed $\Lambda = 2.4$ meV. Further details on this code can be found in Ref. [441]. For illustrative purposes,

[2]These values are determined by using Mathematica's native `NonlinearModelFit` routine.

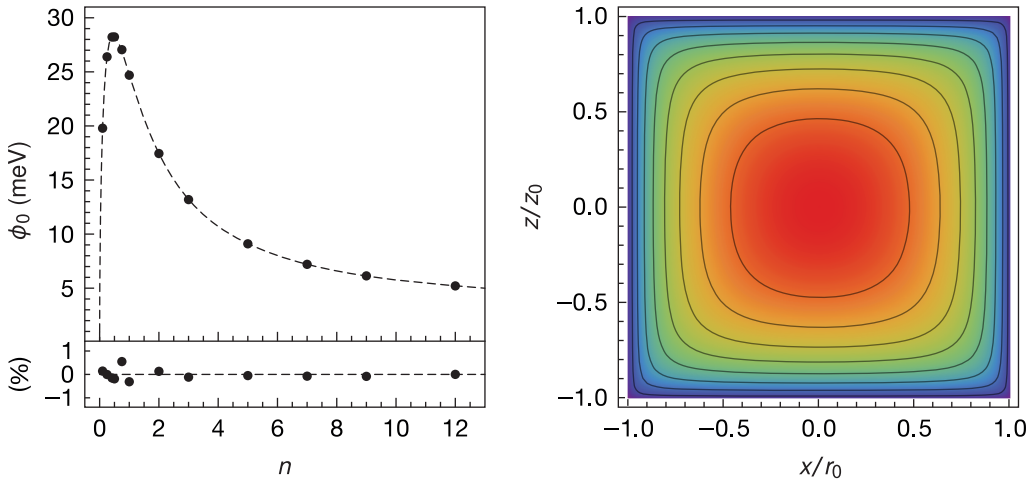


Figure 4.4 Chameleon profile in a cylindrical vacuum cavity. Left panel: Best-fitting analytic approximation (dashed line) to the central field value ϕ_0 for different values of n with $\Lambda = 2.4$ meV, compared with the numerical results (black dots). The lower panel displays the percentage difference between the numerical and analytic results — all points are in agreement to within 1%. Right panel: Chameleon profile in the x–z plane for $n = 1$ and $\Lambda = 2.4$ meV. The field value along the innermost contour is 90% of the value at the origin. Moving outwards, successive contours are 80%, 70%, etc. of the central field value. The field reaches 10% near the boundary of the cavity before quickly plummeting to $\phi \approx 0$ once inside the walls.

the full chameleon profile for $n = 1$ is also presented in Figure 4.4. Profiles for the remaining values of n are qualitatively similar.

Different values of Λ need not be explored numerically when making the perfect-vacuum approximation, since (4.12) admits the scaling symmetry $\Lambda \to s\Lambda$, $\phi \to s^{(4+n)/(2+n)}\phi$ in this limit. In practice, this approximation is valid as long as the energy densities in the cavity are negligible compared to the contribution from the bare potential; i.e., when

$$\frac{\rho_0}{M_c} + \frac{\rho_{\rm em}}{M_\gamma} \ll \frac{n\Lambda^{4+n}}{\phi_0^{n+1}}. \tag{4.39}$$

The zero-skin-depth approximation, meanwhile, is useful for a different reason: It allows us to fix the scalar field at the value ϕ_∞ that minimises its local effective potential once it reaches the walls. In reality, the field cannot achieve this instantaneously, but rather decays to the minimum within a distance set by its local Compton wavelength $\sim m_\infty^{-1} = 1/\sqrt{V''_{\rm eff}(\phi_\infty)}$. Following Elder et al. [441], the zero-skin-depth approximation is taken to be valid whenever the chameleon's Compton wavelength in the walls is much smaller than the grid spacing used in the numerical calculation:

$$m_\infty^{-1} \ll L_{\rm grid}. \tag{4.40}$$

The results in this thesis use $L_{\rm grid} = 0.1$ mm in all three spatial directions.

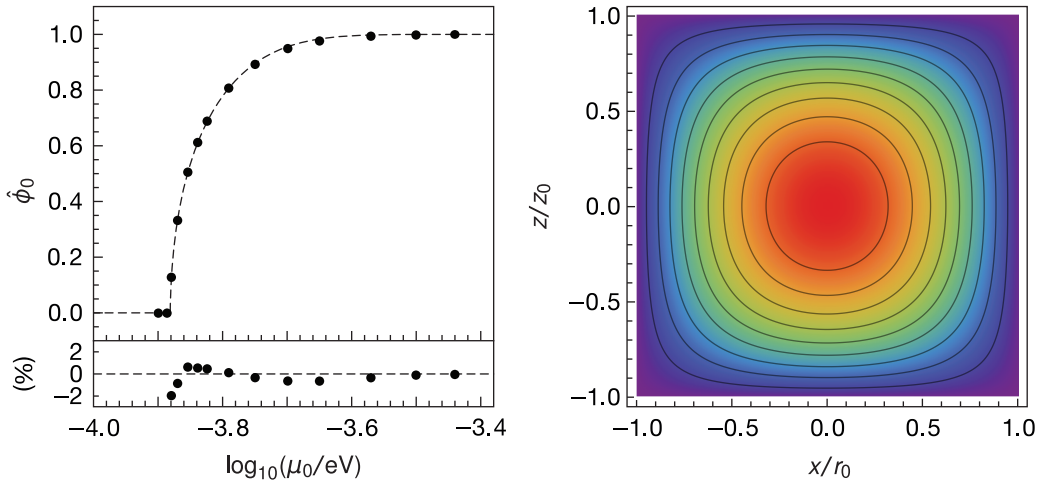


Figure 4.5 Symmetron profile in a cylindrical vacuum cavity. Left panel: Best-fitting analytic approximation (dashed line) to the dimensionless central field value $\hat{\phi}_0$ for different values of μ_0, compared with the numerical results (black dots). The lower panel displays the percentage difference between the numerical and analytic results — all points agree to less than 1%, except the first three near $\mu_0 = 10^{-3.9}$ eV, where $\hat{\phi}_0$ differs from zero only in the eighth (or higher) decimal place. Any discrepancy here is of no concern, since the numerical accuracy is unreliable for such small values of the field. Right panel: Symmetron profile in the x–z plane for $\mu_0 = 10^{-3.82}$ eV. The field value along the innermost contour is 90% of the value at the origin. Moving outwards, successive contours are 80%, 70%, etc. of the central field value. The field reaches $\phi = 0$ once inside the walls.

4.3.2 Symmetron profile in the vacuum cavity

Similar steps may now be repeated for the symmetron. In this case, our starting point is the exact solution in Chapter 2.4.2 for the field profile in a plane-parallel cavity of width 2ℓ, whose central field value ϕ_0 is determined by the self-consistency condition

$$\hat{\phi}_0 \,\mathrm{cd}\left(\mu_0 \ell \frac{\hat{\phi}_0}{\sqrt{2v^2}}, v\right) = 0. \tag{4.41}$$

Recall that we previously defined $\hat{\phi}_0 = \phi_0/(\mu_0/\sqrt{\lambda})$ and $v^2 = \hat{\phi}_0^2/(2-\hat{\phi}_0^2)$, although we must now adjust the definition of μ_0 to include the electromagnetic energy density $\rho_{\rm em}$ in the interior of the Penning trap; namely, let $\mu_0^2 = \mu^2 - \rho_0 M_s^{-2} - \rho_{\rm em} M_\gamma^{-2}$. For simplicity, we shall neglect the subleading contribution from the electric field to this energy density, such that $\rho_{\rm em} \sim \mathbf{B}^2/2$ has a constant value everywhere in the interior. Now, as we did for the chameleon, we deform this solution by introducing a small number of free parameters. Through trial and error, the equation

$$\hat{\phi}_0 \,\mathrm{cd}\left((\mu_0 L)^{1+\delta_0} \frac{\hat{\phi}_0}{\sqrt{2v^2}}, v\right) = 0 \tag{4.42}$$

was found to accurately reproduce the results of full numerical calculations to within 1%; see Figure 4.5.

Roughly speaking, the characteristic length scale L determines the point at which the curve in the left panel starts to rise above zero. The deformation parameter δ_0 then adjusts how quickly this curve reaches its plateau. The best-fit parameters are $L = 1.96$ mm and $\delta_0 = 0.70$.

The numerical results themselves are obtained by using the same under-relaxation method as in Section 4.3.1 and are valid as long as the field profile satisfies the zero-skin-depth approximation, which is used to fix $\phi = 0$ once it reaches the walls. (Note that the perfect-vacuum approximation is not needed in this case because the symmetron's bare and effective potentials have the same functional form.) Following Elder et al. [441] once again, this is taken to be a valid assumption as long as the field's local Compton wavelength in the walls is much smaller than the grid spacing used in the numerical calculation:

$$m_\infty^{-1} \ll L_{\rm grid}. \tag{4.43}$$

For the symmetron, a finer numerical grid with $L_{\rm grid} = 0.05$ mm in all three spatial directions was needed to ensure convergence, especially for solutions with $\hat{\phi}_0$ close to zero.

Appendix 4.A One-loop Feynman diagrams

A chameleon-like scalar contributes to the electron's anomalous magnetic moment a_e at the one-loop level via three Feynman diagrams, as shown in Figure 4.2. In this appendix, we briefly outline the steps involved in the evaluation of these diagrams. The identities used throughout this discussion have been lifted from Srednicki's book [442].

Yukawa diagram We begin with Figure 4.2(a), which involves only the Yukawa-like coupling between the scalar and the electron. Working in d dimensions for now, standard Feynman rules dictate that the contribution of Figure 4.2(a) to the QED vertex function is

$$i\Gamma^\mu_{\rm (a)} = \frac{\beta_m^2 m_e^2}{M_{\rm Pl}^2} \int \frac{{\rm d}^d \ell}{(2\pi)^d} \frac{(-\not{\ell} - \not{p}' + m_e)\gamma^\mu(-\not{\ell} - \not{p} + m_e)}{[(\ell + p')^2 + m_e^2][(\ell + p)^2 + m_e^2][\ell^2 + m_0^2]}. \tag{4.A.1}$$

To perform the integral, we first introduce Feynman parameters (x_1, x_2, x_3) along with the measure

$$\int {\rm d}F_n := (n-1)! \int_0^1 {\rm d}x_1 \, \cdots \int_0^1 {\rm d}x_n \, \delta\left(\textstyle\sum_{i=1}^n x_i - 1\right) \tag{4.A.2}$$

to rewrite $\Gamma^\mu_{(a)}$ in terms of a new integration variable $k = \ell + x_1 p + x_2 p'$:

$$i\Gamma^\mu_{(a)} = \frac{\beta_m^2 m_e^2}{M_{Pl}^2} \int dF_3 \int \frac{d^d k}{(2\pi)^d} \frac{N^\mu_{(a)}}{(k^2 + D_{(a)})^3}. \tag{4.A.3}$$

The objects in the numerator and denominator above are, respectively,

$$N^\mu_{(a)} = (-\not{\ell} - \not{p}' + m_e)\gamma^\mu(-\not{\ell} - \not{p} + m_e)|_{\ell = k - x_1 p - x_2 p'}, \tag{4.A.4}$$

$$D_{(a)} = x_1(1-x_1)p^2 + x_2(1-x_2)p'^2 - 2x_1 x_2 p\cdot p' + (x_1+x_2)m_e^2 + x_3 m_0^2. \tag{4.A.5}$$

As $N^\mu_{(a)}$ sits under an integral over all k, any of its terms that are odd in k^μ can be discarded as they vanish upon integration. For the same reason, we can replace $k^\mu k^\nu \to (k^2/d)\eta^{\mu\nu}$ under the integral as this is the only part that survives. In doing so, the term in $N^\mu_{(a)}$ that is quadratic in k simplifies to

$$\not{k}\gamma^\mu \not{k} \to \frac{k^2}{d}\gamma^\nu\gamma^\mu\gamma_\nu = \left(\frac{d-2}{d}\right) k^2 \gamma^\mu. \tag{4.A.6}$$

Because this term is proportional to γ^μ, it contributes only to the electric form factor F_1. Accordingly, the terms that are presently of interest to us all reside in the k-independent part of $N^\mu_{(a)}$. Let us define

$$n^\mu_{(a)} := N^\mu_{(a)}|_{k=0} = [x_1\not{p} - (1-x_2)\not{p}' + m_e]\gamma^\mu[x_2\not{p}' - (1-x_1)\not{p} + m_e]. \tag{4.A.7}$$

We may simplify this expression further as our interest in Γ^μ extends only as far as to understand how it contributes to the S-matrix element $\bar{u}(p')\Gamma^\mu(p,p')u(p)A_\mu(q)$, where the external photon A_μ is taken to be classical and off shell so as to correspond to a background magnetic field. The external-leg factors for the electron satisfy

$$\not{p}u = -m_e u, \qquad \bar{u}'\not{p}' = -m_e\bar{u}', \tag{4.A.8}$$

where I write $u \equiv u(p)$ and $\bar{u}' \equiv \bar{u}(p')$ for brevity. The spin indices on u and $\bar{u}'$ have also been suppressed as they are inessential here. The name of the game is now to reorder the terms in $n^\mu_{(a)}$ by exploiting the identity $\{\gamma^\mu,\gamma^\nu\} = -2\eta^{\mu\nu}$ until use can be made of (4.A.8). The end result is

$$\begin{aligned}\bar{u}' n^\mu_{(a)} u = {} & [4m_e^2 - (1-x_3)^2 m_e^2 - x_1 x_2 q^2]\,\bar{u}'\gamma^\mu u - m_e(1-x_3)(1+x_3)\,\bar{u}'(p'+p)^\mu u \\ & + m_e[(2x_1 - x_1^2) - (2x_2 - x_2^2)]\,\bar{u}'(p'-p)^\mu u.\end{aligned} \tag{4.A.9}$$

The second line changes sign under the interchange $x_1 \leftrightarrow x_2$ while the remainder of the integrand is unaffected; hence, this second line will vanish upon integration over the Feynman parameters, which are symmetric under $x_1 \leftrightarrow x_2$. For the remaining terms, we use the Gordon identity

$$\bar{u}'(p'+p)^\mu u = \bar{u}'(2m_e\gamma^\mu + 2iS^{\mu\nu}q_\nu)u \tag{4.A.10}$$

to recast them into a form comparable to (4.3). As before, terms proportional to γ^μ contribute only to F_1 and are not of interest to us. The contribution to the magnetic form factor F_2 from Figure 4.2(a) is

$$F_{2,(a)}(q^2) = -i\frac{\beta_m^2 m_e^2}{M_{\rm Pl}^2}\int \mathrm{d}F_3 \int \frac{\mathrm{d}^d k}{(2\pi)^d}\frac{2m_e^2(1-x_3)(1+x_3)}{(k^2+D_{(a)})^3}. \tag{4.A.11}$$

What remains is to perform the integral over k. When employing dimensional regularisation, use can be made of the master formula

$$\int \frac{\mathrm{d}^d k}{(2\pi)^d}\frac{(k^2)^a}{(k^2+D)^b} = i\frac{\Gamma(b-a-d/2)\Gamma(a+d/2)}{(4\pi)^{d/2}\Gamma(b)\Gamma(d/2)}D^{-(b-a-d/2)}, \tag{4.A.12}$$

where the factor of i comes from Wick rotating. The result remains finite in the $d \to 4$ limit, thus

$$F_{2,(a)}(q^2) = \beta_m^2\left(\frac{m_e}{4\pi M_{\rm Pl}}\right)^2 \int \mathrm{d}F_3\,\frac{m_e^2(1-x_3)(1+x_3)}{D_{(a)}(q^2)}. \tag{4.A.13}$$

To read off the anomalous magnetic moment, we put this function on shell by setting $p^2 = p'^2 = -m_e^2$ and $q^2 = 0$. As a result, the term in the denominator reduces to $D_{(a)}(0) = (1-x_3)^2 m_e^2 + x_3 m_0^2$. We may now freely integrate over x_1 and x_2, which generates an extra factor of $(1-x_3)$. Also renaming x_3 to x returns the desired result in (4.5):

$$F_{2,(a)}(0) = 2\beta_m^2\left(\frac{m_e}{4\pi M_{\rm Pl}}\right)^2 I_1(m_0/m_e), \tag{4.A.14}$$

$$I_1(z) = \int_0^1 \mathrm{d}x\,\frac{(1-x)^2(1+x)}{(1-x)^2+xz^2}. \tag{4.A.15}$$

Barr–Zee diagrams The remaining diagrams in Figure 4.2 that involve the photon coupling are often called Barr–Zee diagrams [430, 443]. The contribution from Figure 4.2(b) is

$$i\Gamma^\mu_{(b)} = \frac{\beta_m\beta_\gamma m_e}{M_{\rm Pl}^2}\int \frac{\mathrm{d}^d\ell}{(2\pi)^d}\frac{(\not{\ell}+m_e)\gamma_\nu[(\ell+p)^\mu q^\nu - q\cdot(\ell+p)\eta^{\mu\nu}]}{[\ell^2+m_e^2][(\ell+p')^2+m_0^2](\ell+p)^2}. \tag{4.A.16}$$

Defining k exactly as before, this can be rewritten as

$$i\Gamma^{\mu}_{(\mathrm{b})} = \frac{\beta_m \beta_\gamma m_e}{M_{\mathrm{Pl}}^2} \int \mathrm{d}F_3 \int \frac{\mathrm{d}^d k}{(2\pi)^d} \frac{N^{\mu}_{(\mathrm{b})}}{(k^2 + D_{(\mathrm{b})})^3}, \tag{4.A.17}$$

where
$$N^{\mu}_{(\mathrm{b})} = (\not{\ell} + m_e)\gamma_\nu(\ell + p)^\alpha q^\beta (\delta^\mu_\alpha \delta^\nu_\beta - \eta_{\alpha\beta}\eta^{\mu\nu})|_{\ell = k - x_1 p - x_2 p'}, \tag{4.A.18}$$

$$D_{(\mathrm{b})} = x_1(1-x_1)p^2 + x_2(1-x_2)p'^2 - 2x_1 x_2\, p\cdot p' + x_2 m_0^2 + x_3 m_e^2. \tag{4.A.19}$$

The terms in $N^{\mu}_{(\mathrm{b})}$ that are linear in k vanish upon integration, so we need only pay attention to the terms that are quadratic in or independent of k. The former simplifies to

$$N^{\mu}_{(\mathrm{b})} \supset \not{k}\gamma_\nu k^\alpha q^\beta (\delta^\mu_\alpha \delta^\nu_\beta - \eta_{\alpha\beta}\eta^{\mu\nu}) \to -i\frac{4k^2}{d} S^{\mu\nu} q_\nu \tag{4.A.20}$$

under the integral and contributes a log-divergent piece to F_2, which we regularise by evaluating the integral in $d = 4 - \epsilon$ dimensions. We keep the coupling strengths β_i dimensionless by pulling out an explicit mass dependence: $\beta_i \to \beta_i \mu^{\epsilon/2}$. Then using (4.A.12), one finds that

$$F_{2,(\mathrm{b})}(q^2) \supset \beta_m \beta_\gamma \left(\frac{m_e}{4\pi M_{\mathrm{Pl}}}\right)^2 \int \mathrm{d}F_3 \left[\frac{2}{\epsilon} + \log\left(\frac{\tilde{\mu}^2}{D_{(\mathrm{b})}}\right) + \mathcal{O}(\epsilon)\right] \tag{4.A.21}$$

after also defining a new mass scale $\tilde{\mu}^2 \equiv 4\pi e^{-\gamma_{\mathrm{E}}}\mu^2$ in terms of the Euler–Mascheroni constant γ_{E}. Working on shell, the denominator reduces to $D_{(\mathrm{b})}(0) = x_3^2 m_e^2 + x_2 m_0^2$. Finally, integrating over x_1, renaming x_3 to x, and making the substitution $x_2 = (1-x)y$ brings this into the form

$$F_{2,(\mathrm{b})}(0) \supset 2\beta_m \beta_\gamma \left(\frac{m_e}{4\pi M_{\mathrm{Pl}}}\right)^2 \left[\frac{1}{\epsilon} + \log\left(\frac{\tilde{\mu}}{m_e}\right) + I_2\left(\frac{m_0}{m_e}\right)\right], \tag{4.A.22}$$

$$I_2(z) = \int_0^1 \mathrm{d}x \int_0^1 \mathrm{d}y\, (x-1)\log[x^2 + (1-x)yz^2]. \tag{4.A.23}$$

This is half of the desired result in (4.6). The other half comes from evaluating Figure 4.2(c), which gives exactly the same contribution when on shell.

Having already obtained the desired outcome, it remains to show that the terms independent of k in $N^{\mu}_{(\mathrm{b})}$ and $N^{\mu}_{(\mathrm{c})}$ do not contribute to $F_2(0)$. Taking $N^{\mu}_{(\mathrm{b})}$ as an example, its k-independent terms are

$$n^{\mu}_{(\mathrm{b})} := N^{\mu}_{(\mathrm{b})}|_{k=0} = (\not{\ell} + m_e)\gamma_\nu(\ell + p)^\alpha q^\beta (\delta^\mu_\alpha \delta^\nu_\beta - \eta_{\alpha\beta}\eta^{\mu\nu})|_{\ell = -x_1 p - x_2 p'}. \tag{4.A.24}$$

Once again, we exploit the definition $\{\gamma^\mu, \gamma^\nu\} = -2\eta^{\mu\nu}$ to simplify this expression, all the while applying (4.A.8) judiciously. The end result is

$$\bar{u}' n^\mu_{(\mathrm{b})} u = \frac{1}{2} q^2 (4x_2 - x_3^2) m_e \, \bar{u}' \gamma^\mu u + 2q^2 x_1 x_2 \, \bar{u}' i S^{\mu\nu} q_\nu u, \tag{4.A.25}$$

having also used the Gordon identity in (4.A.10). Observe that these terms yield contributions to both F_1 and F_2 that are proportional to q^2, and so affect neither the electric charge nor the magnetic moment. The same conclusion holds also for the k-independent terms in $N^\mu_{(\mathrm{c})}$.

Feynman parameter integrals The Feynman parameter integrals $I_1(z)$ and $I_2(z)$ can easily be evaluated using numerical methods, but they also turn out to admit closed-form expressions. Let us discuss how these expressions come about. For the integral $I_1(z)$ in (4.A.15), we begin by decomposing its integrand into partial fractions:

$$\frac{(1-x)^2(1+x)}{(1-x)^2 + xz^2} = (1+x) - z^2 + \frac{z^2 + z^2(z^2-3)x}{x^2 + (z^2-2)x + 1}. \tag{4.A.26}$$

The third term can then be massaged to produce a piece that is of the form $f'(x)/f(x)$, which easily integrates to $\log f(x)$. Specifically, we write

$$\begin{aligned}\frac{(1-x)^2(1+x)}{(1-x)^2 + xz^2} = (1+x) - z^2 - \frac{z^2(3-z^2)}{2} \frac{2x + z^2 - 2}{x^2 + (z^2-2)x + 1} \\ - (1-z^2)(4-z^2) \frac{z^2}{2[x^2 + (z^2-2)x + 1]}.\end{aligned} \tag{4.A.27}$$

The three terms in the first line are straightforward to integrate, but more work is needed to evaluate the last term in the second line. To that end, let us define

$$\mathcal{I}_1(z) = \frac{1}{2} \int_0^1 \mathrm{d}x \, \frac{z^2}{x^2 + (z^2-2)x + 1}. \tag{4.A.28}$$

We evaluate this integral by first noting that the quadratic polynomial in the denominator can be written as $(x - a_+)(x - a_-)$ in terms of its roots,

$$a_\pm = 1 - \frac{z^2}{2} \pm \frac{z}{2} \sqrt{z^2 - 4}. \tag{4.A.29}$$

When in this form, the integrand can again be simplified by performing a partial fraction decomposition:

$$\frac{1}{(x-a_+)(x-a_-)} = \frac{1}{a_+ - a_-} \left(\frac{1}{x - a_+} - \frac{1}{x - a_-} \right), \tag{4.A.30}$$

which now allow us to perform the integral over x. Explicitly, we find

$$\begin{aligned}\mathcal{I}_1(z) &= \frac{z^2}{2(a_+ - a_-)}\left[\log\left(\frac{1-a_+}{1-a_-}\right) - \log\left(\frac{a_+}{a_-}\right)\right] \\ &= \frac{1}{\sqrt{1-4/z^2}}\log\left(\frac{z}{2} + \frac{1}{2}\sqrt{z^2-4}\right).\end{aligned} \tag{4.A.31}$$

Note that we are only interested in values of the function $I_1(z)$ when $z \geq 0$. For $z \in [0,2)$, the principal branch should be chosen such that this real-valued function is continuous in the domain of interest. Putting everything together, the final result is[3]

$$I_1(z) = \frac{3}{2} - z^2 - z^2(3-z^2)\log z - \frac{(1-z^2)(4-z^2)}{\sqrt{1-4/z^2}}\log\left(\frac{z}{2} + \frac{1}{2}\sqrt{z^2-4}\right). \tag{4.A.32}$$

Let us now turn our attention to the second Feynman parameter integral $I_2(z)$; cf. (4.A.23). We first perform the integral over y to obtain

$$I_2(z) = \int_0^1 \mathrm{d}x \left(1 - x + \frac{x^2\log x^2}{z^2} - \frac{[x^2 - z^2(x-1)]\log[x^2 - z^2(x-1)]}{z^2}\right). \tag{4.A.33}$$

The remaining integral over x is straightforward to evaluate in the case of the first three terms, but the last term requires more work. Focusing on this last term, define

$$\mathcal{I}_2(z) = \frac{1}{z^2}\int_0^1 \mathrm{d}x\,[x^2 - z^2(x-1)]\log[x^2 - z^2(x-1)]. \tag{4.A.34}$$

The same trick as before turns out to work here as well. Specifically, by factorising the polynomial $x^2 - z^2(x-1) = (x-b_+)(x-b_-)$ in terms of its roots

$$b_\pm = \frac{z^2}{2} \pm \frac{z}{2}\sqrt{z^2-4}, \tag{4.A.35}$$

we are able to write

$$\begin{aligned}z^2\mathcal{I}_2(z) &= \int_0^1 \mathrm{d}x\,(x-b_+)(x-b_-)\log(x-b_+) + (\text{terms with } b_+ \leftrightarrow b_-) \\ &= \frac{1}{36}(6b_+^2 - 6b_+ - 4 - 9z^2) + \frac{1}{6}b_+^2(3z^2 - 4b_+)\log[b_+(b_- - 1)] \\ &\quad + (\text{terms with } b_+ \leftrightarrow b_-)\end{aligned} \tag{4.A.36}$$

after using the identities $(1-b_+)(1-b_-) = 1$ and $b_+b_- = b_+ + b_- = z^2$ to simplify terms.

[3] An alternative but equivalent expression can be found in Ref. [429].

Now substituting this into (4.A.35), we eventually obtain

$$\mathcal{I}_2(z) = -1 - \frac{2}{9z^2} + \frac{z^2}{6} - \frac{z^2(z^2-6)}{6}\log z + \frac{z(z^2-4)^{3/2}}{12}\log\left(\frac{z+\sqrt{z^2-4}}{z-\sqrt{z^2-4}}\right). \tag{4.A.37}$$

As before, the principal branch should be taken when $z \in [0, 2)$. Putting everything together, the end result is

$$I_2(z) = \frac{3}{2} - \frac{z^2}{6} + \frac{z^2(z^2-6)}{6}\log z - \frac{z(z^2-4)^{3/2}}{12}\log\left(\frac{z+\sqrt{z^2-4}}{z-\sqrt{z^2-4}}\right). \tag{4.A.38}$$

Chapter 5
Constraints from precision QED

Our confidence in the accuracy of QED as the fundamental description of light and its interaction with matter is rooted in a number of high-precision experiments. To date, these so-called "precision tests of QED" have established the validity of this theory to within unprecedented error bars of less than one part per billion [6], although many are hoping for a discrepancy to emerge in the near future, as next-generation experiments continue to push for even greater precision. If detected, such deviations could be a tantalising hint of new physics, but for now, the good agreement between theory and experiment can be used to impose constraints on many scenarios beyond the Standard Model. To conclude Part I of this thesis, we will use this agreement to establish new constraints on chameleon-like theories.

Let us begin by reviewing the general method, which we first discussed in Chapter 3. Consider two experiments, X_1 and X_2, that can each provide an independent determination of the fine-structure constant α. Because quantities like frequencies can be measured to much higher precision than the elementary charge e, a value for the fine-structure constant is obtained in practice by equating the measured value of some reference quantity to the corresponding theoretical prediction from QED, and then solving the resulting equation(s) for α. For the ith experiment, let $\overline{\alpha}(X_i)$ represent the value inferred from this procedure. Then, if $\overline{\alpha}(X_1)$ and $\overline{\alpha}(X_2)$ are found to agree to within experimental and theoretical uncertainties, this constitutes strong evidence in support of QED's validity.

The two most precise determinations of the fine-structure constant to date come from measurements of transition frequencies in atomic hydrogen [402, 407, 410] and measurements of the electron's magnetic moment g_e [413, 421, 422]. The difference between these two values is [cf. (3.26) and (4.32)]

$$\overline{\alpha}(\mathrm{H}) - \overline{\alpha}(\mathrm{g}_e) = (8.2 \pm 4.9) \times 10^{-12}, \tag{5.1}$$

where the uncertainty is at the 1σ level. The two values are thus seen to agree to within 2σ.

We can now translate this agreement into a constraint on chameleon-like theories by recognising that if a scalar is active in these experiments, then the values $\overline{\alpha}(X_i)$ are biased because the additional effects from the scalar have not been taken into account. From our earlier analyses in Chapters 3 and 4, $\overline{\alpha}(X_i)$ was seen to differ from the true value α by an amount $\delta\alpha$; i.e., $\alpha = \overline{\alpha}(X_i) - \delta\alpha(X_i)$, where the systematic shifts $\delta\alpha(\mathrm{H})$ and $\delta\alpha(\mathrm{g}_e)$ are given in (3.30) and (4.34), respectively. Now requiring that these shifts do not spoil the spectacular agreement in (5.1) gives us the 2σ constraint

$$|\Delta\alpha + 8.2 \times 10^{-12}| < 9.8 \times 10^{-12}, \tag{5.2a}$$

where I have defined $\Delta\alpha = \delta\alpha(\mathrm{g}_e) - \delta\alpha(\mathrm{H})$. Written out explicitly, we have that

$$\begin{aligned}\Delta\alpha = {} & G m_e^2 \left[2\beta_m^2(\phi_0^{\mathrm{g}_e}) I_1\left(\frac{m_{\mathrm{eff}}(\phi_0^{\mathrm{g}_e})}{m_e}\right) + 4\beta_m(\phi_0^{\mathrm{g}_e})\beta_\gamma(\phi_0^{\mathrm{g}_e}) I_2\left(\frac{m_{\mathrm{eff}}(\phi_0^{\mathrm{g}_e})}{m_e}\right)\right] - \frac{\beta_m(\phi_0^{\mathrm{g}_e}) V'_{\mathrm{eff}}(\phi_0^{\mathrm{g}_e})}{4\pi M_{\mathrm{Pl}} \nu_c^2} \\ & - G m_e m_p \left[2\beta_m^2(\phi_0^{\mathrm{H}}) + 0.17\,\beta_m(\phi_0^{\mathrm{H}})\beta_\gamma(\phi_0^{\mathrm{H}})\,\overline{\alpha}(\mathrm{H})^2 - 0.032 \left(\frac{m_e}{m_p}\right)\beta_\gamma^2(\phi_0^{\mathrm{H}})\,\overline{\alpha}(\mathrm{H})^4 \right]. \end{aligned} \tag{5.2b}$$

The terms in the first line come from $\delta\alpha(\mathrm{g}_e)$ and so must be evaluated with respect to the value of the scalar field $\phi_0^{\mathrm{g}_e}$ at the centre of the Penning trap where g_e is measured. In contrast, the terms in the second line are due to $\delta\alpha(\mathrm{H})$ and so depend on the ambient field value ϕ_0^{H} in the experiment where hydrogen lines are measured. While estimates for the former were obtained in Chapter 4.3, we have yet to discuss how ϕ_0^{H} can be determined and, in fact, will not attempt to do so in this thesis because of the complicated experimental setup; see, e.g., Ref. [444]. Fortunately, such an omission will not have a negative impact on our results. The value of ϕ_0^{H} does not affect the constraints on the chameleon, since both of the coupling strengths in this model are independent of the ambient field value. (Recall that $\beta_m = M_{\mathrm{Pl}}/M_c$ and $\beta_\gamma = M_{\mathrm{Pl}}/M_\gamma$.) As for the symmetron, a conservative bound on the model parameters can still be obtained by simply assuming that the scalar always remains in its unbroken phase in the hydrogen experiment, in which case $\beta_m(\phi_0^{\mathrm{H}}) = \beta_\gamma(\phi_0^{\mathrm{H}}) = 0$.

Having clarified this issue, we now have all of the information necessary to convert (5.2) into a constraint on parameter space. Constraints for the chameleon model are presented in Section 5.1, while those for the symmetron model may be found in Section 5.2.

This page intentionally left blank

5.1 Constraints on the chameleon

The chameleon model contains four free parameters: $(n, \Lambda, M_c, M_\gamma)$. A number of two-dimensional slices in this parameter space are shown in Figure 5.1. (These slices are the most commonly used in the literature.) The shaded regions are excluded at the 95% confidence level.

From Figure 5.1(a), we see that the constraints imposed on the chameleon by precision tests of QED are driven predominantly by $\delta\alpha(\mathrm{H})$, except in a region around $(M_c, \Lambda) \sim (10^{-10} M_{\mathrm{Pl}}, 10^3 \text{ eV})$, where the cavity shift term in $\delta\alpha(\mathrm{g}_e)$ dominates. Also drawn in this figure are two dotted lines, which demarcate the region of parameter space beyond which the perfect-vacuum (PV) and zero-skin-depth (ZSD) approximations break down. These lines correspond to points that saturate the inequalities in (4.39) and (4.40), respectively. Above the ZSD line, our estimate in (4.38) for the central field value $\phi_0^{\mathrm{g}_e}$ in the Penning trap fails to be reliable, and so no constraints can be established in this region. Below the PV line, $\phi_0^{\mathrm{g}_e}$ again cannot be determined reliably, but the constraints from hydrogen are independent of this quantity and so continue to remain valid. These two dotted lines aside, there is also a solid line drawn in Figure 5.1(a), above which the EFT itself is expected to break down. As an order-of-magnitude estimate, the EFT is assumed to hold provided $\phi_0^{\mathrm{g}_e} < \min(M_c, M_\gamma)$. These boundaries continue to be in effect in the remaining panels of Figure 5.1, although the solid and dotted lines have been omitted to avoid overcluttering.

The constraints from precision tests of QED are presented alongside those from a selection of other experiments in Figures 5.1(c) and 5.1(d). Apart from the collider constraints [445], which are self-explanatory, a brief description of each experiment can be found in Chapter 1. (Further details can also be found in the review by Burrage and Sakstein [65].) Note also that the constraint from "astrophysics" combines results from galactic rotation curves [79] and Cepheid/TRGB distance indicators [90]. As was the case in Figure 5.1(a), the EFT is expected to break down in the top left of Figure 5.1(c), but it is difficult to demarcate where exactly this occurs in a universal way. For this reason, the shaded regions excluded by other experiments are drawn as they appear in the original papers, but one should keep in mind that the EFT is not valid for sufficiently large values of Λ and small values of M_c.

From these plots, we see that the constraints imposed by precision tests of QED are not the best in class, but they are still competitive with other experiments in certain regions of parameter space. This is quite remarkable, given that these precision tests were not designed to search for small deviations from Newtonian gravity. Their ability to nevertheless establish meaningful constraints on chameleon models is a testament to the incredible levels of precision that experimentalists have been able to achieve.

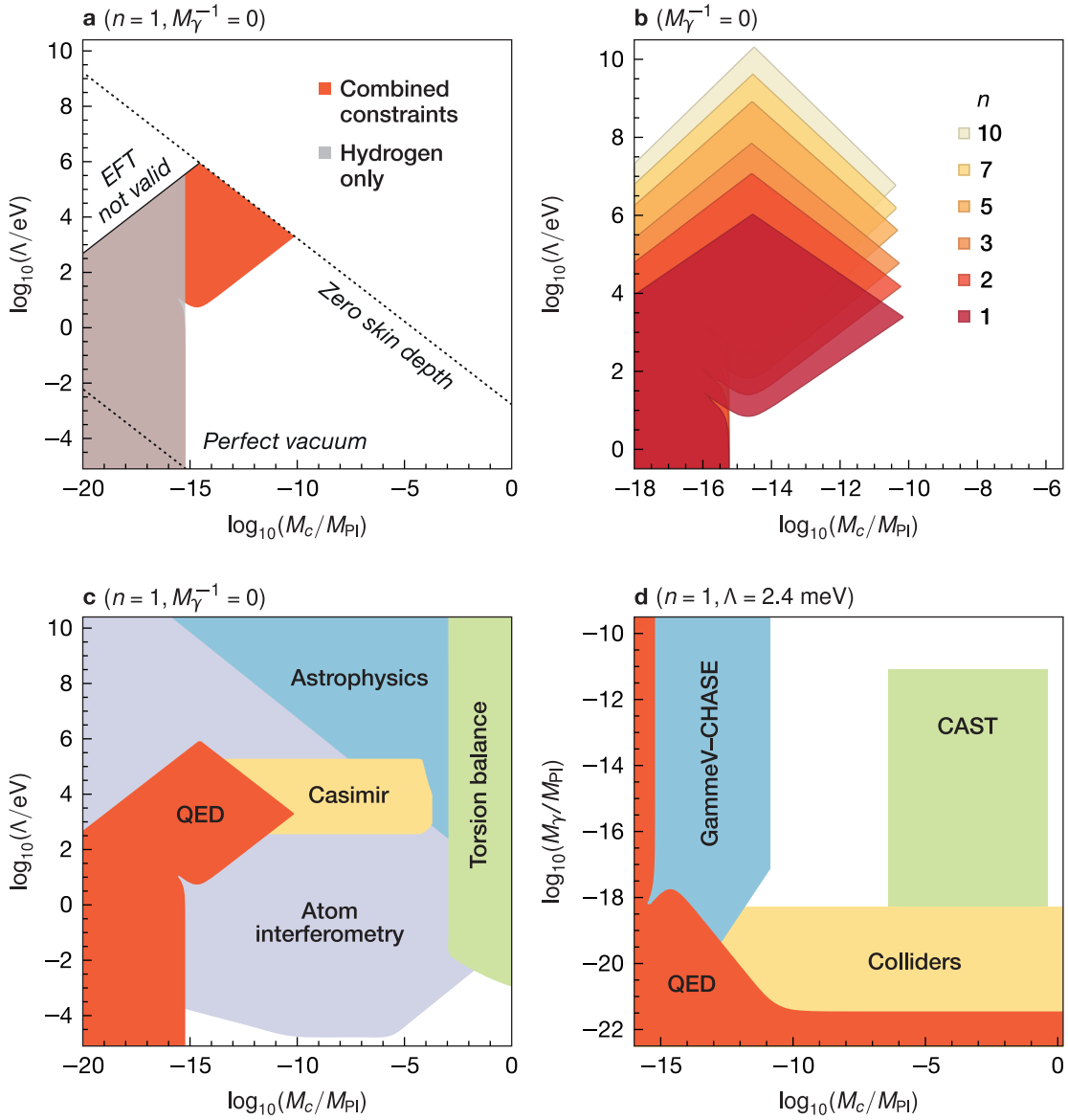


Figure 5.1 Constraints on chameleon models from precision tests of QED. The shaded regions are excluded at the 95% confidence level. (a) Constraints obtained from a combined analysis of hydrogen lines and the electron's magnetic moment are shown in orange, whereas the constraints driven by hydrogen lines only is shown in grey. Numerical limitations mean that the cavity shift can be computed reliably only when both the zero-skin-depth and perfect-vacuum approximations are valid (see Chapter 4.3 for details). This corresponds to the region in between the two dotted lines. Above this region, no constraint can be reliably established, whereas below this region, the constraint is dominated by the contribution from hydrogen lines. (b) Excluded regions in parameter space for different values of n have similar shapes. The cusps stem from the relative sign between the first and second lines of (5.2b). (c,d) The constraining power of precision QED compared with a number of other experiments.

5.2 Constraints on the symmetron

The symmetron model also contains four free parameters: $(\mu, \lambda, M_s, M_\gamma)$. Figures 5.2 to 5.4 present a number of two-dimensional slices in this parameter space, where shaded regions are excluded at the 95% confidence level. As we are assuming that $\phi_0^{\rm H} = 0$ for simplicity, these constraints derive purely from bounding the symmetron's effects on the value of g_e.

In Figure 5.2(a), the contributions from the different terms in $\delta\alpha(g_e)$ are drawn separately. The constraints derived from bounding only the scalar's quantum corrections to the intrinsic value of g_e are shaded in blue, whereas those that come from putting an upper limit on the size of the cavity shift are shaded in grey. These two effects were discussed in Chapters 4.1 and 4.2, respectively, and correspond to the first and second terms in (5.2b). This decomposition serves to draw attention to a number of different boundaries that demarcate the regions in which our approximations are valid.

Both effects require as input a value $\phi_0^{g_e}$ for the scalar field at the centre of the Penning trap. To discuss how this value changes as we move around in parameter space, it will be convenient to recall the definition of the mass scale μ_0, which depends on three of the symmetron's four parameters according to the formula $\mu_0^2 = \mu^2 - \rho_0 M_s^{-2} - \rho_{\rm em} M_\gamma^{-2}$. Given specifics about the trap's dimensions and upper bounds on the gas and electromagnetic energy densities in its interior (i.e., ρ_0 and $\rho_{\rm em}$), we showed in Chapter 4.3 that $\phi_0^{g_e} = 0$ and, consequently, $\delta\alpha(g_e) = 0$ when $\mu_0 \lesssim 10^{-3.88}$ eV. Accordingly, no constraint can be established in this region of parameter space. For larger values of μ_0 in the range $[10^{-3.88}, 10^{-3.39}]$ eV, the symmetron's central field value $\phi_0^{g_e}$ is determined from the semi-empirical formula in (4.42), while for even larger values of μ_0, $\phi_0^{g_e}$ is fixed at $\mu_0/\sqrt{\lambda}$, seeing as the symmetron can now minimise its effective potential in the cavity. (When the symmetron spontaneously breaks its $\mathbb{Z}_2$ symmetry, the positive branch can always be chosen without loss of generality.)

The values for $\phi_0^{g_e}$ at small and large values of μ_0 are robust, but those that depend on the formula in (4.42) rely on the validity of the zero-skin-depth (ZSD) approximation. Consequently, within the intermediate range $\mu_0 \in [10^{-3.88}, 10^{-3.39}]$ eV, no constraint can be reliably established whenever the ZSD approximation breaks down. The boundary at which this occurs is drawn as a dotted line in Figure 5.2(a), and corresponds to points in parameter space that saturate the inequality in (4.43). Also drawn in Figure 5.2(a) is a solid line below which the EFT is expected to break down. As an order-of-magnitude estimate, the EFT is assumed to hold as long as $\max(\phi_0^{g_e}) < \min(M_s, M_\gamma)$, where $\max(\phi_0^{g_e})$ is the maximum possible value that the scalar can attain within the Penning trap. This is simply the value that minimises the symmetron's bare potential; i.e., $\max(\phi_0^{g_e}) = \mu/\sqrt{\lambda}$.

These two boundaries apply to calculations for both the cavity shift and the quantum corrections, but the validity of the latter is subject to an additional two requirements. First, the perturbative approach in

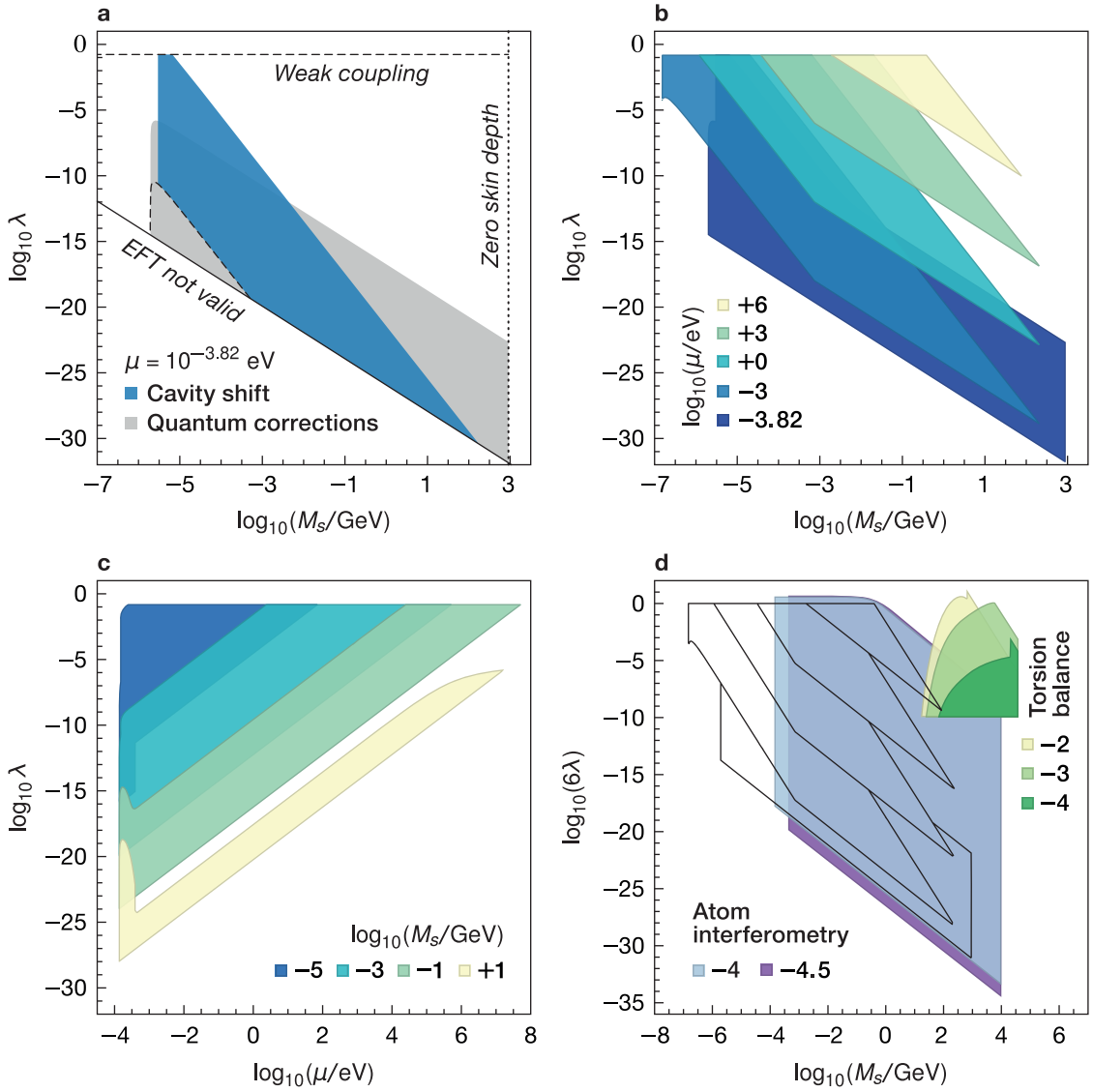


Figure 5.2 Constraints on symmetron models from precision tests of QED in the three-dimensional subspace where $M_\gamma^{-1} = 0$. The shaded regions are excluded at the 95% confidence level. (a) Constraints due to the cavity shift (grey) and quantum corrections (blue) are shown separately for an illustrative value of $\mu = 10^{-3.82}\,\text{eV}$. Numerical limitations mean that these constraints can be computed reliably only when the zero-skin-depth approximation is valid (see Chapter 4.3 for details). This corresponds to the region to the left of the dotted line. Meanwhile, the quantum correction terms are valid only in the weak-coupling regime, corresponding to the region sandwiched between the two dashed lines. Finally, no constraints can be established for sufficiently small values of λ, corresponding to the region below the solid line, where the EFT itself becomes unworkable. (b,c) Constraints from the combined analysis of the cavity shift and quantum corrections are shown for different slices in the M_s–λ and μ–λ planes. (d) Outline of the region excluded by precision tests of QED when compared with the constraining power of a number of other experiments.

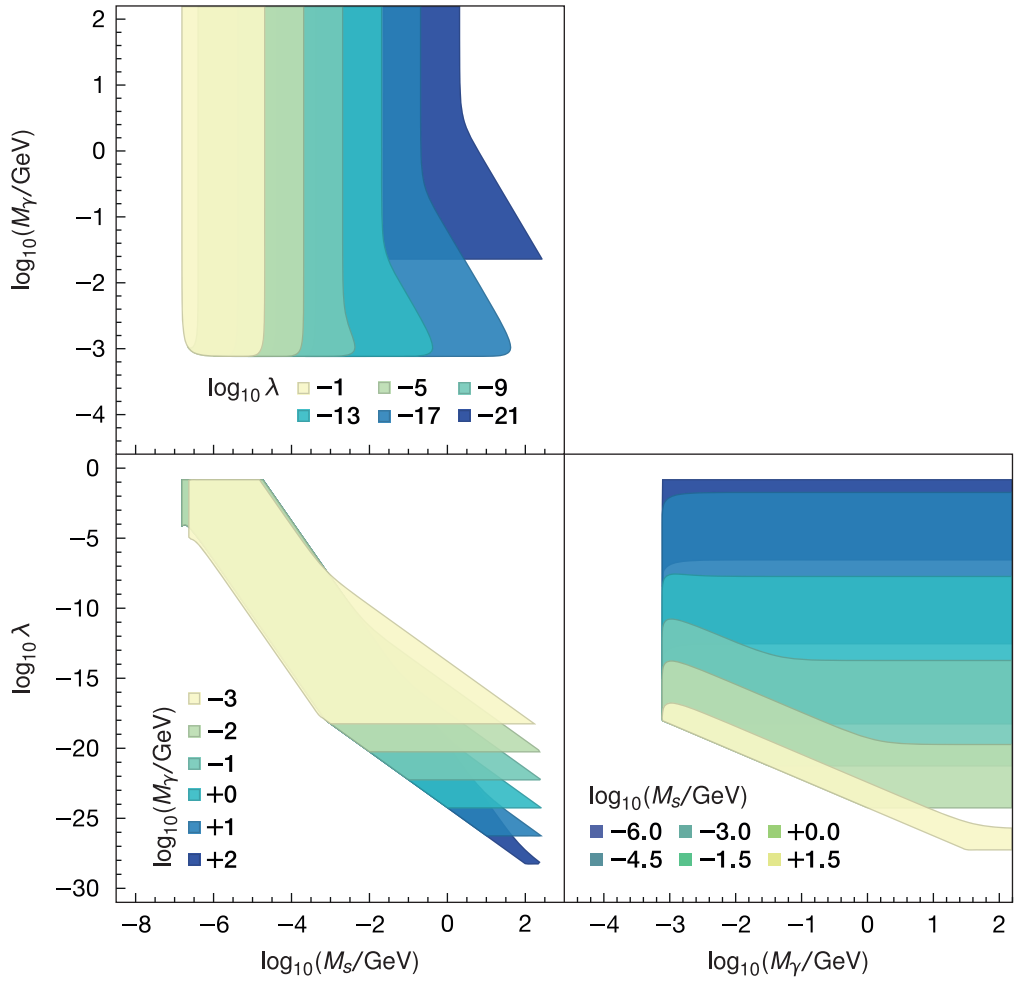


Figure 5.3 Constraints on symmetron models from precision tests of QED in the three-dimensional subspace where $\mu = 10^{-3}$ eV. Top, front, and side views are shown to aid with visualisation. Shaded regions are excluded at the 95% confidence level.

terms of Feynman diagrams that we adopted in Chapter 4.1 is valid only in the weak-coupling regime. This is when the symmetron's couplings to the Standard Model are such that $\beta_m(\phi_0^{ge})\, m_e/M_{\rm Pl} \ll 1$ and $\beta_\gamma(\phi_0^{ge})\, m_e/M_{\rm Pl} \ll 1$, and when the strength of its quartic self-interaction is such that $\lambda \ll 6$.[1] The loci along which these inequalities are saturated are drawn as dashed lines in Figure 5.2(a). Second, the validity of these loop calculations is also contingent on the effective mass of the symmetron excitation being real; i.e., $m_0^2 := m_{\rm eff}^2(\phi_0^{ge}) > 0$. It is worth clarifying that, in this case, the onset of a tachyonic instability is not signalling any kind of pathological behaviour, but merely indicates that another approximation we made in Chapter 4.1 has broken down. Specifically, when $m_0^2 < 0$, we are no longer allowed to neglect spatial variations in the classical background field $\langle\phi\rangle$ when computing loop corrections. A calculation that accounts for these spatial variations is beyond the scope of this thesis,

[1] The unusual factor of 6 stems from the fact this interaction term is normalised as $\lambda/4$ instead of $\lambda/(4!)$.

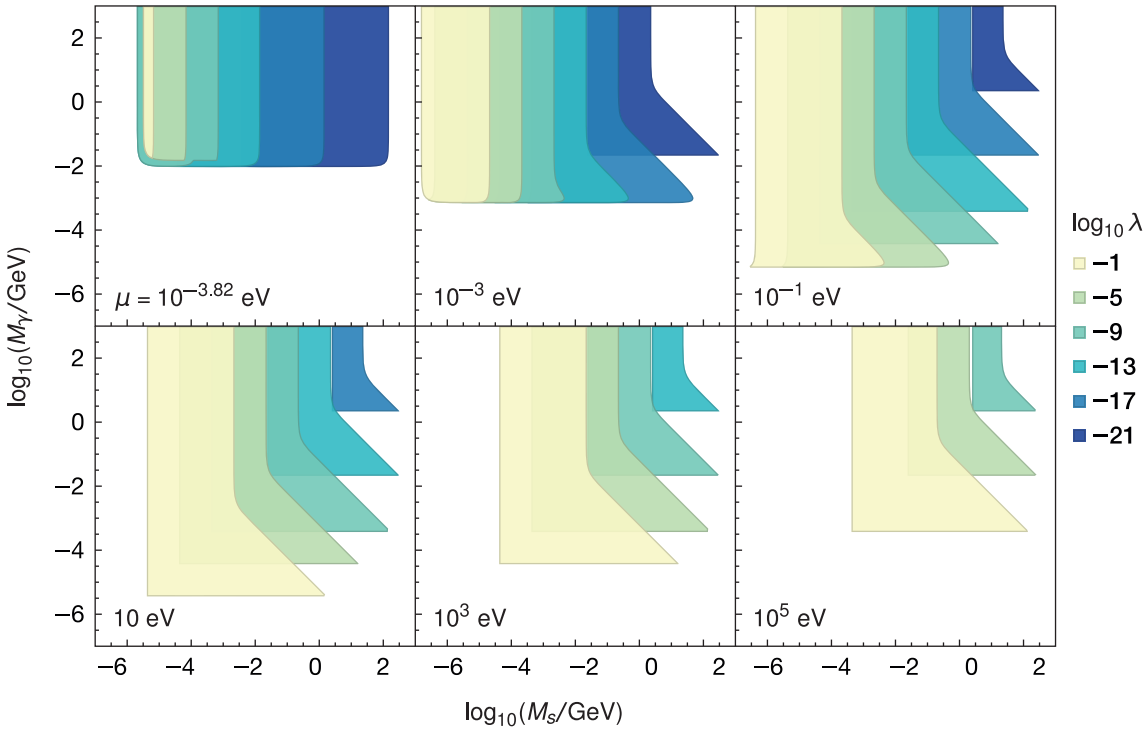


Figure 5.4 Constraints on symmetron models from precision tests of QED. Top-view exclusion plots are shown for a number of three-dimensional subspaces with fixed μ. Shaded regions in the M_s–M_γ plane are excluded at the 95% confidence level for different values of λ.

however, and so we shall simply forgo using the quantum correction terms to establish constraints whenever $m_0^2 < 0$. Luckily, the volume of parameter space in which this tachyonic instability is active is not substantial. No line has been drawn in Figure 5.1(a) to demarcate this region, but its boundary can easily be seen by noting that the blue shaded region does not extend as far to the left as the grey region. (The left edge of the grey region corresponds to the threshold for spontaneous symmetry breaking.) All of these different boundaries continue to be in effect in the remaining panels in Figures 5.2 to 5.4, although the solid, dashed, and dotted lines have been omitted to avoid overcluttering.

One more feature about these exclusion plots is worth highlighting. First, by substituting the explicit expressions for $\beta_m(\phi)$, $\beta_\gamma(\phi)$, and $V_{\mathrm{eff}}(\phi)$ from Chapter 2 into (5.2b), one finds that the cavity shift term is proportional to $(\hat{\phi}_0^{ge})^2[1-(\hat{\phi}_0^{ge})^2]$, and thus switches off whenever $\hat{\phi}_0^{ge} := \phi_0^{ge}/(\mu_0/\sqrt{\lambda}) = 0$ or 1. In terms of the mass scale μ_0, this means that the vacuum bubble exerts an appreciable force on the electron only in the intermediate range $\mu_0 \in [10^{-3.88}, 10^{-3.39}]$ eV. Physically, the lower end of this range corresponds to the threshold for spontaneous symmetry breaking, whereas at the higher end, the local Compton wavelength of the symmetron becomes increasingly small; resulting in a bubble profile that is essentially flat everywhere in the cavity except near the walls. This sensitive dependence on μ_0 is the reason why other laboratory experiments have hitherto left the symmetron parameter space mostly

unexplored. As shown in Figure 5.2(d), atom interferometry experiments [97, 98] place meaningful bounds only in the range $\mu \in [10^{-5}, 10^{-4}]$ eV, whereas an analysis of torsion pendulums [92] has so far constrained the limited range $[10^{-4}, 10^{-2}]$ eV.

Precision QED experiments do not suffer from this same limitation, however. Although the cavity shift turns off at larger values of μ_0, the scalar's quantum corrections to g_e continue to be useful for establishing strong constraints on the model parameters. Indeed, the results presented in this chapter have succeeded in excluding large swathes of the symmetron's parameter space that were previously inaccessible. This success can be seen most clearly in Figure 5.2(c). The "hump" in the bottom left is the constraint imposed by the cavity shift, while the rest of the excluded region, which extends up to $\mu \sim 10^8$ eV, is the result of accounting for quantum corrections. As a final remark, it is also worth noting that the results presented in Figures 5.3 and 5.4 are the first to establish constraints on the symmetron's coupling to photons.

PART II

Black holes

Chapter 6
Inducing black hole hair

Through his work on the renormalisation group [446], Wilson reshaped the way we think about physics. These days, theories that are UV incomplete are no longer seen as pathological, but prove to be indispensable for performing calculations. By recognising that what transpires on microscopic scales affects the outcomes of macroscopic processes only through the coefficients of local operators, these effective field theories allow us to focus on just the most relevant degrees of freedom in any given problem. Recall that we already exploited this idea in Part I when we wrote down models for chameleon-like theories without needing to first possess knowledge about the underlying theory of quantum gravity. Whatever occurs on high-energy scales imprints itself onto the low-energy theory only through the parameters of the model. Now leaving the realm of modified gravity behind, this idea will take centre stage here in Part II as we seek to understand how binary black holes in general relativity interact with ambient, minimally coupled scalar fields.

Before launching into any of the technical details, it is worth painting a general picture of the problem at hand. Let us start by enumerating the different scales in this problem. First, each of the black holes has a characteristic size $\sim GM$, which sets the scale at which strong-gravity effects become important. (For the purposes of this discussion, it will be convenient to assume that both black holes are of comparable mass, although we will generalise to arbitrary mass ratios in later chapters.) The fact that there are two black holes in this system introduces two more scales: the orbital separation a and the orbital frequency Ω, which together characterise the properties of the orbit. Their product defines the typical velocity v of the black holes, and the orbital frequency Ω also sets the energy scale associated with gravitational radiation. If this binary is not isolated in pure vacuum but is surrounded by an ambient scalar field, then (at least) two more scales are relevant: the timescale ω^{-1} on which the field evolves, and the length scale λ over which it varies.

Needless to say, the dynamics of this system are fully captured by the equations of motion associated with the action[1]

$$S[g,\phi] = \int \mathrm{d}^4x\,\sqrt{-g}\left(2m_{\rm Pl}^2 R - \frac{1}{2}(\partial\phi)^2 - V(\phi)\right), \tag{6.1}$$

but confronting these equations head on to track its evolution from inspiral to merger is completely untenable. For this reason, it is necessary that we break the problem up into multiple parts. During the later stages of the inspiral, when $GM \sim a$, numerical relativity methods prove to be the only recourse, given the highly dynamical and strong-field nature of spacetime in this regime. We will have nothing to say about this regime in this thesis, and will instead focus on the early inspiral, when $GM \ll a$. Moreover, as we discussed in Chapter 1, we shall assume that the characteristic length scale λ of the ambient scalar is such that

$$GM \ll a \ll \lambda. \tag{6.2}$$

Working in this "long-wavelength limit" allows us to separate the problem into three distinct zones (see Figure 6.1), each of which is amenable to well-established approximation methods. For each black hole, the region in the immediate vicinity of its horizon is known as the *inner zone* (often also called the near-horizon regime). Here, the spacetime curvature is dominated by the black hole's own self-gravity, while the ambient scalar and the gravitational field sourced by its companion can be treated as weak perturbations about an otherwise stationary Kerr metric. Zooming out, the *near zone* is the regime in which the orbital separation is the most important scale. Having coarse-grained away the inner zones to put ourselves into this regime, each black hole now behaves like an effective point particle whose velocity v is related to the orbital separation a via the virial theorem, $v^2 \sim GM/a$. Consequently, the orbit is nonrelativistic when (6.2) holds, and can thus be solved perturbatively by employing a post-Newtonian approximation [447]. Finally, zooming out even further puts us in the *far zone*, where the binary as a whole behaves like an effective point particle that couples to the scalar and gravitational fields via a set of multipole moments. In this regime, the system is well described by a multipolar post-Minkowskian expansion [448].

The key question now is how to tie these different zones together. We know, for example, that the emission of radiation in the far zone prompts a concomitant decay of the orbit in the near zone and, likewise, absorption of the fields in the inner zone also elicits a decay of the orbit in the near zone. Thus, a complete and self-consistent description of how this system evolves as a whole requires not only that we can solve for its behaviour in each of the different zones, but that we are also able to connect how

[1]Recall that we define $m_{\rm Pl}^2 = 1/(32\pi G)$ in this part of the thesis.

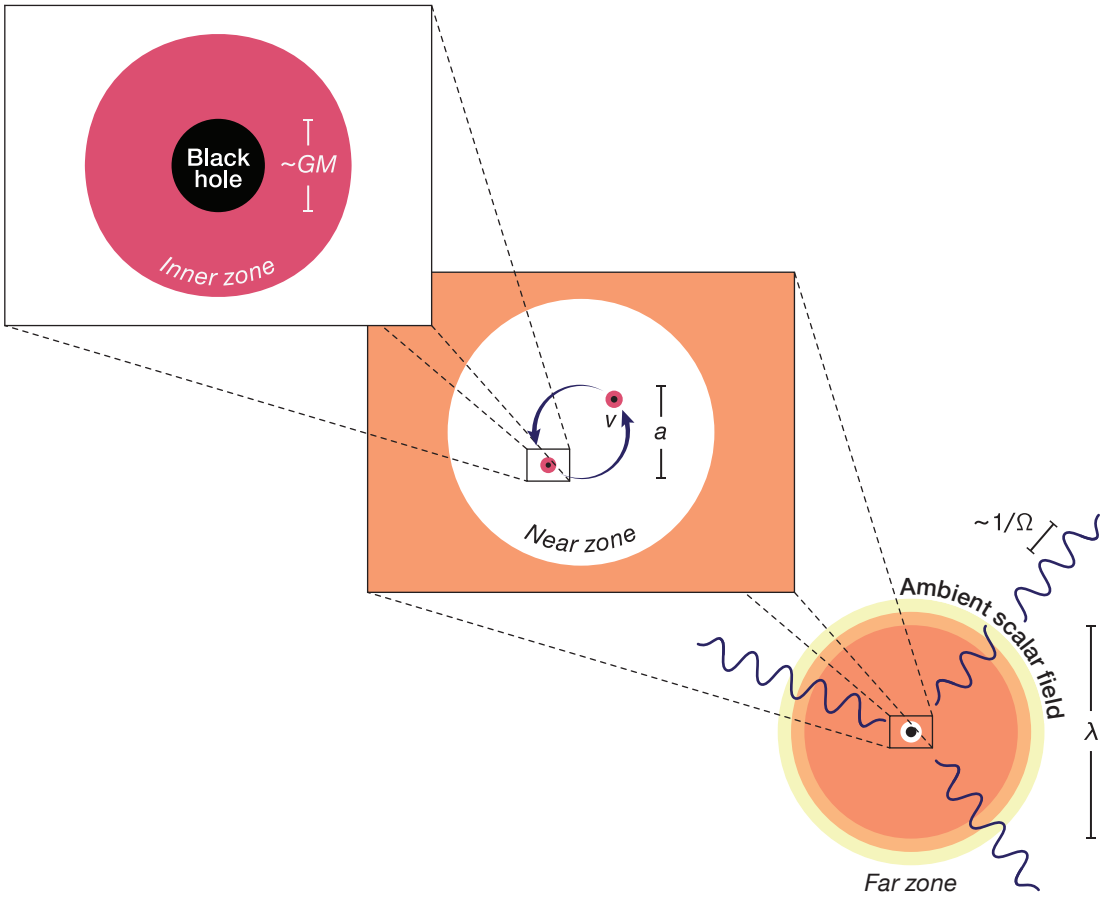


Figure 6.1 Cartoon (not to scale) illustrating the division of a binary black hole spacetime into multiple zones when large separations of scales are present. The inner zones are the regions in the immediate vicinity of each black hole, where spacetime is well described by a perturbed Kerr metric. A post-Newtonian expansion of the spacetime holds in the near zone, where the black holes can be approximated as worldlines that are mutually orbiting one another. Finally, a post-Minkowskian expansion of spacetime is used in the far zone, where the binary as a whole behaves like an effective point particle that is surrounded by radiation and the ambient scalar field.

physical processes in one zone affect those in another. With that in mind, our approach in this thesis will be to construct a tower of effective field theories — one for each zone — that we will then match onto one another à la Goldberger and Rothstein [377–380].

The remainder of this chapter will be concerned with physics in the inner zone. To set the stage, Section 6.1 presents a brief overview of the Schwarzschild and Kerr solutions of general relativity, as well as a proof of the no-hair theorem appropriate to (6.1), which asserts that isolated black holes in these theories are also described by the Kerr metric, and so do not possess any *permanent* scalar charges. The assumptions that go into this theorem — namely, that of stationarity — bear little resemblance to the kinds of systems that we are ultimately interested in, however, and so do not preclude the possibility that a black hole may dynamically develop an *induced* scalar charge as a result of its time-dependent surroundings. With that said, the no-hair theorem continues to be valuable even in these time-dependent cases, for which no exact solution is known, because it guarantees that the stationary Kerr metric is the right starting point about which to construct a perturbative solution. In Section 6.2, we will perform exactly this kind of perturbative calculation to show that a black hole can indeed gain an induced scalar charge (alongside higher multipole moments, more generally) when embedded in an ambient, time-dependent scalar-field profile. These results, which are valid only in the inner zone, will then be matched onto appropriate near- and far-zone EFTs in subsequent chapters.

6.1 Black holes in general relativity

6.1.1 The Schwarzschild solution

The simplest black hole solution to the vacuum Einstein field equations is the Schwarzschild metric,

$$g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu = -\left(1-\frac{2GM}{r}\right)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{1-2GM/r} + r^2(\mathrm{d}\vartheta^2 + \sin^2\vartheta\,\mathrm{d}\chi^2). \tag{6.3}$$

In these coordinates, the spherical symmetry of this spacetime is manifest, as each surface of constant t and r has an induced metric proportional to that of the unit 2-sphere. (To avoid confusion with the scalar fields ϕ and φ, the azimuthal angle on the 2-sphere is denoted by χ.) Algebraically speaking, this symmetry can be associated with the vanishing of the metric's Lie derivatives along the integral curves of three spacelike vectors,

$$m = \partial_\chi, \qquad \ell_1 = \sin\chi\,\partial_\vartheta + \cot\vartheta\cos\chi\,\partial_\chi, \qquad \ell_2 = \cos\chi\,\partial_\vartheta - \cot\vartheta\sin\chi\,\partial_\chi, \tag{6.4}$$

which form a basis for the infinitesimal generators of rotation. Since a vector ξ^μ satisfies the Killing

equation $\nabla_{(\mu}\xi_{\nu)} = 0$ when the Lie derivative $(\mathcal{L}_\xi g)_{\mu\nu} = 0$, these generators are said to be Killing vectors. The Schwarzschild spacetime can readily be seen to admit another Killing vector, $\mathbb{k} = \partial_t$, since it is also invariant under translations of the coordinate time t. Moreover, because hypersurfaces of constant t are orthogonal to $\mathbb{k}$, which is timelike for $r > 2GM$, this spacetime is said to be *static*.

It is also asymptotically flat in the sense that its metric approaches that of Minkowski in the limit $r \to \infty$; meaning test particles in this region experience only weak gravitational effects. Indeed, the geodesic equation for timelike curves reduces to the Newtonian equation of motion $\mathrm{d}^2\mathbf{x}^i/\mathrm{d}t^2 \approx -\partial^i \Phi_\mathrm{N}$ with $\Phi_\mathrm{N} = -GM/r$ at large distances; hence, we may attach to the parameter M, which arises as an integration constant upon solving the vacuum field equations $R_{\mu\nu} = 0$, the interpretation of being the black hole's mass. (A derivation can be found in Chapter 6 of Wald [389].)

As it is written in (6.3), the Schwarzschild metric is singular at $r = 0$, $r = 2GM$, and at $\vartheta = 0$ and π; the last two being trivial singularities associated with working in polar coordinates. The singularity along the Schwarzschild radius $r = 2GM$, as it turns out, is also a coordinate artefact. To see this, swap out the coordinate time t in favour of the null coordinate [449]

$$v = t + r_*(r), \qquad r_*(r) = \int \frac{\mathrm{d}r}{1 - 2GM/r} \tag{6.5}$$

to find that (6.3) becomes

$$g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu = -\left(1 - \frac{2GM}{r}\right)\mathrm{d}v^2 + 2\,\mathrm{d}v\mathrm{d}r + r^2(\mathrm{d}\vartheta^2 + \sin^2\vartheta\,\mathrm{d}\chi^2). \tag{6.6}$$

Notice that when written in these (ingoing) Eddington–Finkelstein coordinates, the Schwarzschild metric is no longer singular at $r = 2GM$. Thus, while (6.3) was valid only in the region $r > 2GM$ (or equivalently, only in the region $0 < r < 2GM$), the metric in (6.6) is now well defined for all positive values of the radial coordinate r.

This is not to say that the surface at $r = 2GM$ is unremarkable. On the contrary, this null hypersurface acts as a one-way membrane that allows things to flow from the exterior region ($r > 2GM$) into the interior region ($r < 2GM$), but not vice versa. We can see this for ourselves by considering the special case of radial null geodesics; i.e., geodesics whose tangent vectors $\dot{x}^\mu = \mathrm{d}x^\mu/\mathrm{d}\sigma$ satisfy the constraints $\dot{\vartheta} = \dot{\chi} = 0$ and $\mathcal{L} := \dot{x}_\mu \dot{x}^\mu = 0$. The presence of the Killing vector $\mathbb{k}$ on this spacetime defines for us another constraint, $\dot{x}^\mu \nabla_\mu \mathcal{E} = 0$ with $\mathcal{E} := -\mathbb{k}_\mu \dot{x}^\mu$, and we may further set $\mathcal{E} = 1$ without loss of generality by exploiting our freedom to rescale the affine parameter σ. Taken together, $\mathcal{L}$ and $\mathcal{E}$ provide us with enough information that we need not solve the geodesic equation directly.

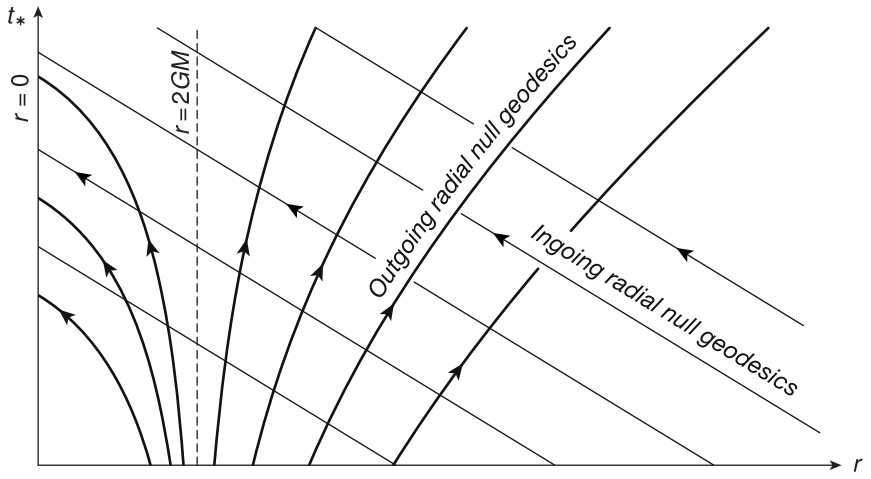


Figure 6.2 Ingoing and outgoing radial null geodesics on the Schwarzschild spacetime.

Working in Eddington–Finkelstein coordinates and writing $X \equiv 1 - 2GM/r$, we have that

$$\mathcal{L} = -X\dot{v}^2 + 2\dot{v}\dot{r} = 0, \qquad \mathcal{E} = X\dot{v} - \dot{r} = 1. \tag{6.7}$$

These equations admit three, qualitatively distinct sets of solutions: the "horizon generators" satisfy $(\dot{v}, \dot{r}) = (1, 0)$ and remain on the hypersurface $r = 2GM$, the ingoing radial null geodesics satisfy $(\dot{v}, \dot{r}) = (0, -1)$ and correspond to curves of constant v, while the outgoing radial null geodesics satisfy $(\dot{v}, \dot{r}) = (2/X, 1)$ and correspond to curves of constant $u = v - 2r_*$.

We can get a sense for the causal structure of this spacetime by drawing lines of constant u and v on a Finkelstein diagram as in Figure 6.2, where $t_* = v - r$ is a convenient time coordinate that puts the ingoing radial null geodesics at a constant 45° angle. Note that each point on this t_*–r plane corresponds to a 2-sphere with area $4\pi r^2$, and that all arrows in this diagram point towards the future.[2] Clearly, Figure 6.2 demonstrates that radial null geodesics can flow into the interior region of the Schwarzschild spacetime, but cannot flow out. This statement holds also for more general causal curves [449, 450]; that is to say: no future-directed, causal curve can emanate from the region $r \leq 2GM$. This region is what we call the black hole, and its boundary at $r = 2GM$ is known as the (future) event horizon.

[2] In the exterior region, the timelike vector k ($= \partial_v$ in Eddington–Finkelstein coordinates) defines for us a time orientation, such that any causal vector U is said to be future-directed if it is in the same lightcone as k (i.e., $U_\mu k^\mu < 0$) and is past-directed otherwise. The notion of a future direction can be extended into the interior region, where k becomes spacelike, by instead using $\mathfrak{z} = -\partial_r$ as our time orientation. This vector is globally null and future-directed, since $\mathfrak{z}_\mu k^\mu < 0$ when $r > 2GM$. It follows that the tangent vector $\dot{x}^\mu_{\text{in}} = (0, -1, 0, 0)$ for the ingoing radial null geodesics is always future-directed, but the tangent vector $\dot{x}^\mu_{\text{out}} = (2/X, 1, 0, 0)$ for the "outgoing" ones is future-directed only in the exterior region, while it is past-directed in the interior region. Accordingly, the arrows in Figure 6.2 for the set of curves $\{u = \text{const.}, r < 2GM\}$ have been drawn such that they point along the $-\dot{x}^\mu_{\text{out}}$ direction.

Once inside the black hole, any future-directed, causal curve reaches $r = 0$ within a finite affine parameter distance, but in contrast to what we found for the event horizon, there is no coordinate transformation that can extend this spacetime into the region $r \leq 0$. Indeed, the curvature invariant $R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma} = 48(GM)^2/r^6$ diverges as we approach $r \to 0$; hence, the (spacelike) surface at $r = 0$ is a genuine physical singularity, where gravitational forces become infinite and the classical laws of general relativity as we know them break down. While we might expect that such singularities will eventually be resolved by quantum-gravity effects, it is reassuring to note that the event horizon shrouds the pathological nature of the Schwarzschild spacetime in this region from view of distant observers like ourselves, who live in the exterior. Thus, in line with the principles of effective field theory, we can make reliable predictions about how black holes behave without first needing a complete quantum theory of gravity, as long as we restrict our attention to the region of spacetime accessible to us.

In light of this, it is worth returning to the metric in (6.3) and asking what would happen if the parameter M were negative. On purely mathematical grounds, this is still a valid solution to the Einstein field equations, but because the components $-g_{tt} = 1/g_{rr} = 1 + 2G|M|/r$ are now regular for all $r > 0$, this metric describes the spacetime around a curvature singularity at $r = 0$ that is not hidden behind an event horizon. Such "naked singularities" are considered unphysical, and roughly speaking, the weak cosmic censorship conjecture asserts that they should never form dynamically from the gravitational collapse of physically realistic matter [321, 451, 452].

6.1.2 The Kerr solution

Schwarzschild's solution is only the simplest example of a spacetime containing a black hole. In more realistic scenarios, those that form dynamically from the gravitational collapse of massive stars are expected to inherit at least some amount of angular momentum from their progenitors, and so must eventually settle down to a spacetime that is rotating but otherwise time-independent. The Kerr metric corresponds to exactly this scenario. Setting $G = 1$ in the rest of this chapter for simplicity, this metric reads

$$g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}t^2 + \frac{2Mr}{\Sigma}(\mathrm{d}t - s\sin^2\vartheta\,\mathrm{d}\chi)^2 + \Sigma\left(\frac{\mathrm{d}r^2}{\Delta} + \mathrm{d}\vartheta^2\right) + (r^2 + s^2)\sin^2\vartheta\,\mathrm{d}\chi^2 \quad (6.8)$$

when written in Boyer–Lindquist coordinates, with $\Sigma = r^2 + s^2\cos^2\vartheta$ and $\Delta = r^2 - 2Mr + s^2$.

Notice, first of all, that the above metric describes a two-parameter family of solutions that reduces to Schwarzschild's in the limit $s = 0$. For more general values of s, an analysis of timelike geodesics in the weak-field region ($r \gg M$) of this asymptotically flat spacetime will tell us that the constant M is

still the black hole's mass, while $|Ms|$ is the magnitude of its angular momentum, which points along the $\vartheta = 0$ direction when $s > 0$. One can again identify $k = \partial_t$ and $m = \partial_\chi$ as Killing vectors on this spacetime, which are associated with time-translation invariance and axial symmetry, respectively. Note, however, that because k is not orthogonal to surfaces of constant t (or, more generally, because k_μ cannot be written in the form of a gradient), this spacetime is said to be *stationary*, but not static.

Setting $s^2 > M^2$ gives rise to a naked singularity; hence, there is an upper limit to the spin of this black hole. Let us discuss the subextremal ($s^2 < M^2$) case first. Aside from the usual, trivial singularities at $\vartheta = 0$ and π associated with working in polar coordinates, the metric in (6.8) is singular whenever Δ or Σ vanishes. The quadratic function Δ admits two roots, $r_\pm = M \pm (M^2 - s^2)^{1/2}$, whose level sets are null hypersurfaces known as the inner and outer horizons ($r = r_-$ and $r = r_+$, respectively). Both of these values of r give rise to mere coordinate singularities, as can be seen by transforming to Kerr coordinates $(v, r, \vartheta, \tilde{\chi})$, where [449]

$$\mathrm{d}v = \mathrm{d}t + \frac{r^2 + s^2}{\Delta}\,\mathrm{d}r, \qquad \mathrm{d}\tilde{\chi} = \mathrm{d}\chi + \frac{s}{\Delta}\,\mathrm{d}r. \tag{6.9}$$

This is the analogue of transforming to Eddington–Finkelstein coordinates in the Schwarzschild case. One can now show that the metric in these new coordinates is regular for all $r > 0$. Moreover, it is possible to further extend this spacetime through the ring-like curvature singularity at $(r, \vartheta) = (0, \pi/2)$, along which Σ vanishes, and into a new, asymptotically flat region where $-\infty < r < 0$.

The extension of the extremal Kerr spacetime ($s^2 = M^2$) proceeds in a similar fashion, although the interior region is qualitatively quite distinct, as the roots r_+ and r_- now coincide. That said, such details will not be important here, as only the exterior region, wherein $r > r_+$, is accessible to distant observers like ourselves. For both the subextremal and extremal cases, charting their causal structure reveals that the surface at $r = r_+$ is indeed the black hole's event horizon.

6.1.3 The no-hair theorem

This brief introduction pays short shrift to many of the interesting features that black hole spacetimes exhibit (the reader may wish to consult Refs. [381, 389, 449, 450] for additional reading), but what is perhaps their most remarkable property is worth discussing in some detail here. As it turns out, the Kerr metric is the only black hole solution to the four-dimensional, vacuum Einstein field equations that is stationary, analytic, and asymptotically flat [163, 164]. Furthermore, this uniqueness extends also to more general field theories, like those described by the action in (6.1). The proof is simple enough to sketch, and so we will conclude this section by showing that stationary, analytic, asymptotically flat

black holes in this class of theories are identical to those in general relativity. (The following arguments closely mirror the original proofs by Hawking [177] and Sotiriou and Faraoni [181].)

Suppose that the field equations descending from the action in (6.1), which read

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi T^{(\phi)}_{\mu\nu}, \tag{6.10a}$$

$$\nabla^\mu \nabla_\mu \phi - V'(\phi) = 0, \tag{6.10b}$$

admit a black hole solution that satisfies the following four properties. First, it should be asymptotically flat, such that the metric $g_{\mu\nu}$ tends to that of Minkowski at large distances, while the scalar ϕ tends to some constant value ϕ_0. This limiting behaviour is consistent with (6.10) only if ϕ_0 is such that $V(\phi_0) = V'(\phi_0) = 0$. Second, assume that this solution is stationary and so admits a Killing vector k that is timelike in the neighbourhood of (null) infinity. Third, assume that it is analytic, and fourth, that the scalar's potential $V(\phi)$ is such that its energy–momentum tensor [cf. (2.7)] satisfies the weak energy condition; i.e., $T^{(\phi)}_{\mu\nu} U^\mu U^\nu \geq 0$ for any causal vector U.

Together, these assumptions automatically imply that any such solution must be either axisymmetric or static [453]. If the former, then it must admit a second Killing vector m that is spacelike near infinity, and one can further show that there is always a linear combination of k and m that remains timelike in the region exterior to the black hole [177,454]. It then follows that the scalar's gradient $\nabla_\mu \phi$ is either spacelike or zero in this region, since it is orthogonal to both k and m by assumption. The same is true also if the spacetime is static, since in this case, k is timelike in the whole of the exterior region.

We will now argue against the possibility of a spacelike gradient. To do so, multiply both sides of (6.10b) by $-V'(\phi)$ and integrate over some volume $\mathcal{V}$ in the exterior region to find

$$\int_{\mathcal{V}} \mathrm{d}^4x \sqrt{-g}\, [-V'(\phi)\nabla^\mu \nabla_\mu \phi + V'(\phi)^2] = 0. \tag{6.11}$$

Integrating the first term by parts then yields

$$\int_{\mathcal{V}} \mathrm{d}^4x \sqrt{-g}\, [V''(\phi)\nabla^\mu \phi \nabla_\mu \phi + V'(\phi)^2] = \int_{\partial\mathcal{V}} \mathrm{d}S^\mu\, V'(\phi)\nabla_\mu \phi, \tag{6.12}$$

where $\mathrm{d}S^\mu$ is the surface element on the boundary ∂V of the integration volume, which points in the direction normal to ∂V. The trick is to now show that the surface integral on the rhs vanishes for suitable choices of $\mathcal{V}$. In particular, let $\mathcal{V}$ be bounded by a portion $\mathcal{N}$ of the event horizon, two partial Cauchy surfaces $\mathcal{S}_1$ and $\mathcal{S}_2$,[3] and a timelike surface $\mathcal{T}$ at infinity. The integral along $\mathcal{T}$ vanishes since $\phi \to \phi_0$

[3] A partial Cauchy surface is a spacelike hypersurface which no causal curve intersects more than once [449].

at large distances, while the integral along $\mathcal{N}$ vanishes because the normal to the horizon must be a linear combination of k and m [453], and so is orthogonal to $\nabla_\mu\phi$. As for $\mathcal{S}_1$ and $\mathcal{S}_2$, first choose $\mathcal{S}_2$ to be the hypersurface obtained by moving each point on $\mathcal{S}_1$ a unit parameter distance along the integral curves of k. Stationarity then implies that the surface integrals over $\mathcal{S}_1$ and $\mathcal{S}_2$ have equal magnitude but opposite sign, since their corresponding surface elements $\mathrm{d}S^\mu$ point in opposite directions, and so cancel each other out.

Having previously argued that $\nabla_\mu\phi$ must be either spacelike or zero (i.e., $\nabla^\mu\phi\nabla_\mu\phi \geq 0$), we now see that if we impose the added assumption that $V''(\phi) \geq 0 \;\forall\, \phi$, the lhs of (6.12) vanishes only if $\phi = \phi_0$ everywhere on $\mathcal{V}$. Of course, the choice of $\mathcal{S}_1$ was arbitrary, and since we require that this solution be analytic, it must be that $\phi = \phi_0$ everywhere on spacetime. Now turning to (6.10a), its trace reverse is seen to reduce to the vacuum equation $R_{\mu\nu} = 0$ when $\phi = \phi_0$, and we already know that the Kerr metric is the only black hole solution to this equation that satisfies all four of our assumptions.

Notice that the above argument hinges on the scalar having a nontrivial potential, and so does not cover the case of a free, massless scalar field, for which $V(\phi) = 0$. Luckily, only a slight modification is needed to prove that the same conclusion holds also for this special case. Instead of $-V'(\phi)$, we multiply (6.10b) by ϕ when $V(\phi) = 0$ and then integrate by parts over the same volume $\mathcal{V}$ to obtain

$$\int_{\mathcal{V}} \mathrm{d}^4x\sqrt{-g}\,\nabla^\mu\phi\nabla_\mu\phi = \int_{\partial V} \mathrm{d}S^\mu\phi\,\nabla_\mu\phi. \tag{6.13}$$

The surface integral on the rhs vanishes for exactly the same reasons, and so we again have that ϕ must be constant everywhere on spacetime. To reiterate, we have thus shown that for potentials satisfying $V(\phi) = 0$ or $V''(\phi) \geq 0$, the only black hole solutions to (6.10) that are stationary, analytic, and asymptotically flat are those with a metric belonging to the Kerr family and a scalar that is everywhere a constant. If $V(\phi) = 0$, its shift symmetry guarantees that any constant value for the scalar is a valid solution, but we must have $\phi = \phi_0$ with $V(\phi_0) = V'(\phi_0) = 0$ otherwise.

6.2 Dynamical multipole moments

The proof of the no-hair theorem in the previous section is predicated on a number of crucial assumptions, whose physical rationale we should now discuss. For starters, requiring that the scalar potential satisfy $V''(\phi) \geq 0$ amounts to imposing a stability condition, as one can indeed find stationary, asymptotically flat black hole solutions with a nontrivial scalar-field profile when $V(\phi)$ admits one or more local maxima [455–458], but these are always linearly unstable [181, 455, 459].

As for stationarity and asymptotic flatness, these are good assumptions in so far as they pertain to the spacetime around an isolated, quiescent black hole, but one that is surrounded by an ambient scalar, which can evolve and persist independently, need no longer be subject to these restrictions. The classic example for this scenario is that of a quintessence-type scalar with a runaway potential, as the absence of a local minimum for $V(\phi)$ means that Minkowski spacetime with a constant (and finite) value of ϕ is never a valid solution to (6.10). Consequently, the simplest solution to these field equations is that of an expanding Universe, and the simplest solution containing a black hole corresponds to one at rest with respect to this Hubble flow. Although the time-dependent nature of the problem makes it too difficult to solve exactly, it is nevertheless possible to construct approximate solutions for this scenario that are valid for finite durations of time [309–312]. These approximate solutions suffice to show that the scalar inevitably develops a radial profile around the black hole which, as we discussed in Chapter 1, may be characterised in terms of an induced scalar charge.

Of course, a quintessence-type scalar generates just one example of a time-dependent environment into which we may embed a black hole. Massive scalars, for instance, form localised, gravitationally bound objects that resist further collapse by oscillating in time [460–464], and those that emerge at the lighter end of the spectrum after string compactification are of particular interest here, because the configurations that they form are large enough to satisfy the hierarchy in (6.2). In this section, we will demonstrate that black holes embedded in this kind of environment also develop an induced scalar charge.

As in the quintessence case, the time-dependent nature of this problem makes solving the field equations exactly untenable, but an approximate solution can be obtained under the following assumptions:

(1) We will assume that the ambient scalar varies on a length scale λ that is much greater than the size of the black hole, as measured by an observer comoving with this black hole.

(2) Similarly, we will assume that the ambient scalar evolves on a timescale ω^{-1} that is much longer than the black hole's light-crossing time.

(3) The Compton wavelength $\sim \mu^{-1}$ for this scalar is also assumed to be much larger than the size of the black hole.

(4) We will assume that the energy density in the scalar is sufficiently dilute that its backreaction onto the spacetime in the immediate vicinity of the event horizon is subdominant to the black hole's own spacetime curvature.

(5) The energy density in this scalar is further assumed to be dilute enough everywhere on spacetime that its self-interactions may be neglected as a first approximation.

It goes without saying that the first three of these assumptions imply that a black hole of mass M must satisfy the conditions $M/\lambda \ll 1$, $M\omega \ll 1$, and $M\mu \ll 1$. In the same vein, it will also be useful to have order-of-magnitude estimates that determine when the other two assumptions are valid. Roughly speaking, the scalar can be said to exert a negligible backreaction onto the spacetime around the black hole if its local energy density ρ is small compared to some measure of the spacetime curvature. Because a vacuum black hole has vanishing Ricci tensor, the simplest measure that we can construct for this purpose (with dimensions of an energy density) is the curvature invariant $(R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma})^{1/2}$, which scales as $1/M^2$ close to the horizon. Thus, our fourth assumption can be taken to be valid as long as

$$\rho \ll 1/M^2 \sim 10^{20}\ M_\odot\,\mathrm{pc}^{-3}\left(\frac{M}{10^{10}\ M_\odot}\right)^{-2}. \tag{6.14}$$

Our fifth assumption can also be interpreted as an upper bound on ρ. For concreteness, suppose that this scalar is an ultralight axion, whose potential may be expanded in powers of ϕ as[4]

$$V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4!}\mathrm{g}_4\phi^4 + \mathcal{O}(\phi^6). \tag{6.15}$$

As the strength of the quartic term is set by the symmetry-breaking scale F_ϕ ($\sim \mu/\sqrt{\mathrm{g}_4}$) associated with this axion [215, 259], demanding that $\mathrm{g}_4\phi^4 \ll \mu^2\phi^2$ amounts to requiring that $\phi \ll F_\phi$. Equivalently, if we take $\rho \sim \mu^2\phi^2$, as is usually the case when the field is dilute, then we must have that

$$\rho \ll \mu^2 F_\phi^2 \sim 10^{11}\ M_\odot\,\mathrm{pc}^{-3}\left(\frac{\mu}{10^{-22}\ \mathrm{eV}}\right)^2\left(\frac{F_\phi}{10^{17}\ \mathrm{GeV}}\right)^2. \tag{6.16}$$

In principle, one also has to check that all of the higher-order interaction terms are similarly small, but we shall assume that $V(\phi)$ is such that this is true whenever $\mathrm{g}_4\phi^4 \ll \mu^2\phi^2$.

For any realistic scenario of the kind that we are interested in, the upper bounds in (6.14) and (6.16) are sufficiently large that they are never violated; hence, the spacetime in the immediate vicinity of the black hole is well described by a Kerr metric that is only weakly perturbed by the ambient scalar. Focusing in on this inner zone and neglecting the small effect from the latter's backreaction onto this spacetime, it now follows that the behaviour of the scalar in this region can be determined by simply solving the Klein–Gordon equation on a fixed Kerr background.

[4] Axion or otherwise, note that a scalar field can always be redefined such that the minimum of its potential (should it exist) sits at $\phi = 0$, and that any realistically sized cosmological-constant-like contribution coming from having $V(0) \neq 0$ can be ignored on the grounds that it has negligible influence on astrophysical scales.

This equation is separable in Boyer–Lindquist coordinates [331,465]; hence, its general solution may be constructed from a linear combination of modes of the form

$$\phi(x) \propto R_{\omega\ell m}(r) S_{\omega\ell m}(\vartheta)\, e^{im\chi - i\omega t}. \tag{6.17}$$

Substituting this ansatz into the Klein–Gordon equation then gives us two ordinary differential equations to solve. The radial function R satisfies [333]

$$\left[\frac{\mathrm{d}}{\mathrm{d}r}\left(\Delta\frac{\mathrm{d}}{\mathrm{d}r}\right) + \frac{\omega^2(r^2+s^2)^2 - 4Msm\omega r + m^2 s^2}{\Delta} - (\omega^2 s^2 + \mu^2 r^2 + \Lambda_{\omega\ell m})\right] R_{\omega\ell m} = 0, \tag{6.18}$$

where s and Δ are defined as in Section 6.1.2, while the angular function S obeys [333]

$$\left[\frac{1}{\sin\vartheta}\frac{\mathrm{d}}{\mathrm{d}\vartheta}\left(\sin\vartheta\frac{\mathrm{d}}{\mathrm{d}\vartheta}\right) + s^2(\omega^2-\mu^2)\cos^2\vartheta - \frac{m^2}{\sin^2\vartheta} + \Lambda_{\omega\ell m}\right] S_{\omega\ell m} = 0. \tag{6.19}$$

The solutions to this latter equation give rise to the spheroidal harmonics [466], with the separation constant $\Lambda_{\omega\ell m}$ determined as an eigenvalue of the problem.

At this stage, it is still not possible to write down exact, analytic solutions to either of these equations, but approximate ones can be obtained under our assumption of a separation of scales, as this allows truncating the equations to first order in the small parameters $M\omega$ and $M\mu$. The angular equation in (6.19) reduces to Legendre's in this limit, and so the angular part of (6.17) must reduce to the spherical harmonics; i.e., $S_{\omega\ell m}(\vartheta)e^{im\chi} \sim Y_{\ell m}(\vartheta,\chi)$ with $\Lambda_{\omega\ell m} \sim \ell(\ell+1)$.

An additional restriction will need to be imposed on the radial equation to render it soluble. We do so by concentrating only on small radii $r \ll \min(\omega^{-1}, \mu^{-1})$ within the inner zone, keeping in mind that the overall result must later be matched onto a solution that is valid at larger distances. This matching will take place in Chapter 7, but for now, limiting our attention to the inner zone is tantamount to discarding terms suppressed by $\mathcal{O}(\omega(r-r_+))$ or $\mathcal{O}(\mu(r-r_+))$ in (6.18). Having done so, the radial equation reduces to [331]

$$\left[\frac{\mathrm{d}}{\mathrm{d}r}\left(\Delta\frac{\mathrm{d}}{\mathrm{d}r}\right) + \frac{\mathscr{P}_{\omega m}^2 (r_+ - r_-)^2}{\Delta} - \ell(\ell+1)\right] R_{\omega\ell m} = 0, \tag{6.20}$$

where

$$\mathscr{P}_{\omega m} := \frac{sm - 2Mr_+\omega}{r_+ - r_-}. \tag{6.21}$$

This definition for $\mathcal{P}$ is useful because the future event horizon at $r = r_+$ is a regular singular point of (6.20) with indicial exponents $\pm i\mathcal{P}$, meaning that this differential equation admits two, linearly independent solutions of the form

$$R_{\omega\ell m}(r) = \left(\frac{r - r_+}{r - r_-}\right)^{\pm i\mathcal{P}_{\omega m}} \psi_{\omega\ell m}(r), \tag{6.22}$$

where $\psi_{\omega\ell m}(r)$ is some function that is analytic about $r = r_+$. Strictly, nothing tells us that the extra factor of $(r - r_-)^{\mp i\mathcal{P}_{\omega m}}$ is needed out front, but its inclusion will prove to be convenient later.

The claim is now that only the $+i\mathcal{P}$ solution is physical. Indeed, substituting (6.22) back into (6.17) and temporarily transforming to Kerr coordinates shows that only this solution is regular on $r = r_+$, whereas the $-i\mathcal{P}$ solution diverges on the future event horizon and should therefore be discarded. As for the remaining function ψ, defining a new variable $u = (r - r_-)/(r_+ - r_-)$ shows that it must be a solution to the hypergeometric differential equation,

$$\left(u(1-u)\frac{\mathrm{d}^2}{\mathrm{d}u^2} + (1 - 2i\mathcal{P}_{\omega m} - 2u)\frac{\mathrm{d}}{\mathrm{d}u} + \ell(\ell+1)\right)\psi_{\omega\ell m}(r) = 0, \tag{6.23}$$

with the boundary condition that it be analytic about $u = 1$. Thus, up to an overall normalisation, the radial solution is

$$R_{\omega\ell m}(r) \propto \left(\frac{r - r_+}{r - r_-}\right)^{i\mathcal{P}_{\omega m}} {}_2F_1(-\ell, \ell+1; 1 + 2i\mathcal{P}_{\omega m}; 1 - u). \tag{6.24}$$

Let us now zoom out on this solution by expanding it in powers of M/r to obtain a coarse-grained description valid at distances $M \ll r \ll \min(\omega^{-1}, \mu^{-1})$. The two dominant terms in this expansion are

$$R_{\omega\ell m}(r) \;\propto\; r^{\ell} + C_{\omega\ell m} r^{-\ell-1} \tag{6.25}$$

with relative coefficients given by

$$C_{\omega\ell m} = (r_+ - r_-)^{2\ell+1} \frac{\ell!}{(2\ell)!} \frac{\Gamma(-1-2\ell)}{\Gamma(-\ell)} \frac{\Gamma(1+\ell+2i\mathcal{P}_{\omega m})}{\Gamma(-\ell+2i\mathcal{P}_{\omega m})}. \tag{6.26}$$

This expression can be simplified further through judicious use of the recurrence relation $z\Gamma(z) = \Gamma(z+1)$, although care must be exercised when dealing with the ratio $\Gamma(-1-2\ell)/\Gamma(-\ell)$. One should first shift $\ell \to \ell + \epsilon$ and take the limit $\epsilon \to 0$ only at the end; otherwise, a spurious factor of 2 will appear [337]. Also using the identity

$$\prod_{j=1}^{\ell} (j^2 + 4\mathcal{P}^2) \equiv \frac{\Gamma(1+\ell+2i\mathcal{P})\Gamma(1+\ell-2i\mathcal{P})}{\Gamma(1+2i\mathcal{P})\Gamma(1-2i\mathcal{P})}, \tag{6.27}$$

the final result is

$$C_{\omega\ell m} = -i\mathcal{P}_{\omega m}\,(r_+ - r_-)^{2\ell+1}\,\frac{(\ell!)^2}{(2\ell)!(2\ell+1)!}\prod_{j=1}^{\ell}(j^2 + 4\mathcal{P}_{\omega m}^2). \tag{6.28}$$

Scalar charge Let us now discuss the physical implications of this general solution on a mode-by-mode basis; keeping in mind that we should project out only the real part of (6.17) if $\phi(x)$ is a real scalar. Beginning with the $\ell = 0$ mode and assuming that it oscillates at a distinct frequency ω, this part of the solution reads

$$\phi|_{\ell=0} = \mathrm{Re}\,\Phi_0 e^{-i\omega t}(1 + C_{\omega 00} r^{-1}) = \mathrm{Re}\,\Phi_0 e^{-i\omega t}(1 + 2Mr_+ i\omega r^{-1}), \tag{6.29}$$

having included an overall amplitude Φ_0 for the field. One can now verify that this result may be written more succinctly as

$$\phi|_{\ell=0} = \Phi(t) + \frac{Q(t)}{4\pi r} \tag{6.30}$$

after defining $\Phi(t) = \Phi_0 \cos\omega t$ and $Q(t) = -A\partial_t\Phi$, where $A = 8\pi M r_+$ is the area of the black hole's event horizon. Written in this way, the first term in (6.30) may be regarded as describing the ambient scalar-field background into which this black hole is embedded, while the second term captures how this field profile must be displaced as a result of the embedding. Physically, this Coulomb-like potential arises simply because the scalar is now forced to undergo absorption at the event horizon, but it is interesting to interpret $Q(t)$ as a charge that the time-dependent background dynamically induces onto the black hole. As we will see in the next chapter, thinking in terms of these scalar charges will be helpful in shaping our intuition for how black holes behave in these kinds of environments.

Dipole moments The $\ell = 0$ solution in (6.30) describes the effect of embedding a black hole in a background that is time-dependent but otherwise spatially homogeneous on the scale of the inner zone. This is, of course, just the simplest possibility. Let us now explore the effect of including the $\ell = 1$ mode,

$$\phi|_{\ell=1} = \sum_{m=-1}^{1} b_m e^{-i\omega t}(r + C_{\omega 1m} r^{-2}) Y_1^m(\vartheta, \chi), \tag{6.31}$$

where the constants b_m scale with the inverse of the characteristic length scale λ over which the ambient background varies. For a concrete example, consider a particular combination of these constants such that the background is given by $\Phi = (\mathbf{b}\cdot\mathbf{x})\cos\omega t$. The constants b_m are then related to the Cartesian

components of the vector $\mathbf{b} = (b_x, b_y, b_z)$ by the equations

$$b_{\pm 1} = \sqrt{\frac{2\pi}{3}}(b_x + i\, b_y), \qquad b_0 = \sqrt{\frac{4\pi}{3}} b_z. \tag{6.32}$$

Unlike the $\ell = 0$ case, the coefficient $C_{\omega 1m}$ has a term that is independent of ω, which is given by $C_{\omega 1m} = -i\, s m M^2/3 + \mathcal{O}(M\omega)$. Substituting this back into (6.31) reveals that, in the presence of a background scalar gradient $\nabla\Phi \equiv \mathbf{b}\cos\omega t$, black holes also acquire a spin-dependent dipole moment:

$$\phi|_{\ell=1} = \Phi + \frac{\mathbf{p}\cdot\hat{\mathbf{x}}}{4\pi r^2}, \qquad \mathbf{p} = \frac{4\pi s M^2}{3}(\hat{\mathbf{S}} \times \nabla\Phi) + \mathcal{O}(M\omega), \tag{6.33}$$

where $\hat{\mathbf{S}}$ is the unit vector along the black hole's spin axis. Notice that this dipole moment $\mathbf{p}$ survives in the static limit $\omega \to 0$ when $s \neq 0$, but the no-hair theorem is still circumvented in this case because a linear gradient $\Phi \sim \mathbf{b}\cdot\mathbf{x}$ violates the assumption of asymptotic flatness.

Spherical black holes can also attain a dipole moment, although the effect is suppressed by one power of $M\omega$ relative to the spinning case. Setting $s = 0$ in (6.28) yields

$$C_{\omega\ell m}|_{s=0} = \frac{(\ell!)^4}{(2\ell)!(2\ell+1)!}(2M)^{2\ell+2} i\omega. \tag{6.34}$$

Upon substitution into (6.31), we find that the spin-independent part of the dipole moment is

$$\mathbf{p}|_{s=0} = -\frac{16\pi M^4}{3}\frac{\mathrm{d}}{\mathrm{d}t}(\nabla\Phi). \tag{6.35}$$

Higher multipole moments The same procedure can be repeated for $\ell \geq 2$; hence, we learn that a black hole embedded in an arbitrary, ambient scalar-field profile generally gains an infinite set of multipole moments. In practice, however, it often suffices to keep only the scalar charge (and perhaps the spin-dependent dipole moment) as the higher multipoles are suppressed by ever greater powers of M/λ ($\ll 1$), making their phenomenological impact increasingly irrelevant.

Chapter 7
Point-particle black holes

We learnt in the last chapter that when a black hole is embedded in an ambient scalar field, absorption across the horizon leads to deformations in the field profile that fall off radially as $r^{-\ell-1}$ ($\ell \geq 0$) at large distances. In a sense, these deformations can be regarded as being sourced by dynamical multipole moments that are induced onto the black hole by the ambient scalar. Although interesting in its own right, the more pressing question for distant observers like ourselves is how these multipole moments go on to influence the trajectory of the black hole and the spectrum of radiation it emits.

As we discussed previously, the evolution of this kind of system is governed by the action

$$S[g,\phi] = \int \mathrm{d}^4x\sqrt{-g}\left(2m_{\rm Pl}^2R - \frac{1}{2}(\partial\phi)^2 - \frac{1}{2}\mu^2\phi^2 + \mathcal{O}(\phi^4)\right), \tag{7.1}$$

although confronting the corresponding field equations head on is completely untenable for the kinds of questions we now wish to address. To make analytic progress, the key assumption we will exploit is the existence of a hierarchy that separates the length and timescales of the black hole from those of its local environment. In this regime, the black hole can be treated approximately as a point particle travelling along the worldline of some effective centre-of-energy coordinate $z^\mu(\tau)$.

From a modern perspective, this coarse-grained description of the system emerges after integrating out short-wavelength modes up to scales on the order of the horizon [377,380]. The result is an effective field theory, whose action is

$$S_{\rm eff} = S[g,\phi] + S_p[z,g,\phi]. \tag{7.2}$$

The first term S is still given by (7.1), but now governs only the remaining long-wavelength modes of the fields (g,ϕ), while the dynamics of the worldline and its interactions with the fields living in the bulk are given by the point-particle action S_p. In general, performing this kind of integral produces infinitely

many terms in S_p, but these can be ordered by relevance when large separations of scales are present. Specifically, this point-particle action can be organised as an expansion in three small parameters:

$$GM/\lambda \ll 1, \qquad GM\omega \ll 1, \qquad \text{and} \qquad GM\mu \ll 1. \tag{7.3}$$

As before, the mass of the black hole is denoted by M, while λ and ω^{-1} are the characteristic length and timescales over which the ambient scalar of mass μ varies. Seeing as only finitely many terms are needed when working to a prescribed level of precision, the goal of this chapter is to present a step-by-step construction and analysis of S_p to leading order in these expansion parameters.

Throughout this discussion, the scenario we should have at the back of our minds is that of some generic, ambient scalar-field configuration establishing a background within which just one black hole is moving. Chapters 8 and 9 will then build on these results to study the more challenging case of a binary black hole, but for now, the general plan is as follows: In Section 7.1, we begin by furnishing the black hole's worldline with a set of composite operators, whose job is to capture finite-size effects associated with the horizon. These dynamical variables are then integrated out within the in–in formalism to produce an effective action expressed in terms of their correlation functions, which we subsequently determine by matching this EFT onto a number of observables computed within the full theory. In contrast to a purely bottom-up approach, the key advantage of this procedure is that it can systematically generate all local interactions — even dissipative ones — that couple the worldline to the long-wavelength fields.

The physical effects stemming from the most relevant of these interactions are then analysed in two stages. In Section 7.2, we will show that the same Wilsonian coefficient that sets the strength of the black hole's scalar charge is also responsible for setting its accretion rate. This inextricable connection, which is fixed by the symmetries of the effective theory, is a manifestation of the fact that induced scalar charges are a direct consequence of absorption at the horizon. In Section 7.3, we then conclude this chapter by examining how this scalar charge influences the trajectory of the black hole.

7.1 The effective action

The finite-size effects associated with a black hole's horizon are captured in this EFT by a set of composite operators $\{q^L(\tau), \dots\}$ localised on the worldline $z^\mu(\tau)$ [378, 467, 468]. Physically, these operators are a proxy for short-wavelength degrees of freedom living near the horizon, which can exchange energy and angular momentum with the long-wavelength fields (g, ϕ) living in the bulk. To construct an effective action from these objects, standard EFT reasoning now dictates that we should

simply write down all possible local interactions that are consistent with the symmetries of the theory. In our case, these are diffeomorphism invariance, worldline reparametrisation invariance, and worldline SO(3) invariance. The last of these is not strictly necessary, of course, but is an assumption that will allow us to restrict our attention to spherical (i.e., nonspinning) black holes for simplicity.

These steps lead to what we will call the *intermediary* point-particle action,

$$I_p = -M \int \mathrm{d}\tau + \sum_{\ell=0}^{\infty} \int \mathrm{d}\tau \, q^L(\tau) \nabla_L \phi(\tau) + \cdots , \tag{7.4}$$

where the scalar $\phi(\tau) \equiv \phi(z(\tau))$ is to be evaluated along the worldline, whose proper time τ is given by $\mathrm{d}\tau = \sqrt{-g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu}$. The first term in (7.4) is the familiar action for a point mass M, while the second accounts for all possible interactions between the black hole and the real scalar field. The ellipsis alludes to the presence of analogous composite operators that couple to the gravitational field [378], although these will play no role in our discussion as they become important only at much higher orders in perturbation theory [377,469]. (Note that conventional multi-index notation is being used [447]. Any tensor with ℓ spatial indices is written as $q^L \equiv q^{\hat{\imath}_1 \cdots \hat{\imath}_\ell}$, whereas $\nabla_L \equiv \nabla_{\hat{\imath}_1} \cdots \nabla_{\hat{\imath}_\ell}$ denotes the action of multiple covariant derivatives. Similarly, a spatial vector repeated ℓ times is written as $\mathbf{x}^L \equiv \mathbf{x}^{\hat{\imath}_1} \cdots \mathbf{x}^{\hat{\imath}_\ell}$. The indices $\hat{\imath} \in \{1, 2, 3\}$ label the three spatial directions in the black hole's rest frame that are mutually orthonormal to one another and to the tangent of the worldline. In the nonrelativistic limit, these reduce to the usual three spatial directions in Minkowski space.)

7.1.1 Worldline operators and linear response theory

By counting the number of independent degrees of freedom, one finds that the composite worldline operators $\{q^L(\tau)\}$ can each be taken to be symmetric and trace free (STF). This follows because traces of $\nabla_L \phi$ are redundant operators [388,470], and so can be absorbed into redefinitions of q^{L-2N}, where n counts the number of traces [467, 468]. Since the set of all STF tensors of rank ℓ generates an irreducible representation of SO(3) [471], these worldline operators now admit a natural interpretation as dynamical multipole moments of the black hole [378]. The $\ell = 0$ operator $q(\tau)$ must therefore be responsible for the black hole's induced scalar charge, the $\ell = 1$ operator $q^{\hat{\imath}}(\tau)$ for its induced dipole moment, and so on.

Being dynamical variables in their own right, these worldline operators, which are associated with short-distance physics transpiring near the black hole's horizon, must come with appropriate kinetic terms that govern their dynamics. However, a long-distance EFT on its own supplies no specific details about what this short-distance physics might be; hence, additional input in the form of a matching

calculation is required. To start with, we should expect these operators to satisfy equations of motion of the form

$$\mathscr{D}_L(\, q(\tau), q^{\hat{i}}(\tau), q^{\hat{i}\hat{j}}(\tau), \ldots) = -\nabla_L \phi(\tau). \tag{7.5}$$

For each ℓ, $\mathscr{D}_L$ is some (possibly nonlinear) function of $\{q^L(\tau)\}$ and their derivatives, while the source term on the rhs comes from varying (7.4) with respect to q^L.

At this stage, the assumed separation of scales between the black hole and its local environment can be used to impose restrictions on the form of (7.5). Recall that one of the key assumptions guaranteeing this separation of scales concerns the energy density in the ambient scalar, which is assumed to be sufficiently dilute that its backreaction onto the spacetime in the inner zone is negligible in the first instance. In the EFT, making this test-field approximation translates into assuming that $\mathscr{D}_L$ is essentially linear in its arguments, in which case, the general solution to (7.5) is given by linear response theory:

$$q^L(\tau) = \langle q^L(\tau) \rangle + \sum_{\ell'=0}^{\infty} \int \mathrm{d}\tau' G_R^{LL'}(\tau, \tau') \nabla_{L'} \phi(\tau'). \tag{7.6}$$

The first term in (7.6) gives rise to a permanent multipole moment if nonzero. Such a term would be present if this were an operator corresponding to a more general compact object, like a neutron star in the DEF model [136–138], but no-hair theorems stipulate that Kerr or, in the present context, Schwarzschild black holes cannot directly source real scalar fields (see Chapter 6.1 or Refs. [174–181]); hence, we must have that $\langle q^L(\tau) \rangle = 0 \; \forall \; \ell$ in our case. Classically, we may think of this one-point correlation function as the solution to the homogeneous part of the differential equation in (7.5), in which case, enforcing the no-hair theorems amounts to specifying appropriate boundary conditions. More generally, these operators $\{q^L(\tau)\}$ may be regarded as quantum mechanical objects living on some Hilbert space with a density matrix $\hat{\rho}$ defined such that $\langle q^L \rangle \equiv \mathrm{tr}(\hat{\rho} q^L) = 0 \; \forall \; \ell$, so as to correspond to a black hole. On its own, this condition does not uniquely pick out a density matrix for the short-wavelength degrees of freedom living near the horizon, but when one is specified correctly, the higher-order correlation functions for $\{q^L(\tau)\}$ can be used to systematically incorporate quantum effects like Hawking radiation [472, 473].

That being said, we will be interested purely in the classical behaviour of black holes in this thesis. It is then unnecessary to specify the exact properties of this density matrix $\hat{\rho}$, since the only other correlation function left in (7.6) is the retarded Green's function

$$G_R^{LL'}(\tau, \tau') := i\theta(\tau - \tau') \langle [q^L(\tau), q^{L'}(\tau')] \rangle. \tag{7.7}$$

This two-point function is necessarily independent of initial conditions when $\mathcal{D}_L$ is linear, which at the quantum level corresponds to saying that the commutator is simply a c-number whose expectation value is thus independent of $\hat{\rho}$.

It remains to discuss the form of this retarded Green's function. As the black hole's worldline is reparametrisation invariant, it is necessarily invariant under proper time translations; hence, $G_R^{LL'}(\tau, \tau')$ depends on its arguments only via their difference, $\tau - \tau'$. Physically, this reflects the fact that the spacetime around an isolated black hole is stationary, and allows us to write

$$G_R^{LL'}(\tau, \tau') = \int \frac{\mathrm{d}\omega}{2\pi} \, \widetilde{G}_R^{LL'}(\omega) \, e^{-i\omega(\tau-\tau')} \tag{7.8}$$

in terms of its Fourier transform. Dimensional analysis and the assumption of spherical symmetry further restrict this function to be of the form

$$\widetilde{G}_R^{LL'}(\omega) = \delta^{LL'}(GM)^{2\ell+1} F_\ell(GM\omega, GM\mu), \tag{7.9}$$

where $\delta^{LL'}$ is the identity on the space of STF tensors of rank ℓ. The dimensionless functions F_ℓ almost certainly depend in a complicated way on their arguments, but under the assumption of a large separation of scales as per (7.3), we can expand in powers of the first argument to obtain

$$\begin{aligned} F_\ell &= \sum_{n=0}^{\infty} \left[F_\ell^{(2n)} (GM\omega)^{2n} + i F_\ell^{(2n+1)} (GM\omega)^{2n+1} \right] \\ &= F_\ell^{(0)} + i F_\ell^{(1)} GM\omega + F_\ell^{(2)} (GM\omega)^2 + \cdots , \end{aligned} \tag{7.10}$$

where the real, dimensionless coefficients $F_\ell^{(n)}$ themselves admit a further expansion in the remaining argument $GM\mu$. Naturally, the size of the black hole sets the UV cutoff for the validity of this expansion. The terms in (7.10) that are even in ω are time-reversal symmetric and constitute what is called the reactive part of the black hole's response. On the other hand, the odd terms that break time-reversal symmetry are responsible for dissipative processes. The values of these Wilsonian coefficients $F_\ell^{(n)}$ are the last remaining piece of the puzzle. They will be determined in Section 7.1.3 via a number of matching calculations; but to do so requires that we first integrate out $\{q^L(\tau)\}$.

7.1.2 Integrating out the worldline operators

As its name suggests, the intermediary point-particle action in (7.4) is not the end of the story. Presently, it is comprised of the UV degrees of freedom $\{q^L(\tau)\}$, which a distant observer cannot directly probe,

and the IR degrees of freedom (z, g, ϕ) that we ultimately care about. While it is possible to perform calculations with this action in its current form (see, e.g., Refs. [378, 468]), for our purposes it will be more convenient — and also instructive — to integrate out the worldline operators and obtain a truly effective point-particle action:

$$S_p[z, g, \phi] = -i \log \int D[q^L] \exp(i I_p[z, g, \phi, q^L]). \tag{7.11}$$

At the classical level, integrating out essentially amounts to substituting the solution in (7.6) back into (7.4), but doing so naively will give rise to acausal equations of motion [474]. To properly implement retarded boundary conditions, the appropriate language required for integrating out these worldline operators is the in–in, or closed time path (CTP), formalism. (See Refs. [475–479] for classic texts on the subject and Refs. [469, 480–482] for applications similar to the present context.) At its heart, this formalism converts the standard version of Hamilton's variational principle, which is inherently a boundary value problem, into an initial value one. It accomplishes this by doubling all dynamical degrees of freedom $\Psi \mapsto (\Psi_1, \Psi_2)$ and allowing both copies to evolve independently subject to appropriate boundary conditions. To elaborate, the first copy Ψ_1 is evolved forwards in time and is allowed to take all paths (in the sense of Feynman) subject to boundary conditions at some initial time, say t_i. At a later time t_f, these paths are connected to the second copy Ψ_2, which is then evolved freely backwards in time subject to boundary conditions at t_i. The physical solution is then obtained by making the identification $\Psi_1 = \Psi_2 = \Psi$ at the end. Following Galley [474], we will refer to this identification as "taking the physical limit."

Fixed worldlines The degrees of freedom in this EFT are $\Psi = \{z^\mu, g_{\mu\nu}, \phi, q^L\}$, and we wish to integrate out q^L. It will be instructive to start by considering a simplified problem in which we fix the metric and worldline to be nondynamical. Under this restriction, (7.4) reads

$$I_p = \int \mathrm{d}\tau \, (q_1\phi_1 - q_2\phi_2) + \cdots \tag{7.12}$$

when recast in the in–in formalism. As the generalisation to higher multipole moments is straightforward, we will focus solely on the $\ell = 0$ operator in our discussion for simplicity.

Introducing CTP indices $a, b \in \{1, 2\}$ allows us to write the integrand in (7.12) more succinctly as $q_1\phi_1 - q_2\phi_2 = c^{ab} q_a \phi_b = q_a \phi^a$. Note that all degrees of freedom $\Psi_a = (\Psi_1, \Psi_2)$ innately come with a lowered index; indices are raised with the CTP metric $c^{ab} = c_{ab} = \mathrm{diag}(1, -1)$. While on the subject of definitions, let us also introduce a different basis $\Psi_A = (\Psi_+, \Psi_-)$, known as the Keldysh

representation. This is related to the original basis via the transformation

$$\Psi_A = \Lambda_A{}^a \Psi_a, \qquad \Lambda_A{}^a = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & -1 \end{pmatrix}. \tag{7.13}$$

Written out explicitly, we see that the components of Ψ_A are exactly the average and difference of the two copies, $\Psi_+ = (\Psi_1 + \Psi_2)/2$ and $\Psi_- = \Psi_1 - \Psi_2$. In the physical limit (PL), $\Psi_+|_{\rm PL} = \Psi$ and $\Psi_-|_{\rm PL} = 0$. Indices are still raised and lowered with the CTP metric, which in this basis reads

$$c^{AB} = c_{AB} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{7.14}$$

We now integrate out $q_a(\tau)$ by evaluating the path integral

$$\exp(iS_p) = \int D[q_\pm] \exp\left(i \int \mathrm{d}\tau \, [\mathcal{L}(q_+, q_-) + q_A \phi^A] \right), \tag{7.15}$$

where $\mathcal{L}(q_+, q_-)$ represents the unknown kinetic terms for this worldline operator. To reproduce the solution in (7.6), we should take $\mathcal{L}(q_+, q_-) = (1/2) q_A(\tau) \mathcal{D}_\tau^{AB} q_B(\tau)$, where $\mathcal{D}_\tau^{AB}$ is some matrix of local, linear differential operators. This gives us a Gaussian path integral that can be evaluated exactly. We complete the square by defining a new variable $q'_A(\tau) = q_A(\tau) + \int \mathrm{d}\tau' G_{AB}(\tau, \tau') \phi^B(\tau')$, where G_{AB} is the matrix of two-point functions that satisfies $\mathcal{D}_\tau^{AB} G_{BC}(\tau, \tau') = -\delta^A_C \delta(\tau - \tau')$. Making this field redefinition leaves us with

$$\exp(iS_p) = \exp\left(\frac{i}{2} \int \mathrm{d}\tau \mathrm{d}\tau' \, G_{AB}(\tau, \tau') \phi^A(\tau) \phi^B(\tau') \right) \int D[q'_\pm] \exp\left(i \int \mathrm{d}\tau \, \mathcal{L}(q'_+, q'_-) \right). \tag{7.16}$$

As the remaining path integral over $q'_\pm$ only contributes an overall normalisation to the partition function, it can be neglected. After also repeating similar steps for the higher multipole operators, we obtain

$$S_p = \frac{1}{2} \sum_{\ell=0}^{\infty} \sum_{\ell'=0}^{\infty} \int \mathrm{d}\tau \mathrm{d}\tau' \, G^{LL'}_{AA'}(\tau, \tau') \nabla_L \phi^A(\tau) \nabla_{L'} \phi^{A'}(\tau') \tag{7.17}$$

as the end result. The matrix of two-point functions above is [477,479]

$$G^{LL'}_{AA'}(\tau, \tau') = \begin{pmatrix} \frac{1}{2} G_H^{LL'}(\tau, \tau') & G_R^{LL'}(\tau, \tau') \\ G_A^{LL'}(\tau, \tau') & 0 \end{pmatrix}, \tag{7.18}$$

where G_H, G_R, and G_A are the Hadamard, retarded, and advanced Green's functions, respectively. From their definitions in Appendix 7.B, the matrix in (7.18) can be seen to have the symmetry property

$$G^{LL'}_{AA'}(\tau,\tau') = G^{L'L}_{A'A}(\tau',\tau). \tag{7.19}$$

Dynamical worldlines Having gained a sense for how this calculation proceeds, we are now in a position to integrate out $\{q^L(\tau)\}$ in the general case when all degrees of freedom are dynamical. Complications arise when there are two copies (z_1, z_2) of the worldline for *one* black hole, since the operators $q^L_1(\tau_1)$ appear to be living on the first copy $z_1(\tau_1)$, whereas $q^L_2(\tau_2)$ live on the second. How, then, should we integrate out these operators? In particular, what does it mean to take the correlation function of operators living on different spaces?

The resolution is to recall that the coordinates z^μ are merely parametrisations in a given coordinate chart. The worldline itself is a map $\gamma : \mathcal{I} \to \mathcal{M}$ from the interval $\mathcal{I} \subset \mathbb{R}$ to the bulk four-dimensional manifold $\mathcal{M}$. When there are two copies z_a, there are also two maps γ_a, but there is still only one underlying manifold $\mathcal{I}$. Let σ be the coordinate on $\mathcal{I}$ used to parametrise both copies of the worldline simultaneously, whose tangents are written as $\dot{z}^\mu_a = \mathrm{d}z^\mu_a/\mathrm{d}\sigma$. The operators $q^L_a(\tau_a)$ are pulled back onto $\mathcal{I}$ via the map

$$\tau_a(\sigma) = \tau_a(\sigma_i) + \int_{\sigma_i}^{\sigma} \mathrm{d}\sigma' \sqrt{-g_{a,\mu\nu}(z_a(\sigma'))\frac{\mathrm{d}z^\mu_a}{\mathrm{d}\sigma'}\frac{\mathrm{d}z^\nu_a}{\mathrm{d}\sigma'}} \qquad \text{(no CTP sum)}, \tag{7.20}$$

where it should be understood that the CTP index a above is acting as a placeholder and is not to be summed over. We are always free to choose the lower integration limit σ_i and the initial value $\tau_a(\sigma_i)$. The intermediary point-particle action thus reads

$$I_p = \int \mathrm{d}\sigma \left[-M\dot{\tau}_1(\sigma) + \dot{\tau}_1(\sigma) q_1(\tau_1(\sigma))\phi_1(z_1(\sigma)) - (1 \leftrightarrow 2) \right] + \cdots . \tag{7.21}$$

As before, we will focus only on the $\ell = 0$ operator, as the generalisation to $\ell \geq 1$ is straightforward.

Clearly, the mess that is (7.21) calls for better notation. To that end, we shall generalise the CTP metric to the set of tensors

$$c^{a_1\cdots a_n} = \begin{cases} +1 & (a_1 = a_2 = \cdots = a_n = 1) \\ -1 & (a_1 = a_2 = \cdots = a_n = 2) \\ 0 & (\text{otherwise}). \end{cases} \tag{7.22}$$

It is then possible to verify by direct evaluation that (7.21) is equivalent to

$$I_p = -M \int \mathrm{d}\sigma\, c^a \dot{\tau}_a(\sigma) + \int \mathrm{d}\sigma\, q_a(\sigma) \mathcal{J}^a(\sigma) \tag{7.23}$$

after defining

$$\mathcal{J}^a(\sigma) := \int \mathrm{d}^4x\, c^{abc}\, \Delta_b(\sigma; x) \phi_c(x), \tag{7.24a}$$

$$\Delta^a(\sigma; x) := \int \mathrm{d}\sigma'\, c^{abcd}\, \delta(\sigma - \tau_b(\sigma'))\, \delta^{(4)}(x - z_c(\sigma'))\, \dot{\tau}_d(\sigma'). \tag{7.24b}$$

In this form, (7.23) is reminiscent of the simplified problem discussed earlier, apart from two minor differences: the manifold $\mathcal{I}$ is parametrised by σ rather than τ, and the scalar field ϕ^a is here replaced by the source $\mathcal{J}^a$. These prove to be no obstacle to evaluating the path integral, which yields

$$S_p = -M \int \mathrm{d}\sigma\, c^a \dot{\tau}_a(\sigma) + \frac{1}{2} \int \mathrm{d}\sigma \mathrm{d}\sigma'\, G_{aa'}(\sigma, \sigma') \mathcal{J}^a(\sigma) \mathcal{J}^{a'}(\sigma'). \tag{7.25}$$

Before proceeding any further, it is worth remarking that the Hadamard propagator $G_H \equiv G_{++}$ appears in this action flanked by two powers of $\mathcal{J}^+$ ($\equiv \mathcal{J}_-$), which vanishes in the physical limit. This implies that when we extremise the full effective action S_{eff} to obtain the (classical) equations of motion for this system, G_H will never contribute; hence, we can just set $G_H = 0$ from now on.

The hard work is over at this point, but (7.25) is not yet in a form that is convenient for calculations. Using the definitions in (7.24), we may write

$$S_p = -M \int \mathrm{d}\sigma\, c^a \dot{\tau}_a(\sigma) + \frac{1}{2} \int \mathrm{d}^4x \mathrm{d}^4x'\, \mathcal{G}^{aa'}(x, x') \phi_a(x) \phi_{a'}(x'), \tag{7.26}$$

where

$$\mathcal{G}^{aa'}(x, x') := \int \mathrm{d}\sigma \mathrm{d}\sigma'\, c^{abc} c^{a'b'c'} \Delta_b(\sigma; x) \Delta_{b'}(\sigma'; x') G_{cc'}(\sigma, \sigma'). \tag{7.27}$$

As a generalisation of (7.19), it is easy to show that

$$\mathcal{G}^{aa'}(x, x') = \mathcal{G}^{a'a}(x', x). \tag{7.28}$$

Finally, let us transform these correlation functions into the Keldysh representation to make their causal structure manifest. Utilising the transformation rule $\mathcal{G}^{AA'}(x, x') = \mathcal{G}^{aa'}(x, x') \Lambda_a{}^A \Lambda_{a'}{}^{A'}$, where the inverse matrix $\Lambda_a{}^A$ satisfies $\Lambda_a{}^A \Lambda_A{}^b = \delta_a^b$, we find

$$\mathcal{G}^{++}(x, x') = \frac{1}{2} \int \mathrm{d}\sigma \mathrm{d}\sigma' \big[2(\Delta_1 \Delta_{1'} - \Delta_2 \Delta_{2'}) G_R - (\Delta_1 + \Delta_2)(\Delta_{1'} - \Delta_{2'}) G_C \big], \tag{7.29a}$$

$$\mathcal{G}^{-+}(x, x') = \frac{1}{4} \int \mathrm{d}\sigma \mathrm{d}\sigma' \big[2(\Delta_1 \Delta_{1'} + \Delta_2 \Delta_{2'}) G_R - (\Delta_1 - \Delta_2)(\Delta_{1'} - \Delta_{2'}) G_C \big]. \tag{7.29b}$$

As for the other components, (7.28) tells us that $\mathcal{G}^{+-}$ can be obtained from (7.29b) by simply relabelling $x \leftrightarrow x'$, while the same argument that allowed us to neglect G_H earlier reveals that we will similarly have no need for $\mathcal{G}^{--}$. For readability, the arguments in (7.29) have been suppressed. The two-point correlation functions all depend on the same arguments, $G \equiv G(\sigma, \sigma')$, and a primed index denotes dependence on primed variables; i.e., $\Delta_1 \equiv \Delta_1(\sigma; x)$ whereas $\Delta_{2'} \equiv \Delta_2(\sigma'; x')$.

Judicious use of (7.B.1) and (7.B.2a) has also been made to express (7.29) only in terms of G_R and G_C (and G_H, which is then discarded). Note that the commutator $G_C(\tau, \tau') := i\langle[q(\tau), q(\tau')]\rangle$ is not independent of G_R, but satisfies the identity

$$G_C^{LL'}(\tau, \tau') = G_R^{LL'}(\tau, \tau') - G_R^{L'L}(\tau', \tau), \tag{7.30}$$

which follows from inverting the definition in (7.7). Thus, once we know G_R, we also know G_C. Indeed, taking the Fourier transform of the above equation yields

$$\widetilde{G}_C(\omega) = \widetilde{G}_R(\omega) - \widetilde{G}_R(-\omega) = 2i\,\mathrm{Im}\,\widetilde{G}_R(\omega). \tag{7.31}$$

7.1.3 Matching the Wilsonian coefficients

We now turn to determine the values of the Wilsonian coefficients $F_\ell^{(n)}$ that appear in the retarded Green's function $G_R^{LL'}(\tau, \tau')$. To make contact with the results in Chapter 6, we ought to compute the scalar-field profile around a black hole. While working with the full fields (g, ϕ) earlier was advantageous for manifestly preserving diffeomorphism invariance, to compute observables, we split

$$\phi_A = \Phi_A + \varphi_A, \qquad g_{A,\mu\nu} = g_{A,\mu\nu} + \frac{h_{A,\mu\nu}}{m_{\rm Pl}}. \tag{7.32}$$

The background fields (g_A, Φ_A) are assumed to be a valid solution to (7.1) that describes an ambient scalar-field profile which persists independently of the black hole. As these background fields will be held fixed, we can immediately set $(g_+, \Phi_+) = (g, \Phi)$ and $(g_-, \Phi_-) = 0$. Being much smaller than the background in which it is embedded, the black hole sources fluctuations (h, φ) in the fields that can be treated perturbatively in the usual way.

At leading order, $\varphi(x)$ is sourced only by terms in S_p that are linear in φ. Moreover, the worldline can be held fixed when computing field expectation values; hence, it suffices to work with the simplified action in (7.17). Decomposing the fields according to (7.32), we find

$$S_p = \sum_{\ell,\ell'} \int \mathrm{d}\tau \mathrm{d}\tau'\, G_R^{LL'}(\tau, \tau') \nabla_L \varphi_-(\tau) \nabla_{L'} \Phi(\tau') + \mathcal{O}(\varphi^2), \tag{7.33}$$

having used (7.19) to simplify terms. Contained in this action is a term proportional to the Wilsonian coefficient $F_0^{(1)}$, which characterises the leading-order, low-frequency dissipative response of the black hole, and is responsible for its induced scalar charge. To see this, let us single out this term, which reads

$$\begin{aligned} S_p \supset & \int \mathrm{d}\tau\, \mathrm{d}\tau' \int \frac{\mathrm{d}\omega}{2\pi}\, F_0^{(1)} (GM)^2 i\omega\, e^{-i\omega(\tau-\tau')} \varphi_-(\tau)\Phi(\tau') \\ = & -\int \mathrm{d}\tau\, F_0^{(1)} (GM)^2 \varphi_-(\tau)\dot{\Phi}(\tau), \end{aligned} \tag{7.34}$$

after integrating by parts. This term sources fluctuations in the scalar field given by

$$\begin{aligned} \varphi(x) = \varphi_+(x)|_{\mathrm{PL}} &= -\int \mathrm{d}\tau\, D_{+-}^{(\varphi)}(x, z(\tau)) F_0^{(1)} (GM)^2 \dot{\Phi}(\tau) \\ &= \frac{F_0^{(1)}}{16\pi} \int \mathrm{d}\tau\, D_R^{(\varphi)}(x, z(\tau)) Q(\tau), \end{aligned} \tag{7.35}$$

where, analogous to (7.18), $D_R^{(\varphi)} \equiv D_{+-}^{(\varphi)}$ is the retarded propagator for the scalar field φ on the background spacetime $g_{\mu\nu}$, and

$$Q(\tau) := -A\dot{\Phi}(z(\tau)) \tag{7.36}$$

is the induced scalar charge of a black hole with area A. Of course, (7.36) is simply a covariantisation of the definition for Q given in Chapter 6. We should expect (7.35) to reduce to the Coulomb-like profile $Q/(4\pi r)$ when viewed in the black hole's rest frame, but more generally, this equation stipulates that a moving, scalar-charged black hole can radiate energy and momentum into scalar waves; thus extending the prediction first due to Horbatsch and Burgess [313] to the case of arbitrary backgrounds $\Phi(x)$.

It remains for us to determine the value of $F_0^{(1)}$. To reproduce the scenario in Chapter 6, we put ourselves in the black hole's rest frame by erecting Fermi normal coordinates $x^\mu = (t, \mathbf{x})$ centred on the black hole [$z^\mu(\tau) = (\tau, \mathbf{0})$ and $\tau \equiv t$ in these coordinates], around which the background is taken to be $\Phi \sim \Phi_0 \cos\omega t$. For a point x in the normal neighbourhood of the origin, $D_R^{(\varphi)}$ is essentially the Klein–Gordon propagator on flat space, up to small corrections associated with the backreaction of Φ onto the spacetime [483,484], which we can neglect. The master integral in Appendix 7.A can then be used to evaluate (7.35). The end result is $\varphi(x) \sim F_0^{(1)} Q(t)/(64\pi^2 r)$ for distances $r \ll \min(\omega^{-1}, \mu^{-1})$, which agrees perfectly with the full-theory calculation of Chapter 6 provided

$$F_0^{(1)} = 16\pi. \tag{7.37}$$

In principle, the same procedure can be repeated for all of the other Wilsonian coefficients. Power counting reveals that the coefficient $F_\ell^{(n)}$ is responsible for effects appearing at order $(GM/\lambda)^\ell (GM\omega)^n$ at the earliest. For example, the coefficient $F_2^{(2)}$ sources an induced quadrupole moment proportional to $(GM)^7 \partial_i \partial_j \ddot{\Phi} \sim (GM/\lambda)^2 (GM\omega)^2 (GM)^3 \Phi$. However, as the full-theory calculation in Chapter 6 is accurate only to first order in $GM\omega$ and $GM\mu$, we do not have sufficient information to determine the value of this coefficient. Indeed, the limitations of Chapter 6 preclude determining the value of any coefficient with $n \geq 2$. This poses no problem when working to low orders in perturbation theory.

The $n = 1$ coefficients can all be determined by following the same procedure that led to (7.37). Moreover, as shown in Appendix 7.B, they have a one-to-one mapping with the objects $C_{\omega\ell m}|_{s=0}$ in Chapter 6, namely,

$$F_\ell^{(1)} = \frac{4\pi C_{\omega\ell m}|_{s=0}}{i\omega \ell!(2\ell-1)!!(GM)^{2\ell+2}} = \frac{16\pi(\ell!)^3 4^\ell}{(2\ell)!(2\ell+1)!(2\ell-1)!!}. \tag{7.38}$$

Taking $\ell = 1$ as an example, the coefficient $F_1^{(1)} = 16\pi/3$ does indeed reproduce the spin-independent dipole moment in (6.35). As for the $n = 0$ coefficients, the vanishing of $C_{\omega\ell m}|_{s=0}$ in the static limit $\omega \to 0$ implies that

$$F_\ell^{(0)} = 0 \quad \forall\, \ell, \tag{7.39}$$

up to possible quadratic-order corrections in $GM\mu$ (since the calculations in Chapter 6 are accurate only to linear order). For $\mu = 0$, these coefficients vanish identically,[1] thus we learn that black holes do not respond at all to static perturbations from a massless scalar.

It is worth remarking that this intriguing property is not unique to scalar fields. It is well known that black holes do not respond to static gravitational perturbations either, as encapsulated by the vanishing of their so-called tidal Love numbers [485–489]. In the EFT, this translates to the vanishing of analogous Wilsonian coefficients that couple the worldline to the gravitational field. This presents an interesting fine-tuning problem, as there is no apparent symmetry in the EFT that would insist on these operators vanishing. (Note that this is unrelated to the no-hair theorems, which only tell us that $\langle q^L \rangle = 0$.) This open problem is discussed further in Refs. [490, 491]. Conceptual issues aside, we will accept (7.39) at face value in what follows, as this is what is required for the EFT to produce the same phenomenological predictions as general relativity in the IR.

[1] Kol and Smolkin arrive at the same result using different methods [485].

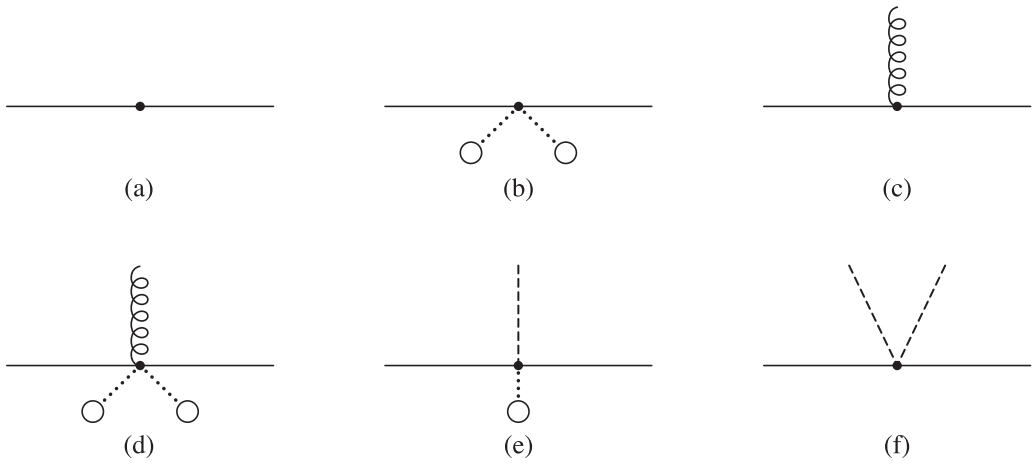


Figure 7.1 Feynman diagrams for a number of interactions localised on the black hole's worldline. The graviton h is drawn as a helical line, the scalar φ as a dashed line, and each insertion of the background scalar Φ is represented by a dotted line terminating in a circle. The solid line stands for the black hole's worldline. The physical processes depicted by each worldline vertex are as follows: (a) kinetic term for the black hole leading to the geodesic equation, (b) correction to the kinetic term due to absorption of the background scalar, (c) a graviton sourced by a black hole of constant mass, (d) correction to the graviton vertex due to mass growth by accretion, (e) induced scalar charge of the black hole, and (f) absorption of φ onto the black hole.

7.2 Worldline vertices

The remainder of this chapter is concerned with the broader phenomenological implications of this EFT. To elucidate the rich physical content that is currently hidden in the correlation functions $\mathcal{G}^{AA'}(x, x')$, we shall substitute the explicit expressions for $\widetilde{G}_R^{LL'}(\omega)$ and $\widetilde{G}_C^{LL'}(\omega)$ into the point-particle action S_p. For simplicity, we will truncate to leading order in the expansion parameters, in which case it suffices to keep only the $F_0^{(1)}$ coefficient. The fact that the $\ell \geq 1$ coefficients do not contribute at this order is the EFT's way of saying that a black hole's absorption cross section is s-wave dominated at low frequencies [492] (i.e., it is the monopolar mode of the field that is being predominantly absorbed). In this limit, the retarded Green's function for $q(\tau)$ is simply

$$G_R(\tau, \tau') = A \int \frac{\mathrm{d}\omega}{2\pi}\, i\omega\, e^{-i\omega(\tau-\tau')}, \tag{7.40}$$

while its commutator G_C is just twice that; cf. (7.31). In fact, when written in this way, (7.40) is valid not only for spherical black holes, but for rotating ones as well.

By substituting these two-point correlation functions into (7.26) and decomposing the fields according to (7.32), one can now write the action S_p as an expansion in powers of (h, φ) and their derivatives. Diagrammatic representations for the first few terms in this series are shown in Figure 7.1.

Scalar terms Those involving the scalar φ all originate from the second term in (7.26), which reads

$$\frac{1}{2}\int \mathrm{d}^4x\,\mathrm{d}^4x'\left[\mathcal{G}^{++}(x,x')\Phi(x)\Phi(x')+2\mathcal{G}^{A+}(x,x')\varphi_A(x)\Phi(x')+\mathcal{G}^{AA'}(x,x')\varphi_A(x)\varphi_{A'}(x')\right] \tag{7.41}$$

after using (7.28) to simplify terms. As $\mathcal{G}^{AA'}$ still depends on the full metric $g=g+h/m_{\rm Pl}$, this can be expanded further to generate an infinite series in powers of h, but we shall here concentrate on the terms in this series that are independent of h.

A discussion of the first term in (7.41), which is represented by the Feynman diagram in Figure 7.1(b), is postponed until Section 7.3. Meanwhile, the second term, which is linear in φ, is responsible for sourcing the black hole's induced scalar charge, and can be rewritten as

$$\text{Fig. 7.1(e)} = \int \mathrm{d}^4x\sqrt{-g}\mathcal{Q}^A(x)\varphi_A(x), \qquad \mathcal{Q}^A(x) := \frac{1}{\sqrt{-g}}\int \mathrm{d}^4x'\,\mathcal{G}_g^{A+}(x,x')\Phi(x'). \tag{7.42}$$

The object $\mathcal{Q}^A(x)$ is the induced charge density, while the subscript on $\mathcal{G}$ instructs us to evaluate the correlation function with respect to the background metric g. Unsurprisingly, these charge densities reduce to

$$\mathcal{Q}_+(x)|_{\rm PL} = \int \mathrm{d}\tau\,\frac{\delta^{(4)}(x-z(\tau))}{\sqrt{-g}}Q(\tau), \qquad \mathcal{Q}_-(x)|_{\rm PL}=0 \tag{7.43}$$

in the physical limit, as shown in Appendix 7.B.

Although not essential for the rest of this thesis, it is interesting to point out that Figure 7.1(f), which corresponds to the third term in (7.41), represents the absorption of the scalar perturbation φ onto the black hole. One arrives at this conclusion after a calculation analogous to the one for Figure 7.1(b).

Graviton terms Two terms in this point-particle action are linear in the metric perturbation $h_{\mu\nu}$, which we will call the "graviton" for short. In both cases, they emerge from expanding the metric contained implicitly in the proper time, $\dot{\tau}_a(g+h/m_{\rm Pl})=\dot{\tau}_a(g)+\delta\dot{\tau}_a+\mathcal{O}(h^2)$. The first-order piece is

$$\delta\dot{\tau}_1(\sigma) = -\frac{1}{2m_{\rm Pl}}\int \mathrm{d}^4x\sqrt{-g}\,h_{1,\mu\nu}(x)t_1^{\mu\nu}(x;\sigma), \tag{7.44}$$

with a similar expression holding for $\delta\dot{\tau}_2$ after relabelling $1\leftrightarrow 2$. Note that

$$t_1^{\mu\nu}(x;\sigma) = \frac{\dot{z}_1^\mu\dot{z}_1^\nu}{\sqrt{-g_{\rho\sigma}(z_1)\dot{z}_1^\rho\dot{z}_1^\sigma}}\frac{\delta^{(4)}(x-z_1(\sigma))}{\sqrt{-g(z_1)}} \tag{7.45}$$

is the contribution to the energy–momentum tensor of a unit point mass when it is at the position $z_1(\sigma)$. Substituting this expansion into the point-mass term $-M\int c^a \dot{\tau}_a \subset S_p$, we get the familiar contribution

$$\text{Fig. 7.1(c)} = \frac{1}{2m_{\rm Pl}} \int \mathrm{d}^4x \sqrt{-g}\, h^a_{\mu\nu}(x) \int \mathrm{d}\sigma\; M\, t_a^{\mu\nu}(x;\sigma), \tag{7.46}$$

which sources the black hole's gravitational potential $\sim GM/r$.

The other term in S_p that is linear in h comes from expanding the metric in

$$\frac{1}{2}\int \mathrm{d}^4x \mathrm{d}^4x'\, \mathcal{G}^{++}(x,x')\Phi(x)\Phi(x'). \tag{7.47}$$

We do so by first substituting (7.29a) into the above expression and integrating over the delta functions contained in Δ_a. Most of the terms will vanish, since G_R is purely dissipative at leading order, and is therefore odd under time reversal. By definition, G_C is also odd under time reversal. Consequently, the only nontrivial part of (7.47) is

$$\frac{1}{2}\int \mathrm{d}\sigma \mathrm{d}\sigma'\, \dot{\tau}_1(\sigma)\dot{\tau}_2(\sigma') G_C(\tau_1(\sigma),\tau_2(\sigma'))\, \Phi(z_1(\sigma))\Phi(z_2(\sigma')). \tag{7.48}$$

We now expand the proper times to first order in h. As this part of the derivation turns out to be rather involved but otherwise uninstructive, the details are relegated to Appendix 7.B. The end result is

$$-\int \mathrm{d}\sigma\, \delta\dot{\tau}^a(\sigma)\big[\delta M_a(\sigma) - \delta M_a(\sigma_f)\big], \tag{7.49}$$

where
$$\delta M_1(\sigma) := A\int_{\sigma_i}^{\sigma} \mathrm{d}\sigma'\, \dot{\Phi}(z_1(\sigma')) \int_{\sigma_i}^{\sigma_f} \mathrm{d}\sigma''\, \dot{\Phi}(z_2(\sigma''))\, \delta(\tau_1(\sigma') - \tau_2(\sigma'')) \tag{7.50}$$

and δM_2 is obtained by interchanging $1 \leftrightarrow 2$ in the above definition. The integration limits (σ_i, σ_f) in these formulae are the initial and final times at which appropriate boundary conditions are specified according to the in–in formalism.

Using the expression for $\delta\dot{\tau}_a$ in (7.44), the first term in (7.49) is

$$\text{Fig. 7.1(d)} = \int \mathrm{d}^4x \sqrt{-g}\, \frac{h_{a,\mu\nu}(x)}{2m_{\rm Pl}} \int \mathrm{d}\sigma\, c^{abc}\delta M_b(\sigma) t_c^{\mu\nu}(x;\sigma). \tag{7.51}$$

By comparing this with (7.46), we see that this vertex couples a graviton to a black hole whose mass is

slowly growing due to absorption of the background scalar. Indeed, in the physical limit, this change in mass is given by

$$\delta M(\tau) = \delta M_+|_{\rm PL} = A\int_{\tau(\sigma_i)}^{\tau} \mathrm{d}\tau'\,\dot{\Phi}^2(z(\tau')), \tag{7.52}$$

which is exactly what we would predict from the full theory by calculating the flux of the scalar across the horizon [303, 309, 311, 312]. What is remarkable here is that we did not put this result in by hand. After performing a matching calculation to reproduce the correct behaviour of the scalar charge, the EFT immediately gives us the correct accretion rate for free. This is evidence that this formalism is working correctly and, more importantly, that the physics governing these two effects are one and the same. Indeed, their magnitudes are both set by the same Wilsonian coefficient $F_0^{(1)} = 16\pi$. Interestingly, this coefficient manifests as a scalar charge when it appears in the retarded propagator G_R, but is responsible for setting the accretion rate when appearing in the commutator G_C. In this light, the relation between a black hole's scalar charge and its accretion rate can be viewed as a special case of the fluctuation–dissipation theorem.

What about the second term in (7.49)? It is a constant contribution to the black hole mass, but one that generically diverges in the limit $\sigma_f \to \infty$. Physically, this singularity is signalling the breakdown of the EFT at late times. This makes intuitive sense, since an increase in the black hole's mass must be compensated for by a concomitant depletion of the surrounding scalar-field environment. Eventually, the black hole will grow to be nearly as massive as its dwindling environment, at which point there is no longer a good separation of scales. We should therefore only trust this EFT for a limited duration of time. Within its period of validity, it is safe to just absorb the constant $\delta M_+(\sigma_f)$ into a renormalisation of the parameter M appearing in the action, such that M represents the mass of the black hole at the point when initial conditions are specified.

Another way to see that this EFT cannot be valid for all times is to differentiate (7.52) to obtain the accretion rate

$$\delta\dot{M}(\tau) = A\dot{\Phi}^2(z(\tau)). \tag{7.53}$$

Notice that the horizon area A appearing on the rhs is that defined at some fixed time. This is only a good approximation provided $\delta M \ll M$. A more precise formula would see the constant A replaced by the instantaneous area $A(\tau)$, but doing it properly would require a resummation involving higher-order terms. Exploring how to do so is an interesting direction for future work, but in practice the systems studied in this thesis will have Φ dilute enough that (7.53) is a good enough approximation.

7.3 Worldline dynamics

Two classes of observables are worth calculating in this EFT: field expectation values, which tell us about gravitational and scalar radiation, and the equation of motion for the worldline. The general method for computing the former has already been discussed in Section 7.1.3. For instance, (7.35) describes the profile of scalar waves radiated at leading order by a black hole travelling along some worldline $z^\mu(\tau)$. To determine the trajectory of this worldline, we must now integrate out the bulk fields entirely to obtain a new effective action

$$\Gamma[z_\pm] = -i \log \int D[h_\pm, \varphi_\pm] \exp(i S_{\rm eff}). \tag{7.54}$$

The equation of motion for $z^\mu(\tau)$ is then obtained from the extremisation condition [469, 480, 481]

$$\left.\frac{\delta\Gamma}{\delta z_-^\mu}\right|_{\rm PL} = 0. \tag{7.55}$$

This new effective action Γ can be represented as a sum over connected Feynman diagrams with no external h or φ legs. The first two terms in this sum are shown in Figures 7.1(a) and 7.1(b). Also in this sum are diagrams involving internal scalar or graviton lines (see, e.g., Figure 7.2), which lead to self-force effects like radiation reaction. That being said, there is little to be gained from discussing these latter kinds of diagrams in generality here, as they are better understood through examples and so are left to be explored in the next chapter. For now, we shall concentrate on Figures 7.1(a) and 7.1(b), whose effects apply universally to black holes embedded in any ambient scalar-field profile. This part of the action reads

$$\Gamma \supset -M \int \mathrm{d}\sigma\, c^a \dot{\tau}_a + \frac{1}{2} \int \mathrm{d}^4x\, \mathrm{d}^4x'\, \mathcal{G}_g^{++}(x, x') \Phi(x) \Phi(x'). \tag{7.56}$$

Note that this effective action is a functional of $z_+ := (z_1 + z_2)/2$ and $z_- := z_1 - z_2$, which give the average and difference of the coordinates of the two worldline copies (z_1, z_2), but do not themselves correspond to worldlines. Of course, the average coordinate tends to a description of the

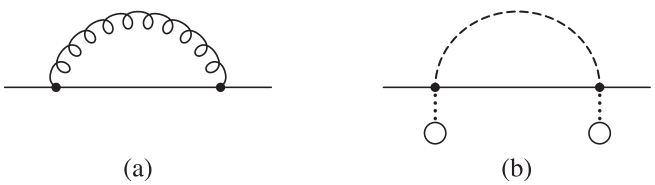


Figure 7.2 Feynman diagrams representing the first-order (a) gravitational and (b) scalar self-force.

physical worldline, $z_+|_{\rm PL} = z$, whereas $z_-|_{\rm PL} = 0$. The latter suggests that we can easily solve (7.55) by Taylor expanding this action in powers of z_- and reading off the linear coefficient. Performing this expansion for $\dot{\tau}_1$ (note $z_1 = z_+ + z_-/2$) yields $\dot{\tau}_1(z_1) = \dot{\tau}_1(z) + \delta\dot{\tau}_1 + \mathcal{O}(z_-^2)$, where

$$\delta\dot{\tau}_1 = \frac{1}{2}\left(a_\mu z_-^\mu - \frac{\mathrm{d}}{\mathrm{d}\tau}(u_\mu z_-^\mu)\right). \tag{7.57}$$

Being interested only in the physical limit, I have already sent $z_+ \to z$ and parametrised it by its proper time τ. The 4-velocity $u^\mu \equiv \mathrm{d}z^\mu/\mathrm{d}\tau$ of this worldline is normalised such that $u_\mu u^\mu = -1$, and $a^\mu := u^\nu \nabla_\nu u^\mu$ denotes its acceleration. The result for $\delta\dot{\tau}_2$ is similar up to the change of sign $z_- \to -z_-$. Using this expansion, the point-mass term in the action gives

$$\Gamma \supset -M \int \mathrm{d}\sigma\, c^a \dot{\tau}_a = -M \int \mathrm{d}\tau\, a_\mu z_-^\mu + \mathcal{O}(z_-^2). \tag{7.58}$$

As for the second term in (7.56), we previously argued in Section 7.2 that it simplifies to

$$\frac{1}{2}\int \mathrm{d}\sigma \mathrm{d}\sigma'\, \dot{\tau}_1(\sigma)\dot{\tau}_2(\sigma')\, G_{\boldsymbol{C}}(\tau_1(\sigma), \tau_2(\sigma'))\, \Phi(z_1(\sigma))\Phi(z_2(\sigma')). \tag{7.59}$$

We now have to expand this to first order in z_-. There are two routes from which z_- emerges: from expanding the proper times $\dot{\tau} \to \dot{\tau} + \delta\dot{\tau}$ and from expanding the arguments of the background scalar Φ. The method for performing the first of these expansions has already been established, with the final result given in (7.49). After renormalising the IR-divergent part, we find

$$\Gamma \supset -\int \mathrm{d}\sigma\, \delta\dot{\tau}^a \delta M_a = -\int \mathrm{d}\tau\, (a_\mu \delta M - u_\mu \delta\dot{M}) z_-^\mu. \tag{7.60}$$

Second, we expand the arguments of Φ and use the antisymmetry property of $G_{\boldsymbol{C}}$ to obtain

$$\Gamma \supset -\frac{1}{2}\int \mathrm{d}\tau \mathrm{d}\tau'\, G_{\boldsymbol{C}}(\tau,\tau')\, \Phi(z(\tau))\partial_\mu \Phi(z(\tau'))\, z_-^\mu = \int \mathrm{d}\tau\, Q(\tau)\partial_\mu \Phi(z(\tau))\, z_-^\mu. \tag{7.61}$$

Finally, combining (7.58), (7.60), and (7.61), we learn that the equation of motion for this black hole's worldline (neglecting self-force effects) is

$$[M + \delta M(\tau)]\, a^\mu = -\delta\dot{M}(\tau)\, u^\mu + Q(\tau)\, g^{\mu\nu}\partial_\nu \Phi. \tag{7.62}$$

The terms involving δM administer a drag force due to accretion, whereas the remaining term involving a derivative on Φ must be interpreted as a scalar "fifth" force. This second term is somewhat odd given

what we learnt in Part I. There, the fifth force appeared in the equation of motion as $Q(g^{\mu\nu}+u^\mu u^\nu)\partial_\nu\Phi$. In fact, (7.62) can easily be put into such a form, since $\delta\dot{M} = A\dot{\Phi}^2 = -Qu^\nu\partial_\nu\Phi$ by definition. Thus, an equivalent way of writing (7.62) is

$$[M + \delta M(\tau)]a^\mu = Q(\tau)(g^{\mu\nu} + u^\mu u^\nu)\partial_\nu\Phi. \tag{7.63}$$

In Section 7.2, we saw that the physics of the scalar charge and of accretion were one and the same, having both emerged from the same parent term in the point-particle action. Here, this connection is made manifest at the level of the equations of motion: The scalar fifth force associated with this charge includes the drag force due to accretion, but additionally predicts that a black hole's trajectory will also be influenced by scalar-field gradients. This result reveals that a black hole embedded in an ambient scalar-field profile behaves in a similar way to how scalarised compact objects behave in alternative theories of gravity. There are two key distinctions, however. Whereas the fifth force mediated between permanently charged objects in a scalar–tensor theory is the result of a fundamental interaction, the fifth force in (7.63) is an emergent phenomenon that arises dynamically due to the absorptive nature of the black hole's horizon. Second, the phenomenology in this latter case can generally be much richer due to the spacetime dependence of $Q(\tau)$. We will explore this in greater detail in the next two chapters.

Appendix 7.A Master integrals

Many results in the main text require the solution for a Klein–Gordon field generated by a point source. The general procedure for performing such a calculation is reviewed in this appendix.

We begin by considering a monopolar, time-dependent source $J(t)$ localised at the origin. The spacetime everywhere else is taken to be flat. The scalar-field profile generated by this source is given by the integral

$$\varphi(x) = \int \mathrm{d}t'\; D_R^{(\varphi)}(x,x')J(t')\Big|_{\mathbf{x}'=\mathbf{0}}, \tag{7.A.1}$$

where the retarded propagator for the scalar field of mass μ can be written as

$$D_R^{(\varphi)}(x,x') = \int \frac{\mathrm{d}^4p}{(2\pi)^4}\,\frac{e^{ip\cdot(x-x')}}{-(p^0+i\epsilon)^2+\mathbf{p}^2+\mu^2}. \tag{7.A.2}$$

Substituting (7.A.2) into (7.A.1) and writing $p^0 \equiv \omega$, we obtain

$$\begin{aligned}\varphi(x) &= \int \frac{\mathrm{d}\omega}{2\pi} \int \mathrm{d}t' J(t') e^{-i\omega(t-t')} \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3} \frac{e^{i\mathbf{p}\cdot\mathbf{x}}}{\mathbf{p}^2 + \mu^2 - (\omega + i\epsilon)^2} \\ &= \int \frac{\mathrm{d}\omega}{2\pi} \int \mathrm{d}t' J(t') e^{-i\omega(t-t')} \left(\frac{-i}{2\pi r}\right) \int_{-\infty}^{\infty} \frac{\mathrm{d}p}{2\pi} \frac{p e^{ipr}}{p^2 - k^2} \end{aligned} \tag{7.A.3}$$

after transforming to spherical coordinates and integrating over the angular variables.

As $r > 0$, the remaining integral over p can be performed by closing the integration contour in the upper half of the complex-p plane. Which of the poles we encircle depends on the sign of the real part of $k^2 \equiv (\omega + i\epsilon)^2 - \mu^2$. When $\mathrm{Re}(k^2) < 0$, we pick up the pole at $p = i\sqrt{\mu^2 - \omega^2}$, which leads to a Yukawa-like potential. If instead $\mathrm{Re}(k^2) \geq 0$, there are two poles along the real line,

$$p = \pm \left(\sqrt{\omega^2 - \mu^2} + \frac{i\omega\epsilon}{\omega^2 - \mu^2} + \mathcal{O}(\epsilon^2) \right), \tag{7.A.4}$$

which correspond to spherical waves. To enforce retarded boundary conditions, the poles have been shifted off the real axis by an amount $\propto i\epsilon$. For a given value of ω, only one of these poles will be enclosed by the integration contour. Finally, note that the special case $k = 0$ is also included in this prescription as the limit $\omega \to \pm\mu^{\pm}$.

Putting everything together, the end result can be written in the particularly simple form

$$\varphi(x) = \frac{1}{4\pi r} \int \frac{\mathrm{d}\omega}{2\pi} e^{-i\omega t + ik(\omega)r} \int \mathrm{d}t' J(t') e^{i\omega t'}, \tag{7.A.5}$$

provided we define the momentum in the radial direction by

$$k(\omega) := \begin{cases} i\sqrt{\mu^2 - \omega^2} & (\omega^2 < \mu^2) \\ \mathrm{sgn}(\omega)\sqrt{\omega^2 - \mu^2} & (\omega^2 \geq \mu^2). \end{cases} \tag{7.A.6}$$

It is worth remarking that this same master integral can also be used to determine the (linearised) metric perturbation $h_{\mu\nu}$ generated by a point source, since its retarded propagator has a similar form, up to some index structure and setting $\mu = 0$.

Let us now consider a more general scenario in which there are a number of multipolar sources $J^L(t)$ localised at the origin. The corresponding integral for the scalar-field profile is

$$\varphi(x) = \int \mathrm{d}^4x' \, D_R^{(\varphi)}(x, x') \sum_{\ell=0}^{\infty} (-1)^\ell \frac{\partial}{\partial x'^L} \delta^{(3)}(\mathbf{x}') J^L(t'). \tag{7.A.7}$$

Integrating by parts to move the spatial derivatives onto the propagator and using the fact that it is Lorentz invariant to write $\partial D_R^{(\varphi)}/\partial x'^{\mu} = -\partial D_R^{(\varphi)}/\partial x^{\mu}$, this simplifies to

$$\varphi(x) = \sum_{\ell}(-1)^{\ell}\frac{\partial}{\partial x^{L}}\int \mathrm{d}t'\, D_R^{(\varphi)}(x,x')\, J^{L}(t')\Big|_{\mathbf{x}'=\mathbf{0}}. \tag{7.A.8}$$

The result in (7.A.5) can now be used to show that

$$\varphi(x) = \sum_{\ell}(-1)^{\ell}\partial_{L}\left(\frac{1}{4\pi r}\int\frac{\mathrm{d}\omega}{2\pi}\, e^{-i\omega t + ik(\omega)r}\int \mathrm{d}t'\, J^{L}(t')e^{i\omega t'}\right). \tag{7.A.9}$$

To proceed, we will assume that the tensors $J^{L}(t)$ are all symmetric and trace free (STF), such that each one corresponds to a distinct multipole. We now introduce the set of tensors $\mathcal{Y}^{\ell m}_{L} \equiv \mathcal{Y}^{\ell m}_{i_1\cdots i_\ell}$ labelled by the integer $m \in \{-\ell, -\ell+1, \ldots, \ell\}$, which form a basis for the linear vector space of STF tensors of rank ℓ. They also generate the spherical harmonics, $Y_{\ell m}(\hat{\mathbf{x}}) = \mathcal{Y}^{\ell m}_{L}\hat{\mathbf{x}}^{L}$, and satisfy the orthogonality relation [471]

$$(\mathcal{Y}^{\ell m}_{L})^{*}\mathcal{Y}^{\ell m'}_{L} = \frac{(2\ell+1)!!}{4\pi\ell!}\delta^{mm'}. \tag{7.A.10}$$

These properties can be used to expand any STF tensor in terms of this basis, thus we can write

$$J^{L}(t) = -4\pi i\sum_{m}\mathcal{Y}^{\ell m}_{L}J_{\ell m}(t), \qquad J_{\ell m}(t) \equiv \frac{i\ell!}{(2\ell+1)!!}(\mathcal{Y}^{\ell m}_{L})^{*}J^{L}(t), \tag{7.A.11}$$

where the factor of $-4\pi i$ is included purely for convenience. Using this decomposition, we have that

$$\varphi(x) = \int\frac{\mathrm{d}\omega}{2\pi}\, e^{-i\omega t}\sum_{\ell,m}(-1)^{\ell}\mathcal{Y}^{\ell m}_{L}\partial_{L}\left(\frac{e^{ikr}}{ir}\right)\int \mathrm{d}t'\, J_{\ell m}(t')\, e^{i\omega t'}. \tag{7.A.12}$$

A final set of identities can be used to put this result into a form that makes its physical significance manifest. First, we use the fact that [448]

$$\mathcal{Y}^{\ell m}_{L}\partial_{L}f(r) = \mathcal{Y}^{\ell m}_{L}\hat{\mathbf{x}}^{L}r^{\ell}\left(\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\right)^{\ell}f(r) \equiv Y_{\ell m}(\hat{\mathbf{x}})\, r^{\ell}\left(\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\right)^{\ell}f(r) \tag{7.A.13}$$

for any spherical function $f(r)$. Then noting that the spherical Hankel functions can be defined by [471]

$$h^{\pm}_{\ell}(z) := z^{\ell}(-1)^{\ell}\left(\frac{1}{z}\frac{\mathrm{d}}{\mathrm{d}z}\right)^{\ell}\left(\frac{e^{\pm iz}}{\pm iz}\right), \tag{7.A.14}$$

we conclude that

$$\varphi(x) = \sum_{\ell,m} \int \frac{\mathrm{d}\omega}{2\pi} e^{-i\omega t} \mathcal{A}_{\omega\ell m} h^{+}_{\ell}(kr) Y_{\ell m}(\hat{\mathbf{x}}), \tag{7.A.15}$$

where

$$\mathcal{A}_{\omega\ell m} = k^{\ell+1} \int \mathrm{d}t' J_{\ell m}(t') e^{i\omega t'}. \tag{7.A.16}$$

In this form, the solution can be seen to be given by a linear superposition of outgoing waves and Yukawa-like potentials, with the amplitude of each mode set by the coefficient $\mathcal{A}_{\omega\ell m}$.

Appendix 7.B Deriving the effective action

In constructing the effective point-particle action, many of the more technical aspects of the derivation were sidestepped in the interest of readability. Those details are presented here.

7.B.1 Review of two-point correlation functions

Let us begin by reviewing a few key properties of two-point correlation functions. Consider a dynamical system comprised of a set of Hermitian operators $\{q^L(\tau)\}$ living on a one-dimensional worldline. We will assume that the system is "free" in the sense that its dynamics is fully characterised by its two-point correlation functions, which are defined as follows:

Positive-frequency Wightman function	$-iG^{LL'}_{+}(\tau,\tau') := \langle q^L(\tau) q^{L'}(\tau')\rangle,$	(7.B.1a)
Negative-frequency Wightman function	$-iG^{LL'}_{-}(\tau,\tau') := \langle q^{L'}(\tau') q^{L}(\tau)\rangle,$	(7.B.1b)
Feynman propagator	$-iG^{LL'}_{F}(\tau,\tau') := \langle T q^L(\tau) q^{L'}(\tau')\rangle,$	(7.B.1c)
Dyson propagator	$-iG^{LL'}_{D}(\tau,\tau') := \langle T^* q^L(\tau) q^{L'}(\tau')\rangle,$	(7.B.1d)
Hadamard propagator	$-iG^{LL'}_{H}(\tau,\tau') := \langle \{q^L(\tau), q^{L'}(\tau')\}\rangle,$	(7.B.1e)
Commutator	$-iG^{LL'}_{C}(\tau,\tau') := \langle [q^L(\tau), q^{L'}(\tau')]\rangle,$	(7.B.1f)
Retarded propagator	$G^{LL'}_{R}(\tau,\tau') := \theta(\tau-\tau') G^{LL'}_{C}(\tau,\tau'),$	(7.B.1g)
Advanced propagator	$G^{LL'}_{A}(\tau,\tau') := -\theta(\tau'-\tau) G^{LL'}_{C}(\tau,\tau').$	(7.B.1h)

Note that $\theta(x)$ is the Heaviside step function, while T and T^* denote the time-ordering and anti-time-ordering operators, respectively. Notice also from their definitions that not all of these two-point functions are independent. In particular,

$$G^{LL'}_{+}(\tau,\tau') = G^{L'L}_{-}(\tau',\tau), \tag{7.B.2a}$$

while the identity $\theta(x) + \theta(-x) = 1$ can be used to show that

$$G_R = G_F - G_- = G_+ - G_D, \tag{7.B.2b}$$

$$G_A = G_F - G_+ = G_- - G_D, \tag{7.B.2c}$$

$$G_H = G_F + G_D = G_+ + G_-. \tag{7.B.2d}$$

In these last three equations, all two-point functions have the same indices LL' and arguments (τ, τ'), which have been suppressed for readability.

To make contact with the path-integral approach adopted in the main text, it is useful to see how these correlation functions can be obtained from a generating functional. We start by considering how the vacuum $|0\rangle$ of the free theory evolves under the influence of external perturbations. In fact, suppose that we have two identical copies of this system, and we couple each one to a different set of external sources, say $J_1^L(\tau)$ and $J_2^L(\tau)$. In the interaction picture, the vacuum state of the first copy at some time τ is given by Dyson's formula,

$$|\Omega_1(\tau)\rangle = T \exp\left(i \sum_\ell \int_{-\infty}^{\tau} \mathrm{d}\tau'\, q^L(\tau') J_{1,L}(\tau') \right) |0\rangle\,, \tag{7.B.3}$$

and a similar expression holds for the vacuum state $|\Omega_2(\tau)\rangle$ of the second copy. An interesting quantity to calculate is the overlap between these two vacuum states. Making the comparison in the asymptotic future $\tau \to \infty$ defines for us the generating functional

$$\begin{aligned} \mathcal{Z}[J_1, J_2] &:= \lim_{\tau\to\infty} \langle\, \Omega_2(\tau) \,|\, \Omega_1(\tau) \,\rangle \\ &= \langle 0|\, T^* \exp\left(- i \sum_{\ell'} \int_{-\infty}^{\infty} \mathrm{d}\tau'\, q^{L'}(\tau') J_{2,L'}(\tau') \right) \\ &\quad \times T \exp\left(i \sum_\ell \int_{-\infty}^{\infty} \mathrm{d}\tau\, q^L(\tau) J_{1,L}(\tau) \right) |0\rangle\,, \end{aligned} \tag{7.B.4}$$

whose functional derivatives give us vacuum expectation values of $q^L(\tau)$; for instance,

$$(-i)^2 \frac{\delta}{\delta J_L^1(\tau)} \frac{\delta}{\delta J_{L'}^1(\tau')} \mathcal{Z}[J_1, J_2] \bigg|_{J_1 = J_2 = 0} = \langle 0|\, T q^L(\tau) q^{L'}(\tau') \,|0\rangle\,. \tag{7.B.5}$$

It is now straightforward to generalise this procedure so as to obtain the Green's functions in (7.B.1), whose expectation values are defined with respect to some density matrix $\hat{\rho}$; i.e., $\langle \cdots \rangle \equiv \mathrm{tr}(\cdots \hat{\rho})$.

Introducing a new generating functional

$$\mathcal{Z}[J_1, J_2; \hat{\rho}] := \mathrm{tr}\left\{ T^* \exp\left(-i \sum_{\ell'} \int_{-\infty}^{\infty} \mathrm{d}\tau' \, q^{L'}(\tau') J_{2,L'}(\tau') \right) \right.$$
$$\left. \times T \exp\left(i \sum_{\ell} \int_{-\infty}^{\infty} \mathrm{d}\tau \, q^{L}(\tau) J_{1,L}(\tau) \right) \hat{\rho} \right\} \tag{7.B.6}$$

and using the CTP index notation introduced earlier, one can easily verify that

$$-i G^{LL'}_{ab}(\tau, \tau') := (-i)^2 \frac{\delta}{\delta J^a_L(\tau)} \frac{\delta}{\delta J^b_{L'}(\tau')} \mathcal{Z}[J_1, J_2; \hat{\rho}] \Big|_{J_1 = J_2 = 0}$$
$$= \begin{pmatrix} -i G^{LL'}_F(\tau, \tau') & -i G^{LL'}_-(\tau, \tau') \\ -i G^{LL'}_+(\tau, \tau') & -i G^{LL'}_D(\tau, \tau') \end{pmatrix}. \tag{7.B.7}$$

This matrix can then be transformed into the Keldysh representation by using the transformation rule $G^{LL'}_{AB}(\tau, \tau') = \Lambda_A{}^a \Lambda_B{}^b G^{LL'}_{ab}(\tau, \tau')$, where the Jacobian $\Lambda_A{}^a$ is given in (7.13). The result in (7.18) then follows after also using the identities in (7.B.2a).

To complete this discussion, note that the generating functional $\mathcal{Z}[J_1, J_2; \hat{\rho}]$ also admits a representation in terms of a path integral [477,479]. Just as we did with the sources, we double the set of operators $q^L \mapsto (q^L_1, q^L_2)$ in order to write

$$\mathcal{Z}[J_1, J_2; \hat{\rho}] = \int D[q^L_1, q^L_2] \exp\left(i \int \mathrm{d}\tau \, \mathcal{L}(q^L_1, q^L_2) \right.$$
$$\left. + i \sum_{\ell} \int \mathrm{d}\tau \, \left[q^L_1(\tau) J_{1,L}(\tau) - q^L_2(\tau) J_{2,L}(\tau) \right] \right), \tag{7.B.8}$$

where $\mathcal{L}$ is the Lagrangian for this system and the minus sign in front of the $q^L_2 J_{2,L}$ term is needed to be consistent with the minus sign in (7.B.6). In the context of Section 7.1, the sources $J^a_L(\tau) \equiv \nabla_L \phi^a(\tau)$ are identified as the multipoles of a bulk scalar field centred on the black hole. The terms in the exponent above then correspond to the scalar sector of the intermediary point-particle action I_p; cf. (7.15).

7.B.2 A generalised matching calculation

In this part of the appendix, we will prove the formula in (7.38) for the Wilsonian coefficients $F^{(1)}_\ell$. What follows is a straightforward generalisation of the discussion in Section 7.1.3 to arbitrary ℓ. To reproduce the scenario in Chapter 6, we put ourselves in the black hole's rest frame by erecting Fermi normal coordinates $x^\mu = (t, \mathbf{x})$ centred on the black hole's position. In a small neighbourhood about the

origin, any generic background field $\Phi(x)$ can be written in terms of the following mode decomposition:

$$\Phi(x) \sim \mathrm{Re}\sum_{\ell,m}\int \frac{\mathrm{d}\omega}{2\pi}\, e^{-i\omega t}\Phi_{\omega\ell m} r^{\ell} Y_{\ell m}(\hat{\mathbf{x}}). \tag{7.B.9}$$

As always, the length and timescales of this background are assumed to be much greater than those of the black hole, in which case, $\Phi_{\omega\ell m} \sim 1/\lambda^{\ell}$ with $\lambda \gg GM$. Moreover, there must exist a maximum frequency $\omega_{\max} \ll (GM)^{-1}$ such that $\Phi_{\omega\ell m} = 0$ for $|\omega| \geq \omega_{\max}$. The expansion in (7.B.9) is then valid for distances $GM \ll r \ll \min(\omega_{\max}^{-1}, \mu^{-1})$. Notice that this is simply a generalisation of the type of backgrounds considered in Chapter 6.

This background generates induced multipole moments via the interaction terms

$$\begin{aligned} S_p &\supset \sum_{\ell=0}^{\infty}\int \mathrm{d}\tau\mathrm{d}\tau' G_R^{LL'}(\tau,\tau')\nabla_L\varphi_-(\tau)\nabla_{L'}\Phi(\tau') \\ &\supset \sum_{\ell=0}^{\infty}\int \mathrm{d}\tau\mathrm{d}\tau' \int \frac{\mathrm{d}\omega}{2\pi}\, e^{-i\omega(\tau-\tau')} F_{\ell}^{(1)}(GM)^{2\ell+2} i\omega\, \nabla_L\varphi_-(\tau)\nabla_L\Phi(\tau') \\ &= -\sum_{\ell=0}^{\infty}\int \mathrm{d}\tau\mathrm{d}\tau' F_{\ell}^{(1)}(GM)^{2\ell+2}\, \nabla_L\varphi_-(\tau)\nabla_L\dot{\Phi}(\tau), \end{aligned} \tag{7.B.10}$$

where I have kept only the leading, nontrivial coefficient for each ℓ. These terms source fluctuations in the scalar field given by

$$\varphi(x) = -\sum_{\ell=0}^{\infty}\int \mathrm{d}^4y\, F_{\ell}^{(1)}(GM)^{2\ell+2}\frac{\partial D_R^{(\varphi)}(x,y)}{\partial y^L}\frac{\partial\dot{\Phi}(y)}{\partial y^L}\delta^{(3)}(\mathbf{y}), \tag{7.B.11}$$

where $\tau \equiv y^0$ in Fermi normal coordinates, since the metric at the origin is flat. For the same reason, $D_R^{(\varphi)}$ is essentially the flat-space Klein–Gordon propagator when expressed in these coordinates, provided x is in the normal neighbourhood of y [483, 484]. Now using the fact that $D_R^{(\varphi)}$ is Lorentz invariant, (7.B.11) can be simplified to yield

$$\varphi(x) = \sum_{\ell=0}^{\infty}(-1)^{\ell}\frac{\partial}{\partial x^L}\int \mathrm{d}^4y\, D_R^{(\varphi)}(x,y)J^L(y^0)\delta^{(3)}(\mathbf{y}), \tag{7.B.12}$$

where

$$\begin{aligned} J^L(t) &= -F_{\ell}^{(1)}(GM)^{2\ell+2}\partial_L\dot{\Phi}(t,\mathbf{0}) \\ &= F_{\ell}^{(1)}\ell!(GM)^{2\ell+2}\,\mathrm{Re}\sum_m\int\frac{\mathrm{d}\omega}{2\pi}\, i\omega\, e^{-i\omega t}\Phi_{\omega\ell m}\mathcal{Y}_L^{\ell m} \end{aligned} \tag{7.B.13}$$

are the induced multipole moments of the black hole. For $\ell = 0$, one can easily verify that the induced

scalar charge is recovered. Written in this form, (7.B.12) can be evaluated using the master integral in Appendix 7.A, which returns

$$\varphi(x) = \mathrm{Re}\sum_{\ell,m}\int\frac{\mathrm{d}\omega}{2\pi}\,e^{-i\omega t}\Phi_{\omega\ell m}\left[-\frac{\omega}{4\pi}F_\ell^{(1)}\ell!(GM)^{2\ell+2}k^{\ell+1}h_\ell^+(kr)\right]Y_{\ell m}(\hat{\mathbf{x}}). \tag{7.B.14}$$

As we are allowed to use the flat-space propagator only when x is in the normal neighbourhood of the origin, this result is valid only for distances $kr \ll 1$. Using the fact that the spherical Hankel functions have the asymptotic form $h_\ell^+(z) \sim -i(2\ell-1)!!z^{-\ell-1}$ when $z \to 0$ [397], we have that

$$\varphi(x) \sim \mathrm{Re}\sum_{\ell,m}\int\frac{\mathrm{d}\omega}{2\pi}e^{-i\omega t}\Phi_{\omega\ell m}\left[\frac{i\omega}{4\pi}F_\ell^{(1)}\ell!(2\ell-1)!!(GM)^{2\ell+2}r^{-\ell-1}\right]Y_{\ell m}(\hat{\mathbf{x}}). \tag{7.B.15}$$

Finally, noting that the full scalar-field profile around the black hole is $\phi(x) = \Phi(x) + \varphi(x)$, identifying the expression in square brackets above with the second term in (6.25) gives us the desired result:

$$F_\ell^{(1)} = \frac{4\pi C_{\omega\ell m}|_{s=0}}{i\omega\ell!(2\ell-1)!!(GM)^{2\ell+2}}. \tag{7.B.16}$$

7.B.3 Worldline vertex functions

This part of the appendix supplements the discussion in Section 7.2 by deriving explicit expressions for the black hole's charge density and accretion rate.

Charge density In (7.42), we defined the black hole's induced charge density as

$$\mathcal{Q}^A(x) := \frac{1}{\sqrt{-g}}\int \mathrm{d}^4x'\,\mathcal{G}_g^{A+}(x,x')\Phi(x'). \tag{7.B.17}$$

To prove the end result in (7.43), all we have to do is substitute in explicit expressions for $\mathcal{G}_g^{A+}$ and simplify. Using (7.29a) and integrating over the delta functions in $\Delta_a(\sigma';x')$, one finds

$$\begin{aligned}\sqrt{-g}\mathcal{Q}^+(x) = \int \mathrm{d}\sigma\mathrm{d}\sigma'\Big\{&\big[G_R(\sigma,\tau_1(\sigma'))\Delta_1(\sigma;x) - G_C(\sigma,\tau_1(\sigma'))\Delta_+(\sigma;x)\big]\,\dot\tau_1(\sigma')\Phi(z_1(\sigma'))\\ &-\big[G_R(\sigma,\tau_2(\sigma'))\Delta_2(\sigma;x) - G_C(\sigma,\tau_2(\sigma'))\Delta_+(\sigma;x)\big]\,\dot\tau_2(\sigma')\Phi(z_2(\sigma'))\Big\},\end{aligned} \tag{7.B.18}$$

having defined $\Delta_+ := (\Delta_1 + \Delta_2)/2$.

Now use the explicit expressions for G_R and G_C in (7.31) and (7.40) to obtain

$$\sqrt{-g}\mathcal{Q}^+(x) = A\int \mathrm{d}\sigma\mathrm{d}\sigma' \int \frac{\mathrm{d}\omega}{2\pi}\big[\Delta_1(\sigma;x)\dot{\tau}_2(\sigma')\, i\omega\, e^{-i\omega[\sigma-\tau_2(\sigma')]}\Phi(z_2(\sigma'))-(1\leftrightarrow 2)\big]. \quad (7.B.19)$$

Then recognising that $\dot{\tau}_2(\sigma')i\omega$ can be rewritten as a derivative $\mathrm{d}/\mathrm{d}\sigma'$ acting on the exponential, and likewise for $1\leftrightarrow 2$, we find

$$\sqrt{-g}\mathcal{Q}^+(x) = -A\int \mathrm{d}\sigma\mathrm{d}\sigma'\big[\Delta_1(\sigma;x)\dot{\Phi}(z_2(\sigma'))\,\delta(\sigma-\tau_2(\sigma'))-(1\leftrightarrow 2)\big] \quad (7.B.20)$$

after integrating by parts. Finally, integrating over the remaining delta functions in $\Delta_a(\sigma;x)$ yields

$$\mathcal{Q}^+(x) = -A\int \mathrm{d}\sigma \frac{\delta^{(4)}(x-z_1(\sigma))}{\sqrt{-g}}\dot{\tau}_1(\sigma)\int \mathrm{d}\sigma'\dot{\Phi}(z_2(\sigma'))\,\delta(\sigma-\tau_2(\sigma'))-(1\leftrightarrow 2). \quad (7.B.21)$$

Also repeating analogous steps for $\mathcal{Q}^-(x)$, the following pattern becomes apparent: If we define

$$\mathcal{Q}_1(x) := -A\int \mathrm{d}\sigma \frac{\delta^{(4)}(x-z_1(\sigma))}{\sqrt{-g}}\dot{\tau}_1(\sigma)\int \mathrm{d}\sigma'\dot{\Phi}(z_2(\sigma'))\,\delta(\tau_1(\sigma)-\tau_2(\sigma')) \quad (7.B.22)$$

and define $\mathcal{Q}_2(x)$ by interchanging $1\leftrightarrow 2$ in the above equation, then the charge densities $\mathcal{Q}^{\mp}\equiv\mathcal{Q}_{\pm}$ in the Keldysh representation are obtained through the usual transformation rule; i.e., $\mathcal{Q}_+ = (\mathcal{Q}_1+\mathcal{Q}_2)/2$ and $\mathcal{Q}_- = \mathcal{Q}_1-\mathcal{Q}_2$. The result in (7.43) then follows immediately.

Accretion rate To derive the accretion rate as given in (7.50), our starting point is the term

$$S_p \supset \frac{1}{2}\int \mathrm{d}\sigma\mathrm{d}\sigma'\dot{\tau}_1(\sigma)\dot{\tau}_2(\sigma')\,G_C(\tau_1(\sigma),\tau_2(\sigma'))\,\Phi(z_1(\sigma))\Phi(z_2(\sigma')) \quad (7.B.23)$$

in the point-particle action. As we did in the main text, we now perturb the proper time by replacing $\tau_a\to\tau_a+\delta\tau_a$ in the above expression. The terms that are linear in $\delta\tau_a$ are

$$\frac{1}{2}\int \mathrm{d}\sigma\mathrm{d}\sigma'\dot{\tau}_1(\sigma)\dot{\tau}_2(\sigma')\,G_C(\tau_1(\sigma),\tau_2(\sigma'))\,\Phi(z_1(\sigma))\Phi(z_2(\sigma'))\left(\frac{\delta\dot{\tau}_1(\sigma)}{\dot{\tau}_1(\sigma)}+\frac{\delta\dot{\tau}_2(\sigma')}{\dot{\tau}_2(\sigma')}\right.$$

$$\left.+\frac{\mathrm{d}\log G_C(\tau_1(\sigma),\tau_2(\sigma'))}{\mathrm{d}\tau_1}\int_{\sigma_i}^{\sigma}\mathrm{d}\widetilde{\sigma}\,\delta\dot{\tau}_1(\widetilde{\sigma})+\frac{\mathrm{d}\log G_C(\tau_1(\sigma),\tau_2(\sigma'))}{\mathrm{d}\tau_2}\int_{\sigma_i}^{\sigma'}\mathrm{d}\widetilde{\sigma}\,\delta\dot{\tau}_2(\widetilde{\sigma})\right). \quad (7.B.24)$$

Substituting the explicit expression for G_C into the above expression then yields

$$A\int \mathrm{d}\sigma\mathrm{d}\sigma'\Phi(z_1(\sigma))\Phi(z_2(\sigma'))\int\frac{\mathrm{d}\omega}{2\pi}e^{-i\omega[\tau_1(\sigma)-\tau_2(\sigma')]}\bigg(\delta\dot\tau_1(\sigma)\dot\tau_2(\sigma')i\omega+\delta\dot\tau_2(\sigma')\dot\tau_1(\sigma)i\omega$$

$$-\dot\tau_1(\sigma)\dot\tau_2(\sigma')(i\omega)^2\int_{\sigma_i}^{\sigma}\mathrm{d}\widetilde\sigma\,\delta\dot\tau_1(\widetilde\sigma)+\dot\tau_1(\sigma)\dot\tau_2(\sigma')(i\omega)^2\int_{\sigma_i}^{\sigma'}\mathrm{d}\widetilde\sigma\,\delta\dot\tau_2(\widetilde\sigma)\bigg). \tag{7.B.25}$$

Just as we did when deriving the charge density, notice that each factor of $\dot\tau_1(\sigma)i\omega$ can be replaced by a derivative $-\mathrm{d}/\mathrm{d}\sigma$ acting on the exponential, and likewise each factor of $\dot\tau_2(\sigma')i\omega$ can be replaced by $\mathrm{d}/\mathrm{d}\sigma'$. Having done so, (7.B.25) can be written more succinctly as

$$A\int \mathrm{d}\sigma\mathrm{d}\sigma'\Phi(z_1(\sigma))\Phi(z_2(\sigma'))\bigg[\frac{\mathrm{d}}{\mathrm{d}\sigma}\bigg(\int_{\sigma_i}^{\sigma}\mathrm{d}\widetilde\sigma\,\delta\dot\tau_1(\widetilde\sigma)\frac{\mathrm{d}}{\mathrm{d}\sigma'}\bigg)$$

$$-\frac{\mathrm{d}}{\mathrm{d}\sigma'}\bigg(\int_{\sigma_i}^{\sigma'}\mathrm{d}\widetilde\sigma\,\delta\dot\tau_2(\widetilde\sigma)\frac{\mathrm{d}}{\mathrm{d}\sigma}\bigg)\bigg]\delta(\tau_1(\sigma)-\tau_2(\sigma')) \tag{7.B.26}$$

after also integrating by parts. Integrating by parts once more then yields

$$A\int_{\sigma_i}^{\sigma_f}\mathrm{d}\sigma\,\dot\Phi(z_1(\sigma))\int_{\sigma_i}^{\sigma}\mathrm{d}\widetilde\sigma\,\delta\dot\tau_1(\widetilde\sigma)\int_{\sigma_i}^{\sigma_f}\mathrm{d}\sigma'\dot\Phi(z_2(\sigma'))\,\delta(\tau_1(\sigma)-\tau_2(\sigma'))-(1\leftrightarrow 2). \tag{7.B.27}$$

Note that (σ_i,σ_f) correspond to the initial and final times at which boundary conditions are to be specified according to the in–in formalism. The final result in (7.50) is obtained after swapping the integration limits on σ and $\widetilde\sigma$ by using the identity

$$\int_{\sigma_i}^{\sigma_f}\mathrm{d}\sigma\int_{\sigma_i}^{\sigma}\mathrm{d}\widetilde\sigma=\int_{\sigma_i}^{\sigma_f}\mathrm{d}\widetilde\sigma\int_{\widetilde\sigma}^{\sigma_f}\mathrm{d}\sigma=\int_{\sigma_i}^{\sigma_f}\mathrm{d}\widetilde\sigma\left(\int_{\sigma_i}^{\sigma_f}\mathrm{d}\sigma-\int_{\sigma_i}^{\widetilde\sigma}\mathrm{d}\sigma\right). \tag{7.B.28}$$

Chapter 8
Binary black holes in fuzzy dark matter halos

We have thus far learnt that a black hole embedded in an ambient scalar field experiences an emergent fifth force and gains the ability to radiate scalar waves due to the onset of an induced scalar charge. To attach some quantitative predictions to these ideas, it will be instructive to consider a concrete example of such a system. This chapter explores the impact of a galactic fuzzy dark matter (FDM) halo on the evolution of a binary black hole. The calculations to be carried out will be valid for astrophysical black holes of any size, although the effects will be largest for supermassive black holes, since the scalar charge Q grows with the black hole's area A.

Galactic halos in FDM models consist of a central, (pseudo)solitonic core surrounded by an envelope of fluctuating density granules that arise from wave interference [493–496]. The core resists further gravitational collapse by coherently oscillating in time at a frequency that is approximately equal to the scalar's mass μ, up to small corrections from a nonrelatistivic binding energy that we will neglect [460–464]. The scalar's de Broglie wavelength λ sets the characteristic length scale over which this central core varies:

$$\lambda \sim 400 \text{ pc} \left(\frac{\mu}{10^{-22} \text{ eV}}\right)^{-1} \left(\frac{v_{\text{vir}}}{300 \text{ km s}^{-1}}\right)^{-1}, \tag{8.1}$$

with v_{vir} denoting its virial velocity. In this chapter, we will focus on the early inspiral of a binary black hole embedded in the central core of such a halo, and will assume a hierarchy of scales as per (6.2), such that the binary's orbital separation $a \ll \lambda$.

Even a gargantuan $10^{10}\ M_\odot$ black hole has a radius that extends to only a few milliparsecs, thus it is easy to envision comfortably fitting not just one black hole, but a binary of supermassive black holes

inside this central core. Calculations are straightforward in this regime because the constituents of the binary perceive a local environment that is effectively spatially homogeneous. Erecting Fermi normal coordinates $(t, \mathbf{x})$ around the binary, the background scalar field can be taken to be

$$\Phi \approx \Phi_0 \cos(\mu t + \Upsilon) \tag{8.2}$$

in the neighbourhood of its barycentre, which coincides with the origin of our coordinate system. The constant Υ is some arbitrary phase.

Let $\varepsilon = \Phi_0/m_{\rm Pl}$ be a dimensionless parameter that characterises the local dark matter density. Typical FDM halos have $\varepsilon \ll 1$ [see also (8.25) later]; hence, the scalar backreacts onto the spacetime only weakly. We can therefore expand the background metric as $g_{\mu\nu} = \eta_{\mu\nu} + H_{\mu\nu}$ about Minkowski space, where $H_{\mu\nu} \sim \mathcal{O}(\varepsilon^2)$ is the gravitational potential of the halo. Provided that its gradients ∂H are not strong enough to tidally disrupt the binary, the black holes remain in a bound orbit as a result of their mutual gravitational attraction. Their typical size $\sim GM$, their characteristic velocity v, and their orbital separation a are then related by the virial theorem, $v^2 \sim GM/a$. The fact that $v \ll 1$ during the early inspiral facilitates working in the nonrelativistic limit.

These considerations stipulate that an EFT for this system should be organised as an expansion in three small parameters: the orbital velocity v, the "separation of scale" parameter $GM\mu$, and the background weak-field expansion parameter ε. Working perturbatively in powers of v leads to the usual post-Newtonian (PN) expansion [447], which is here augmented by additional expansions in powers of $GM\mu$ and ε, whose job is to characterise the binary's interactions with the ambient scalar. Notice that the two other expansion parameters in (7.3), $GM\omega$ and GM/λ, do not feature here as we are neglecting any spatial variations in Φ and have set $\omega = \mu$.

The remainder of this chapter proceeds as follows: We begin by establishing Feynman rules for this EFT in Section 8.1 so as to facilitate systematic calculations in both the near and far zones of the binary. The capabilities of this approach are then illustrated in Section 8.2, where we identify and compute a number of key physical effects that arise as a result of embedding a binary black hole in an FDM halo. Finally, these calculations are confronted with astronomical data in Section 8.3 to assess if these effects can be meaningfully constrained.

8.1 Feynman rules

The gravitational and scalar perturbations that a black hole sources when it is a part of a binary system exhibit two distinct kinds of behaviour, depending on their wavelengths. First, fluctuations that vary on length scales on the order of the orbital separation a mediate attractive forces between the binary's

constituents. These *potential modes* are always off shell, and so do not appear as propagating degrees of freedom in the EFT. Second, the binary also sources *radiation modes* that vary on length scales $\sim a/v$, which do go on shell and propagate to infinity. When v is small, these scales are widely separated; hence, the different modes can be treated hierarchically by constructing a tower of EFTs, as we discussed in Chapter 6 (see also Refs. [377,497,498]). To that end, we write

$$g_{\mu\nu} = (\underbrace{\eta_{\mu\nu} + H_{\mu\nu}}_{g_{\mu\nu}}) + \frac{\bar{h}_{\mu\nu}}{m_{\rm Pl}} + \frac{h_{\mu\nu}}{m_{\rm Pl}}, \qquad \phi = \Phi + \bar{\varphi} + \varphi, \tag{8.3}$$

where the potential modes are represented by (h, φ) and the radiation modes by $(\bar{h}, \bar{\varphi})$.

To construct a dynamical model for these fields, our starting point is the effective action in Chapter 7. Under the assumption that the orbital separation is much greater than the sizes of the individual black holes, one obtains the effective action $S_{\rm NZ}$ for the binary in its near zone by simply adding a second copy of the point-particle action S_p to the rhs of (7.2). In other words,

$$S_{\rm NZ} = S[g, \phi] + \sum_K S_{p,K}[z_K, g, \phi], \tag{8.4}$$

where the index $K \in \{1, 2\}$ labels the individual members of the binary, and CTP indices have been suppressed for readability. The potential modes are integrated out first to obtain a new effective action for the far zone,

$$S_{\rm FZ} = -i \log \int D[h, \varphi] \, \exp(i S_{\rm NZ}). \tag{8.5}$$

In this regime, the binary as a whole behaves like an effective point particle that is coupled to the remaining radiative degrees of freedom. The flux of radiation off to infinity can be calculated at this stage using $S_{\rm FZ}$. Finally, the effective action for the self-consistent motion of the worldlines is obtained after also integrating out the radiation modes,

$$\Gamma = -i \log \int D[\bar{h}, \bar{\varphi}] \, \exp(i S_{\rm FZ}). \tag{8.6}$$

As with most perturbative EFTs, these path integrals can be performed with the help of Feynman diagrams. The Feynman rules for this theory are presented below.

8.1.1 Feynman rules for potential modes

Because general relativity is invariant under general coordinate transformations, it is necessary to fix a gauge when integrating out the graviton. For its potential modes, we do so by supplementing the

near-zone effective action $S_{\rm NZ}$ with the gauge-fixing term

$$S_{\rm gf}[h;\bar{g}] = -\int \mathrm{d}^4x\,\sqrt{-\bar{g}}\,c_{AB}\bar{g}^{\mu\nu}G^A_\mu[h;\bar{g}]G^B_\nu[h;\bar{g}] \tag{8.7}$$

à la Fadeev and Popov, which enforces the gauge condition $G^A_\mu \approx 0$ for h with respect to the background $\bar{g} = \eta + H + \bar{h}/m_{\rm Pl}$. Note that the inclusion of $\bar{h}$ in (8.7) ensures that $S_{\rm FZ}$ remains gauge invariant. When we subsequently integrate out $\bar{h}$, we will need to gauge-fix again, but more on that later. For our purposes, it is most convenient to work in the Lorenz gauge

$$G^A_\mu[h;\bar{g}] = \bar{\nabla}^\nu\Big(h^A_{\mu\nu} - \frac{1}{2}h^A\bar{g}_{\mu\nu}\Big), \tag{8.8}$$

where $\bar{\nabla}$ is the covariant derivative defined with respect to $\bar{g}$.

Using the field decomposition in (8.3), the terms in the gauge-fixed, near-zone effective action can be grouped as follows:

$$\begin{aligned} S_{\rm NZ} + S_{\rm gf}[h;\bar{g}] = S[\bar{g},\Phi+\bar{\varphi}] + \sum_K S_{p,K}[z_K,\bar{g},\Phi+\bar{\varphi}] \\ + S_{(2)}[h,\varphi] + S^{\rm int}_{\rm NZ}[h,\varphi;z_K,H,\Phi,\bar{h},\bar{\varphi}]. \end{aligned} \tag{8.9}$$

The terms in the first line are independent of the potential modes, and so factor out of the path integral when evaluating (8.5). In the second line, the propagators for the potential modes are specified by

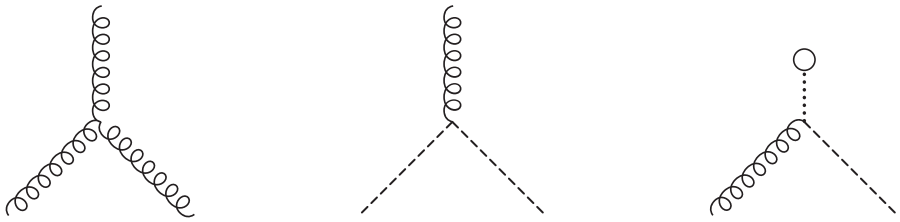
$$S_{(2)}[h,\varphi] = \frac{1}{2}c_{AB}\int \mathrm{d}^4x\,\big[h^A_{\mu\nu}P^{\mu\nu\rho\sigma}\Box h^B_{\rho\sigma} + \varphi^A(\Box-\mu^2)\varphi^B\big], \tag{8.10}$$

where $P^{\mu\nu\rho\sigma} = (\eta^{\mu\rho}\eta^{\sigma\nu} + \eta^{\mu\sigma}\eta^{\rho\nu} - \eta^{\mu\nu}\eta^{\rho\sigma})/2$ is some index structure for the graviton propagator and $\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu$ is the wave operator. Notice that these propagators are defined on flat space; the effects of the background fields (H,Φ) can be treated perturbatively as interaction terms when $\varepsilon \ll 1$. Analogous to (7.18), the action $S_{(2)}$ gives rise to the following propagator matrices:

$$D^{(h)\,\mu\nu\rho\sigma}_{AA'}(x,x') = P^{\mu\nu\rho\sigma}\begin{pmatrix} \frac{1}{2}D^{(h)}_H(x,x') & D^{(h)}_R(x,x') \\ D^{(h)}_A(x,x') & 0 \end{pmatrix}, \tag{8.11a}$$

$$D^{(\varphi)}_{AA'}(x,x') = \begin{pmatrix} \frac{1}{2}D^{(\varphi)}_H(x,x') & D^{(\varphi)}_R(x,x') \\ D^{(\varphi)}_A(x,x') & 0 \end{pmatrix}. \tag{8.11b}$$

The functions D_H, D_R, and D_A are the Hadamard, retarded, and advanced propagators, respectively.



(a) (b) (c)

Figure 8.1 Feynman diagrams for a number of interaction vertices in the bulk of the spacetime. The graviton h, the scalar perturbation φ, and the background scalar Φ are drawn as helical lines, dashed lines, and dotted lines terminating in a circle, respectively.

Returning to (8.9), the last term contains all possible interactions involving the potential modes (h, φ). These include their couplings to the black holes via the worldline vertices discussed in Chapter 7.2, as well as their self-interactions and couplings to the other fields in the bulk; the latter coming from expanding the field action $S[g, \phi]$. A few examples of bulk vertices are shown in Figure 8.1. In later parts of this chapter, we will make explicit use of the vertex in Figure 8.1(c), whose corresponding term in the action is

$$S^{\rm int}_{\rm NZ} \supset \int \mathrm{d}^4x\, h^A_{\mu\nu}(x) V^{\mu\nu}_{h\varphi\Phi}(x) \varphi_A(x), \tag{8.12}$$

$$V^{\mu\nu}_{h\varphi\Phi}(x) = \frac{1}{m_{\rm Pl}} \left(\partial^\mu \Phi \partial^\nu - \frac{1}{2} \eta^{\mu\nu} (\partial^\alpha \Phi \partial_\alpha + \mu^2 \Phi) \right). \tag{8.13}$$

In general, the interaction vertices all have the schematic form $\sim \int \mathrm{d}^4x\, V\, (h^{n_h}/n_h!)\, (\varphi^{n_\varphi}/n_\varphi!)$, where the vertex function V is usually a differential operator, the integers (n_h, n_φ) count the number of potential modes participating in the interaction, and all indices have been suppressed. We may now summarise the position-space Feynman rules for integrating out these potential modes:

(1) For each internal graviton or scalar line, we pick up a factor of the appropriate propagator matrix: either $-iD^{(h)\mu\nu\rho\sigma}_{AA'}$ or $-iD^{(\varphi)}_{AA'}$, respectively.

(2) Each interaction vertex gives a factor of iV.

(3) All CTP and spacetime indices are to be summed over, and all spacetime points are to be integrated over, except those corresponding to external legs.

(4) Divide each diagram by the appropriate symmetry factor.

(5) When a diagram has no external legs, multiply by an additional factor of $-i$, such that it corresponds to a term in the effective action Γ, rather than one in $i\Gamma$.

8.1.2 Feynman rules for radiation modes

To integrate out the radiation modes, we must once again fix a gauge. As before, we enforce the Lorenz gauge condition by supplementing $S_{\rm FZ}$ with the gauge-fixing term $S_{\rm gf}[\bar{h}; g]$, which is now specified with respect to the background metric $g = \eta + H$. The terms in this gauge-fixed effective action fall into one of three categories:

$$S_{\rm FZ} + S_{\rm gf}[\bar{h}; g] = S[g, \Phi] + S_{(2)}[\bar{h}, \bar{\varphi}] + S^{\rm int}_{\rm FZ}[\bar{h}, \bar{\varphi}; z_K, H, \Phi]. \tag{8.14}$$

The first term on the rhs is the action for the background fields (g, Φ), which can be ignored since they are fixed. The second term gives rise to the propagators for the radiation modes, which are identical to those for the potential modes. Finally, the third term represents all possible interactions. The most relevant interaction vertices, which couple the radiation modes to the binary as a whole, are discussed in Section 8.2.5. The same Feynman rules listed earlier also apply to these radiation modes.

8.2 Phenomenology

We shall now discuss five distinct physical effects that arise when a binary black hole is embedded in an FDM halo. Due to the rapid increase in the number of Feynman diagrams at each successive PN order, a comprehensive and systematic expansion of Γ is far beyond the scope of this thesis. Instead, we will limit ourselves to deriving only the leading-order expression for each effect.

8.2.1 Inverse-square law forces

The dominant interaction holding the binary together is the exchange of a graviton between the black holes, shown in Figure 8.2(a), which gives rise to Newton's inverse-square law force. To see this, we make use of the worldline vertex in Figure 7.1(c), which reduces to

$$S_{p,K} \supset \frac{M_K}{2m_{\rm Pl}} \int {\rm d}^4x \, \delta^A_K(x) h_{A,00}(x) \tag{8.15}$$

in the nonrelativistic limit. The delta function $\delta_{a,K}(x) := \delta^{(3)}(\mathbf{x} - \mathbf{z}_{a,K}(t))$ localises the integral to be along the ath copy of the Kth worldline. In the Keldysh representation,

$$\delta^+_K(x) = -\mathbf{z}^i_{K,-} \frac{\partial}{\partial \mathbf{x}^i} \delta^{(3)}(\mathbf{x} - \mathbf{z}_K(t)) + \mathcal{O}(\mathbf{z}^2_-), \tag{8.16a}$$

$$\delta^-_K(x) = \delta^{(3)}(\mathbf{x} - \mathbf{z}_K(t)) + \mathcal{O}(\mathbf{z}_-), \tag{8.16b}$$

having expanded about the physical limit. These expressions will be essential for calculating equations

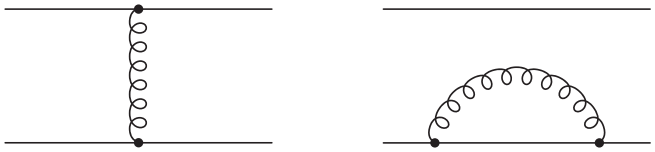


Figure 8.2 Feynman diagrams depicting (a) the exchange of a graviton between the two black holes and (b) a correction to the black hole's self-energy. The first leads to Newton's gravitational inverse-square law, whereas the second is pure counterterm. Its mirror inverse, in which the graviton propagates to and from the top worldline, is included implicitly as the diagram does not differentiate between the two solid lines.

of motion, which (as we discussed in Chapter 7) follow from simply expanding Γ in powers of $\mathbf{z}_-$ and reading off the linear coefficient. Now using the Feynman rules, we obtain

$$\begin{aligned}\text{Fig. 8.2} &= \frac{1}{2}\sum_{K,K'}\frac{M_K M_{K'}}{4m_{\rm Pl}^2}\int \mathrm{d}^4x\,\mathrm{d}^4x'\,\delta_K^A(x)\delta_{K'}^{A'}(x')P^{0000}D_{AA'}^{(h)}(x,x')\\ &= \sum_{K,K'}\frac{M_K M_{K'}}{8m_{\rm Pl}^2}\int \mathrm{d}^4x\,\mathrm{d}^4x'\,\delta_K^+(x)\delta_{K'}^-(x')D_R^{(h)}(x,x'),\end{aligned} \tag{8.17}$$

where the factor of $1/2$ in the first line is a symmetry factor. To obtain the second line, first note that the Hadamard propagator $D_H^{(h)} \equiv D_{++}^{(h)}$ is flanked by two powers of δ^+ ($\equiv\delta_-$), which vanishes in the physical limit. Thus, much like with G_H in Chapter 7, we learn that $D_H^{(h)}$, and similarly $D_H^{(\varphi)}$, never contribute to the classical equations of motion. For convenience, we shall simply set these to zero hereafter. Second, note that the propagator matrices in (8.11) are symmetric under the simultaneous interchange of the arguments $x \leftrightarrow x'$ and the CTP indices $A \leftrightarrow A'$; hence, an appropriate relabelling of dummy indices and integration variables can always be used to replace any advanced propagator appearing in a Feynman diagram with the retarded propagator.

To proceed, it will be instructive to temporarily hold the black holes at rest at their respective positions, such that the delta functions $\delta_K^\pm(x)$ are independent of time. In this case, the integral over x' can be evaluated easily by exploiting the Lorentz invariance of the propagator and the master integral in Appendix 7.A. After also using (8.16), the result is

$$\begin{aligned}\text{Fig. 8.2} &= \sum_{K,K'}\int \mathrm{d}t\int \mathrm{d}^3\mathbf{x}\,\mathrm{d}^3\mathbf{x}'\,\delta_K^+(\mathbf{x})\delta_{K'}^-(\mathbf{x}')\frac{GM_K M_{K'}}{|\mathbf{x}-\mathbf{x}'|}\\ &= \sum_{K,K'}\int \mathrm{d}t\,\frac{\partial}{\partial\mathbf{z}_K^i}\left(\frac{GM_K M_{K'}}{|\mathbf{z}_K-\mathbf{z}_{K'}|}\right)\mathbf{z}_{K,-}^i + \mathcal{O}(\mathbf{z}_{K,-}^2).\end{aligned} \tag{8.18}$$

There are two types of terms in this sum. The cross terms with $K \neq K'$ are depicted in Figure 8.2(a), and lead to the mediation of a gravitational force between the black holes. Additionally, the Feynman

rules tell us that we should also get terms with $K = K'$ when integrating out the graviton. These UV-divergent contributions, drawn in Figure 8.2(b), represent a correction to the self-energy of the point mass and are pure counterterm. Moreover, they can be seen to vanish identically in dimensional regularisation by Veltman's formula [377].

The equation of motion for the Kth black hole has the general form $M_K \mathbf{a}_K = \mathbf{F}_K$. As Figure 8.2(a) constitutes a term in the effective action Γ, its contribution to $\mathbf{F}_K$ is obtained by differentiating it with respect to $\mathbf{z}_{-,K}$ and taking the physical limit; cf. (7.55). As expected, the force on the first black hole is found to be

$$\mathbf{F}_1 \supset -\frac{GM_1M_2}{r^2}\hat{\mathbf{r}}, \tag{8.19}$$

where $\mathbf{r} = \mathbf{z}_1 - \mathbf{z}_2$, $r = |\mathbf{r}|$, and $\hat{\mathbf{r}} = \mathbf{r}/r$. Naturally, interchanging the labels $1 \leftrightarrow 2$ gives us the force on the second black hole.[1]

This gravitational force is always there, of course. What is more interesting in the present context is the fifth force that arises due to the FDM halo inducing scalar charges onto the black holes. This effect is shown in Figure 8.3(a). The self-energy diagram in Figure 8.3(b), meanwhile, is unphysical and vanishes for the same reason as before. Evaluating the former diagram requires the worldline vertex in Figure 7.1(e), which is given by

$$S_{p,K} \supset \int \mathrm{d}^4x\, Q_K(t)\delta^A_K(x)\varphi_A(x) \tag{8.20}$$

at leading order in v. For the background in (8.2), the scalar charge reads $Q_K(t) = A_K\mu\Phi_0\sin(\mu t + \Upsilon)$. Having shown explicitly how this kind of diagrammatic calculation is performed in the case of the graviton, it is now straightforward to repeat the steps for the scalar field. One finds that

$$\begin{aligned}\text{Fig. 8.3(a)} &= \sum_{K\neq K'}\int \mathrm{d}^4x\mathrm{d}^4x'\, Q_K(t)Q_{K'}(t')\delta^+_K(x)\delta^-_{K'}(x')D^{(\varphi)}_R(x,x') \\ &= \sum_{K\neq K'}\int \mathrm{d}t\,\frac{\partial}{\partial\mathbf{z}^i_{K,-}}\left(\frac{Q_K(t)Q_{K'}(t)}{4\pi|\mathbf{z}_K - \mathbf{z}_{K'}|}\right)\mathbf{z}^i_{K,-} + \mathcal{O}(\mathbf{z}^2_{K,-}),\end{aligned} \tag{8.21}$$

and thus

$$\mathbf{F}_1 \supset -\frac{Q_1(t)Q_2(t)}{4\pi r^2}\hat{\mathbf{r}}. \tag{8.22}$$

Notice that this fifth force is not Yukawa suppressed, despite the scalar being massive. The reason is

[1] Note that these labels now distinguish between the members of the binary. Equations of motion are always given in the physical limit, so there are no longer any CTP indices floating around.

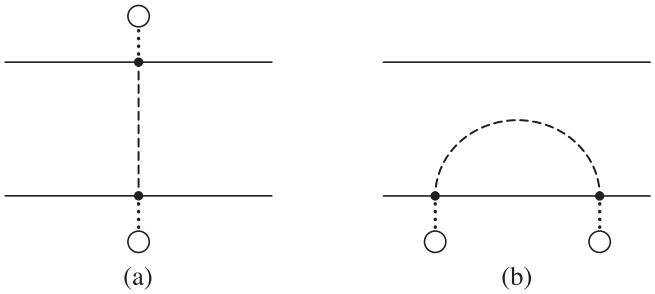


Figure 8.3 Analogous Feynman diagrams depicting the (a) exchange of a scalar between two black holes with induced scalar charges and (b) a further correction to the black hole's self-energy. The first results in the mediation of a fifth force, whereas the second is pure counterterm. As in Figure 8.2(b), its mirror inverse is included implicitly.

that its mass has been "eaten up" by the time dependence of the background $\Phi(t)$, which oscillates at the same frequency $\omega_0 = \mu$.[2]

Recall that our job is not yet complete, however, as these results were obtained by artificially holding the black holes at rest. What happens should we relax this assumption and allow them to move freely? Since any departure from the static case must depend on the velocity v, nothing changes at this order in the PN expansion, but when going to higher orders, a more systematic approach that makes power counting in v manifest would be desirable. This can be accomplished by expanding the momentum-space propagators for the potential modes as

$$\widetilde{D}_R^{(h)}(p) = \frac{1}{p^2} = \frac{1}{\mathbf{p}^2}\left(1 + \frac{(p^0)^2}{\mathbf{p}^2} + \cdots\right), \tag{8.23a}$$

$$\widetilde{D}_R^{(\varphi)}(p) = \frac{1}{p^2 + \mu^2} = \frac{1}{\mathbf{p}^2}\left(1 + \frac{(p^0)^2 - \mu^2}{\mathbf{p}^2} + \cdots\right). \tag{8.23b}$$

In both cases, the leading $1/\mathbf{p}^2$ term corresponds to an instantaneously propagating field, which is responsible for mediating the inverse-square law forces in (8.19) and (8.22). The typical 3-momentum for these potential modes is therefore $|\mathbf{p}| \sim 1/a$. Of course, neither of them actually propagate instantaneously, but since this is apparent only when the black holes are moving, it must be that the remaining terms in the expansion are subleading. For the graviton h, it follows that $p^0 \sim v/a$ [377].

The energy for the scalar φ is slightly more complicated, as it has two pieces that scale differently: $p^0 \sim v/a + \mu$, or equivalently, $[(p^0)^2 - \mu^2]/\mathbf{p}^2 \sim v^2 + \mu a v$. The first piece is due to the time-dependent motion of the black holes as before, while the second is introduced by the oscillating

[2]To be more precise, the background oscillates at a frequency $\omega_0 = \mu - |\Delta E|$, where $|\Delta E|$ ($\ll \mu$) is the binding energy of the halo's central core. Because this frequency is not exactly equal to the mass of the scalar, the fifth force in (8.22) does actually exhibit a very slight Yukawa suppression $\sim e^{-|\Delta E| r}$. However, as the associated length scale $|\Delta E|^{-1} \gg \lambda$, this exponential factor can be neglected for orbital separations $a \ll \lambda$.

background Φ, which enters into the worldline vertices via the charges $Q_K(t)$. The expansion in (8.23b) is therefore valid only when $v^2 \ll 1$ and $\mu a v \ll 1$. The first condition is always satisfied during the early inspiral phase, but the second stipulates that the scalar field can be approximated as propagating instantaneously only when the binary is not too widely separated. Using the virial theorem to rewrite this as $a \ll 1/(GM\mu^2)$, we require that

$$a \ll 10\ \text{pc} \left(\frac{M}{10^{10}\ M_\odot}\right)^{-1} \left(\frac{\mu}{10^{-22}\ \text{eV}}\right)^{-2}. \tag{8.24}$$

This condition can be regarded as an IR cutoff for the validity of this near-zone EFT.

Equations (8.23a) and (8.23b) make it possible to organise our perturbative series in such a way that, at least at low orders, each Feynman diagram scales homogeneously with the EFT's expansion parameters. Specifically, each term in the effective action Γ scales as $L^{1-\ell}\, v^{2n} \varepsilon^{p_1} (GM\mu)^{p_2}$, where $L \sim Mav$ is the characteristic angular momentum of the binary. (As in Chapter 6, we will assume that the black holes have comparable masses for the purposes of power counting, but the results in this chapter are valid for arbitrary mass ratios.) The integer ℓ counts the number of loops in a given Feynman diagram, and since $L \gg 1$ for astrophysical black holes, only tree-level diagrams are needed [377]. The integer or half-integer n, meanwhile, counts the order in the PN expansion, and since the terms with $p_1 = p_2 = 0$ give rise to the standard PN equations for a binary in vacuum [447], they need not be revisited here. Due to its $\phi \to -\phi$ symmetry, effects involving the scalar field first appear when $p_1 = p_2 = 2$. In this chapter, we work to leading order in ε and $GM\mu$.

Let us now establish power counting rules for the potential modes. (The rules developed herein and later in this chapter are summarised in Table 8.1.) As the potential-mode graviton has typical momentum scaling as $(p^0, |\mathbf{p}|) \sim (v/a, 1/a)$, time derivatives acting on h scale as v/a while spatial derivatives scale as $1/a$. It then follows that the instantaneous propagator $\langle hh \rangle \sim \int \mathrm{d}^4 p / \mathbf{p}^2 \sim p^0 |\mathbf{p}| \sim v/a^2$; hence, $h \sim \sqrt{v}/a$ when appearing as an internal line in a Feynman diagram. After also using the virial theorem to show that $M/m_{\text{Pl}} \sim \sqrt{Lv}$, one finds that Fig. 8.2 $\sim [(M/m_{\text{Pl}}) \int \mathrm{d}t]^2 \langle hh \rangle \sim Lv^0$, having

Table 8.1 Power counting rules for a binary black hole in a fuzzy dark matter halo. All derivatives ∂_μ scale in the same way, except spatial derivatives acting on the potential modes, which are denoted by the 3-momentum $\mathbf{p}$, and spatial derivatives on the background scalar Φ, which vanish. Note that these rules are valid only to leading order in ε and $GM\mu$. Note also that the rules for the radiation modes further assume $\Omega \gg \mu$ for simplicity.

h, φ	$\bar{h}, \bar{\varphi}$	Φ/m_{Pl}	H	∂_μ	$\mathbf{p}$	M/m_{Pl}	$\delta M/m_{\text{Pl}}$	Q
$\frac{\sqrt{v}}{a}$	$\frac{v}{a}$	$\frac{\varepsilon\mu a}{v}$	$\frac{(\varepsilon\mu a)^2}{v^2}$	$\frac{v}{a}$	$\frac{1}{a}$	$\sqrt{Lv}$	$\frac{\sqrt{Lv}(\varepsilon GM\mu)^2}{v^3}$	$\sqrt{Lv}(\varepsilon GM\mu)$

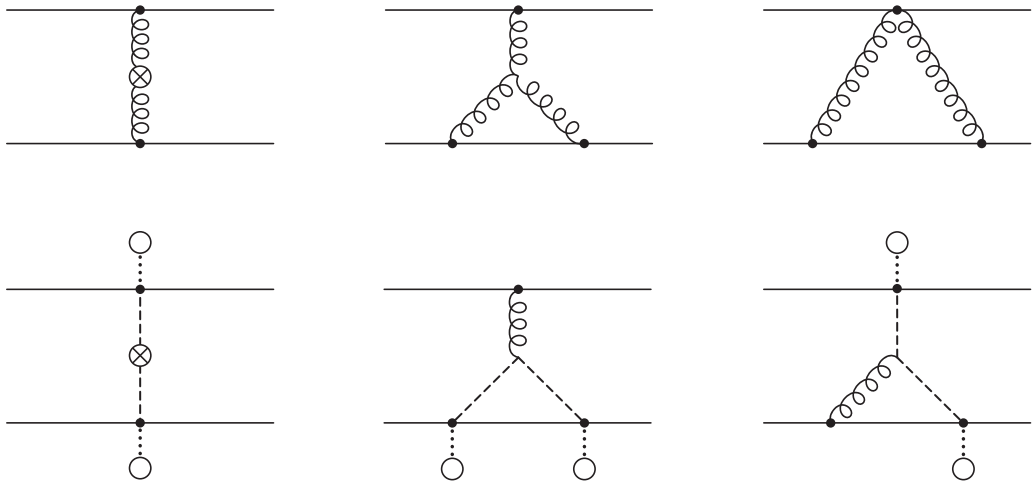


Figure 8.4 Examples of Feynman diagrams contributing to the effective action Γ at 1PN order. The diagrams in the second row are suppressed relative to those in the first by $(\varepsilon GM\mu)^2$. All of the propagators are evaluated using their instantaneous limits, except those in the diagrams of the leftmost column. The "$\otimes$" symbol instructs us to use the $\mathcal{O}(v^2)$ term in the propagator instead.

set $\int \mathrm{d}t \sim a/v$, since the orbital period of the binary is the most relevant timescale in this problem. Given that a term in Γ is said to be of nPN order if it scales as v^{2n}, the Newtonian gravitational force in Figure 8.2 is, unsurprisingly, a Newtonian-order effect (0PN).

Now consider the scalar fifth force between the black holes. Time derivatives can be arranged to scale in the same way when acting on all fields by choosing $\Phi/m_{\mathrm{Pl}} \sim \varepsilon\mu a/v$, such that $\partial_t \Phi \sim \mu\varepsilon m_{\mathrm{Pl}}$, and consequently, $Q \sim \sqrt{Lv}(\varepsilon GM\mu)$. Since this charge is already linear in $GM\mu$, and it is not possible to draw a tree-level diagram involving the scalar with fewer than two powers of Q, we can neglect the μ dependence in $D_R^{(\varphi)}$ entirely at leading order in $GM\mu$.[3] As a result, the potential-mode scalar also has typical momentum scaling as $(p^0, |\mathbf{p}|) \sim (v/a, 1/a)$, and thus $\varphi \sim \sqrt{v}/a$ when appearing as an internal line. It now follows that Fig. 8.3 $\sim (Q\int \mathrm{d}t)^2 \langle\varphi\varphi\rangle \sim Lv^0(\varepsilon GM\mu)^2$. The scalar fifth force is also a Newtonian-order effect, albeit suppressed by two powers of $\varepsilon GM\mu$.

These power counting rules make it easy to determine ahead of any detailed calculation the order at which a given diagram appears in perturbation theory. For instance, the diagrams in Figure 8.4 can all be shown to contribute at 1PN order. However, as the diagrams involving the scalar are all suppressed relative to their purely-gravitational counterparts by at least

$$(\varepsilon GM\mu)^2 \sim 2 \times 10^{-16} \left(\frac{\rho}{100\ M_\odot\,\mathrm{pc}^{-3}}\right)\left(\frac{M}{10^{10}\ M_\odot}\right)^2, \tag{8.25}$$

[3] At higher orders, it becomes necessary to factor out the μ dependence explicitly; for instance, by working with the complex field ψ defined from $\varphi(x) \propto [\psi(x)e^{-i\mu t} + \psi^*(x)e^{i\mu t}]$ instead.

which is a very small number for typical values of the local FDM density $\rho \equiv \mu^2\Phi_0^2/2$ [307], going to higher PN orders is unproductive.

What is interesting about this system, however, is the emergence of Feynman diagrams that are formally of *pre*-Newtonian order; i.e., diagrams that scale with negative powers of v. These can partially counteract the smallness of $\varepsilon GM\mu$ and lead to effects that are enhanced (but still subleading to the dominant Newtonian interaction in Figure 8.2) when the binary's orbital separation is large. Let us now turn our attention to some of these pre-Newtonian effects.

8.2.2 Accretion

We already encountered the drag force from accretion in Chapter 7 in a fully relativistic setting. When expanded in powers of v, the leading term in (7.59) is proportional to v^2 and is depicted in Figure 8.5(a). Schematically, Fig. 8.5(a) $\sim \int \mathrm{d}t\, \delta M\, \mathbf{v}^2 \sim Lv^{-3}(\varepsilon GM\mu)^2$. This is formally a -1.5PN effect, but it is still subleading to the Newtonian-order interaction in Figure 8.2 ($\sim Lv^0$) due to suppression by two powers of $\varepsilon GM\mu$.

In the presence of a second black hole, an additional diagram contributes at this order: Figure 8.5(b) accounts for the change in the gravitational force between the black holes due to their gradually increasing masses. Notice that only one of the black holes is accreting in this diagram. The diagram in which both are accreting first appears at $\mathcal{O}(\varepsilon^4)$ and should therefore be neglected when working to leading order. Even without detailed calculation, it is easy to correctly intuit that the combined effect of the diagrams in Figure 8.5 is the force

$$\mathbf{F}_1 \supset -\delta\dot{M}_1\mathbf{v}_1 - \delta M_1\mathbf{a}_1 - \frac{G(M_1\delta M_2 + M_2\delta M_1)}{r^2}\hat{\mathbf{r}}. \tag{8.26}$$

As always, the force on the second black hole can be obtained by interchanging the labels $1 \leftrightarrow 2$, and note that $\hat{\mathbf{r}}$ changes sign under this interchange.

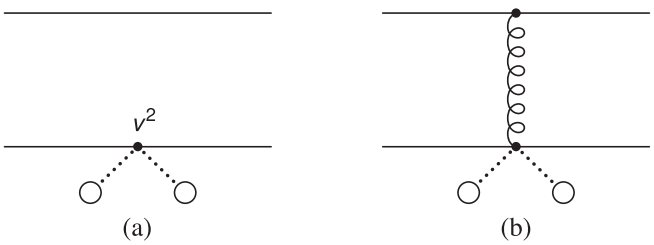


Figure 8.5 Leading-order Feynman diagrams accounting for (a) the drag force and (b) the change in the strength of the gravitational interaction due to accretion of the background scalar. As with earlier diagrams, their mirror inverses, in which the background scalar interacts with the top worldline, are included implicitly.

8.2.3 Dynamical friction

Dynamical friction is a universal effect that arises when any massive body is immersed in a background medium. The gravitational field of this massive body — in our case, a black hole — perturbs the medium through which it moves, forming a wake in the latter that then exerts a gravitational pull back on the object. Classic studies on dynamical friction have focused on collisionless or gaseous media [499–502], although recent work has begun to explore what modifications are needed to account for the wave-like nature of fuzzy dark matter [259, 273, 503].

The EFT formalism employed herein provides a natural language for calculating the force that dynamical friction exerts on a massive body. Figure 8.6 depicts a black hole interacting with its gravitationally induced wake, and evaluates to

$$\begin{aligned}\text{Fig. 8.6} = \left(\frac{M}{2m_{\rm Pl}}\right)^2 \int \mathrm{d}^4x\, \mathrm{d}^4x'\, \mathrm{d}^4y\, \mathrm{d}^4y'\, \delta^+(x)\delta^-(x') \\ \times \int \frac{\mathrm{d}^4p}{(2\pi)^4}\frac{\mathrm{d}^4p'}{(2\pi)^4}\frac{\mathrm{d}^4q}{(2\pi)^4}\,\frac{e^{ip\cdot(x-y)}}{\mathbf{p}^2}\frac{e^{iq\cdot(y-y')}}{\mathbf{q}^2}\frac{e^{ip'\cdot(y'-x')}}{\mathbf{p}'^2} \\ \times P_{00\mu\nu}V^{\mu\nu}_{h\varphi\Phi}(y^0,q^0)\,P_{00\rho\sigma}V^{\rho\sigma}_{h\varphi\Phi}(y'^0,-q^0)\end{aligned} \tag{8.27}$$

at leading order in v. As this force acts independently on each member of the binary, the index K has been suppressed for the sake of readability. Each of the vertex functions in the third line read

$$P_{00\mu\nu}V^{\mu\nu}_{h\varphi\Phi}(t,\omega) = -\frac{2i\omega\,\dot{\Phi}(t)+\mu^2\Phi(t)}{2m_{\rm Pl}}. \tag{8.28}$$

Although the two terms in the numerator scale with different powers of $GM\mu$ and v, it will be instructive to keep both of them around in this derivation. It turns out that the second term, $\mu^2\Phi$, provides no contribution whatsoever to the force.

We proceed by first performing a number of trivial integrations to simplify (8.27). Integrating over p^0 and p'^0 produces delta functions that enforce the conditions $y^0 = x^0$ ($\equiv t$) and $y'^0 = x'^0$ ($\equiv t'$),

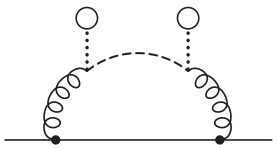


Figure 8.6 Feynman diagram depicting the drag force due to dynamical friction.

respectively. Moreover, integrating over $\mathbf{y}$ and $\mathbf{y}'$ enforces the conservation of 3-momentum along the entire diagram; i.e., $\mathbf{p} = \mathbf{p}' = \mathbf{q}$. The result is

$$\text{Fig. 8.6} = \frac{M^2}{16 m_{\rm Pl}^4} \int \mathrm{d}^4x\, \mathrm{d}^4x'\, \delta^+(x)\delta^-(x') \int \frac{\mathrm{d}^4q}{(2\pi)^4} \frac{e^{iq\cdot(x-x')}}{\mathbf{q}^6}\, W(q^0;t,t'), \tag{8.29}$$

where $$W(q^0;t,t') = \left[-2iq^0\dot{\Phi}(t) - \mu^2\Phi(t)\right]\left[2iq^0\dot{\Phi}(t') - \mu^2\Phi(t')\right]. \tag{8.30}$$

Next, we perform the integral over $\mathbf{q}$ by utilising the standard identity [504]

$$\int \frac{\mathrm{d}^d\mathbf{q}}{(2\pi)^d} \frac{e^{i\mathbf{q}\cdot\mathbf{r}}}{(\mathbf{q}^2)^\alpha} = \frac{1}{(4\pi)^{d/2}} \frac{\Gamma(d/2-\alpha)}{\Gamma(\alpha)} \left(\frac{\mathbf{r}^2}{4}\right)^{\alpha-d/2}, \tag{8.31}$$

while the integral over q^0 is performed by replacing each factor of iq^0 in $W(q^0;t,t')$ with a derivative $\mathrm{d}/\mathrm{d}t'$ acting on $e^{-iq^0(t-t')}$ and then integrating by parts. These steps lead to the expression

$$\begin{aligned}\text{Fig. 8.6} = \frac{M^2}{16 m_{\rm Pl}^4} \int \mathrm{d}t\, \mathrm{d}t' \delta(t-t') \Bigg(& W_0(t,t') + W_1(t,t') \frac{\mathrm{d}}{\mathrm{d}t'} \\ & + W_2(t,t') \frac{\mathrm{d}^2}{\mathrm{d}t'^2} \Bigg) \int \mathrm{d}^3\mathbf{x}\, \mathrm{d}^3\mathbf{x}'\, \delta^+(x)\delta^-(x') |\mathbf{x}-\mathbf{x}'|^3,\end{aligned} \tag{8.32}$$

with $W_2(t,t') = -4\dot{\Phi}(t)\dot{\Phi}(t')$. Determining expressions for W_0 and W_1 will not be necessary. We now expand $\delta^\pm(x)$ in powers of $\mathbf{z}_-$ according to (8.16). To linear order in $\mathbf{z}_-$, the term involving W_2 yields

$$\text{Fig. 8.6} \supset -\frac{8\pi}{3}(GM)^2 \int \mathrm{d}t\, \mathrm{d}t' \delta(t-t') \dot{\Phi}(t)\dot{\Phi}(t')\, \mathbf{z}_-^i(t) \frac{\mathrm{d}^2}{\mathrm{d}t'^2} \left(\frac{\partial |\mathbf{z}_+(t) - \mathbf{z}(t')|^3}{\partial \mathbf{z}_+^i(t)} \right)_{\rm PL}. \tag{8.33}$$

At this stage, observe that the integrand is proportional to the delta function $\delta(t-t')$; hence, if we substitute $t' = t - s$ and expand in powers of s, we see that only the $\sim \mathcal{O}(s^0)\delta(s)$ term contributes. This allows us to simplify the derivatives in the integrand, such that

$$\frac{\mathrm{d}^2}{\mathrm{d}t'^2} \left(\frac{\partial |\mathbf{z}_+(t) - \mathbf{z}(t')|^3}{\partial \mathbf{z}_+^i(t)} \right)_{\rm PL} = \frac{\mathrm{d}^2}{\mathrm{d}s^2} \left(3s^2 |\mathbf{v} + \mathcal{O}(s)|^2 \frac{[\mathbf{v} + \mathcal{O}(s)]_i}{|\mathbf{v} + \mathcal{O}(s)|} \right) = 6|\mathbf{v}|^2 \hat{\mathbf{v}} + \mathcal{O}(s), \tag{8.34}$$

where $\mathbf{v} \equiv \dot{\mathbf{z}}(t)$. Putting things together, the term in Figure 8.6 involving W_2 simplifies to

$$\text{Fig. 8.6} \supset -16\pi (GM)^2 \int \mathrm{d}t\, \dot{\Phi}^2 \mathbf{v}^2\, \hat{\mathbf{v}} \cdot \mathbf{z}_-. \tag{8.35}$$

As a final step, note that the terms involving W_0 and W_1 do not contribute. It is easy to convince

ourselves that this must the case, since

$$W_0(t,t')\left(\frac{\partial|\mathbf{z}_+(t)-\mathbf{z}(t')|^3}{\partial\mathbf{z}^i_+(t)}\right)_{\rm PL}\sim\mathcal{O}(s^2), \tag{8.36}$$

$$W_1(t,t')\frac{\mathrm{d}}{\mathrm{d}t'}\left(\frac{\partial|\mathbf{z}_+(t)-\mathbf{z}(t')|^3}{\partial\mathbf{z}^i_+(t)}\right)_{\rm PL}\sim\mathcal{O}(s). \tag{8.37}$$

To obtain the force on the Kth black hole due to dynamical friction, we now differentiate (8.35) with respect to $\mathbf{z}_-$ and then reinstate the label K to find

$$\mathbf{F}_K\supset -16\pi(GM_K)^2\dot{\Phi}^2\mathbf{v}_K^2\,\hat{\mathbf{v}}_K. \tag{8.38}$$

Power counting reveals that this force is a -1PN order effect; i.e., Fig. 8.6 $\sim Lv^{-2}(\varepsilon GM\mu)^2$.

As this result relies on the assumption that both the graviton and the scalar propagate instantaneously at leading PN order, it is valid only if the binary is tight enough to satisfy the condition $\mu av\ll 1$; cf. (8.24). In Refs. [259, 273, 503], dynamical friction within an FDM halo is studied in the opposite regime, $\mu av\gtrsim 1$. (Stars orbiting the centre of the galaxy, for instance, satisfy this condition.) Consequently, a direct comparison between the result in (8.38) and that of Refs. [259, 273, 503] is not possible and, in fact, they need not agree. Nonetheless, one can verify that this EFT approach correctly reproduces the results in Ref. [273] when working under similar assumptions.

8.2.4 The halo's gravitational potential

The FDM halo in which the binary resides weakly curves the spacetime geometry and so exerts an external gravitational force on the black holes. As the halo's gravitational potential varies on a length scale $\sim\lambda$ that is much greater than the orbital separation a, we may expand

$$H_{\mu\nu}(t,\mathbf{x})=H_{\mu\nu}(t,\mathbf{0})+\mathbf{x}^i\partial_i H_{\mu\nu}(t,\mathbf{0})+\frac{1}{2}\mathbf{x}^i\mathbf{x}^j\partial_i\partial_j H_{\mu\nu}(t,\mathbf{0})+\cdots \tag{8.39}$$

about the binary's barycentre, which (recall) coincides with the origin of our coordinate system.

Note that $H_{\mu\nu}$ must depend on the spatial coordinates, despite us having assumed that Φ is spatially homogeneous, if it is to be a consistent solution at $\mathcal{O}(\varepsilon^2)$ to the background field equation

$$R_{\mu\nu}=8\pi G\left(\partial_\mu\Phi\,\partial_\nu\Phi+\frac{1}{2}\eta_{\mu\nu}\mu^2\Phi^2\right). \tag{8.40}$$

For the power counting rules, this equation enforces the relation $R\sim\partial\partial H\sim\varepsilon^2\mu^2$, which is satisfied provided all derivatives acting on $H\equiv H(t,\mathbf{0})$ scale as $\partial_\mu\sim v/a$, while taking $H\sim(\varepsilon\mu a/v)^2$.

While it is possible to stick with the general multipole expansion in (8.39), it is far more convenient if we pick a gauge. In Fermi normal coordinates, we have that $H_{\mu\nu}(t,\mathbf{0})$ and $\partial_i H_{\mu\nu}(t,\mathbf{0})$ are both equal to zero (the linear gradients $\partial_i H$ would be nontrivial if $\partial_i \Phi \neq 0$, cf. Chapter 3.2), whereas [403]

$$\frac{1}{2}\partial_i\partial_j H_{00}(t,\mathbf{0}) = -R_{0i0j}(t,\mathbf{0}), \tag{8.41a}$$

$$\frac{1}{2}\partial_i\partial_j H_{0k}(t,\mathbf{0}) = -\frac{2}{3}R_{0ikj}(t,\mathbf{0}), \tag{8.41b}$$

$$\frac{1}{2}\partial_i\partial_j H_{k\ell}(t,\mathbf{0}) = -\frac{1}{3}R_{ki\ell j}(t,\mathbf{0}). \tag{8.41c}$$

At leading order in the PN expansion, the only contribution involving $H_{\mu\nu}$ comes from expanding the point-mass term in the action:

$$\Gamma \supset c^a \sum_K \int \mathrm{d}t\, \frac{1}{2} M_K H_{00}(t, \mathbf{z}_{a,K}(t)) = -\sum_K \int \mathrm{d}t\, M_K R_{0i0j}(t,\mathbf{0}) \mathbf{z}^i_{-,K} \mathbf{z}^j_{+,K}, \tag{8.42}$$

which gives rise to the force

$$\mathbf{F}^i_K \supset -R^i{}_{0j0}(t,\mathbf{0}) M_K \mathbf{z}^j_K(t). \tag{8.43}$$

For the background in (8.2), the Riemann tensor evaluates to $R^i{}_{0j0}(t,\mathbf{0}) = -4\pi G\rho \cos(2\mu t + 2\Upsilon)\delta^i_j$.

The effect of this kind of external, periodic force has previously been studied by Blas et al. [296] in the context of pulsar binaries, although its impact on binary black holes is completely analogous. In particular, the binary experiences a secular variation in its orbital period as a result of (8.43) that is most pronounced for highly eccentric orbits. (The effect vanishes for circular orbits.) Furthermore, the transfer of energy between the binary and the halo is resonantly amplified when the background spacetime's oscillation frequency 2μ is close to an integer multiple of the binary's orbital frequency Ω. We will have more to say about this force in Section 8.3.

For now, power counting tells us that this external force first appears at -3PN order; i.e., the term in (8.42) scales as $Lv^{-6}(\varepsilon GM\mu)^2$. At leading order in ε and $GM\mu$, this term scales with the most negative power of v, and so establishes a second, independent IR cutoff for this system [the first being (8.24)]. To see this, recall that our perturbative expansion is predicated on the virial relation $v^2 \sim GM/a$, which holds only if the Newtonian-order interactions ($\sim Lv^0$) are the dominant terms in the action. This demands that the binary satisfy the condition $Lv^{-6}(\varepsilon GM\mu)^2 \ll Lv^0$, which can be written equivalently as $a^3 \ll GM/(\varepsilon\mu)^2$, or most transparently as $(\varepsilon\mu a/v)^2 \sim H \ll 1$. For small enough velocities and large enough orbital separations, this scaling analysis naively suggests that H can attain values of order one, at which point it stops being a weak perturbation to the Minkowski metric.

Before this can happen, spatial variations of Φ become relevant and must be taken into account. Thus, a power series expansion of the background fields is valid only if

$$a \ll 80 \text{ pc} \left(\frac{\rho}{100\ M_\odot \text{ pc}^{-3}} \right)^{-1/3} \left(\frac{M}{10^{10}\ M_\odot} \right)^{1/3}. \tag{8.44}$$

Which of the two cutoffs in (8.24) and (8.44) is more relevant will depend on the specifics of the particular system, but for supermassive binary black holes in typical FDM halos, it turns out that the former establishes the more stringent condition.

8.2.5 Scalar radiation

All of the effects discussed so far emerge from either integrating out potential modes or from direct couplings of the background fields to the point-particle action. To complete this discussion on the phenomenology of binary black holes in FDM halos, we now turn our attention to the radiative sector of the theory. The radiation-mode scalar $\bar{\varphi}(x)$ couples to the black holes via the term

$$S_{\text{FZ}}^{\text{int}} \supset \sum_K \int \mathrm{d}^4x\ Q_K(t) \delta_K^A(x) \bar{\varphi}_A(x). \tag{8.45}$$

As the radiation modes vary on length scales much greater than the size of the binary, we may expand

$$\bar{\varphi}_A(x) = \bar{\varphi}_A(t, \mathbf{0}) + \mathbf{x}^i \partial_i \bar{\varphi}_A(t, \mathbf{0}) + \frac{1}{2} \mathbf{x}^i \mathbf{x}^j \partial_i \partial_j \bar{\varphi}_A(t, \mathbf{0}) + \cdots \tag{8.46}$$

about the binary's barycentre, just as we did in (8.39). Substituting this back into (8.45), integrating over $\mathbf{x}$, and focusing on the terms that are independent of $\mathbf{z}_-$, we get

$$S_{\text{FZ}}^{\text{int}} \supset \int \mathrm{d}t \left[\sum_K Q_K(t) \right] \bar{\varphi}_-(t, \mathbf{0}) + \int \mathrm{d}t \left[\sum_K Q_K(t) \mathbf{z}_K^i(t) \right] \partial_i \bar{\varphi}_-(t, \mathbf{0}) + \cdots . \tag{8.47}$$

In effect, this expansion in powers of $\mathbf{x}$ amounts to a coarse-graining procedure that zooms out from the binary and replaces it by a new effective point particle localised at the origin. Each square bracket in (8.47) corresponds to a multipole moment of the binary as a whole. (The expressions are accurate to leading PN order). The first term is the global monopole moment, which simply describes the total scalar charge of the binary. This mode sources a nonpropagating field, and so does not lead to any transport of energy. The dominant channel for scalar radiation is therefore the global dipole moment, which is given by the second term in (8.47).

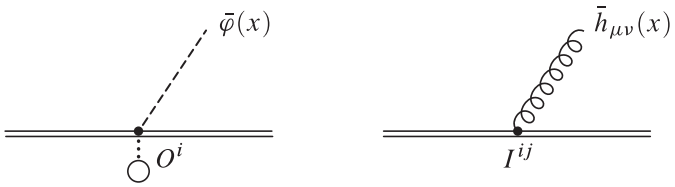


Figure 8.7 Feynman diagrams representing the emission of (a) scalar waves sourced by the dipole moment $O^i = \sum_K Q_K \mathbf{z}_K^i$ and (b) gravitational waves sourced by the quadrupole moment $I^{ij} = \sum_K M_K \mathbf{z}_K^{\langle i} \mathbf{z}_K^{j\rangle}$. The binary as a whole behaves like an effective point particle from the perspective of the long-wavelength radiation modes, and is depicted by double solid lines.

At leading order, the scalar field sourced by this dipole moment is

$$\bar{\varphi}(x) = -\int \mathrm{d}^4x'\, D_R^{(\bar{\varphi})}(x,x') \frac{\partial \delta^{(3)}(\mathbf{x}')}{\partial x'^i} \left[\sum_K Q_K(t') \mathbf{z}_K^i(t') \right]. \tag{8.48}$$

Diagrammatically, this solution is represented by Figure 8.7(a). Note that, in contrast to the potential-mode φ, the propagator $D_R^{(\bar{\varphi})}$ for this radiation mode must not be expanded about its instantaneous limit to allow for the fact that $\bar{\varphi}$ can go on shell. Now using the master integral in Appendix 7.A, we find that

$$\bar{\varphi}(x) = \sum_m \int \frac{\mathrm{d}\omega}{2\pi}\, e^{-i\omega t} \mathcal{A}_{\omega 1m}\, h_1^+(kr) Y_{1m}(\hat{\mathbf{x}}). \tag{8.49}$$

Recall that the radial momentum k satisfies $k^2 = \omega^2 - \mu^2$, while $h_1^+(z)$ is a spherical Hankel function. By projecting the dipole moment $O^i(t) := \sum_K Q_K(t)\, \mathbf{z}_K^i(t)$ onto a basis of spherical harmonics $Y_{\ell m}(\hat{\mathbf{x}})$ as per (7.A.11), the amplitude $\mathcal{A}_{\omega 1m}$ for scalar dipole radiation is found to be

$$\mathcal{A}_{\omega 1m} = \frac{i}{3} k^2 \sum_K \int \mathrm{d}t'\, Q_K(t')\, |\mathbf{z}_K(t')|\, Y_{1m}^*(\hat{\mathbf{z}}_K(t'))\, e^{i\omega t'}. \tag{8.50}$$

The rate at which these scalar waves carry energy off to infinity can then be computed by integrating the (t,r) component of this scalar's energy–momentum tensor over a spherical shell of radius r and taking the limit $r \to \infty$. Looking ahead, a result that we will prove in Appendix 9.C tells us that the energy flux $\dot{E}_{\mathrm{SW}}$ radiated into scalar waves is related to the amplitude $\mathcal{A}_{\omega 1m}$ by

$$\int_{-\infty}^{\infty} \mathrm{d}t\; \dot{E}_{\mathrm{SW}} = \sum_m \int \frac{\mathrm{d}\omega}{2\pi} \frac{\omega}{k}\, \theta(k^2) |\mathcal{A}_{\omega 1m}|^2. \tag{8.51}$$

Dividing this equation by the total time $\int \mathrm{d}t \equiv 2\pi\delta(0)$ then gives us the time-averaged scalar flux

$$\langle \dot{E}_{\mathrm{SW}} \rangle = \frac{(\varepsilon G M \mu)^2 (GM\Omega)^{8/3} \nu^2}{6G} \left(\frac{M_1 - M_2}{M} \right)^2 \left[v_+^3 \frac{\Omega_+^4}{\Omega^4} + \theta(\Omega_- - \mu) v_-^3 \frac{\Omega_-^4}{\Omega^4} \right] \tag{8.52}$$

emitted by a binary of total mass $M = M_1 + M_2$ and symmetric mass ratio $\nu := M_1 M_2 / M^2$ ($\leq 1/4$),

having assumed for simplicity that the two black holes are spherical (i.e., nonspinning) and that they are travelling on a circular orbit of frequency Ω.

Three worthy observations can be made here. First, the terms in square brackets indicate that scalar waves emanate at two distinct frequencies, $\Omega_\pm = \Omega \pm \mu$. This is to be expected, since the Fourier transform of the dipole moment $O^i(t)$ is built from the convolution of Q_K and $\mathbf{z}_K$. The two waves travel with different group velocities, $v_\pm = (1 - \mu^2/\Omega_\pm^2)^{1/2}$. Second, the appearance of the step function θ indicates that the larger-frequency mode Ω_+ is radiated throughout the entire history of the inspiral, whereas the lower-frequency mode Ω_- is radiated only when $\Omega_- > \mu$, or equivalently, when $\Omega > 2\mu$. This discontinuity stems from the fact that only sources with frequencies greater than the scalar's mass can deposit energy into on-shell modes. Third, observe that the flux vanishes entirely in the equal-mass limit. This property can be traced back to the definition of the dipole moment, $O^i = \sum_K Q_K \mathbf{z}_K^i$, which vanishes in our coordinate system when $M_1 = M_2$. More generally, $O^i \to Q_1(t)\mathbf{X}^i$ in the equal-mass limit, where the vector $\mathbf{X}^i$ marks the position of the binary's barycentre. This is a constant vector at leading PN order; hence, the dipole moment now oscillates at the single frequency $\sim \mu$, which is not large enough to generate on-shell radiation modes.

It is instructive to compare this dipolar flux of scalar waves with the leading-order, quadrupolar flux of gravitational radiation. The latter process is depicted in Figure 8.7(b), and it is well known that the energy flux radiated into gravitational waves by a circular binary is[4]

$$\mathcal{F}_{\rm GW} = \frac{32}{5}\frac{G^4 M_1^2 M_2^2 (M_1 + M_2)}{a^5} = \frac{32}{5G}(GM\Omega)^{10/3}\nu^2 \tag{8.53}$$

at leading PN order [447, 505]. Taking the limit $\Omega \gg \mu$ for simplicity, it follows that

$$\frac{\langle \dot{E}_{\rm SW}\rangle}{\mathcal{F}_{\rm GW}} \sim \frac{5}{96}\frac{(\varepsilon GM\mu)^2}{v^2}\left(\frac{M_1 - M_2}{M}\right)^2. \tag{8.54}$$

For illustrative values of $v \sim 0.1$ and $(\varepsilon GM\mu)^2 \sim 2 \times 10^{-16}$ [cf. (8.25)], this ratio is at most $\sim 10^{-15}$; hence, the amount of scalar radiation that a binary black hole emits when it is embedded in a realistic FDM halo is always negligible.

The same conclusion can also be drawn by establishing power counting rules for the radiation modes. First, it is important to highlight that the result in (8.52) hinges on the validity of the multipole expansion in (8.46). This holds if the mode with the larger frequency Ω_+ has a wavelength $1/|\mathbf{p}_+| \gg a$. Writing this as $a^2\mathbf{p}_+^2 = a^2\Omega^2 + 2a^2\mu\Omega \ll 1$, we obtain two necessary conditions: $a^2\Omega^2 \ll 1$ and $a^2\mu\Omega \ll 1$.

[4] A word on notation: Both $\dot{E}$ and $\mathcal{F}$ are being used to denote energy fluxes radiated to infinity. I use $\dot{E}$ to refer to fluxes that arise only when the binary interacts with an ambient field, whereas $\mathcal{F}$ refers to fluxes that are inextricable from the binary itself. Being able to make this distinction will be useful in the next chapter.

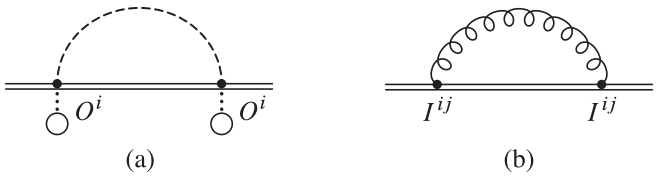


Figure 8.8 Leading-order Feynman diagrams for (a) scalar and (b) gravitational radiation reaction.

The virial theorem can then be used to rewrite these as $v^2 \ll 1$ and $\mu a v \ll 1$, respectively, which we should recognise from earlier as being exactly the conditions that allowed us to expand the potential-mode scalar's propagator about its instantaneous limit; cf. the discussion above (8.24).

To simplify the power counting rules, let us concentrate on the later stages of the inspiral ($\Omega \gg \mu$), when radiative effects are most pronounced. The radiation-mode scalar $\bar{\varphi}$ then has typical momentum scaling as $(p^0, \mathbf{p}) \sim \Omega \sim v/a$. Consequently, its propagator $\langle \bar{\varphi}\bar{\varphi} \rangle \sim \int \mathrm{d}^4 p / p^2 \sim (v/a)^2$, from which we deduce that $\bar{\varphi} \sim v/a$ when appearing inside a Feynman diagram. Identical reasoning implies that $\bar{h} \sim v/a$. In position space, these arguments tell us that derivatives acting on the radiation modes all scale as $\partial_\mu \sim v/a$. Combined with the power counting rules established earlier (see Table 8.1), it follows that the vertex in Figure 8.7(a) scales as $\int \mathrm{d}t \, O^i \partial_i \bar{\varphi} \sim \sqrt{L} v^{3/2} (\varepsilon G M \mu)$.

When integrating out the radiation modes, two copies of this vertex linked by a propagator generate a term in Γ that scales as $L v^3 (\varepsilon G M \mu)^2$ [see Figure 8.8(a)]; hence, scalar radiation reaction first appears at 1.5PN order, albeit suppressed by two powers of $\varepsilon G M \mu$. In contrast, gravitational radiation reaction is a 2.5PN-order effect, or in other words, Fig. 8.8(b) $\sim L v^5$. Taking the ratio of these two diagrams, we find that scalar radiation reaction is suppressed by $\sim (\varepsilon G M \mu)^2 / v^2$ relative to its gravitational counterpart, in agreement with (8.54).

8.3 Observational constraints

Of the five effects discussed in the previous section, those that were said to be of pre-Newtonian order can be seen to exert the greatest influence on the binary black hole, owing to their scaling with negative powers of v. For binary systems of supermassive black holes, future space-based gravitational-wave detectors like LISA are unlikely to be sensitive to these effects, however, as they are most pronounced during the early stages of the inspiral, which may occur outside the LISA band.[5] Accordingly, pulsar timing arrays or other astronomical observations may prove to be more suitable for probing these effects. To conclude this chapter, we will explore how well current astronomical data on supermassive binary black holes constrain the local density of fuzzy dark matter.

[5]LISA will be much more sensitive to the early inspiral of stellar-mass binaries [506], but see also Ref. [507].

The quasar OJ 287 is a bright source of doubly-peaked quasiperiodic flares, whose most plausible explanation is a supermassive binary black hole central engine [314–316]. Each peak in the light curve coincides with the passage of the secondary black hole punching through the primary's accretion disk, which occurs twice per orbit. Apart from the negligible fraction of energy lost during this impact, the evolution of the inspiral is well described by standard post-Newtonian theory in vacuum. By fitting these equations to the observed flares, Dey et al. deduce to within an uncertainty of 6% that the binary's orbital period decays at a rate $\dot{P} \sim 10^{-3}$ [316].

If we now suppose that this binary is located in the central core of an FDM halo, it must be that the additional effects due to the scalar neither hasten nor stall its inspiral by more than $\Delta\dot{P} = 6 \times 10^{-5}$ (assuming, for simplicity, that there are no degeneracies between these scalar effects and the predictions of PN theory in vacuum). This condition can then be translated into an upper bound on the FDM density ρ in the vicinity of the quasar. Although PN corrections are needed to accurately predict how the inspiral evolves as a result of gravitational radiation, it suffices to treat the effects involving the scalar field as first-order perturbations to the binary's Keplerian orbit when calculating their contribution to $\dot{P}$. The general method for performing this kind of calculation is reviewed in Appendix 8.A. Here, we will simply quote the relevant formula for the secular decay of the binary's orbital period,

$$\langle \dot{P} \rangle = \frac{3GM}{(GM\Omega)^{4/3}} \int_0^{2\pi} \mathrm{d}u \, [(\mathbf{f} \cdot \hat{\mathbf{v}}_\perp)\sqrt{1-e^2} + (\mathbf{f} \cdot \hat{\mathbf{r}})\, e \sin u]. \tag{8.55}$$

The external force (per unit mass) acting on this binary is denoted by $\mathbf{f}$, while Ω and e are, respectively, the frequency and eccentricity of its unperturbed orbit. The unit vector $\hat{\mathbf{r}}$ points along the displacement $\mathbf{r} = \mathbf{z}_1 - \mathbf{z}_2$ between the two black holes, whereas $\hat{\mathbf{v}}_\perp$ points along the direction orthogonal to both $\hat{\mathbf{r}}$ and the binary's angular momentum vector. The integral on the rhs is evaluated with respect to the eccentric anomaly u, which is related to the coordinate time t via Kepler's equation, $\Omega(t - t_0) = u - e \sin u$. Finally, the orbital parameter t_0 is a constant called the time of pericentre passage, and can be set to zero in this calculation without loss of generality.

Of the five effects discussed in Section 8.2, the forces due to accretion and the background gravitational potential of the halo provide the largest contributions to $\dot{P}$, since they scale with the most negative powers of v. Respectively, they exert the forces

$$\mathbf{f}_{\mathrm{acc}} = -\left(\frac{\delta\dot{M}_1}{M_1^2} + \frac{\delta\dot{M}_2}{M_2^2}\right) M v \dot{\mathbf{r}} - \frac{G(\delta M_1 + \delta M_2)}{r^2}\hat{\mathbf{r}}, \tag{8.56}$$

$$\mathbf{f}^i_{\mathrm{bkg}} = -R^i{}_{0j0}(t, \mathbf{0})\, \mathbf{r}^j. \tag{8.57}$$

Table 8.2 Parameters for the supermassive binary black hole in quasar OJ 287, reproduced from Dey et al. [316]. The intrinsic period P is determined by rescaling the value measured on Earth by the cosmological scale factor $(1+z)^{-1}$ [314]. Errors have been omitted for any quantity accurate to at least three significant figures.

Parameter		Value
Redshift	z	0.306
Primary black hole mass	M_1	$1.83 \times 10^{10}\ M_\odot$
Secondary black hole mass	M_2	$1.50 \times 10^{8}\ M_\odot$
Spin of the primary black hole	S_1	$0.381\ GM_1^2$
Eccentricity	e	0.657
Intrinsic orbital period	P	9.24 yr
Orbital period decay (1σ uncertainty)	$\dot{P}$	$(99 \pm 6) \times 10^{-5}$

Substituting these into (8.55) gives us an expression for $\langle \dot{P} \rangle$ that is a function of the local density ρ, which we wish to constrain, the known orbital parameters as summarised in Table 8.2, and one unknown: the phase factor Υ of the background relative to our zero of time. Not knowing what value this parameter ought to take, a conservative estimate for $\langle \dot{P} \rangle$ can be obtained by marginalising over Υ assuming a uniform prior.[6] The resulting expectation value is

$$\mathbb{E}[\langle \dot{P} \rangle] = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\Upsilon\, \langle \dot{P} \rangle. \tag{8.58}$$

This procedure automatically eliminates any contribution coming from the halo's gravitational potential, as $|\mathbf{f}_{\rm bkg}| \propto \cos(2\mu t + 2\Upsilon)$. That being said, one can still use this effect to extract a meaningful constraint by manually tuning Υ to give the maximum possible value of $|\langle \dot{P} \rangle|$ (as is done for the case of pulsar binaries in Ref. [296]), but for our purposes, it will suffice to just focus on the force due to accretion. As it turns out, the constraint derived from $\mathbf{f}_{\rm acc}$ is several orders of magnitude better than what one would get from $\mathbf{f}_{\rm bkg}$, even though the latter is formally of lower PN order, because the orbital decay $\langle \dot{P} \rangle$ is much more sensitive to forces that have a component along $\hat{\mathbf{v}}_\perp$. Said differently, $\mathbf{f}_{\rm acc}$ has a greater impact on $\langle \dot{P} \rangle$ because it has a component that is always opposing the binary's motion.

After combining (8.55), (8.56), and (8.58) with additional formulae provided in Appendix 8.A, one finds that the orbital period decay due to accretion is

$$\mathbb{E}[\langle \dot{P} \rangle]_{\rm acc} = -\frac{48\pi G^2 M P (2\nu - e)\rho}{1-e}. \tag{8.59}$$

This result assumes that the black holes are spherical, even though the spin of the primary black hole

[6] By randomly sampling values of $\Upsilon \in [0, 2\pi)$ and observing how they affect the value of $\langle \dot{P} \rangle$, one can check that this assumption of a uniform prior does not bias the conclusions.

has been measured (see Table 8.2), as this is good enough for deriving an order-of-magnitude constraint. Now requiring that (8.59) have a magnitude less than $\Delta\dot{P} = 6 \times 10^{-5}$ imposes the 1σ upper bound

$$\rho \lesssim 2 \times 10^9 \; M_\odot \, \mathrm{pc}^{-3} \tag{8.60}$$

in the vicinity of OJ 287. As a final step, it is necessary to check that the parameters of this binary are within the EFT's regime of validity. The IR cutoff in (8.44) is easily satisfied, as the binary's orbital separation $a \approx 56$ mpc, but the other cutoff in (8.24) tells us that our calculations are valid only for scalar fields with a mass $\mu \ll 8 \times 10^{-22}$ eV.

The constraint in (8.60) is very weak, as typical FDM halos are expected to have core densities of around 100 $M_\odot \, \mathrm{pc}^{-3}$ when $\mu \sim 10^{-22}$ eV [307]. One therefore concludes that these halos are too dilute to leave any observable imprint on the inspiral of a binary black hole, at least in the regime amenable to our approximations.[7] Although this is disappointing, it is also not surprising, given the large separations of scales inherent in this system. Nonetheless, the work in this chapter has been useful for illustrating how quantitative predictions can be made within our EFT framework. In the next chapter, we will further demonstrate the predictive power of this EFT by exploring an interesting phenomenon that arises when the ambient scalar carries a considerable amount of angular momentum.

Appendix 8.A Orbital dynamics and the perturbed Kepler problem

Borrowing heavily from Chapter 3 of Poisson and Will [508], this appendix provides a brief review on the general theory for solving the perturbed Kepler problem,

$$\ddot{\mathbf{r}} + \frac{GM}{r^2}\hat{\mathbf{r}} = \mathbf{f}\,. \tag{8.A.1}$$

The constant M is the total mass of the binary, $\mathbf{r} = \mathbf{z}_1 - \mathbf{z}_2$ is the displacement vector between its two constituents, and $\mathbf{f}$ is some external force (per unit mass), which we will treat as a small perturbation.

8.A.1 Keplerian orbits

Let us set $\mathbf{f} = \mathbf{0}$ and solve the unperturbed problem first. It is useful to project this equation onto an orthonormal basis comprised of $\hat{\mathbf{r}}$ and two other unit vectors. One of these is taken to be along the angular momentum vector $\mathbf{L} = \mathbf{r} \times \dot{\mathbf{r}}$, from which we define $L = |\mathbf{L}|$ and $\hat{\mathbf{L}} \equiv \mathbf{L}/L$. The remaining

[7] Recent studies suggest that binary black holes outside the central core may find their orbital inspiral stalled at kiloparsec scales due to interactions with FDM fluctuations, which pump energy into the orbit [259, 503].

basis vector can then be taken to be $\hat{\mathbf{v}}_\perp := \hat{\mathbf{L}} \times \hat{\mathbf{r}}$. It is well known that the angular momentum of this system is conserved; hence, without loss of generality, we may orient our coordinate system such that $\mathbf{L}$ points along the positive z axis; i.e., $\hat{\mathbf{L}} = \mathbf{e}_z$. In cylindrical polar coordinates (r, ϑ, z), we then have that

$$\mathbf{r} = r\,\hat{\mathbf{r}}, \qquad \dot{\mathbf{r}} = \dot{r}\,\hat{\mathbf{r}} + r\dot{\vartheta}\,\hat{\mathbf{v}}_\perp, \qquad \ddot{\mathbf{r}} = (\ddot{r} - r\dot{\vartheta}^2)\,\hat{\mathbf{r}}. \tag{8.A.2}$$

Three conserved quantities for this system can be written down: the energy (per unit mass), angular momentum (per unit mass), and Runge–Lenz vector are, respectively,

$$E = \frac{\dot{r}^2}{2} + \frac{L^2}{2r^2} - \frac{GM}{r}, \tag{8.A.3a}$$

$$\mathbf{L} = r^2\dot{\vartheta}\,\hat{\mathbf{L}}, \tag{8.A.3b}$$

$$\mathbf{A} = \frac{\dot{\mathbf{r}} \times \mathbf{L}}{GM} - \hat{\mathbf{r}}. \tag{8.A.3c}$$

Conic sections Projecting (8.A.1) with $\mathbf{f} = \mathbf{0}$ along the radial direction yields

$$\ddot{r} - \frac{L^2}{r^3} + \frac{GM}{r} = 0, \tag{8.A.4}$$

having used (8.A.3b) to eliminate $\dot{\vartheta}$. The same equation can be used again to also eliminate t in favour of having ϑ as the independent variable. One then arrives at the equation

$$\frac{\mathrm{d}^2}{\mathrm{d}\vartheta^2}\left(\frac{1}{r}\right) + \frac{1}{r} = \frac{1}{\ell}, \tag{8.A.5}$$

which is just the equation for a harmonic oscillator; hence, the general solution is

$$r = \frac{\ell}{1 + e\cos(\vartheta - \psi)}. \tag{8.A.6}$$

The constant $\ell \equiv L^2/(GM)$ is called the semi-latus rectum, while the integration constants e and ψ are called the eccentricity and longitude of pericentre, respectively. Note that solutions with (e, ψ) are identical to those with $(-e, \psi + \pi)$; hence, we need only consider solutions with positive eccentricity. In what follows, we will focus purely on bound orbits, which have $0 \leq e < 1$.

The longitude of pericentre ψ specifies the orientation of the orbit in the x–y plane, since r is at its minimum value $r_{\min} = \ell/(1 + e)$ when $\vartheta = \psi$. The maximum value, at which point the orbit is said to be at apocentre, is given by $r_{\max} = \ell/(1 - e)$. The average of $r_{\min}$ and $r_{\max}$ defines for us the semi-major axis

$$a := \frac{r_{\max} + r_{\min}}{2} = \frac{\ell}{1 - e^2}. \tag{8.A.7}$$

Differentiating (8.A.6) with respect to time and using (8.A.3b), one finds that

$$\dot{r} = \sqrt{\frac{GM}{\ell}} e \sin(\vartheta - \psi), \tag{8.A.8a}$$

$$\dot{\vartheta} = \sqrt{\frac{GM}{\ell^3}} [1 + e\cos(\vartheta - \psi)]^2. \tag{8.A.8b}$$

From these solutions, it can be shown that $L = \sqrt{GM\ell}$ and $E = -GM/(2a)$; signifying that ℓ is a proxy for the angular momentum in the system and a the total energy.

Eccentric anomaly It remains for us to determine how the coordinates (r, ϑ) evolve with time. In principle, this is done by integrating (8.A.8b) and inverting the result to get $\vartheta \equiv \vartheta(t)$, although the integral cannot be evaluated directly in its current form. To proceed, we define a new variable u called the eccentric anomaly, which is related to the true anomaly $w := \vartheta - \psi$ by the relations

$$\cos w = \frac{\cos u - e}{1 - e\cos u}, \qquad \sin w = \frac{\sqrt{1-e^2}\sin u}{1 - e\cos u}. \tag{8.A.9a}$$

These can be inverted to read

$$\cos u = \frac{\cos w + e}{1 + e\cos w}, \qquad \sin u = \frac{\sqrt{1-e^2}\sin w}{1 + e\cos w}. \tag{8.A.9b}$$

With this substitution, (8.A.8b) integrates to

$$t - t_0 = \sqrt{\frac{a^3}{GM}} (u - e\sin u). \tag{8.A.10}$$

This is known as Kepler's equation, and the integration constant t_0 is called the time of pericentre passage. As the eccentric anomaly u increases by 2π when the system makes a complete orbit, it is easy to see that the orbital period is given by $P = 2\pi\sqrt{a^3/(GM)}$. This is Kepler's third law.

Replacing ϑ with u and substituting a for ℓ, the solutions in (8.A.6) and (8.A.8) have the following useful expressions:

$$r = a(1 - e\cos u), \tag{8.A.11a}$$

$$\dot{r} = \sqrt{\frac{GM}{a}} \frac{e\sin u}{1 - e\cos u}, \tag{8.A.11b}$$

$$\dot{\vartheta} = \sqrt{\frac{GM}{a^3}} \frac{\sqrt{1-e^2}}{(1 - e\cos u)^2}, \tag{8.A.11c}$$

$$\dot{u} = \sqrt{\frac{GM}{a^3}} \frac{1}{1 - e\cos u}. \tag{8.A.11d}$$

8.A.2 Osculating orbital elements

Now consider introducing an external force (per unit mass) $\mathbf{f}$ acting on this system. While it is certainly possible to solve the problem directly — either by analytic methods in the rare instances when $\mathbf{f}$ is sufficiently simple, or numerically otherwise — such an approach provides little insight into how the evolution of the orbit is affected by $\mathbf{f}$. Instead, we shall favour a clever reformulation of the problem known as the method of osculating orbital elements.

In the unperturbed case, let us denote the solution for the vector $\mathbf{r}$ and its velocity $\mathbf{v} \equiv \dot{\mathbf{r}}$ by

$$\mathbf{r}(t) = \mathbf{r}_{\rm Kepler}(t, \Pi), \qquad \mathbf{v}(t) = \mathbf{v}_{\rm Kepler}(t, \Pi), \tag{8.A.12}$$

where Π refers to the set of six orbital elements, $\Pi = (\ell, e, t_0, \psi, \iota, \Theta)$, the first four of which we have already defined. The angular variables ι and Θ are the inclination and longitude of the ascending node, respectively. Together with ψ, they specify the orientation of the orbit relative to an observer with Cartesian coordinates (X, Y, Z). Writing $\mathrm{C}_x \equiv \cos x$ and $\mathrm{S}_x = \sin x$ for brevity, the transformation between the orbit's body-fixed frame (x, y, z) and the space-fixed frame (X, Y, Z) is given by

$$\begin{pmatrix} \mathbf{e}_x \\ \mathbf{e}_y \\ \mathbf{e}_z \end{pmatrix} = \begin{pmatrix} \mathrm{C}_\Theta \mathrm{C}_\psi - \mathrm{C}_\iota \mathrm{S}_\Theta \mathrm{S}_\psi & \mathrm{S}_\Theta \mathrm{C}_\psi + \mathrm{C}_\iota \mathrm{C}_\Theta \mathrm{S}_\psi & \mathrm{S}_\iota \mathrm{S}_\psi \\ -\mathrm{C}_\Theta \mathrm{S}_\psi - \mathrm{C}_\iota \mathrm{S}_\Theta \mathrm{C}_\psi & -\mathrm{S}_\Theta \mathrm{S}_\psi + \mathrm{C}_\iota \mathrm{C}_\Theta \mathrm{C}_\psi & \mathrm{S}_\iota \mathrm{C}_\psi \\ \mathrm{S}_\iota \mathrm{S}_\Theta & -\mathrm{S}_\iota \mathrm{C}_\Theta & \mathrm{C}_\iota \end{pmatrix} \begin{pmatrix} \mathbf{e}_X \\ \mathbf{e}_Y \\ \mathbf{e}_Z \end{pmatrix}. \tag{8.A.13}$$

Moving forward, it will be convenient to adjust the orientation of the orbit in the (x, y, z) frame such that $w = 0$ coincides with the particle crossing the x axis (rather than $\vartheta = 0$ previously). With this minor modification, we have that $\hat{\mathbf{r}} = \cos w\, \mathbf{e}_x + \sin w\, \mathbf{e}_y$ and $\hat{\mathbf{v}}_\perp = -\sin w\, \mathbf{e}_x + \cos w\, \mathbf{e}_y$. It then follows from using (8.A.13) that

$$\begin{aligned} \hat{\mathbf{r}} = {} & [\cos\Theta \cos(\psi + w) - \cos\iota \sin\Theta \sin(\psi + w)]\, \mathbf{e}_X \\ & + [\sin\Theta \cos(\psi + w) + \cos\iota \cos\Theta \sin(\psi + w)]\, \mathbf{e}_Y \\ & + \sin\iota \sin(\psi + w)\, \mathbf{e}_Z, \end{aligned} \tag{8.A.14a}$$

$$\begin{aligned} \hat{\mathbf{v}}_\perp = {} & [-\cos\Theta \sin(\psi + w) - \cos\iota \sin\Theta \cos(\psi + w)]\, \mathbf{e}_X \\ & + [-\sin\Theta \sin(\psi + w) + \cos\iota \cos\Theta \cos(\psi + w)]\, \mathbf{e}_Y \\ & + \sin\iota \cos(\psi + w)\, \mathbf{e}_Z. \end{aligned} \tag{8.A.14b}$$

In the presence of an external force, the method of osculating orbital elements instructs us to use the Keplerian solutions for $\mathbf{r}(t)$ and $\mathbf{v}(t)$ in (8.A.12) as an ansatz while promoting the orbital elements to functions of time, $\Pi \to \Pi(t)$. Differential equations for the evolution of these orbital elements can then be obtained by appealing to the definitions of the angular momentum $\mathbf{L}$ and the Runge–Lenz vector $\mathbf{A}$. Both vectors are conserved and can be written as $\mathbf{L} = \sqrt{GM\ell}\,\mathbf{e}_z$ and $\mathbf{A} = e\,\mathbf{e}_x$ when $\mathbf{f} = 0$, but in the presence of an external force, a torque is exerted on the system such that

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \mathbf{r} \times \mathbf{f}, \tag{8.A.15}$$

$$GM\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} = \mathbf{f} \times \mathbf{L} + \mathbf{v} \times (\mathbf{r} \times \mathbf{f}), \tag{8.A.16}$$

where the second equation comes from differentiating (8.A.3c) and substituting in (8.A.1) and (8.A.15).

The components of (8.A.15) and (8.A.16) give us a total of six, first-order equations for the evolution of the orbital elements. After decomposing $\mathbf{f} = f_r\,\hat{\mathbf{r}} + f_v\,\hat{\mathbf{v}}_\perp + f_z\,\mathbf{e}_z$, one finds

$$\frac{\mathrm{d}\ell}{\mathrm{d}t} = \sqrt{\frac{\ell^3}{GM}}\frac{2f_v}{1+e\cos w}, \tag{8.A.17a}$$

$$\frac{\mathrm{d}\iota}{\mathrm{d}t} = \sqrt{\frac{\ell}{GM}}\frac{f_z\cos(\psi+w)}{1+e\cos w}, \tag{8.A.17b}$$

$$\frac{\mathrm{d}\Theta}{\mathrm{d}t} = \operatorname{cosec}\iota\sqrt{\frac{\ell}{GM}}\frac{f_z\sin(\psi+w)}{1+e\cos w}, \tag{8.A.17c}$$

$$\frac{\mathrm{d}e}{\mathrm{d}t} = \sqrt{\frac{\ell}{GM}}\left(f_r\sin w + f_v\,\frac{2\cos w + e(1+\cos^2 w)}{1+e\cos w}\right), \tag{8.A.17d}$$

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \frac{1}{e}\sqrt{\frac{\ell}{GM}}\left(-f_r\cos w + f_v\sin w\frac{2+e\cos w}{1+e\cos w} - f_z\,e\cot\iota\frac{\sin(\psi+w)}{1+e\cos w}\right). \tag{8.A.17e}$$

The first three of these equations are obtained by successively projecting (8.A.15) along $\mathbf{e}_z$, $\mathbf{e}_Z$, and $\mathbf{e}_X$, while the remaining two come from projecting (8.A.16) along $\mathbf{e}_x$ and $\mathbf{e}_X$, respectively. We require one final equation for the evolution of t_0, although in practice it is more convenient to close this system of equations by obtaining an expression for $\mathrm{d}w/\mathrm{d}t$ instead. By definition, $\cos w = \hat{\mathbf{r}}\cdot\mathbf{e}_x$; hence, differentiating this and using the projection of (8.A.16) onto the plane perpendicular to $\mathbf{e}_x$ yields

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \sqrt{\frac{GM}{\ell^3}}(1+e\cos w)^2 + \frac{1}{e}\sqrt{\frac{\ell}{GM}}\left(f_r\cos w - f_v\sin w\frac{2+e\cos w}{1+e\cos w}\right). \tag{8.A.17f}$$

First-order approximation As it stands, this system of equations is exactly equivalent to — and no less difficult to solve than — the original equation of motion in (8.A.1), but its utility becomes apparent when the components (f_r, f_v, f_z) of $\mathbf{f}$ are suitably small. In this case, an approximate solution can be obtained by using perturbation theory.

The first five equations in (8.A.17) all have the general form

$$\frac{\mathrm{d}\Pi^a}{\mathrm{d}t} = \mathcal{F}^a(w(t); \Pi(t)), \tag{8.A.18}$$

where $\mathcal{F}^a$ $(a = 1, \ldots, 5)$ are functions of the true anomaly $w \equiv w(t)$ and the orbital elements $\Pi \equiv \Pi(t)$. The first-order solution to this equation is

$$\Pi^a(t) = \int \mathrm{d}t\, \mathcal{F}^a(w(t), \Pi), \tag{8.A.19}$$

where the orbital elements appearing on the rhs are to be evaluated using their constant, zeroth-order values. These integrals can still be challenging to perform, however, and so one further simplification is often made. Typically, it is found that orbital elements undergo two types of changes due to an external force. First, they exhibit oscillations that vary on a timescale related to the orbital period P, which are often uninteresting. Second, the orbital elements also undergo a steady drift that does not average out after a few orbital cycles, but accumulates and can eventually lead to large deviations from the initial orbit. We shall focus on these latter, *secular* changes by time-averaging over one orbital cycle.

The secular rate of change of the orbital element Π^a is defined as

$$\langle \dot{\Pi}^a \rangle := \frac{1}{P} \int_0^P \mathrm{d}t\, \mathcal{F}^a(w(t), \Pi). \tag{8.A.20}$$

Rather than integrate over t, it is more convenient to change variables and integrate over u. By the chain rule, we have that

$$\langle \dot{\Pi}^a \rangle = \frac{1}{P} \int_0^{2\pi} \mathrm{d}u \left(\frac{\mathrm{d}t}{\mathrm{d}u} \right) \mathcal{F}^a(w(u), \Pi), \tag{8.A.21}$$

where $\mathrm{d}t/\mathrm{d}u$ is given in (8.A.11d), while (8.A.9) provides the expressions for $w \equiv w(u)$.

8.A.3 Orbital period decay

To conclude this appendix, we should derive the formula in (8.55) for the secular decay of the binary's orbital period. As $\mathrm{d}P/\mathrm{d}t$ is not one of the equations in (8.A.17), the desired result must be pieced

together in several steps. We begin by differentiating (8.A.7) to obtain

$$\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{1}{1-e^2}\frac{\mathrm{d}\ell}{\mathrm{d}t} + \frac{2e\ell}{(1-e^2)^2}\frac{\mathrm{d}e}{\mathrm{d}t} = \sqrt{\frac{a^3}{GM}}\,\frac{2[f_r\, e\sin w + f_v(1+e\cos w)]}{\sqrt{1-e^2}}, \tag{8.A.22}$$

where the second expression follows from substituting in (8.A.17b) and (8.A.17e), and then using (8.A.7) to eliminate ℓ in favour of a. The transformations in (8.A.9) should now be used to replace the coordinate time t with the eccentric anomaly u. The result is

$$\frac{\mathrm{d}a}{\mathrm{d}u} = \frac{\mathrm{d}t}{\mathrm{d}u}\frac{\mathrm{d}a}{\mathrm{d}t} = (1-e\cos u)\sqrt{\frac{a^3}{GM}}\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{2a^3}{GM}(f_v\sqrt{1-e^2} + f_r\, e\sin u). \tag{8.A.23}$$

Finally, Kepler's third law is used to eliminate a in favour of P:

$$\frac{\mathrm{d}P}{\mathrm{d}u} = \frac{\mathrm{d}P}{\mathrm{d}a}\frac{\mathrm{d}a}{\mathrm{d}u} = 3\pi\sqrt{\frac{a}{GM}}\frac{\mathrm{d}a}{\mathrm{d}u} = \frac{3GMP}{(GM\Omega)^{4/3}}(f_v\sqrt{1-e^2} + f_r\, e\sin u), \tag{8.A.24}$$

where $\Omega \equiv 2\pi/P$ is the binary's orbital frequency. The secular rate of change of P is then obtained by integrating the above equation over one orbital cycle and dividing by P:

$$\langle \dot{P} \rangle = \frac{3GM}{(GM\Omega)^{4/3}}\int_0^{2\pi} \mathrm{d}u\,(f_v\sqrt{1-e^2} + f_r\, e\sin u). \tag{8.A.25}$$

For a background scalar field Φ of the form in (8.2), its backreaction onto the spacetime exerts a force $\mathbf{f}_{\rm bkg}$ on the binary, whose components along $\hat{\mathbf{v}}_\perp$ and $\hat{\mathbf{r}}$ are given by

$$f_{v,\rm bkg} = 0, \tag{8.A.26a}$$

$$f_{r,\rm bkg} = 4\pi G\rho\cos(2\mu t + 2\Upsilon)\, r. \tag{8.A.26b}$$

In contrast, the drag force $\mathbf{f}_{\rm acc}$ due to the accretion of Φ onto the black holes (assumed to be spherical for simplicity) has components

$$f_{v,\rm acc} = -64\pi G\rho\nu\,(GM)\sin^2(\mu t + \Upsilon)\, r\dot{\vartheta}, \tag{8.A.27a}$$

$$f_{r,\rm acc} = f_{v,\rm acc}\left(\frac{\dot{r}}{r\dot{\vartheta}}\right) + 8\pi G\rho(2\nu-1)(GM)^2\left(\frac{2\mu t - \sin(2\mu t + 2\Upsilon) + \sin(2\Upsilon)}{\mu r^2}\right). \tag{8.A.27b}$$

In both of these cases, Kepler's equation in (8.A.10) provides the expression for $t \equiv t(u)$, while expressions for r, $\dot{r}$, and $\dot{\vartheta}$ as functions of u are given in (8.A.11).

Chapter 9
Superradiance on binary black hole spacetimes

The goal of this chapter is to describe a novel mechanism by which a long-wavelength scalar field can extract energy and angular momentum from a binary black hole. At the heart of this phenomenon is the absorptive nature of the binary's constituents: Because the ambient scalar field must always remain regular on their event horizons, it is forced to be "dragged" alongside the black holes as they orbit one another. In the time that it takes for a part of this field to irreversibly cross the horizons and deposit energy into the individual black holes, what remains in the bulk of the spacetime is agitated by the motion of the binary and can either gain or lose momentum as a result. Given this description, the close parallels between this process and the reflection of electromagnetic waves off moving conductive boundaries [509–511] is unsurprising, although the underlying mathematics also shares many features in common with superradiant phenomena in general [329, 512, 513]. In this light, it is natural to regard this mechanism as a novel variant of superradiance that is fuelled by orbital motion. Let us give this mechanism a name for ease of reference: In what follows, we will call it *orbital superradiance*.

To illustrate its most salient features, the kinds of scenarios we will consider in this chapter differ from those of Chapters 7 and 8 in two important respects. First, while we have so far been preoccupied with the motion of black holes embedded in much more massive scalar-field environments, we will now consider the opposite limit, in which diffuse configurations of a scalar field evolve around a much more massive binary black hole. Second, while spatial variations in the field have mostly been neglected thus far, they will have an essential role to play in this chapter. Fortunately, these distinctions pose no obstacle to our effective field theory approach, which is sufficiently robust that it can be applied to the study of these systems as well.

The case for orbital superradiance will be made as follows: In Section 9.1, we begin by using the EFT to derive an effective equation of motion for a long-wavelength scalar field propagating on top of a binary black hole background. Owing to the inherent separations of scales in this problem, an approximate solution can be obtained perturbatively via the method of Green's functions, and we do so in Section 9.2 to first order in perturbation theory. The result is then discussed in two stages. In Section 9.3, we first examine the evolution of a long-wavelength scalar cloud that is diffuse enough to engulf the binary as a whole. Not unlike how bosonic fields behave around single rotating black holes, we will find that these clouds are generally comprised of a number of quasibound states that either grow or decay exponentially due to their interactions with the binary. Unlike the case of single black holes, however, the particular geometry of a binary spacetime leads to a number of unique features like mode mixing. In Section 9.3, we then show that the periodic forcing exerted by the binary inevitably converts a fraction of these bound states into outgoing scalar radiation. Also in this section, we show that scalar waves incident on the binary can extract energy from the orbit and undergo amplification when given the right initial conditions. Finally, we conclude in Section 9.5 by revisiting the problem from a thermodynamic viewpoint to gain additional insight.

9.1 Far-zone effective action

Stated broadly, our goal in this chapter is to study the evolution of a real Klein–Gordon field $\phi(x)$ around a binary black hole. Two simplifying assumptions will enable us to address this problem analytically. First, we will assume that the energy density in this field is always suitably dilute that its backreaction onto the spacetime can be neglected as a first approximation. Second, we will restrict our attention to scalar-field configurations whose characteristic length scale λ is much greater than the binary's orbital separation a. Given that the individual black holes cannot be resolved by such a long-wavelength field, the binary as a whole behaves like an effective point particle that interacts with the scalar via a set of dynamical multipole moments.

These interactions are encoded in an effective point-particle action for the binary of the form

$$S_{\rm BBH} = -\int \mathrm{d}\tau\, M + \sum_{\ell=0}^{\infty} \int \mathrm{d}\tau\, O^L(\tau) \nabla_L \phi(\tau) + \cdots . \tag{9.1}$$

The reader should recognise (9.1) as being identical to the point-particle action in (7.4), up to appropriate identifications. In the present context, the constant $M = M_1 + M_2$ is the total mass of the binary, the scalar $\phi(\tau) \equiv \phi(z(\tau))$ is to be evaluated along the position of the binary's barycentre $z^\mu(\tau)$, and the composite operators $\{O^L(\tau)\}$ are dynamical variables that characterise how the ambient scalar interacts

with the binary as a whole. The ellipsis in (9.1) alludes to the presence of analogous composite operators that couple to the gravitational field [379], although these will play no role in our discussion.

9.1.1 Interaction terms from matching

By virtue of being associated with short-distance physics transpiring on length scales $\ll \lambda$ (namely, absorption of the scalar across the black holes' horizons), the composite operators $\{O^L(\tau)\}$ are governed by kinetic terms for which the effective action in (9.1) provides no specific details; hence, additional input in the form of a matching calculation is required. Specifically, the far-zone effective action in (9.1) should be matched onto a post-Newtonian description of the binary that is valid in its near zone. The scalar's interactions with the binary in this latter regime can be described by simply taking two copies of the point-particle action S_p in Chapter 7 — one for each black hole. In fact, as we are treating the scalar as a test field, it suffices to use the simplified version of S_p in (7.17). Written out explicitly in the in–in formalism, the interaction terms for this near-zone EFT are

$$S_{\rm int} = \sum_K \sum_{\ell,\ell'} \int {\rm d}\tau_K {\rm d}\tau'_K \, G^{LL'}_{R,K}(\tau_K, \tau'_K) \nabla_L \phi_-(\tau_K) \nabla_{L'} \phi_+(\tau'_K), \tag{9.2}$$

where $G^{LL'}_{R,K}$ is the retarded Green's function associated with the composite operators $\{q^L_K(\tau_K)\}$, which are localised on the worldline of the Kth black hole. Recall that this Green's function was itself obtained by matching the near-zone effective action onto observables calculated in the full theory.

Further simplification is needed to put (9.2) into a workable form. To that end, we shall assume that the binary is in the early stage of its inspiral, during which the orbital separation a is much greater than the typical size $\sim GM$ of the black holes. As in earlier chapters, it will be convenient to assume that the black holes have comparable masses for the purposes of power counting, although the EFT itself is valid for arbitrary mass ratios as long as $a \gg \max(GM_1, GM_2)$. In this limit, we can expand in powers of the orbital velocity v, allowing us to choose a coordinate chart $x^\mu = (t, \mathbf{x})$ such that $z^\mu_K = (t, \mathbf{z}_K(t))$ and $\tau_K = t + \mathcal{O}(v^2)$. To leading order in v, (9.2) reduces to

$$S_{\rm int} = \sum_K \sum_{\ell,\ell'} \int {\rm d}t {\rm d}t' G^{LL'}_{R,K}(t, t') \partial_L \phi_-(t, \mathbf{z}_K(t)) \partial_{L'} \phi_+(t', \mathbf{z}_K(t')). \tag{9.3}$$

Under the assumption that $a \ll \lambda$, we may now zoom out from this near-zone description by Taylor expanding the scalar field about the binary's barycentre,

$$\partial_L \phi_\pm(t, \mathbf{z}_K(t)) = \sum_{\ell'=0}^{\infty} \frac{1}{\ell'!} \mathbf{z}^{L'}_K(t) \partial_{L+L'} \phi_\pm(t, \mathbf{0}), \tag{9.4}$$

which we have taken to coincide with the origin of our coordinate system. Substituting this expansion back into (9.3) and regrouping terms, the result can be rewritten as

$$S_{\rm int} = \sum_{\ell,\ell'} \int {\rm d}t\,{\rm d}t'\, \mathcal{G}_R^{LL'}(t,t')\, \partial_L \phi_-(t,\mathbf{0})\, \partial_{L'} \phi_+(t',\mathbf{0}), \tag{9.5}$$

where

$$\mathcal{G}_R^{LL'}(t,t') = \sum_K \sum_{p=0}^{\ell} \sum_{p'=0}^{\ell'} G_{R,K}^{PP'}(t,t') \frac{\mathbf{z}_K^{L-P}(t)}{(\ell-p)!} \frac{\mathbf{z}_K^{L'-P'}(t')}{(\ell'-p')!}. \tag{9.6}$$

A comparison between (9.2) and (9.5) reveals that the object $\mathcal{G}_R^{LL'}$ should be interpreted as the retarded Green's function for the composite operators $\{O^L(t)\}$. By analogy with (7.6), this correlation function characterises their dynamical response to external perturbations:

$$O^L(t) = \sum_{\ell'=0}^{\infty} \int {\rm d}t'\, \mathcal{G}_R^{LL'}(t,t')\, \partial_{L'} \phi(t',\mathbf{0}). \tag{9.7}$$

Let us make three further simplifications so as to pick out only the most relevant terms in (9.6). First, power counting tells us that the $(p,p') = (0,0)$ terms always provide the dominant contribution at a given multipole order (ℓ,ℓ'). To see this, note that $|\mathbf{z}| \sim a$ and $G_R^{PP'} \sim (GM)^{p+p'+1}$ [cf. (7.9)]; hence, a term of the form $G_R^{P+1,P'+1}\mathbf{z}^{L-P-1}\mathbf{z}^{L'-P'-1}$ is suppressed by two powers of GM/a relative to the term $G_R^{PP'}\mathbf{z}^{L-P}\mathbf{z}^{L'-P'}$, for instance. As I pointed out in Chapter 7, this scaling analysis indicates that a black hole's absorption cross section is s-wave dominated in the long-wavelength limit [492].

Having discarded terms of higher order in p and p', we are left with just the retarded Green's function $G_{R,K}$ for the monopolar worldline operators q_K. Now assuming that the scalar evolves on a timescale ω^{-1} that is much greater than the black holes' light-crossing times $\sim GM$, this Green's function may be written as an expansion in powers of $GM\omega$. To leading order, $G_{R,K}$ is given by (7.40). Upon substitution back into (9.6), we find that the retarded Green's function for O^L is

$$\mathcal{G}_R^{LL'}(t,t') = \frac{1}{\ell!\ell'!} \sum_K \int \frac{{\rm d}\omega}{2\pi}\, A_K\, i\omega\, e^{-i\omega(t-t')}\, \mathbf{z}_K^L(t)\, \mathbf{z}_K^{L'}(t'), \tag{9.8}$$

where A_K is the area of the Kth black hole. Note that at leading $(p,p') = (0,0)$ order, this correlation function does not capture any of the effects associated with the black holes' spin vectors (which are suppressed by powers of v), but is otherwise valid for black holes of any spin.

We have one last simplification to make. The claim is that only the symmetric, trace-free (STF) projection of the indices L and L' are needed to leading order in a/λ. To see this, consider the

particular case $(\ell,\ell') = (0,2)$ as an example. This correlation function gives rise to the terms

$$S_{\rm int} \supset \int \mathrm{d}t\,\mathrm{d}t' \left[\frac{1}{2}\sum_K \int \frac{\mathrm{d}\omega}{2\pi}\, A_K i\omega\, e^{-i\omega(t-t')}\, \mathbf{z}_K^i(t')\mathbf{z}_K^j(t')\right]\phi_-(t,\mathbf{0})\partial_i\partial_j\phi_+(t',\mathbf{0})$$

$$= \frac{1}{2}\sum_K \int \mathrm{d}t\,\mathrm{d}t' \int \frac{\mathrm{d}\omega}{2\pi}\, A_K i\omega\, e^{-i\omega(t-t')}\phi_-\Big(\mathbf{z}_K^{\langle i}\mathbf{z}_K^{j\rangle}\partial_i\partial_j + \frac{1}{3}|\mathbf{z}_K|^2\partial^i\partial_i\Big)\phi_+ \tag{9.9}$$

in the effective action. In the second line, angled brackets denote the STF projection of a tensor, and all arguments have been suppressed for readability. Now notice that because the scalar satisfies the free Klein–Gordon equation in the absence of these interactions, the second term in (9.9) would be unchanged at leading order were we to replace $\partial^i\partial_i\phi_+ \to (\partial_t^2+\mu^2)\phi_+$. (On a technical note, this sort of removal of redundant operators is achieved by making appropriate field redefinitions [388,470].) As the new term $(\partial_t^2+\mu^2)\phi_+$ no longer depends on spatial derivatives, it can be absorbed into a redefinition of the correlation function $\mathcal{G}_R^{LL'}$ with $(\ell,\ell') = (0,0)$. Power counting tells us that $|\mathbf{z}_K|^2(\partial_t^2+\mu^2) \sim (a/\lambda)^2$; hence, this is a subleading correction. More generally, one finds that traces of $\mathcal{G}_R^{LL'}$ can always be converted into subleading corrections to $\mathcal{G}_R^{L-2N,L'-2N'}$, where n and n' count the number of traces; hence, at leading order in a/λ, it suffices to substitute only the STF part of (9.8) into (9.7). Integrating over ω and t' then gives us the final result:

$$O^L(t) = -\sum_{K=1}^{2}\sum_{\ell'=0}^{\infty}\frac{A_K}{\ell!\ell'!}\,\mathbf{z}_K^{\langle L\rangle}(t)\,\frac{\mathrm{d}}{\mathrm{d}t}\Big[\mathbf{z}_K^{\langle L'\rangle}(t)\,\partial_{L'}\phi(t,\mathbf{0})\Big]. \tag{9.10}$$

Now extremising the total effective action with respect to $\phi_-(x)$ and taking the physical limit, we obtain the equation of motion

$$\left(\Box - \mu^2 + \frac{2GM\mu^2}{r}\right)\phi(x) = -\sum_{\ell=0}^{\infty}(-1)^\ell O^L(t)\partial_L\delta^{(3)}(\mathbf{x}), \tag{9.11}$$

where $\Box-\mu^2$ is the Klein–Gordon operator on flat space. Also included on the lhs is the leading contribution from the binary's gravitational potential. Together, these terms govern the dynamics of the long-wavelength scalar in the bulk of the spacetime, whereas the terms on the rhs are responsible for capturing its interactions with the binary localised at the origin. As the composite operators $O^L(t)$ are symmetric and trace free, the sum over ℓ corresponds to a decomposition of these interactions into distinct multipole moments.

It is worth remarking that the delta function on the rhs of (9.11) inevitably leads to singularities. Likewise, singularities also arise from the operators $O^L(t)$, whose solutions are functions of the scalar

field and its derivatives evaluated at the origin; cf. (9.10). These UV divergences, which in this case originate from the point-particle approximation of the binary, are commonplace in EFTs and can be dealt with in the usual way by introducing a convenient regulator and an appropriate renormalisation scheme [377,401]. This procedure will turn out to be unnecessary, however, when working to first order in perturbation theory.

9.1.2 Expansion parameters

For the sake of clarity, let us reiterate where each of the four expansion parameters in this EFT come from. Two of them are associated with the post-Newtonian approximation of the binary in its near zone — the retarded Green's function in (9.8) is only the leading term in a series organised as an expansion in the binary's orbital velocity $v \sim \sqrt{GM/a}$ and the ratio of timescales $GM\omega$. Treating the entire binary as an effective point particle in the far zone introduces two more expansion parameters: the post-Minkowskian parameter GM/λ and the ratio of length scales a/λ. The first of these characterises the nonlinearity of a given term in the solution due to self-interactions of the gravitational field. In this chapter, we will work to first order in GM/λ and will have to treat it nonperturbatively to capture the bound states of $\phi(x)$, but otherwise all higher-order corrections will be neglected. As for the parameter a/λ, its smallness is what facilitates a decomposition of the binary's interactions with the long-wavelength scalar into distinct multipoles. The solution for each of the Green's functions $\mathcal{G}_R^{LL'}$ is accurate only to leading order in a/λ, but otherwise, terms that scale with different powers of a/λ will be treated on equal footing, despite higher multipole interactions being parametrically suppressed. This is rather unusual from the point of view of an EFT, but doing so will allow us to keep track of the mixing between different angular momentum modes, which leads to interesting consequences.

These four parameters control different aspects of the perturbative expansion, but enforcing the two conditions $v \ll 1$ and $a/\lambda \ll 1$ is often sufficient to ensure that we are inside the EFT's regime of validity. For a scalar cloud that is gravitationally bound to the binary, its characteristic frequency is set by the scalar field's mass, $\omega \simeq \mu$, up to some nonrelativistic binding energy $\sim GM\mu/\lambda_{\rm dB}$. Appearing in the denominator is the de Broglie wavelength $\lambda_{\rm dB} \sim (GM\mu^2)^{-1}$, which sets the characteristic length scale of the cloud. This may be used to recast the condition $a/\lambda \ll 1$ into an upper bound on the scalar's mass,

$$\mu/\Omega \ll 1/v^2, \tag{9.12}$$

which follows after also using $v^2 \sim GM/a$ and $v^3 \sim GM\Omega$, where Ω is the binary's orbital frequency. As a rough guide, (9.12) says that a scalar field should have a mass $\mu \ll 10^{-11}\ \text{eV}\,(v/0.1)\,(M_\odot/M)$ if it is to fully engulf a binary of total mass M and orbital velocity v.

An additional upper bound must be established for freely propagating scalar waves that impinge on the binary with frequency ω and momentum $k \sim 1/\lambda$. For high-momentum modes, we shall choose to replace the necessary condition $a/\lambda \ll 1$ with the sufficient condition $a\omega \ll 1$ for simplicity. The latter equivalently reads

$$\omega/\Omega \ll 1/v. \tag{9.13}$$

For low-momentum modes, the quantity a/λ can be arbitrarily small, so now the UV cutoff for this EFT is set by another expansion parameter, namely, $GM\omega$. Since $\omega \simeq \mu$ in this limit, the condition $GM\omega \ll 1$ is equivalent to $\mu/\Omega \ll v^{-3}$. This upper bound is weaker than that of (9.12), thus we are guaranteed to remain in the EFT's regime of validity as long as the UV cutoffs in (9.12) and (9.13) are both respected.

9.2 Perturbative solution via Green's function

Eliminating $O^L(t)$ from the system of equations in (9.10) and (9.11) gives us a linear, differential equation for $\phi(x)$, but this effective equation of motion is still too difficult to solve exactly. As such, we will obtain an approximate solution by treating the interaction terms on the rhs of (9.11) as small perturbations. Denoting the differential operator on its lhs by D_x^{-1}, this entails looking for a solution of the form $\phi = \phi^{(0)} + \phi^{(1)} + \cdots$, where the zeroth-order piece is an exact solution to the noninteracting theory, $D_x^{-1}\phi^{(0)}(x) = 0$, while the first-order correction is

$$\phi^{(1)}(x) = \int \mathrm{d}^4x'\, D_R(x,x') \sum_{\ell'=0}^{\infty} (-1)^{\ell'} O^{(0)}_{L'}(t')\, \partial_{L'}\delta^{(3)}(\mathbf{x}') + \phi^{(1)}_{\mathrm{cf}}(x). \tag{9.14}$$

In the first term, the retarded Green's function D_R is the inverse of the differential operator D_x^{-1}.

Aside from being analytically tractable, this perturbative approach also has the benefit of an intuitive physical interpretation: The zeroth-order solution $\phi^{(0)}$ behaves in a similar way to the background field Φ in previous chapters in that it induces scalar multipole moments onto the black holes. Having truncated the correlation function in (9.6) to leading order in GM/a ($\sim v^2$), only the induced scalar charges survive in the effective theory. When viewed at large distances, these charges — which are separated by a finite distance a — manifest as induced multipole moments $O^{(0)}_L(t)$ for the binary as whole. These multipole moments then go on to source corrections $\phi^{(1)}$ to the scalar field according to the particular integral in the first term of (9.14). Meanwhile, the second term in this equation is the complementary function, $D_x^{-1}\phi^{(1)}_{\mathrm{cf}}(x) = 0$, whose inclusion may be necessary to ensure that our choice of boundary conditions is satisfied.

In this section, we will discuss the three ingredients that make up the particular integral. We begin by writing down the general solution $\phi^{(0)}$ to the noninteracting theory, which is then fed into (9.10) to obtain an explicit expression for the induced multipoles $O_L^{(0)}(t)$. Finally, taking their convolution with the retarded Green's function $D_R(x, x')$ produces the end result.

9.2.1 The noninteracting theory

To establish some nomenclature and introduce the basis for our perturbative approach, let us begin with a brief review of the Coulomb functions [514, 515]. As the noninteracting theory is time-translation invariant and spherically symmetric, we may look for solutions of the form $\phi(x) \propto R(r)Y_{\ell m}(\hat{\mathbf{x}})e^{-i\omega t}$, where $Y_{\ell m}(\hat{\mathbf{x}})$ are the usual spherical harmonics. The radial part $R(r)$ must then satisfy the equation

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}r^2} + k^2 + \frac{2GM\mu^2}{r} - \frac{\ell(\ell+1)}{r^2}\right) rR(r) = 0, \tag{9.15}$$

where $k^2 \equiv \omega^2 - \mu^2$. The resulting set of solutions can be divided into three categories depending on the value of this quantity. As in Appendix 7.A, the momentum $k \equiv k(\omega)$ in the radial direction is defined as the appropriate root of k^2, namely

$$k(\omega) := \begin{cases} \operatorname{sgn}(\omega)\sqrt{\omega^2 - \mu^2} & (k^2 \geq 0) \\ i\sqrt{\mu^2 - \omega^2} & (k^2 < 0). \end{cases} \tag{9.16}$$

Radiation modes Consider the case $k^2 \geq 0$ to start with. Defining $\zeta := -GM\mu^2/k$, two linearly independent solutions to (9.15) are

$$R_\ell^\pm(k, r) := \frac{H_\ell^\pm(\zeta, kr)}{\pm ikr}, \tag{9.17a}$$

where $H_\ell^\pm$ are Coulomb functions. (We follow the notation in Ref. [397].) From their asymptotic forms at large r, given in (9.A.1), we can deduce that these solutions correspond to ingoing ($-$) and outgoing ($+$) spherical waves. For later purposes, it will also be useful to define a particular linear combination of these *radiation modes*. Let

$$R_\ell(k, r) := \frac{1}{2}[R_\ell^+(k, r) + R_\ell^-(k, r)] \equiv \frac{F_\ell(\zeta, kr)}{kr}, \tag{9.17b}$$

where F_ℓ is another Coulomb function. This solution describes a superposition of ingoing and outgoing

waves in equal measure and is regular at the origin as a result. Multiplying these radial solutions by an appropriate spherical harmonic gives us the mode functions

$$\phi^{\pm}_{k\ell m}(\mathbf{x}) := R^{\pm}_{\ell}(k,r)Y_{\ell m}(\hat{\mathbf{x}}), \tag{9.18a}$$

$$\phi_{k\ell m}(\mathbf{x}) := R_{\ell}(k,r)Y_{\ell m}(\hat{\mathbf{x}}), \tag{9.18b}$$

which describe the three-dimensional spatial profile of these scalar waves.

Yukawa modes Solutions to (9.15) for the case $k^2 < 0$ may now be obtained by analytic continuation. Represented in terms of Whittaker functions [397], they read

$$R^{\pm}_{\ell}(k,r) = (\mp i)^{\ell+1} e^{\pi\zeta/2 \pm i\sigma_\ell(\zeta)} \frac{W_{\mp i\zeta,\ell+1/2}(\mp 2ikr)}{kr}, \tag{9.19a}$$

$$R_{\ell}(k,r) = \frac{C_\ell(\zeta)}{(-2i)^{\ell+1}} \frac{M_{-i\zeta,\ell+1/2}(-2ikr)}{kr}, \tag{9.19b}$$

where the Coulomb phase shift $\sigma_\ell(\zeta)$ and the Gamow factor $C_\ell(\zeta)$ are given by [397, 514]

$$\sigma_\ell(\zeta) := \frac{1}{2i}[\log\Gamma(\ell+1+i\zeta) - \log\Gamma(\ell+1-i\zeta)], \tag{9.20a}$$

$$C_\ell(\zeta) := \frac{\Gamma(\ell+1-i\zeta)}{\Gamma(2\ell+2)} 2^\ell e^{i\sigma_\ell(\zeta)-\pi\zeta/2}. \tag{9.20b}$$

For imaginary k defined according to (9.16), the $R^{-}_{\ell}(k,r)$ solution is seen to grow exponentially with r and is therefore unphysical. In contrast, the function $R^{+}_{\ell}(k,r)$ describes a nonpropagating field profile with characteristic size $\lambda \sim 1/|k|$. Depending on the value of $k \in i\mathbb{R}_{>0}$, this solution can be either singular or regular at $r = 0$. The set of singular solutions includes the $\omega = \ell = 0$ mode, which has the asymptotic form

$$R^{+}_{0}(i\mu,r) \sim \frac{e^{-\mu r}}{r} e^{GM\mu\log(2\mu r)}[1 + \mathcal{O}(r^{-1})] \tag{9.21}$$

at large r, up to some constant prefactor. One easily recognises this as the Yukawa potential sourced by a point charge at the origin, albeit with corrections coming from the gravitational potential of the point mass M. In general, we shall refer to this set of singular solutions as the *Yukawa modes*. Because these modes correspond to having pure imaginary k in the continuous domain $i\mathbb{R}_{>0}$ modulo a discrete set of points to be discussed, the term *continuum states* will be used to refer to the combined set of radiation modes and Yukawa modes.

Bound states The solution $R^{+}_{\ell}(k,r)$ is regular at the origin when its radial momentum k takes special values such that $-i\zeta = n$ is an integer and $n \geq \ell + 1$. These regular solutions have corresponding frequencies $\omega = \pm E_n$ given by

$$E_n = \mu\left(1 - \frac{(GM\mu)^2}{n^2}\right)^{1/2} \simeq \mu - \frac{\mu(GM\mu)^2}{2n^2}, \tag{9.22}$$

which is reminiscent of the Rydberg formula for the bound states in atomic hydrogen. Accordingly, this set of regular solutions will be called the *bound states*. The connection to the hydrogen atom can be made even more explicit by defining $R_{n\ell}(r)$ to be a rescaled version of the solution $R^{+}_{\ell}(k,r)$ when evaluated at $\zeta = in$.[1] In terms of the Whittaker functions, it reads

$$\begin{aligned} R_{n\ell}(r) &:= \frac{(-1)^{n-\ell-1}(-2ik)^{3/2}}{\sqrt{2n(n+\ell)!(n-\ell-1)!}} \frac{W_{-i\zeta,\ell+1/2}(-2ikr)}{-2ikr}\bigg|_{\zeta=in} \\ &\equiv \sqrt{\frac{(n+\ell)!(-2ik)^3}{2n(n-\ell-1)!}} \frac{M_{-i\zeta,\ell+1/2}(-2ikr)}{(2\ell+1)!(-2ikr)}\bigg|_{\zeta=in}. \end{aligned} \tag{9.23}$$

The second line follows from equation (13.14.32) of Ref. [397], which reveals that R^{+}_{ℓ} and R_{ℓ} become proportional to one another when $\zeta \to in$. Note also that $-ik = GM\mu^2/n$ in this limit, thus $R_{n\ell}(r)$ is a real-valued function. With this definition, the orthonormal mode functions

$$\psi_{n\ell m}(\mathbf{x}) := R_{n\ell}(r) Y_{\ell m}(\hat{\mathbf{x}}) \tag{9.24}$$

are exactly the hydrogen wavefunctions,[2] albeit with $GM\mu$ in place of the fine-structure constant.

General solution Because we are treating the black holes' absorption of the scalar field perturbatively via interaction terms, the origin is devoid of sinks or sources in the noninteracting theory. Consequently, the zeroth-order solution $\phi^{(0)}$ must be regular at $r = 0$. This boundary condition precludes the existence of Yukawa modes at this order and, moreover, prohibits a net flux of radiation into or out of the origin.

[1] This rescaling is also necessary on mathematical grounds, as the original solution $R^{+}_{\ell}(k,r)$ vanishes when $\zeta = in$. This behaviour can be traced back to the Coulomb phase shift, as the first gamma function in (9.20a) is being evaluated at one of its poles.

[2] To be precise, I mean the hydrogen wavefunctions for the nonrelativistic Schrödinger equation. These can be obtained from the hydrogen wavefunctions associated with the relativistic Dirac equation, which we discussed earlier in Chapter 3, by performing a Foldy–Wouthuysen transformation. See, e.g., Ref. [516] for details.

The general solution is therefore given by the linear combination

$$\phi^{(0)}(x) = \sum_{\ell,m} \int \frac{\mathrm{d}\omega}{2\pi} 2\mathcal{I}^{>}_{\omega\ell m} \phi_{k\ell m}(\mathbf{x}) e^{-i\omega t} + \frac{1}{\sqrt{2\mu}} \sum_{n,\ell,m} c^{(0)}_{n\ell m} \psi_{n\ell m}(\mathbf{x}) e^{-iE_n t} + \text{c.c.} \tag{9.25}$$

The first term is a sum over a superposition of ingoing and outgoing waves [the factor of 2 follows from the definitions in (9.17b) and (9.18)], where a given mode (ω, ℓ, m) has an ingoing amplitude specified by the coefficient $\mathcal{I}^{>}_{\omega\ell m}$. The "$>$" symbol is used to emphasise that this function can be chosen without loss of generality to have support only in the domain $\omega \geq 0$, corresponding to positive-frequency modes. The negative-frequency modes are then automatically taken into account by the complex conjugate (c.c.) terms. Further note that the boundary conditions on $\phi^{(0)}$ require that $\mathcal{I}^{>}_{\omega\ell m}$ vanishes for $\omega \in (-\mu, \mu)$. Meanwhile, the second term in (9.25) is a sum over bound states, with a conventional prefactor of $1/\sqrt{2\mu}$ included to render the coefficients $c^{(0)}_{n\ell m}$ dimensionless.

9.2.2 Induced multipole moments

The zeroth-order solution in (9.25) may now be fed into (9.10) to obtain the binary's induced multipole moments. Going forward, it will be convenient to introduce some compact notation: Let $u \equiv (n, \ell, m)$ collectively refer to the three integers that specify a bound state and, likewise, let $w \equiv (\omega, \ell, m)$ refer to the parameters for a given continuum state. We then write

$$\sum_u \equiv \sum_{n,\ell,m} \quad \text{and} \quad \sum_w \equiv \sum_{\ell,m} \int \frac{\mathrm{d}\omega}{2\pi} \tag{9.26}$$

to denote summing over the bound and continuum states, respectively. In this new shorthand, the zeroth-order solution reads

$$\phi^{(0)}(x) = \sum_w 2\mathcal{I}^{>}_w \phi_{k\ell m}(\mathbf{x})\, e^{-i\omega t} + \frac{1}{\sqrt{2\mu}} \sum_u c^{(0)}_u \psi_u(\mathbf{x})\, e^{-iE_n t} + \text{c.c.} \tag{9.27}$$

Rather than substitute this directly into (9.10) to produce a set of tensorial objects, it will be more convenient to project $O^L(t)$ onto a basis of STF tensors. As in Appendix 7.A, we do so by utilising the basis vectors $\mathcal{Y}^{\ell m}_L \equiv \mathcal{Y}^{\ell m}_{i_1\cdots i_\ell}$, which generate the spherical harmonics [i.e., $Y_{\ell m}(\hat{\mathbf{x}}) = \mathcal{Y}^{\ell m}_L \hat{\mathbf{x}}^L$] and satisfy the orthogonality relation [471]

$$(\mathcal{Y}^{\ell m}_L)^* \mathcal{Y}^{\ell m'}_L = \frac{(2\ell+1)!!}{4\pi \ell!} \delta^{mm'}. \tag{9.28}$$

Each of the $2\ell + 1$ independent degrees of freedom in $O^L(t)$ can then be extracted via the projection

$$O_{\ell m}(t) = -\frac{1}{4\pi i}\frac{4\pi \ell!}{(2\ell+1)!!}(\mathcal{Y}^{\ell m}_L)^* O^L(t), \tag{9.29}$$

where the prefactor of $-4\pi i$ is included purely for convenience. To reconstruct $O^L(t)$, one simply inverts this relation to find

$$O^L(t) = -4\pi i \sum_{m=-\ell}^{\ell} \mathcal{Y}^{\ell m}_L O_{\ell m}(t). \tag{9.30}$$

Circular orbits At the moment, the formula in (9.10) for these induced multipoles does not make any assumptions about the black holes' trajectories, apart from requiring that $|\mathbf{z}_K(t)| \ll \lambda$. For simplicity, in this chapter we shall restrict our attention to (quasi)circular orbits of frequency Ω, whose angular momentum vectors can be oriented to point along the positive z axis without loss of generality. We may then write

$$\mathbf{z}^i_K(t) = r_K R^i{}_j(\Omega t)\mathbf{d}^j, \tag{9.31}$$

where $r_1 = aM_2/M$ and $r_2 = -aM_1/M$ are the displacements of the black holes from the origin, $\mathbf{d}$ is a unit vector in the $z = 0$ plane that specifies their positions at a reference time $t = 0$, and the appropriate rotation matrix is

$$R^i{}_j(\vartheta) = \begin{pmatrix} \cos\vartheta & -\sin\vartheta & 0 \\ \sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{9.32}$$

Direct evaluation can be used to prove the identity [468]

$$\mathcal{Y}^{\ell m}_{i_1\cdots i_\ell} R^{i_1}{}_{j_1}(\Omega t)\cdots R^{i_\ell}{}_{j_\ell}(\Omega t) = \mathcal{Y}^{\ell m}_{j_1\cdots j_\ell} e^{im\Omega t}, \tag{9.33}$$

which enables us to expand the STF products $\mathbf{z}^{\langle L\rangle}_K$ appearing in (9.10) as

$$\mathbf{z}^{\langle L\rangle}_K(t) = \frac{4\pi \ell!}{(2\ell+1)!!} r^\ell_K \sum_{m=-\ell}^{\ell} Y^*_{\ell m}(\mathbf{d})\mathcal{Y}^{\ell m}_L e^{-im\Omega t}. \tag{9.34}$$

Putting everything together, the components $O_{\ell m}(t)$ are given by

$$O_{\ell m}(t) = \sum_{\ell', m'} \frac{4\pi Y^*_{\ell m}(\mathbf{d}) Y_{\ell' m'}(\mathbf{d}) B_{\ell\ell'}}{(2\ell+1)!!(2\ell'+1)!!} e^{-i(m-m')\Omega t}(m'\Omega - i\partial_t)(\mathcal{Y}^{\ell' m'}_{L'})^* \partial_{L'}\phi(t, \mathbf{0}), \tag{9.35}$$

where the interaction strength between the binary and the scalar is characterised by the quantity

$$B_{\ell\ell'} := \sum_K A_K r_K^{\ell+\ell'}. \tag{9.36}$$

Positive-frequency components The induced multipoles $O^L(t)$ are necessarily real by construction; hence, the definition in (9.30) can be used to show that its components must satisfy the constraint

$$O^*_{\ell m}(t) = -(-1)^m O_{\ell,-m}(t), \tag{9.37}$$

which follows from the identity for the complex conjugate of a spherical harmonic; cf. (9.A.12). This motivates writing

$$O_{\ell m}(t) = O^{>}_{\ell m}(t) - (-1)^m O^{>*}_{\ell,-m}(t) \tag{9.38}$$

such that (9.37) is automatically satisfied for any function $O^{>}_{\ell m}(t)$. It is then easy to show that for a real scalar-field solution of the form $\phi(x) = \phi^{>}(x) + \text{c.c.}$, one obtains $O^{>}_{\ell m}(t)$ by evaluating the rhs of (9.10) using only $\phi^{>}(x)$ rather than $\phi(x)$. In other words, one finds

$$\begin{aligned} O^{(0)>}_{\ell m}(t) = \sum_{w'} \frac{Y^*_{\ell m}(\mathbf{d}) Y_{\ell' m'}(\mathbf{d}) B_{\ell\ell'}}{(2\ell+1)!!} \mathfrak{R}_{\ell'}(k')(m'\Omega - \omega') 2\mathcal{I}^{>}_{w'} e^{-i[\omega' + (m-m')\Omega]t} \\ + \frac{1}{\sqrt{2\mu}} \sum_{u'} \frac{Y^*_{\ell m}(\mathbf{d}) Y_{\ell' m'}(\mathbf{d}) B_{\ell\ell'}}{(2\ell+1)!!} \mathfrak{R}_{n'\ell'}(m'\Omega - E_{n'}) c^{(0)}_{u'} e^{-i[E_{n'} + (m-m')\Omega]t} \end{aligned} \tag{9.39}$$

after substituting only the positive-frequency part of (9.27) into (9.35). In obtaining this result, I have made use of the identities in (9.A.10) and have defined

$$\mathfrak{R}_\ell(k) := \lim_{r\to 0} \frac{1}{\ell!} \frac{\mathrm{d}^\ell}{\mathrm{d}r^\ell} R_\ell(k, r), \qquad \mathfrak{R}_{n\ell} := \lim_{r\to 0} \frac{1}{\ell!} \frac{\mathrm{d}^\ell}{\mathrm{d}r^\ell} R_{n\ell}(r); \tag{9.40}$$

explicit expressions for which are provided in (9.A.11). The result in (9.39) can now be used to determine the first-order correction $\phi^{(1)}$ via the method of Green's functions.

9.2.3 Integration contours for Green's function

The retarded Green's function in (9.14) is defined by the equation $D_x^{-1} D_R(x, x') = -\delta^{(4)}(x - x')$. The time-translation and rotational symmetries of the noninteracting theory allow us to express it in terms of its inverse Fourier transform,

$$D_R(x, x') = \sum_{\ell,m} \int \frac{\mathrm{d}\omega}{2\pi} \widetilde{D}_{\omega\ell}(r, r') Y_{\ell m}(\hat{\mathbf{x}}) Y^*_{\ell m}(\hat{\mathbf{x}}') e^{-i\omega(t-t')}. \tag{9.41}$$

Defining $r_> := \max(r, r')$ and $r_< := \min(r, r')$, the radial part is given by [517, 518]

$$\widetilde{D}_{\omega\ell}(r, r') = (-2ik) \frac{\Gamma(\ell + 1 + i\zeta)}{(2\ell+1)!} \frac{W_{-i\zeta,\ell+1/2}(-2ikr_>)}{-2ikr_>} \frac{M_{-i\zeta,\ell+1/2}(-2ikr_<)}{-2ikr_<}. \tag{9.42}$$

To evaluate the particular integral in (9.14), it helps to notice that the delta function imposes the restrictions $r_< = r'$ and $r_> = r$, which make integrating over $\mathbf{x}'$ relatively straightforward. First integrating by parts to move the derivatives $\partial_{L'}$ onto $G(x, x')$, we obtain

$$\phi^{(1)}(x) \supset \sum_{\ell,m} \int \frac{\mathrm{d}\omega}{2\pi} \left\{ \frac{\Gamma(\ell+1+i\zeta)}{(2\ell+1)!} \sum_{\ell'=0}^{\infty} \int \mathrm{d}^3\mathbf{x}' \delta^{(3)}(\mathbf{x}') \partial_{L'} \left(\frac{M_{-i\zeta,\ell+1/2}(-2ikr')}{-2ikr'} Y^*_{\ell m}(\hat{\mathbf{x}}') \right) \right.$$
$$\left. \times (-2ik) \int \mathrm{d}t' O^{(0)}_{L'}(t') e^{-i\omega t'} \right\} \frac{W_{-i\zeta,\ell+1/2}(-2ikr)}{-2ikr} Y_{\ell m}(\hat{\mathbf{x}}) e^{-i\omega t}. \tag{9.43}$$

Now integrating over $\mathbf{x}'$ while using (9.28), (9.30), and (9.A.9), this can be shown to simplify to

$$\phi^{(1)}(x) \supset \sum_{\ell,m} \int \frac{\mathrm{d}\omega}{2\pi} \left\{ \frac{\Gamma(\ell+1+i\zeta)}{(2\ell)!!} (-i)(-2ik)^{\ell+1} \int \mathrm{d}t' O^{(0)}_{\ell m}(t') e^{i\omega t'} \right\}$$
$$\times \frac{W_{-i\zeta,\ell+1/2}(-2ikr)}{-2ikr} Y_{\ell m}(\hat{\mathbf{x}}) e^{-i\omega t} \tag{9.44}$$

after also using the identity $n! \equiv n!!(n-1)!!$ for the double factorial. Given that $(2\ell)!! = 2^\ell \ell!$, an equivalent expression for the particular integral is

$$\phi^{(1)}(x) \supset \sum_{\ell,m} \int \frac{\mathrm{d}\omega}{2\pi} \left\{ \frac{\Gamma(\ell+1+i\zeta)}{\Gamma(\ell+1)} (-ik)^{\ell+1} \int \mathrm{d}t' O^{(0)}_{\ell m}(t') e^{i\omega t'} \right\}$$
$$\times \frac{W_{-i\zeta,\ell+1/2}(-2ikr)}{kr} Y_{\ell m}(\hat{\mathbf{x}}) e^{-i\omega t}. \tag{9.45}$$

Both expressions will turn out to be useful in later sections.

It remains to perform the integrals over t' and ω. Care must be exercised with the latter because the Green's function contains poles at $\omega = \pm E_n$ and branch points at $\omega = \pm\mu$ and at infinity [517]. To proceed, we split the integral over ω into two parts: Its principal value along the real line gives rise to the continuum states, while the bound states are obtained by integrating over closed contours that encircle each of the poles; see Figure 9.1. The complete first-order solution is thus a sum of three parts:

$$\phi^{(1)} = \phi^{(1)}_b + \phi^{(1)}_c + \phi^{(1)}_{\mathrm{cf}}. \tag{9.46}$$

The first term contains the bound states while the second contains the continuum states. These will be discussed in Sections 9.3 and 9.4, respectively. Finally, recall from our earlier discussion that the third term is a complementary function that may be required to satisfy boundary conditions.

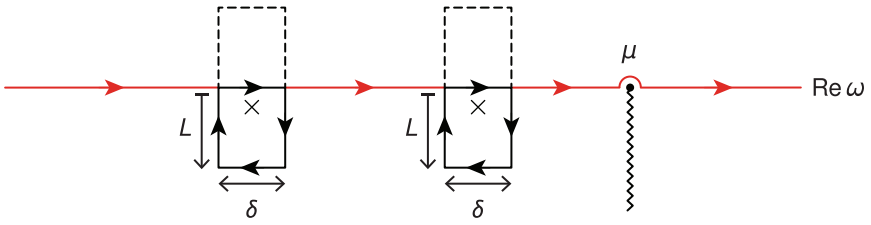


Figure 9.1 Illustration (not to scale) showing the different contributions to the ω integral. The principal value (red line) gives rise to the continuum states, while the bound states come from closed contours (black lines) that encircle each of the poles of the Green's function. The poles (cross marks) are shifted off the real axis to enforce retarded boundary conditions and thus contribute to the solution only when the black contours are closed in the lower half of the complex plane. The limit $\delta \to 0$ and $L \to \infty$ should be taken at the end. Note also the presence of branch points at $\omega = \pm\mu$ and at infinity.

9.3 Bound states

The bound states surrounding a binary evolve in an intricate manner as a result of their interaction with the black holes. In our perturbative approach, this evolution can be regarded as being sourced by the induced multipoles $O_L^{(0)}(t)$. The first-order result $\phi_b^{(1)}$ is obtained by performing the ω integral in (9.44) over closed contours that each encircle one of the poles $\omega = \pm E_n$ of the Green's function.

When $t > t'$, these contours should be closed in the lower half of the complex plane; see Figure 9.1. In the limit $\delta \to 0$ and $L \to \infty$, the integrals along the vertical paths cancel each other out, while the integral over the bottom horizontal path vanishes. The residue theorem can then be applied to show that

$$\phi_b^{(1)}(x) = \sum_{s=\pm1}\sum_{u}(-i)^2\frac{\mathrm{Res}[\Gamma(\ell+1+i\zeta)]}{(2\ell)!!}(-2ik)^{\ell+1}\int \mathrm{d}t'\theta(t-t')O^{(0)}_{\ell m}(t')e^{i\omega t'} \times \frac{W_{-i\zeta,\ell+1/2}(-2ikr)}{-2ikr}Y_{\ell m}(\hat{\mathbf{x}})e^{-i\omega t}\Big|_{\omega=sE_n-i\epsilon}, \tag{9.47}$$

where the poles have been shifted by an amount $-i\epsilon$ to enforce retarded boundary conditions, and the sum over $s = \pm1$ is used to account for both the positive- and negative-frequency solutions. Note also the presence of the step function $\theta(t-t')$, which follows from the fact that the contour should be closed in the upper half of the complex plane when $t < t'$.

To determine the residue of the gamma function at the pole $\omega = sE_n$, use the standard Laurent expansion $\Gamma(-j+z) = (-1)^j/(j!z) + \mathcal{O}(z^0)$ valid for any nonnegative integer j to obtain

$$\mathrm{Res}[\Gamma(\ell+1+i\zeta)]_{\omega=sE_n} = -\frac{(-1)^{n-\ell-1}(GM\mu^2)^2}{(n-\ell-1)!n^3sE_n}. \tag{9.48}$$

Substituting this back into (9.47) and also using (9.23) to replace the Whittaker function in favour of

the radial solution $R_{n\ell}(r)$, one finds

$$\phi_b^{(1)}(x) = \sum_{s=\pm 1}\sum_u \frac{\mathfrak{R}_{n\ell}}{2sE_n}(2\ell+1)!!\,\psi_u(\mathbf{x})\int_{-\infty}^{\infty} \mathrm{d}t'\theta(t-t')O_{\ell m}^{(0)}(t')e^{-i(sE_n - i\epsilon)(t-t')}, \tag{9.49}$$

which can be further simplified to read

$$\phi_b^{(1)}(x) = \sum_u \frac{\mathfrak{R}_{n\ell}}{2E_n}(2\ell+1)!!\,\psi_u(\mathbf{x})\int^{t} \mathrm{d}t' O_{\ell m}^{(0)}(t')e^{-iE_n(t-t')} + \text{c.c.} \tag{9.50}$$

This latter expression follows after making two observations. First, that the $s=-1$ terms in (9.49) are exactly the complex conjugates of the $s=+1$ terms, which one can check by using (9.37) and (9.A.13) along with the freedom to relabel $m \to -m$, as it is being summed over. Second, that the lower bound of the t' integral in (9.49) has vanishing contribution due to the $-i\epsilon$ term in the exponent, thus we need only keep track of the result from the upper bound in (9.50). Discarding this lower bound constitutes no loss in generality, as the freedom to specify initial conditions for the bound states at some initial time, say $t=0$, is provided by our freedom to choose the complementary function $\phi_{\text{cf}}^{(1)}$.

The result in (9.50) is not yet in a useful form because the functions $O_{\ell m}^{(0)}(t')$ contain both the positive- and negative-frequency parts of the zeroth-order solution $\phi^{(0)}$, whose separate contributions we would like to make manifest. To do this, use the decomposition in (9.38) to write

$$\phi_b^{(1)}(x) = \sum_u \frac{\mathfrak{R}_{n\ell}}{2E_n}(2\ell+1)!!\,\psi_u(\mathbf{x})\int^{t} \mathrm{d}t'\big[O_{\ell m}^{(0)>}(t') - (-1)^m O_{\ell,-m}^{(0)>*}(t')\big]e^{-iE_n(t-t')} + \text{c.c.} \tag{9.51}$$

The freedom to swap the terms involving $O_{\ell,-m}^{(0)>*}(t)$ with their complex conjugates and to subsequently relabel $m \to -m$ gives us our final expression:

$$\phi_b^{(1)}(x) = \sum_u \frac{\mathfrak{R}_{n\ell}}{2E_n}(2\ell+1)!!\,\psi_u(\mathbf{x})\int^{t} \mathrm{d}t'\big[O_{\ell m}^{(0)>}(t') - O_{\ell m}^{(0)>}(t)e^{2iE_n(t-t')}\big]e^{-iE_n(t-t')} + \text{c.c.} \tag{9.52}$$

It is now apparent that the full solution $\phi = \phi^{(0)} + \phi^{(1)} + \cdots$ for the bound states has the form

$$\phi_b(x) = \frac{1}{\sqrt{2\mu}}\sum_u \big[c_u(t)\psi_u(\mathbf{x})e^{-iE_n t} + \text{c.c.}\big] \tag{9.53}$$

with $c_u(t) \equiv c_u^{(0)} + c_u^{(1)}(t) + c_{\text{cf},u}^{(1)} + \cdots$. The time evolution of the amplitude is given to first order by

$$c_u^{(1)}(t) = \sqrt{2\mu}\,\frac{\mathfrak{R}_{n\ell}}{2E_n}(2\ell+1)!!\int^{t} \mathrm{d}t' O_{\ell m}^{(0)>}(t')\big(e^{iE_n t'} - e^{2iE_n t}e^{-iE_n t'}\big), \tag{9.54}$$

while $c_{\text{cf},u}^{(1)}$ is a constant term coming from $\phi_{\text{cf}}^{(1)}$ that we tune in order to choose initial conditions. Finally,

substituting the expression for $O^{(0)>}_{\ell m}(t)$ in (9.39) into (9.54) leads to the explicit solution

$$c^{(1)}_u(t) = \sum_{u'} \frac{V_{uu'}}{2E_n}(m'\Omega - E_{n'})c^{(0)}_{u'} \int^t \mathrm{d}t' \big(e^{i\Delta_{uu'}t'} - e^{2iE_nt}e^{i(\Delta_{uu'}-2E_n)t'}\big) + \frac{1}{2\mu}\sum_{w'} \frac{V_{uw'}}{2E_n}(m'\Omega - \omega')2\mathcal{I}^{>}_{w'} \int^t \mathrm{d}t' \big(e^{i\Delta_{uw'}t'} - e^{2iE_nt}e^{i(\Delta_{uw'}-2E_n)t'}\big), \tag{9.55}$$

which is written in terms of the energy differences

$$\Delta_{uu'} = E_n - E_{n'} - (m - m')\Omega, \tag{9.56a}$$

$$\Delta_{uw'} = E_n - \omega' - (m - m')\Omega \tag{9.56b}$$

and the matrix elements

$$V_{uu'} = Y^*_{\ell m}(\mathbf{d})Y_{\ell' m'}(\mathbf{d})B_{\ell\ell'}\mathfrak{R}_{n\ell}\mathfrak{R}_{n'\ell'}, \tag{9.57a}$$

$$V_{uw'} = Y^*_{\ell m}(\mathbf{d})Y_{\ell' m'}(\mathbf{d})B_{\ell\ell'}\mathfrak{R}_{n\ell}\mathfrak{R}_{\ell'}(k')(2\mu)^{3/2}. \tag{9.57b}$$

Note the extra normalisation factor of $1/(2\mu)$ in the second line of (9.55) has been included because the sum over continuum states $\sum_w$ is dimensionful; cf. (9.26). Written in this way, both $V_{uu'}$ and $V_{uw'}$ have dimensions of energy.

9.3.1 Growth rates

We are now in a position to discuss the physical implications of this result. To begin with, suppose that $c^{(0)}_u \neq 0$ and consider its contribution to $c^{(1)}_u(t)$. Since $\Delta_{uu} = 0$, we have that

$$c^{(1)}_u(t) \supset \frac{V_{uu}}{2E_n}(m\Omega - E_n)c^{(0)}_u \int^t \mathrm{d}t' \big(1 - e^{2iE_n(t-t')}\big) = c^{(0)}_u \Gamma_u t + \text{const.} \tag{9.58}$$

As the constant term may be removed by an appropriate choice of $c^{(1)}_{\mathrm{cf},u}$, we see that the physical effect of the diagonal element V_{uu} is to cause the bound state u to grow at the rate

$$\Gamma_u := \frac{V_{uu}}{2E_n}(m\Omega - E_n). \tag{9.59}$$

Thus, the bound states of the noninteracting theory turn into quasibound states once their interaction with the binary is taken into account.

It is worth remarking that the linear growth in (9.58) is an approximation that is valid only at early times $t \ll 1/\Gamma_u$. Once $\Gamma_u t$ becomes of order unity, terms of the form $\sim(\Gamma_u t)^p$ that appear at higher orders in perturbation theory all become relevant. One might naturally expect that resumming these

polynomials to all orders will lead to an exponentially growing solution $c_u(t) \propto \exp(\Gamma_u t)$ and, indeed, this turns out to be the case. The details of this resummation procedure, while interesting on theoretical grounds, have been relegated to Appendix 9.B as they will not be relevant to this chapter's main line of discussion. In what follows, it will suffice to work with the linear approximation in (9.58).

Written out explicitly, the growth rate for the $u \equiv (n, \ell, m)$ mode reads

$$\Gamma_u = \left| \frac{Y_{\ell m}(\mathbf{d})}{(2\ell+1)!} \right|^2 \frac{(n+\ell)!(2GM\mu^2)^{2\ell+3}}{(n-\ell-1)!4n^{2\ell+4}E_n} B_{\ell\ell}(m\Omega - E_n). \tag{9.60}$$

It is instructive to first compare this result with the growth rate of a long-wavelength scalar cloud around a *single* rotating black hole [331, 367, 468]. Strikingly, after identifying the total mass M of the binary with the mass of the single black hole and, likewise, identifying the binary's orbital frequency Ω with the angular frequency of the horizon, the expressions for the two growth rates are seen to be equivalent up to an overall factor associated with differences in the geometry. That these two results are so closely related is not a coincidence, but is a reflection of the fact that a binary and a single black hole both effectively behave like point particles in the long-wavelength limit. Indeed, at leading order in the expansion parameters, the bound (and continuum) states of a Klein–Gordon field on these two spacetimes are mathematically identical. All differences between these two variants of superradiance can therefore be attributed to differences in the corresponding operators $O^L(t)$ that are localised at the origin. (For an EFT approach to single-black-hole superradiance along these lines, see Ref. [468].)

This is not to say that there is nothing novel about orbital superradiance, however. Indeed, the particular "dumbbell" geometry of the binary establishes a selection rule that requires $\ell + m$ to be even if the mode is to interact with the binary. Otherwise, the growth rate Γ_u vanishes. This property can be traced back to the spherical harmonic in (9.60), or more generally to the spherical harmonics in the matrix elements of (9.57), which are being evaluated with respect to the unit vector $\mathbf{d}$ that is confined to be in the $z = 0$ plane. The vanishing of these spherical harmonics has an intuitive physical interpretation: Modes with $\ell + m \notin 2\mathbb{Z}$ correspond to field profiles that are concentrated away from the $z = 0$ plane and are therefore unappreciable in the neighbourhood of the binary; hence, no interaction can occur. If instead $\ell + m \in 2\mathbb{Z}$, a given mode grows if $0 < E_n < m\Omega$ and decays otherwise.

9.3.2 Mode mixing

Growth rates aside, the geometric properties of a binary also give rise to a number of other interesting effects. From the general solution in (9.55), it is clear that even if a given mode u has zero amplitude

at an initial time $t = 0$, the presence of another bound state $u' \neq u$ will seed the growth of u as long as $V_{uu'} \neq 0$. Likewise, energy from ingoing radiation can also be captured and converted into bound states, as the second line in (9.55) demonstrates. In fact, this is the reason why I have chosen to call the objects in (9.57) matrix elements. In analogy with quantum mechanics, they characterise the overlap between different modes as a result of their interaction with the binary.

The appearance of mode mixing is unsurprising in this context, as the underlying spacetime is neither time-translation invariant nor is it rotationally symmetric. Accordingly, the modes $u \equiv (n, \ell, m)$ and $w \equiv (\omega, \ell, m)$ do not remain eigenstates of the interacting theory. Actually, this last statement needs refining because the matrix elements in (9.57) vanish when at least one of $\ell + m$ or $\ell' + m'$ is odd, for reasons already discussed. Thus, modes with $\ell + m \notin 2\mathbb{Z}$ are effectively blind to the presence of the black holes and are conserved (at this order in perturbation theory), while modes with $\ell + m \in 2\mathbb{Z}$ interact with the binary and get mixed into one another.

Let us highlight another consequence of mode mixing: As the solution in (9.55) shows, a given mode u will oscillate not just at its natural frequency E_n, but also at the secondary frequencies $|E_n - \Delta_{uu'}|$, which results in a beating pattern when viewed in the time domain. Crucially, note that some of these secondary frequencies are much greater than the scalar field's mass. If a particle of the same mass were to have an energy given by one of these high frequencies, we would naturally expect it to escape the gravitational potential of the binary and travel off to infinity. Indeed, the same thing happens in the case of a long-wavelength scalar field, as we will show in Section 9.4.

9.3.3 Backreaction and energy extraction

We previously discussed in Chapter 1 that, in many well-motivated scenarios, an ultralight scalar couples only very weakly to the Standard Model. If so, the formation of a scalar cloud and its subsequent evolution around a binary black hole are not processes that can be observed directly, but rather, must be inferred through more indirect means. One possibility is to examine the way it influences the binary's inspiral. At leading order, such effects can be determined without the need to explicitly calculate the scalar's backreaction onto the black holes' trajectories. Instead, we may simply appeal to an energy balance argument: Any secular increase in the total energy E_b of the bound states must have been extracted from the orbital energy stored in the binary.

To gain a sense for how much energy is extracted during this process, consider a simplified scenario in which only a single mode $\hat{u} \equiv (\hat{n}, \hat{\ell}, \hat{m})$ is populated initially. After time-averaging over a period much longer than the other timescales in the problem, one finds that the rate at which energy is extracted

from the binary into this bound state is [339, 353, 357][3]

$$\dot{E}_b = 2\Gamma_{\hat{u}} M_{\hat{u}} \tag{9.61}$$

at first order in the interactions, where $M_{\hat{u}} := \mu \left| c^{(0)}_{\hat{u}} \right|^2$ is the total initial energy in the scalar cloud. Note that the time-averaging procedure eliminates any contribution from mode mixing at this order.

If $\Gamma_{\hat{u}} > 0$, the growth of this mode extracts energy from the binary and causes it to inspiral more rapidly than it would in pure vacuum. On the other hand, if $\Gamma_{\hat{u}} < 0$, this decaying mode injects energy into the orbit while it is being absorbed and will decelerate the inspiral as a result. Whether either of these effects leave an observable imprint depends on the magnitude of (9.61). A useful measure is to compare it to the flux of energy $\mathcal{F}_{\rm GW}$ emitted by the binary into gravitational waves; cf. (8.53). Taking the ratio of (9.61) to (8.53), we find

$$\frac{\dot{E}_b}{\mathcal{F}_{\rm GW}} = \frac{M_{\hat{u}}}{E_{\rm orb}} \frac{2\Gamma_{\hat{u}}}{\Gamma_{\rm GW}}, \tag{9.62}$$

where $E_{\rm orb} = GM_1M_2/(2a)$ is the magnitude of the binary's orbital energy and

$$\Gamma_{\rm GW} = \frac{64}{5} \frac{G^3 M_1 M_2 (M_1 + M_2)}{a^4} \tag{9.63}$$

is the rate at which the orbit shrinks due to gravitational radiation.

A cursory glance at (9.36) and (9.60) will confirm that, for fixed component masses M_1 and M_2, the scalar cloud's growth rate is largest when both black holes are spherical (i.e., nonspinning). In this limit, the ratio $\Gamma_u/\Gamma_{\rm GW}$ can be expressed as a function of three dimensionless quantities: the binary's orbital velocity v, the symmetric mass ratio ν ($= M_1M_2/M^2$), and the ratio μ/Ω. An explicit formula is provided in Appendix 9.D, where I also argue that the precise value of ν has little effect on the conclusions below. For this reason, we will consider only equal-mass binaries in what follows.

The value of $\Gamma_u/\Gamma_{\rm GW}$ as a function of μ/Ω is shown in Figure 9.2 for an equal-mass binary composed of spherical black holes. A value of $v = 0.1$ has been chosen for the orbital velocity, which is large enough that it is at the limit of validity of the post-Newtonian expansion. As the curves in Figure 9.2 would all move downwards for smaller values of v, they represent the largest possible rates that we can calculate reliably with this EFT. For small values of μ, the $u = (2, 1, 1)$ mode has the largest growth rate, which reaches a maximum of $\Gamma_{211}/\Gamma_{\rm GW} \sim 2 \times 10^{-22}$ when $\mu \simeq 9\Omega/10$. Above the threshold at $\mu \simeq \Omega$ (note $E_n \simeq \mu$), it turns into a decaying mode and leaves the $(3, 2, 2)$ mode to take over as

[3] At leading order, a number of results from the literature on the superradiant instability of a single Kerr black hole can be adapted to the study of orbital superradiance because, as I mentioned earlier, the mathematics describing the bound states of a Klein–Gordon field in the long-wavelength limit is identical in both cases.

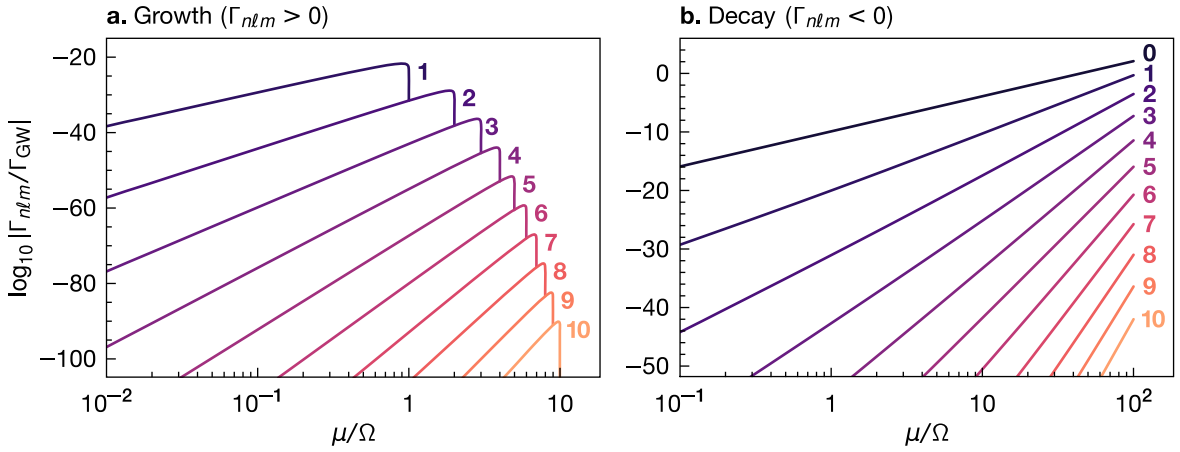


Figure 9.2 The rate $\Gamma_u \equiv \Gamma_{n\ell m}$ at which different quasibound states grow or decay around a binary black hole. On the horizontal axis is the scalar field's mass μ in units of the binary's orbital frequency Ω, while the rates on the vertical axis are normalised in units of $\Gamma_{\rm GW}$, which is the rate at which the orbit shrinks due to gravitational radiation. The binary itself is taken to be composed of spherical black holes of equal mass travelling with orbital velocity $v = 0.1$. The growth rates of the $(n, \ell, m) = (\ell + 1, \ell, \ell)$ modes are shown for $\ell \in [1, 10]$ in the left panel (a), whereas the decay rates for the $(\ell + 1, \ell, -\ell)$ modes are shown for $\ell \in [0, 10]$ in the right panel (b). Note that for a given angular momentum mode (ℓ, m), a larger value of n would result in a lower rate $\Gamma_{n\ell m}$.

the fastest-growing mode, until it too becomes a decaying mode at the next threshold, $\mu \simeq 2\Omega$. This pattern continues for increasing values of μ, with $(\ell + 1, \ell, \ell)$ being the fastest-growing mode when $(\ell - 1) \lesssim \mu/\Omega \lesssim \ell$. The overall trend in Figure 9.2(a) clearly shows that the maximum value that Γ_u can attain decreases rapidly as μ increases. Although the EFT breaks down as we approach $\mu/\Omega \sim v^{-2}$, this trend strongly suggests that orbital superradiance is always grossly inefficient. Consequently, the exponential growth of a long-wavelength scalar cloud is unlikely to leave any measurable impact on the evolution of a binary black hole.

In contrast, the decay rates can become much larger as μ increases [see Figure 9.2(b)], but their observational viability rests on $M_{\hat{u}}$ being comparable to $E_{\rm orb}$. The fact that the growth rates are so small implies that clouds with such high densities are unlikely to have formed dynamically around binary black holes that start off in pure vacuum. With that said, there may be a possibility that other processes could generate these clouds, particularly in alternative theories of gravity wherein the scalar field is nonminimally coupled to matter. For instance, does the core collapse of a massive star into a black hole remnant leave behind an appreciable scalar cloud? If so, could successive supernova events in a stellar binary lead to a pair of black holes enveloped by a common scalar cloud (assuming an optimal value for μ)? It has been shown that a large amount of scalar radiation can be produced during core collapse in a certain class of scalar–tensor theories [519–522], although current numerical methods are unable to determine if a scalar cloud can develop around a black hole remnant [523]. Exploring these open questions presents an exciting opportunity for future work.

9.4 Outgoing radiation

The periodic forcing that a binary exerts onto a surrounding cloud inevitably leads to a fraction of the scalar field being upscattered and ejected as outgoing radiation. Additionally, ingoing radiation can scatter off this binary and undergo amplification when given the right initial conditions. In our perturbative approach, both of these phenomena are encoded in the principal value of the ω integral in (9.45). Using (9.19) to replace the Whittaker function with the radial solution $R^{+}_{\ell}(k,r)$, the result is

$$\phi^{(1)}_{c}(x) = \sum_{\ell,m} \int_{\omega\in\mathbb{R}\setminus\{\pm E_n\}} \frac{\mathrm{d}\omega}{2\pi}\, \mathcal{A}^{(1)}_{w}\phi^{+}_{k\ell m}(\mathbf{x})e^{-i\omega t}, \tag{9.64}$$

where the first-order correction to the outgoing amplitude for a given mode $w \equiv (\omega,\ell,m)$ is

$$\mathcal{A}^{(1)}_{w} = s_{\ell}(\zeta)\, k^{\ell+1} \int \mathrm{d}t\; O^{(0)}_{\ell m}(t) e^{i\omega t}, \tag{9.65a}$$

with

$$s_{\ell}(\zeta) = \frac{\Gamma(\ell+1+i\zeta)}{\Gamma(\ell+1)} e^{-\pi\zeta/2 - i\sigma_{\ell}(\zeta)}. \tag{9.65b}$$

It is worth highlighting that the solution in (9.64) is not of the form $\phi = \phi^{>} + \text{c.c.}$; hence, $\mathcal{A}^{(1)}_{w}$ implicitly accounts for both positive- and negative-frequency modes. Because the overall solution must be real, these coefficients will have to satisfy a constraint analogous to (9.37). Combining (9.A.17a) with the freedom to relabel $\omega \to -\omega$ and $m \to -m$ as they are being integrated and summed over, respectively, we arrive at the constraint

$$\mathcal{A}^{(1)}_{\omega\ell m} = e^{-\pi\zeta}(-1)^{\ell+m}\mathcal{A}^{(1)*}_{-\omega\ell-m}. \tag{9.66}$$

One can now verify that the solution in (9.65a) adheres to this constraint after using (9.37) along with the identities $[\Gamma(z)]^{*} \equiv \Gamma(z^{*})$ and $k^{*}(\omega) \equiv -k(-\omega)$; the latter following directly from the definition in (9.16).

To compute the rate at which energy is carried away from this system in the form of scalar waves, it is useful to first recast the zeroth-order solution in (9.27) into a form that is similar to (9.64). The identity in (9.A.17b) can be used to rewrite the sum over continuum states as

$$\phi^{(0)}_{c}(x) = \sum_{w} 2\mathcal{I}_{w}\phi_{k\ell m}(\mathbf{x})e^{-i\omega t}, \tag{9.67}$$

where

$$\mathcal{I}_{w} = \mathcal{I}^{>}_{\omega\ell m} + e^{\pi\zeta}(-1)^{\ell+m}\mathcal{I}^{>*}_{-\omega\ell-m} \tag{9.68}$$

implicitly accounts for the ingoing amplitudes of both the positive- and negative-frequency modes, and

can be seen to satisfy the requisite constraint

$$\mathcal{I}_{\omega\ell m} = e^{\pi\zeta}(-1)^{\ell+m}\mathcal{I}^{*}_{-\omega\ell-m}. \tag{9.69}$$

By combining this with (9.64), we see that the full solution $\phi = \phi^{(0)} + \phi^{(1)} + \cdots$ for the continuum states has the form

$$\phi_c(x) = \sum_w \left[\mathcal{I}_w \phi^{-}_{k\ell m}(\mathbf{x})e^{-i\omega t} + \mathcal{R}_w \phi^{+}_{k\ell m}(\mathbf{x})e^{-i\omega t}\right], \tag{9.70}$$

where $\mathcal{R}_w := \mathcal{I}_w + \mathcal{A}_w$ is the total outgoing amplitude, with $\mathcal{A}_w$ given to first order in (9.65a).

Now integrating the (t, r) component of this scalar's energy–momentum tensor over a spherical shell of radius r and taking the limit $r \to \infty$, one finds that the power lost to scalar waves can be written as the difference between the energy fluxes flowing into and out of the system, $\dot{E}_{\rm SW} = \dot{E}^{\rm out}_{\rm SW} - \dot{E}^{\rm in}_{\rm SW}$. As shown in Appendix 9.C, these quantities have particularly simple expressions when integrated over all time:

$$\int_{-\infty}^{\infty} \mathrm{d}t\ \dot{E}^{\rm out}_{\rm SW} = \sum_w \frac{\omega}{k}\theta(k^2)|\mathcal{R}_w|^2, \tag{9.71a}$$

$$\int_{-\infty}^{\infty} \mathrm{d}t\ \dot{E}^{\rm in}_{\rm SW} = \sum_w \frac{\omega}{k}\theta(k^2)|\mathcal{I}_w|^2. \tag{9.71b}$$

9.4.1 Ejection of bound states

To better understand the physical implications of (9.64), let us start by supposing — as we did in the previous section — that there is no ingoing radiation and only a single bound state $\hat{u}$ is populated at zeroth order. In this case, the flux $\dot{E}_{\rm SW}$ represents the energy in the cloud that is being ejected out of the system as scalar waves. After dividing (9.71a) by the total time $\int \mathrm{d}t \equiv 2\pi\delta(0)$ and setting $\mathcal{I}_w = 0$, we obtain the expression

$$\langle \dot{E}_{\rm SW}\rangle = \frac{1}{2\pi\delta(0)}\sum_w \frac{\omega}{k}\theta(k^2)\left|\mathcal{A}^{(1)}_w\right|^2 \tag{9.72}$$

for the time-averaged scalar flux. We will work exclusively with time-averaged fluxes in what follows, so let us drop the angled brackets henceforth to declutter the notation.

When only a single bound state $\hat{u}$ is populated, a close inspection of (9.39) reveals that mode mixing generates continuum states with frequencies in a discrete set given by $\omega_m = |E_{\hat{n}} + (m - \hat{m})\Omega|$. As we argued at the end of Section 9.3.2, these continuum states are generated alongside newly populated quasibound states $u \neq \hat{u}$, which oscillate at the same set of frequencies $|E_n - \Delta_{u\hat{u}}| \equiv |E_{\hat{n}} + (m - \hat{m})\Omega|$. The subset of these continuum states with $\omega_m > \mu$ are radiation modes that propagate to infinity.

Substituting (9.39) into (9.72) shows that the power carried away in these radiation modes is

$$\dot{E}_{\rm SW} = \frac{1}{2\mu}\sum_{\ell,m}\left|\frac{Y^*_{\ell m}(\mathbf{d})Y_{\hat{\ell}\hat{m}}(\mathbf{d})}{(2\ell+1)!!}B_{\ell\hat{\ell}}\mathfrak{R}_{\hat{n}\hat{\ell}}(\hat{m}\Omega - E_{\hat{n}})c^{(0)}_{\hat{u}}\right|^2 \times \theta(k_m^2)[S_\ell(\zeta_m) + S_\ell(-\zeta_m)]\omega_m k_m^{2\ell+1}, \tag{9.73}$$

having written $S_\ell(\zeta) := |s_\ell(\zeta)|^2$, $k_m \equiv k(\omega_m)$, and $\zeta_m \equiv \zeta(\omega_m)$ for short. (The intermediate steps for this calculation are presented in Appendix 9.C.) Note that although this result is a sum over infinitely many modes, it suffices to keep only the lowest few values of ℓ to obtain a good approximation, as the prefactor $|Y^*_{\ell m}(\mathbf{d})/(2\ell+1)!!|^2$ that is common to every term decays rapidly with ℓ like $\sim 1/(2\ell+1)!$. In other words, the contributions to this sum from the high-ℓ modes are factorially suppressed.

With that said, (9.73) is still not particularly illuminating. It is perhaps most instructive to compare it to (9.61), in which case one finds

$$\frac{\dot{E}_{\rm SW}}{\dot{E}_b} = \sum_{\ell,m}\left|\frac{Y_{\ell m}(\mathbf{d})}{(2\ell+1)!!}\right|^2 \frac{E_{\hat{n}}B^2_{\ell\hat{\ell}}}{2\mu^2 B_{\hat{\ell}\hat{\ell}}}(\hat{m}\Omega - E_{\hat{n}})\,\theta(k_m^2)[S_\ell(\zeta_m) + S_\ell(-\zeta_m)]\omega_m k_m^{2\ell+1}. \tag{9.74}$$

To proceed, consider first the limiting behaviour of a given term in (9.74) for which $k_m^2 > 0$ and $\zeta_m \gg 1$. Standard identities for the gamma function can be used to show that

$$S_\ell(\zeta) := |s_\ell(\zeta)|^2 = \frac{1}{(\ell!)^2}\frac{\pi\zeta e^{-\pi\zeta}}{\sinh(\pi\zeta)}\prod_{j=1}^{\ell}(j^2+\zeta^2), \tag{9.75}$$

which one might recognise as the Sommerfeld enhancement factor [524–526]. The sum of these factors appearing in (9.74) can thus be seen to have the asymptotic form $S_\ell(\zeta) + S_\ell(-\zeta) \sim 2\pi|\zeta|^{2\ell+1}$ when $\zeta \gg 1$. On a mathematical level, this Sommerfeld factor arises naturally in our calculations because the solution to $O^L(t)$ involves derivatives of Coulomb functions evaluated at the origin. We run into difficulties, however, if we attempt to assign to this factor its usual physical interpretation. In the present context, there is no analogous process that occurs on flat space, since bound states cannot form in the absence of the binary's gravitational potential. Accordingly, the sum $S_\ell(\zeta) + S_\ell(-\zeta)$ in (9.74) is perhaps best regarded as simply an inevitable part of the result, rather than engendering some kind of enhancement in the power carried away by low-momentum modes (recall $\zeta \propto 1/k$).

The interpretation of $S_\ell(\zeta)$ notwithstanding, the corresponding term in (9.74) can be seen to reduce to

$$\left|\frac{Y_{\ell m}(\mathbf{d})}{(2\ell+1)!!}\right|^2 \frac{E_{\hat{n}}B^2_{\ell\hat{\ell}}}{2\mu B_{\hat{\ell}\hat{\ell}}}\,2\pi(GM\mu^2)^{2\ell+1}(\hat{m}\Omega - E_{\hat{n}}) \tag{9.76}$$

in the low-momentum limit, $\zeta_m \gg 1$. Power counting now reveals that this term scales with the EFT's expansion parameters as $v^5(a/\lambda_{\rm dB})^{2\ell+1}(\hat{m}-\mu/\Omega)$; hence, for clouds with $\hat{m}\sim\mathcal{O}(1)$, and assuming $\mu/\Omega \ll v^{-2}$ [cf. (9.12)], the rate at which energy is carried away by low-momentum radiation is always parametrically suppressed relative to $\dot{E}_b$. Indeed, each term in (9.74) is a monotonically increasing function of k_m, so most of the energy is carried away in high-momentum modes, for which $\zeta_m \ll 1$.

As $S_\ell(\zeta_m)\sim 1$ in this opposite limit, the corresponding term in (9.74) reduces to

$$\left|\frac{Y_{\ell m}(\mathbf{d})}{(2\ell+1)!!}\right|^2 \frac{E_{\hat{n}}B^2_{\ell\hat{\ell}}}{\mu^2 B_{\hat{\ell}\hat{\ell}}}\omega_m k_m^{2\ell+1}(\hat{m}\Omega - E_{\hat{n}}). \tag{9.77}$$

To assess the typical size of this term, it is convenient to express it in terms of the ratio $f_\mu := \mu/\Omega$. Also using the definition $\omega_m = |E_{\hat{n}} + (m-\hat{m})\Omega|$ and approximating $E_{\hat{n}} \simeq \mu$, (9.77) becomes

$$\frac{1}{f_\mu}\left|\frac{Y_{\ell m}(\mathbf{d})}{(2\ell+1)!!}\right|^2 \frac{B^2_{\ell\hat{\ell}}}{B_{\hat{\ell}\hat{\ell}}}\,\Omega^{2\ell+2}(\hat{m}-f_\mu)(f_\mu+\Delta m)(2f_\mu\Delta m+\Delta m^2)^{\ell+1/2}, \tag{9.78}$$

where $\Delta m = m-\hat{m}$. Notice that this term scales as $v^{2\ell+6}$ when f_μ, $\hat{m}$, and Δm are all of order unity; meaning that the rate at which energy is carried away by high-momentum radiation is, in this case, also parametrically suppressed relative to $\dot{E}_b$. However, when $f_\mu \ll 1$, the $1/f_\mu$ prefactor enhances this term such that power loss to radiation can become significant in comparison to $\dot{E}_b$. More precisely, a given high-momentum mode will extract energy from the cloud at a rate greater than $\dot{E}_b$ if[4]

$$\mu/\Omega \lesssim v^{2\ell+6} \leq v^6. \tag{9.79}$$

One may conclude from this simple scaling analysis that a scalar cloud cannot form dynamically around a binary black hole when the scalar field's mass μ is sufficiently light, as the rate at which the cloud is depleted via scalar radiation is greater than the rate at which it grows due to orbital superradiance. As a rough guide, this occurs when $\mu \lesssim 10^{-19}\ \text{eV}\,(v/0.1)^9(M_\odot/M)$.

Gravitational waves It is worth briefly remarking that a scalar cloud will also emit gravitational waves due to the oscillatory nature of its backreaction onto the spacetime. When only the single bound state $\hat{u}$ is populated (and further assuming that $\hat{\ell}=\hat{m}$ for simplicity), the flux of gravitational waves

[4]This is a conservative upper bound that does not take the possibility that $B^2_{\ell\hat{\ell}}$ can vanish into consideration. As an example, if an equal-mass binary is surrounded by a cloud comprised of only the $\hat{u}=(2,1,1)$ mode, one finds $B_{01}=0$; hence, energy is predominantly radiated away in the $\ell=1$ modes. In this case, $\dot{E}_{\rm SW}$ becomes larger than $\dot{E}_b$ only when $\mu/\Omega \lesssim v^8$.

emitted by the cloud is [339, 357]

$$\dot{E}_{\rm GW} \simeq \frac{\mathcal{C}_{\hat{u}}}{G}\left(\frac{M_{\hat{u}}}{M}\right)^2 (GM\mu)^{4\hat{\ell}+10}, \tag{9.80}$$

where $\mathcal{C}_{\hat{u}}$ (<1) is some dimensionless prefactor whose exact form will not be important to us, but note that $\mathcal{C}_{n00} = 0$. Comparing this with (9.61), we find $\dot{E}_{\rm GW}/\dot{E}_b \sim (M_{\hat{u}}/M) f_\mu^4 v^{4\hat{\ell}+9}$. Given that we should expect $M_{\hat{u}} < M$ if the scalar is to behave like a test field around the binary, we see that $\dot{E}_{\rm GW}$ is generally much smaller than $\dot{E}_b$. Moreover, unless $\hat{\ell} = 1$ and f_μ is close to the UV cutoff of this EFT [cf. (9.12)], a comparison with (9.78) reveals that $\dot{E}_{\rm GW}$ is also typically smaller than $\dot{E}_{\rm SW}$.

9.4.2 Superradiant scattering

Let us now turn our attention to a different setup in which the zeroth-order solution is given by a steady stream of radiation. In realistic scenarios, we would expect an incident wave to be essentially planar on the scales of the binary, but such a configuration turns out to be difficult to analyse in the present context. To elaborate, because a plane wave can always be written as a linear combination of spherical waves, this zeroth-order solution will contain infinitely many modes that subsequently mix into one another as a result of their interactions with the binary.

To render the following discussion more tractable, we shall consider a simpler, albeit more artificial setup that should nevertheless suffice to illustrate the most salient features of orbital superradiance. Specifically, we will consider a steady stream of ingoing radiation peaked at the single mode $\hat{w} \equiv (\hat{\omega}, \hat{\ell}, \hat{m})$. This corresponds to making the choice

$$\mathcal{I}^{>}_{\omega\ell m} = \Phi_{\hat{w}} 2\pi\delta(\omega - \hat{\omega})\delta_{\ell\hat{\ell}}\,\delta_{m\hat{m}}, \tag{9.81}$$

where $\Phi_{\hat{w}}$ is in general some complex-valued coefficient with dimensions of energy.

Amplification factor From (9.39), we learn that a single ingoing mode $\hat{w}$ will scatter into multiple outgoing modes with frequencies in a discrete set given by $\omega_m = |\hat{\omega} + (m - \hat{m})\Omega|$. Included in this spectrum is the original, or "primary," mode with frequency $\hat{\omega} \equiv \omega_{\hat{m}}$, which typically comprises a majority of the outgoing energy flux as a result of its interference with the ingoing radiation. To see this, we expand (9.71) in powers of $\mathcal{A}_w$ to find

$$\int \mathrm{d}t\ \dot{E}_{\rm SW} = \sum_w \frac{\omega}{k}\theta(k^2) 2\,\mathrm{Re}\,\mathcal{I}^*_w \mathcal{A}^{(1)}_w + \mathcal{O}(\mathcal{A}^2). \tag{9.82}$$

Naturally, only the $\hat{w}$ mode contributes to the sum at linear order in $\mathcal{A}$, since $\mathcal{I}^{>}_w = 0\ \forall\ w \neq \hat{w}$.

Let us begin by focusing on this linear term. Substituting (9.39) into the above formula and making use of the symmetries in (9.66) and (9.69), the rate at which the scalar gains energy as it scatters off the binary black hole is given by

$$\dot{E}^{(1)}_{\rm SW} = 2\,{\rm Re}\,\frac{\left|2Y_{\hat{\ell}\hat{m}}(\mathbf{d})\right|^2}{(2\hat{\ell}+1)!!}B_{\hat{\ell}\hat{\ell}}|\Phi_{\hat{w}}|^2 s_{\hat{\ell}}(\hat{\zeta})C_{\hat{\ell}}(\hat{\zeta})\hat{\omega}\hat{k}^{2\hat{\ell}}(\hat{m}\Omega-\hat{\omega}) \tag{9.83}$$

to first order in $\mathcal{A}$. (The details of this calculation are presented in Appendix 9.C.) As the Gamow factor $C_\ell(\zeta) \equiv s^*_\ell(\zeta)/(2\ell+1)!!$ when $\zeta \in \mathbb{R}$, the above expression is already real and can be further simplified to read

$$\dot{E}^{(1)}_{\rm SW} = \left|\frac{2Y_{\hat{\ell}\hat{m}}(\mathbf{d})}{(2\hat{\ell}+1)!!}\right|^2 2B_{\hat{\ell}\hat{\ell}}|\Phi_{\hat{w}}|^2 S_{\hat{\ell}}(\hat{\zeta})\hat{\omega}\hat{k}^{2\hat{\ell}}(\hat{m}\Omega-\hat{\omega}). \tag{9.84}$$

To gain a sense for how much energy is exchanged during this process, it is instructive to compare (9.84) to the total flux of ingoing radiation $\dot{E}^{\rm in}_{\rm SW}$, an expression for which is also derived in Appendix 9.C [cf. (9.C.18)]. This dimensionless ratio defines for us the amplification factor $Z := \dot{E}^{(1)}_{\rm SW}/\dot{E}^{\rm in}_{\rm SW}$ to first order in the interactions. In the case of a single ingoing mode $\hat{w}$, it reads

$$Z = \left(\frac{2S_{\hat{\ell}}(\hat{\zeta})}{1+e^{-2\pi\hat{\zeta}}}\right)\left|\frac{2Y_{\hat{\ell}\hat{m}}(\mathbf{d})}{(2\ell+1)!!}\right|^2 B_{\hat{\ell}\hat{\ell}}\,\hat{k}^{2\hat{\ell}+1}(\hat{m}\Omega-\hat{\omega}). \tag{9.85}$$

If this mode is such that $\hat{k} \gg GM\mu^2$ (i.e., $\hat{\zeta} \ll 1$), the Sommerfeld factor $S_{\hat{\ell}}(\hat{\zeta})$ and the exponential $e^{-2\pi\hat{\zeta}}$ both reduce to unity. This limiting behaviour indicates that the binary's long-range gravitational potential has negligible influence on the amplification or absorption of high-momentum modes, as one might naturally expect. Taking this limit also facilitates a comparison between (9.85) and known results in the literature for the amplification of a long-wavelength scalar field incident on a single rotating black hole [317, 329, 468]. As was the case for the growth rates in Section 9.3, one finds that the expressions for Z are identical in these two cases, up to a geometric factor and appropriate identifications of M and Ω.

For more general values of $\hat{k}$, (9.75) may be used to show that

$$\frac{2S_\ell(\zeta)}{1+e^{-2\pi\zeta}} = \frac{1}{(\ell!)^2}\frac{2\pi\zeta}{\sinh(2\pi\zeta)}\prod_{j=1}^{\ell}(j^2+\zeta^2), \tag{9.86}$$

which has the asymptotic form $\sim|\zeta|^{2\ell+1}e^{-2\pi|\zeta|}$ when $\zeta \to \infty$. Interestingly, while we might naively expect that the appearance of $S_\ell(\zeta)$ in (9.84) should lead to a Sommerfeld enhancement of low-momentum modes, the result in (9.86) reveals that it actually engenders an exponential suppression.

We interpret this as follows: In the classical analogue of Sommerfeld enhancement, we imagine a stream of particles impinging on a star of radius $R_\star$ [526]. In the absence of gravity, the geometric cross section $\sigma_0 = \pi R_\star^2$ of this star provides a measure of the fraction of particles that collide into it and are subsequently absorbed. However, the actual cross section σ for this interaction can be much larger, especially for particles with low momenta, because the star's attractive gravitational potential is able to pull in particles that have impact parameters greater than $R_\star$.

With this picture in mind, one should now expect no analogous enhancement to occur in the present scenario. In our setup, the ingoing mode $\hat{w}$ is a spherical wave that is already directed straight at the origin; hence, the presence of the binary's gravitational potential does nothing to affect the amount of radiation that reaches it. While this argues for the lack of Sommerfeld enhancement, it remains to explain the suppression of low-momentum modes observed in (9.86). Admittedly, the physical origin of this suppression is still not fully understood, although the most likely explanation is that it is due to the conversion of radiation modes into bound states [cf. the second line in (9.55)], which is enhanced at low momenta. A full quantitative analysis is still needed to validate this interpretation, although such a task is beyond the scope of this thesis.

Putting this conceptual issue aside, let us discuss the likelihood of observing this energy exchange between the binary and the scalar. As I mentioned earlier in Section 9.3, the particular geometry of the binary prevents it from interacting with any long-wavelength mode whose angular momentum is such that $\ell + m \notin 2\mathbb{Z}$. For the remaining modes, amplification occurs if $0 < \hat{\omega} < \hat{m}\Omega$, in which case the binary loses energy and inspirals more rapidly as a result. Otherwise, there is a net absorption of the scalar by the binary, which then gains energy and experiences a slowing down, or even a possible reversal, of its inspiral. The feasibility of observing either of these effects depends on the magnitude of $\dot{E}_{\rm SW}$ when compared to the outgoing flux of gravitational radiation $\mathcal{F}_{\rm GW}$. As a rough estimate, we should expect to observe the influence of this energy exchange on the orbital motion only if the ratio $\dot{E}^{(1)}_{\rm SW}/\mathcal{F}_{\rm GW} \equiv Z\,\dot{E}^{\rm in}_{\rm SW}/\mathcal{F}_{\rm GW}$ is not too much smaller than unity.

In Figure 9.3, the amplification factor Z is shown as a function of the ingoing frequency for different values of the scalar field's mass. As in Section 9.3.3, the binary is assumed to be composed of spherical black holes of equal mass travelling with orbital velocity $v = 0.1$. The same reasoning as before justifies limiting ourselves to this specific case: First, the curves in Figure 9.3 would all move downwards for smaller values of v, so once again they represent the largest possible values that can be reliably calculated using this EFT. Moreover, the precise value of the binary's mass ratio has little effect on our overall conclusions, as Appendix 9.D argues.

In Figure 9.3(a), we see that amplification is most pronounced for the $\ell = m = 1$ mode, which

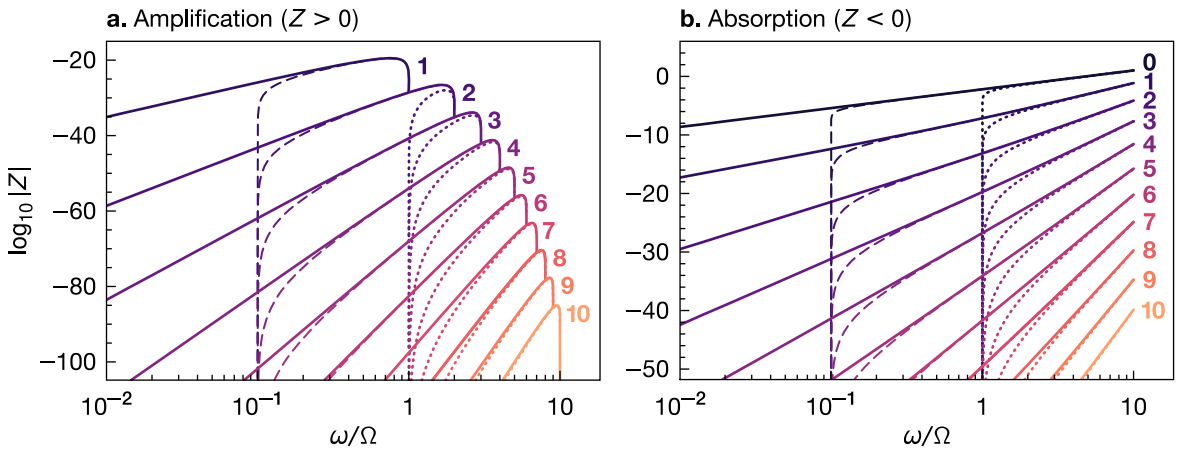


Figure 9.3 The amplification factor Z for a single ingoing radiation mode as a function of its frequency ω in units of binary's orbital frequency Ω. The binary itself is taken to be composed of spherical black holes of equal mass travelling with orbital velocity $v = 0.1$. Values of Z for the $\ell = m$ modes are shown for $\ell \in [1, 10]$ in the left panel (a), while the right panel (b) shows the corresponding values for the $\ell = -m$ modes in the range $\ell \in [0, 10]$. In both panels, the amplification factors are shown for three illustrative values of the scalar field's mass: $\mu = 0$ (solid lines), $\mu = 0.1\Omega$ (dashed lines), and $\mu = \Omega$ (dotted lines).

reaches a maximum value $Z \sim 4 \times 10^{-10}$ when $\omega = 3\Omega/4$. As the higher $\ell = m$ modes are less efficiently amplified, the overall trend suggests that this orbital superradiant mechanism continues to become increasingly insignificant even for large frequencies $\omega \gtrsim \Omega/v$ beyond the EFT's regime of validity. Given the smallness of Z and the unlikelihood that the ingoing flux $\dot{E}^{\rm in}_{\rm SW}$ of scalar waves would match or exceed the outgoing gravitational-wave flux $\mathcal{F}_{\rm GW}$ in realistic astrophysical scenarios, we deduce that the amplification of long-wavelength scalar fields is observationally inaccessible. Granted, this conclusion is based on a rather artificial setup, although it seems unlikely to change were we to consider the more realistic case of plane waves.

In contrast, Figure 9.3(b) demonstrates that absorption continues to become more efficient as ω increases, naturally prompting us to ask: Is there a regime (possibly outside the EFT's regime of validity) in which $|Z|$ and $\dot{E}^{\rm in}_{\rm SW}$ are both large enough that they can leave a measurable imprint on the evolution of the binary? On a more theoretical level, it is also interesting to ask: What is the maximum amount of radiation that can be absorbed by a binary black hole in a given amount of time? Because the amplification factor is bounded from below ($Z \geq -1$), there are two possibilities for what might occur in the high-frequency regime. Either Z gradually tends to a minimum value (meaning absorption would be most pronounced at high frequencies), or it has a turning point (i.e., there is a critical frequency beyond which absorption becomes less efficient again). The fact that moving black holes can amplify high-frequency radiation [527] via what is essentially the slingshot effect is a hint that the latter may be more likely. These questions point to potential directions for future work.

Secondary modes To complete our discussion on the scattering of scalar waves, we ought to discuss the additional energy carried away by the secondary modes $w \neq \hat{w}$, which are generated by mode mixing. Although their contribution to the energy flux is typically subleading because they first appear in $\dot{E}_{\rm SW}$ at $\mathcal{O}(\mathcal{A}^2)$, the energy they carry can exceed that of the primary mode if $\hat{\omega}$ is sufficiently small, as we will now show.

Let us denote this second-order contribution to the energy flux by $\dot{E}^{(2)}_{\rm SW}$. The calculation is almost identical to that in Section 9.4.1, and the end result is found to be

$$\dot{E}^{(2)}_{\rm SW} = \frac{1}{2\pi\delta(0)} \sum_w \frac{\omega}{k} \theta(k^2) \left|\mathcal{A}^{(1)}_w\right|^2 \tag{9.87}$$

$$= \sum_{\ell,m} \left| \frac{2\Phi_{\hat{w}} Y^*_{\ell m}(\mathbf{d}) Y_{\hat{\ell}\hat{m}}(\mathbf{d})}{(2\ell+1)!!(2\hat{\ell}+1)!!} \right|^2 B^2_{\ell\hat{\ell}} S_{\hat{\ell}}(\hat{\xi}) \hat{k}^{2\hat{\ell}} (\hat{m}\Omega - \hat{\omega})^2$$

$$\times \theta(k_m^2) [S_\ell(\zeta_m) + S_\ell(-\zeta_m)] \omega_m k_m^{2\ell+1}, \tag{9.88}$$

where the frequencies of the modes being summed over are given by $\omega_m = |\hat{\omega} + (m - \hat{m})\Omega|$. To see that this second-order contribution can be much larger than the outgoing flux at $\mathcal{O}(\mathcal{A})$ in (9.84), we simply divide one by the other to find

$$\frac{\dot{E}^{(2)}_{\rm SW}}{\dot{E}^{(1)}_{\rm SW}} = \sum_{\ell,m} \frac{1}{\hat{\omega}} \left| \frac{Y_{\ell m}(\mathbf{d})}{(2\ell+1)!!} \right|^2 \frac{B^2_{\ell\hat{\ell}}}{B_{\hat{\ell}\hat{\ell}}} (\hat{m}\Omega - \hat{\omega}) \theta(k_m^2) [S_\ell(\zeta_m) + S_\ell(-\zeta_m)] \omega_m k_m^{2\ell+1}. \tag{9.89}$$

Observe that (9.89) has the same mathematical structure as (9.74); hence, the analysis proceeds in a largely similar fashion. First, recall that while sums of this kind are to be taken over infinitely many modes, a good approximation can be obtained by keeping only the lowest few values of ℓ, since the higher multipoles are factorially suppressed. Next, the fact that each term in (9.89) is a monotonically increasing function of k_m implies that most of the energy carried away will be in the form of high-momentum modes. We shall therefore concentrate on a given term in (9.89) and suppose that $\zeta_m \ll 1$. In terms of the dimensionless ratio $f_{\hat{\omega}} := \hat{\omega}/\Omega$, this term reads

$$\frac{1}{f_{\hat{\omega}}} \left| \frac{Y_{\ell m}(\mathbf{d})}{(2\ell+1)!!} \right|^2 \frac{B^2_{\ell\hat{\ell}}}{B_{\hat{\ell}\hat{\ell}}} \Omega^{2\ell+2} (\hat{m} - f_{\hat{\omega}})(\Delta m + f_{\hat{\omega}}) [(\Delta m + f_{\hat{\omega}})^2 - (\mu/\Omega)^2]^{\ell+1/2}, \tag{9.90}$$

where $\Delta m = m - \hat{m}$. When $f_{\hat{\omega}}$, $\hat{m}$, and Δm are all of order unity, this term scales as $v^{2\ell+6}$ and is thus parametrically suppressed. However, if instead $f_{\hat{\omega}} \ll 1$ (and necessarily $\mu/\Omega < f_{\hat{\omega}}$ if $\hat{w}$ is to be a radiation mode), this term can become arbitrarily large. As a result, the energy carried away in a

secondary mode of frequency ω_m will dominate over that of the primary mode $\hat{w}$ when

$$\hat{\omega}/\Omega \lesssim v^{2\ell+6} \leq v^6. \tag{9.91}$$

This phenomenon is particularly interesting if $\hat{w}$ is a counter-rotating mode satisfying $\hat{m}\Omega - \hat{\omega} < 0$, since in this case we would predict a total amplification factor $Z = \dot{E}_{\rm SW}/\dot{E}^{\rm in}_{\rm SW} < 0$ when truncating to $\mathcal{O}(\mathcal{A})$. However, if this ingoing mode has $\hat{\omega} \lesssim v^6\Omega$, the outgoing energy flux is dominated by the $\mathcal{O}(\mathcal{A}^2)$ term, which is positive definite; cf. (9.87). Thus, we learn that the energy in a scalar wave is always amplified when it scatters off a binary black hole if the ingoing frequency is low enough. To be clear, while $\dot{E}^{(2)}_{\rm SW}$ can be very large relative to $\dot{E}^{(1)}_{\rm SW}$, its magnitude is still small in comparison to $\dot{E}^{\rm in}_{\rm SW}$ and further decreases as $\hat{\omega} \to 0$, meaning the actual amount of energy that a low-frequency, counter-rotating mode extracts from the binary is always negligible. Nonetheless, this calculation illustrates the kinds of rich physics that can arise as a consequence of mode mixing.

9.5 Thermodynamics and the fifth force

To conclude this chapter, let us say a few more words about the underlying mechanism that drives this novel variant of superradiance. We have already discussed that the absence of time-translation invariance and rotational symmetry is at the heart of many of the unique features that distinguish orbital superradiance from its more familiar cousins; namely, the appearance of mode mixing, beating patterns, and the upscattering of bound states into scalar waves. The kinds of binary spacetimes we have been considering are not completely generic, however, as we have restricted our attention to circular orbits and, moreover, have neglected the orbital decay due to gravitational radiation. The latter is a valid approximation at first order in perturbation theory,[5] and so the goal is not to disparage it here. Instead, what is interesting to point out is that these simplifying assumptions lead to a residual helical symmetry, which we can exploit to glean further insight into the onset of superradiance.

At low post-Newtonian orders, the approximate spacetime around a binary black hole evolving in a perennial, circular orbit admits the helical Killing vector $\xi = \partial_t + \Omega\,\partial_\chi$, which loosely speaking generates time-translation symmetry in the frame corotating with the binary. (Recall that χ denotes the azimuthal angle.) In the presence of this Killing vector, the underlying spacetime satisfies what is known as the first law of binary black hole mechanics [529,530]:

$$\mathrm{d}\mathcal{M} - \Omega\,\mathrm{d}\mathcal{J} = \sum_K \frac{\kappa_K}{8\pi G}\mathrm{d}A_K. \tag{9.92}$$

[5] In the same way that it is valid to hold the orbit fixed while using the quadrupole formula to compute the flux of gravitational waves [447,528].

Respectively, $\mathcal{M}$ and $\mathcal{J}$ are the total (Arnowitt–Deser–Misner) mass and angular momentum of this binary, while κ_K is the surface gravity of the Kth black hole.

Now suppose that there is a long-wavelength scalar field living on top of this spacetime. The first law can then be seen to restrict the way in which this ambient scalar interacts with the binary. For concreteness, we will focus on the amplification of scalar waves, although the same argument will follow through for the growth of a scalar cloud as well. When a single ingoing mode $(\hat{\omega}, \hat{\ell}, \hat{m})$ is incident on the binary, the change in its angular momentum $\mathrm{d}L_\phi$ is related to the change in its energy $\mathrm{d}E_\phi$ by the constraint $\mathrm{d}L_\phi/\mathrm{d}E_\phi = \hat{m}/\hat{\omega}$, up to subleading corrections from mode mixing, which we will neglect. Seeing as these changes in the energy and angular momentum must come at the expense of the binary, it follows that $\mathrm{d}\mathcal{J}/\mathrm{d}\mathcal{M} = \mathrm{d}L_\phi/\mathrm{d}E_\phi = \hat{m}/\hat{\omega}$; hence, the lhs of (9.92) equivalently reads $\hat{\omega}^{-1}(\hat{\omega} - \hat{m}\Omega)\,\mathrm{d}\mathcal{M}$. Because the second law of black hole mechanics ($\mathrm{d}A_K \geq 0$) renders the rhs of this equation positive semidefinite, the binary necessarily loses energy (i.e., $\mathrm{d}\mathcal{M} < 0$) whenever $0 < \hat{\omega} < \hat{m}\Omega$ and $\mathrm{d}A_K > 0$. (This kind of thermodynamic argument was first made by Bekenstein in 1973 for the case of single-black-hole superradiance [531].)

Three comments are worth making at this stage. First, the sum over K in (9.92) [or, equivalently, in (9.36)] signifies that only one black hole is needed for superradiance to occur — the other member of the binary need not interact with the scalar at all. Second, nowhere in the above argument did we have to assume that the ambient field was a scalar. Any long-wavelength bosonic field will satisfy $\mathrm{d}\mathcal{J}/\mathrm{d}\mathcal{M} = \hat{m}/\hat{\omega}$ (provided that corrections from mode mixing are subleading and can be neglected); hence, any long-wavelength bosonic field will be susceptible to orbital superradiance. Third, in the case of scalar fields, the first law can be used to do much more than just derive the superradiant condition $0 < \hat{\omega} < \hat{m}\Omega$. Combining it with the results from Chapter 7 gives us an alternative method for computing the growth rate Γ_u and the amplification factor Z.

To see this, we shall again focus on the amplification of scalar waves for simplicity, although the calculation for the growth rate is entirely analogous. Under the assumption that the binary is in a circular orbit, the term $\Omega\,\mathrm{d}\mathcal{J}$ in (9.92) is exactly the work done on the binary by the scalar, $\mathrm{d}W = \sum_K \mathbf{F}_K \cdot \mathrm{d}\mathbf{z}_K$. The force that this scalar exerts onto the Kth black hole is given by (7.63), which reduces to $\mathbf{F}^i_K = Q_K\,\partial^i\phi$ in the nonrelativistic limit, where $Q_K = -A_K\dot{\phi}$ is the black hole's induced scalar charge. Also using the fact that a black hole's absorption cross section is s-wave dominated in the long-wavelength limit, one can show that $\kappa_K\,\mathrm{d}A_K/(8\pi G) = \mathrm{d}M_K = A_K\dot{\phi}^2\,\mathrm{d}t$, where the last equality follows from (7.53). Taken together, these equations imply that the total energy gained by the binary in a time Δt is

$$\Delta\mathcal{M} = \sum_K \int_{-\Delta t/2}^{\Delta t/2} \mathrm{d}t\, \left(\dot{M}_K + \mathbf{F}_K \cdot \dot{\mathbf{z}}_K\right) = \sum_K \int_{-\Delta t/2}^{\Delta t/2} \mathrm{d}t\, A_K\dot{\phi}\left(\dot{\phi} - \dot{\mathbf{z}}^i_K\partial_i\phi\right). \tag{9.93}$$

Note that in each term, the scalar $\phi \equiv \phi(t, \mathbf{z}_K(t))$ is to be evaluated along the worldline $\mathbf{z}_K(t)$.

It will suffice to substitute the zeroth-order solution $\phi^{(0)}$ into (9.93) to obtain the energy change $\Delta\mathcal{M}$ at first order; hence, our starting point is (9.67). After using the identities in (9.19), (9.28), (9.34), and (9.A.3), we arrive at the expression

$$\phi^{(0)}(t, \mathbf{z}_K(t)) = \sum_w 2\mathcal{I}_w C_\ell(\zeta) k^\ell r_K^\ell Y_{\ell m}(\mathbf{d})\, e^{-i(\omega - m\Omega)t} \tag{9.94}$$

for a general sum over long-wavelength radiation modes ($k r_K \ll 1$). Differentiating this expression with respect to time pulls down a factor of $-i(\omega - m\Omega)$ under the sum over w, whereas the derivative $\dot{\mathbf{z}}_K^i \partial_i$ along the azimuthal direction pulls down a factor of $im\Omega$. Also taking the limit $\Delta t \to \infty$, we find that the total energy gained by the binary during this scattering process is

$$\begin{aligned}\Delta\mathcal{M} = \sum_K \int \mathrm{d}t\; A_K \left\{ \sum_w -i\omega\, 2\mathcal{I}_w C_\ell(\zeta) k^\ell r_K^\ell Y_{\ell m}(\mathbf{d})\, e^{-i(\omega - m\Omega)t} \right\} \\ \times \left\{ \sum_{w'} -i(\omega' - m'\Omega)\, 2\mathcal{I}_{w'} C_{\ell'}(\zeta') k'^{\ell'} r_K^{\ell'} Y_{\ell' m'}(\mathbf{d})\, e^{-i(\omega' - m'\Omega)t} \right\}.\end{aligned} \tag{9.95}$$

As $\phi(x)$ is a real scalar, either of the curly brackets above can replaced by their complex conjugate without affecting the result. Finally, a rearranging of terms can be used to show that

$$-\Delta\mathcal{M} = \sum_w \frac{\omega}{k} \theta(k^2) 2\,\mathrm{Re}\, \mathcal{I}_w^* \mathcal{A}_w^{(1)}, \tag{9.96}$$

where

$$\begin{aligned}\mathcal{A}_w^{(1)} = \int \mathrm{d}t \sum_{w'} Y_{\ell m}^*(\mathbf{d}) Y_{\ell' m'}(\mathbf{d}) C_\ell^*(\zeta) C_{\ell'}(\zeta') k^{\ell+1} k'^{\ell'} \\ \times \sum_K A_K r_K^{\ell+\ell'} (m'\Omega - \omega') 2\mathcal{I}_{w'} e^{i[\omega - \omega' - (m-m')\Omega]t}.\end{aligned} \tag{9.97}$$

The amplitude in (9.97) is exactly what we would obtain from (9.65a). Notice also that in (9.96), the "Re" symbol and the step function are not strictly necessary because (9.95) was real to begin with and the coefficients $\mathcal{I}_w$ have vanishing support over the domain $k^2 < 0$ by construction. Nevertheless, they have been included to make it obvious that (9.96) is equivalent to (9.82).

This calculation therefore teaches us three things. First, we have now verified explicitly that the energy gained by a scalar as it scatters off a binary black hole is indeed equal to the energy that the binary loses, $\int \mathrm{d}t\, \dot{E}_{\mathrm{SW}} = -\Delta\mathcal{M}$. An analogous calculation will show that $\int \mathrm{d}t\, \dot{E}_b = -\Delta\mathcal{M}$ in the case of bound states. Second, this result gives us the opportunity to refine our understanding of what the "fifth" force in (7.63) means physically: If it was not clear before, it should be clear now that the rhs

of (7.63) is the *backreaction* force that a long-wavelength scalar exerts onto a black hole's trajectory as it crosses the horizon. Third, this calculation provides an intuitive explanation as to why the amplification factor Z and the growth rate Γ_u are proportional to the areas of the black holes. For fixed components masses M_1 and M_2, one might naively guess that orbital superradiance is least pronounced for spherical black holes due to their larger absorption cross sections, but the opposite is true because the fifth force is also proportional to this cross section.

Appendix 9.A Properties of the radial solutions

This appendix provides a collection of useful identities for the radial solutions to (9.15).

Limiting forms The identities in this first part have all been reproduced or adapted from Ref. [397]. At large distances ($r \to \infty$), the $R^{\pm}_{\ell}$ solutions have the asymptotic forms

$$R^{\pm}_{\ell}(k,r) = \frac{H^{\pm}_{\ell}(\zeta,kr)}{\pm ikr} \sim \frac{e^{\pm i\theta_\ell(\zeta,kr)}}{\pm ikr}[1+\mathcal{O}(r^{-1})], \tag{9.A.1}$$

where

$$\theta_\ell(\zeta,kr) = kr - \zeta\log(2kr) - \frac{\ell\pi}{2} + \sigma_\ell(\zeta). \tag{9.A.2}$$

If instead $r \to 0$, the behaviour of the radial solutions around the origin may be inferred from the limiting forms of the Whittaker functions. The solutions that are regular at the origin are all proportional to

$$M_{-i\zeta,\ell+1/2}(z) = z^{\ell+1}[1+\mathcal{O}(z)], \tag{9.A.3}$$

whereas the irregular solutions are proportional to

$$W_{-i\zeta,\ell+1/2}(z) = \frac{\Gamma(2\ell+1)}{\Gamma(\ell+1+i\zeta)} \times \begin{cases} z^{-\ell}[1+\mathcal{O}(z)] & (\ell \geq 1) \\ 1+\mathcal{O}(z\log z) & (\ell = 0). \end{cases} \tag{9.A.4}$$

It is also useful to know the limiting behaviour of these solutions for small and large values of ζ. For $\zeta \to 0$ with k held fixed (corresponding to a removal of the gravitational potential), one has

$$\lim_{\zeta\to 0} R^{\pm}_{\ell}(k,r) = h^{\pm}_{\ell}(kr), \qquad \lim_{\zeta\to 0} R_{\ell}(k,r) = j_{\ell}(kr), \tag{9.A.5}$$

where $h^{\pm}_{\ell}$ are the spherical Hankel functions while j_ℓ is the spherical Bessel function of the first kind.

Taking the low-momentum limit $k \to 0$ (i.e., $\zeta \to \infty$ with M and μ held fixed) instead, one recovers the usual Bessel functions:

$$\lim_{k\to 0} \frac{(-\zeta)^\ell e^{\pi\zeta/2 - i\sigma_\ell(\zeta)}}{\Gamma(\ell+1-i\zeta)} R_\ell(k,r) = \frac{J_{2\ell+1}(2\sqrt{2GM\mu^2 r})}{\sqrt{2GM\mu^2 r}}, \tag{9.A.6}$$

$$\lim_{k\to 0} \frac{e^{\mp i\sigma_\ell(\zeta)}\Gamma(\ell+1\pm i\zeta)}{\pm 2\pi i(-\zeta)^{\ell+1}e^{\pi\zeta/2}} R^\pm_\ell(k,r) = \frac{Y_{2\ell+1}(2\sqrt{2GM\mu^2 r})}{\sqrt{2GM\mu^2 r}}. \tag{9.A.7}$$

Note that the prefactor multiplying R^+_ℓ on the lhs of (9.A.7) has exactly the same k dependence as the outgoing amplitude $\mathcal{A}^{(1)}_w$ in (9.65a). This provides a good sanity check that the solution in (9.64) is well behaved for all values of k.

Derivatives The result for the induced multipoles in (9.39) requires computing derivatives of the scalar field evaluated at the origin. Because $\partial_{L'}\phi(t,\mathbf{0})$ is contracted with the STF product $\mathbf{z}_K^{\langle L'\rangle}(t)$ in (9.10), only the STF part of the derivative contributes. This can be written as

$$\partial_{\langle L\rangle}\phi(t,\mathbf{0}) \equiv \int \mathrm{d}^3\mathbf{x}\, \delta^{(3)}(\mathbf{x})\partial_{\langle L\rangle}\phi(t,\mathbf{x}). \tag{9.A.8}$$

Since both $R_\ell(k,r)$ and $R_{n\ell}(r)$ are proportional to the Whittaker function M, a good starting point is

$$\int \mathrm{d}^3\mathbf{x}\, \delta^{(3)}(\mathbf{x})\partial_{\langle L'\rangle}\left(\frac{M_{-i\zeta,\ell+1/2}(-2ikr)}{-2ikr}Y_{\ell m}(\hat{\mathbf{x}})\right) = (-2ik)^\ell \ell!\, \mathcal{Y}^{\ell m}_{L'}\delta_{\ell\ell'}, \tag{9.A.9}$$

which follows from (9.A.3) and the identity $\partial_L \mathbf{x}^L = \ell!$. The definitions in (9.19) and (9.23) can then be used to show that

$$\frac{4\pi}{(2\ell'+1)!!}(\mathcal{Y}^{\ell' m'}_{L'})^*\partial_{L'}\phi_{k\ell m}(\mathbf{0}) = \mathfrak{R}_\ell(k)\delta^{\ell\ell'}\delta^{mm'}, \tag{9.A.10a}$$

$$\frac{4\pi}{(2\ell'+1)!!}(\mathcal{Y}^{\ell' m'}_{L'})^*\partial_{L'}\psi_{n\ell m}(\mathbf{0}) = \mathfrak{R}_{n\ell}\delta^{\ell\ell'}\delta^{mm'}, \tag{9.A.10b}$$

where the coefficients on the rhs are given by

$$\mathfrak{R}_\ell(k) = C_\ell(\zeta)k^\ell, \tag{9.A.11a}$$

$$\mathfrak{R}_{n\ell} = \sqrt{\frac{(n+\ell)!}{2(n-\ell-1)!}}\frac{(2GM\mu^2)^{\ell+3/2}}{(2\ell+1)!n^{\ell+2}}. \tag{9.A.11b}$$

Complex conjugates Several instances in the main text exploit identities for the complex conjugates of the mode functions to obtain simplified expressions. These identities are derived here. First, combining the well-known identity

$$Y^*_{\ell m}(\hat{\mathbf{x}}) \equiv (-1)^m Y_{\ell,-m}(\hat{\mathbf{x}}) \tag{9.A.12}$$

with the fact that $R_{n\ell}(r)$ is a real function tells us that the complex conjugate of a bound-state mode function is

$$\psi^*_{n\ell m}(\mathbf{x}) \equiv (-1)^m \psi_{n\ell -m}(\mathbf{x}). \tag{9.A.13}$$

As for the continuum states, the identity $[H^{\pm}_{\ell}(\zeta, z)]^* \equiv H^{\mp}_{\ell}(\zeta^*, z^*)$ [532] can be used to show that

$$[R^{\pm}_{\ell}(k,r)]^* \equiv R^{\mp}_{\ell}(k^*, r), \qquad [R_{\ell}(k,r)]^* \equiv R_{\ell}(k^*, r). \tag{9.A.14}$$

These can be written in a more useful form by utilising the circuital relations [532, 533]

$$F_{\ell}(\zeta, z) \equiv -e^{i\pi(\ell+i\zeta)} F_{\ell}(-\zeta, z e^{-i\pi}), \tag{9.A.15a}$$

$$H^{+}_{\ell}(\zeta, z) \equiv e^{-i\pi(\ell+i\zeta)} H^{-}_{\ell}(-\zeta, z e^{-i\pi}) \tag{9.A.15b}$$

in conjunction with the identity $k^*(\omega) \equiv -k(-\omega)$, which is a consequence of the definition in (9.16). Only the results for R^{+}_{ℓ} and R_{ℓ} are relevant for physical solutions. They are

$$[R^{+}_{\ell}(k,r)]^* \equiv e^{+i\pi(\ell+i\bar{\zeta})} R^{+}_{\ell}(\bar{k}, r), \tag{9.A.16a}$$

$$[R_{\ell}(k,r)]^* \equiv e^{-i\pi(\ell+i\bar{\zeta})} R_{\ell}(\bar{k}, r), \tag{9.A.16b}$$

where I write $\bar{k} \equiv k(-\omega)$ and $\bar{\zeta} \equiv \zeta(-\omega)$ as shorthand. Combined with (9.A.12), the complex conjugates of the continuum-state mode functions are

$$\phi^{+*}_{k\ell m}(\mathbf{x}) \equiv e^{-\pi\bar{\zeta}} (-1)^{\ell+m} \phi^{+}_{\bar{k}\ell -m}(\mathbf{x}), \tag{9.A.17a}$$

$$\phi^{*}_{k\ell m}(\mathbf{x}) \equiv e^{+\pi\bar{\zeta}} (-1)^{\ell+m} \phi_{\bar{k}\ell -m}(\mathbf{x}). \tag{9.A.17b}$$

Appendix 9.B Resummation and late-time behaviour

The growth of $\Gamma_u t$ in (9.58) invalidates our naive perturbative approach once it becomes of order unity, even though the expansion parameters enumerated in Section 9.1.2 all remain small. This kind of secular growth turns out to be generic in any system with an interaction Hamiltonian that persists for all times [534]. For the scenario studied in this chapter, this late-time breakdown of perturbation theory

poses no threat because the binary will typically coalesce well before $\Gamma_u t \sim 1$. That being said, on theoretical grounds, it is interesting to explore how we might obtain an approximate solution to the system of equations in (9.10) and (9.11) that remains valid at late times. The general results may find application in studies of other open systems whose lifetimes exceed $1/\Gamma_u$.

The key is to carefully resum the dominant polynomial behaviour $\propto t^p$ at each order p in perturbation theory while neglecting subleading terms. Included in this set of terms we will neglect are backreaction effects from the outgoing radiation and Yukawa modes (see Section 9.4), because they contribute to $c_u(t)$ beginning only at second order. Additionally, higher-order corrections to the formula for the induced multipoles in (9.10), which are suppressed by extra powers of v, $GM\omega$, and a/λ, can also be neglected. For added simplicity, we will also assume no ingoing radiation in this appendix.

With these considerations in mind, resummation amounts to looking for a solution of the form $\phi(x) = \sum_{p=0}^{\infty} \phi^{(p)}(x)$, where each term in this series is sourced by the previous term via the iteration

$$\phi^{(p+1)}(x) = \int \mathrm{d}^4x' G(x,x') \sum_{\ell=0}^{\infty} (-1)^\ell O_L^{(p)}(t') \partial_L \delta^{(3)}(\mathbf{x}') + \phi_{\mathrm{cf}}^{(p)}(x). \tag{9.B.1}$$

This is, of course, simply a generalisation of (9.14). Accordingly, the coefficients for the bound states at order $p+1$ are given by

$$c_u^{(p+1)}(t) = \sum_{u'} \frac{V_{uu'}}{2E_n} \int^t \mathrm{d}t' \left[(m'\Omega - E_{n'}) c_{u'}^{(p)} - i\dot{c}_{u'}^{(p)}\right] \left(e^{i\Delta_{uu'}t'} - e^{2iE_n t} e^{i(\Delta_{uu'} - 2E_n)t'}\right) \tag{9.B.2}$$

after suitably generalising (9.55). Rather than perform this string of integrals, the trick is to now differentiate twice to obtain

$$\ddot{c}_u^{(p+1)} - 2iE_n \dot{c}_u^{(p+1)} = -i \sum_{u'} V_{uu'} [(m'\Omega - E_{n'}) c_{u'}^{(p)} - i\dot{c}_{u'}^{(p)}] e^{i\Delta_{uu'}t}. \tag{9.B.3}$$

In the same way that (9.B.1) is an iterative solution to the equation of motion in (9.11), this set of differential equations in (9.B.3) can be viewed as establishing an iterative method (assuming V is suitably small) for solving the master equation[6]

$$\ddot{c}_u - 2iE_n \dot{c}_u = -i \sum_{u'} V_{uu'} [(m'\Omega - E_{n'}) c_{u'} - i\dot{c}_{u'}] e^{i\Delta_{uu'}t}. \tag{9.B.4}$$

[6]With the benefit of hindsight, this master equation can be seen to follow more easily from substituting the ansatz $\phi(x) \propto \sum_u [c_u(t)\psi_u(\mathbf{x}) e^{-iE_n t} + \text{c.c.}]$ directly into (9.10) and (9.11). However, doing so obscures the fact that the bound states alone are not a complete solution to the problem. As we discussed in Section 9.4.1, the production of outgoing radiation is inevitable in this system.

This equation is strongly reminiscent of time-dependent perturbation theory in quantum mechanics, albeit with two key differences. First, the terms on the rhs can be regarded as arising from some interaction Hamiltonian for the system, whereby $\langle u|H_{\rm int}|u'\rangle \propto -iV_{uu'}e^{-i(m-m')\Omega t}$. Given that the diagonal elements V_{uu} are real, the prefactor of $-i$ indicates that $H_{\rm int}$ is not Hermitian. This is to be expected if the system is to exhibit exponential growth — a hallmark of nonunitary evolution. The other key difference is that (9.B.4) is clearly a set of second-order, rather than first-order, differential equations; reflecting the relativistic nature of this system.

To perform the requisite resummation, we now treat the terms involving the diagonal elements V_{uu} on the rhs of (9.B.4) nonperturbatively. Moving them over to the lhs, the master equation may be rewritten as

$$\ddot{c}_u - (2iE_n - V_{uu})\dot{c}_u + 2iE_n\Gamma_u c_u = -iJ_u, \tag{9.B.5}$$

where

$$J_u(t) = \sum_{u'\neq u} V_{uu'}[(m'\Omega - E_{n'})c_{u'} - i\dot{c}_{u'}]e^{i\Delta_{uu'}t} \tag{9.B.6}$$

is independent of c_u and can therefore be regarded as a source term. To solve this equation, begin by noting that it is of the form

$$\ddot{c}_u - (\gamma_+ + \gamma_-)\dot{c}_u + \gamma_+\gamma_- c_u = -iJ_u, \tag{9.B.7}$$

where $\gamma_\pm$ are the two zeros of the characteristic polynomial

$$\gamma_\pm^2 - (2iE_n - V_{uu})\gamma_\pm + 2iE_n\Gamma_u. \tag{9.B.8}$$

As $V \ll E_n$, it suffices to use the approximate solutions

$$\gamma_+ \simeq 2iE_n - (\Gamma_u + V_{uu}) = 2iE_n + \Gamma_{\bar{u}}, \tag{9.B.9a}$$

$$\gamma_- \simeq \Gamma_u, \tag{9.B.9b}$$

having defined $\bar{u} \equiv (n, \ell, -m)$ as shorthand. For later purposes, it will also be useful to have defined $\gamma_u := (\gamma_+ - \gamma_-)/2$. Now choosing boundary conditions such that $c_u(0) = c_u^{(0)}$ and $c_u(t) \to c_u^{(0)}\ \forall\, t$ in the limit $V \to 0$, the solution to (9.B.7) is

$$c_u(t) = c_u^{(0)}e^{\gamma_- t} + \frac{i}{2\gamma_u}\int_0^t \mathrm{d}t'\big(e^{\gamma_-(t-t')} - e^{\gamma_+(t-t')}\big)J_u(t'). \tag{9.B.10}$$

This is only a formal solution because J_u depends on the other bound states $u' \neq u$, whose solutions are also given by (9.B.10). To obtain an explicit result, we should iterate (9.B.10) in powers of $V/\gamma \ll 1$. Starting with $c_u(t) = c_u^{(0)} e^{\Gamma_u t} + \mathcal{O}(V/\gamma)$, we find

$$c_u(t) = c_u^{(0)} e^{\Gamma_u t} + \sum_{u' \neq u} \frac{V_{uu'}}{2\gamma_u} (m'\Omega - \widetilde{E}_{u'}) \left[\frac{e^{i\Delta_{uu'}t} e^{\Gamma_{u'}t} - e^{\Gamma_u t}}{\widetilde{\Delta}_{uu'}} \right.$$

$$\left. - \frac{e^{i\Delta_{uu'}t} e^{\Gamma_{u'}t} - e^{2iE_n t} e^{\Gamma_{\bar{u}}t}}{\widetilde{\Delta}_{uu'} + 2i\gamma_u} \right] c_{u'}^{(0)} + \mathcal{O}((V/\gamma)^2) \tag{9.B.11}$$

after one iteration, where $\widetilde{E}_u = E_n + i\Gamma_u$ is the complex frequency of the quasibound state and $\widetilde{\Delta}_{uu'}$ is defined in the same way as $\Delta_{uu'}$ in (9.56a), except with $\widetilde{E}_u$ in place of E_u. Having carefully resummed the leading polynomial growth to all orders, this solution is valid at late times $t \gg 1/\Gamma_u$ while still being organised as a perturbative expansion in the small parameter V/γ.

This resummed solution also brings with it a new prediction. Let us denote the fastest-growing mode by $u_\star$ and its growth rate by $\Gamma_\star$. The sum over u' in (9.B.11), which quantifies the leading effects due to mode mixing, then tells us that all modes satisfying $V_{uu_\star} \neq 0$ will grow at the same rate $\Gamma_\star$ at late times, even if they were initially decaying.[7] While this is a nontrivial result for the system of equations under study, it is irrelevant in the case of a scalar cloud around a binary black hole because even the shortest e-folding time $1/\Gamma_\star$ is always orders of magnitude greater than the orbital decay timescale $1/\Gamma_{\rm GW}$. It would therefore be interesting to explore if there are other open systems that could survive long enough to exhibit this universal growth rate at late times.

Appendix 9.C Scalar-wave flux

Multiple instances in this thesis call for the evaluation of the power radiated to infinity in scalar waves. For the sake of efficiency, we will here derive a general formula for this energy flux. To begin with, recall from our discussion in Section 9.4 that a generic linear combination of continuum states is given by the sum [cf. (9.70)]

$$\phi_c(x) = \sum_w \left[\mathcal{I}_w R_\ell^-(k,r) + \mathcal{R}_w R_\ell^+(k,r) \right] Y_{\ell m}(\hat{\mathbf{x}})\, e^{-i\omega t}. \tag{9.C.1}$$

At any point in time, the rate at which energy is carried off to infinity by these scalar waves is given by integrating the (t,r) component of this scalar's energy–momentum tensor over a spherical shell of

[7] This phenomenon occurs when the $\mathcal{O}(V/\gamma)$ terms dominate over the first term in (9.B.11). Nonetheless, our perturbative expansion is still valid because the $\mathcal{O}((V/\gamma)^2)$ terms remain subleading.

radius r and taking the limit $r \to \infty$:

$$\dot{E}_{\rm SW} = \lim_{r\to\infty} -r^2 \int \mathrm{d}^2\Omega\, \partial_t\phi\, \partial_r\phi. \tag{9.C.2}$$

(The measure $\mathrm{d}^2\Omega$ is the infinitesimal solid angle element and is not to be confused with the binary's orbital frequency.) Substituting (9.C.1) into (9.C.2) and using the fact that ϕ_c is real then yields

$$\begin{aligned}\dot{E}_{\rm SW} = \lim_{r\to\infty} -r^2 \sum_{w,w'} \int \mathrm{d}^2\Omega\, Y_{\ell m}(\hat{\mathbf{x}})\, Y^*_{\ell' m'}(\hat{\mathbf{x}})\, (-i\omega)(-ik')^* e^{-i(\omega-\omega')t} \\ \times \big[\mathcal{I}_w R^-_\ell(k,r) + \mathcal{R}_w R^+_\ell(k,r)\big]\big[\mathcal{I}_{w'} R^-_{\ell'}(k',r) - \mathcal{R}_{w'} R^+_{\ell'}(k',r)\big]^*.\end{aligned} \tag{9.C.3}$$

Use has also been made of the fact that only the radiation modes with $k^2 \geq 0$ survive in the limit $r \to \infty$. Their asymptotic forms in (9.A.1) can then be used to show that $\partial_r R^\pm_\ell(k,r) \sim \pm ik R^\pm_\ell(k,r)$.

The result is still a complicated integral over ω and ω' at this stage, but it will simplify drastically if we now apply a time-averaging procedure. To elaborate, notice that if we integrated the above expression for the instantaneous flux $\dot{E}_{\rm SW}$ over a duration Δt and then divide by Δt to obtain the time-averaged flux $\langle \dot{E}_{\rm SW}\rangle$, the exponential in the integrand would be replaced by the function $\mathrm{sinc}((\omega-\omega')\Delta t/2)$, up to a phase factor. Taking the limit in which Δt is much longer than any of the other timescales in the problem then allows us to discard all terms except those with $\omega \approx \omega'$. Equivalently, if we just integrated over some duration Δt and took the limit $\Delta t \to \infty$, the exponential in the integrand turns into a delta function that enforces the condition $\omega = \omega'$. Performing the integral over ω' then gives us a particularly simple expression for the total energy radiated over all time:

$$\begin{aligned}\int_{-\infty}^{\infty} \mathrm{d}t\, \dot{E}_{\rm SW} = \lim_{r\to\infty} \sum_w \omega k \big(|\mathcal{R}_w|^2 - |\mathcal{I}_w|^2\big) R^+_\ell(k,r) R^-_\ell(k,r) \\ + \lim_{r\to\infty} \sum_w \omega k r^2 \Big\{\mathcal{I}_w \mathcal{R}^*_w [R^-_\ell(k,r)]^2 - \mathcal{I}^*_w \mathcal{R}_w [R^+_\ell(k,r)]^2\Big\}.\end{aligned} \tag{9.C.4}$$

In obtaining this expression, I have made use of the identity in (9.A.14) and have also exploited the orthogonality of the spherical harmonics when integrating over the angular variables. Two more steps are needed to arrive at the desired result in (9.71). First, notice that the cross terms in the second line of (9.C.4) are pure imaginary and so must necessarily vanish since the lhs of the equation is real by construction. We are then left with the first line, which simplifies to

$$\int_{-\infty}^{\infty} \mathrm{d}t\, \dot{E}_{\rm SW} = \sum_w \frac{\omega}{k} \theta(k^2) \big(|\mathcal{R}_w|^2 - |\mathcal{I}_w|^2\big) \tag{9.C.5}$$

after substituting in (9.A.1). Dividing by the total time $\int \mathrm{d}t \equiv 2\pi\delta(0)$ then gives us the time-averaged flux $\langle \dot{E}_{\mathrm{SW}} \rangle$. As in the main text, we will omit writing the angled brackets hereafter.

Special cases As it currently stands, the result in (9.C.5) is valid for any generic sum over radiation modes. In the remainder of this appendix, we will simplify it further by specialising to the kinds of scenarios explored in Section 9.4. For these specific cases, the components of the induced multipoles take on the general form

$$O^{(0)}_{\ell m}(t) = o_{\ell m} e^{-i\omega_m t} - (-1)^m o^*_{\ell,-m} e^{i\omega_{-m} t}. \tag{9.C.6}$$

Comparing this expression with the result in (9.39) tells us that the complex coefficients $o_{\ell m}$ and the real frequencies ω_m are given by

$$o_{\ell m} = \frac{1}{\sqrt{2\mu}} \frac{Y^*_{\ell m}(\mathbf{d}) Y_{\hat{\ell}\hat{m}}(\mathbf{d}) B_{\ell\hat{\ell}}}{(2\ell+1)!!} \mathfrak{R}_{\hat{n}\hat{\ell}}(\hat{m}\Omega - E_{\hat{n}}) c^{(0)}_{\hat{u}}, \tag{9.C.7a}$$

$$\omega_m = E_{\hat{n}} + (m - \hat{m})\Omega \tag{9.C.7b}$$

for the case of a scalar cloud comprised of a single bound state $\hat{u} \equiv (\hat{n}, \hat{\ell}, \hat{m})$. If we instead suppose that the zeroth-order solution contains only the single ingoing radiation mode $\hat{w} \equiv (\hat{\omega}, \hat{\ell}, \hat{m})$, with $\mathcal{I}^{>}_{w}$ given by (9.81), we would then have

$$o_{\ell m} = \frac{Y^*_{\ell m}(\mathbf{d}) Y_{\hat{\ell}\hat{m}}(\mathbf{d}) B_{\ell\hat{\ell}}}{(2\ell+1)!!} \mathfrak{R}_{\hat{\ell}}(\hat{k})(\hat{m}\Omega - \hat{\omega}) 2\Phi_{\hat{w}}, \tag{9.C.8a}$$

$$\omega_m = \hat{\omega} + (m - \hat{m})\Omega. \tag{9.C.8b}$$

By substituting (9.C.6) into (9.65a), we see that these induced multipole moments go on to source corrections to the outgoing amplitude given by

$$\mathcal{A}^{(1)}_{w} = s_\ell(\zeta) k^{\ell+1} \left[2\pi\delta(\omega - \omega_m) o_{\ell m} - (-1)^m 2\pi\delta(\omega + \omega_{-m}) o^*_{\ell,-m} \right] \tag{9.C.9}$$

at first order in perturbation theory. Because the energy flux is quadratic in the scalar field, there are contributions at first and second order in $\mathcal{A}^{(1)}_{w}$. Written out explicitly, the total energy that the binary loses to scalar radiation is

$$\int \mathrm{d}t\ \dot{E}_{\mathrm{SW}} = \sum_{w} \frac{\omega}{k} \theta(k^2) \left(2\,\mathrm{Re}\, \mathcal{I}^*_{w} \mathcal{A}^{(1)}_{w} + \left| \mathcal{A}^{(1)}_{w} \right|^2 \right). \tag{9.C.10}$$

First order Let us begin by evaluating the term in (9.C.10) that is linear in $\mathcal{A}_w^{(1)}$. After substituting in (9.C.9), we find that the energy flux at first order is

$$\int \mathrm{d}t\ \dot{E}_{\mathrm{SW}}^{(1)} = 2\,\mathrm{Re}\sum_w \theta(k^2) s_\ell(\zeta)\big[2\pi\delta(\omega-\omega_m)o_{\ell m} - (-1)^m 2\pi\delta(\omega+\omega_{-m})o^*_{\ell,-m}\big]\mathcal{I}^*_w \omega k^\ell. \quad (9.C.11)$$

To proceed, we deduce from the definition of $s_\ell(\zeta)$ in (9.65b) that $s_\ell(-\zeta) = s^*_\ell(\zeta)e^{\pi\zeta}$ for radiation modes with $\zeta \in \mathbb{R}$. Combined with the freedom to relabel $\omega \to -\omega$ and $m \to -m$ as they are being integrated and summed over, respectively, we find

$$\int \mathrm{d}t\ \dot{E}_{\mathrm{SW}}^{(1)} = 2\,\mathrm{Re}\sum_w \theta(k^2) 2\pi\delta(\omega-\omega_m)\big[\mathcal{I}^*_{\omega\ell m} s_\ell(\zeta) o_{\ell m} + e^{\pi\zeta}(-1)^{\ell+m}\mathcal{I}^*_{-\omega\ell-m} s^*_\ell(\zeta) o^*_{\ell m}\big]\omega k^\ell. \quad (9.C.12)$$

The identity in (9.69) may now be used to show that the two terms in square brackets are complex conjugates of one another, thus this expression further simplifies to

$$\int \mathrm{d}t\ \dot{E}_{\mathrm{SW}}^{(1)} = 4\,\mathrm{Re}\sum_w \theta(k^2) 2\pi\delta(\omega-\omega_m) o_{\ell m} s_\ell(\zeta)\omega k^\ell \mathcal{I}^*_w. \quad (9.C.13)$$

Because this formula assumes that the zeroth-order solution $\phi^{(0)}$ is comprised of just a single bound state $\hat{u}$ or a single ingoing radiation mode $\hat{w}$, (9.C.13) is nonvanishing only in the latter case. Substituting in the expression for $\mathcal{I}^>_w$ in (9.81) and dividing by a common factor of $\int \mathrm{d}t \equiv 2\pi\delta(0)$ on both sides ultimately yields

$$\dot{E}_{\mathrm{SW}}^{(1)} = 4\,\mathrm{Re}\, s_{\hat{\ell}}(\hat{\zeta})\hat{\omega}\hat{k}^{\hat{\ell}} o_{\hat{\ell}\hat{m}} \Phi^*_{\hat{w}}. \quad (9.C.14)$$

Second order Let us now turn our attention to the $\mathcal{O}(\mathcal{A}^2)$ term in the energy flux. Taking the absolute square of (9.C.9), one finds

$$\big|\mathcal{A}_w^{(1)}\big|^2 = S_\ell(\zeta) k^{2(\ell+1)} 2\pi\delta(0)\big[2\pi\delta(\omega-\omega_m)|o_{\ell m}|^2 + 2\pi\delta(\omega+\omega_{-m})|o_{\ell,-m}|^2\big]. \quad (9.C.15)$$

To arrive at this result, use the fact that the cross terms proportional to $\delta(\omega-\omega_m)\delta(\omega+\omega_{-m})$ may be discarded as they have nonoverlapping support. (This is because the frequencies generally satisfy the condition $\omega_m \neq -\omega_{-m}$ except when $\hat{m}\Omega - E_{\hat{n}} = 0$ or $\hat{m}\Omega - \hat{\omega} = 0$, but the coefficient $o_{\ell m}$ vanishes in this case.) It then follows that

$$\int \mathrm{d}t\ \dot{E}_{\mathrm{SW}}^{(2)} = 2\pi\delta(0)\sum_w \theta(k^2) S_\ell(\zeta)\omega k^{2\ell+1}|o_{\ell m}|^2[2\pi\delta(\omega-\omega_m) + 2\pi\delta(\omega+\omega_m)] \quad (9.C.16)$$

after also using the freedom to relabel $m \to -m$. To simplify this result one step further, note that the product $\omega k^{2\ell+1}$ is invariant under the transformation $\omega \to -\omega$, whereas ζ changes sign; hence, the integral over ω can easily be performed, with the end result being

$$\dot{E}^{(2)}_{\rm SW} = \sum_{\ell,m} \theta(k_m^2)\big[S_\ell(\zeta_m) + S_\ell(-\zeta_m)\big]|o_{\ell m}|^2 \omega_m k_m^{2\ell+1}, \tag{9.C.17}$$

after dividing by $\int {\rm d}t \equiv 2\pi\delta(0)$ and defining $k_m \equiv k(\omega_m)$, and likewise $\zeta_m \equiv \zeta(\omega_m)$.

Ingoing flux The ratio of $\dot{E}_{\rm SW}$ to $\dot{E}^{\rm in}_{\rm SW}$ is often a useful measure. For a single ingoing mode $\hat{w}$, the latter is given by substituting (9.81) into (9.71b). After neglecting cross terms that involve products of delta functions with nonoverlapping support, the end result is

$$\dot{E}^{\rm in}_{\rm SW} = \frac{\hat{\omega}}{\hat{k}}|\Phi_{\hat{w}}|^2\big(1 + e^{-2\pi\hat{\zeta}}\big), \tag{9.C.18}$$

where the exponential arises from the identity in (9.69).

Appendix 9.D Arbitrary mass ratios

For a binary composed of spherical (sph) black holes, the growth rate for the bound state $u \equiv (n, \ell, m)$ may be written as

$$\left(\frac{\Gamma_u}{\Gamma_{\rm GW}}\right)_{\rm sph} = \frac{\mathfrak{n}_\ell(\nu)}{\nu}\left|\frac{Y_{\ell m}(\mathbf{d})}{(2\ell+1)!}\right|^2 \frac{5\pi(n+\ell)!}{(n-\ell-1)!4n^{2\ell+4}} v^{8\ell+10} f_\mu^{4\ell+5}(m - f_\mu) \tag{9.D.1}$$

in terms of the dimensionless parameters v, $\nu := M_1M_2/M^2$, and $f_\mu := \mu/\Omega$. Meanwhile, the amplification factor for the primary mode $w \equiv (\omega, \ell, m)$ is

$$Z_{\rm sph} = \mathfrak{n}_\ell(\nu)\left(\frac{2S_\ell(\zeta)}{1 + e^{-2\pi\zeta}}\right)\left|\frac{Y_{\ell m}(\mathbf{d})}{(2\ell+1)!!}\right|^2 \frac{32\pi}{4^\ell} v^{2\ell+6}(f_\omega^2 - f_\mu^2)^{\ell+1/2}(m - f_\omega). \tag{9.D.2}$$

In this case, the expression depends on the four dimensionless quantities v, ν, f_μ, and $f_\omega := \omega/\Omega$. Also note that $\zeta \equiv -v^3(f_\omega^2/f_\mu^2 - 1)^{-1/2}$. Crucially, in both formulae, the effect of adjusting the symmetric mass ratio ν is encapsulated by the same function

$$\mathfrak{n}_\ell(\nu) := 8\nu^2\big[(1 + \sqrt{1-4\nu})^{2\ell-2} + (1 - \sqrt{1-4\nu})^{2\ell-2}\big], \tag{9.D.3}$$

which is normalised such that $\mathfrak{n}_\ell(1/4) = 1$.

The additional prefactor of $1/\nu$ in (9.D.1) causes the ratio $\Gamma_u/\Gamma_{\rm GW}$ to diverge in the limit $\nu \to 0$. This singularity is unrelated to Γ_u and is due entirely to $\Gamma_{\rm GW}$. Physically, it is reflecting the fact that the timescale over which the orbit shrinks becomes infinite in the limit of a test particle around a host black hole. Ignoring this behaviour, it follows that Γ_u and Z are both proportional to $\mathfrak{n}_\ell(\nu)$. We shall focus on understanding the properties of this function.

It presents three different classes of behaviour depending on the value of ℓ. For $\ell = 0$, the largest value of $\mathfrak{n}_0(\nu) = 2 - 4\nu$ in the domain $\nu \in (0, 1/4]$ is $\mathfrak{n}_0(0) = 2$. This behaviour has a simple physical interpretation: For a binary of fixed total mass M, a smaller symmetric mass ratio gives rise to a larger combined area for the black holes' horizons; hence, that $\mathfrak{n}_0(\nu)$ is maximised when $\nu = 0$ simply corroborates the fact that absorption is more efficient when there is a larger surface area. Note, however, that the value of this function only changes by a factor of 2 in the domain $\nu \in (0, 1/4]$. For $\ell = 1$, one finds $\mathfrak{n}_1(\nu) = 16\nu^2$, which has maximum value of $\mathfrak{n}_1(1/4) = 1$ when the binary's components are of equal mass. For $\ell \geq 2$, this function always has a maximum somewhere in the domain $\nu \in (0, 1/4]$. Numerically, I find $\max \mathfrak{n}_\ell(\nu) \sim \exp(1.4\ell - 2.0 \log \ell - 1.5)$ when $\ell \gg 1$, which can be quite a large number. For instance, $\max \mathfrak{n}_{20}(\nu) \sim 10^9$.

What does this mean for the conclusions in the main text? Given that $\mathfrak{n}_1(\nu)$ is maximised for equal-mass binaries, the growth rates for the $\ell = 1$ modes (shown in Figure 9.2) and the amplification factors for the same modes (shown in Figure 9.3) are indeed the largest values possible within the EFT's regime of validity. For larger values of ℓ, carefully selecting an optimal value for ν can enhance the growth rates and amplification factors relative to the equal-mass case, but this enhancement grows exponentially with ℓ at best, which is still no match for the factorials in the denominators of (9.D.1) and (9.D.2). Consequently, the general trend remains unchanged: The maximum value that Γ_u or Z can attain decreases rapidly as we increase μ or ω.

Chapter 10
Conclusion

Light scalar fields have become ubiquitous in modern theoretical physics as potential solutions to some of the most important open problems. In this thesis alone, we have seen how the properties of these fields make them appealing candidates for both dark matter and dark energy, as well as compelling solutions to outstanding puzzles in particle physics like the strong *CP* problem. Moreover, candidate quantum-gravity theories like string theory invariably predict that large numbers of these scalar fields should exist in our Universe; hence, the discovery of a new, light scalar degree of freedom would have far-reaching implications for our prevailing notions of what might lie beyond the Standard Model. It is only fitting, then, that efforts to detect or otherwise rule out these hypothetical scalars have now flourished into an active area of ongoing research.

In this thesis, we have made two key contributions in support of this search for new physics. First, we established new constraints on chameleon-like scalars in Part I by quantifying their impact on the energy levels of electrons in two high-precision laboratory experiments. Because each one delivers an independent measurement of the fine-structure constant that is in good agreement with the other, we were able to impose an upper limit on the size of the effects that a chameleon-like scalar can introduce. Second, in Part II, we developed a novel worldline effective field theory to better understand the influence that light, minimally coupled scalar fields exert on binary black holes. The resulting framework is the first of its kind to systematically incorporate the dissipative effects associated with a black hole's horizon at the level of the action and, as a result, has facilitated analytic calculations that have proven prohibitively difficult for more traditional approaches, like that of matched asymptotic expansions. To draw this thesis to a close, let us summarise our main findings.

In Chapter 3, we showed that a chameleon-like scalar introduces systematic corrections to the gross, fine, and hyperfine structure of atomic hydrogen and other one-electron atoms in two different ways.

First, the mass and electromagnetic fields of the nucleus source perturbations to the ambient scalar-field profile, which go on to exert an attractive fifth force on the electron. Second, the nonminimal interaction between this scalar-field profile and the electromagnetic fields of the nucleus triggers a polarisation of the vacuum already at the classical level, whose effect on the electron's energy levels is comparable in size to that of the fifth force. Also in this chapter, we examined the response of hydrogen atoms to external perturbations from a background scalar-field gradient, as might be found in unscreened astrophysical environments. Interestingly, simple arguments based on parity were sufficient to prove that their spectral lines are insensitive (at first order in G) to even the largest gradients imaginable.

In Chapter 4, we then turned our attention to the prospect of using the electron's magnetic dipole moment g_e as a probe of chameleon-like scalars. Two qualitatively distinct effects were seen to arise. First, a chameleon-like scalar increases the intrinsic value of g_e by generating additional loop corrections to the QED vertex function. Because its screening mechanism allows the scalar to dynamically alter its effective mass and/or couplings to the Standard Model based on changes in the ambient density, the magnitudes of these quantum corrections vary accordingly as a function of the local environment. Second, the formation of a bubble profile within the vacuum cavity in which g_e is measured introduces a systematic bias into the experiment unless accounted for. This "cavity shift," as we called it, arises because the electron's magnetic moment is determined experimentally from measurements of its eigenfrequencies in the cavity, but these eigenfrequencies are shifted when the scalar bubble exerts a fifth force on the electron.

Accurately evaluating the size of both of these effects required knowledge about the scalar-field profile at the centre of the vacuum cavity, which we ascertained by using a novel, semi-empirical approach. In short, we constructed estimates for the scalar's central field value ϕ_0 by starting with known, analytic solutions for bubble profiles in a number of one-dimensional setups, which we then smoothly deformed by introducing free parameters that account for geometric differences between the actual experiment and our one-dimensional toy models. After calibrating just two of these free parameters to match a small set of full three-dimensional solutions that were obtained numerically, these semi-empirical formulae were shown to offer a quick and reliable way of determining ϕ_0 across large swathes of parameter space.

In Chapter 5, we imposed new constraints on chameleon-like scalars by demanding that their effects should not jeopardise the good agreement between the values of the fine-structure constant inferred from these experiments. For the chameleon model, placing limits on the allowed strength of its fifth force in atomic hydrogen translates into the upper bound $\beta_m = M_{\mathrm{Pl}}/M_c \lesssim 10^{15}$ on its coupling strength

to matter,[1] while a restriction on the amount of vacuum polarisation that is allowed establishes the upper bound $\beta_\gamma = M_{\rm Pl}/M_\gamma \lesssim 10^{21}$ on its coupling to photons. Both of these constraints are universal for all values of the other two parameters n and Λ, as long as the EFT remains valid, but the cavity shift associated with the electron magnetic moment experiment establishes the more stringent bound $M_{\rm Pl}/M_c \lesssim 10^{10}$ in the region around $\Lambda \approx 2$ keV for $n = 1$. (For constraints on models with $n \geq 2$, see Figure 5.1). While this part of parameter space has already been ruled out by other laboratory experiments, this bound represents the tightest constraint yet from an experiment that was not originally designed to search for fifth forces.

The same phenomenon also imposes constraints on the symmetron model, albeit only in a very small range of the mass parameter μ. This limitation is unsurprising and is, in fact, generic to any laboratory experiment that establishes constraints by probing the symmetron's fifth force. The reason is that the symmetron is obliged to remain in its unbroken phase when its Compton wavelength is too large to fit within the size of the experimental setup, meaning that the cavity shift vanishes for values of μ below a certain threshold. At the other end of the spectrum, the symmetron faces no difficulty with the spontaneous breaking of its $\mathbb{Z}_2$ symmetry when μ is large, but its fifth force is now strongly Yukawa suppressed, thereby resulting in a field profile that is essentially flat everywhere in the cavity except near the walls. Decidedly, the sweet spot lies somewhere in between these two extremes and, indeed, we showed that placing an upper limit on the size of the cavity shift excludes a small region of the symmetron's parameter space around $\mu \approx 0.15$ meV.

That said, what sets this work apart from other laboratory experiments that have hitherto been used to probe the symmetron are the quantum corrections to g_e, which allowed us to establish constraints almost independently of μ, provided only that it is large enough to elicit spontaneous symmetry breaking but still small enough that the EFT remains valid. As Figures 5.2 to 5.4 demonstrate, bounding the size of these quantum corrections has decisively ruled out a large and previously unexplored region of parameter space spanning 12 orders of magnitude in μ for couplings (M_s, M_γ) around the GeV scale. On the whole, the results in Part I of this thesis therefore paint a clearer picture of the space of chameleon-like theories that remain viable in our Universe. On top of that, they point to an interesting direction for future work: While dedicated fifth-force experiments like atom interferometry may well provide the best sensitivities in a given mass range, it will be interesting to explore if other experiments that exploit the quantum nature of the symmetron could be used to cover large regions of parameter space more efficiently.

[1] This effect and the resulting bound were first derived by Brax and Burrage [399], but are included in this summary for completeness. The remaining results discussed herein are all original to this thesis.

Leaving the confines of high-vacuum laboratory experiments behind, in Part II of this thesis we turned to an investigation of astrophysical black holes and their dynamical response to ambient, long-wavelength scalar fields. Although stringent no-hair theorems place certain limits on the kinds of interactions that can occur, we demonstrated through explicit calculation that these systems still exhibit an immensely rich phenomenology due to the finite-size effects associated with a black hole's horizon. Central to this achievement was the effective field theory we constructed in Chapters 6 and 7, which treats a black hole as a point particle moving along the worldline of some effective centre-of-energy coordinate. Through a novel combination of integrating out short-distance modes localised on the worldline and matching to observables computed from black hole perturbation theory, we were able to derive an effective action that accurately models how black holes in general relativity interact with long-wavelength scalar fields.

Interestingly, these interactions can all be understood in terms of dynamical multipole moments that the scalar field induces onto the black hole. While the idea itself dates back to much earlier work by Jacobson [309], adopting the viewpoint of an EFT spurred two key advancements in our understanding of this phenomenon. First, we have shown that it is completely generic: any ambient scalar-field profile will induce a charge (and higher multipoles, more generally) onto a black hole, provided only that there is a hierarchy separating the length and timescales of the black hole from those of its local environment. The onset of this charge then enables the black hole to radiate energy and momentum into scalar waves as it moves through this external scalar field. Second, we found that the same terms in the effective action that describe this phenomenon also capture the leading backreaction of the scalar onto the black hole — at no extra cost. Consequently, once the Wilsonian coefficients in the action are fixed to give the correct magnitude for the induced scalar charge, the EFT automatically specifies the resulting force that an ambient scalar exerts onto a moving black hole.

At leading order, this force is made up of two terms: a familiar term that describes the drag force due to accretion of the scalar onto the black hole, and a novel term, which stipulates that a black hole's trajectory is also influenced by scalar-field gradients. Remarkably, when taken together, the overall force has the same mathematical structure as the fifth force in Part I of this thesis. A crucial distinction must be drawn, however. While the fifth force in Part I is the result of a fundamental interaction that nonminimally couples the scalar to matter fields, the force in this present context is an emergent one pertaining to scalar fields that are minimally coupled to gravity. Physically, it describes the backreaction that a long-wavelength scalar exerts on the motion of a black hole as it crosses the horizon.

In Chapter 8, we applied this newly minted EFT to study the impact of embedding a binary black hole in the central core of a fuzzy dark matter halo. Scalar radiation and the "fifth force" aside, we found

that the EFT also provides a natural language for computing other effects, like dynamical friction, that are not unique to black holes, but which influence the motion of any massive body. Ultimately, however, typical halos in these fuzzy dark matter models were seen to be too dilute to leave any observable imprint on the binary's inspiral. Nevertheless, the work in this chapter proved useful for illustrating how quantitative predictions can be made within this framework.

The predictive capabilities of this EFT were on even greater display in Chapter 9, where we used explicit calculations to show that long-wavelength scalar fields evolving on binary black hole spacetimes are susceptible to a novel variant of superradiance. Not unlike what we are used to in the case of a single rotating black hole, this phenomenon — which we dubbed "orbital superradiance"— triggers the amplification of incident, low-frequency radiation and fuels the exponential growth of scalar clouds that are gravitationally bound to the binary as whole. Unique to this phenomenon, however, is the inherent lack of symmetry in the underlying spacetime, which brings about a number of other interesting effects like mode mixing, beating patterns, and the upscattering of bound states into outgoing radiation. These effects highlight the rich phenomenology that can arise in systems with horizons (or dissipative channels, more broadly) when time-translation invariance and rotational symmetry are weakly[2] broken, but unfortunately, plugging in numbers revealed that orbital superradiance is grossly inefficient. The energy extracted from a binary black hole to drive these processes is always negligible for systems in the long-wavelength limit. Moreover, the way these effects scale with the EFT's expansion parameters suggests that this conclusion holds also for scalar fields with higher frequencies or larger masses.

Although clearly disappointing from an observational standpoint, this should not be allowed to detract from the progress that has been made. For the phenomenologist, these results still constitute useful information about which effects play an important role during a binary's lifetime. As for the theorist, the calculations underpinning these results demonstrate the efficacy of using modern EFT techniques to study complex systems with multiple, largely separated scales. Even so, it is almost certain that we have only just scratched the surface in terms of exploring what this kind of EFT approach can do for us. Now, looking towards the future, let us conclude by speculating on a number of promising ways in which the results presented in Part II of this thesis could be extended.

First, on the subject of orbital superradiance, it will be interesting to generalise the calculations in Chapter 9 to the case of higher-spin fields. As we discussed in Chapter 1, the amplification of electromagnetic and gravitational waves by a single rotating black hole is substantially more pronounced than it is for scalars, and likewise, the corresponding growth rates for massive spin-1 and spin-2 fields are much greater than that of a scalar cloud. Establishing if the same trend holds also for fields

[2]In the sense that that the interaction terms which break these symmetries are treated perturbatively in the EFT.

around a binary black hole, or if it is the opposite, carries interesting implications for our theoretical understanding of general relativity either way.

Superradiance aside, natural next steps would also include relaxing some of the assumptions we made in Chapters 8 and 9. For instance, it is conceivable that generalising to the case of eccentric orbits, or accounting for the orbital decay due to gravitational radiation, will teach us something new about how binary black holes interact with external fields. Additionally, fields that carry substantial amounts of angular momentum are likely to have interesting effects on the precession of the binary's orbital plane (and presumably on the black holes' spins as well, though to a lesser degree), but such effects have yet to be explored in any detail.

Perhaps the most important assumption to relax, however, is that of the long-wavelength limit. All work to date, including the upward trends in Figures 9.2(b) and 9.3(b), point to the likelihood of ambient fields with higher frequencies or larger masses having a more dramatic impact on the orbital evolution of a binary black hole. Of particular interest is the case of an ambient field whose characteristic size λ is comparable to the binary's orbital separation a, as resonant excitations of the orbit may occur in this regime. To clarify, the EFT we constructed in Chapter 7 is valid as long as the typical size of the black holes is much smaller than both a and λ; the added assumption that $a \ll \lambda$ was made purely for convenience.[3] Thus, at least in principle, this EFT can be used to describe how a binary interacts with an ambient field when $\lambda \sim a$, but whether the corresponding equations of motion will be useful in practice, or if this regime is only amenable to numerical methods, remains to be seen.

Finally, going beyond black holes and general relativity, it will be interesting to also apply the novel techniques developed herein to study how different astrophysical objects interact with external scalar, vector, or tensor fields. In this and many other respects, modern EFT techniques are poised to play an important role as we seek to gain further insight into the nature of gravity, and our Universe as a whole.

[3] We also require that the black holes are much smaller than the Compton wavelength of the ambient field and its characteristic timescale; see the discussion around (7.3) for details.

References

[1] L. K. Wong and A.-C. Davis, One-electron atoms in screened modified gravity, Phys. Rev. D **95**, 104010 (2017).

[2] P. Brax, A.-C. Davis, B. Elder, and L. K. Wong, Constraining screened fifth forces with the electron magnetic moment, Phys. Rev. D **97**, 084050 (2018).

[3] L. K. Wong, A.-C. Davis, and R. Gregory, Effective field theory for black holes with induced scalar charges, Phys. Rev. D **100**, 024010 (2019).

[4] L. K. Wong, Superradiant scattering by a black hole binary, Phys. Rev. D **100**, 044051 (2019).

[5] L. K. Wong, Evolution of diffuse scalar clouds around binary black holes, Phys. Rev. D **101**, 124049 (2020).

[6] G. Gabrielse, The standard model's greatest triumph, Phys. Today **66**, 64 (2013).

[7] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Binary black hole mergers in the first Advanced LIGO observing run, Phys. Rev. X **6**, 041015 (2016); Erratum ibid **8**, 039903 (2018).

[8] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs, Phys. Rev. X **9**, 031040 (2019).

[9] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Tests of general relativity with GW150914, Phys. Rev. Lett. **116**, 221101 (2016); Erratum ibid **121**, 129902 (2018).

[10] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), Tests of general relativity with the binary black hole signals from the LIGO-Virgo Catalog GWTC-1, Phys. Rev. D **100**, 104036 (2019).

[11] O. Lahav and A. R. Liddle, Cosmological parameters, in M. Tanabashi et al. (Particle Data Group), Review of Particle Physics, Phys. Rev. D **98**, 030001 (2018) and 2019 update.

[12] S. Weinberg, Ultraviolet divergences in quantum theories of gravitation, in *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, 1979) pp. 790–831.

[13] J. F. Donoghue, General relativity as an effective field theory: The leading quantum corrections, Phys. Rev. D **50**, 3874 (1994).

[14] C. P. Burgess, Quantum gravity in everyday life: General relativity as an effective field theory, Living Rev. Relativity **7**, 5 (2004).

[15] A. G. Riess, L. Macri, S. Casertano, M. Sosey, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha, W. Li, R. Chornock, and D. Sarkar, A redetermination of the Hubble constant with the Hubble Space Telescope from a differential distance ladder, Astrophys. J. **699**, 539 (2009).

[16] A. G. Riess, L. Macri, S. Casertano, H. Lampeitl, H. C. Ferguson, A. V. Filippenko, S. W. Jha, W. Li, and R. Chornock, A 3% solution: Determination of the Hubble constant with the Hubble Space Telescope and Wide Field Camera 3, Astrophys. J. **730**, 119 (2011); Errata ibid **732**, 129 (2011) and ibid **752**, 76 (2012).

[17] G. Efstathiou, H_0 revisited, Mon. Not. R. Astron. Soc. **440**, 1138 (2014).

[18] A. G. Riess, L. M. Macri, S. L. Hoffmann, D. Scolnic, S. Casertano, A. V. Filippenko, B. E. Tucker, M. J. Reid, D. O. Jones, J. M. Silverman, R. Chornock, P. Challis, W. Yuan, P. J. Brown, and R. J. Foley, A 2.4% determination of the local value of the Hubble constant, Astrophys. J. **826**, 56 (2016).

[19] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic, Large Magellanic Cloud Cepheid standards provide a 1% foundation for the determination of the Hubble constant and stronger evidence for physics beyond ΛCDM, Astrophys. J. **876**, 85 (2019).

[20] P. Bull et al., Beyond ΛCDM: Problems, solutions, and the road ahead, Phys. Dark Univ. **12**, 56 (2016).

[21] M. Ishak, Testing general relativity in cosmology, Living Rev. Relativity **22**, 1 (2019).

[22] S. Weinberg, The cosmological constant problem, Rev. Mod. Phys. **61**, 1 (1989).

[23] B. Ratra and P. Peebles, Cosmological consequences of a rolling homogeneous scalar field, Phys. Rev. D **37**, 3406 (1988).

[24] R. Caldwell, R. Dave, and P. J. Steinhardt, Cosmological imprint of an energy component with general equation of state, Phys. Rev. Lett. **80**, 1582 (1998).

[25] S. Tsujikawa, Quintessence: A review, Class. Quantum Gravity **30**, 214003 (2013).

[26] D. Lovelock, The uniqueness of the Einstein field equations in a four-dimensional space, Arch. Ration. Mech. Anal. **33**, 54 (1969).

[27] D. Lovelock, The Einstein tensor and its generalizations, J. Math. Phys. **12**, 498 (1971).

[28] D. Lovelock, The four-dimensionality of space and the Einstein tensor, J. Math. Phys. **13**, 874 (1972).

[29] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, Modified gravity and cosmology, Phys. Rep. **513**, 1 (2012).

[30] T. P. Sotiriou, Gravity and scalar fields, in *Modifications of Einstein's Theory of Gravity at Large Distances*, edited by E. Papantonopoulos (Springer International Publishing, 2015) pp. 3–24.

[31] H. Ruegg and M. Ruiz-Altaba, The Stueckelberg field, Int. J. Mod. Phys. A **19**, 3265 (2004).

[32] K. Hinterbichler, Theoretical aspects of massive gravity, Rev. Mod. Phys. **84**, 671 (2012).

[33] C. de Rham, Massive gravity, Living Rev. Relativity **17**, 7 (2014).

[34] M. Ostrogradsky, Mémoires sur les équations différentielles, relatives au problème des isopérimètres, Mem. Acad. St. Petersbourg **6**, 385 (1850).

[35] R. P. Woodard, Avoiding dark energy with $1/R$ modifications of gravity, Lect. Notes Phys. **720**, 403 (2007).

[36] T. P. Sotiriou and V. Faraoni, $f(R)$ theories of gravity, Rev. Mod. Phys. **82**, 451 (2010).

[37] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, 1972) p. 17.

[38] C. H. Brans and R. H. Dicke, Mach's principle and a relativistic theory of gravitation, Phys. Rev. **124**, 925 (1961).

[39] C. H. Brans, Mach's principle and a relativistic theory of gravitation. II, Phys. Rev. **125**, 2194 (1962).

[40] J. N. Wilford (1986, January 8). Hints of 5th force in Universe challenge Galileo's findings. *The New York Times*, p. 1.

[41] E. Fischbach, D. Sudarsky, A. Szafer, C. Talmadge, and S. H. Aronson, Reanalysis of the Eötvös experiment, Phys. Rev. Lett. **56**, 3 (1986); Erratum ibid **56**, 1427 (1986).

[42] R. V. Eötvös, D. Pekár, and E. Fekete, Beiträge zum gesetze der proportionalität von trāgheit und gravität, Ann. Phys. (Leipzig) **68**, 11 (1922).

[43] S. H. Aronson, G. J. Bock, H.-Y. Cheng, and E. Fischbach, Energy dependence of the fundamental parameters of the K^0–$\overline{K}^0$ system. I. Experimental analysis, Phys. Rev. D **28**, 476 (1983).

[44] S. H. Aronson, G. J. Bock, H.-Y. Cheng, and E. Fischbach, Energy dependence of the fundamental parameters of the K^0–$\overline{K}^0$ system. II. Theoretical formalism, Phys. Rev. D **28**, 495 (1983).

[45] F. D. Stacey and G. J. Tuck, Geophysical evidence for non-Newtonian gravity, Nature **292**, 230 (1981).

[46] S. C. Holding and G. J. Tuck, A new mine determination of the Newtonian gravitational constant, Nature **307**, 714 (1984).

[47] E. Fischbach, G. T. Gillies, D. E. Krause, J. G. Schwan, and C. Talmadge, Non-Newtonian gravity and new weak forces: an index of measurements and theory, Metrologia **29**, 213 (1992).

[48] E. Fischbach and C. Talmadge, *The Search for Non-Newtonian Gravity* (Springer-Verlag New York, 1998).

[49] E. Fischbach and C. Talmadge, Ten years of the fifth force, in *Dark Matter in Cosmology, Quantum Measurements, Experimental Gravitation*, Proceedings of the 31st Rencontres de Moriond (16th Moriond Workshop), edited by R. Ansari, U. Giraud-Héraud, and J. Trân Thanh Vân (Editions Frontiéres, 1996) pp. 443–451.

[50] E. Fischbach, The fifth force: A personal history, Eur. Phys. J. H **40**, 385 (2015).

[51] A. Franklin and E. Fischbach, *The Rise and Fall of the Fifth Force* (Springer International Publishing, 2016).

[52] L. Perivolaropoulos, PPN parameter γ and solar system constraints of massive Brans-Dicke theories, Phys. Rev. D **81**, 047501 (2010).

[53] J. Alsing, E. Berti, C. M. Will, and H. Zaglauer, Gravitational radiation from compact binary systems in the massive Brans-Dicke theory of gravity, Phys. Rev. D **85**, 064041 (2012).

[54] F. Hofmann and J. Müller, Relativistic tests with lunar laser ranging, Class. Quantum Gravity **35**, 035015 (2018).

[55] B. Bertotti, L. Iess, and P. Tortora, A test of general relativity using radio links with the Cassini spacecraft, Nature **425**, 374 (2003).

[56] E. G. Adelberger, J. H. Gundlach, B. R. Heckel, S. Hoedl, and S. Schlamminger, Torsion balance experiments: A low-energy frontier of particle physics, Prog. Part. Nucl. Phys. **62**, 102 (2009).

[57] C. M. Will, The confrontation between general relativity and experiment, Living Rev. Relativity **17**, 4 (2014).

[58] T. Damour, Experimental tests of gravitational theory, in M. Tanabashi et al. (Particle Data Group), Review of Particle Physics, Phys. Rev. D **98**, 030001 (2018) and 2019 update.

[59] M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory, Vol. 1: Introduction* (Cambridge University Press, 1987).

[60] M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory, Vol. 2: Loop Amplitudes, Anomalies and Phenomenology* (Cambridge University Press, 1987).

[61] K. Becker, M. Becker, and J. H. Schwarz, *String Theory and M-Theory* (Cambridge University Press, 2006).

[62] A. Joyce, B. Jain, J. Khoury, and M. Trodden, Beyond the cosmological standard model, Phys. Rep. **568**, 1 (2015).

[63] K. Koyama, Cosmological tests of modified gravity, Rept. Prog. Phys. **79**, 046902 (2016).

[64] A. Joyce, L. Lombriser, and F. Schmidt, Dark energy versus modified gravity, Annu. Rev. Nucl. Part. Sci. **66**, 95 (2016).

[65] C. Burrage and J. Sakstein, Tests of chameleon gravity, Living Rev. Relativity **21**, 1 (2018).

[66] T. Baker, A. Barreira, H. Desmond, P. G. Ferreira, B. Jain, K. Koyama, B. Li, L. Lombriser, A. Nicola, J. Sakstein, and F. Schmidt, The Novel Probes Project — Tests of gravity on astrophysical scales, arXiv:1908.03430.

[67] B. Li and K. Koyama, *Modified Gravity* (World Scientific, 2019).

[68] E. Berti et al., Testing general relativity with present and future astrophysical observations, Class. Quantum Gravity **32**, 243001 (2015).

[69] J. Khoury and A. Weltman, Chameleon fields: Awaiting surprises for tests of gravity in space, Phys. Rev. Lett. **93**, 171104 (2004).

[70] J. Khoury and A. Weltman, Chameleon cosmology, Phys. Rev. D **69**, 044026 (2004).

[71] K. Hinterbichler and J. Khoury, Symmetron fields: Screening long-range forces through local symmetry restoration, Phys. Rev. Lett. **104**, 231301 (2010).

[72] K. Hinterbichler, J. Khoury, A. Levy, and A. Matas, Symmetron cosmology, Phys. Rev. D **84**, 103521 (2011).

[73] P. Brax, C. van de Bruck, A.-C. Davis, and D. Shaw, The dilaton and modified gravity, Phys. Rev. D **82**, 063519 (2010).

[74] B. Jain and J. VanderPlas, Tests of modified gravity with dwarf galaxies, J. Cosmol. Astropart. Phys. 2011 (2011) 032.

[75] J. Sakstein, *Astrophysical Tests of Modified Gravity*, Ph.D. thesis, University of Cambridge (2014).

[76] T. Tamaki and S. Tsujikawa, Revisiting chameleon gravity: Thin-shell and no-shell fields with appropriate boundary conditions, Phys. Rev. D **78**, 084028 (2008).

[77] C. Burrage, E. J. Copeland, and J. Stevenson, Ellipticity weakens chameleon screening, Phys. Rev. D **91**, 065030 (2015).

[78] L. Hui, A. Nicolis, and C. Stubbs, Equivalence principle implications of modified gravity models, Phys. Rev. D **80**, 104002 (2009).

[79] V. Vikram, J. Sakstein, C. Davis, and A. Neil, Astrophysical tests of modified gravity: Stellar and gaseous rotation curves in dwarf galaxies, Phys. Rev. D **97**, 104055 (2018).

[80] H. Desmond, P. G. Ferreira, G. Lavaux, and J. Jasche, Fifth force constraints from galaxy warps, Phys. Rev. D **98**, 083010 (2018).

[81] H. Desmond, P. G. Ferreira, G. Lavaux, and J. Jasche, Fifth force constraints from the separation of galaxy mass components, Phys. Rev. D **98**, 064015 (2018).

[82] W. A. Hellwing, A. Barreira, C. S. Frenk, B. Li, and S. Cole, A clear and measurable signature of modified gravity in the galaxy velocity field, Phys. Rev. Lett. **112**, 221102 (2014).

[83] C. Burrage, E. J. Copeland, and P. Millington, Radial acceleration relation from symmetron fifth forces, Phys. Rev. D **95**, 064050 (2017); Erratum ibid **95**, 129902 (2017).

[84] C. A. J. O'Hare and C. Burrage, Stellar kinematics from the symmetron fifth force in the Milky Way disk, Phys. Rev. D **98**, 064019 (2018).

[85] A. P. Naik, E. Puchwein, A.-C. Davis, and C. Arnold, Imprints of chameleon $f(R)$ gravity on galaxy rotation curves, Mon. Not. R. Astron. Soc. **480**, 5211 (2018).

[86] A. P. Naik, E. Puchwein, A.-C. Davis, D. Sijacki, and H. Desmond, Constraints on chameleon $f(R)$-gravity from galaxy rotation curves of the SPARC sample, Mon. Not. R. Astron. Soc. **489**, 771 (2019).

[87] A. Naik, N. Evans, E. Puchwein, H. Zhao, and A.-C. Davis, Stellar streams in chameleon gravity, arXiv:2002.05738.

[88] J. Sakstein, Stellar oscillations in modified gravity, Phys. Rev. D **88**, 124013 (2013).

[89] R. L. Beaton, G. Bono, V. F. Braga, M. Dall'Ora, G. Fiorentino, I. S. Jang, C. E. Martínez-Vázquez, N. Matsunaga, M. Monelli, J. R. Neeley, and M. Salaris, Old-aged primary distance indicators, Space Sci. Rev. **214**, 113 (2018).

[90] B. Jain, V. Vikram, and J. Sakstein, Astrophysical tests of modified gravity: Constraints from distance indicators in the nearby Universe, Astrophys. J. **779**, 39 (2013).

[91] A. Upadhye, Dark energy fifth forces in torsion pendulum experiments, Phys. Rev. D **86**, 102003 (2012).

[92] A. Upadhye, Symmetron dark energy in laboratory experiments, Phys. Rev. Lett. **110**, 031301 (2013).

[93] P. Brax, C. van de Bruck, A.-C. Davis, D. F. Mota, and D. J. Shaw, Detecting chameleons through Casimir force measurements, Phys. Rev. D **76**, 124034 (2007).

[94] P. Brax and A.-C. Davis, Casimir, gravitational and neutron tests of dark energy, Phys. Rev. D **91**, 063503 (2015).

[95] B. Elder, V. Vardanyan, Y. Akrami, P. Brax, A.-C. Davis, and R. S. Decca, Classical symmetron force in Casimir experiments, Phys. Rev. D **101**, 064065 (2020).

[96] P. Hamilton, M. Jaffe, P. Haslinger, Q. Simmons, H. Müller, and J. Khoury, Atom-interferometry constraints on dark energy, Science **349**, 849 (2015).

[97] C. Burrage, A. Kuribayashi-Coleman, J. Stevenson, and B. Thrussell, Constraining symmetron fields with atom interferometry, J. Cosmol. Astropart. Phys. 12 (2016) 041.

[98] M. Jaffe, P. Haslinger, V. Xu, P. Hamilton, A. Upadhye, B. Elder, J. Khoury, and H. Müller, Testing sub-gravitational forces on atoms from a miniature, in-vacuum source mass, Nat. Phys. **13**, 938 (2017).

[99] D. O. Sabulsky, I. Dutta, E. Hinds, B. Elder, C. Burrage, and E. J. Copeland, Experiment to detect dark energy forces using atom interferometry, Phys. Rev. Lett. **123**, 061102 (2019).

[100] P. Brax and G. Pignol, Strongly coupled chameleons and the neutronic quantum bouncer, Phys. Rev. Lett. **107**, 111301 (2011).

[101] T. Jenke, G. Cronenberg, J. Burgdörfer, L. A. Chizhova, P. Geltenbort, A. N. Ivanov, T. Lauer, T. Lins, S. Rotter, H. Saul, U. Schmidt, and H. Abele, Gravity resonance spectroscopy constrains dark energy and dark matter scenarios, Phys. Rev. Lett. **112**, 151105 (2014).

[102] G. Cronenberg, H. Filter, M. Thalhammer, T. Jenke, H. Abele, and P. Geltenbort, A gravity of Earth measurement with a qBOUNCE Experiment, in *The European Physical Society Conference on High Energy Physics*, Vol. 234 (Proceedings of Science, 2015) p. 408.

[103] G. Cronenberg, P. Brax, H. Filter, P. Geltenbort, T. Jenke, G. Pignol, M. Pitschmann, M. Thalhammer, and H. Abele, Acoustic Rabi oscillations between gravitational quantum states and impact on symmetron dark energy, Nat. Phys. **14**, 1022 (2018).

[104] A. I. Vainshtein, To the problem of nonvanishing gravitation mass, Phys. Lett. B **39**, 393 (1972).

[105] H. van Dam and M. Veltman, Massive and mass-less Yang-Mills and gravitational fields, Nucl. Phys. B **22**, 397 (1970).

[106] V. I. Zakharov, Linearized gravitation theory and the graviton mass, JETP Lett. **12**, 312 (1970).

[107] Y. Iwasaki, Consistency condition for propagators, Phys. Rev. D **2**, 2255 (1970).

[108] G. R. Dvali, G. Gabadadze, and M. Porrati, 4D gravity on a brane in 5D Minkowski space, Phys. Lett. B **485**, 208 (2000).

[109] A. Nicolis, R. Rattazzi, and E. Trincherini, The galileon as a local modification of gravity, Phys. Rev. D **79**, 064036 (2009).

[110] C. Deffayet, G. Esposito-Farèse, and A. Vikman, Covariant galileon, Phys. Rev. D **79**, 084003 (2009).

[111] C. Deffayet, S. Deser, and G. Esposito-Farèse, Generalized galileons: All scalar models whose curved background extensions maintain second-order field equations and stress-tensors, Phys. Rev. D **80**, 064015 (2009).

[112] G. W. Horndeski, Second-order scalar-tensor field equations in a four-dimensional space, Int. J. Theor. Phys. **10**, 363 (1974).

[113] E. Babichev, C. Deffayet, and R. Ziour, k-Mouflage gravity, Int. J. Mod. Phys. D **18**, 2147 (2009).

[114] C. Deffayet, X. Gao, D. A. Steer, and G. Zahariade, From k-essence to generalised galileons, Phys. Rev. D **84**, 064039 (2011).

[115] C. Deffayet and D. A. Steer, A formal introduction to Horndeski and galileon theories and their generalizations, Class. Quantum Gravity **30**, 214006 (2013).

[116] E. Babichev and C. Deffayet, An introduction to the Vainshtein mechanism, Class. Quantum Gravity **30**, 184001 (2013).

[117] B. P. Abbott et al. (LIGO Scientific and Virgo Collaborations), GW170817: Observation of gravitational waves from a binary neutron star inspiral, Phys. Rev. Lett. **119**, 161101 (2017).

[118] A. Goldstein et al., An ordinary short gamma-ray burst with extraordinary implications: *Fermi*-GBM detection of GRB 170817A, Astrophys. J. **848**, L14 (2017).

[119] B. P. Abbott et al. (Virgo, Fermi-GBM, INTEGRAL, and LIGO Scientific Collaborations), Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A, Astrophys. J. **848**, L13 (2017).

[120] P. Creminelli and F. Vernizzi, Dark energy after GW170817 and GRB170817A, Phys. Rev. Lett. **119**, 251302 (2017).

[121] J. M. Ezquiaga and M. Zumalacárregui, Dark energy after GW170817: Dead ends and the road ahead, Phys. Rev. Lett. **119**, 251304 (2017).

[122] J. Sakstein and B. Jain, Implications of the neutron star merger GW170817 for cosmological scalar-tensor theories, Phys. Rev. Lett. **119**, 251303 (2017).

[123] T. Baker, E. Bellini, P. G. Ferreira, M. Lagos, J. Noller, and I. Sawicki, Strong constraints on cosmological gravity from GW170817 and GRB 170817A, Phys. Rev. Lett. **119**, 251301 (2017).

[124] L. Heisenberg, A systematic approach to generalisations of general relativity and their cosmological implications, Phys. Rep. **796**, 1 (2019).

[125] J. M. Ezquiaga and M. Zumalacárregui, Dark energy in light of multi-messenger gravitational-wave astronomy, Front. Astron. Space Sci. **5**, 44 (2018).

[126] R. Kase and S. Tsujikawa, Dark energy in Horndeski theories after GW170817: A review, Int. J. Mod. Phys. D **28**, 1942005 (2019).

[127] C. de Rham and S. Melville, Gravitational rainbows: LIGO and dark energy at its cutoff, Phys. Rev. Lett. **121**, 221101 (2018).

[128] P. Brax, C. van de Bruck, A.-C. Davis, J. Khoury, and A. Weltman, Detecting dark energy in orbit: The cosmological chameleon, Phys. Rev. D **70**, 123518 (2004).

[129] J. Wang, L. Hui, and J. Khoury, No-go theorems for generalized chameleon field theories, Phys. Rev. Lett. **109**, 241301 (2012).

[130] P. Brax, A.-C. Davis, and J. Sakstein, Pulsar constraints on screened modified gravity, Class. Quantum Gravity **31**, 225001 (2014).

[131] X. Zhang, T. Liu, and W. Zhao, Gravitational radiation from compact binary systems in screened modified gravity, Phys. Rev. D **95**, 104027 (2017).

[132] T. Liu, X. Zhang, W. Zhao, K. Lin, C. Zhang, S. Zhang, X. Zhao, T. Zhu, and A. Wang, Waveforms of compact binary inspiral gravitational radiation in screened modified gravity, Phys. Rev. D **98**, 083023 (2018).

[133] C. de Rham, A. J. Tolley, and D. H. Wesley, Vainshtein mechanism in binary pulsars, Phys. Rev. D **87**, 044025 (2013).

[134] C. de Rham, A. Matas, and A. J. Tolley, Galileon radiation from binary systems, Phys. Rev. D **87**, 064024 (2013).

[135] F. Dar, C. De Rham, J. T. Deskins, J. T. Giblin, and A. J. Tolley, Scalar gravitational radiation from binaries: Vainshtein mechanism in time-dependent systems, Class. Quantum Gravity **36**, 025008 (2019).

[136] T. Damour and G. Esposito-Farèse, Nonperturbative strong field effects in tensor–scalar theories of gravitation, Phys. Rev. Lett. **70**, 2220 (1993).

[137] T. Damour and G. Esposito-Farèse, Tensor–scalar gravity and binary pulsar experiments, Phys. Rev. D **54**, 1474 (1996).

[138] T. Damour and G. Esposito-Farèse, Gravitational-wave versus binary-pulsar tests of strong-field gravity, Phys. Rev. D **58**, 042001 (1998).

[139] T. A. de Pirey Saint Alby and N. Yunes, Cosmological evolution and Solar System consistency of massive scalar-tensor gravity, Phys. Rev. D **96**, 064040 (2017).

[140] D. Anderson, N. Yunes, and E. Barausse, Effect of cosmological evolution on Solar System constraints and on the scalarization of neutron stars in massless scalar-tensor theories, Phys. Rev. D **94**, 104064 (2016).

[141] T. Anson, E. Babichev, C. Charmousis, and S. Ramazanov, Cosmological instability of scalar-Gauss-Bonnet theories exhibiting scalarization, J. Cosmol. Astropart. Phys. 06 (2019) 023.

[142] N. Franchini and T. P. Sotiriou, Cosmology with subdominant Horndeski scalar field, Phys. Rev. D **101**, 064068 (2020).

[143] T. Anson, E. Babichev, and S. Ramazanov, Reconciling spontaneous scalarization with cosmology, Phys. Rev. D **100**, 104051 (2019).

[144] G. Antoniou, L. Bordin, and T. P. Sotiriou, Compact object scalarization with general relativity as a cosmic attractor, arXiv:2004.14985.

[145] E. Barausse, C. Palenzuela, M. Ponce, and L. Lehner, Neutron-star mergers in scalar-tensor theories of gravity, Phys. Rev. D **87**, 081506 (2013).

[146] C. Palenzuela, E. Barausse, M. Ponce, and L. Lehner, Dynamical scalarization of neutron stars in scalar-tensor gravity theories, Phys. Rev. D **89**, 044024 (2014).

[147] M. Shibata, K. Taniguchi, H. Okawa, and A. Buonanno, Coalescence of binary neutron stars in a scalar-tensor theory of gravity, Phys. Rev. D **89**, 084005 (2014).

[148] K. Taniguchi, M. Shibata, and A. Buonanno, Quasiequilibrium sequences of binary neutron stars undergoing dynamical scalarization, Phys. Rev. D **91**, 024033 (2015).

[149] N. Sennett and A. Buonanno, Modeling dynamical scalarization with a resummed post-Newtonian expansion, Phys. Rev. D **93**, 124004 (2016).

[150] N. Sennett, L. Shao, and J. Steinhoff, Effective action model of dynamically scalarizing binary neutron stars, Phys. Rev. D **96**, 084019 (2017).

[151] M. Khalil, N. Sennett, J. Steinhoff, and A. Buonanno, Theory-agnostic framework for dynamical scalarization of compact binaries, Phys. Rev. D **100**, 124013 (2019).

[152] T. Damour and G. Esposito-Farèse, Tensor–multi–scalar theories of gravitation, Class. Quantum Gravity **9**, 2093 (1992).

[153] P. C. C. Freire, N. Wex, G. Esposito-Farèse, J. P. W. Verbiest, M. Bailes, B. A. Jacoby, M. Kramer, I. H. Stairs, J. Antoniadis, and G. H. Janssen, The relativistic pulsar–white dwarf binary PSR J1738+0333 – II. The most stringent test of scalar–tensor gravity, Mon. Not. R. Astron. Soc. **423**, 3328 (2012).

[154] P. Kanti, N. Mavromatos, J. Rizos, K. Tamvakis, and E. Winstanley, Dilatonic black holes in higher curvature string gravity, Phys. Rev. D **54**, 5049 (1996).

[155] T. Torii, H. Yajima, and K. Maeda, Dilatonic black holes with Gauss-Bonnet term, Phys. Rev. D **55**, 739 (1997).

[156] S. Alexeev and M. Pomazanov, Black hole solutions with dilatonic hair in higher curvature gravity, Phys. Rev. D **55**, 2110 (1997).

[157] P. Pani and V. Cardoso, Are black holes in alternative theories serious astrophysical candidates? The case for Einstein-Dilaton-Gauss-Bonnet black holes, Phys. Rev. D **79**, 084031 (2009).

[158] T. P. Sotiriou and S.-Y. Zhou, Black hole hair in generalized scalar-tensor gravity, Phys. Rev. Lett. **112**, 251102 (2014).

[159] T. P. Sotiriou and S.-Y. Zhou, Black hole hair in generalized scalar-tensor gravity: An explicit example, Phys. Rev. D **90**, 124063 (2014).

[160] H. O. Silva, J. Sakstein, L. Gualtieri, T. P. Sotiriou, and E. Berti, Spontaneous scalarization of black holes and compact stars from a Gauss-Bonnet coupling, Phys. Rev. Lett. **120**, 131104 (2018).

[161] D. D. Doneva and S. S. Yazadjiev, New Gauss-Bonnet black holes with curvature-induced scalarization in extended scalar-tensor theories, Phys. Rev. Lett. **120**, 131103 (2018).

[162] G. Antoniou, A. Bakopoulos, and P. Kanti, Evasion of no-hair theorems and novel black-hole solutions in Gauss-Bonnet theories, Phys. Rev. Lett. **120**, 131102 (2018).

[163] D. C. Robinson, Four decades of black hole uniqueness theorems, in *The Kerr Spacetime: Rotating Black Holes in General Relativity*, edited by D. L. Wiltshire, M. Visser, and S. M. Scott (Cambridge University Press, 2009) pp. 115–143.

[164] P. T. Chruściel, J. L. Costa, and M. Heusler, Stationary black holes: Uniqueness and beyond, Living Rev. Relativity **15**, 7 (2012).

[165] R. P. Kerr, Gravitational field of a spinning mass as an example of algebraically special metrics, Phys. Rev. Lett. **11**, 237 (1963).

[166] E. T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence, Metric of a rotating, charged mass, J. Math. Phys. **6**, 918 (1965).

[167] S. D. Majumdar, A class of exact solutions of Einstein's field equations, Phys. Rev. **72**, 390 (1947).

[168] A. Papapetrou, A static solution of the equations of the gravitational field for an arbitrary charge-distribution, Proc. R. Ir. Acad. A **51**, 191 (1947).

[169] J. B. Hartle and S. W. Hawking, Solutions of the Einstein-Maxwell equations with many black holes, Commun. Math. Phys. **26**, 87 (1972).

[170] D. M. Eardley and W. H. Press, Astrophysical processes near black holes., Annu. Rev. Astron. Astrophys. **13**, 381 (1975).

[171] G. Gibbons, Vacuum polarization and the spontaneous loss of charge by black holes, Commun. Math. Phys. **44**, 245 (1975).

[172] R. Blandford and R. Znajek, Electromagnetic extractions of energy from Kerr black holes, Mon. Not. R. Astron. Soc. **179**, 433 (1977).

[173] R. Ruffini, G. Vereshchagin, and S.-S. Xue, Electron-positron pairs in physics and astrophysics: From heavy nuclei to black holes, Phys. Rep. **487**, 1 (2010).

[174] K. S. Thorne and J. J. Dykla, Black holes in the Dicke-Brans-Jordan theory of gravity, Astrophys. J. **166**, L35 (1971).

[175] J. D. Bekenstein, Nonexistence of Baryon number for static black holes, Phys. Rev. D **5**, 1239 (1972).

[176] S. L. Adler and R. B. Pearson, "No-hair" theorems for the Abelian Higgs and Goldstone models, Phys. Rev. D **18**, 2798 (1978).

[177] S. W. Hawking, Black holes in the Brans-Dicke theory of gravitation, Commun. Math. Phys. **25**, 167 (1972).

[178] T. Zannias, Black holes cannot support conformal scalar hair, J. Math. Phys. **36**, 6970 (1995).

[179] J. D. Bekenstein, Novel "no-scalar-hair" theorem for black holes, Phys. Rev. D **51**, R6608 (1995).

[180] A. Saa, New no-scalar-hair theorem for black holes, J. Math. Phys. **37**, 2346 (1996).

[181] T. P. Sotiriou and V. Faraoni, Black holes in scalar-tensor gravity, Phys. Rev. Lett. **108**, 081103 (2012).

[182] L. Hui and A. Nicolis, No-hair theorem for the galileon, Phys. Rev. Lett. **110**, 241104 (2013).

[183] A. A. H. Graham and R. Jha, Nonexistence of black holes with noncanonical scalar fields, Phys. Rev. D **89**, 084056 (2014); Erratum ibid **92**, 069901 (2015).

[184] S. Alexander and N. Yunes, Chern-Simons modified general relativity, Phys. Rep. **480**, 1 (2009).

[185] M. S. Volkov, Hairy black holes in the XX-th and XXI-st centuries, in *Proceedings of the Fourteenth Marcel Grossmann Meeting on General Relativity*, edited by M. Bianchi, R. T. Jantzen, and R. Ruffini (World Scientific, 2017) pp. 1779–1798.

[186] K. Yagi, L. C. Stein, N. Yunes, and T. Tanaka, Post-Newtonian, quasicircular binary inspirals in quadratic modified gravity, Phys. Rev. D **85**, 064022 (2012); Erratum ibid **93**, 029902 (2016).

[187] H. Witek, L. Gualtieri, P. Pani, and T. P. Sotiriou, Black holes and binary mergers in scalar Gauss-Bonnet gravity: Scalar field dynamics, Phys. Rev. D **99**, 064035 (2019).

[188] F.-L. Julié and E. Berti, Post-Newtonian dynamics and black hole thermodynamics in Einstein-scalar-Gauss-Bonnet gravity, Phys. Rev. D **100**, 104061 (2019).

[189] K. Yagi, A new constraint on scalar Gauss-Bonnet gravity and a possible explanation for the excess of the orbital decay rate in a low-mass X-ray binary, Phys. Rev. D **86**, 081504 (2012).

[190] N. Yunes, K. Yagi, and F. Pretorius, Theoretical physics implications of the binary black-hole mergers GW150914 and GW151226, Phys. Rev. D **94**, 084002 (2016).

[191] S. Tahura, K. Yagi, and Z. Carson, Testing gravity with gravitational waves from binary black hole mergers: Contributions from amplitude corrections, Phys. Rev. D **100**, 104001 (2019).

[192] R. Nair, S. Perkins, H. O. Silva, and N. Yunes, Fundamental physics implications for higher-curvature theories from binary black hole signals in the LIGO-Virgo catalog GWTC-1, Phys. Rev. Lett. **123**, 191101 (2019); Erratum ibid **124**, 169904 (2020).

[193] R. Peccei and H. R. Quinn, CP conservation in the presence of pseudoparticles, Phys. Rev. Lett. **38**, 1440 (1977).

[194] F. Wilczek, Problem of strong P and T invariance in the presence of instantons, Phys. Rev. Lett. **40**, 279 (1978).

[195] S. Weinberg, A new light boson?, Phys. Rev. Lett. **40**, 223 (1978).

[196] J. E. Kim, Weak interaction singlet and strong CP invariance, Phys. Rev. Lett. **43**, 103 (1979).

[197] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Can confinement ensure natural CP invariance of strong interactions?, Nucl. Phys. B **166**, 493 (1980).

[198] M. Dine, W. Fischler, and M. Srednicki, A simple solution to the strong CP problem with a harmless axion, Phys. Lett. B **104**, 199 (1981).

[199] A. R. Zhitnitsky, On possible suppression of the axion hadron interactions, Yad. Fiz. **31**, 497 (1980) [Sov. J. Nucl. Phys. **31**, 260 (1980)].

[200] E. Witten, Some properties of O(32) superstrings, Phys. Lett. B **149**, 351 (1984).

[201] P. Svrcek and E. Witten, Axions in string theory, J. High Energy Phys. 06 (2006) 051.

[202] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, String axiverse, Phys. Rev. D **81**, 123530 (2010).

[203] J. Halverson, C. Long, and P. Nath, Ultralight axion in supersymmetry and strings and cosmology at small scales, Phys. Rev. D **96**, 056025 (2017).

[204] T. C. Bachlechner, K. Eckerle, O. Janssen, and M. Kleban, Axion landscape cosmology, J. Cosmol. Astropart. Phys. 09 (2019) 062.

[205] J. E. Moody and F. Wilczek, New macroscopic forces?, Phys. Rev. D **30**, 130 (1984).

[206] G. Raffelt, Limits on a CP-violating scalar axion-nucleon interaction, Phys. Rev. D **86**, 015001 (2012).

[207] M. Kuster, G. Raffelt, and B. Beltrán, Axions, Lect. Notes Phys. **741**, 1 (2008).

[208] J. E. Kim and G. Carosi, Axions and the strong CP problem, Rev. Mod. Phys. **82**, 557 (2010); Erratum ibid **91**, 049902 (2019).

[209] S. Asztalos, Axion searches, in *Particle Dark Matter: Observations, Models and Searches*, edited by G. Bertone (Cambridge University Press, 2010) pp. 467–488.

[210] J. Jaeckel and A. Ringwald, The low-energy frontier of particle physics, Annu. Rev. Nucl. Part. Sci. **60**, 405 (2010).

[211] A. Ringwald, Exploring the role of axions and other WISPs in the dark Universe, Phys. Dark Univ. **1**, 116 (2012).

[212] A. Ringwald, Searching for axions and ALPs from string theory, J. Phys. Conf. Ser. **485**, 012013 (2014).

[213] E. Pajer and M. Peloso, A review of axion inflation in the era of Planck, Class. Quantum Gravity **30**, 214002 (2013).

[214] M. Kawasaki and K. Nakayama, Axions: Theory and cosmological role, Annu. Rev. Nucl. Part. Sci. **63**, 69 (2013).

[215] D. J. E. Marsh, Axion cosmology, Phys. Rep. **643**, 1 (2016).

[216] P. W. Graham, I. G. Irastorza, S. K. Lamoreaux, A. Lindner, and K. A. van Bibber, Experimental searches for the axion and axion-like particles, Annu. Rev. Nucl. Part. Sci. **65**, 485 (2015).

[217] I. G. Irastorza and J. Redondo, New experimental approaches in the search for axion-like particles, Prog. Part. Nucl. Phys. **102**, 89 (2018).

[218] A. Ringwald, L. J. Rosenberg, and G. Rybka, Axions and other similar particles, in M. Tanabashi et al. (Particle Data Group), Review of Particle Physics, Phys. Rev. D **98**, 030001 (2018) and 2019 update.

[219] P. Sikivie, Experimental tests of the "invisible" axion, Phys. Rev. Lett. **51**, 1415 (1983).

[220] G. Grilli di Cortona, E. Hardy, J. Pardo Vega, and G. Villadoro, The QCD axion, precisely, J. High Energy Phys. 01 (2016) 034.

[221] A. Anselm, Axion $\leftrightarrow$ photon oscillations in a steady magnetic field, Yad. Fiz. **42**, 1480 (1985).

[222] K. Van Bibber, N. R. Dagdeviren, S. E. Koonin, A. K. Kerman, and H. N. Nelson, Proposed experiment to produce and detect light pseudoscalars, Phys. Rev. Lett. **59**, 759 (1987).

[223] G. Ruoso, R. Cameron, G. Cantatore, A. C. Melissinos, Y. Semertzidis, H. J. Halama, D. M. Lazarus, A. G. Prodell, F. Nezrick, C. Rizzo, and E. Zavattini, Search for photon regeneration in a magnetic field, Z. Phys. C **56**, 505 (1992).

[224] R. Cameron, G. Cantatore, A. C. Melissinos, G. Ruoso, Y. Semertzidis, H. J. Halama, D. M. Lazarus, A. G. Prodell, F. Nezrick, C. Rizzo, and E. Zavattini, Search for nearly massless, weakly coupled particles by optical techniques, Phys. Rev. D **47**, 3707 (1993).

[225] M. Fouché, C. Robilliard, S. Faure, C. Rizzo, J. Mauchain, M. Nardone, R. Battesti, L. Martin, A.-M. Sautivet, J.-L. Paillard, and F. Amiranoff, Search for photon oscillations into massive particles, Phys. Rev. D **78**, 032013 (2008).

[226] A. Afanasev, O. Baker, K. Beard, G. Biallas, J. Boyce, M. Minarni, R. Ramdon, M. Shinn, and P. Slocum, Experimental limit on optical-photon coupling to light neutral scalar bosons, Phys. Rev. Lett. **101**, 120401 (2008).

[227] K. Ehret et al. (ALPS Collaboration), New ALPS results on hidden-sector lightweights, Phys. Lett. B **689**, 149 (2010).

[228] A. S. Chou et al. (GammeV Collaboration), Search for axion-like particles using a variable baseline photon regeneration technique, Phys. Rev. Lett. **100**, 080402 (2008).

[229] R. Ballou et al. (OSQAR Collaboration), New exclusion limits on scalar and pseudoscalar axionlike particles from light shining through a wall, Phys. Rev. D **92**, 092002 (2015).

[230] J. Redondo and A. Ringwald, Light shining through walls, Contemp. Phys. **52**, 211 (2011).

[231] K. van Bibber, P. M. McIntyre, D. E. Morris, and G. G. Raffelt, Design for a practical laboratory detector for solar axions, Phys. Rev. D **39**, 2089 (1989).

[232] D. M. Lazarus, G. C. Smith, R. Cameron, A. C. Melissinos, G. Ruoso, Y. K. Semertzidis, and F. A. Nezrick, Search for solar axions, Phys. Rev. Lett. **69**, 2333 (1992).

[233] S. Moriyama, M. Minowa, T. Namba, Y. Inoue, Y. Takasu, and A. Yamamoto, Direct search for solar axions by using strong magnetic field and X-ray detectors, Phys. Lett. B **434**, 147 (1998).

[234] Y. Inoue, T. Namba, S. Moriyama, M. Minowa, Y. Takasu, T. Horiuchi, and A. Yamamoto, Search for sub-electronvolt solar axions using coherent conversion of axions into photons in magnetic field and gas helium, Phys. Lett. B **536**, 18 (2002).

[235] Y. Inoue, Y. Akimoto, R. Ohta, T. Mizumoto, A. Yamamoto, and M. Minowa, Search for solar axions with mass around 1 eV using coherent conversion of axions into photons, Phys. Lett. B **668**, 93 (2008).

[236] M. Arik et al. (CAST Collaboration), New solar axion search using the CERN Axion Solar Telescope with ^{4}He filling, Phys. Rev. D **92**, 021101 (2015).

[237] V. Anastassopoulos et al. (CAST Collaboration), New CAST limit on the axion–photon interaction, Nat. Phys. **13**, 584 (2017).

[238] P. Sikivie, Detection rates for "invisible"-axion searches, Phys. Rev. D **32**, 2988 (1985); Erratum ibid **36**, 974 (1987).

[239] S. DePanfilis, A. C. Melissinos, B. E. Moskowitz, J. T. Rogers, Y. K. Semertzidis, W. U. Wuensch, H. J. Halama, A. G. Prodell, W. B. Fowler, and F. A. Nezrick, Limits on the abundance and coupling of cosmic axions at $4.5 < m_a < 5.0$ μeV, Phys. Rev. Lett. **59**, 839 (1987).

[240] W. U. Wuensch, S. De Panfilis-Wuensch, Y. K. Semertzidis, J. T. Rogers, A. C. Melissinos, H. J. Halama, B. E. Moskowitz, A. G. Prodell, W. B. Fowler, and F. A. Nezrick, Results of a laboratory search for cosmic axions and other weakly coupled light particles, Phys. Rev. D **40**, 3153 (1989).

[241] C. Hagmann, P. Sikivie, N. S. Sullivan, and D. B. Tanner, Results from a search for cosmic axions, Phys. Rev. D **42**, 1297 (1990).

[242] R. Bradley, J. Clarke, D. Kinion, L. Rosenberg, K. van Bibber, S. Matsuki, M. Muck, and P. Sikivie, Microwave cavity searches for dark-matter axions, Rev. Mod. Phys. **75**, 777 (2003).

[243] T. Braine et al. (ADMX Collaboration), Extended search for the invisible axion with the Axion Dark Matter Experiment, Phys. Rev. Lett. **124**, 101303 (2020).

[244] P. Brax, C. Burrage, A.-C. Davis, D. Seery, and A. Weltman, Anomalous coupling of scalars to gauge fields, Phys. Lett. B **699**, 5 (2011).

[245] P. Brax, C. Burrage, A.-C. Davis, D. Seery, and A. Weltman, Higgs production as a probe of chameleon dark energy, Phys. Rev. D **81**, 103524 (2010).

[246] P. Brax and K. Zioutas, Solar chameleons, Phys. Rev. D **82**, 043007 (2010).

[247] V. Anastassopoulos et al. (CAST Collaboration), Search for chameleons with CAST, Phys. Lett. B **749**, 172 (2015).

[248] H. Gies, D. F. Mota, and D. J. Shaw, Hidden in the light: Magnetically induced afterglow from trapped chameleon fields, Phys. Rev. D **77**, 025016 (2008).

[249] M. Ahlers, A. Lindner, A. Ringwald, L. Schrempp, and C. Weniger, Alpenglow: A signature for chameleons in axionlike particle search experiments, Phys. Rev. D **77**, 015018 (2008).

[250] A. Upadhye, J. H. Steffen, and A. Weltman, Constraining chameleon field theories using the GammeV afterglow experiments, Phys. Rev. D **81**, 015013 (2010).

[251] A. S. Chou, W. Wester, A. Baumbaugh, H. R. Gustafson, Y. Irizarry-Valle, P. O. Mazur, J. H. Steffen, R. Tomlin, A. Upadhye, A. Weltman, X. Yang, and J. Yoo, Search for chameleon particles using a photon-regeneration technique, Phys. Rev. Lett. **102**, 030402 (2009).

[252] J. H. Steffen, A. Upadhye, A. Baumbaugh, A. S. Chou, P. O. Mazur, R. Tomlin, A. Weltman, and W. Wester, Laboratory constraints on chameleon dark energy and power-law fields, Phys. Rev. Lett. **105**, 261803 (2010).

[253] A. Upadhye, J. H. Steffen, and A. S. Chou, Designing dark energy afterglow experiments, Phys. Rev. D **86**, 035006 (2012).

[254] G. Rybka, M. Hotz, L. J. Rosenberg, S. J. Asztalos, G. Carosi, C. Hagmann, D. Kinion, K. van Bibber, J. Hoskins, C. Martin, P. Sikivie, D. B. Tanner, R. Bradley, and J. Clarke, Search for chameleon scalar fields with the Axion Dark Matter Experiment, Phys. Rev. Lett. **105**, 051801 (2010).

[255] W. H. Press, B. S. Ryden, and D. N. Spergel, Single mechanism for generating large scale structure and providing dark missing matter, Phys. Rev. Lett. **64**, 1084 (1990).

[256] S. Ji and S. Sin, Late-time phase transition and the galactic halo as a Bose liquid. II. The effect of visible matter, Phys. Rev. D **50**, 3655 (1994).

[257] W. Hu, R. Barkana, and A. Gruzinov, Fuzzy cold dark matter: The wave properties of ultralight particles, Phys. Rev. Lett. **85**, 1158 (2000).

[258] L. Amendola and R. Barbieri, Dark matter from an ultra-light pseudo-Goldsone-boson, Phys. Lett. B **642**, 192 (2006).

[259] L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten, Ultralight scalars as cosmological dark matter, Phys. Rev. D **95**, 043541 (2017).

[260] L. A. Ureña-López, Brief review on scalar field dark matter models, Front. Astron. Space Sci. **6**, 47 (2019).

[261] E. G. M. Ferreira, Ultra-light dark matter, arXiv:2005.03254.

[262] J. S. Bullock and M. Boylan-Kolchin, Small-scale challenges to the ΛCDM paradigm, Annu. Rev. Astron. Astrophys. **55**, 343 (2017).

[263] F. Governato, A. Zolotov, A. Pontzen, C. Christensen, S. H. Oh, A. M. Brooks, T. Quinn, S. Shen, and J. Wadsley, Cuspy no more: How outflows affect the central dark matter and baryon distribution in λ cold dark matter galaxies, Mon. Not. R. Astron. Soc. **422**, 1231 (2012).

[264] A. Di Cintio, C. B. Brook, A. V. Macciò, G. S. Stinson, A. Knebe, A. A. Dutton, and J. Wadsley, The dependence of dark matter profiles on the stellar-to-halo mass ratio: A prediction for cusps versus cores, Mon. Not. R. Astron. Soc. **437**, 415 (2013).

[265] A. Pontzen and F. Governato, Cold dark matter heats up, Nature **506**, 171 (2014).

[266] E. Tollet, A. V. Macciò, A. A. Dutton, G. S. Stinson, L. Wang, C. Penzo, T. A. Gutcke, T. Buck, X. Kang, C. Brook, A. Di Cintio, B. W. Keller, and J. Wadsley, NIHAO – IV: Core creation and destruction in dark matter density profiles across cosmic time, Mon. Not. R. Astron. Soc. **456**, 3542 (2016).

[267] A. Fitts, M. Boylan-Kolchin, O. D. Elbert, J. S. Bullock, P. F. Hopkins, J. Oñorbe, A. Wetzel, C. Wheeler, C.-A. Faucher-Giguère, D. Kereš, E. D. Skillman, and D. R. Weisz, FIRE in the field: Simulating the threshold of galaxy formation, Mon. Not. R. Astron. Soc. **471**, 3547 (2017).

[268] A. Zolotov, A. M. Brooks, B. Willman, F. Governato, A. Pontzen, C. Christensen, A. Dekel, T. Quinn, S. Shen, and J. Wadsley, Baryons matter: Why luminous satellite galaxies have reduced central masses, Astrophys. J. **761**, 71 (2012).

[269] T. Sawala, C. S. Frenk, A. Fattahi, J. F. Navarro, R. G. Bower, R. A. Crain, C. D. Vecchia, M. Furlong, J. C. Helly, A. Jenkins, K. A. Oman, M. Schaller, J. Schaye, T. Theuns, J. Trayford, and S. D. M. White, The APOSTLE simulations: Solutions to the Local Group's cosmic puzzles, Mon. Not. R. Astron. Soc. **457**, 1931 (2016).

[270] Q. Zhu, F. Marinacci, M. Maji, Y. Li, V. Springel, and L. Hernquist, Baryonic impact on the dark matter distribution in Milky Way-sized galaxies and their satellites, Mon. Not. R. Astron. Soc. **458**, 1559 (2016).

[271] S. Y. Kim, A. H. G. Peter, and J. R. Hargis, Missing satellites problem: Completeness corrections to the number of satellite galaxies in the Milky Way are consistent with cold dark matter predictions, Phys. Rev. Lett. **121**, 211302 (2018).

[272] J. P. Ostriker, E. Choi, A. Chow, and K. Guha, Mind the gap: Is the too big to fail problem resolved?, Astrophys. J. **885**, 97 (2019).

[273] V. Lora, J. Magaña, A. Bernal, F. J. Sánchez-Salcedo, and E. K. Grebel, On the mass of ultra-light bosonic dark matter from galactic dynamics, J. Cosmol. Astropart. Phys. 02 (2012) 011.

[274] D. J. E. Marsh and A.-R. Pop, Axion dark matter, solitons and the cusp–core problem, Mon. Not. R. Astron. Soc. **451**, 2479 (2015).

[275] S.-R. Chen, H.-Y. Schive, and T. Chiueh, Jeans analysis for dwarf spheroidal galaxies in wave dark matter, Mon. Not. R. Astron. Soc. **468**, 1338 (2017).

[276] A. X. González-Morales, D. J. E. Marsh, J. Peñarrubia, and L. A. Ureña-López, Unbiased constraints on ultralight axion mass from dwarf spheroidal galaxies, Mon. Not. R. Astron. Soc. **472**, 1346 (2017).

[277] E. Calabrese and D. N. Spergel, Ultra-light dark matter in ultra-faint dwarf galaxies, Mon. Not. R. Astron. Soc. **460**, 4397 (2016).

[278] D. J. E. Marsh and J. Silk, A model for halo formation with axion mixed dark matter, Mon. Not. R. Astron. Soc. **437**, 2652 (2014).

[279] I. De Martino, T. Broadhurst, S.-H. H. Tye, T. Chiueh, and H.-Y. Schive, Dynamical evidence of a solitonic core of $10^9 M_\odot$ in the Milky Way, Phys. Dark Univ. **28**, 100503 (2020).

[280] E. Leung, T. Broadhurst, J. Lim, J. M. Diego, T. Chiueh, H.-Y. Schive, and R. Windhorst, Magnification bias of distant galaxies in the Hubble Frontier Fields: Testing wave versus particle dark matter predictions, Astrophys. J. **862**, 156 (2018).

[281] R. Hložek, D. J. E. Marsh, and D. Grin, Using the full power of the cosmic microwave background to probe axion dark matter, Mon. Not. R. Astron. Soc. **476**, 3063 (2018).

[282] B. Bozek, D. J. E. Marsh, J. Silk, and R. F. G. Wyse, Galaxy UV-luminosity function and reionization constraints on axion dark matter, Mon. Not. R. Astron. Soc. **450**, 209 (2015).

[283] H.-Y. Schive, T. Chiueh, T. Broadhurst, and K.-W. Huang, Contrasting galaxy formation from quantum wave dark matter, ψDM, with ΛCDM, using Planck and Hubble data, Astrophys. J. **818**, 89 (2016).

[284] A. Sarkar, R. Mondal, S. Das, S. Sethi, S. Bharadwaj, and D. J. E. Marsh, The effects of the small-scale DM power on the cosmological neutral hydrogen (HI) distribution at high redshifts, J. Cosmol. Astropart. Phys. 2016 (2016) 012–012.

[285] P. S. Corasaniti, S. Agarwal, D. J. E. Marsh, and S. Das, Constraints on dark matter scenarios from measurements of the galaxy luminosity function at high redshifts, Phys. Rev. D **95**, 083512 (2017).

[286] V. Iršič, M. Viel, M. G. Haehnelt, J. S. Bolton, and G. D. Becker, First constraints on fuzzy dark matter from Lyman-α forest data and hydrodynamical simulations, Phys. Rev. Lett. **119**, 031302 (2017).

[287] E. Armengaud, N. Palanque-Delabrouille, C. Yèche, D. J. E. Marsh, and J. Baur, Constraining the mass of light bosonic dark matter using SDSS Lyman-α forest, Mon. Not. R. Astron. Soc. **471**, 4606 (2017).

[288] M. Nori, R. Murgia, V. Iršič, M. Baldi, and M. Viel, Lyman α forest and non-linear structure characterization in fuzzy dark matter cosmologies, Mon. Not. R. Astron. Soc. **482**, 3227 (2019).

[289] N. Bar, D. Blas, K. Blum, and S. Sibiryakov, Galactic rotation curves versus ultralight dark matter: Implications of the soliton-host halo relation, Phys. Rev. D **98**, 083027 (2018).

[290] N. Bar, K. Blum, J. Eby, and R. Sato, Ultralight dark matter in disk galaxies, Phys. Rev. D **99**, 103020 (2019).

[291] B. V. Church, P. Mocz, and J. P. Ostriker, Heating of Milky Way disc stars by dark matter fluctuations in cold dark matter and fuzzy dark matter paradigms, Mon. Not. R. Astron. Soc. **485**, 2861 (2019).

[292] D. J. E. Marsh and J. C. Niemeyer, Strong constraints on fuzzy dark matter from ultrafaint dwarf galaxy Eridanus II, Phys. Rev. Lett. **123**, 051103 (2019).

[293] A. Khmelnitsky and V. Rubakov, Pulsar timing signal from ultralight scalar dark matter, J. Cosmol. Astropart. Phys. 2014 (2014) 019.

[294] I. De Martino, T. Broadhurst, S.-H. H. Tye, T. Chiueh, H.-Y. Schive, and R. Lazkoz, Recognizing axionic dark matter by Compton and de Broglie scale modulation of pulsar timing, Phys. Rev. Lett. **119**, 221103 (2017).

[295] N. K. Porayko et al. (PPTA Collaboration), Parkes Pulsar Timing Array constraints on ultralight scalar-field dark matter, Phys. Rev. D **98**, 102002 (2018).

[296] D. Blas, D. L. Nacir, and S. Sibiryakov, Ultralight dark matter resonates with binary pulsars, Phys. Rev. Lett. **118**, 261102 (2017).

[297] D. Blas, D. López Nacir, and S. Sibiryakov, Secular effects of ultralight dark matter on binary pulsars, Phys. Rev. D **101**, 063016 (2020).

[298] D. Grin, M. A. Amin, V. Gluscevic, R. Hlǒzek, D. J. E. Marsh, V. Poulin, C. Prescod-Weinstein, and T. L. Smith, Gravitational probes of ultra-light axions, arXiv:1904.09003.

[299] L. Barack et al., Black holes, gravitational waves and fundamental physics: A roadmap, Class. Quantum Gravity **36**, 143001 (2019).

[300] B. Sathyaprakash et al., Extreme gravity and fundamental physics, arXiv:1903.09221.

[301] G. Bertone et al., Gravitational wave probes of dark matter: Challenges and opportunities, arXiv:1907.10610.

[302] E. Barausse et al., Prospects for fundamental physics with LISA, arXiv:2001.09793.

[303] L. A. Ureña López and A. R. Liddle, Supermassive black holes in scalar field galaxy halos, Phys. Rev. D **66**, 083005 (2002).

[304] A. Cruz-Osorio, F. S. Guzmán, and F. D. Lora-Clavijo, Scalar field dark matter: Behavior around black holes, J. Cosmol. Astropart. Phys. 2011 (2011) 029–029.

[305] L. Hui, D. Kabat, X. Li, L. Santoni, and S. S. Wong, Black hole hair from scalar dark matter, J. Cosmol. Astropart. Phys. 06 (2019) 038.

[306] K. Clough, P. G. Ferreira, and M. Lagos, Growth of massive scalar hair around a Schwarzschild black hole, Phys. Rev. D **100**, 063014 (2019).

[307] M. Bošković, F. Duque, M. C. Ferreira, F. S. Miguel, and V. Cardoso, Motion in time-periodic backgrounds with applications to ultralight dark matter halos at galactic centers, Phys. Rev. D **98**, 024037 (2018).

[308] N. Bar, K. Blum, T. Lacroix, and P. Panci, Looking for ultralight dark matter near supermassive black holes, J. Cosmol. Astropart. Phys. 07 (2019) 045.

[309] T. Jacobson, Primordial black hole evolution in tensor-scalar cosmology, Phys. Rev. Lett. **83**, 2699 (1999).

[310] S. Chadburn and R. Gregory, Time dependent black holes and scalar hair, Class. Quantum Gravity **31**, 195006 (2014).

[311] R. Gregory, D. Kastor, and J. Traschen, Black hole thermodynamics with dynamical lambda, J. High Energy Phys. 10 (2017) 118.

[312] R. Gregory, D. Kastor, and J. Traschen, Evolving black holes in inflation, Class. Quantum Gravity **35**, 155008 (2018).

[313] M. W. Horbatsch and C. P. Burgess, Cosmic black-hole hair growth and quasar OJ287, J. Cosmol. Astropart. Phys. 05 (2012) 010.

[314] A. Sillanpää, S. Haarala, M. J. Valtonen, B. Sundelius, and G. G. Byrd, OJ 287: Binary pair of supermassive black holes, Astrophys. J. **325**, 628 (1988).

[315] M. J. Valtonen et al., A massive binary black-hole system in OJ 287 and a test of general relativity, Nature **452**, 851 (2008).

[316] L. Dey et al., Authenticating the presence of a relativistic massive black hole binary in OJ 287 using its general relativity centenary flare: Improved orbital parameters, Astrophys. J. **866**, 11 (2018).

[317] A. A. Starobinsky, Amplification of waves during reflection from a rotating "black hole", Zh. Eksp. Teor. Fiz. **64**, 48 (1973) [Sov. Phys. JETP **37**, 28 (1973)].

[318] A. A. Starobinsky and S. M. Churilov, Amplification of electromagnetic and gravitational waves scattered by a rotating "black hole", Zh. Eksp. Teor. Fiz. **65**, 3 (1973) [Sov. Phys. JETP **38**, 1 (1974)].

[319] S. A. Teukolsky, Rotating black holes: Separable wave equations for gravitational and electromagnetic perturbations, Phys. Rev. Lett. **29**, 1114 (1972).

[320] S. A. Teukolsky and W. H. Press, Perturbations of a rotating black hole. III. Interaction of the hole with gravitational and electromagnetic radiation, Astrophys. J. **193**, 443 (1974).

[321] R. Penrose, Gravitational collapse: The role of general relativity, Riv. Nuovo Cimento **1**, 252 (1969) [Gen. Relativ. Gravit. **34**, 1141 (2002)].

[322] R. Penrose and R. M. Floyd, Extraction of rotational energy from a black hole, Nat. Phys. Sci. **229**, 177 (1971).

[323] Y. B. Zel'dovich, Generation of waves by a rotating body, Pis'ma Zh. Eksp. Teor. Fiz. **14**, 270 (1971) [JETP Lett. **14**, 180 (1971)].

[324] Y. B. Zel'dovich, Amplification of cylindrical electromagnetic waves reflected from a rotating body, Zh. Eksp. Teor. Fiz **62**, 2076 (1972) [Sov. Phys. JETP **35**, 1085 (1972)].

[325] T. Torres, S. Patrick, A. Coutant, M. Richartz, E. W. Tedford, and S. Weinfurtner, Rotational superradiant scattering in a vortex flow, Nat. Phys. **13**, 833 (2017).

[326] M. Richartz and A. Saa, Superradiance without event horizons in general relativity, Phys. Rev. D **88**, 044008 (2013).

[327] V. Cardoso, R. Brito, and J. L. Rosa, Superradiance in stars, Phys. Rev. D **91**, 124026 (2015).

[328] V. Cardoso, P. Pani, and T.-T. Yu, Superradiance in rotating stars and pulsar-timing constraints on dark photons, Phys. Rev. D **95**, 124056 (2017).

[329] R. Brito, V. Cardoso, and P. Pani, Superradiance, Lect. Notes Phys. **906**, 1 (2015).

[330] T. J. M. Zouros and D. M. Eardley, Instabilities of massive scalar perturbations of a rotating black hole, Ann. Phys. (N.Y.) **118**, 139 (1979).

[331] S. Detweiler, Klein-Gordon equation and rotating black holes, Phys. Rev. D **22**, 2323 (1980).

[332] V. Cardoso, O. J. C. Dias, J. P. S. Lemos, and S. Yoshida, Black-hole bomb and superradiant instabilities, Phys. Rev. D **70**, 044039 (2004); Erratum ibid **70**, 049903 (2004).

[333] S. R. Dolan, Instability of the massive Klein-Gordon field on the Kerr spacetime, Phys. Rev. D **76**, 084001 (2007).

[334] S. R. Dolan, Superradiant instabilities of rotating black holes in the time domain, Phys. Rev. D **87**, 124026 (2013).

[335] S. R. Dolan, Instability of the Proca field on Kerr spacetime, Phys. Rev. D **98**, 104006 (2018).

[336] P. Pani, V. Cardoso, L. Gualtieri, E. Berti, and A. Ishibashi, Black-hole bombs and photon-mass bounds, Phys. Rev. Lett. **109**, 131102 (2012).

[337] P. Pani, V. Cardoso, L. Gualtieri, E. Berti, and A. Ishibashi, Perturbations of slowly rotating black holes: Massive vector fields in the Kerr metric, Phys. Rev. D **86**, 104017 (2012).

[338] H. Witek, V. Cardoso, A. Ishibashi, and U. Sperhake, Superradiant instabilities in astrophysical systems, Phys. Rev. D **87**, 043513 (2013).

[339] R. Brito, V. Cardoso, and P. Pani, Black holes as particle detectors: Evolution of superradiant instabilities, Class. Quantum Gravity **32**, 134001 (2015).

[340] H. Okawa, H. Witek, and V. Cardoso, Black holes and fundamental fields in numerical relativity: Initial data construction and evolution of bound states, Phys. Rev. D **89**, 104032 (2014).

[341] M. Zilhão, H. Witek, and V. Cardoso, Nonlinear interactions between black holes and Proca fields, Class. Quantum Gravity **32**, 234003 (2015).

[342] M. Baryakhtar, R. Lasenby, and M. Teo, Black hole superradiance signatures of ultralight vectors, Phys. Rev. D **96**, 035019 (2017).

[343] W. E. East and F. Pretorius, Superradiant instability and backreaction of massive vector fields around Kerr black holes, Phys. Rev. Lett. **119**, 041101 (2017).

[344] W. E. East, Superradiant instability of massive vector fields around spinning black holes in the relativistic regime, Phys. Rev. D **96**, 024004 (2017).

[345] W. E. East, Massive boson superradiant instability of black holes: Nonlinear growth, saturation, and gravitational radiation, Phys. Rev. Lett. **121**, 131104 (2018).

[346] V. Cardoso, Ó. J. C. Dias, G. S. Hartnett, M. Middleton, P. Pani, and J. E. Santos, Constraining the mass of dark photons and axion-like particles through black-hole superradiance, J. Cosmol. Astropart. Phys. 03 (2018) 043.

[347] R. Brito, V. Cardoso, and P. Pani, Massive spin-2 fields on black hole spacetimes: Instability of the Schwarzschild and Kerr solutions and bounds on the graviton mass, Phys. Rev. D **88**, 023514 (2013).

[348] R. Brito, S. Grillo, and P. Pani, Black hole superradiant instability from ultralight spin-2 fields, Phys. Rev. Lett. **124**, 211101 (2020).

[349] C. A. R. Herdeiro and E. Radu, Kerr black holes with scalar hair, Phys. Rev. Lett. **112**, 221101 (2014).

[350] C. A. R. Herdeiro, E. Radu, and H. Rúnarsson, Kerr black holes with Proca hair, Class. Quantum Gravity **33**, 154001 (2016).

[351] C. A. R. Herdeiro and E. Radu, Dynamical formation of Kerr black holes with synchronized hair: An analytic model, Phys. Rev. Lett. **119**, 261101 (2017).

[352] J. C. Degollado, C. A. R. Herdeiro, and E. Radu, Effective stability against superradiance of Kerr black holes with synchronised hair, Phys. Lett. B **781**, 651 (2018).

[353] A. Arvanitaki and S. Dubovsky, Exploring the string axiverse with precision black hole physics, Phys. Rev. D **83**, 044026 (2011).

[354] A. Arvanitaki, M. Baryakhtar, S. Dimopoulos, S. Dubovsky, and R. Lasenby, Black hole mergers and the QCD axion at Advanced LIGO, Phys. Rev. D **95**, 043001 (2017).

[355] 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**, 083006 (2018).

[356] H. Davoudiasl and P. B. Denton, Ultralight boson dark matter and Event Horizon Telescope observations of M87*, Phys. Rev. Lett. **123**, 021102 (2019).

[357] H. Yoshino and H. Kodama, Gravitational radiation from an axion cloud around a black hole: Superradiant phase, Prog. Theor. Exp. Phys. **2014**, 043E02 (2014).

[358] H. Yoshino and H. Kodama, Probing the string axiverse by gravitational waves from Cygnus X-1, Prog. Theor. Exp. Phys. **2015**, 061E01 (2015).

[359] R. Brito, S. Ghosh, E. Barausse, E. Berti, V. Cardoso, I. Dvorkin, A. Klein, and P. Pani, Stochastic and resolvable gravitational waves from ultralight bosons, Phys. Rev. Lett. **119**, 131101 (2017).

[360] R. Brito, S. Ghosh, E. Barausse, E. Berti, V. Cardoso, I. Dvorkin, A. Klein, and P. Pani, Gravitational wave searches for ultralight bosons with LIGO and LISA, Phys. Rev. D **96**, 064050 (2017).

[361] M. C. Ferreira, C. F. B. Macedo, and V. Cardoso, Orbital fingerprints of ultralight scalar fields around black holes, Phys. Rev. D **96**, 083017 (2017).

[362] O. A. Hannuksela, K. W. K. Wong, R. Brito, E. Berti, and T. G. F. Li, Probing the existence of ultralight bosons with a single gravitational-wave measurement, Nat. Astron. **3**, 447 (2019).

[363] J. Zhang and H. Yang, Gravitational floating orbits around hairy black holes, Phys. Rev. D **99**, 064018 (2019).

[364] J. Zhang and H. Yang, Dynamic signatures of black hole binaries with superradiant clouds, Phys. Rev. D **101**, 043020 (2020).

[365] S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein, Science with the space-based interferometer LISA. V: Extreme mass-ratio inspirals, Phys. Rev. D **95**, 103012 (2017).

[366] D. Baumann, H. S. Chia, and R. A. Porto, Probing ultralight bosons with binary black holes, Phys. Rev. D **99**, 044001 (2019).

[367] D. Baumann, H. S. Chia, J. Stout, and L. ter Haar, The spectra of gravitational atoms, J. Cosmol. Astropart. Phys. 12 (2019) 006.

[368] D. Baumann, H. S. Chia, R. A. Porto, and J. Stout, Gravitational collider physics, Phys. Rev. D **101**, 083019 (2020).

[369] E. Berti, R. Brito, C. F. B. Macedo, G. Raposo, and J. L. Rosa, Ultralight boson cloud depletion in binary systems, Phys. Rev. D **99**, 104039 (2019).

[370] K. Akiyama et al. (Event Horizon Telescope Collaboration), First M87 Event Horizon Telescope results. I. The shadow of the supermassive black hole, Astrophys. J. **875**, L1 (2019).

[371] P. V. Cunha, C. A. R. Herdeiro, and E. Radu, EHT constraint on the ultralight scalar hair of the M87 supermassive black hole, Universe **5**, 220 (2019).

[372] R. Roy and U. A. Yajnik, Evolution of black hole shadow in the presence of ultralight bosons, Phys. Lett. B **803**, 135284 (2020).

[373] G. Creci, S. Vandoren, and H. Witek, Evolution of black hole shadows from superradiance, Phys. Rev. D **101**, 124051 (2020).

[374] L. Bernard, V. Cardoso, T. Ikeda, and M. Zilhão, Physics of black hole binaries: Geodesics, relaxation modes, and energy extraction, Phys. Rev. D **100**, 044002 (2019).

[375] J. Healy, T. Bode, R. Haas, E. Pazos, P. Laguna, D. M. Shoemaker, and N. Yunes, Late inspiral and merger of binary black holes in scalar-tensor theories of gravity, Class. Quantum Gravity **29**, 232002 (2012).

[376] E. Berti, V. Cardoso, L. Gualtieri, M. Horbatsch, and U. Sperhake, Numerical simulations of single and binary black holes in scalar-tensor theories: Circumventing the no-hair theorem, Phys. Rev. D **87**, 124020 (2013).

[377] W. D. Goldberger and I. Z. Rothstein, Effective field theory of gravity for extended objects, Phys. Rev. D **73**, 104029 (2006).

[378] W. D. Goldberger and I. Z. Rothstein, Dissipative effects in the worldline approach to black hole dynamics, Phys. Rev. D **73**, 104030 (2006).

[379] W. D. Goldberger and A. Ross, Gravitational radiative corrections from effective field theory, Phys. Rev. D **81**, 124015 (2010).

[380] R. A. Porto, The effective field theorist's approach to gravitational dynamics, Phys. Rep. **633**, 1 (2016).

[381] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W. H. Freeman and Company, 1973).

[382] P. Brax and J. Martin, Moduli fields as quintessence and the chameleon, Phys. Lett. B **647**, 320 (2007).

[383] K. Hinterbichler, J. Khoury, and H. Nastase, Towards a UV completion for chameleon scalar theories, J. High Energy Phys. 03 (2011) 061; Erratum ibid 06 (2011) 072.

[384] P. G. Bergmann, Comments on the scalar tensor theory, Int. J. Theor. Phys. **1**, 25 (1968).

[385] R. V. Wagoner, Scalar tensor theory and gravitational waves, Phys. Rev. D **1**, 3209 (1970).

[386] S. Kamefuchi, L. O'Raifeartaigh, and A. Salam, Change of variables and equivalence theorems in quantum field theories, Nucl. Phys. **28**, 529 (1961).

[387] R. E. Kallosh and I. V. Tyutin, The equivalence theorem and gauge invariance in renormalizable theories, Yad. Fiz. **17**, 190 (1973) [Sov. J. Nucl. Phys. **17**, 98 (1973)].

[388] H. Georgi, On-shell effective field theory, Nucl. Phys. B **361**, 339 (1991).

[389] R. M. Wald, *General Relativity* (The University of Chicago Press, 1984).

[390] T. Damour and J. F. Donoghue, Equivalence principle violations and couplings of a light dilaton, Phys. Rev. D **82**, 084033 (2010).

[391] D. J. Griffiths, *Introduction to Electrodynamics*, 4th ed. (Cambridge University Press, 2017).

[392] C. Wetterich, Cosmology and the fate of dilatation symmetry, Nucl. Phys. B **302**, 668 (1988).

[393] S. S. Gubser and J. Khoury, Scalar self-interactions loosen constraints from fifth force searches, Phys. Rev. D **70**, 104001 (2004).

[394] P. Brax, C. van de Bruck, A.-C. Davis, D. F. Mota, and D. J. Shaw, Testing chameleon theories with light propagating through a magnetic field, Phys. Rev. D **76**, 085010 (2007).

[395] P. Brax, G. Pignol, and D. Roulier, Probing strongly coupled chameleons with slow neutrons, Phys. Rev. D **88**, 083004 (2013).

[396] P. Brax and M. Pitschmann, Exact solutions to nonlinear symmetron theory: One- and two-mirror systems, Phys. Rev. D **97**, 064015 (2018).

[397] NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.0.25 of 2019-12-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[398] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison–Wesley, 1964).

[399] P. Brax and C. Burrage, Atomic precision tests and light scalar couplings, Phys. Rev. D **83**, 035020 (2011).

[400] S. K. Suslov, Expectation values in relativistic Coulomb problems, J. Phys. B **42**, 185003 (2009).

[401] C. P. Burgess, P. Hayman, M. Williams, and L. Zalavari, Point-particle effective field theory I: Classical renormalization and the inverse-square potential, J. High Energy Phys. 04 (2017) 106.

[402] P. J. Mohr, D. B. Newell, and B. N. Taylor, CODATA recommended values of the fundamental physical constants: 2014, Rev. Mod. Phys. **88**, 035009 (2016).

[403] E. Poisson, A. Pound, and I. Vega, The motion of point particles in curved spacetime, Living Rev. Relativity **14**, 7 (2011).

[404] L. Parker, One electron atom as a probe of space-time curvature, Phys. Rev. D **22**, 1922 (1980).

[405] L. Parker and L. O. Pimentel, Gravitational perturbation of the hydrogen spectrum, Phys. Rev. D **25**, 3180 (1982).

[406] E. Gill, G. Wunner, M. Soffel, and H. Ruder, On hydrogen-like atoms in strong gravitational fields, Class. Quantum Gravity **4**, 1031 (1987).

[407] B. de Beauvoir, C. Schwob, O. Acef, L. Jozefowski, L. Hilico, F. Nez, L. Julien, A. Clairon, and F. Biraben, Metrology of the hydrogen and deuterium atoms: Determination of the Rydberg constant and Lamb shifts, Eur. Phys. J. D **12**, 61 (2000).

[408] K. Pachucki, D. Leibfired, M. Weitz, A. Huber, W. Konig, and T. W. Hansch, Theory of the energy levels and precise two-photon spectroscopy of atomic hydrogen and deuterium, J. Phys. B **29**, 177 (1996).

[409] M. I. Eides, H. Grotch, and V. A. Shelyuto, Theory of light hydrogen-like atoms, Phys. Rep. **342**, 63 (2001).

[410] S. G. Karshenboim, Precision physics of simple atoms: QED tests, nuclear structure and fundamental constants, Phys. Rep. **422**, 1 (2005).

[411] M. E. Rose, *Relativistic Electron Theory* (John Wiley, 1961).

[412] M. E. Rose, *Elementary Theory of Angular Momentum* (Dover, 1995).

[413] T. Aoyama, T. Kinoshita, and M. Nio, Revised and improved value of the QED tenth-order electron anomalous magnetic moment, Phys. Rev. D **97**, 036001 (2018).

[414] G. Audi, M. Wang, A. Wapstra, F. Kondev, M. MacCormick, X. Xu, and B. Pfeiffer, The ame2012 atomic mass evaluation, Chin. Phys. C **36**, 1287 (2012).

[415] M. Wang, G. Audi, A. Wapstra, F. Kondev, M. MacCormick, X. Xu, and B. Pfeiffer, The ame2012 atomic mass evaluation, Chin. Phys. C **36**, 1603 (2012).

[416] P. Clade, E. de Mirandes, M. Cadoret, S. Guellati-Khelifa, C. Schwob, F. Nez, L. Julien, and F. Biraben, Determination of the fine structure constant based on Bloch oscillations of ultracold atoms in a vertical optical lattice, Phys. Rev. Lett. **96**, 033001 (2006).

[417] M. Cadoret, E. de Mirandes, P. Clade, S. Guellati-Khelifa, C. Schwob, F. Nez, L. Julien, and F. Biraben, Combination of Bloch oscillations with a Ramsey-Borde interferometer: New determination of the fine structure constant, Phys. Rev. Lett. **101**, 230801 (2008).

[418] R. Bouchendira, P. Clade, S. Guellati-Khelifa, F. Nez, and F. Biraben, New determination of the fine structure constant and test of the quantum electrodynamics, Phys. Rev. Lett. **106**, 080801 (2011).

[419] F. Schwabl, R. Hilton, and A. Lahee, *Advanced Quantum Mechanics* (Springer-Verlag, 2013).

[420] G. S. Adkins, Dirac–Coulomb energy levels and expectation values, Am. J. Phys. **76**, 579 (2008).

[421] D. Hanneke, S. Fogwell, and G. Gabrielse, New measurement of the electron magnetic moment and the fine structure constant, Phys. Rev. Lett. **100**, 120801 (2011).

[422] D. Hanneke, S. Fogwell Hoogerheide, and G. Gabrielse, Cavity control of a single-electron quantum cyclotron: Measuring the electron magnetic moment, Phys. Rev. A **83**, 052122 (2011).

[423] P. Agrawal, $g-2$ in QED and beyond, in *The Allure of Ultrasensitive Experiments*, Fermilab Academic Lectures (2013).

[424] J. Schwinger, On quantum-electrodynamics and the magnetic moment of the electron, Phys. Rev. **73**, 416 (1948).

[425] J. Schwinger, Quantum electrodynamics. III. The electromagnetic properties of the electron—Radiative corrections to scattering, Phys. Rev. **76**, 790 (1949).

[426] R. Jackiw and S. Weinberg, Weak-interaction corrections to the muon magnetic moment and to muonic-atom energy levels, Phys. Rev. D **5**, 2396 (1972).

[427] F. Jegerlehner and A. Nyffeler, The muon $g-2$, Phys. Rep. **477**, 1 (2009).

[428] G. F. Giudice, P. Paradisi, and M. Passera, Testing new physics with the electron $g-2$, J. High Energy Phys. 11 (2012) 113.

[429] C.-Y. Chen, H. Davoudiasl, W. J. Marciano, and C. Zhang, Implications of a light "dark Higgs" solution to the $g_\mu-2$ discrepancy, Phys. Rev. D **93**, 035006 (2016).

[430] W. J. Marciano, A. Masiero, P. Paradisi, and M. Passera, Contributions of axionlike particles to lepton dipole moments, Phys. Rev. D **94**, 115033 (2016).

[431] A. Upadhye, W. Hu, and J. Khoury, Quantum stability of chameleon field theories, Phys. Rev. Lett. **109**, 041301 (2012).

[432] P. Brax, A.-C. Davis, and J. Sakstein, SUPER-Screening, Phys. Lett. B **719**, 210 (2013).

[433] C. Burrage, E. J. Copeland, and P. Millington, Radiative screening of fifth forces, Phys. Rev. Lett. **117**, 211102 (2016).

[434] G. Gabrielse, X. Fei, L. A. Orozco, R. L. Tjoelker, J. Haas, H. Kalinowsky, T. A. Trainor, and W. Kells, Thousandfold improvement in the measured antiproton mass, Phys. Rev. Lett. **65**, 1317 (1990).

[435] L. S. Brown and G. Gabrielse, Geonium theory: Physics of a single electron or ion in a Penning trap, Rev. Mod. Phys. **58**, 233 (1986).

[436] L. S. Brown and G. Gabrielse, Precision spectroscopy of a charged particle in an imperfect Penning trap, Phys. Rev. A **25**, 2423 (1982).

[437] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, Tenth-order QED contribution to the electron $g-2$ and an improved value of the fine structure constant, Phys. Rev. Lett. **109**, 111807 (2012).

[438] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, Tenth-order electron anomalous magnetic moment: Contribution of diagrams without closed lepton loops, Phys. Rev. D **91**, 033006 (2015); Erratum ibid **96**, 019901 (2017).

[439] G. A. Baker and P. Graves-Morris, *Padé Approximants*, 2nd ed. (Cambridge University Press, 1996).

[440] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. (Cambridge University Press, 2007).

[441] B. Elder, J. Khoury, P. Haslinger, M. Jaffe, H. Müller, and P. Hamilton, Chameleon dark energy and atom interferometry, Phys. Rev. D **94**, 044051 (2016).

[442] M. Srednicki, *Quantum Field Theory* (Cambridge University Press, 2007).

[443] S. M. Barr and A. Zee, Electric dipole moment of the electron and of the neutron, Phys. Rev. Lett. **65**, 21 (1990); Erratum ibid **65**, 2920 (1990).

[444] C. G. Parthey, A. Matveev, J. Alnis, R. Pohl, T. Udem, U. D. Jentschura, N. Kolachevsky, and T. W. Hänsch, Precision measurement of the hydrogen-deuterium $1S-2S$ isotope shift, Phys. Rev. Lett. **104**, 233001 (2010).

[445] P. Brax, C. Burrage, A.-C. Davis, D. Seery, and A. Weltman, Collider constraints on interactions of dark energy with the Standard Model, J. High Energy Phys. 09 (2009) 128.

[446] K. G. Wilson and J. Kogut, The renormalization group and the ϵ expansion, Phys. Rep. **12**, 75 (1974).

[447] L. Blanchet, Gravitational radiation from post-Newtonian sources and inspiralling compact binaries, Living Rev. Relativity **17**, 2 (2014).

[448] L. Blanchet and T. Damour, Radiative gravitational fields in general relativity I. General structure of the field outside the source, Phil. Trans. R. Soc. A. **320**, 379 (1986).

[449] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, 1973).

[450] H. Reall, *Black Holes* (2020), http://www.damtp.cam.ac.uk/user/hsr1000/black_holes_lectures_2020.pdf.

[451] R. M. Wald, "Weak" cosmic censorship, PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association **1992**, 181 (1992).

[452] R. Penrose, The question of cosmic censorship, J. Astrophys. Astr. **20**, 233 (1999).

[453] S. W. Hawking, Black holes in general relativity, Comm. Math. Phys. **25**, 152 (1972).

[454] B. Carter, Killing horizons and orthogonally transitive groups in space-time, J. Math. Phys. **10**, 70 (1969).

[455] U. Nucamendi and M. Salgado, Scalar hairy black holes and solitons in asymptotically flat spacetimes, Phys. Rev. D **68**, 044026 (2003).

[456] K. G. Zloshchastiev, Coexistence of black holes and a long-range scalar field in cosmology, Phys. Rev. Lett. **94**, 121101 (2005).

[457] M. Cadoni and E. Franzin, Asymptotically flat black holes sourced by a massless scalar field, Phys. Rev. D **91**, 104011 (2015).

[458] A. Anabalón and J. Oliva, Exact hairy black holes and their modification to the universal law of gravitation, Phys. Rev. D **86**, 107501 (2012).

[459] A. Anabalón and N. Deruelle, Mechanical stability of asymptotically flat black holes with minimally coupled scalar hair, Phys. Rev. D **88**, 064011 (2013).

[460] D. A. Feinblum and W. A. McKinley, Stable states of a scalar particle in its own gravational field, Phys. Rev. **168**, 1445 (1968).

[461] P. Jetzer, Boson stars, Phys. Rep. **220**, 163 (1992).

[462] E. Seidel and W.-M. Suen, Oscillating soliton stars, Phys. Rev. Lett. **66**, 1659 (1991).

[463] S. L. Liebling and C. Palenzuela, Dynamical boson stars, Living Rev. Relativity **20**, 5 (2017).

[464] L. Visinelli, S. Baum, J. Redondo, K. Freese, and F. Wilczek, Dilute and dense axion stars, Phys. Lett. B **777**, 64 (2018).

[465] D. J. Rowan and G. Stephenson, The Klein-Gordon equation in a Kerr-Newman background space, J. Phys. A **10**, 15 (1977).

[466] E. Berti, V. Cardoso, and M. Casals, Eigenvalues and eigenfunctions of spin-weighted spheroidal harmonics in four and higher dimensions, Phys. Rev. D **73**, 024013 (2006); Erratum ibid **73**, 109902 (2006).

[467] B. Kol, The delocalized effective degrees of freedom of a black hole at low frequencies, Gen. Relativ. Gravit. **40**, 2061 (2008) [Int. J. Mod. Phys. D **17**, 2617 (2008)].

[468] S. Endlich and R. Penco, A modern approach to superradiance, J. High Energy Phys. 05 (2017) 052.

[469] C. R. Galley and B. L. Hu, Self-force on extreme mass ratio inspirals via curved spacetime effective field theory, Phys. Rev. D **79**, 064002 (2009).

[470] W. D. Goldberger, Effective field theories and gravitational radiation, in *Particle Physics and Cosmology: The Fabric of Spacetime*, edited by F. Bernardeau, C. Grojean, and J. Dalibard, pp. 351–396, Lecture Notes of the Les Houches Summer School 2006.

[471] K. S. Thorne, Multipole expansions of gravitational radiation, Rev. Mod. Phys. **52**, 299 (1980).

[472] W. D. Goldberger and I. Z. Rothstein, An effective field theory of quantum mechanical black hole horizons, J. High Energy Phys. 04 (2020) 056.

[473] W. D. Goldberger and I. Z. Rothstein, Virtual Hawking radiation, arXiv:2007.00726.

[474] C. R. Galley, Classical mechanics of nonconservative systems, Phys. Rev. Lett. **110**, 174301 (2013).

[475] J. Schwinger, Brownian motion of a quantum oscillator, J. Math. Phys. **2**, 407 (1961).

[476] L. V. Keldysh, Diagram technique for nonequilibrium processes, Zh. Eksp. Teor. Fiz. **47**, 1515 (1965) [Sov. Phys. JETP **20**, 1018 (1965)].

[477] K.-C. Chou, Z.-B. Su, B.-L. Hao, and L. Yu, Equilibrium and nonequilibrium formalisms made unified, Phys. Rep. **118**, 1 (1985).

[478] R. D. Jordan, Effective field equations for expectation values, Phys. Rev. D **33**, 444 (1986).

[479] E. Calzetta and B. L. Hu, Closed-time-path functional formalism in curved spacetime: Application to cosmological back-reaction problems, Phys. Rev. D **35**, 495 (1987).

[480] C. R. Galley and M. Tiglio, Radiation reaction and gravitational waves in the effective field theory approach, Phys. Rev. D **79**, 124027 (2009).

[481] C. R. Galley, A nonlinear scalar model of extreme mass ratio inspirals in effective field theory: I. Self-force through third order, Class. Quantum Gravity **29**, 015010 (2012).

[482] O. Birnholtz, S. Hadar, and B. Kol, Theory of post-Newtonian radiation and reaction, Phys. Rev. D **88**, 104037 (2013).

[483] T. S. Bunch and L. Parker, Feynman propagator in curved space-time: A momentum space representation, Phys. Rev. D **20**, 2499 (1979).

[484] M. Rinaldi, Momentum-space representation of Green's functions with modified dispersion relations on general backgrounds, Phys. Rev D. **78**, 024025 (2008).

[485] B. Kol and M. Smolkin, Black hole stereotyping: Induced gravito-static polarization, J. High Energy Phys. 02 (2012) 010.

[486] T. Damour and A. Nagar, Relativistic tidal properties of neutron stars, Phys. Rev. D **80**, 084035 (2009).

[487] T. Binnington and E. Poisson, Relativistic theory of tidal Love numbers, Phys. Rev. D **80**, 084018 (2009).

[488] S. Chakrabarti, T. Delsate, and J. Steinhoff, Effective action and linear response of compact objects in Newtonian gravity, Phys. Rev. D **88**, 084038 (2013).

[489] S. Chakrabarti, T. Delsate, and J. Steinhoff, New perspectives on neutron star and black hole spectroscopy and dynamic tides, arXiv:1304.2228.

[490] I. Z. Rothstein, Progress in effective field theory approach to the binary inspiral problem, Gen. Relativ. Gravit. **46**, 1726 (2014).

[491] R. A. Porto, The tune of love and the nature(ness) of spacetime, Fortschr. Phys. **64**, 723 (2016).

[492] J. A. H. Futterman, F. A. Handler, and R. A. Matzner, *Scattering from Black Holes* (Cambridge University Press, Cambridge, England, 1988).

[493] H.-Y. Schive, T. Chiueh, and T. Broadhurst, Cosmic structure as the quantum interference of a coherent dark wave, Nat. Phys. **10**, 496 (2014).

[494] H.-Y. Schive, M.-H. Liao, T.-P. Woo, S.-K. Wong, T. Chiueh, T. Broadhurst, and W.-Y. Pauchy Hwang, Understanding the core-halo relation of quantum wave dark matter from 3D simulations, Phys. Rev. Lett. **113**, 261302 (2014).

[495] B. Schwabe, J. C. Niemeyer, and J. F. Engels, Simulations of solitonic core mergers in ultralight axion dark matter cosmologies, Phys. Rev. D **94**, 043513 (2016).

[496] J. Veltmaat and J. C. Niemeyer, Cosmological particle-in-cell simulations with ultralight axion dark matter, Phys. Rev. D **94**, 123523 (2016).

[497] J. Huang, M. C. Johnson, L. Sagunski, M. Sakellariadou, and J. Zhang, Prospects for axion searches with Advanced LIGO through binary mergers, Phys. Rev. D **99**, 063013 (2019).

[498] A. Kuntz, F. Piazza, and F. Vernizzi, Effective field theory for gravitational radiation in scalar-tensor gravity, J. Cosmol. Astropart. Phys. 1905 (2019) 052.

[499] S. Chandrasekhar, Dynamical friction I. General considerations: The coefficient of dynamical friction., Astrophys. J. **97**, 255 (1943).

[500] J. Binney and S. Tremaine, *Galactic Dynamics*, 2nd ed. (Princeton University Press, 2008).

[501] E. C. Ostriker, Dynamical friction in a gaseous medium, Astrophys. J. **513**, 252 (1999).

[502] H. Kim and W.-T. Kim, Dynamical friction of a circular-orbit perturber in a gaseous medium, Astrophys. J. **665**, 432 (2007).

[503] B. Bar-Or, J.-B. Fouvry, and S. Tremaine, Relaxation in a fuzzy dark matter halo, Astrophys. J. **871**, 28 (2019).

[504] J. B. Gilmore and A. Ross, Effective field theory calculation of second post-Newtonian binary dynamics, Phys. Rev. D **78**, 124021 (2008).

[505] P. C. Peters and J. Mathews, Gravitational radiation from point masses in a Keplerian orbit, Phys. Rev. **131**, 435 (1963).

[506] D. Gerosa, S. Ma, K. W. K. Wong, E. Berti, R. O'Shaughnessy, Y. Chen, and K. Belczynski, Multiband gravitational-wave event rates and stellar physics, Phys. Rev. D **99**, 103004 (2019).

[507] C. J. Moore, D. Gerosa, and A. Klein, Are stellar-mass black-hole binaries too quiet for LISA?, Mon. Not. R. Astron. Soc. **488**, L94 (2019).

[508] E. Poisson and C. M. Will, *Gravity: Newtonian, Post-Newtonian, Relativistic* (Cambridge University Press, 2014).

[509] J. Van Bladel, Electromagnetic fields in the presence of rotating bodies, Proc. IEEE **64**, 301 (1976).

[510] J. Cooper, Scattering of electromagnetic fields by a moving boundary: The one-dimensional case, IEEE Trans. Antennas Propag. **28**, 791 (1980).

[511] J. Cooper, Long-time behavior and energy growth for electromagnetic waves reflected by a moving boundary, IEEE Trans. Antennas Propag. **41**, 1365 (1993).

[512] J. D. Bekenstein and M. Schiffer, The many faces of superradiance, Phys. Rev. D **58**, 064014 (1998).

[513] M. Richartz, S. Weinfurtner, A. J. Penner, and W. G. Unruh, Generalized superradiant scattering, Phys. Rev. D **80**, 124016 (2009).

[514] J. Humblet, Analytical structure and properties of Coulomb wave functions for real and complex energies, Ann. Phys. (N.Y.) **155**, 461 (1984).

[515] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Springer, 1977).

[516] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, 1964) pp. 52–57.

[517] L. Hostler, Coulomb Green's functions and the Furry approximation, J. Math. Phys. **5**, 591 (1964).

[518] R. A. Mapleton, Characteristic Coulomb Green's function and its eigenfunction expansion, J. Math. Phys. **2**, 478 (1961).

[519] D. Gerosa, U. Sperhake, and C. D. Ott, Numerical simulations of stellar collapse in scalar-tensor theories of gravity, Class. Quantum Gravity **33**, 135002 (2016).

[520] U. Sperhake, C. J. Moore, R. Rosca, M. Agathos, D. Gerosa, and C. D. Ott, Long-lived inverse chirp signals from core-collapse in massive scalar-tensor gravity, Phys. Rev. Lett. **119**, 201103 (2017).

[521] R. Rosca-Mead, C. J. Moore, M. Agathos, and U. Sperhake, Inverse-chirp signals and spontaneous scalarisation with self-interacting potentials in stellar collapse, Class. Quantum Gravity **36**, 134003 (2019).

[522] P. C.-K. Cheong and T. G. F. Li, Numerical studies on core collapse supernova in self-interacting massive scalar-tensor gravity, Phys. Rev. D **100**, 024027 (2019).

[523] U. Sperhake (private communication).

[524] R. Iengo, Sommerfeld enhancement: General results from field theory diagrams, J. High Energy Phys. 05 (2009) 024.

[525] S. Cassel, Sommerfeld factor for arbitrary partial wave processes, J. Phys. G **37**, 105009 (2010).

[526] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, A theory of dark matter, Phys. Rev. D **79**, 015014 (2009).

[527] V. Cardoso and R. Vicente, Moving black holes: Energy extraction, absorption cross section, and the ring of fire, Phys. Rev. D **100**, 084001 (2019).

[528] M. Maggiore, *Gravitational Waves: Theory and Experiments*, Vol. 1 (Oxford University Press, 2007).

[529] J. L. Friedman, K. Uryū, and M. Shibata, Thermodynamics of binary black holes and neutron stars, Phys. Rev. D **65**, 064035 (2002); Erratum ibid **70**, 129904 (2004).

[530] A. Le Tiec, L. Blanchet, and B. F. Whiting, First law of binary black hole mechanics in general relativity and post-Newtonian theory, Phys. Rev. D **85**, 064039 (2012).

[531] J. D. Bekenstein, Extraction of energy and charge from a black hole, Phys. Rev. D **7**, 949 (1973).

[532] D. Gaspard, Connection formulas between Coulomb wave functions, J. Math. Phys. **59**, 112104 (2018).

[533] A. Dzieciol, S. Yngve, and P. O. Fröman, Coulomb wave functions with complex values of the variable and the parameters, J. Math. Phys. **40**, 6145 (1999).

[534] C. P. Burgess, J. Hainge, G. Kaplanek, and M. Rummel, Failure of perturbation theory near horizons: The Rindler example, J. High Energy Phys. 10 (2018) 122.