THE UNIVERSITY OF CHICAGO

PHENOMENOLOGICAL MODEL BUILDING IN PARTICLE COSMOLOGY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
AURORA NICOLE IRELAND

CHICAGO, ILLINOIS
JUNE 2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you from the bottom of my heart to my wonderful panel of mentors — my advisor Marcela and my close collaborators Dan, Gordan, Stefano, and Carlos. You have been so generous with your time and wisdom, and are all such inspirations to me. I look forward to making you proud.

Thank you also to my wonderful partner Marisa, for her endless patience, support, and love. (Not to mention her tolerance of my black hole ramblings.)

Thank you to my parents, who have always encouraged me to follow my passions and whose support has made that possible.

Finally, thank you to my cats Finnian and Tortilla, for their constructive comments on an earlier version of this manuscript.

ABSTRACT

The Standard Model of particle physics and the ΛCDM cosmological paradigm have been largely successful in describing the current state of the universe and how this came to be. And yet, we know both descriptions to be incomplete, as there exist a great number of open questions left unaddressed by both. This thesis presents several model building efforts towards resolving these shortcomings in our current picture of particle cosmology. After reviewing the current state of affairs and the puzzles that persist, we turn to the events of the early universe. Focusing on two exceptionally promising probes — primordial black holes and their accompanying gravitational waves — we show how constraints on both can lead to inferences about inflation, the subsequent expansion history, and more. Next, we turn to the cosmological phase transitions which may have occurred during this period, with particular emphasis on strongly first-order phase transitions accompanied by gravitational wave signatures. We address a major source of uncertainty in predictions of such phase transitions and present novel theoretical bounds on a model which is in principle capable of making the electroweak phase transition strongly first-order. Finally, we examine three mysteries of late-time cosmology, informed by present day observations and data. First, in the context of dark matter model building, we present a new scheme of asymmetric reheating which can reconcile dark sectors with light degrees of freedom with precision observables. Next, in light of recent observations of supermassive black holes at surprisingly high redshifts, we explore the possibility of supermassive black holes of a primordial origin. Finally, we present a novel mechanism for generating the primordial magnetic fields required to seed the cosmological magnetic fields observed today.

CHAPTER 1

INTRODUCTION

"The universe is the poor man's accelerator". As the words of Zel'dovich would suggest, there exists a rich symbiosis between theoretical particle physics and cosmology. On the one hand, fundamental particle physics has been extremely successful in describing certain aspects of the early universe, such as predicting the abundances of the light elements and matching the observed spectrum of the cosmic microwave background. At the same time, the early universe presents a valuable testing ground for particle physics, allowing us to probe energies far beyond those accessible in terrestrial experiments.

Both the Standard Model of particle physics and the standard ΛCDM cosmological paradigm are incomplete. The former lacks particle content to account for the dark matter and dark energy, an explanation for the matter-antimatter asymmetry of the universe, a mechanism for generating neutrino masses, and much more, in addition to experiencing a number of fine-tuning "problems". Meanwhile, the latter suffers from various observational anomalies and cosmological tensions. There is good reason to believe that studying these models in parallel could lead to extensions that address shortcomings in both.

This thesis presents several efforts in this direction. After reviewing the current state of the Standard Model and the ΛCDM cosmological paradigm in Chapter 2, we turn to the early universe, about which much remains to be understood. In Chapter 3, we highlight two particularly promising early universe probes: primordial black holes and gravitational waves. Primordial black holes form from the collapse of large density perturbations in the first fractions of a second following the Big Bang, and so constitute relics of our universe's earliest epoch. They are often accompanied by distinctive gravitational wave signatures from their formation, Hawking radiation, and mergers, which travel effectively unimpeded through the early universe, offering a remarkably clean probe to be compared against measured spectra.

Primordial black holes and gravitational waves could therefore play a vital role in shedding

light on inflation and the subsequent epochs of cosmological expansion. We begin in Sec. 3.1 by reviewing in great detail primordial black hole formation and the enhancement of small scale power from features in the inflationary potential. Then in Sec. 3.2, we turn to primordial black hole evaporation and show how constraints on gravitational waves can allow us to make inferences about the expansion history of the universe. Analogously, in Sec. 3.3 we show how they may be used to put constraints on extra-dimensional scenarios.

In Chapter 4, we turn to the cosmological phase transitions which occurred as the universe expanded and cooled. In particular, we will be interested in strongly first-order phase transitions proceeding through the nucleation of true vacuum bubbles, since these are accompanied by gravitational wave signatures. This includes the electroweak phase transition in certain extensions of the Standard Model, as well as potential first-order phase transitions in dark sectors. A strongly first-order electroweak phase transition is an especially interesting possibility since it is a prerequisite for electroweak baryogenesis — one of the leading candidates for establishing the matter-antimatter asymmetry of the universe. Further, upcoming space-based interferometers are expected to be sensitive to the gravitational wave signal from a first-order electroweak phase transition with frequency peaking in the mHz range.

We begin in Sec. 4.1 by presenting an in-depth review of first-order cosmological phase transitions, including the thermodynamics of the transition, the hydrodynamic description of true vacuum bubbles expanding in a thermal plasma, and the production of gravitational waves. Predictions of thermal phase transition parameters and thereby of gravitational wave signals are plagued by large theoretical uncertainties. In Sec. 4.2, we address one source of uncertainties — the resummation of the finite temperature effective potential. We discuss the accuracy of different resummation prescriptions and point out conceptual issues with certain approximations made in the literature. Finally in Sec. 4.3, we look at 2 Higgs doublet models, which provide the low energy effective description of many well-motivated Standard Model extensions, and further, may lead to a strongly first-order electroweak phase transition and

possibly even a successful baryogenesis.

In Chapter 5, we examine several open questions in cosmology persisting today. Despite constituting roughly 26% of the universe's energy budget, the identity of the dark matter remains elusive. There exist a number of compelling particle dark matter candidates, though model building in dark sectors with light new species is constrained by precision measurements of cosmological parameters. In Sec. 5.1 we highlight a novel scheme of *asymmetric reheating* that can reconcile models featuring light dark species with observation. Another puzzle arises in light of recent observations of the high-redshift universe revealing a surprising number of supermassive black holes. These findings have challenged the standard picture of supermassive black hole growth from stellar-mass black hole seeds through accretion at the Eddington-limited rate. In light of this, in Sec. 5.2 we explore the possibility of supermassive black holes of a *primordial* origin. Finally, the origin of the tiny primordial magnetic fields which seed the cosmological magnetic fields present within galaxies and galaxy clusters today poses yet another outstanding mystery. Very few processes in the early universe could have given rise to magnetic fields with sufficiently long correlation lengths to survive the diffusive processes of the hot primordial plasma. In Sec. 5.3, we propose a novel mechanism which makes use of a population of charged, spinning black holes in the early universe.

It should be noted that some of the material in this thesis has either previously appeared in publication or is awaiting publication by this author. In Table 1.1, we summarize this author's previous works and the sections where pieces of them may appear.

Publication	**Ref.**	**Appears in**
Supermassive Primordial Black Holes From Inflation	[24]	Sec. 3.1.1, Sec. 5.2
Primordial Gravitational Waves from Black Hole Evaporation in Standard and Nonstandard Cosmologies	[25]	Sec. 3.2
Gravitational Waves from Primordial Black Hole Evaporation with Large Extra Dimensions	[26]	Sec. 3.3
Improved Thermal Resummation for Multi-Field Potentials	—	Sec. 4.2
New Tools for Dissecting the General 2HDM	[27]	Sec. 4.3
Asymmetric Reheating via Inverse Symmetry Breaking	[28]	Sec. 5.1
Cosmological Magnetic Fields from Primordial Kerr-Newman Black Holes	[29]	Sec. 5.3
$T\overline{T}$ *Deformed* YM_2 *on General Backgrounds from a Integral Transformation*	[30]	—

Table 1.1: A listing of this author's publications and the sections in which material from these publications appears.

CHAPTER 2

ΛCDM AND THE STANDARD MODEL

This thesis presents several examples of model building efforts to address extant mysteries in particle physics and cosmology. In order to motivate these models, we first need to review the current state of the standard cosmological paradigm and the Standard Model of particle physics, in particular highlighting the unresolved questions they fail to satisfactorily address. This is the goal of this chapter. We begin in Sec. 2.1 by reviewing the ΛCDM cosmological model, as well as the abundance of observational data which has informed it. In Sec. 2.2, we provide the same treatment for the Standard Model of particle physics. Finally in Sec. 2.3, we highlight the shortcomings of both and indicate directions for future model building.

2.1 The Standard Cosmology

It is now essentially universally accepted that the universe expanded from an initial state of incredibly high temperature and density — the Big Bang singularity[1]. The hot Big Bang model was first put forth in late 1920s by Georges Lemaître, who inferred that the recession of nearby galaxies could be explained by an expanding universe — a proposal which was confirmed shortly afterwards by Edwin Hubble. Extrapolating this expansion backwards in time, he proposed that at some finite time in the distant past, all mass in the universe must have been concentrated in a single "primeval atom" [34, 35]. Unlike the prevailing steady-state model of the day, the Big Bang model predicted that we should see a uniform background of radiation, a remnant of our universe's high-temperature past. With the serendipitous discovery of the cosmic microwave background (CMB) in 1964 [36], this

1. Technically there need not have been a literal physical singularity at the origin of cosmic time. The Big Bang theory is predicated on the equations of classical general relativity, which break down in the extreme conditions of the very early universe. A true theory of quantum gravity is needed to extrapolate farther backwards, and is generally expected to resolve this apparent singular behavior. Alternatives to the Big Bang singularity include emergent universe scenarios [31], bouncing universes [32], and cyclic cosmologies [33].

prediction was confirmed and the Big Bang model became the leading candidate for our universe's origins.

The following decades saw a slow trickle of observational data confirming that a hot Big Bang predicts the correct light element abundances [37, 38, 39, 40], further substantiating the theory. These years also saw a mounting body of evidence for dark matter, which had been postulated by Fritz Zwicky while studying the Coma cluster of galaxies some years prior in 1933 [41]. With measurements of spiral galaxy rotation curves by Vera Rubin and Kent Ford in the 1970s [42] and the observation of gravitational lensing by galaxy clusters in the 1980s, it eventually came to be accepted that a large component of the universe's energy density exists in the form of dark matter. In 1992 the Cosmic Background Explorer (COBE) measured the first deviations from isotropy, or anisotropies, in the CMB [13], believed to be the seeds of structure. Computer simulations of structure formation revealed that these seeds could conceivably have produced the large scale structure observed today provided the dark matter was slow-moving with a small free streaming length — so-called *cold dark matter*.

A few years later in 1998, observations of Type 1a supernovae suggested that the expansion of space is actually accelerating [43, 44]. Following corroboration from several other independent lines of evidence, it has become generally accepted that we are living in a period of accelerated expansion driven by some hitherto unknown component dominating the universe's energy budget — *dark energy*. Perhaps the simplest explanation for dark energy is that it is an intrinsic energy of the spacetime vacuum itself. Mathematically, this corresponds to a positive value for the cosmological constant Λ. Thus, we come to a picture of the universe as consisting predominantly of dark energy and cold dark matter components, in addition to a subdominant fraction of ordinary baryonic matter — the ΛCDM model of Big Bang cosmology.

The technological advances of the past three decades have ushered cosmology into a

golden age, rich with data and observations. While increasingly precise measurements have slightly shifted the values of certain cosmological parameters, the ΛCDM model has remained steadfast as the leading cosmological paradigm. Nevertheless, many questions remain, and cosmological tensions between values of parameters measured in different ways[2] have challenged the assumption of a homogeneous FLRW-based cosmology. In the following subsections, we first recap the major sources of data in observational cosmology which have culminated in the ΛCDM model, before reviewing the ΛCDM model. We defer our discussion of the open questions unaddressed by this model to Sec. 2.3.

2.1.1 The Data

Here we briefly review the main sources of cosmological data which have informed our current ΛCDM paradigm, including: the light element abundances, the cosmic microwave background, and large scale structure. See e.g. [47] for further discussion.

The Light Element Abundances

The lightest[3] "elements" — hydrogen, deuterium, helium, and lithium — are now understood to have formed through nuclear fusion in the hot, dense conditions of the first seconds to minutes after the Big Bang in a process called *Big Bang Nucleosynthesis* (BBN). This time period corresponds to a temperature range of $T \sim 1 - 0.01\,\mathrm{MeV}$, which is sufficiently energetic for the strong force to bind together free protons and neutrons but not so energetic that the resultant nucleons are immediately broken apart by high-energy photons.

Predictions for the light element abundances are calculable from a knowledge of Standard Model physics and depend on the neutron-to-proton ratio as well as the baryon-to-photon ratio. These calculations were originally presented, albeit erroneously, in the famous 1948

2. Notably, the infamous Hubble tension and S_8 tension. See Refs. [45, 46] for a comprehensive overview.

3. Forming heavier elements requires stellar fusion, and so did not take place until much later — approximately 50-200 Myr after the Big Bang.

"Alpha-Beta-Gamma" paper [48]. Corrected calculations had been worked out by the 1960s [37, 49]. These predictions have since been compared to measurements of the light element abundances, in particular measurements of the primordial deuterium-to-hydrogen ratio $(^2\mathrm{H/H})_p = (2.569 \pm 0.27) \times 10^{-5}$, the primordial helium mass fraction $Y_p = 0.245 \pm 0.003$, and the primordial lithium-to-hydrogen ratio $(\mathrm{Li/H})_p = (1.6 \pm 0.3) \times 10^{-10}$ [50]. With the exception of the lithium abundance[4], these measurements agree precisely with the predicted values, robustly evincing Big Bang theory over the steady state model that had persisted until the 1950s.

These measurements of the light element abundances are invaluable probes of the early universe, informing us about the universe's energy density and expansion rate as early as a second after the Big Bang. They also give us insight into the various events that took place around this time-frame, including neutrino decoupling ($t \sim 1\,\mathrm{s}$, $T \sim 1\,\mathrm{MeV}$), proton-neutron freeze-out ($t \sim 1\,\mathrm{s}$, $T \sim 1\,\mathrm{MeV}$), free neutron decay ($t \sim 15\,\mathrm{min}$, $T \sim 30\,\mathrm{keV}$), and more.

The Cosmic Microwave Background

The cosmic microwave background is arguably our most significant source of information about the primordial universe, which is somewhat surprising as its 1965 discovery by Arno Penzias and Robert Wilson was completely accidental [51]. Intuitively, the CMB is a snapshot of the universe as it appeared just $\sim 380,000$ years after the Big Bang. By this point, the universe had cooled sufficiently ($T \sim 0.3\,\mathrm{eV}$) for electrons and nucleons to combine into neutral atoms (an event confusingly referred to as *recombination*), allowing photons to decouple from the thermal plasma. They have been free streaming through the universe ever since, and today these redshifted photons constitute a thermal background of microwave radiation.

4. The measured abundance of ^{7}Li is roughly 3 times less than the predicted value — the so-called *cosmological lithium problem*. It is unclear whether this discrepancy is due to astrophysical uncertainties (namely the depletion rate of lithium in stars) or physics beyond the Standard Model.

The spectrum of the CMB is almost that of a perfect blackbody, indicating that the photons were in kinetic equilibrium at the time of decoupling. The average temperature of this near-blackbody today is $\langle T \rangle \equiv T_0 = 2.72548 \pm 0.00057\,\mathrm{K}$ [52]. This temperature is not completely uniform throughout space, however; the CMB features small variations in temperature as a function of the angle on the sky,

$$\frac{\delta T}{T}(\theta, \phi) = \frac{T(\theta, \phi) - \langle T \rangle}{\langle T \rangle} \,. \tag{2.1}$$

These CMB *anisotropies* were first detected by the Cosmic Background Explorer (COBE) satellite in 1992, and in the following years, the root mean square of these fluctuations, $\left\langle (\delta T/T)^2 \right\rangle^{1/2} \simeq 1.1 \times 10^{-5}$, has been measured[5] with increasing precision [53]. This magnitude is consistent with what would be expected given our understanding of the physical processes giving rise to these anisotropies, principal among them the Sachs-Wolfe effect [54].

Because we understand the physical origin of the CMB anisotropies, it is possible to extract information about the universe's geometry, composition, and expansion from measurements of their distribution. In practice, one begins by constructing the temperature angular power spectrum. The 2-point correlation function for the temperature fluctuation at two points on the sky separated by an angle θ is

$$C(\theta) = \frac{1}{T^2} \left\langle \delta T(\hat{n})\, \delta T(\hat{n}') \right\rangle \,, \tag{2.2}$$

where $\hat{n}$ and $\hat{n}'$ are unit vectors denoting two directions on the sky with scalar product $\hat{n} \cdot \hat{n}' = \cos\theta$ and the brackets signify an average performed over all such pairs of directions with angular separation θ. By expanding the temperature anisotropies in a series of spherical

5. Note that this is the value after subtracting off the uninteresting dipole anisotropy due to the Doppler shift caused by the peculiar velocity of the Earth relative to the cosmic rest frame.

harmonics $Y_{\ell m}(\theta, \phi)$,

$$\frac{\delta T}{T}(\theta, \phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\theta, \phi)\,, \tag{2.3}$$

and making use of the completeness and orthogonality of the spherical harmonics on the unit sphere to define C_ℓ via

$$\langle a_{\ell m}\, a_{\ell' m'} \rangle = \delta_{\ell,\ell'} \delta_{m,-m'} C_\ell\,, \tag{2.4}$$

one can show that

$$C(\theta) = \sum_{\ell=0}^{\infty} \frac{(2\ell+1)}{4\pi}\, C_\ell\, P_\ell(\cos\theta)\,, \tag{2.5}$$

where $P_\ell(x)$ are the Legendre polynomials. All of the information about the distribution of CMB anisotropies is thus encoded in the C_ℓ's. The top panel of Fig. 2.1 shows these C_ℓ's (technically, it shows $\mathcal{D}_\ell = T^2 \ell(\ell+1) C_\ell / 2\pi$) as a function of multipole ℓ — the temperature angular power spectrum.

The temperature anisotropies are not the only piece of information contained in the CMB, though; the distribution of photon polarizations also encodes vital information about the state of the universe at recombination. One can perform a similar treatment to obtain the polarization angular power spectrum, whose C_ℓ's are denoted[6] C_ℓ^{EE}. Analogously, the C_ℓ's of the temperature angular power spectrum are denoted C_ℓ^{TT}. Because the temperature and E-mode polarization signals are correlated, vital information is also contained in the mixed temperature-polarization angular power spectrum, whose C_ℓ's are denoted C_ℓ^{TE}. The measured spectra of the C_ℓ^{TE}'s and the C_ℓ^{EE}'s are shown in the left and right bottom panels, respectively, of Fig. 2.1.

By precisely measuring the locations and amplitudes of the peaks, as well as ratios between peaks, one can extract cosmological parameters from these power spectra. For ex-

6. The CMB's polarization tensor can be decomposed into curl-free E-modes and divergence-free B-modes. The C_ℓ^{EE}'s correspond to the former.

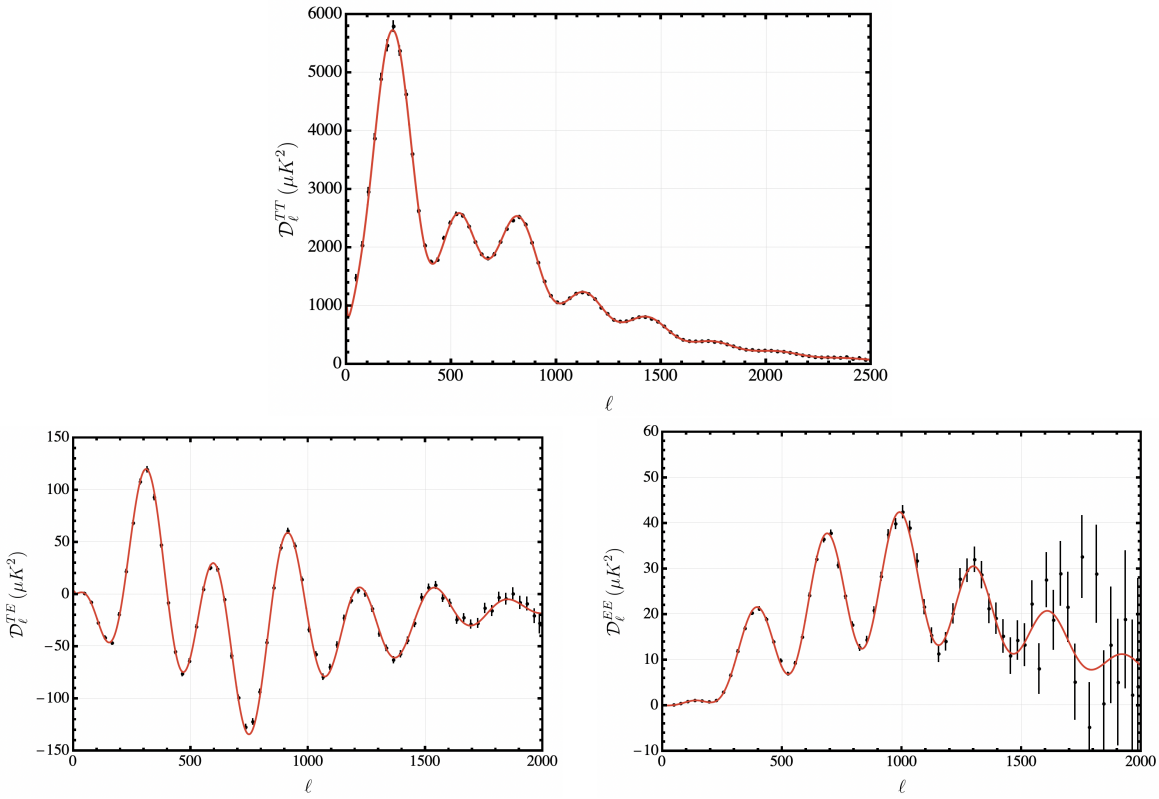

Figure 2.1: Black data points: Baseline high-ℓ Planck TT (top), TE (bottom left), and EE (bottom right) binned power spectra [1]. Note that we plot $\mathcal{D}_\ell = T^2\ell(\ell+1)C_\ell/2\pi$. Red line: Power spectra generated using Planck best fit cosmological parameters (TT,TE,EE+lowE+lensing+BAO 68% limits) [2]. Plots generated using the Cosmic Linear Anisotropy Solving System (CLASS) [3, 4].

ample, the location of the first peak in the temperature angular power spectrum is quite sensitive to curvature and has allowed us to establish that the universe is to great approximation spatially flat, $|\Omega_\kappa| \simeq 0$. The locations and amplitudes of the first few peaks, as well the ratios between odd and even peaks, has allowed us to establish the physical density of baryons $\Omega_b h^2$ and dark matter $\Omega_c h^2$ to high accuracy. See Table 2.1 for values of these and other cosmological parameters from *Planck* 2018 [2].

Parameter	Description	*Planck* 2018 Values
$\Omega_b h^2$	Physical baryon density parameter	0.02242 ± 0.00014
$\Omega_c h^2$	Physical dark matter density parameter	0.11933 ± 0.00091
$100\,\theta_{\rm MC}$	Angular size of sound horizon at recombination	1.04101 ± 0.00029
τ	Optical depth to reionization	0.0561 ± 0.0071
n_s	Scalar spectral index	0.9665 ± 0.0038
$\ln(10^{10} A_s)$	Amplitude of curvature perturbations at k_p	3.047 ± 0.014

Table 2.1: Base ΛCDM parameters fitted from *Planck* CMB TT,TE,EE+lowE power spectra in combination with lensing reconstruction and baryon acoustic oscillation (BAO) measurements [2]. Other cosmological parameters can be derived from this independent set.

Large Scale Structure

While the universe began in a remarkably homogeneous, isotropic state, it did not remain that way. Today, the universe remains homogeneous only on very large scales $\gtrsim 100\,\text{Mpc}$, with smaller scales characterized by matter clumped together in galaxy clusters, galaxies, and compact objects. In particular, baryonic matter today is largely concentrated in galaxies while dark matter resides in halos surrounding these objects. Structure formation originates in the simple fact that gravity has a tendency to collapse overdense regions, but the specifics involved in the distribution of large scale structure today reflects a much more intricate interplay between gravity, pressure, the universe's expansion and evolution through various epochs, and the properties of dark matter. As such, studying the present day distribution of large scale structure can allow us insight into the universe's expansion history and the nature of dark matter.

Due to the effects of pressure, gravitational collapse is only efficient on scales larger than the *Jeans length* λ_J, which can be expressed in terms of the Hubble rate H as

$$\lambda_J = \sqrt{\frac{8\pi^2}{3}} \frac{c_s}{H}\,, \tag{2.6}$$

where c_s is the sound speed. During radiation domination, the large value of the sound

speed (squared) $c_s^2 = 1/3$ prevented the gravitational collapse of photons and baryons on scales smaller than the Hubble radius $1/H$. Even after the onset of matter domination ($t \sim 51,000\,\mathrm{yr}$), baryonic structures were prevented from collapsing due to their tight coupling with photons, which persisted until recombination and photon decoupling ($t \sim 380,000\,\mathrm{yr}$). It was only afterwards that the formation of gravitationally collapsed baryonic structures could really begin. Dark matter, in contrast, decoupled from the photon thermal bath (if it was ever in equilibrium to begin with) much earlier. This allowed dark matter to form smaller structures at earlier times, the precise size of which depend on the dark matter mass and its temperature at decoupling. Consequently, dark matter acted as a kind of scaffolding for the growth of later structures. Observations of large scale structure suggest that dark matter is *cold*, with velocities much smaller than the speed of light. Cold dark matter forms structure hierarchically, with structures forming on smaller scales first and large scale structures formed by the continuous merger of such objects.

The large scale structure observed today was seeded by tiny primordial density perturbations δ in the early universe. In the Newtonian approximation (valid on subhorizon scales), the evolution of these density perturbations in an expanding universe is governed by

$$\ddot{\delta} + 2H\dot{\delta} + \frac{4\pi\bar{\rho}}{M_{\mathrm{Pl}}^2}\left[\left(\frac{\lambda_J}{\lambda_p}\right)^2 - 1\right]\delta = 0\,, \tag{2.7}$$

where $\bar{\rho}$ is the background energy density and λ_p is the proper wavelength of the perturbation mode under consideration. Note the critical presence of the "Hubble friction" term $2H\dot{\delta}$, which stems from the expansion of the universe. Without this term, which slows the growth of perturbations, modes with $\lambda_p \gg \lambda_J$ would experience unphysical exponential growth. Meanwhile, modes with $\lambda_p \ll \lambda_J$ exhibit oscillatory rather than growing behavior.

More generally for a universe consisting of multiple components, the evolution of a density

perturbation of component i evolves as

$$\ddot{\delta}_i + 2H\dot{\delta}_i + \frac{4\pi\bar{\rho}}{M_{\rm Pl}^2}\left(\frac{\lambda_{J,i}}{\lambda_p}\right)^2 \delta_i - \frac{4\pi\bar{\rho}}{M_{\rm Pl}^2}\sum_j \epsilon_j \delta_j = 0\,, \tag{2.8}$$

where ϵ_j is the fractional contribution of component j to the total energy density. We have already argued that for baryonic matter during radiation domination, λ_J is too large for modes to grow on subhorizon scales. For dark matter perturbations $\delta_{\rm DM}$ however, which do not have to contend with this pressure, the solution during radiation domination is[7]

$$\delta_{\rm DM} \simeq A + B\ln t\,, \tag{2.9}$$

indicating that subhorizon perturbations grow logarithmically slowly during radiation domination. During matter domination, on the other hand,

$$\delta_{\rm DM} \simeq At^{2/3} + Bt^{-1}\,. \tag{2.10}$$

The first term (the "growing mode") scales as $\delta_{\rm DM} \propto t^{2/3}$, and so density perturbations grow much more quickly during matter domination. In fact, over the duration of matter domination, these perturbations can increase by a factor $\sim 10^3$, leading to the large scale structure observed today. As for superhorizon perturbations, solving for their evolution requires a full general relativistic treatment, so here we simply cite the surprising result that

$$\delta_{\rm superhorizon} \sim \begin{cases} t & \text{radiation domination}\,, \\ t^{2/3} & \text{matter domination}\,. \end{cases} \tag{2.11}$$

That is, even during radiation domination, superhorizon perturbations grow rapidly — much more rapidly than the logarithmic growth of subhorizon modes during radiation domination,

7. Throughout this subsection we work in Newtonian gauge.

and even faster than their $\sim t^{2/3}$ growth during matter domination.

Finally, we briefly comment on experimental data. Information about the statistical distribution of our universe's matter is encoded in the matter power spectrum, which describes the density contrast δ as a function of scale. From the matter power spectrum, one can extract cosmological parameters as well as information about the geometry of spacetime, dark matter, dark energy, and our universe's expansion history. One way to measure the matter power spectrum is through galaxy surveys, or more generally redshift surveys like the Sloan Digital Sky Survey (SDSS) [55, 56]. The matter power spectrum can also be inferred from observations of the CMB, since this encodes information about the distribution of the density perturbations that seeded structure formation. The matter power spectrum follows in a predictable way from the spectrum of primordial density perturbations, whose large scale behavior can be constrained by CMB measurements. In particular, the primordial power spectrum has been measured to be approximately scale invariant on large scales, a fact which we will return to when we discuss inflation.

2.1.2 ΛCDM Cosmology

The ΛCDM model is the simplest mathematical model of Big Bang cosmology that can account for the wealth of data described in the previous section — that is, the CMB, the observed light element abundances, large scale structure, and the late-time accelerated expansion of the universe. It emerged in the late 1990s amongst a confusion about the energy makeup of the universe as a *concordance cosmology*. ΛCDM posits a universe of 3 major components: 1) dark energy, in the form of the cosmological constant Λ; 2) cold dark matter (CDM); and 3) ordinary baryonic matter. These account for 69%, 26%, and 5%, respectively, of the universe's energy budget [2]. Like essentially any modern cosmological theory, ΛCDM presumes[8] the cosmological principle — the notion that on sufficiently large scales,

8. Recent discoveries of mega-structures like the Big Ring [57], the Giant Arc [58], and the Hercules-Corona Borealis Great Wall [59] challenge the statistical expectations from the cosmological principle.

the universe's matter distribution is isotropic (looks the same in all directions) and homogeneous (looks the same from all locations). This is a powerful starting principle, allowing for a significant simplification of the mathematical formulas describing the universe's evolution. ΛCDM also takes general relativity as a starting point, as well as the observational fact that the universe is expanding. Before detailing the model further, we quickly review the Friedmann equations and the evolution of various energy density components in an expanding universe.

The Friedmann Equations

The most general spacetime metric consistent with a homogeneous, isotropic, expanding (or contracting) universe is given by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric $g_{\mu\nu}$ corresponding to the line element $ds^2 = g_{\mu\nu}dx^\mu dx^\nu$, with[9]

$$ds^2 = -dt^2 + a(t)^2 \left[\frac{dr^2}{1-\kappa r^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right] . \tag{2.12}$$

The scale factor $a(t)$ describes how space expands or contracts with time and the coefficient κ can take values $+1$, 0, or -1 depending on whether spatial slices have positive, vanishing, or negative curvature, respectively. The time evolution of the scale factor can be obtained by solving the Friedmann equations, which are derived from the Einstein field equations with FLRW metric and isotropic stress-energy tensor. Explicitly, the Einstein equations read

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi}{M_{\rm Pl}^2} T_{\mu\nu} \,, \tag{2.13}$$

where $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$ is the Einstein tensor, defined in terms of the Ricci tensor $R_{\mu\nu}$ and Ricci scalar $R = g^{\mu\nu}R_{\mu\nu}$, Λ is the cosmological constant, $M_{\rm Pl} = 1.22 \times 10^{19}$ GeV is the Planck mass, and $T_{\mu\nu}$ is the stress-energy tensor. Taking the metric $g_{\mu\nu}$ to have the FLRW

9. I work with metric signature $(-,+,+,+)$.

form of Eq. (2.12), it follows that the non-vanishing components of the Ricci tensor are

$$R_{00} = -3\frac{\ddot{a}}{a}\,, \;\; R_{ii} = \left(\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} + 2\frac{\kappa}{a^2}\right) g_{ii}\,, \tag{2.14}$$

while the Ricci scalar is

$$R = 6\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{\kappa}{a^2}\right)\,, \tag{2.15}$$

where the overhead dots denote derivatives with respect to time[10], "0" is the temporal index, and "i" are spatial indices. Meanwhile, the assumptions of isotropy and homogeneity demand that the stress-energy tensor take the perfect fluid form,

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}\,, \tag{2.16}$$

where ρ and p are the energy density and pressure, respectively, of the fluid, and u^μ is the fluid 4-velocity, which in the local rest frame of a given spacetime point takes the form $u^\mu = (1, 0, 0, 0)$. This implies that the non-vanishing components of the covariant form of the stress-energy tensor are

$$T_{00} = \rho\,, \;\; T_{ii} = pg_{ii}\,. \tag{2.17}$$

Combining Eqs. (2.14), (2.15), and (2.17), the "00" component of the Einstein equation yields the first Friedmann equation

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2} = \frac{8\pi\rho}{3M_{\rm Pl}^2} + \frac{\Lambda}{3}\,. \tag{2.18}$$

10. The time coordinate t is called cosmic time, and it has the property that at each moment in time, the universe has the same density everywhere. It is the homogeneity property of the universe that makes the definition of such preferred time coordinate possible.

Meanwhile, the "ii" component, combined with the equation above, yields the second Friedmann equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3M_{\rm Pl}^2}(\rho + 3p) + \frac{\Lambda}{3}\,. \tag{2.19}$$

Going forward, we will set $\kappa = 0$, in accordance with the observational fact that our universe seems to have negligible curvature.

Finally, we remark that by placing the cosmological constant on the right-hand side of these equations, we take the view that it should be regarded as a source term, with stress-energy $T^{\Lambda}_{\mu\nu} = -\frac{M_{\rm Pl}^2}{8\pi}\Lambda g_{\mu\nu}$. This corresponds to a perfect fluid with energy density

$$\rho_\Lambda = \frac{M_{\rm Pl}^2}{8\pi}\Lambda\,, \tag{2.20}$$

and pressure

$$p_\Lambda = -\rho_\Lambda\,. \tag{2.21}$$

Going forward, we will include the contribution from the cosmological constant as part of the total stress-energy tensor, $T^{\mu\nu} = T^{\mu\nu}_{\rm matter} + T^{\mu\nu}_{\Lambda}$.

The Energy Content

Conservation of the stress-energy tensor in an expanding universe corresponds to

$$\nabla_\mu T^{\mu\nu} = \partial_\mu T^{\mu\nu} + \Gamma^\mu_{\mu\alpha}T^{\alpha\nu} + \Gamma^\nu_{\mu\alpha}T^{\mu\alpha} = 0\,, \tag{2.22}$$

where the second expression is the definition of the action of the covariant derivative ∇_μ and $\Gamma^\mu_{\alpha\beta} = \frac{1}{2}g^{\mu\nu}\left(\partial_\beta g_{\alpha\nu} + \partial_\alpha g_{\beta\nu} - \partial_\nu g_{\alpha\beta}\right)$ are the Christoffel symbols. Looking at the $\nu = 0$ component and using the fact that $\Gamma^0_{ii} = \frac{\dot{a}}{a}g_{ii}$ and $\Gamma^i_{0i} = \frac{\dot{a}}{a}$, we arrive at the equation

governing the evolution of the energy density

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0\,. \tag{2.23}$$

Eqs. (2.18), (2.19), and (2.23) describe the dynamics of an evolving FLRW universe. In order to use this system of equations, though, we need to know about the energy density and pressure of the universe's various components. For a particle species with energy E, mass m, and internal degrees of freedom g in kinetic equilibrium at temperature T and chemical potential μ, the energy density is given by

$$\rho = \frac{g}{2\pi^2}\int_m^\infty dE\,\frac{E^2\sqrt{E^2-m^2}}{e^{(E-\mu)/T}\mp 1}\,, \tag{2.24}$$

where the $-$ $(+)$ in the phase space distribution function is for bosons (fermions). A similar expression yields the pressure

$$p = \frac{g}{6\pi^2}\int_m^\infty dE\,\frac{(E^2-m^2)^{3/2}}{e^{(E-\mu)/T}\mp 1}\,. \tag{2.25}$$

For highly relativistic particles with $T \gg m, \mu$, which we will refer to as *radiation*, these integrals can be evaluated explicitly as

$$\rho_{\rm rad} = \begin{cases} \frac{\pi^2}{30}gT^4 & (\text{boson})\,, \\ \frac{7}{8}\frac{\pi^2}{30}gT^4 & (\text{fermion})\,, \end{cases} \tag{2.26}$$

and

$$p_{\rm rad} = \begin{cases} \frac{\pi^2}{90}gT^4 & (\text{boson})\,, \\ \frac{7}{8}\frac{\pi^2}{90}gT^4 & (\text{fermion})\,. \end{cases} \tag{2.27}$$

This tells us that the *equation of state* $w \equiv p/\rho$ for radiation is $w_{\rm rad} = 1/3$. Returning to

Eq. (2.23), it also tells us that the energy density of a relativistic species evolves as

$$\dot{\rho}_{\rm rad} + 4\frac{\dot{a}}{a}\rho_{\rm rad} = 0\,. \tag{2.28}$$

This in turns implies that the quantity $\rho_{\rm rad}a^4$ is conserved in time

$$\frac{\partial}{\partial t}(\rho_{\rm rad}a^4) = 4\rho_{\rm rad}a^3\dot{a} + \dot{\rho}_{\rm rad}a^4 = 4\rho_{\rm rad}a^3\dot{a} - 4\rho_{\rm rad}a^3\dot{a} = 0\,. \tag{2.29}$$

The energy density of radiation thus dilutes as four powers of the scale factor,

$$\rho_{\rm rad} \propto a^{-4}\,. \tag{2.30}$$

Three of these can be understood as originating from geometric dilution while the last comes from cosmological redshift. In cosmological contexts, particularly in the hot, dense conditions of the early universe, it is useful to consider an entire ensemble of relativistic particle species. The energy density of such an ensemble is given by summing over those of the components $\rho = \sum_i \rho_i$, or

$$\rho = \frac{\pi^2}{30}g_\star(T)T^4\,. \tag{2.31}$$

The quantity $g_\star(T)$, the so-called *number of relativistic degrees of freedom*, is defined as

$$g_\star(T) = \sum_B g_B\left(\frac{T_B}{T}\right)^4 + \frac{7}{8}\sum_F g_F\left(\frac{T_F}{T}\right)^4\,, \tag{2.32}$$

with T the temperature of the photon bath, T_i the temperature of a given species, B = bosons, and F = fermions. For a bath in kinetic equilibrium ($T_i = T$ for all i), $g_\star$ is simply the sum of all bosonic particle degrees of freedom and all fermionic particle degrees of freedom weighted by 7/8. Species not in equilibrium with the rest of the bath have a diminished contribution to $g_\star$. As the universe expanded and cooled and different Standard

Model species became now relativistic, the number of relativistic degrees of freedom evolved from an initial value of $g_\star(T \gg 100\,\mathrm{GeV}) = 106.75$ to a final value today of $g_\star(T_0) \simeq 3.36$.

Now consider the opposite limit $T \ll m$ of a highly non-relativistic species (henceforth referred to as *matter*). In this case, the integrals for energy density and pressure evaluate to

$$\rho_\mathrm{m} = g m^{5/2} \left(\frac{T}{2\pi}\right)^{3/2} e^{-(m-\mu)/T}\,, \tag{2.33}$$

and

$$p_\mathrm{m} = g T^{5/2} \left(\frac{m}{2\pi}\right)^{3/2} e^{-(m-\mu)/T}\,. \tag{2.34}$$

These can be written in terms of the number density of the non-relativistic species n_m as $\rho_\mathrm{m} = n_\mathrm{m} m$ and $p_\mathrm{m} = n_\mathrm{m} T$. Note that since $m \gg T$ for such a species, $\rho_\mathrm{m} \gg p_\mathrm{m}$ and so pressure can usually be neglected[11] $p = 0$. Thus matter is characterized by the equation of state $w_\mathrm{m} = 0$. From Eq. (2.23), the energy density of a non-relativistic species then evolves as

$$\dot{\rho}_\mathrm{m} + 3\frac{\dot{a}}{a}\rho_\mathrm{m} = 0\,. \tag{2.35}$$

This implies the comoving energy density $\rho_\mathrm{m} a^3$ is conserved in time

$$\frac{\partial}{\partial t}(\rho_\mathrm{m} a^3) = 3\rho_\mathrm{m} a^2 \dot{a} + \dot{\rho}_\mathrm{m} a^3 = 3\rho_\mathrm{m} a^2 \dot{a} - 3\rho_\mathrm{m} a^2 \dot{a} = 0\,. \tag{2.36}$$

Thus the energy density of matter dilutes as three powers of the scale factor,

$$\rho_\mathrm{m} \propto a^{-3}\,. \tag{2.37}$$

This can be understood as simply a geometric effect as the volume of the universe expands.

Finally before moving on, we note that the result of considering Eq. (2.23) for arbitrary

11. Pressureless matter is also often referred to as *dust*.

equation of state w is the general scaling

$$\rho \propto a^{-3(1+w)} \,. \tag{2.38}$$

In particular for the case of the cosmological constant with $w_\Lambda = -1$, which follows from Eq. (2.21), this yields

$$\rho_\Lambda = \text{constant} \,. \tag{2.39}$$

The fact that ρ_Λ remains constant even as spacetime expands and other components dilute is why the cosmological constant is often considered to be a kind of *vacuum energy*.

The Expansion History

It is customary to recast the first Friedmann equation Eq. (2.18) in terms of the Hubble rate $H \equiv \dot{a}/a$,

$$H^2 = \frac{8\pi\rho}{3M_{\rm Pl}^2} \,, \tag{2.40}$$

which can be understood as the rate at which the universe is expanding. Note that the total energy density appearing on the right-hand side includes contributions from radiation, non-relativistic matter, and the cosmological constant: $\rho = \rho_{\rm rad} + \rho_{\rm m} + \rho_\Lambda$. It is often convenient to parameterize the contributions from the various components ρ_i in terms of their fractional energy densities today $\Omega_{i,0} \equiv \rho_{i,0}/\rho_{\rm crit,0}$, where here the subscript "0" denotes present day quantities. These are normalized with respect to the critical density today $\rho_{\rm crit,0}$, defined as the value required in order that the universe be flat ($\kappa = 0$),

$$\rho_{\rm crit,0} = \frac{3M_{\rm Pl}^2}{8\pi} H_0^2 \,. \tag{2.41}$$

The Hubble rate today H_0 is called the *Hubble constant*, and its value remains hotly contested. Determinations relying on early-time probes like the CMB (and thereby tacitly

the assumption of standard ΛCDM cosmology) consistently predict values for H_0 which are markedly lower than those determined from local measurements of distances and redshifts. For example, the early-time prediction from *Planck* 2018 at 68% confidence level (CL) is $H_0 = 67.4 \pm 0.5\,\mathrm{km\,s^{-1}Mpc^{-1}}$ [2]. In contrast, the most recent local measurement from the SH0ES team (R20) finds $H_0 = 73.2 \pm 1.3\,\mathrm{km\,s^{-1}Mpc^{-1}}$ at 68% CL [60]. The 4σ to 6σ disagreement between early- and late-time measurements of H_0 is the famous *Hubble tension* — one of the most pressing crises in cosmology today.

In terms of $\rho_{\mathrm{crit},0}$ and $\Omega_{i,0}$, the total energy density at arbitrary time can be written

$$\rho = \rho_{\mathrm{crit},0}\left(\Omega_{\mathrm{m},0}\, a^{-3} + \Omega_{\mathrm{rad},0}\, a^{-4} + \Omega_{\Lambda,0}\right), \tag{2.42}$$

where we have also taken the value of the scale factor today to be $a_0 = 1$. The present day values of the fractional abundances are $\Omega_{\mathrm{m},0} = 0.315$, $\Omega_{\mathrm{rad},0} = 9.26\times10^{-5}$, and $\Omega_{\Lambda,0} = 0.685$ [2]. Clearly our universe is dark energy dominated today, but before that it was matter dominated and even farther back it was radiation dominated. To obtain the times when these transitions took place and thereby the expansion history of the universe, we can solve the Friedmann equation Eq. (2.18) with energy density Eq. (2.42),

$$H = H_0\sqrt{\Omega_{\mathrm{m},0}\, a^{-3} + \Omega_{\mathrm{rad},0}\, a^{-4} + \Omega_{\Lambda,0}}\,. \tag{2.43}$$

First to build some intuition, consider a single component universe. In the case of a radiation dominated universe, $\rho \propto a^{-4}$ by Eq. (2.30), and so solving the Friedmann equation yields $a(t) \propto t^{1/2}$ and therefore $H = \frac{1}{2t}$. For a matter dominated universe, $\rho \propto a^{-3}$ by Eq. (2.37), and so $a(t) \propto t^{2/3}$ and $H = \frac{2}{3t}$. Finally for a dark energy dominated universe, $\rho =$ constant by Eq. (2.39), leading to $a(t) = e^{Ht}$. These rough scalings are reflected in Fig. 2.2, which was obtained by numerically solving Eq. (2.43) for $a(t)$. The point t_0 marks

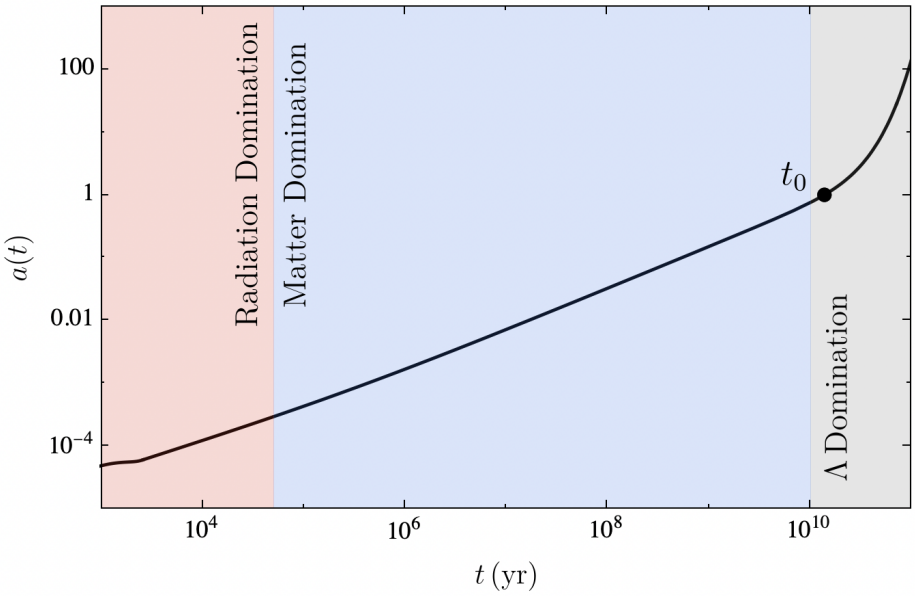

Figure 2.2: Evolution of scale factor $a(t)$ obtained by solving the Friedmann equations with $\Omega_{\mathrm{m},0} = 0.315$, $\Omega_{\mathrm{rad},0} = 9.26 \times 10^{-5}$, $\Omega_{\Lambda,0} = 0.685$, and $H_0 = 67.4\,\mathrm{km\,s^{-1}Mpc^{-1}}$.

the present day age of the universe, which was found by integrating

$$t_0 = \int_0^1 \frac{da}{a\sqrt{\Omega_{\mathrm{m},0}\,a^{-3} + \Omega_{\mathrm{rad},0}\,a^{-4} + \Omega_{\Lambda,0}}} \simeq 13.8\,\mathrm{Gyr}\,, \tag{2.44}$$

where we have assumed the Planck value of $H_0 = 67.4\,\mathrm{km\,s^{-1}Mpc^{-1}}$. By using the solution for $a(t)$ and equating the appropriate energy components, we can also determine that the transition between radiation and matter domination occurred around $51,000$ years after the Big Bang and that the transition between matter and dark energy domination occurred much more recently, around 1.0×10^{10} years after the Big Bang. Note that at present we are currently living in an era of dark energy domination, with the negative pressure associated with the cosmological constant driving an accelerating expansion rate.

Initial Conditions

The ΛCDM model describes a flat universe which is both homogeneous and isotropic on large scales, in accordance with observational data. Although it is not clear how these seemingly fine-tuned initial conditions were established, perhaps the most commonly accepted explanation[12] is that at some point in the first fractions of a second following the Big Bang, the universe underwent a period of exponential expansion — *inflation* [61, 62].

In order to drive this exponential expansion, the universe must have been dominated by some substance with approximately constant energy density, corresponding to an equation of state $w \approx -1$. Note that the comoving cosmological horizon $R_H = (aH)^{-1}$ generically evolves as

$$R_H \propto a^{(1+3w)/2} \,. \tag{2.45}$$

During such a period with $w = -1$, it would shrink as $R_H \propto a^{-1}$, such that at sufficiently early times the entire observable universe could have been in causal contact. This would explain the homogeneity of the observable universe, providing an explanation for why regions which naively appear to have been casually disconnected nevertheless look the same and thereby resolving the horizon problem. It also resolves the flatness problem; even if the universe initially had some degree of curvature, this would dilute as $|\Omega_\kappa| \propto a^{-2}$ during inflation, leading to the apparent flatness observed today.

There remains the question of how to obtain such an equation of state. We cannot simply posit a cosmological constant-like term at early times, since the universe needs to transition to radiation and eventually matter domination at later times. Instead, we need a transient source that will eventually decay to reheat the thermal Standard Model bath. A

12. Inflation also provides a resolution to the "magnetic monopole problem", or more generically the "heavy relics problem". Many grand unified theories predict the copious production of heavy, stable particles or topological defects like magnetic monopoles which are not observed in nature. This could be explained by a period of exponential expansion following their production, which would serve to drastically lower the density of such objects.

dynamical mechanism to generate such a term is to invoke a slowly evolving scalar field — the *inflaton* ϕ. As the inflaton slowly rolls down its nearly flat potential, its approximately constant energy density behaves as an effective cosmological constant. More quantitatively, the energy density and pressure of a canonical scalar field ϕ can be arranged to obtain the equation of state

$$w_\phi = \frac{p_\phi}{\rho_\phi} = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)}\,. \tag{2.46}$$

The limit in which the scalar field is dominated by its potential rather than kinetic term $V \gg \frac{1}{2}\dot{\phi}^2$ then corresponds to a de Sitter-like period of exponential expansion with $w_\phi \simeq 1$. Inflation ends when the inflaton gathers speed and rolls down its potential, entering into a series of oscillations about the minimum of its potential. Eventually when its decay rate is of order the Hubble rate $\Gamma_\phi \simeq H$, the inflaton decays to reheat[13] the universe.

During inflation, tiny quantum fluctuations — either from the inflaton itself or from any other light spectator fields present — are stretched to superhorizon scales to become the classical density perturbations needed to seed the structure observed today. Most viable theories of inflation predict a spectrum of primordial density perturbations which is approximately scale invariant, consistent with large scale observations e.g. of the CMB [64, 65, 66, 67]. It is customary to parameterize the scalar power spectrum of primordial perturbations $\mathcal{P}_\mathcal{R}(k)$ as

$$\mathcal{P}_\mathcal{R}(k) = A_s \left(\frac{k}{k_p}\right)^{n_s - 1}\,, \tag{2.47}$$

where k_p is the CMB pivot scale, A_s is the amplitude measured at the pivot scale, and n_s is the scalar spectral index. In particular, $n_s = 1$ corresponds to a scale invariant power spectrum. Presuming ΛCDM and taking the pivot scale as $k_p = 0.05\,\mathrm{Mpc}^{-1}$, the Planck collaboration reports $A_s = (2.10 \pm 0.10) \times 10^{-9}$ and $n_s = 0.965 \pm 0.004$ [2]. The latter

13. Reheating can also occur in a non-perturbative manner during a stage of parameteric resonance called *preheating* [63]. Intuitively, quantum fluctuations of fields coupled to the coherently oscillating classical inflaton condensate can experience resonant amplification, leading to an exponential growth in their occupation numbers.

is consistent with the nearly[14] scale invariant spectrum predicted by inflation, though it features slightly more power on large scales.

Though it does face some criticisms and alternatives have been proposed, for the most part the inflationary paradigm is now widely accepted. However, its precise particle physics realization remains an open question, with many viable single- and multi-field models having been proposed. The qualitative and quantitative nature of the inflaton potential will likely not be resolved until we are able to resolve the small scale features of the primordial power spectrum, a topic which will be discussed in more detail in Chapter 3.

2.1.3 Summary

Since its emergence in the late 1990s, the ΛCDM model has remained steadfast as the standard cosmological paradigm. It is a model of Big Bang cosmology predicated on the cosmological principle and general relativity, and describes a universe which is approximately homogeneous and isotropic on large scales, spatially flat, and expanding. Mathematically, such a spacetime is described by the FLRW metric with $\kappa = 0$. The universe's expansion history is captured by the scale factor $a(t)$, whose time evolution is governed by the Friedmann equations. These in turn depend on the energy content of the universe. ΛCDM posits that the universe today is dominated by the following three components,

- dark energy (Λ) with $\Omega_\Lambda^0 \simeq 0.689$,
- cold dark matter with $\Omega_{\text{CDM}}^0 \simeq 0.262$,
- ordinary baryonic matter with $\Omega_b^0 \simeq 0.049$,

in addition to a sub-dominant radiation component. These constituents have traded off in their relative contributions to the universe's energy budget as it expanded and evolved.

14. The evolution of the inflaton leads to a departure from a strictly scale invariant spectrum, tilting n_s towards values slightly less than 1.

Following the Big Bang, potentially a period of inflation, and the subsequent reheating, the universe progressed through eras of radiation domination, matter domination, and now finally dark energy domination (in the form of the cosmological constant Λ) — see Fig. 2.2. The rate of expansion throughout these eras is described by the Hubble rate H, whose present day value $H_0 = (67 - 73)\,\mathrm{km\,s^{-1}Mpc^{-1}}$ remains a subject of intense debate. Based on this picture, the age of the universe is estimated to be $t_0 \simeq 13.8\,\mathrm{Gyr}$.

Remarkably, standard ΛCDM requires as input only six independent parameters which must be measured in order to completely specify the model; all other parameters can readily by derived from these six. One particularly convenient set[15] is listed in Table 2.1 below. With just this minimal input, ΛCDM is able to precisely account for the light element abundances, the observed patterns of the CMB, large scale structure, and the late-time accelerated expansion of the universe. In spite of these numerous successes, the standard ΛCDM cosmological paradigm leaves a number of open questions. Since these questions share considerable overlap with the outstanding puzzles in particle physics, we defer their discussion to Sec. 2.3, after we review the structure and status of the Standard Model.

2.2 The Standard Model of Particle Physics

In contrast to cosmology, which remained a data-starved field until only very recently, particle physics was already enjoying its golden age in the 1900s. The birth of modern particle physics can be traced back to J.J. Thomson's discovery of the electron e^- in 1897 [68]. Almost two decades later, Rutherford's famous scattering experiments established evidence of the proton [69]. These were organized into the first modern picture of the atom by Bohr, who was able to use his model to correctly predict the emission spectrum of hydrogen [70, 71]. Finally with Chadwick's 1932 discovery of the neutron [72], all of the ingredients necessary for our

15. Other popular choices include swapping out the angular size of the sound horizon at recombination θ_{MC} for either the dark energy density Ω_Λ^0 or the age of the universe t_0.

classical picture of ordinary matter were in place.

Against this backdrop of experimental triumphs, the modern theory of quantum mechanics was being born. The first steps were laid in 1900 when Max Planck introduced quanta in order to explain blackbody radiation [73], a feat for which he received the Nobel prize not even two decades later. In 1905, Einstein took this a step further and argued that quantization was a property of the electromagnetic field itself, a fact which he used to explain the photoelectric effect [74] (another Nobel prize-worthy endeavor). Finally in 1923, Compton's scattering experiments revealed that light behaved as a particle in addition to a wave [75], and thus the photon, now understood to be the mediator of the electromagnetic force, was discovered. About a decade later, the Japanese physicist Yukawa proposed the first tenable theory of the strong force, responsible for binding nucleons. Because this force was short-ranged, he reasoned that it would have to be mediated by a massive particle, which he named the meson (meaning "middle-weight") [76].

A few years later saw the first steps towards a relativistic quantum mechanics, with the introduction of the Klein-Gordon equation in 1926 and the Dirac equation in 1928 [77]. Curiously, the latter admitted both positive and negative energy solutions, now understood to correspond to particles and *antiparticles*, respectively. The positron e^+, the antiparticle of the electron, was discovered by Anderson just four years later in 1931 [78]. The next few years also saw the prediction of Pauli's neutrino and the Fermi theory of β decay, but for the most part the *leptonic* sector (i.e. the sector of particles which do not interact via the strong force) was mostly under control by the 1950s, barring some confusion about neutrinos.

The *hadronic* sector, consisting of particles interacting via the strong force, was another story. Yukawa's meson (the *pion*) turned out not to be the only meson; in 1947, the neutral kaon was discovered, followed by a menagerie of other mesons and *baryons* (which are fermions, in contrast to the bosonic mesons). This onslaught of new hadronic particles was precipitated by the introduction of modern particle accelerators, starting with the

Brookhaven Cosmotron in 1952. By the 1960s, hadronic physics was in a state of great confusion — a whole zoo of particles had been discovered with seemingly no organizational principle. Soon, though, Murray Gell-Mann discovered a way to arrange the mesons and baryons into geometric patterns (octets and decuplets) which bespoke a deeper structure: the *Eightfold Way* [79]. From this organizational principle came the quark model [80], wherein it was understood that hadrons are composed of more fundamental constituents called *quarks*. In particular, mesons are composed of quark-antiquark pairs and baryons are composed of triplets of quarks. The fact that these quarks always appeared confined in hadrons was first deeply mysterious, but was eventually understood once it was realized that the strong force is described mathematically by the non-Abelian gauge theory $SU(3)_c$ [81].

The next decade came with another surprise: at high energies (and therefore at high temperatures in the early universe), the electromagnetic and weak forces are actually unified in a single *electroweak theory* described by the gauge group $SU(2)_L \times U(1)_Y$. Glashow, Salam, and Weinberg (GSW) were awarded the 1979 Nobel prize for their efforts in this direction [82, 83, 84]. The theory was validated in 1983 with the discovery of the weak charged $W^{\pm}$ and neutral Z bosons [85, 86, 87, 88], whose masses matched the prediction of GSW. The spontaneous breaking of electroweak symmetry $SU(2)_L \times U(1)_Y \to U(1)_{\mathrm{EM}}$ and the generation of masses for the weak gauge bosons proceeds through the *Higgs mechanism*, independently postulated by three different groups in 1964: Higgs [89]; Brout & Englert [90]; and Guralnik, Hagen, & Kibble [91]. This mechanism required the introduction of a new field — a scalar $SU(2)$ doublet which develops a vacuum expectation value. Though the experimental confirmation of the Higgs mechanism via the observation of the Higgs boson, a byproduct of the spontaneous symmetry breaking process, at the Large Hadron Collider did not come until 2012 [92], it is safe to say that all the ingredients for the Standard Model in its modern incarnation were already in place by the end of the century.

In Fig. 2.3, we present an overview of the fields of the Standard Model and their classi-

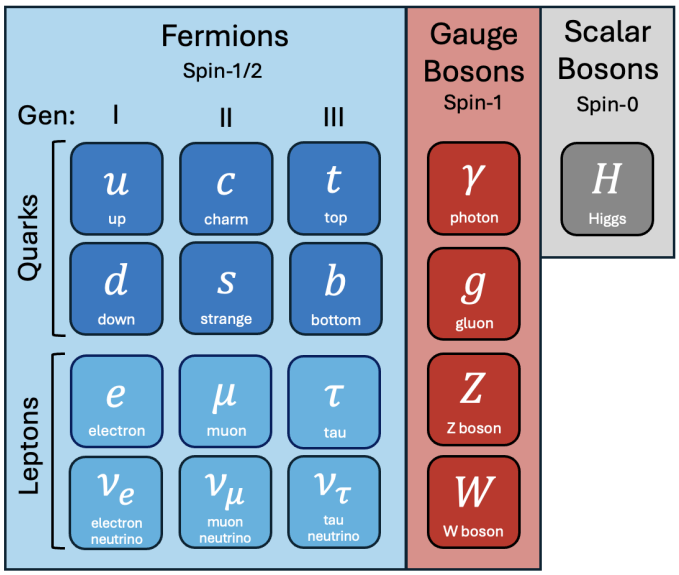

Figure 2.3: An overview of the particle content of the Standard Model. The spin-1/2 fermions constitute the matter sector and come in three generations. They include the quarks, which are charged under the strong force and exist confined in hadrons, and the leptons, which are not charged under $SU(3)$. The spin-1 vector bosons constitute the gauge sector and include photons (the mediators of the electromagnetic force), 8 gluons (the mediators of the strong force), and the neutral Z and charged $W^{\pm}$ bosons (the mediators of the weak force). Finally, the Higgs sector consists of the spin-0 Higgs boson, which plays a vital role in electroweak symmetry breaking.

fications. There are three main classes of particles: 1) **the gauge sector**, consisting of the spin-1 vector bosons mediating the fundamental forces; 2) **the matter sector**, consisting of the spin-1/2 fermions which constitute ordinary matter; and 3) **the Higgs sector**, containing the spin-0 Higgs boson which is the byproduct of spontaneous symmetry breaking. We will elucidate each of the above in the coming subsections, after first reviewing the gauge structure of the Standard Model.

2.2.1 Gauge Invariance

Underlying the seemingly random collection of particles in Fig. 2.3 is a deep mathematical structure at the heart of the Standard Model — *gauge symmetry*. More precisely, the

Standard Model is a spontaneously broken, non-Abelian gauge theory based on the group

$$SU(3) \times SU(2) \times U(1)\,. \tag{2.48}$$

Under these symmetries, the spin-1/2 fermions of the matter sector and the spin-0 Higgs transform in the fundamental representation while the spin-1 gauge bosons transform in the adjoint. To unpack what all this means, we will review the concept of gauge symmetry in the context of the simplest example: the Abelian group $U(1)$. See for example Refs. [93, 94, 95] for further review.

$U(1)$ Gauge Symmetry

Suppose we would like to write down the Lagrangian $\mathcal{L}$ for a theory of fermions $\psi(x)$ which is invariant under $U(1)$, the unitary group of degree 1. In general, the unitary group of degree N, $U(N)$, is the group of all unitary $N \times N$ matrices (with matrix multiplication as the group operation). Recall that an invertible, complex square matrix U is said to be unitary if $U^\dagger U = UU^\dagger = \mathbb{I}$, where the dagger denotes the conjugate transpose and $\mathbb{I}$ is the identity matrix. The group $U(1)$ is then simply the group of all complex numbers with norm 1 (under multiplication). It is convenient to parameterize these group elements as $e^{i\alpha}$, with $\alpha \in [0, 2\pi)$ labeling the elements of the group. Geometrically this corresponds to the circle, with multiplication among the group elements corresponding to rotation.

In physics contexts, there are two options with regard to the parameter α: When α is a constant independent of the position in spacetime, the symmetry is said to be *global*. On the other hand when $\alpha(x)$ depends on spacetime position x^μ, then the symmetry is said to be a *gauge symmetry*. Such symmetries are also called local symmetries, since the transformation depends on one's local position in spacetime. In a sense, they are not symmetries at all, but rather *redundancies in description*, indicating that certain states should be identified.

Returning to the task at hand, the statement that our theory should be invariant under a gauge $U(1)$ means that it should be invariant under transformations of the fermion fields of the form

$$\psi(x) \to e^{i\alpha(x)}\psi(x)\,. \tag{2.49}$$

It is immediately clear that mass terms like $m\bar{\psi}\psi$ have this property, since the complex phases cancel between the original and conjugate fields

$$m\bar{\psi}\psi \to m\bar{\psi}\psi\,. \tag{2.50}$$

Terms containing derivatives like the kinetic term $\bar{\psi}i\gamma^\mu\partial_\mu\psi \equiv \bar{\psi}i\partial\!\!\!/\psi$, however, pose a problem, since

$$\bar{\psi}i\partial\!\!\!/\psi \to \bar{\psi}i\partial\!\!\!/\psi - \bar{\psi}\gamma^\mu\psi\partial_\mu\alpha\,. \tag{2.51}$$

In order to write down a fermion kinetic term which is invariant, it is necessary to construct a generalization of the derivative called the *covariant derivative* D_μ. To cancel off the unwanted additional term, this should take the form

$$D_\mu = \partial_\mu + igA_\mu\,, \tag{2.52}$$

where we have introduced a new vector field A_μ defined by its transformation property

$$A_\mu \to A_\mu - \frac{1}{g}\partial_\mu\alpha\,. \tag{2.53}$$

One can show that the new kinetic term $\bar{\psi} i \not{D} \psi$ is now invariant under $U(1)$,

$$\begin{aligned} \bar{\psi} i \not{D} \psi = \bar{\psi} i \gamma^\mu (\partial_\mu + igA_\mu)\psi \rightarrow & \bar{\psi} e^{-i\alpha(x)} i\gamma^\mu (\partial_\mu + igA_\mu - i\partial_\mu \alpha) e^{i\alpha(x)} \psi \\ & = \bar{\psi} i \gamma^\mu (\partial_\mu \psi + igA_\mu \psi + i\psi \partial_\mu \alpha - i\psi \partial_\mu \alpha) \\ & = \bar{\psi} i \gamma^\mu (\partial_\mu + igA_\mu)\psi \,. \end{aligned} \tag{2.54}$$

Note that postulating the existence of a local symmetry has *forced* us to introduce the *gauge boson* A_μ. This new vector boson is not gauge invariant in and of itself, but from it we can construct the gauge invariant field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \,. \tag{2.55}$$

One can check that $F_{\mu\nu}$ can equivalently be defined from the commutator of covariant derivatives as

$$[D_\mu, D_\nu] = igF_{\mu\nu} \,, \tag{2.56}$$

where $[A, B] = AB - BA$. Note that A_μ is sometimes referred to as the *connection*, in analogy with general relativity where the connection $\Gamma^\alpha_{\mu\nu}$ is used to construct the covariant derivative ∇_μ. Just as the Riemann curvature tensor is constructed as the commutator of covariant derivatives $[\nabla_\mu, \nabla_\nu] = R_{\mu\nu}$, the fact that $F_{\mu\nu}$ can be constructed out of covariant derivatives suggests that it can also be thought of as a kind of curvature.

Returning to our task of constructing a locally $U(1)$ invariant Lagrangian, note that A_μ is a dynamical[16] field in its own right, and so requires a kinetic term. This can be formulated in a gauge-invariant manner using the newly introduced field strength tensor as

$$\mathcal{L}_{\text{kin,gauge}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \,. \tag{2.57}$$

16. Technically, the "0" component A_0 is not dynamical, since the antisymmetric nature of $F_{\mu\nu}$ means there is no way of constructing a gauge invariant kinetic term for it.

Finally, note that there is one more gauge invariant Lorentz scalar we can construct out of $F_{\mu\nu}$; by contracting with it with the completely antisymmetric tensor $\epsilon^{\alpha\beta\mu\nu}$, we can write $\mathcal{L} \supset \epsilon^{\alpha\beta\mu\nu} F_{\alpha\beta} F_{\mu\nu}$. Note that this term is a total derivative and so will not affect the classical equations of motion, though it can have consequences in the quantum theory. Moreover, it violates both parity P and time reversal T symmetries, and so we can neglect it if we postulate these symmetries. Doing so, we arrive at the most general $U(1)$ invariant Lagrangian for ψ and A_μ,

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi} i \not{D} \psi - m \bar{\psi} \psi \,. \tag{2.58}$$

Expanding out the covariant derivative, this can equivalently be written

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi} i \not{\partial} \psi - m \bar{\psi} \psi - g \bar{\psi} \gamma^\mu \psi A_\mu \,. \tag{2.59}$$

This latter form makes it clear that the covariant derivative in the fermion kinetic term actually induces an interaction between fermions ψ and the gauge boson A_μ. Finally, we comment that in the case the $U(1)$ under consideration is electromagnetism $U(1)_{\text{EM}}$, the charge g is e, A_μ is interpreted as the photon, and $F_{\mu\nu}$ encodes the E and B fields. In particular, one can show that the components are

$$F^{\text{EM}}_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix} . \tag{2.60}$$

Gauge Fixing

You should be skeptical of my claim that for $U(1)_{\text{EM}}$, A_μ should be regarded as the photon after quantization, given that the photon has 2 degrees of freedom (2 physical polarizations) while A_μ appears to be a 4-component object with 4 degrees of freedom. To resolve this appearant discrepancy, first note that A_0 has no kinetic term, owing to the antisymmetric nature of $F_{\mu\nu}$, and so is not dynamical. Further, its value is completely fixed by the equation of motion $\nabla \cdot \vec{E} = 0$, and so it is not an independent degree of freedom at all. This eliminates one degree of freedom, leaving us with an apparent 3.

The next thing to realize is that gauge symmetry means that there is great freedom in defining the vector potential. For each A_μ, we could equivalently define $A_\mu \to A_\mu + \partial_\mu \lambda$ for any function λ and arrive at the same $F_{\mu\nu}$ and hence $\vec{E}$ and $\vec{B}$, which are the true physical[17] gauge invariant quantities. This suggests that A_μ and $A_\mu + \partial_\mu \lambda$ should be regarded as the same physical state. More generally, any two states related by a gauge symmetry should be identified. This is precisely what was meant earlier by the comment that gauge symmetry should be regarded not as a true symmetry, but rather a *redundancy in description.*

The picture then is that the phase space of classical electromagnetism is larger than the space of physical states. One can think of this phase space as being foliated by "gauge orbits" — the equivalence class of states related by a gauge transformation. It suffices then to pick a representative from each gauge orbit, a procedure known as *fixing the gauge.* In a sense, picking a gauge is like[18] picking a coordinate system, and one is free to pick the gauge which makes the problem at hand simplest. Gauge fixing is implemented by introducing a

17. Even though A_μ cannot be directly measured, there is still a sense in which it is "physical". This is apparent in certain quantum mechanical phenomena like the Aharonov-Bohm effect [96], wherein electrically charged particles are affected by the presence of a non-zero electromagnetic potential A_μ despite being in a region where both E and B fields are zero.

18. And there is a sense in which picking a coordinate system *is* fixing the gauge! The gauge symmetry of general relativity is *diffeomorphism invariance*: invariance under differentiable coordinate transformations.

constraint on A_μ. Two popular choices include Lorentz gauge,

$$\partial_\mu A^\mu = 0\,, \tag{2.61}$$

and Coulomb gauge

$$\nabla \cdot \vec{A} = 0\,. \tag{2.62}$$

Lorentz gauge is nice because it is Lorentz invariant. However, this choice does not completely fix the gauge, since we still have the freedom to make further gauge transformations provided $\Box\lambda \equiv \partial_\mu \partial^\mu \lambda = 0$. This *residual* gauge freedom can be fixed by choosing Coulomb gauge.

Note that fixing the gauge eliminates one more degree of freedom, leaving A_μ with just 2 physical degrees of freedom. We should identify these as the 2 physical polarizations of the photon. Finally having fixed the gauge, the theory is ready[19] for quantization. We refer the interested reader to e.g. [93, 97] for a review of quantization in quantum field theory.

$SU(N)$ Gauge Symmetry

The $U(1)$ example discussed previously was particularly simple in part because this group is *Abelian*. A general group G is said to be Abelian if the group operation is commutative. In the present context, the group operation is matrix multiplication, and so a group will be Abelian if the commutator of any two group elements is vanishing $[\mathcal{G}, \mathcal{G}'] = 0$ for all $\mathcal{G}, \mathcal{G}' \in G$. The group $U(1)$ was trivially Abelian since its elements are just complex numbers, which of course commute.

More generally in physics, many of the groups we are interested in are *non-Abelian* — that is, their elements do not all commute. Examples include the rotation group in 3-dimensions $SO(3)$, the group $SU(2)$ (which describes quantum spin, among other things), and the group

19. Technically, it is not necessary to gauge fix before quantization, though this is the more convenient option for most purposes.

$SU(3)$ (which describes the strong force). These groups are also all[20] examples of *Lie groups* — continuous groups whose elements are described by one or more smooth parameters, meaning that the group has the structure of a smooth manifold. Lie groups are generated by *Lie algebras*, which represent infinitesimal transformations. We will first review these concepts generically for the group $G = SU(N)$, the group of $N \times N$ unitary matrices of determinant 1. We will then specialize to $SU(2)$ for concreteness.

The Lie algebra $\mathfrak{g} = \mathfrak{su}(N)$ consists of generators T^a, with $a = 1, 2, ...\dim(G) = (N^2 - 1)$, satisfying

$$[T^a, T^b] = if^{abc}T^c \,, \tag{2.63}$$

where f^{abc} are the fully antisymmetric structure constants, which depend on the specific algebra under consideration. Elements $\mathcal{G}$ of the group are obtained by exponentiating the algebra generators as

$$\mathcal{G} = \exp\left(i\theta^a T^a\right) \,. \tag{2.64}$$

Note that for $SU(N)$, whose elements are unitary matrices with determinant 1, the generators should be Hermitian ($T = T^\dagger$) and traceless ($\mathrm{Tr}\,[\,T\,] = 0$). It is also customary to normalize the generators; for the fundamental (minimal) representation, the normalization is

$$\mathrm{Tr}[T^a T^b] = \frac{1}{2}\delta^{ab} \,. \tag{2.65}$$

In the context of gauge theories, for each algebra generator T^a we introduce a gauge field A^a_μ, often conveniently packaged in the Lie algebra-valued gauge potential

$$A_\mu = A^a_\mu T^a \,, \tag{2.66}$$

which, like the generators, are traceless $N \times N$ Hermitian matrices. Analogously to the $U(1)$

20. The Abelian $U(1)$ example discussed earlier is also a Lie group.

case, we should also introduce a (Lie algebra-valued) field strength tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - ig[A_\mu, A_\nu]\,. \tag{2.67}$$

Equivalently, we can pull out the generator and write $F_{\mu\nu} = F^a_{\mu\nu} T^a$, where

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu\,. \tag{2.68}$$

For the purposes of coupling to matter fields, we can also introduce the covariant derivative D_μ, which acts on matter in the fundamental[21] representation ψ as

$$D_\mu \psi = \partial_\mu \psi - igT^a A^a_\mu \psi\,. \tag{2.69}$$

As in the Abelian case, one should interpret $F_{\mu\nu}$ as the curvature and so can[22] construct it as the commutator of covariant derivatives

$$[D_\mu, D_\nu] = -igF_{\mu\nu}\,. \tag{2.70}$$

Finally, when building Lagrangians for $SU(N)$ invariant gauge theories, the form of Eq. (2.57) should be generalized to

$$\mathcal{L}_{\text{kin,gauge}} = -\frac{1}{2}\text{Tr}[F_{\mu\nu}F^{\mu\nu}] = -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a\,, \tag{2.71}$$

where the trace is over the group indices and we have made use of the normalization of Eq. (2.65).

21. All of the fermionic matter fields in the Standard Model transform in the fundamental representation of the groups under which they are charged (and the trivial representation of the groups under which they are not). The gauge fields, on the other hand, transform in the *adjoint*.

22. This is technically the more fundamental definition of $F_{\mu\nu}$, from which the explicit form shown in Eq. (2.67) was derived.

To see how this works explicitly, consider now $SU(2)$ — the group of 2×2 unitary matrices with determinant 1. A generic complex 2×2 matrix has 8 degrees of freedom, but the unitary condition eliminates 4 of these and the condition that the determinant be 1 eliminates 1 more, leaving just 3 degrees of freedom. The algebra $\mathfrak{su}(2)$ then has 3 generators, a convenient choice for which is the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \,, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \,, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \,. \tag{2.72}$$

These are indeed traceless and Hermitian, as required for generators of a special unitary group. Additionally, they are involutary $\sigma_i^2 = \mathbb{I}$, have determinant $\det \sigma_i = -1$, have eigenvalues ± 1, satisfy commutation relations

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k \,, \tag{2.73}$$

and satisfy anticommutation relations

$$\{\sigma_i, \sigma_j\} = 2\delta_{ij}\mathbb{I} \,. \tag{2.74}$$

In keeping with the normalization of Eq. (2.65), we should rescale these by a factor $1/2$ in order to obtain the generators for $SU(2)$, which we will denote by S_i

$$S_i = \frac{\sigma_i}{2} \,, \quad i = 1, 2, 3 \,. \tag{2.75}$$

These then generate the $\mathfrak{su}(2)$ Lie algebra

$$[S_i, S_j] = i\epsilon_{ijk}S_k \,, \tag{2.76}$$

and so we identify the totally antisymmetric epsilon tensor ϵ_{ijk} as the structure constant for

$\mathfrak{su}(2)$. The elements of $SU(2)$ are obtained from these generators via the exponential map of Eq. (2.64). Letting $\vec{\theta} = \theta\,\hat{n}$, where $\hat{n}$ has unit norm, be a real vector,

$$\begin{aligned}
\mathcal{G} &\equiv \exp\left(i\vec{\theta}\cdot\vec{S}\right) \\
&= \exp\left(i\frac{\theta}{2}(\hat{n}\cdot\vec{\sigma})\right) \\
&= \mathbb{I} + i(\theta/2)(\hat{n}\cdot\vec{\sigma}) - \frac{(\theta/2)^2}{2}(\hat{n}\cdot\vec{\sigma})^2 - \frac{i(\theta/2)^3}{3!}(\hat{n}\cdot\vec{\sigma})^3 + ... \\
&= \mathbb{I}\left(1 - \frac{(\theta/2)^2}{2} + \frac{(\theta/2)^4}{4!} - ...\right) + i(\hat{n}\cdot\vec{\sigma})\left((\theta/2) - \frac{(\theta/2)^3}{3!} + \frac{(\theta/2)^5}{5!} + ...\right) \\
&= \mathbb{I}\,\cos\frac{\theta}{2} + i(\hat{n}\cdot\vec{\sigma})\,\sin\frac{\theta}{2}\,,
\end{aligned} \tag{2.77}$$

where we have used the fact that $(\hat{n}\cdot\vec{\sigma})^2 = \mathbb{I}$. Since any element $\mathcal{G}$ of $SU(2)$ can be written in this form, we verify that the S_i indeed generate $SU(2)$.

2.2.2 The Gauge Sector

Now that we have reviewed gauge symmetries, we are ready to discuss the gauge sector of the Standard Model; that is, the sector containing the gauge bosons which mediate three[23] of the four fundamental forces: the electromagnetic force, the strong force, and the weak force. See Fig. 2.2 for a comparison of the relative strengths of these forces at the nuclear scale.

The Strong Force

The Standard Model is a quantum field theory based on the gauge group of Eq. (2.48), or more precisely

$$SU(3)_c \times SU(2)_L \times U(1)_Y\,. \tag{2.78}$$

23. The Standard Model does not account for gravity.

Force	Strength	Mediator	Spin	Mass (GeV)
Strong	1	Gluon g	1	0
Electromagnetism	10^{-3}	Photon γ	1	0
Weak	10^{-8}	$W^{\pm}$, Z	1	80.4, 91.2
Gravity	10^{-37}	Graviton G	2	0

Table 2.2: Fundamental forces and their properties after electroweak symmetry breaking. Relative strengths are evaluated at typical nuclear scales ($\sim 10^{-15}$ m). Note that gravity is not described by the Standard Model, but we include it nevertheless for comparison purposes.

The first factor $SU(3)_c$ is the symmetry group of the strong force, the theory of which is called *quantum chromodynamics* and the charge of which is called *color charge* (hence the "c" subscript on $SU(3)_c$). $SU(3)$ has 8 generators, typically taken to be the Gell-Mann matrices λ^a (or more accurately, $\lambda^a/2$), with $a = 1...8$. Correspondingly, it has 8 gauge bosons G^a_μ — the *gluons*, so-called because they mediate the force responsible for binding quarks together in hadrons. The gluons are massless spin-1 bosons that carry color charge[24], which follows from the fact that they lie in the adjoint representation of $SU(3)_c$. From these gluons, we construct the gluon field strength tensor

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g f^{abc} G^b_\mu G^c_\nu \,. \tag{2.79}$$

This enters into the gluon kinetic term in the usual way, i.e. $\mathcal{L}_{\rm QCD} \supset -\frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a$. Note that this kinetic term generates 3- and 4-gluon vertices, meaning that gluons interact amongst themselves, in contrast to Abelian gauge bosons like photons. Gluon fields also couple to matter in the form of quark fields in the fundamental of $SU(3)$ by means of the covariant derivative

$$D_\mu = \partial_\mu - i g_s \frac{\lambda^a}{2} G^a_\mu \,, \tag{2.80}$$

24. Specifically, they carry a color-anticolor pair.

where g_s is the coupling of the strong force. Finally, one can also construct out of the field strength tensor a θ term

$$\mathcal{L}_{\text{QCD}} \supset \theta \frac{g_s^2}{64\pi^2} \epsilon^{\mu\nu\alpha\beta} G^a_{\mu\nu} G^a_{\alpha\beta} \,, \tag{2.81}$$

which manifestly violates CP symmetry. This term appears in the QCD Lagrangian so in principle CP-violating strong interactions *could* occur, yet none has ever been experimentally observed. This implies that θ must be very small, which requires unnatural fine-tuning — the *strong CP "problem"*.

The Electroweak Force

The remaining factor in Eq. (2.78), $SU(2)_L \times U(1)_Y$, is the symmetry group of the electroweak sector. The three generators corresponding to the $SU(2)_L$ factor go by the name of *weak isospin* $T^i = \sigma^i/2$, where $i = 1, 2, 3$ and σ_i are the Pauli matrices of Eq. (2.72). The subscript "L" refers to the fact that this $SU(2)$ only acts on *left* chiral fermions, a fact which will be discussed further in Sec. 2.2.3. There are three corresponding weak isospin gauge bosons, denoted w^i_μ, which are massless. These enter into the field strength tensors $W^i_{\mu\nu}$ as

$$W^i_{\mu\nu} = \partial_\mu w^i_\nu - \partial_\nu w^i_\mu + g\epsilon^{ijk} w^j_\mu w^k_\nu \,, \tag{2.82}$$

where g is the $SU(2)_L$ coupling strength. The $U(1)_Y$ factor is generated by *weak hypercharge* Y, with corresponding massless gauge boson B_μ and field strength tensor

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \,. \tag{2.83}$$

Finally, the gauge covariant derivative for the electroweak sector reads

$$D_\mu = \partial_\mu - ig' Y B_\mu - igT^i w^i_\mu \,, \tag{2.84}$$

where g' is the $U(1)_Y$ coupling constant.

This spectrum is clearly not what we observe. This is because at low energies electroweak symmetry is spontaneously broken to $U(1)_{\rm EM}$ via the Higgs mechanism, such that the symmetry group of the Standard Model becomes

$$SU(3)_c \times U(1)_{\rm EM}\,. \tag{2.85}$$

Electroweak symmetry breaking was an event which occurred at some point in our universe's history, likely around $200\,\text{GeV}$ ($t \sim 10^{-11}\,\text{s}$) presuming only the known particle content of the Standard Model. Only above this temperature in the hot conditions of the early universe is the unified electroweak description applicable. We will discuss electroweak symmetry breaking in further detail when we discuss the Higgs mechanism in Sec. 2.2.4. For now, we simply comment on the gauge structure and spectrum post-symmetry breaking.

After symmetry breaking, three linear combinations of the original electroweak gauge bosons (corresponding to the broken generators) become massive. These mass eigenstates include the charged $W_\mu^\pm$ bosons,

$$W_\mu^\pm = \frac{1}{\sqrt{2}}\left(w_\mu^1 \mp i w_\mu^2\right)\,, \tag{2.86}$$

and the neutral Z_μ,

$$Z_\mu = \frac{1}{\sqrt{g^2 + {g'}^2}}\left(g w_\mu^3 - g' B_\mu\right)\,. \tag{2.87}$$

These have masses

$$m_W = \frac{gv}{2}\,, \tag{2.88}$$

and

$$m_Z = \frac{\sqrt{g^2 + {g'}^2}\,v}{2}\,, \tag{2.89}$$

respectively, where v is the Higgs vacuum expectation value. The linear combination orthogonal to Z_μ,

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}} \left(g' w_\mu^3 + g B_\mu \right) , \tag{2.90}$$

remains massless and is identified with the mediator of the electromagnetic force — the photon! In terms of these mass eigenstates, the electroweak gauge covariant derivative of Eq. (2.84) becomes

$$D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}} \left(T^+ W_\mu^+ + T^- W_\mu^- \right) - i\frac{1}{\sqrt{g^2 + g'^2}} (g^2 T^3 - g'^2 Y) Z_\mu - i\frac{g g'}{\sqrt{g^2 + g'^2}} (T^3 + Y) A_\mu \,, \tag{2.91}$$

where we have defined $T^\pm = (T^1 \pm i T^2)$. By defining the fundamental unit of electric charge e as

$$e = \frac{g g'}{\sqrt{g^2 + g'^2}} \,, \tag{2.92}$$

and the electric charge quantum number Q as

$$Q = T^3 + Y \,, \tag{2.93}$$

the last term in Eq. (2.91) becomes $-ieQA_\mu$, just as one would expect for the coupling to the electromagnetic field.

2.2.3 The Matter Sector

We now turn to the matter content of the Standard Model — the collection of fermionic fields which transform under representations of the Standard Model's gauge group Eq. (2.78).

Chirality

The first thing to note is that the Standard Model is a *chiral* theory, with left- and right-handed fields transforming differently under the electroweak $SU(2)_L \times U(1)_Y$. A field is said to have left-handed (right-handed) chirality[25] if it transforms in the left-handed (right-handed) representation of the Poincaré group, the symmetry group of special relativity consisting of spacetime translations, rotations, and boosts. A chiral theory is one in which there is an asymmetry in the transformations of left- and right-handed fields.

Consider a 4-component Dirac spinor ψ consisting of 2-component spinors ψ_L and ψ_R as

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} . \tag{2.94}$$

One can project out the left- and right-handed components using the projection operators $P_{L,R}$, defined as

$$P_L = \frac{1}{2}\left(\mathbb{I} - \gamma^5\right) , \text{ where } \gamma^5 = \begin{pmatrix} -\mathbb{I} & 0 \\ 0 & \mathbb{I} \end{pmatrix} , \tag{2.95}$$

as

$$\psi_R = P_R\psi , \;\; \psi_L = P_L\psi . \tag{2.96}$$

Using such a decomposition, one can show that fermion mass terms like $m\bar{\psi}\psi$ have the structure

$$m\bar{\psi}\psi = m\left(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R\right) , \tag{2.97}$$

which mixes left- and right-handed fields. Such mass terms are thus forbidden in chiral theories like the Standard Model, where left- and right-handed fields transform under different representations of the gauge group. We will return to the subject of writing down

25. For massive particles, chirality is distinct from *helicity*, the projection of the particle's spin onto the direction of its momentum. For massless particles, these notions coincide.

gauge-invariant mass terms for the fermions in the coming Sec. 2.2.4.

Leptons

There are two major classes of fundamental fermions, or spin-1/2 particles, in the Standard Model, differentiated by their charge under $SU(3)_c$. We first consider the *leptons*, which are singlets **1** under $SU(3)_c$ and so do not interact via the strong force. There are 6 different varieties organized into 3 generations — electron, muon, and tau — with each generation consisting of a charged lepton $\{e^-, \mu^-, \tau^-\}$ and a neutrino $\{\nu_e, \nu_\mu, \nu_\tau\}$. The left-handed fields of each generation live in the fundamental representation of $SU(2)$, **2**, and so are packaged into $SU(2)_L$ doublets L_L

$$L_L^i = \begin{pmatrix} \nu_L^i \\ \ell_L^i \end{pmatrix} , \tag{2.98}$$

with $i = 1, 2, 3$ labeling the generation. Explicitly, these are

$$L_L = \left\{ \begin{pmatrix} \nu_{e,L} \\ e_L^- \end{pmatrix}, \begin{pmatrix} \nu_{\mu,L} \\ \mu_L^- \end{pmatrix}, \begin{pmatrix} \nu_{\tau,L} \\ \tau_L^- \end{pmatrix} \right\} . \tag{2.99}$$

The right-handed fields ℓ_R^i are all singlets under $SU(2)_L$, and we denote them as the set

$$\ell_R = \{e_R^-, \mu_R^-, \tau_R^-\} . \tag{2.100}$$

Note that we have not included any right-handed neutrinos, which transform trivially under the entire gauge group of the Standard Model. Since they do not interact via electromagnetism, the weak force, or the strong force, they have never been detected and so are not included in the mathematical description of the Standard Model. As for the rest of the leptons, their charges under $U(1)_Y$ and $U(1)_{\text{EM}}$ are listed in Tables 2.3 and 2.4, respectively.

Quarks

In contrast, the quark fields transform in the fundamental representation **3** of $SU(3)_c$. They come in 6 different "flavors"; up u, down d, charm c, strange s, top t, and bottom b; again organized into 3 generations. The left-handed quark fields transform in the fundamental representation **2** of $SU(2)_L$ and so sit in the doublets

$$Q_L = \left\{ \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \right\}, \tag{2.101}$$

while the right-handed quark fields q_R are $SU(2)_L$ singlets

$$q_R = \{u_R, d_R, c_R, s_R, t_R, b_R\}. \tag{2.102}$$

See Tables 2.3 and 2.4, respectively, for their charges under $U(1)_Y$ and $U(1)_{\text{EM}}$.

Summary

In the following two tables, we summarize the fermionic fields constituting the matter sector and their charges under the symmetries of the Standard Model. In particular, Table 2.3 shows their representations under the full symmetry group $SU(3)_c \times SU(2)_L \times U(1)_Y$ while Table 2.4 shows their quantum numbers under the weak isospin generator T^3, the weak hypercharge generator Y, and the resultant electric charge $Q = T^3 + Y$, corresponding to the unbroken $U(1)_{\text{EM}}$ after electroweak symmetry breaking.

2.2.4 The Higgs Mechanism

So far we have described gauge and fermion sectors populated by massless fields, which clearly is not the particle spectrum we observe. The missing ingredient is the Higgs field H,

Field	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$
L_L	**1**	**2**	$-1/2$
e_R^-	**1**	**1**	-1
Q_L	**3**	**2**	$+1/6$
u_R	**3**	**1**	$+2/3$
d_R	**3**	**1**	$-1/3$

Table 2.3: The fermionic fields of the Standard Model and their representations under the full symmetry group $SU(3)_c \times SU(2)_L \times U(1)_Y$.

Field		T^3	Y	Q
L_L	ν_L	$+1/2$	$-1/2$	0
	e_L^-	$-1/2$	$-1/2$	-1
e_R^-		0	-1	-1
Q_L	u_L	$+1/2$	$+1/6$	$+2/3$
	d_L	$-1/2$	$+1/6$	$-1/3$
u_R		0	$+2/3$	$+2/3$
d_R		0	$-1/3$	$-1/3$

Table 2.4: The fermionic fields of the Standard Model and their quantum numbers corresponding to the weak isospin generator T^3, the weak hypercharge generator Y, and their resultant electric charge $Q = T^3 + Y$.

a spin-0 $SU(2)_L$ doublet with a charge assignment

$$H : (\mathbf{1}, \mathbf{2}, +1/2)\,, \tag{2.103}$$

under the Standard Model's $SU(3)_c \times SU(2)_L \times U(1)_Y$ symmetry group. We will parameterize the doublet in terms of the complex charged Higgs H^+ and neutral Higgs H^0 as

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} H^+ \\ H_0 \end{pmatrix}\,. \tag{2.104}$$

This is the field responsible for effecting electroweak symmetry breaking, and by extension generating mass terms for the fermions and massive gauge bosons in a gauge invariant way. The Lagrangian of the Higgs sector is

$$\mathcal{L}_{\text{Higgs}} = -(D_\mu H)^\dagger (D^\mu H) - V(H)\,, \tag{2.105}$$

where the covariant derivative is[26]

$$D_\mu = \partial_\mu - igT^i w_\mu^i - ig'YB_\mu\,, \tag{2.106}$$

and the potential is

$$V(H) = -\mu^2 H^\dagger H + \lambda\left(H^\dagger H\right)^2\,. \tag{2.107}$$

Before discussing the Higgs mechanism for the electroweak sector, we will review it in the simpler context of an Abelian $U(1)$.

The Abelian Higgs Mechanism

We consider a complex scalar field ϕ charged under a local $U(1)$ with Lagrangian

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - (D_\mu\phi)^*(D^\mu\phi) - V(\phi)\,, \tag{2.108}$$

where $D_\mu = \partial_\mu - igA_\mu$ is the covariant derivative and the potential is

$$V(\phi) = -\mu^2\phi^*\phi + \frac{\lambda}{2}(\phi^*\phi)^2\,. \tag{2.109}$$

26. Note that the generators and gauge bosons of $SU(3)_c$ do not appear since H transforms trivially under this group.

Extremizing the potential with respect to ϕ and ϕ^*, we arrive at the minimization condition

$$\langle|\phi|\rangle = \sqrt{\frac{\mu^2}{\lambda}}\,. \tag{2.110}$$

Since we are free to choose this vacuum expectation value (vev) to lie along any direction related by the $U(1)$, let us choose the real direction and write

$$\langle\phi\rangle = \sqrt{\frac{\mu^2}{\lambda}} \equiv v\,, \tag{2.111}$$

where we have defined the vev v. Now consider expanding the potential in small fluctuations about this new vacuum,

$$\phi \to \phi' = \langle\phi\rangle + \phi = v + \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)\,, \tag{2.112}$$

where we have parameterized the complex fluctuation as $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$. Upon expanding the potential, we find

$$V = -\frac{\mu^4}{2\lambda} + \mu^2\phi_1^2 + \sqrt{\frac{\mu^2\lambda}{2}}\phi_1(\phi_1^2 + \phi_2^2) + \frac{\lambda}{8}(\phi_1^2 + \phi_2^2)^2\,. \tag{2.113}$$

The main term of interest is the second, which tells us that we have generated a mass m_1 for ϕ_1

$$m_1 = \sqrt{2}\mu\,. \tag{2.114}$$

This real massive remnant of spontaneous symmetry breaking will be the analog of the Higgs boson when we turn to the non-Abelian electroweak case. The remaining ϕ_2 does *not* acquire a mass and is identified as the massless *Goldstone boson*, which can be thought of as an excitation of the field in the direction of the broken symmetry generator. Finally, note that we have also generated cubic and quartic interaction terms.

We should also expand the kinetic term for ϕ in small fluctuations about the new vacuum. Doing so, we find

$$\begin{aligned} -|D_\mu \phi|^2 = &-\frac{1}{2}\left(\partial_\mu \phi_1\right)^2 - \frac{1}{2}\left(\partial_\mu \phi_2\right)^2 + \sqrt{2} g v A_\mu \partial^\mu \phi_2 - g^2 v^2 A_\mu A^\mu \\ &- g\left(\phi_1 \partial_\mu \phi_2 + \phi_2 \partial_\mu \phi_1\right) A^\mu - \frac{g^2}{2}\left(\phi_1^2 + \phi_2^2\right) A_\mu A^\mu - \sqrt{2} g^2 v \phi_1 A_\mu A^\mu \,. \end{aligned} \tag{2.115}$$

The first line consists of kinetic terms for ϕ_1 and ϕ_2, a mixing between A_μ and ϕ_2, and a mass term for A_μ while the second consists of cubic and quartic interaction terms. Note that we can eliminate the ϕ-A_μ mixing term by making the gauge transformation

$$A_\mu \to A'_\mu = A_\mu + \frac{1}{\sqrt{2} g v} \partial_\mu \phi_2 \,. \tag{2.116}$$

The expanded kinetic term then becomes

$$-|D_\mu \phi|^2 = -\frac{1}{2}\left(\partial_\mu \phi_1\right)^2 - g^2 v^2 A_\mu A^\mu + \text{interactions} \,. \tag{2.117}$$

Notice that the massless Goldstone ϕ_2 no longer has a kinetic term, and so is not a dynamical degree of freedom! We can now identify the mass term for the photon as $-\frac{1}{2} m_A^2 A_\mu A^\mu = -g^2 v^2 A_\mu A^\mu$, indicating that A^μ has acquired a mass

$$m_A = \sqrt{2} g v \,. \tag{2.118}$$

This gauge choice that we have used to eliminate[27] the unphysical Goldstone mode is called *unitary gauge* and it is convenient for degree of freedom counting. Prior to spontaneous symmetry breaking, we had a massless photon (2 polarization degrees of freedom) and a complex scalar ϕ (2 degrees of freedom) for a total of 4 degrees of freedom. After, our spectrum consists of a massive photon (3 polarization degrees of freedom) and a real scalar

27. It is sometimes said that the unphysical Goldstone mode is *eaten* to give the gauge boson a mass.

ϕ_1 (1 degree of freedom), again for a total of 4 degrees of freedom. This is summarized in Table 2.5. Note that we start and end with the same number of degrees of freedom, just organized in a different way.

Degrees of freedom before SSB	Degrees of freedom after SSB
Massless photon: 2 dof	Massive photon: 3 dof
Complex scalar: 2 dof	Real scalar: 1 dof
Total: 4 dof	Total: 4 dof

Table 2.5: Distribution of degrees of freedom (dof) amongst the particles of the spectrum before and after spontaneous symmetry breaking (SSB) in the Abelian Higgs example.

Electroweak Symmetry Breaking

This mechanism wherein spontaneous symmetry breaking generates a mass for the gauge boson in a gauge-invariant way can easily be generalized to the non-Abelian case. Specifically, we are interested in the Higgs mechanism in the context of electroweak symmetry breaking, wherein the electroweak group of the Standard Model $SU(2)_L \times U(1)_Y$ is broken to electromagnetism $U(1)_{\text{EM}}$,

$$SU(2)_L \times U(1)_Y \xrightarrow[\text{mechanism}]{\text{Higgs}} U(1)_{\text{EM}} \,. \tag{2.119}$$

Now we consider an $SU(2)_L$ doublet Higgs field with charge assignment given in Eq. (2.103), parameterization given in Eq. (2.104), Lagrangian given in Eq. (2.105), covariant derivative given in Eq. (2.106), and potential given in Eq. (2.107). As in the Abelian example, for this choice of potential with negative mass squared, the Higgs acquires a vev. We can use the freedom of $SU(2)$ rotations to choose this to lie in the direction

$$\langle H \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \,. \tag{2.120}$$

Expanding the potential $V(H)$ in small fluctuations about this new vacuum,

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1^+ + i\phi_2^+ \\ v + h + i\phi_3^0 \end{pmatrix} , \tag{2.121}$$

we find one real scalar field h with mass [98]

$$m_h = \sqrt{2}\mu = 125.25 \pm 0.17\,\text{GeV} , \tag{2.122}$$

and three massless Goldstone bosons ϕ_1^+, ϕ_2^+, ϕ_3^0. The former is the titular Higgs boson, discovered at the Large Hadron Collider in 2012 [92]. As before, the gauge boson masses come from expanding the scalar kinetic term $\mathcal{L}_{H,\text{kin}} = -(D_\mu H)^\dagger (D^\mu H)$. We again choose unitary gauge in order to eliminate the unphysical Goldstones, which become the longitudinal polarizations of the massive gauge bosons $W_\mu^\pm$ and Z_μ, whose linear combinations in terms of the original electroweak generators were given in Eqs. (2.86) and (2.87), respectively. Their masses are [98]

$$m_W = \frac{gv}{2} = 80.377 \pm 0.012\,\text{GeV} , \tag{2.123}$$

and

$$m_Z = \frac{\sqrt{g^2 + g'^2}\, v}{2} = 91.1876 \pm 0.0021\,\text{GeV} . \tag{2.124}$$

Note that though $\langle H \rangle$ breaks $SU(2)_L \times U(1)_Y$, the fact that this vev is charge neutral means that the generator $Q = T^3 + Y$ corresponding to $U(1)_\text{EM}$ remains unbroken. The corresponding gauge boson A_μ, whose linear combination in terms of the electroweak generators was given in Eq. (2.90), thus remains massless. We summarize in Table 2.6 the distribution of degrees of freedom before and after electroweak symmetry breaking.

Degrees of freedom before EWSB	Degrees of freedom after EWSB
3 massless w_μ^i of $SU(2)_L$: 6 dof	3 massive gauge bosons $W_\mu^\pm$, Z: 9 dof
1 massless B_μ of $U(1)_Y$: 2 dof	1 massless photon A_μ: 2 dof
1 complex scalar doublet H: 4 dof	1 real scalar h: 1 dof
Total: 12 dof	Total: 12 dof

Table 2.6: Distribution of degrees of freedom (dof) before and after electroweak symmetry breaking (EWSB).

Fermion Mass Terms

Recall that the obstruction to writing down a fermion mass term like $m\bar{\psi}\psi$ was that this coupled left- and right-handed fields as $m(\bar{\psi}_L\psi_R+\bar{\psi}_R\psi_L)$, which transform differently under $SU(2)_L \times U(1)_Y$ in the Standard Model. Fortunately, the Higgs mechanism also provides a way to generate fermion mass terms in a gauge invariant manner. Consider the following so-called *Yukawa coupling* between the left-handed lepton doublet, Higgs, and right-handed electron

$$\mathcal{L}_{\text{Yuk},e} = -y_e \bar{L}_L H e_R + h.c. \tag{2.125}$$

Note that since the Higgs hypercharge is $Y_H = +1/2$, the Dirac adjoint lepton doublet hypercharge is $Y_{L_L} = +1/2$, and the right-handed electron hypercharge is $Y_{e_R} = -1$, this term is a singlet under $SU(2)_L \times U(1)_Y$, and so completely permissible to add to the Lagrangian. Crucially, after electroweak symmetry breaking this Yukawa term becomes

$$\mathcal{L}_{\text{Yuk},e} \to -\frac{y_e v}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L) = -m_e(\bar{e}_L e_R + \bar{e}_R e_L)\,, \tag{2.126}$$

spontaneously generating a mass $m_e = y_e v/\sqrt{2}$ for the electron. If there exists a right-handed neutrino ν_R, one could in principle also write down a Dirac mass term of neutrino

$$\begin{aligned}\mathcal{L}_{\text{Yuk},\nu} &= -y_\nu \bar{L}_L \tilde{H} \nu_R + h.c. \\ &\to -\frac{y_\nu v}{\sqrt{2}}(\bar{\nu}_L \nu_R + \bar{\nu}_R \nu_L) \\ &= -m_\nu(\bar{\nu}_L \nu_R + \bar{\nu}_R \nu_L)\,,\end{aligned} \tag{2.127}$$

where $\tilde{H} = i\sigma_2 H^*$ and $m_\nu = y_\nu v/\sqrt{2}$. Analogously for the up and down quark fields,

$$\begin{aligned}\mathcal{L}_{\text{Yuk},q} &= -y_d \bar{Q}_L H d_R - y_u \bar{Q}_L \tilde{H} u_R + h.c. \\ &\to -\frac{y_d v}{\sqrt{2}}(\bar{d}_L d_R + \bar{d}_R d_L) - \frac{y_u v}{\sqrt{2}}(\bar{u}_L u_R + \bar{u}_R u_L) \\ &= -m_d(\bar{d}_L d_R + \bar{d}_R d_L) - m_u(\bar{u}_L u_R + \bar{u}_R u_L)\,,\end{aligned} \tag{2.128}$$

where $m_d = y_d v/\sqrt{2}$ and $m_u = y_u v/\sqrt{2}$. Recall that each of these fields actually comes in 3 generations, so writing the fields as vectors $u_L^i = (u_L, c_L, t_L)$, $d_L^i = (d_L, s_L, b_L)$, $e_L^i = (e_L, \mu_L, \tau_L)$, and analogous for the right-handed fields, we can write the Standard Model Yukawa sector as

$$\begin{aligned}\mathcal{L}_{\text{Yuk}} &= -y^e_{ij} \bar{L}^i_L H e^j_R - y^d_{ij} \bar{Q}^i_L H d^j_R - y^u_{ij} \bar{Q}^i_L \tilde{H} u^j_R + h.c. \\ &\to -M^e_{ij} \bar{e}^i_L e^j_R - M^d_{ij} \bar{d}^i_L d^j_R - M^u_{ij} \bar{u}^i_L u^j_R + h.c.\end{aligned} \tag{2.129}$$

The y_{ij} and $M_{ij} = y_{ij} v/\sqrt{2}$ are now completely general complex-valued 3×3 matrices, which need not be Hermitian or symmetric. Nevertheless, they can still be diagonalized by means of a biunitary transformation. To do so for the quark sector, consider introducing the unitary matrices V^u_L, V^u_R, V^d_L, and V^d_R. The diagonalized quark mass matrices are then obtained as

$$M^u_{\text{diag}} = V^u_L y^u V^{u\,\dagger}_R \;\text{ and }\; M^d_{\text{diag}} = V^d_L y^d V^{d\,\dagger}_R\,, \tag{2.130}$$

where $M^u_{\text{diag}} = \text{diag}(m_u, m_c, m_t)$ and $M^d_{\text{diag}} = \text{diag}(m_d, m_s, m_b)$. These matrices $V^{u/d}_{L/R}$ can also be used to go between the mass and weak eigenstates. The combination $V^u_L V^{d\,\dagger}_L \equiv V_{\text{CKM}}$ defines the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which enters into the charged current interactions mediated by $W^{\pm}$. This 3×3 unitary matrix is parameterized by three mixing angles and a single CP violating phase, responsible for CP-violation in flavor-changing processes in the Standard Model quark sector.

2.2.5 Summary

The Standard Model is a gauge theory based on the non-Abelian groups $SU(3)_c \times SU(2)_L \times U(1)_Y$. The first factor, $SU(3)$ "color", is the symmetry group of strong interactions, mediated by the 8 massless spin-1 gluons. Under $SU(3)_c$, the spin-1/2 quark fields (u, d, c, s, t, b) transform in the fundamental representation **3** while the remainder of the Standard Model's matter content is uncharged.

The second factor, $SU(2)_L \times U(1)_Y$, is the symmetry group of the electroweak sector. At high energies when these symmetry is unbroken, it is mediated by the massless $w^{i=1,2,3}_\mu$ bosons (corresponding to the $SU(2)_L$ generators) and the massless B_μ boson (corresponding to hypercharge $U(1)_Y$). Left-handed particles transform in the fundamental **2** of $SU(2)_L$ while right-handed particles are singlets. The hypercharge assignments are collected in Table 2.3.

At low energies, electroweak symmetry is spontaneously broken by means of the Higgs mechanism to electromagnetism $U(1)_{\text{EM}}$. Certain linear combinations of the original electroweak bosons become the massive charged $W^{\pm}_\mu$ and neutral Z_μ bosons. The linear combination corresponding to the unbroken generator $Q = T^3 + Y$ is identified as the photon A_μ, the mediator of electromagnetism.

The fermionic fields and their representations under the various symmetry groups are organized in Tables 2.3 and 2.4. They include the quarks, which interact via the strong force

and exist confined in hadrons, and the leptons, consisting of the charged leptons (e^-, μ^-, τ^-) and neutrinos $(\nu_e, \nu_\mu, \nu_\tau)$. Each of these fields also has an associated antiparticle.

Finally, the Higgs mechanism is responsible for breaking electroweak symmetry and bestowing masses unto the massive gauge bosons and fermions. The Higgs field H is a spin-0 $SU(2)_L$ doublet with hypercharge $+1/2$, and it acquires a vacuum expectation value v at low energies. The Goldstone bosons become longitudinal polarizations for the massive $W^\pm$ and Z bosons, and the Higgs boson h is the remaining real scalar physical degree of freedom. Its 2012 discovery by the ATLAS and CMS experiments at the Large Hadron Collider robustly evinced the Higgs mechanism.

2.3 Open Questions

The Standard Model is extremely successful — frustratingly so. It has predicted the existence and properties of the $W^\pm$ and Z bosons, Higgs boson h, top t and charm c quarks, and several other particles years before their discoveries, and also allows us to compute quantities like the fine-structure constant α with *incredible* precision [99]. With the 2012 discovery of the Higgs boson [92], however, the particle spectrum predicted by the Standard Model is complete, and the lack of new particles observed at colliders has left theorists without experimental guidance. The Standard Model may be "complete", but it *cannot*[28] be the ultimate theory for describing fundamental particles and their interactions, as the following list of open problems makes clear.

Meanwhile, the ΛCDM model has been a relatively successful framework for explaining and predicting the evolution and structure of the universe on large scales. It has faced challenges in explaining observations on smaller length scales $\lesssim 1\,\mathrm{Mpc}$, however. Further, ΛCDM is in some sense a phenomenological model; it describes and allows for predictions, but it fails to actually *explain* the physics of dark matter and dark energy, which constitute

28. Not to mention the fact that it completely omits gravity.

the vast majority of the universe's energy budget.

The following subsections describe some of the most pressing problems in particle cosmology and astronomy today. This list is not meant to be exhaustive, but rather a sampling of the puzzles left unaddressed by the current paradigms. Note also that not all of the questions included in this list will be addressed at length in this thesis.

2.3.1 Problems in Particle Cosmology

We begin with a discussion of open questions at the intersection of particle physics and cosmology. Problems of a strictly cosmological or astrophysical nature are deferred to Sec. (2.3.2).

Identity of the Dark Matter

Clearly none of the Standard Model particles described in Sec. 2.2 can constitute the dark matter, which is uncharged under the gauge group of the Standard Model and so interacts with ordinary matter only gravitationally. The evidence for dark matter is irrefutable and comes from many sources, including observations of the CMB, gravitational lensing, large scale structure, galaxy cluster dynamics, galaxy rotation curves, and much more. From these lines of evidence, we know that it constitutes roughly 26% of the universe's energy density today. The hierarchical formation of structure also suggests that it is cold, with a slow-moving velocity and a small free streaming length. Popular classes of particle dark matter candidates include weakly interacting massive particles (WIMPs), axions, sterile neutrinos, and supersymmetric particles such as neutralinos [100]. Primordial black holes are a very plausible non-particle dark matter candidate [101, 9].

Nature of Dark Energy

Our universe has recently entered into a period of dark energy domination, with dark energy constituting $\sim 68\%$ of the current energy budget. The hallmark of this dark energy is that it does not appreciably dilute as the universe expands, corresponding to an equation of state $w = -1$ and leading to an accelerated expansion rate [43, 44]. But what is the physical substance actually driving the observed accelerated late-time expansion? And moreover, is this apparent[29] acceleration genuine?

Presuming that it is, the simplest explanation is the presence of a small, positive cosmological constant Λ. In general relativity, Λ is a free parameter[30]; in quantum field theory, however, it would be identified as the zero point energy of the vacuum. Estimating this quantity in QFT, one finds

$$\Lambda_{\rm QFT} \sim 8\pi M_{\rm Pl}^2 \simeq 4 \times 10^{39}\,{\rm GeV}^2\,. \tag{2.131}$$

Meanwhile, the observed value of the cosmological constant is

$$\Lambda_{\rm obs} = \frac{8\pi}{M_{\rm Pl}^2}\Omega_\Lambda \rho_{\rm crit} \simeq 4 \times 10^{-84}\,{\rm GeV}^2\,, \tag{2.132}$$

leading to the ratio

$$\frac{\Lambda_{\rm QFT}}{\Lambda_{\rm obs}} \sim 10^{123}\,. \tag{2.133}$$

This discrepancy of roughly 123 orders of magnitude is the famous *cosmological constant problem*, also known as the most embarrassing theoretical prediction in physics. This incon-

29. Some have proposed that the averaged effect of small scale inhomogeneities could alter observational relations on large scales, resulting in the illusion of an accelerated expansion. See e.g. [102].

30. Interestingly, string theory does not seem to admit any stable de Sitter vacua [103]. If we are living in a universe with a positive cosmological constant and if string theory does turn out to be the ultimate description of nature, then for these facts to be compatible the cosmological "constant" would have to be elevated to a dynamical field.

gruous prediction may reflect the necessity of a true theory of quantum gravity. Alternatively, the cosmological "constant" may not be a constant at all, but rather a dynamical field, as in theories of so-called *dynamical dark energy* [104].

Origin of the Matter-Antimatter Asymmetry

It is readily apparent that matter far exceeds antimatter in the observable universe. This asymmetry can be quantified by the baryon asymmetry parameter η, defined as the net baryon number density $n_B = n_b - n_{\bar{b}}$ normalized to the number density of photons n_γ,

$$\eta \equiv \frac{n_B}{n_\gamma} = \frac{n_b - n_{\bar{b}}}{n_\gamma} \,. \tag{2.134}$$

Alternatively, the net baryon density may be normalized to the entropy density s to define

$$Y_B \equiv \frac{n_B}{s} = \frac{45\zeta(3)}{\pi^4 g_{\star,S}} \eta \,. \tag{2.135}$$

Measurements of both the light element abundances and the CMB reveal a value of $Y_B \simeq 8.8 \times 10^{-11}$ [50]. The fact that these measurements coincide indicates that the asymmetry was established prior to both events, within the first second of our universe's history. *Baryogenesis* is the name given to the process in the early universe which resulted in this overabundance of baryons relative to antibaryons. While the mechanism of baryogenesis remains unclear, it has been established that a successful baryogenesis requires the following three conditions, the *Sakharov conditions* [105], to be satisfied: 1) Baryon number violation; 2) C and CP violation; and 3) Out-of-equilibrium conditions. One very popular prospect is for baryogenesis to have occurred during the electroweak phase transition. While the particle content of the Standard Model lacks the requisite amount of CP violation and also fails to make the phase transition strongly first-order (as required to fulfill the out-of-equilibrium condition), this can be accomplished in minimal extensions beyond the Standard Model. We

return to this problem in Chapter 4.

Fine-Tuned Initial Conditions

The universe appears to have had very finely-tuned initial conditions. One manifestation of this is the *horizon problem* — the mystery of why regions which were apparently never in casual contact with one another nevertheless appear homogeneous. Another is the *flatness problem* — the fact that the universe's geometry appears to a very good approximation flat (alternatively, that the energy density appears very close to critical), which is surprising since curvature increases as the universe expands. In order to avoid having to introduce very finely-tuned initial conditions to explain these observational facts of large scale flatness and homogeneity, cosmologists typically invoke an early period of exponential expansion — *inflation* [61, 62]. Since the comoving cosmological horizon shrinks as $R_H \propto a^{-1}$ during such a period, it is plausible that for an inflation lasting sufficiently long, the entire observable universe could have once been in causal contact, explaining the observed homogeneity. Such an exponential expansion would have also diluted any initial curvature as $|\Omega_\kappa| \propto a^{-2}$, explaining the flatness observed today.

We have included this open question in the "particle cosmology" section as the most commonly proposed explanation — inflation — requires a particle physics realization. Whatever is driving this initial period of accelerated expansion must eventually disappear, ruling out vacuum energy, and further must lead to a reheated bath of Standard Model particles. A slowly rolling scalar field ϕ (the *inflaton*) whose dynamics are dominated by its potential rather than kinetic energy gives rise to an equation of state $w = -1$ and so can lead to inflation. Moreover, the inflaton's eventual decay can serve to populate the Standard Model bath. Inflation can also explain the large scale structure observed today; tiny quantum fluctuations generated during inflation are subsequently stretched to superhorizon scales, becoming the classical density perturbations seeding structure formation. The nearly scale-invariant

spectrum predicted by inflation matches large scale observations remarkably well. That being said, inflation does face some challenges and alternatives to inflation do exist. Further, even if inflation is the resolution to this problem of fine-tuned initial conditions, its particle physics realization remains an open question. We will return to inflation in Chapter 3 in the context of primordial black hole formation.

Origin and Nature of Neutrino Masses

The Standard Model describes all particles transforming in nontrivial representations of the symmetry group $SU(3)_c \times SU(2)_L \times U(1)_Y$. As such, it does not include right-handed neutrinos, which are singlets under this symmetry group. More generally, the term *sterile neutrino* is used to describe such Standard Model gauge singlets. The lack of right-handed neutrinos prevents one from writing down a Yukawa term of the form $\mathcal{L}_{\text{Yuk},\nu} = -y_\nu \bar{L}_L \tilde{H} \nu_R + h.c.$ and by extension a Dirac mass term. Neutrinos in the Standard Model are then massless. Neutrinos in nature are not. The observation of neutrino oscillations — the quantum mechanical phenomenon whereby a neutrino created in one flavor eigenstate is later measured after propagating through space in a different flavor eigenstate — has robustly demonstrated that neutrinos *do* have mass.

There are two distinct gauge invariant ways to construct neutrino mass terms in extensions beyond the Standard Model. One can generate a *Dirac* mass term from Yukawa interactions following electroweak symmetry breaking in the manner described in Sec. 2.2.4

$$\mathcal{L}_{M_D} = -y^\nu_{ij} \bar{L}^i_L \tilde{H} \nu^j_s + h.c. \rightarrow -M^D_{ij} \bar{\nu}^i_L \nu^j_s + h.c. \tag{2.136}$$

where ν^j_s, $j = 1...m$ is a vector of m sterile neutrinos and M_D is a complex $3 \times m$ matrix. Such a Dirac mass term conserves lepton number but may break lepton flavor number. The

other option is to add a *Majorana* mass term of the form

$$\mathcal{L}_{M_M} = -\frac{1}{2} M_{ij}^{M} \bar{\nu}_s^i \nu_s^{c\,j} + h.c.\,, \tag{2.137}$$

where M_M is a symmetric $m \times m$ matrix and ν_s^c is the charge conjugate sterile neutrino. Note that since this term is a singlet under the Standard Model gauge group, and so can be added to the Standard Model Lagrangian as a bare mass term. Note also that such a term violates lepton number (by two units). The entire mass matrix for the neutrinos can in principle include both terms

$$\mathcal{L}_{\nu,\text{mass}} = -\frac{1}{2} \begin{pmatrix} \overline{\vec{\nu}_L^c}\,, & \overline{\vec{\nu}_s} \end{pmatrix} \begin{pmatrix} 0 & M_D^T \\ M_D & M_M \end{pmatrix} \begin{pmatrix} \vec{\nu}_L \\ \vec{\nu}_s^c \end{pmatrix} + h.c. \tag{2.138}$$

The physical mass eigenstates would then be obtained by diagonalizing this mass matrix. If the mass eigenvalues in M_M are much larger than those in M_D, the result is three light "active" neutrinos and m heavy neutrinos N. This mechanism of generating small masses for the active neutrinos is referred to as the *see-saw mechanism*.

The question of whether neutrinos are Majorana or Dirac, their absolute mass scale and mass hierarchy, the amount of CP violation in the neutrino sector, the number and properties of any sterile neutrinos, and much more remains unknown about these mysterious particles. A large impediment to progress thus far is the weakness of the active neutrinos' interactions with matter. Nevertheless, terrestrial experimental efforts like the upcoming Deep Underground Neutrino Experiment (DUNE) at Fermilab [106] promise to made headway on such questions. Cosmological observations also offer a complimentary way to probe the neutrino sector [107]. Neutrinos occupy a unique role as the only known particle species which is relativistic at early times (during the formation of the CMB) but non-relativistic later (by structure formation). Measurements of the CMB and light element abundances have allowed

us to put tight constraints on the number of relativistic neutrino species present in the early universe N_{eff}, while late time measurements of large scale structure allow us to bound the sum of neutrino masses $\sum m_\nu$.

The Hierarchy Problem

The hierarchy problem is the question of why the Higgs mass $m_h \simeq 125\,\text{GeV}$ is so much lighter than the Planck mass $M_{\text{Pl}} = 1.22 \times 10^{19}\,\text{GeV}$. The Higgs mass recieves quantum corrections from everything it couples to. For example, a new Dirac fermion f coupling to H via the operator $-\lambda_f H \bar{f} f$ leads to a correction

$$\Delta m_h^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{\text{UV}}^2 + ... \tag{2.139}$$

where Λ_{UV} is an ultraviolet (UV) momentum cutoff used to regulate the loop integral and the subleading terms are logarithmic and even slower growing in Λ_{UV}. This quadratic sensitivity to Λ_{UV} is unique to scalars and makes them very sensitive to the presence of new physics. Λ_{UV} should be thought of as the smallest energy scale where this new physics enters to alter the high-energy behavior of the theory. If the Standard Model is all there is then $\Lambda_{\text{UV}} \sim M_{\text{Pl}}$ is of order the Planck scale, in which case these quantum corrections are *significantly* larger than the Higgs bare mass. The fact that the observed Higgs is so light suggests a fine-tuned cancellation of any new contributions, which seems unnatural. Such cancellations can be made natural in theories like supersymmetry, where quantum corrections from new bosons and fermions come with equal magnitude and opposite sign so as to precisely cancel and thereby protect the tiny Higgs mass against quantum corrections. Other proposed solutions to the hierarchy problem include extra-dimensional and brane world scenarios [108] as well as a breakdown of effective field theory reasoning due to UV/IR mixing [109].

The Strong CP Problem

Recall from Sec. 2.2.1 that the most general force for the QCD Lagrangian included a term

$$\begin{aligned}\mathcal{L}_\theta &= \theta \frac{g_s^2}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{\mu\nu}_a \\ &= \theta \frac{g_s^2}{64\pi^2} \epsilon^{\mu\nu\alpha\beta} G^a_{\mu\nu} G^a_{\alpha\beta} \,,\end{aligned} \tag{2.140}$$

where $\tilde{G}^{\mu\nu}_a = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta} G^a_{\alpha\beta}$ is the dual field strength tensor and $\epsilon^{\mu\nu\alpha\beta}$ is the Levi-Civita (pseudo)tensor. Crucially, this changes sign under parity $\epsilon^{\mu\nu\alpha\beta} \to -\epsilon^{\mu\nu\alpha\beta}$, and so the QCD θ term explicitly violates CP. Surprisingly however, CP violation has never been observed in the strong interactions, implying that θ must be very small. From measurements of the neutron electric dipole moment d_n, which restrict [110]

$$\begin{aligned}|d_n| &\simeq 3.6 \times 10^{-16} \, |\theta| \, e \, \text{cm} \\ &< 3 \times 10^{-26} \, e \, \text{cm} \,,\end{aligned} \tag{2.141}$$

we have that $|\theta| \lesssim 10^{-10}$. This degree of fine-tuning is not technically natural; one cannot even introduce a new symmetry to enforce $\theta = 0$ since θ receives contributions both from the structure of QCD and the quark masses, and there is no good reason that these contributions should precisely cancel. This is the famous strong CP problem. Perhaps the most famous solution is the Peccei-Quinn mechanism [111], which introduces a new complex scalar field φ as well as an anomalous global symmetry $U(1)_{\text{PQ}}$ under which φ is charged. The potential is such that φ develops a vev at low energies, giving rise to a pseudo-Goldstone boson a called the axion. Among other couplings, the axion enters into the Lagrangian with a *color anomaly* term of the form

$$\mathcal{L}_a = \xi \frac{a}{f_a} \frac{g_s^2}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{\mu\nu}_a \,, \tag{2.142}$$

where f_a is the axion decay constant and ξ is related to the $U(1)_{\rm PQ}$ charge assignments for the quarks. The θ angle is then modified to an effective

$$\theta_{\rm eff} = \theta + \xi \frac{a}{f_a} \,, \tag{2.143}$$

which in turn modifies the effective potential due to instanton effects. The result is that in order to minimize the ground state energy, a assumes the vev $\langle a \rangle = -\theta f_a/\xi$, relaxing $\theta_{\rm eff}$ to zero and thereby resolving the strong CP problem. While the original PQ axion has long been experimentally excluded, there exist modifications of this scenario that remain potentially viable.

2.3.2 Astrophysical and Cosmological Puzzles

In addition to the above questions, which can be argued to fall within the domain of particle cosmology, the following questions are of a more strictly cosmological or astrophysical nature. Nevertheless, particle physics may play a role in their eventual resolution.

The Hubble Tension

There are two primary methods used to determine the Hubble constant H_0, the present-day rate at which the universe is expanding. One can either:

- Infer the value of H_0 from *early-time* probes, in particular from CMB and BAO data
- Infer the value of H_0 from *local* measurements of distances and redshifts, in particular Type 1a supernovae

Note that the first method tacitly assumes standard ΛCDM cosmology. The problem is that these measurements do not agree; the latter method based on local measurements consistently predicts values considerably higher than the former based on early-time data.

For example, the most recent local measurement from the SH0ES team (R20) determined $H_0 = 73.2 \pm 1.3\,\mathrm{km\,s^{-1}Mpc^{-1}}$ at 68% CL [60] while the value from *Planck* 2018 at 68% CL is $H_0 = 67.4 \pm 0.5\,\mathrm{km\,s^{-1}Mpc^{-1}}$ [2].

The *Hubble tension* refers to this 4σ to 6σ tension between early- and late-time measurements of H_0. At the root of this tension is really a tension in distance measurements: different data sets do not agree about how far from us things are. As far as a resolution to the Hubble tension goes, there are then two broad options. Either there is a source of systematic error in one or both methods, leading the distance measurements to be incorrect, or the cosmological model (i.e. ΛCDM) used to fit these data sets is incorrect. The majority of cosmologists are of the opinion the latter is more plausible. Still, there have been *many* proposed solutions to the Hubble tension, with none fully satisfactory. See Ref. [46] for an exhaustive list.

Robustness of the Cosmological Principle

The cosmological principle is the statement that the universe is statistically homogeneous and isotropic when viewed on large enough scales, where in practice "large enough" roughly equates to $\sim (100 - 150)\,\mathrm{Mpc}$. The current ΛCDM cosmological paradigm is predicated on the premise that the cosmological principle is true. Recent observations, however, have brought this assumption into question.

In particular, the presumption of a homogeneous universe has been challenged by the discovery of mega-structures like the *Big Ring*, with a diameter of $\sim 400\,\mathrm{Mpc}$ [57]; the *Giant Arc*, spanning $\sim 1000\,\mathrm{Mpc}$ [58]; and the *Hercules-Corona Borealis Great Wall*, extending $\sim 3000\,\mathrm{Mpc}$ in length [59]. These discoveries do not necessarily violate the cosmological principle, but their statistical improbability does bring it into question.

Meanwhile, the presumption of an isotropic universe has been challenged by the presence of large features in the CMB at scales $\gtrsim 4000\,\mathrm{Mpc}$ which appear to be aligned with the solar

system's motion and orientation — the so-called "axis of evil" [112]. Admittedly the statistical significance of this alignment remains a matter of debate, however. If these evidences do signal a genuine breakdown of the cosmological principle, then clearly it is inappropriate to model the universe with an FLRW metric at late times, and the ΛCDM model would require serious revision.

Origin of Cosmological Magnetic Fields

The magnetic fields present within galaxies and galaxy clusters today stem from pre-existing "seed" fields which have been amplified through the dynamo mechanism [113]. The origin of these seed fields, which must have been in place by the onset of structure formation, remains the subject of much speculation, however. Some possibilities for generating small primordial magnetic fields include phase transitions in the early universe or during inflation [114]. Each of these mechanisms comes with challenges, however, principle among these that the field's initial correlation length must be sufficiently large in order to have survived the effects of early magnetic dissipation and diffusion. We will return to this question in Sec. 5.3, wherein we propose a novel means of generating these primordial magnetic fields.

High-Redshift Supermassive Black Holes

The standard assumption is that supermassive black holes grow from relatively low mass seeds ($\lesssim 100\,M_\odot$) through accretion and mergers. At the Eddington limited rate, a seed of mass $M_{\rm BH} \sim 100\,M_\odot$ would require $\sim 0.8\,$Gyr to grow to a supermassive black hole of mass $M_{\rm BH} \sim 10^{10}\,M_\odot$. Given that a redshift of $z = 6 - 7$ corresponds to only $\sim 0.7 - 0.9\,$Gyr after the Big Bang, the recent observations of supermassive black holes at these and higher redshifts is quite surprising. Some examples include the $7.8 \times 10^8\,M_\odot$ ULAS J1342+0928 at $z = 7.54$ [115], the $2 \times 10^9\,M_\odot$ QSO J0313-1806 at $z = 7.642$ [116], and most recently the $2 \times 10^6\,M_\odot$ GN-z11 observed by the James Webb Space Telescope at $z = 10.6$ [117].

Explaining how these supermassive black holes came to be so large on such a short timescale remains an open challenge. Some cosmologists have suggested an enhanced role of mergers or super-Eddington accretion as possible explanations [118]. However sufficiently frequent mergers would require very heavily clustered initial populations of seed black holes, which is unlikely for population III star seeds. Further, mergers can disrupt supermassive black hole growth by knocking them out of the material-rich centers of galaxies, resulting in a net negative impact on their growth. Super-Eddington growth is possible and even likely in high-redshift galaxies with large reservoirs of turbulent gas. However super-Eddington growth is not sustainable and only occurs in transient spurts. In fact, feedback effects from these super-Eddington periods can actually also have a net deterimental effect on supermassive black hole growth [119].

Another possibility is that rather than being seeded by population III stars, which do not form until $z \sim 30 - 20$, these early supermassive black holes might instead be seeded by primordial black holes formed from the collapse of primordial overdensities [120]. As primordial black holes form much earlier, the timing problem would then be relaxed. We return to this question in Sec. 5.2, wherein we consider an even more radical proposal — supermassive black holes from the direct collapse of overdensities seeded by inflation.

CHAPTER 3

PROBING THE EARLY UNIVERSE

Much remains to be understood about the earliest moments of our universe's evolution, with the precise sequence of events spanning from the Big Bang up through the first second only indirectly and weakly constrained. Progress thus far has been hindered by the fact that there exist few direct probes prior to the synthesis of the light elements ($\sim 1\,\mathrm{s}$) and the formation of the CMB ($\sim 380,000\,\mathrm{yr}$) that could shed light on these dark ages.

This is situation is rapidly changing as we enter the age of multi-messenger cosmology and gravitational wave astronomy. Already, observations of black hole binary mergers by the LIGO/Virgo collaboration [121] and the recent detection of a signal consistent with the stochastic gravitational wave background by the NANOGrav collaboration [122] and other pulsar timing arrays [123, 124, 125] have inspired much excitement and optimism. With the influx of data from the next generation of ground-based gravitational wave observatories, space-based interferometers, increasingly precise CMB experiments, and a multitude of cosmic surveys, we are now in a better position than ever before to unravel the mysteries of these earliest times.

In the coming subsections, we will examine two exceptionally promising probes of the early universe: primordial black holes and their gravitational wave signatures. See [126, 127] for recent reviews. After examining primordial black hole formation and how bounds on present day abundances can be used to constrain the small scale primordial power spectrum (and thereby inflation), we will turn to primordial black hole evaporation and see how their evaporation products can be used to constrain non-standard expansion histories and even scenarios with extra dimensions.

3.1 Primordial Black Holes and Inflation

It is widely believed that within the first fractions of a second following the Big Bang, the universe underwent a period of exponential expansion called inflation (see Sec. 2.1 for a review). During this period, tiny quantum fluctuations were stretched to superhorizon scales and became the density perturbations needed to seed the structure observed today. Inflation sources a spectrum of fluctuations measured on large scales to be approximately scale invariant with an amplitude $\sim 2.1 \times 10^{-9}$ on CMB scales, which is consistent with the single-field slow-roll paradigm. On small scales, however, the primordial power spectrum remains largely unconstrained, permitting a much richer landscape of inflationary models.

Amplified power on these small scales arises very generically in models which deviate even briefly from slow-roll behavior, and can lead to primordial black hole formation. More generally, primordial black holes form from the gravitational collapse of large overdensities in the radiation dominated[1] early universe. They may also be produced during cosmological phase transitions [128] or in the collapse of topological defects like cosmic strings [129], though we will not consider such mechanisms in this work.

The mass of a primordial black hole is upper bounded by the size of the cosmological horizon at the time of its formation, and so can range from the Planck scale $M_{\rm Pl}$ all the way up to $\sim 10^3\, M_\odot$ (or possibly even higher, as we will explore in Sec. 5.2). Because they form as early as $\sim 10^{-42}\,$s after the Big Bang, they constitute relics of our universe's earliest epoch. In addition to their evaporation products and other signatures that could be used to make inferences about the early universe, large abundances of primordial black holes can have a profound impact on cosmological evolution, as demonstrated in Fig. 3.1. For example, primordial black holes can serve to seed supermassive black holes and thereby accelerate early galaxy formation. This possibility is especially exciting in light of recent data from the

1. Primordial black hole formation can also occur during periods of early matter domination in non-Standard cosmologies. In this case, the major impediment to formation is not the density threshold, but rather deviations of the perturbation from a spherically symmetric profile.

James Webb Space Telescope (JWST), which has reported a number of surprisingly luminous high-redshift galaxy candidates [130, 131, 132, 133] whose existence poses a challenge to the standard ΛCDM paradigm. These massive early galaxies could conceivably be explained if primordial black holes accelerated galaxy formation in the early universe.

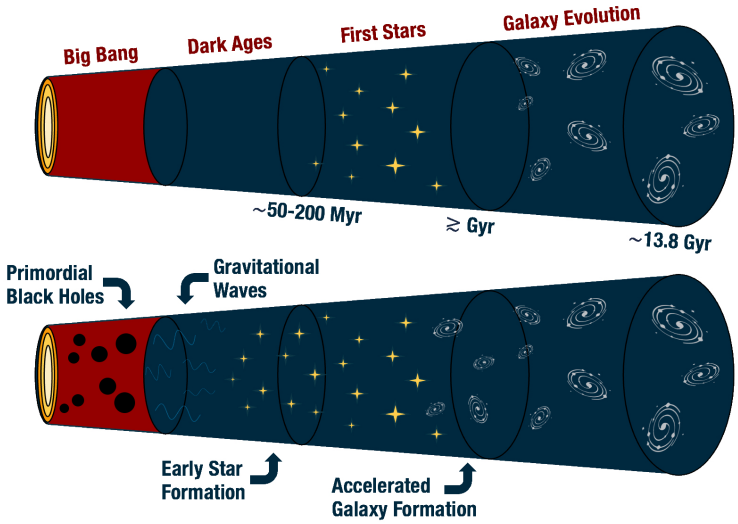

Figure 3.1: Potential effects of primordial black holes on cosmological evolution. The top depicts the standard cosmological history, while the bottom depicts how this may be altered given an initial abundance of primordial black holes. In addition to sourcing gravitational waves and other evaporation products, these could seed supermassive black holes and accelerate early galaxy formation.

Given these motivations for considering primordial black holes, we now turn to review their formation in Sec. 3.1.1 before discussing inflationary mechanisms to enhance the primordial curvature perturbation in Sec. 3.1.2.

3.1.1 Primordial Black Hole Formation

Consider a perturbation mode δ_k with comoving length scale k^{-1}. Because primordial black hole formation is ultimately a *casual* process, the length scale that this should be compared

with is that of the comoving horizon $R_H \equiv (aH)^{-1}$, which evolves with scale factor as

$$R_H \propto a^{(1+3w)/2}\,, \tag{3.1}$$

for generic equation of state w. See Fig. 3.2. During inflation ($w = -1$), the causal horizon

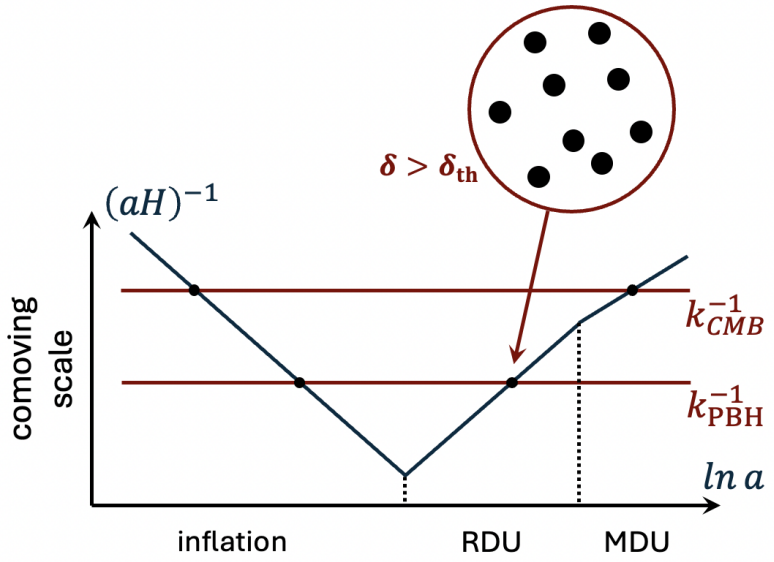

Figure 3.2: Evolution of the comoving horizon $R_H = (aH)^{-1}$ during inflation, radiation domination, and matter domination. This is compared with sample comoving length scales $k_{\rm CMB}^{-1}$ (associated with the CMB) and $k_{\rm PBH}^{-1}$ (associated with primordial black hole formation). When $k_{\rm PBH} = aH$ and the associated perturbation mode re-enters the horizon, primordial black hole formation will occur provided the density contrast exceeds the threshold value, $\delta > \delta_{\rm th}$.

shrinks as $R_H \propto a^{-1}$, while afterwards during radiation domination ($w = 1/3$) it begins to grow again as $R_H \propto a$. Perturbation modes typically originate deep in the horizon, where they are highly oscillatory. We say that such a mode with $k > aH$ is *subhorizon*. As the horizon shrinks during inflation, a mode will eventually exit the horizon to become *superhorizon* ($k < aH$), with the *horizon crossing* condition set by $k = aH$. Superhorizon modes freeze-out and become essentially classical. Later, once the universe has entered radiation domination and the comoving horizon begins growing once more, these modes will

re-enter and become subhorizon again. As can be seen in Fig. 3.2, smaller scale modes (corresponding to larger k) exit later and re-enter earlier.

Horizon re-entry is significant because this is the time at which causal information can first propagate across the scale k^{-1} in a Hubble time. If the density contrast δ associated with k^{-1} is sufficiently large, this can then be communicated gravitationally such that the entire overdense region can begin to collapse. In order to form a primordial black hole, the density contrast must exceed a critical threshold value $\delta_{\rm th}$. Provided this condition is satisfied at horizon re-entry, the entire cosmological horizon will collapse to form a primordial black hole. See Fig. 3.3. The initial mass of the primordial black hole is then set by the

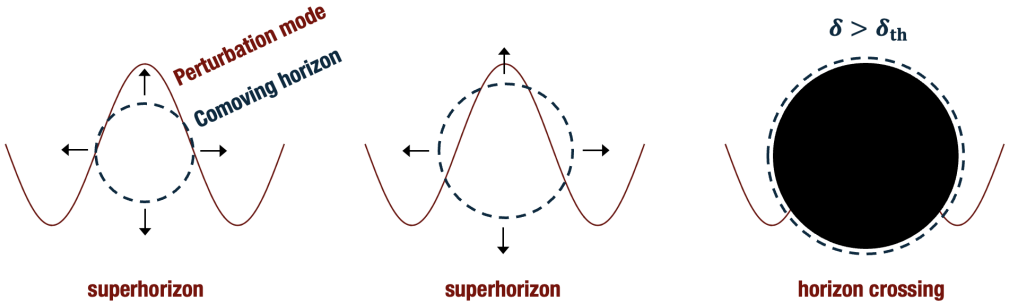

Figure 3.3: As the comoving horizon (blue) expands after inflation, superhorizon modes (red) come to re-enter the horizon when $k = aH$. Provided the density contrast associated with the mode exceeds the critical threshold $\delta > \delta_{\rm th}$, primordial black hole formation will occur upon horizon re-entry.

mass of the cosmological horizon at collapse,

$$M_{\rm PBH} = 10^8\,{\rm g}\left(\frac{\gamma}{0.2}\right)\left(\frac{106.75}{g_\star(T)}\right)^{1/2}\left(\frac{4\times10^{11}\,{\rm GeV}}{T}\right)^2\,, \tag{3.2}$$

where γ is an efficiency factor which takes the value $\simeq 0.2$ during radiation domination and $g_\star(T)$ is the number of effective degrees of freedom in the primordial plasma.

In addition to the mass, the other pertinent piece of information one might be interested

in the is initial abundance,

$$\beta = \frac{\rho_{\rm PBH}}{\rho_{\rm tot}}\,, \tag{3.3}$$

sometimes also called the *mass fraction at formation*. In the standard[2] treatment based on the Press-Schechter formalism [138], β is computed by integrating the probability distribution function[3] $P[\delta]$ for the coarse-grained density contrast,

$$\delta = \frac{\delta\rho}{\bar{\rho}}\,, \tag{3.4}$$

over all values greater than the critical threshold for collapse, $\delta_{\rm th}$

$$\beta \simeq 2\int_{\delta_{\rm th}}^{\infty} d\delta\, P[\delta]\,. \tag{3.5}$$

The factor of 2 is customarily introduced[4] to compensate for the undercounting that otherwise arises [138]. This prescription has the benefit of having a simple implementation and an intuitive interpretation — we are simply summing over the fraction of regions with a sufficient overdensity to collapse to form a primordial black hole.

The resulting initial abundance will depend sensitively on the value for the collapse threshold $\delta_{\rm th}$. Intuitively, we expect a perturbation to be able to collapse during radiation domination if the size of the overdensity at maximum expansion exceeds the Jeans length, or equivalently if it exceeds radiation pressure. This intuition led to Carr's original estimate of $\delta_{\rm th} \simeq c_s^2 = 1/3$ [139], where c_s is the sound speed of density perturbations. If one includes

2. One can also calculate the black hole abundance using peak theory [134, 135, 136, 137]. Unlike Press-Schechter, where the overdensity must simply exceed the threshold, peak theory further demands that it be a local maximum. This formalism has been demonstrated to be more appropriate when perturbations exist on multiple scales. For sharply peaked spectra, the simpler Press-Schechter prescription suffices.

3. The probability distribution function $P[\delta]$, denoted with a non-script "P", should *not* be confused with the power spectrum $\mathcal{P}_\delta(k)$.

4. It is unclear whether this factor should still be included when considering asymmetric probability distribution functions, as in the case of non-Gaussianities. We retain it nevertheless since this is an ultimately inconsequential $\mathcal{O}(1)$ effect.

general relativistic effects, the criterion is slightly modified too $\delta_{\rm th} \simeq 0.4$. A more careful treatment that is applicable for arbitrary equation of state w finds [140]

$$\delta_{\rm th} = \frac{3(1+w)}{5+3w} \sin^2\left(\frac{\pi\sqrt{w}}{1+3w}\right) , \tag{3.6}$$

which yields $\delta_{\rm th} \simeq 0.414$ during radiation domination. It should be noted that there is considerable uncertainty[5] in the proper choice for the collapse threshold, since this should in principle depend on the spatial profile of the overdensity. The value $\delta_{\rm th} \simeq 0.414$ seems to reproduce the results of numerical simulations fairly well for strictly spherical collapses, but an overdensity profile need not be spherical. This is especially the case in the presence of non-Gaussianities, which generically arise in models that produce local amplifications of small scale power. These can then also change the threshold for collapse [143, 144, 145].

While the coarse-grained density contrast is the proper object to consider when computing the probability of primordial black hole formation, it is often convenient to work directly with the curvature perturbation ζ, since its statistics are more easily computed from underlying inflationary models. When large perturbations exist only on one scale, as is the case for the sharply peaked power spectra, only a minimal amount of error is incurred by making this approximation. On superhorizon scales, ζ is related to the density contrast field as [146]

$$\delta = -\frac{2(1+w)}{5+3w}\left(\frac{1}{aH}\right)^2 e^{-2\zeta}\left(\nabla^2\zeta + \frac{1}{2}(\partial_i\zeta)^2\right) . \tag{3.7}$$

5. A more precise quantity which does away with this uncertainty is the compaction function $\mathcal{C}$, defined as twice the local mass excess over radius of the comoving area [141, 142]. The value of the compaction at its peak is interpreted as the threshold.

Working to linear[6] order in radiation domination, this simplifies to the Fourier space relation

$$\delta_k \simeq \frac{4}{9}\left(\frac{k}{aH}\right)^2 \zeta_k \,, \tag{3.8}$$

which implies that their power spectra, defined for arbitrary Fourier variable f_k via the 2-point function as $\langle f_k f_{k'}\rangle = \frac{2\pi^2}{k^3}\mathcal{P}_f(k)\delta^{(3)}(\vec{k}+\vec{k}')$, are related as

$$\mathcal{P}_\delta(k) \simeq \frac{16}{81}\left(\frac{k}{aH}\right)^4 \mathcal{P}_\zeta(k)\,. \tag{3.9}$$

Note that at the time of horizon crossing $k = aH$ when a perturbation can collapse to form a black hole[7], the density contrast and curvature perturbation are linearly related, $\delta \simeq \frac{4}{9}\zeta$. We can then assume that peaks in δ also correspond to peaks in ζ, and work directly with the curvature perturbation. In terms of ζ, the initial abundance in the Press-Schechter formalism reads

$$\beta \simeq 2\int_{\zeta_{\rm th}}^{\infty} d\zeta\, P[\zeta]\,, \tag{3.10}$$

where the collapse threshold $\zeta_{\rm th} \simeq \frac{9}{4}\delta_{\rm th}$ follows from $\sigma_\delta^2 \simeq \frac{16}{81}\sigma_\zeta^2$, since $\mathcal{P} \sim \sigma^2$ for a sharply peaked spectrum. More precisely, the variance of ζ smoothed on the scale $R \simeq (aH)^{-1} \simeq k^{-1}$ can be computed from the power spectrum as [148]

$$\sigma_\zeta^2(R) \equiv \langle\zeta\rangle_R^2 = \int_0^\infty \frac{dk}{k}\tilde{W}^2(k,R)\mathcal{P}_\zeta(k)\,, \tag{3.11}$$

where $\tilde{W}(k,R)$ is the Fourier transform of the (real space) window function used to coarse-grain δ. It is unclear what functional form for the window function most accurately repro-

6. A certain degree of non-Gaussianity inevitably arises in the density contrast field due to this nonlinear relation. This implies that even if the statistics of ζ were perfectly Gaussian, those of δ would not be. See Ref. [146] for discussion.

7. There is additional uncertainty in the calculation of primordial black hole abundances coming from non-linear effects around the time of horizon crossing. See Ref. [147].

duces the actual relation between the primordial black hole abundance and power spectrum, but popular choices in the literature include the volume normalized Gaussian, as well as real and k-space top hats. See [149, 150] for a discussion of the resultant uncertainties.

In principle, the probability distribution function will depend on the model under consideration. For a long time the standard in the literature was to simply assume a Gaussian form for the probability distribution,

$$P_G[\zeta] = \frac{1}{\sqrt{2\pi}\sigma_\zeta} \exp\left(-\frac{\zeta^2}{2\sigma_\zeta^2}\right) , \tag{3.12}$$

in which case the mass fraction at formation evaluates explicitly to

$$\beta = \text{erfc}\left(\frac{\zeta_{\text{th}}}{\sqrt{2}\sigma_\zeta}\right) \simeq \sqrt{\frac{2}{\pi}}\frac{\sigma_\zeta}{\zeta_{\text{th}}} \exp\left(-\frac{\zeta_{\text{th}}^2}{2\sigma_\zeta^2}\right) , \tag{3.13}$$

where erfc is the complimentary error function and the second approximation holds for $\zeta_{\text{th}} \gg \sigma_\zeta$, which is generically the true for all physical cases of interest. Since the threshold is fixed, the initial abundance is determined solely by the variance of the power spectrum. In order to have $\beta = 10^{-20}$, we see we need $\sigma_\zeta^2 \simeq 0.01$. This corresponds to a peak in the power spectrum of $\mathcal{P}_\zeta \sim 0.01$, which is seven orders of magnitude greater than the value measured on CMB scales! Since the amplification of $\mathcal{P}_\zeta$ needed for PBH formation depends on β only logarithmically, this degree of enhancement is a generic requirement for any non-vanishing initial abundance.

Keep in mind however that this discussion has assumed a Gaussian form for the probability distribution function $P[\zeta]$. In Sec. 5.2, we will see how this story changes upon considering non-Gaussian probability distributions. While we will find that the amplification required can be reduced, it will turn out to be the case that one still needs a substantial amplification of power over values on CMB scales. We discuss in the next subsection how such an enhancement may be engineered in inflationary models. For further reading on primordial

black hole formation, Refs [151, 152, 153] may prove useful.

3.1.2 Enhanced Curvature Perturbations from Inflation

For simplicity, we begin by considering single field inflation sourced by a minimally coupled scalar field ϕ with canonical kinetic term and potential $V(\phi)$. The Friedmann equations in combination with the scalar's equation of motion form the following set of equations

$$H^2 = \frac{8\pi}{3M_{\rm Pl}^2}\left(\frac{1}{2}\dot{\phi}^2 + V(\phi)\right)\,, \tag{3.14a}$$

$$\dot{H} = -\frac{4\pi}{M_{\rm Pl}^2}\dot{\phi}^2\,, \tag{3.14b}$$

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\partial V}{\partial \phi} = 0\,, \tag{3.14c}$$

which govern the evolution of the background values of ϕ and a. From the solution to this set of equations, we can compute the duration of inflation, usually quantified in terms of the number of e-folds ΔN

$$\Delta N = \int_{a_i}^{a_f} d\ln a = \int_{t_i}^{t_f} dt\, H\,, \tag{3.15}$$

where a_i (a_f) is the value of the scale factor at the beginning (end) of inflation and t_i (t_f) is the corresponding time. Recall that in order to resolve the horizon and flatness problems, inflation needs to last sufficiently long: $\Delta N \simeq 50 - 60$. Inflation can technically last longer, but fluctuations generated earlier than ~ 60 e-folds before the end of inflation are unobservable.

As for what sort of potential is appropriate, recall that the equation of state in the scalar-dominated universe is

$$w = \frac{p_\phi}{\rho_\phi} = \frac{\frac{1}{2}\dot{\phi}^2 - V}{\frac{1}{2}\dot{\phi}^2 + V}\,. \tag{3.16}$$

From this, we see that if we want a quasi-de Sitter phase with $w \simeq -1$, then the scalar's

potential energy needs to dominate and we should choose a potential which is sufficiently flat. See Fig. 3.4 and ignore for now the small Gaussian bump. Note that the inflaton must be allowed to eventually roll down to its minimum, such that inflation is not eternal and reheating can occur.

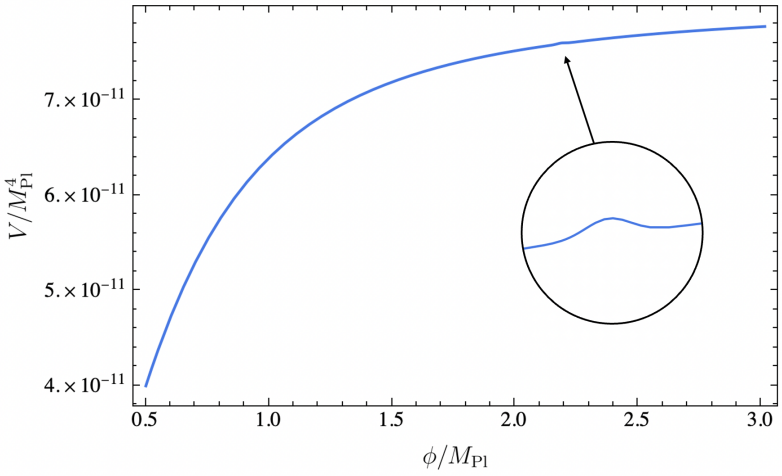

Figure 3.4: Potential of Eq. (3.32), with KKLT base potential V_0 in Eq. (3.33) and a small, localized Gaussian bump of the form Eq. (3.34). Parameters: $A = (8.4 \times 10^{-11})\, M_{\rm Pl}^4$, $M = M_{\rm Pl}/2$, $\delta A = 1.17036 \times 10^{-3}$, $\phi_b = 2.188\, M_{\rm Pl}$, $\sigma_b = 0.0159\, M_{\rm Pl}$.

During the slow-roll phase of inflation, the Hubble friction term dominates and the $\ddot{\phi}$ term is very small. The inflaton's equation of motion then approximately reduces to

$$3H\dot{\phi} \simeq -\frac{\partial V}{\partial \phi}\,. \tag{3.17}$$

It is useful to define the slow-roll parameters[8] ϵ and η as [154]

$$\epsilon \equiv -\frac{\dot{H}}{H^2} = \frac{4\pi}{M_{\rm Pl}^2}\frac{\dot{\phi}^2}{H^2}\,, \tag{3.18}$$

and

$$\eta \equiv -\frac{\ddot{\phi}}{H\dot{\phi}}\,. \tag{3.19}$$

8. These are not to be confused with the *potential* slow-roll parameters ϵ_V and η_V.

The slow-roll regime is then defined by simultaneously satisfying the conditions

$$\epsilon < 1\,, \quad |\eta| < 1\,. \tag{3.20}$$

Quantum fluctuations during inflation lead to the spectrum of primordial density perturbations observed today in the CMB. There are two gauge invariant fields excited during inflation: the scalar $\mathcal{R}$, which is called the *comoving curvature perturbation*, and the transverse traceless tensor h_{ij}. Note that $\mathcal{R}$ is closely related to but not identical to ζ, which we also called the "curvature perturbation" in the previous subsection. More precisely, ζ is the curvature perturbation on *uniform density hypersurfaces* while $\mathcal{R}$ is the curvature perturbation on *comoving hypersurfaces*. On superhorizon scales $k \ll aH$ and during the slow-roll phase of inflation, they are equal [154].

The power spectrum of comoving curvature perturbations $\mathcal{P}_\mathcal{R}$ can be approximated during the slow-roll phase as

$$\mathcal{P}_\mathcal{R} \simeq \frac{4\pi}{M_{\rm Pl}^2}\left(\frac{H}{2\pi}\right)^2 \frac{1}{\epsilon}\bigg|_{k=aH}\,, \tag{3.21}$$

where the right-hand side should be evaluated at horizon crossing. Since we will be interested in deviations from slow-roll, however, we must use the general expression for $\mathcal{P}_\mathcal{R}$. To derive this, it is first necessary to solve the Mukhanov-Sasaki equation [155, 156]

$$v_k'' + \left(k^2 - \frac{z''}{z}\right) v_k = 0\,, \tag{3.22}$$

where $v = z\mathcal{R}$, $z = a\dot{\phi}/H$, and primes denote derivatives with respect to conformal time. To derive this equation, one starts with the action for the minimally coupled scalar

$$S = \int d^4x\,\sqrt{-g}\left(\frac{M_{\rm Pl}^2}{16\pi}\,R - \frac{1}{2}(\partial\phi)^2 - V(\phi)\right)\,, \tag{3.23}$$

and fixes the gauge as

$$\delta\phi = 0\,, \;\; g_{ij} = a^2[(1-2\mathcal{R})\delta_{ij} + h_{ij}]\,, \;\; \partial_i h_{ij} = h_i^i = 0\,. \tag{3.24}$$

Then upon expanding this action to second order in $\mathcal{R}$, one finds

$$S^{(2)}[\mathcal{R}] = \frac{1}{2}\int d^4x\, a^3 \frac{\dot{\phi}^2}{H^2}\left(\dot{\mathcal{R}}^2 - \frac{1}{a^2}(\partial_i\mathcal{R})^2\right)\,. \tag{3.25}$$

Upon defining the Mukhanov-Sasaki variable

$$v = z\mathcal{R}\,, \;\; \text{with} \;\; z = a\frac{\dot{\phi}}{H}\,, \tag{3.26}$$

this takes the very simple form

$$S^{(2)}[v] = \frac{1}{2}\int d^4x\, \left((v')^2 + (\partial_i v)^2 + \frac{z''}{z}v^2\right)\,, \tag{3.27}$$

where primes denote derivatives with respect to conformal time. The Fourier modes v_k then clearly obey the Mukhanov-Sasaki equation of Eq. (3.22). It is possible to express the "mass term" z''/z exactly in terms of a hierarchy of slow-roll parameters [157]

$$\frac{z''}{z} = a^2H^2\left(2 - \epsilon_1 + \frac{3}{2}\epsilon_2 + \frac{1}{4}\epsilon_2^2 - \frac{1}{2}\epsilon_1\epsilon_2 + \frac{1}{2}\epsilon_2\epsilon_3\right)\,, \tag{3.28}$$

where $\epsilon_1 = \epsilon = -\dot{H}/H^2$ and

$$\epsilon_{n+1} = -\frac{d\ln\epsilon}{dN}\,. \tag{3.29}$$

Note that in spite of the appearance of slow-roll parameters, this expression is *exact*.

To solve the Mukhanov-Sasaki equation, we must supply boundary conditions for v_k. It is customary to take the vacuum at early times ($\tau \to -\infty$) when the mode is deep in the

horizon $k \gg aH$ to be the Bunch-Davies vacuum [158], corresponding to

$$\lim_{\tau \to -\infty} v_k = \frac{e^{-ik\tau}}{\sqrt{2k}} \,. \tag{3.30}$$

At early times, the solution for v_k is highly oscillatory, while at late times once the mode becomes superhorizon $k \ll aH$, $|v_k|$ approaches a constant value. The primordial power spectrum for $\mathcal{R}$ is then given by substituting this solution for v_k in

$$\mathcal{P}_{\mathcal{R}} = \frac{k^3}{2\pi^2} \frac{|v_k|^2}{z^2}\bigg|_{k \ll aH} \,, \tag{3.31}$$

where the right-hand side is evaluated once the mode is superhorizon and frozen-out.

Now that this formalism is in place, we are ready to see how choosing a potential with localized features which result in a brief deviation from slow-roll can lead to in amplified power. We consider a general potential of the form

$$V(\phi) = V_0(\phi)\big(1 + \delta V(\phi)\big) \,, \tag{3.32}$$

where the base potential V_0 should be chosen to reproduce observations on CMB scales and $\delta V(\phi)$ is a localized feature. For concreteness, let us consider as our base potential the string-inspired KKLT potential with $n = 2$ [159]

$$V_0(\phi) = A\left(\frac{\phi^2}{\phi^2 + M^2}\right) \,. \tag{3.33}$$

The overall amplitude A must be chosen such that $\mathcal{P}_{\mathcal{R}}(k_p) = 2.1 \times 10^{-9}$, with k_p the CMB pivot scale. As for the deviation, let us chose a Gaussian bump located at ϕ_b with amplitude δA and variance σ_b^2.

$$\delta V(\phi) = \delta A \exp\left(-\frac{1}{2}\frac{(\phi - \phi_b)^2}{\sigma_b^2}\right) \,. \tag{3.34}$$

This potential, including the impact of the bump, is shown in Fig. 3.4. By numerically solving the Mukhanov-Sasaki equation for v_k and substituting in Eq. (3.31), we arrive at the power spectrum of curvature perturbations shown in Fig. 3.5. We also show the power spectrum obtained using the slow-roll approximation of Eq. (3.21) to demonstrate how this miscalculates both the size and position of the peak.

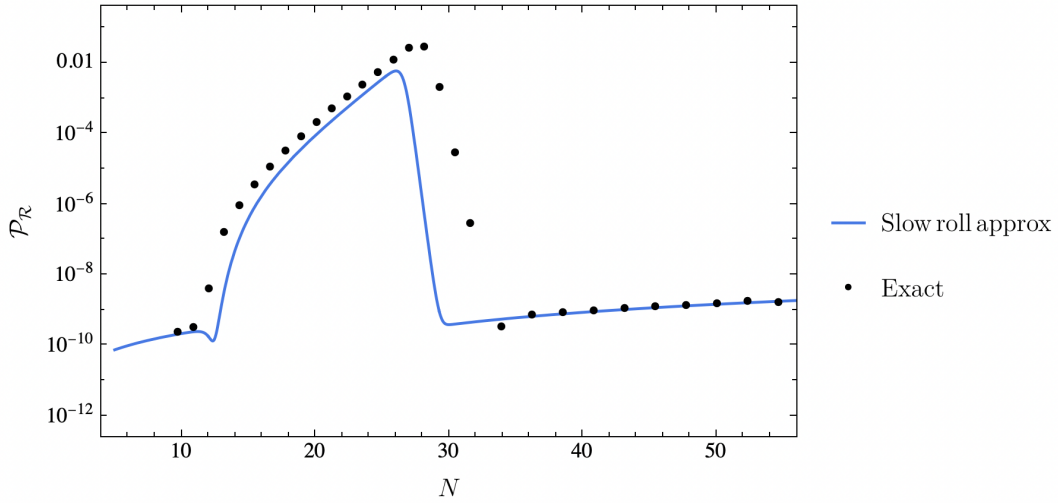

Figure 3.5: Power spectrum of curvature perturbations $\mathcal{P}_\mathcal{R}$ obtained for the potential of Eq. (3.32) with KKLT base Eq. (3.33) and a small Gaussian bump of the form Eq. (3.34). Same parameters as in Fig. 3.4. The solid dots represent the exact solution obtained by solving the Mukhanov-Sasaki equation for v_k and using Eq. (3.31) while the blue line is the approximate slow-roll expression of Eq. (3.21).

This growth can ultimately be traced back to the fact that the slow-roll condition is violated due to the presence of the bump, as can be seen in Fig. 3.6. The first slow-roll parameter remains small throughout, but the second slow-roll parameter $|\eta|$ transiently becomes larger than 1 both when the inflaton decelerates to climb the bump and when the inflaton accelerates once more after surmounting it. As for why a violation of slow-roll results in an enhancement in the power spectrum of curvature perturbations, it turns out that the mode which is the decaying mode during slow-roll gets revived and begins to grow. Since this is a generic result whenever one deviates from the slow-roll attractor, other means of generating amplified power in the single-field case include localized features like

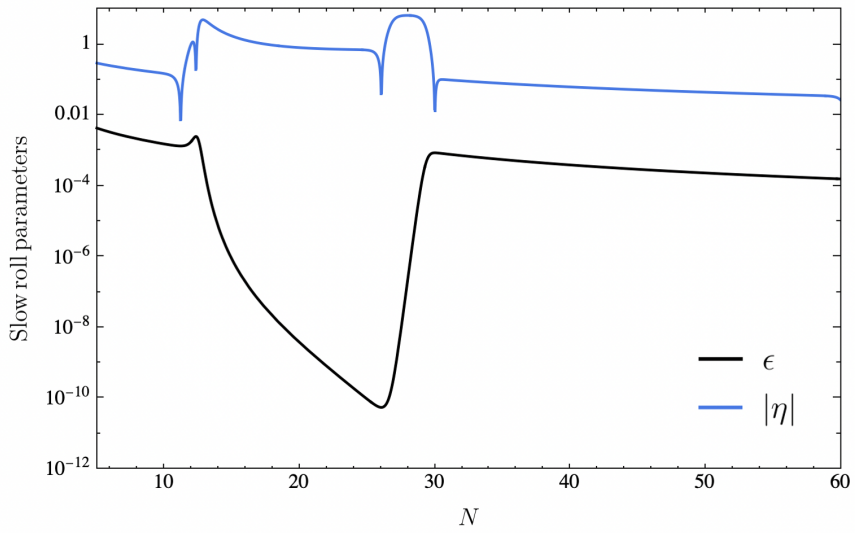

Figure 3.6: Evolution of slow-roll parameters ϵ and $|\eta|$ throughout inflation for the potential Eq. (3.32) with V_0 given by Eq. (3.33) and δV given by Eq. (3.34). Same parameters as in Figs. 3.4 and 3.5. Slow-roll is transiently broken when $|\eta|$ becomes larger than 1.

bumps, dips, steps, and kinks; transient periods of ultra-slow-roll; inflection points; and more [160, 161, 162, 163, 164, 165]. Finally, we comment that given that $\mathcal{P}_{\mathcal{R}}|_{\max} \sim 10^{-2}$ for the sample case presented in this section, this model would be capable of generating a non-trivial number of primordial black holes, even in the absence of non-Gaussianities.

3.2 Constraining Non-Standard Expansion Histories

In the previous section, we saw how deviations from slow-roll behavior during inflation can give rise to amplified power on small scales, potentially resulting in primordial black hole formation. Thus, constraints on primordial black hole abundances today allow us to make inferences about the shape of the inflationary potential. In this section, we turn to primordial black hole evaporation, and ask what constraints on these evaporation products allow us to infer about the expansion history of the universe. Note that much of the material in this section shares considerable overlap with [25], which was recently published by this author.

In particular, we will look at constraints on the expansion history coming from gravita-

tional waves sourced by primordial black holes. Gravitons radiated off of light, evaporating black holes contribute to the stochastic background of gravitational waves. The spectrum of such emission depends on both the mass and the spin of the black holes, as well as on the redshifting that occurs between the black hole formation and today. This, in turn, depends on the expansion history of the universe. Here, we study the features of the stochastic background of gravitational waves from black hole evaporation under a broad range of possible early cosmological histories. We will see that the resulting gravitational wave signals typically peak at very high frequencies, offering opportunities for proposed ultra-high frequency gravitational wave detectors. Lower-frequency peaks are also possible, albeit with a suppressed intensity that is likely well below the threshold of detectability. We find that the largest intensity peaks correspond to cosmologies dominated by fluids with "stiff" equations of state. Such scenarios can be constrained on the basis of violation of $\Delta N_{\rm eff}$ bounds.

3.2.1 Overview

Black holes emit quasi-thermal radiation via the well-known process of Hawking evaporation [166, 167], through which they evaporate at calculable rates into all physical degrees of freedom with mass around or below the associated black hole temperature [167] — including gravitons. As such, black hole evaporation directly produces gravitational radiation, as pointed out long ago [168]. This possibility is especially intriguing for light primordial black holes. In fact, this gravitational wave signal provides one of very few ways to probe light primordial black holes with masses $M \lesssim 5 \times 10^8\,\mathrm{g}$, which evaporate prior to BBN and are otherwise completely unconstrained.

Ref. [169] and [170] first studied the production of a stochastic gravitational wave background from light, evaporating primordial black holes, including the possibility of a phase of early matter domination. We note, however, that both of these studies neglected the black hole angular momentum and its evolution, as well as the corresponding large devia-

tions from blackbody emission [168]. Ref. [171] studied gravitational wave production from a number of mechanisms, including mergers and Hawking evaporation, but for the latter case assumed both instantaneous black hole decay and a blackbody spectrum – assumptions which, as we will demonstrate, are inadequate. Ref. [172] performed a precision study of gravitational wave production from near-extremal Kerr black holes in a standard cosmological setting. Refs. [173] and [174] investigated the contribution to the effective number of neutrino species due to the thermal background of gravitons from evaporating Kerr PBH. Finally, Refs. [175] and [176] studied generic redshift effects in particle production from Kerr PBH, in particular exploring the effects of extended black hole spin and mass distributions.

As we discuss in detail below, the generic expectation for the stochastic gravitational wave background produced by evaporating *Schwarzschild* black holes with a *standard cosmological history* is two-fold:

1. The peak frequency for gravitational wave emission is (see Eq. (3.64) below) $f_{\rm peak} \simeq (1.8 \times 10^{16}\,{\rm Hz})(M/10^5\,{\rm g})^{1/2}$, and thus, even for very light black holes with masses close to the Planck scale, the signal is at ultra-high frequencies.

2. The peak gravitational wave emission has an absolute maximum energy density $\Omega_{\rm GW}h^2\big|_{\rm peak} \simeq 4.2 \times 10^{-7}$ (see Eq. (3.66) below). As such, the gravitational wave emission is possibly large enough to be constrained by measurements of the number of relativistic degrees of freedom (see e.g. the recent detailed study presented in Ref. [174]). However, the signal itself likely remains out of reach of both currently operating gravitational wave telescopes and planned high-frequency detectors (see [6] for a review).

Here, we examine how assumptions about the very early universe affect the expectations summarized above. First, the peak gravitational wave frequency depends critically on how the emitted gravitons redshift, especially at very early times. Secondly, both the peak amplitude and location depend quite sensitively on the spin of the primordial black holes (see e.g. [174]). Spinning black holes radiate gravitons both more abundantly and with

energies peaking at lower frequencies [177]. Finally, the cosmological history drastically affect the maximal gravitational wave intensity, to the level of enhancing it by several orders of magnitude, making graviton emission a prime target for future high-frequency gravitational wave searches.

Using state-of-the-art tools such as the `BlackHawk` package [174], we explore the features of the stochastic background of gravitational waves stemming from Hawking evaporation of light primordial black holes with non-standard, non-radiation-dominated cosmologies at early times. An especially well-motivated scenario is the possibility of an early phase of matter domination [169, 174]; more generically, prior to BBN, the universe's energy density could have been dominated by a species[9] ϕ with a generic equation of state.

In Sec. 3.2.2, we first review primordial black hole evolution and evaporation for generic spinning *Kerr* black holes. Sec. 3.2.2 examines, both numerically and analytically, the gravitational wave production from Hawking evaporation of gravitons, and elucidates the impact of the blackbody and instant decay approximations. In Sec. 3.2.4, we turn to well-motivated non-standard cosmological histories and examine their impact on gravitational waves from the Hawking evporation of gravitons. Finally, Sec. 3.2.4 discusses observational prospects and constraints.

3.2.2 Hawking Radiation and Black Hole Evolution

Kerr Black Holes

In the present analysis, we will be interested in Kerr black holes, which have non-vanishing angular momentum J and therefore preferentially emit higher spin particles, like the spin-2 graviton. In Boyer-Lindquist coordinates, the geometry of a Kerr black hole is described by

9. The species ϕ is **not** to be confused with the inflaton of the previous section.

the metric [178]

$$ds^2 = -\frac{\Delta}{\rho^2}(dt - \alpha \sin^2\theta\, d\phi)^2 + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2 + \frac{\sin^2\theta}{\rho^2}\left[(r^2+\alpha^2)d\phi - \alpha dt\right]^2 , \quad (3.35)$$

where M is the black hole mass, $\alpha = J/M$ is the spin parameter, and we have defined

$$\rho^2 = r^2 + \alpha^2\cos^2\theta\,, \quad \Delta = r^2 + \alpha^2 - \frac{2Mr}{M_{\rm Pl}^2}\,. \quad (3.36)$$

It will also be convenient to define the dimensionless spin parameter $\alpha_\star = \frac{M_{\rm Pl}^2}{M^2}J$, which can take values $\alpha_\star \in [0,1]$, with $\alpha_\star = 1$ corresponding to the extremal case. Kerr black holes have two horizons $r_\pm$, with the outer horizon located at

$$r_+ = \frac{M}{M_{\rm Pl}^2}(1 + \sqrt{1-\alpha_\star^2})\,. \quad (3.37)$$

The Hawking temperature associated with this horizon is

$$T_{\rm BH} = \frac{M_{\rm Pl}^2}{4\pi M}\frac{\sqrt{1-\alpha_\star^2}}{1+\sqrt{1-\alpha_\star^2}}\,. \quad (3.38)$$

Note that this reduces to the temperature of the Schwarzschild black hole $T = \frac{M_{\rm Pl}^2}{8\pi M}$ in the limit of vanishing spin $\alpha_\star \to 0$, and tends to 0 in the extremal limit $\alpha_\star \to 1$. The Kerr black hole is also characterized by an angular velocity $\Omega_{\rm BH}$ given by

$$\Omega_{\rm BH} = \frac{M_{\rm Pl}^2}{2M}\frac{\alpha_\star}{1+\sqrt{1-\alpha_\star^2}}\,. \quad (3.39)$$

Hawking Radiation

The flux spectrum for the emission of a single particle degree of freedom[10] of a species i with frequency ω and spin s is [179, 177, 180]

$$\frac{dN_i(\omega)}{dt} = \sum_{\ell,m} \frac{\sigma^{(s)}_{\ell m}(\omega)}{e^{(\omega - m\Omega)/T_{\rm BH}} - (-1)^{2s}} \frac{d^3k}{(2\pi)^3}\,, \tag{3.40}$$

where the sum runs over the total ℓ and axial m angular momenta of the emitted mode[11]. This spectrum is almost that of a perfect blackbody, with the deviation captured by the "greybody factor" $\sigma^{(s)}_{\ell m}(\omega)$, which is related to the probability that a given mode will be able to surmount the gravitational potential barrier and escape to spatial infinity.

Since the emission of particles with masses greater than the black hole temperature is exponentially suppressed, it often suffices to include in the sum for the total flux only those degrees of freedom lighter than the black hole. In this case we can evaluate the phase space factor to obtain the simplified expression for the emission of a massless degree of freedom per frequency interval

$$\frac{dN_i}{dtd\omega} = \frac{1}{2\pi} \sum_{\ell,m} \frac{\Gamma^{(s)}_{\ell m}(\omega)}{e^{(\omega - m\Omega)/T_{\rm BH}} - (-1)^{2s}}\,, \tag{3.41}$$

where $\Gamma^{(s)}_{\ell m} = \frac{\omega^2}{\pi}\sigma^{(s)}_{\ell m}$ is the absorption probability[12].

Each emitted particle carries off units of energy ω and of angular momentum m about the black hole axis. Note that $m\Omega$ acts as an effective chemical potential, biasing the emission of particles whose angular momentum is aligned with that of the black hole. In this manner,

10. In order to obtain the total flux per particle species i, one would sum over the polarization and charge degrees of freedom.

11. We neglect here the effect of the charge of the particle on $\sigma^{(s)}_{\ell m}$.

12. The absorption probability $\Gamma^{(s)}_{\ell m}(\omega) = \frac{\omega^2}{\pi}\sigma^{(s)}_{\ell m}$ is sometimes referred to in the literature as the "greybody factor". We refrain from doing so here, as they are distinct.

a black hole sheds both mass and angular momentum, and evolves toward a non-rotating state. The power emitted in a given frequency interval per particle degree of freedom is

$$\frac{dE_i}{dtd\omega} = \frac{1}{2\pi}\sum_{\ell,m}\frac{\omega\Gamma^{(s)}_{\ell m}(\omega)}{e^{(\omega-m\Omega)/T_{\rm BH}}-(-1)^{2s}}\,. \tag{3.42}$$

This enters into the rate at which the black hole loses mass as [177]

$$\frac{dM}{dt} = -\sum_i \int d\omega \left(\frac{dE_i}{dtd\omega}\right)\,, \tag{3.43}$$

where the sum runs over all degrees of freedom i emitted by the black hole. Similarly, angular momentum is lost at a rate

$$\frac{dJ}{dt} = -\sum_i \int d\omega \left(m\frac{dN_i}{dtd\omega}\right)\,. \tag{3.44}$$

In practice, it is sometimes convenient to track the evolution by introducing the dimensionless "Page factors" f and g, defined implicitly via [180]

$$f(M,\alpha_\star) = -M^2\frac{dM}{dt}\,, \tag{3.45a}$$

$$g(M,\alpha_\star) = -\frac{M}{\alpha_\star}\frac{dJ}{dt}\,. \tag{3.45b}$$

Given explicit forms for the greybody factors of all relevant particle species, the contributions to f and g from each species can be numerically evaluated and the values tabulated. These can then be interpolated for the functions $f(M,\alpha_\star)$ and $g(M,\alpha_\star)$, from which one can solve Eq. (3.45) to obtain the black hole mass and angular momentum as a function of time. Sample evolutions are shown Fig. 3.7.

Fig. 3.7 illustrates that the black hole mass remains roughly constant near its initial value until the very end of its lifetime, at which point it falls off dramatically. Angular

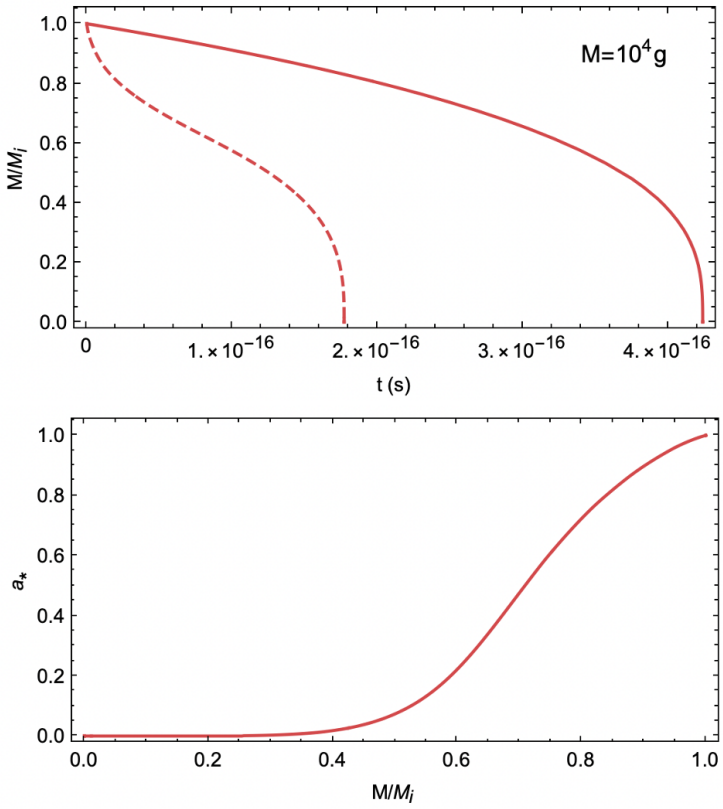

Figure 3.7: Top: Evolution of black hole mass as a function of time for the case of a Schwarzschild black hole (solid red line) as compared with a near extremal ($\alpha_\star = 0.999$) Kerr black hole (dashed line). The lifetime of a rapidly spinning black hole is reduced by a factor ~ 2. $M_i = 10^4$ g is taken as a benchmark, but the above behavior is generic. Bottom: Evolution of dimensionless spin parameter $\alpha_\star$ as a function of M/M_i.

momentum serves to reduce the black hole lifetime by an $\mathcal{O}(1)$ factor, with rapidly spinning black holes evaporating more quickly than their Schwarzschild counterparts. Because angular momentum decreases more rapidly than mass does, Kerr black holes finish shedding angular momentum and transition to a Schwarzschild phase before evaporating completely, as can be seen in Fig. 3.7. Once the black hole has spun down (seen in the lower panel to occur around $M/M_i \sim 0.4$), it evolves identically to the Schwarzschild case, as can be seen by the identical slopes at low $M/M_i \lesssim 0.4$ in the upper panel.

Greybody Factors

While approximating the spectrum as a perfect blackbody usually suffices for estimating the black hole lifetime, computing the gravitational wave signal will require a precise knowledge of the greybody factors. In general, these depend on the frequency, angular momentum, and spin of the emitted particle species, as well as the structure of spacetime about the black hole. Computation of the greybody factors is quite non-trivial and involves solving the relevant equation of motion for a given particle species on a curved black hole background with appropriate boundary conditions — the Teukolsky equations [181, 182]. The absorption probability for a given mode, and thereby the greybody factor, is then determined by taking ratios of the amplitudes for incoming and outgoing waves at infinity.

The Teukolsky equations generically need to be solved numerically[13], though analytic approximations exist in the low frequency limit, $M\omega/M_{\rm Pl}^2 \ll 1$. In particular for the $s = 2$ graviton, the greybody factor, summed over angles, in the low frequency limit reads [187, 177]

$$\sum_{\ell,m} \sigma^{(2)}_{\ell m} \xrightarrow{\omega\sim 0} \frac{16A}{225}\left(5\frac{M_{\rm Pl}^2}{M^2} + \frac{5}{2}\alpha_\star^2 + \alpha_\star^4\right)\left(\frac{M\omega}{M_{\rm Pl}^2}\right)^4 , \tag{3.46}$$

where $A = 4\pi r_+^2$ is the black hole area. Note that this is highly suppressed, scaling with the frequency as ω^4. In contrast, $\sum \sigma^{(0)}_{\ell m} \sim \omega^0$ for scalars, $\sum \sigma^{(1/2)}_{\ell m} \sim \omega^0$ for fermions, and $\sum \sigma^{(1)}_{\ell m} \sim \omega^2$ for vector bosons. The suppression at low frequencies in the graviton case can be understood by recognizing that the dominant contribution to σ comes from the mode of lowest ℓ, and since $\ell \geq s$ this is $\ell = 2$ for the graviton. Meanwhile in the high frequency limit $M\omega/M_{\rm Pl}^2 \gg 1$, the greybody factors for all particle species approach the geometric optics limit, which is essentially the emitting area of the black hole.

13. Numerical solutions typically have issues with convergence at the black hole horizon and spatial infinity. For this reason, `BlackHawk` employs the methods outlined in Ref. [183, 184, 185, 186] to transform the Teukolsky equations into Schrödinger-like wave equations with appropriately chosen short-range potentials.

3.2.3 Standard Cosmological Evolution

Analytical Estimates

We are interested here in the present-day energy density in the form of gravitational waves from evaporating primordial black holes, as parameterized by the spectral density parameter $\Omega_{\rm GW}$, defined as

$$\Omega_{\rm GW} = \frac{1}{\rho_{\rm crit}} \frac{d\rho_{\rm GW}}{d\ln f} \,, \tag{3.47}$$

with $\rho_{\rm crit}$ the critical energy density today. Before turning to numerics, we will demonstrate how this can be computed starting from the instantaneous spectrum of graviton emission, $\frac{dE_{\rm grav}}{dtd\omega}$. For the sake of having analytic expressions, we will restrict ourselves to Schwarzschild black holes ($\alpha_\star = 0$) in the blackbody approximation, for which the greybody factor is simply the frequency-independent area of the black hole $\sum \sigma^{(s)}_{\ell m} = 4\pi r_s^2$. We will also presume instantaneous decay, taking the black hole mass and temperature to be constants up until the moment of evaporation at $\tau_{\rm BH}$. These approximations will be all relaxed in the subsequent sections where we present our numerical results.

Our starting point is the instantaneous energy flux expression of Eq. (3.42). Taking $\sum \sigma_{\ell m} = 4\pi r_s^2 = 16\pi \frac{M^2}{M_{\rm Pl}^4}$ and multiplying by $g_i = 2$ for the two graviton polarizations, this becomes

$$\frac{dE_{\rm grav}}{dtd\omega} \simeq \frac{16}{\pi} \frac{M^2}{M_{\rm Pl}^4} \frac{\omega^3}{e^{\omega/T_{\rm BH}} - 1} \,. \tag{3.48}$$

To obtain the rate of graviton emission for an entire population of evaporating primordial black holes, we multiply by the number density $n_{\rm BH}(t)$

$$\frac{d\rho_{\rm GW}}{dtd\omega} \simeq n_{\rm BH}(t) \frac{dE_{\rm grav}}{dtd\omega} \,. \tag{3.49}$$

This should then be integrated over the black hole lifetime in order to determine the total amount of energy density in the form of gravitational waves at the time of evaporation.

Let t_i be the time of black hole formation, when graviton emission commences, and let $t_* = t_i + \tau_{\rm BH} \simeq \tau_{\rm BH}$ be the time of black hole evaporation, which, for a Schwarzschild black hole, is approximately

$$\tau_{\rm BH} \simeq \frac{10240\pi}{g_{\star,H}} \frac{M^3}{M_{\rm Pl}^4} \,, \tag{3.50}$$

where $g_{\star,H} \simeq 108$ the number of effective degrees of freedom, since we restrict ourselves to light black holes evaporating before BBN, $M \lesssim 5 \times 10^8\,{\rm g}$. Since evaporation is occurring in an expanding universe, the black hole number density and graviton energy density and frequency are not fixed quantities, but rather experience cosmological redshift. In particular, they evolve as $n_{\rm BH} \sim a^{-3}$, $\rho_{\rm GW} \sim a^{-4}$, and $\omega \sim a^{-1}$, respectively. For ease of integration, we can isolate the time dependence by relating the graviton frequency and energy density to their values at the time of evaporation, which we denote by a star

$$\rho_{\rm GW} = \rho_{\rm GW}^* \left(\frac{a_*}{a}\right)^4 \,, \;\; \omega = \omega_* \left(\frac{a_*}{a}\right) \,. \tag{3.51}$$

For the number density, it is more convenient to relate to the initial value

$$n_{\rm BH} = n_{{\rm BH},i} \left(\frac{a_i}{a}\right)^3 \,, \tag{3.52}$$

which, in turn, can be related to the initial black hole mass and mass fraction[14] $\Omega_{{\rm BH},i} = \rho_{{\rm BH},i}/\rho_{{\rm crit},i}$, presuming formation via the collapse of density perturbations in the radiation dominated early universe:

$$n_{{\rm BH},i} = \frac{3M_{\rm Pl}^6}{32\pi M^3} \Omega_{{\rm BH},i} \,. \tag{3.53}$$

Finally converting from frequency interval to logarithmic frequency interval $\frac{d}{d\ln\omega} = \omega \frac{d}{d\omega}$,

14. Note that this is the quantity we called "β" in the previous section.

the energy density in the form of gravitational waves at the time of evaporation is

$$\frac{d\rho^*_{\rm GW}}{d\ln\omega} \simeq \frac{16 n_{{\rm BH},i} M^2 \omega_*^4}{\pi M_{\rm Pl}^4} \int_{t_i}^{t_*} dt\, \frac{(a_i/a)^3}{e^{\omega_* a_*/a T_{\rm BH}} - 1}\,. \tag{3.54}$$

As for the time dependence of the scale factor[15], initially during radiation domination it scales as $a \sim t^{1/2}$. If the initial energy density in black holes is sufficiently large and the black holes are sufficiently long lived, then they will eventually come to dominate the energy density of the universe at a time

$$t_{\rm eq} \simeq \left(\frac{1-\Omega_{{\rm BH},i}}{\Omega_{{\rm BH},i}}\right)^2 \frac{M}{M_{\rm Pl}^2}\,. \tag{3.55}$$

The condition on the initial energy density and mass for this to occur is $t_* > t_{\rm eq}$, or

$$\left(\frac{M}{10^5\,{\rm g}}\right)^2 \left(\frac{\Omega_{{\rm BH},i}}{10^{-11}}\right)^2 \geq 1\,. \tag{3.56}$$

When this is satisfied, the universe will undergo a brief period of early matter domination from $t_{\rm eq}$ until t_*, during which the scale factor scales as $a \sim t^{2/3}$. Thus the scale factor appearing in Eq. (3.54) is

$$a(t) = \begin{cases} a_i \left(\frac{t}{t_i}\right)^{1/2} & t \lesssim t_{\rm eq} \\ a_i \left(\frac{t_{\rm eq}}{t_i}\right)^{1/2} \left(\frac{t}{t_{\rm eq}}\right)^{2/3} & t_{\rm eq} \lesssim t \lesssim t_*\,. \end{cases} \tag{3.57}$$

One can also express $a(t)$ in terms of a_* by noting that

$$\frac{a_i}{a_*} = \left(\frac{t_i}{t_{\rm eq}}\right)^{1/2} \left(\frac{t_{\rm eq}}{t_*}\right)^{2/3}\,. \tag{3.58}$$

15. When we turn to the numerical calculation, we will actually solve the Friedmann equations for the precise background evolution.

Finally to translate the gravitational wave spectrum from evaporation to today, we need to account for the dilution of energy density and redshifting of frequency due to cosmological expansion. The energy density in the form of gravitational waves today is related to that at evaporation as

$$\frac{d\rho^0_{\rm GW}}{d\ln\omega_0} = \frac{d\rho^*_{\rm GW}}{d\ln\omega_*}\left(\frac{a_*}{a_0}\right)^4 , \tag{3.59}$$

where $a_0 = a(t_0)$ is the scale factor today, which we take to be $a_0 = 1$. Explicit factors of the frequency appearing in this expression should be translated to their redshifted values today as $\omega_0 = \omega_* a_*$. Following black hole evaporation, the universe undergoes the usual epoch of radiation domination, and it is convenient to express the ratio of scale factors in terms of the plasma temperature and effective degrees of freedom in entropy, obtained via conservation of entropy $g_{\star,s}a^3T^3 = \text{constant}$ as

$$a_* = \left(\frac{g_{\star,s}(T_0)}{g_{\star,s}(T_{\rm RH})}\right)^{1/3}\frac{T_0}{T_{\rm RH}} , \tag{3.60}$$

where $T_0 = 0.235\,\text{meV}$ is the temperature of the CMB today and $g_{\star,s}(T_0) = 3.91$. The reheating temperature for the SM plasma $T_{\rm RH} \equiv T(t_*)$ can be obtained by equating the energy density in the form of primordial black holes immediately before decay with the energy density in radiation immediately afterwards. Presuming black holes come to dominate prior to decay, this is approximately

$$T_{\rm RH} = 450\left(\frac{g_\star(T_{\rm RH})}{106.75}\right)^{-1/4}\left(\frac{M}{10^5\,\text{g}}\right)^{-3/2}\text{GeV} . \tag{3.61}$$

Substituting the spectral energy density today in the definition of Eq. (3.47), we finally arrive at the prediction for the spectral density parameter today

$$\Omega_{\rm GW} \simeq \frac{\Omega_{{\rm BH},i}}{H_0^2 M}\,\omega_0^4\, I(\omega_0) , \tag{3.62}$$

where $H_0 = 100h\,\mathrm{km}\cdot\mathrm{s}^{-1}\cdot\mathrm{Mpc}^{-1}$ is the Hubble rate, with $h \simeq 0.67 - 0.73$. The non-trivial frequency dependence lies in the integral

$$I(\omega_0) = \int_{t_i}^{t_*} dt\, \frac{(a_i/a)^3}{e^{\omega_0/aT_{\rm BH}} - 1}\,, \tag{3.63}$$

which generically needs to be evaluated numerically for each ω_0. Sample spectra are shown in Fig. 3.8.

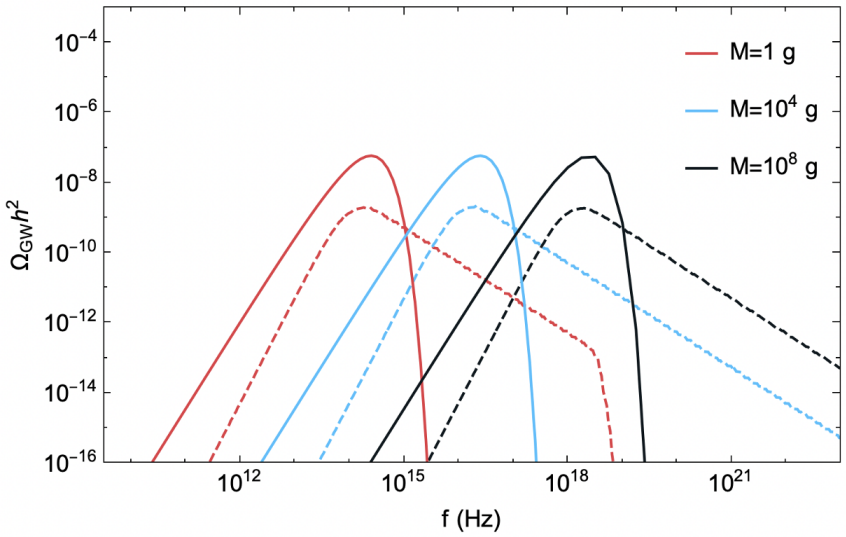

Figure 3.8: Semi-analytic estimate (solid lines) for the spectral density parameter $\Omega_{\rm GW}h^2$ today, presuming a monochromatic spectrum of Schwarzschild black holes of initial mass M and $\Omega_{{\rm BH},i}$ sufficiently large that the PBH eventually dominate the universe energy density, i.e. satisfying Eq. (3.56), such that the black holes come to dominate before decay. We work in the blackbody approximation and presume instantaneous decay. This can be compared with the exact numerical solution (dashed lines), for which these assumptions are relaxed.

The gravitational wave spectrum from graviton production off of Hawking evaporation is almost thermal, but it features more power at low frequencies due to the redshifting of higher frequency modes into lower frequencies. The peak is generically located at very high frequencies — far out of the range of current and near-future gravitational wave detectors. Decreasing the black hole mass shifts the peak frequency to lower values. This is because smaller black holes correspond to earlier formation times and evaporate more promptly,

leading to a much longer period of cosmological redshift which serves to shift the spectrum to lower frequencies. Even saturating the mass bound by considering Planck scale black holes, however, we remain outside of detector sensitivity.

By extremizing $\Omega_{\rm GW}$ with respect to ω_0, one can show that the frequency today peaks at $\omega_{\rm peak} \simeq 2.8 a_* T_{\rm BH}$, or more explicitly

$$f_{\rm peak} \simeq (1.8 \times 10^{16}\,{\rm Hz}) \left(\frac{M}{10^5\,{\rm g}}\right)^{1/2}\,, \tag{3.64}$$

where we have converted to linear frequency $f = \omega/2\pi$. This is consistent with the peak positions in the sample spectra of Fig. 3.8. Evaluating $I(f_0)$ at the peak frequency, which dominates the contribution to the integral, one can show that the following empirical relation holds

$$I(f_{\rm peak}) \simeq (2.3 \times 10^{-33}\,{\rm GeV}^{-1}) \left(\frac{M}{10^5\,{\rm g}}\right)^{-1} \Omega_{{\rm BH},i}^{-1}\,. \tag{3.65}$$

Combining this with Eq. (3.62) evaluated at $f_{\rm peak}$, we find an estimate for the maximal value of the spectral density parameter which is, somewhat surprisingly, independent of both the initial black hole mass and fractional energy density, so long as these are sufficiently large that the black holes come to dominate the energy density of the universe before decay

$$\Omega_{\rm GW} h^2\big|_{\rm peak} \simeq 4.2 \times 10^{-7}\,. \tag{3.66}$$

Note that we have extracted the reduced Hubble rate h to alleviate the associated uncertainty in its value. This estimate is consistent with the peak amplitude shown in Fig. 3.8 as well as with [171]. Comparing with current gravitational wave sensitivities (see e.g. Fig. A3 of [5]), we see that the magnitude of this signal at its peak is within reach of several current and proposed experiments; however this peak occurs as ultra-high frequencies far outside the current regime of observability.

In the coming sections, we will see how a prolonged phase of early matter domination can actually give rise to extra cosmological redshift, which in turn serves to shift the peak emission to lower frequencies. This additional redshift, however, also has the effect of diluting the gravitational wave signal. In fact, since the energy density falls off as four powers of the scale factor $\rho_{\rm GW} \sim a^{-4}$ while the frequency scales with just one $f \sim a^{-1}$, the effect on $\Omega_{\rm GW}$ is much more significant. Thus, to retain a detectable signal, one would need to enhance graviton emission. Recall that this estimate considered Schwarzschild black holes, for which only approximately 1% of energy is emitted as gravitons. In contrast, Kerr black holes preferentially emit particles of higher spin, like the spin-2 graviton. For this reason, we consider Kerr black holes in the remainder of this work.

Finally by comparing with the exact spectra, obtained via `BlackHawk` and shown also in Fig. 3.8 as dashed lines, we see that the blackbody and instantaneous decay approximations are not adequate, although the peak positions coincide almost perfectly (which makes sense as the graviton peak frequency is largely fixed by the black hole temperature independently of greybody factors). As for the peak height, however, we see that the semi-analytic calculation overestimates the amplitude by several orders of magnitude due to the neglect of greybody factors, which otherwise suppress higher angular moments. Interestingly, however, the peak in the case of near-extremal Kerr black holes nearly matches the blackbody Schwarzschild estimate. Finally, the exact spectral shape also differs from the semi-analytic estimate in that it features an extended high frequency tail. As the black hole evaporates, it grows smaller and hotter, and the resultant gravitons emitted towards the end of the black hole lifetime have a higher initial frequency, leading to the formation of the high-frequency tail.

Beyond Blackbody & Instant Decay: Numerical Results

In order to obtain the exact instantaneous graviton spectrum, we use the publicly available code `BlackHawk` [188, 189], which goes beyond both the blackbody and instantaneous decay

approximations. `BlackHawk` uses tabulated and appropriately interpolated greybody factors to precisely compute the emission rates of Eq. (3.41) for all primary particle species. These are then used to solve for the black hole mass and angular momentum loss rates of Eqs. (3.43) and (3.44) in order to obtain the overall black hole evolution. The result is time-dependent spectra which incorporate both greybody factors and the evolution of black hole mass and temperature. While `BlackHawk` does allow for the study of black hole populations with extended mass and spin distributions, we presume monochromatic spectra for simplicity, as our primary focus will be on the effect of modified cosmological expansion histories.

We denote by $Q_{\rm GW}(t,\omega) \equiv \frac{dN_{\rm grav}}{dtd\omega}$ the instantaneous graviton flux, which is an output of `BlackHawk`. The corresponding instantaneous power is $\frac{dE_{\rm grav}}{dtd\omega} = \frac{\omega}{2\pi}Q_{\rm GW}(t,\omega)$ and the instantaneous energy density emitted in the form of gravitational waves from an evaporating population of primordial black holes with number density $n_{\rm BH}(t)$ is

$$\frac{d\rho_{\rm GW}}{dtd\omega} = n_{\rm BH}(t)\frac{\omega}{2\pi}Q_{\rm GW}(t,\omega)\,. \tag{3.67}$$

In order to obtain the total energy emitted in the form of gravitational waves, we need to integrate this quantity over the course of the black hole lifetime, from formation[16] at $t_i \simeq \frac{M}{M_{\rm Pl}^2}$ to evaporation at t_*, which are also outputs of `BlackHawk`. The energy density per logarithmic frequency interval at evaporation then looks like

$$\frac{d\rho^*_{\rm GW}}{d\ln\omega_*} = n_{{\rm BH},i}\left(\frac{a_i}{a_*}\right)^3\frac{\omega_*^2}{2\pi}\int_{t_i}^{t_*}dt\,\frac{a_*}{a(t)}Q_{\rm GW}\big(t,\omega_*\frac{a_*}{a(t)}\big)\,, \tag{3.68}$$

where, once again, we denote quantities evaluated at evaporation with a "$*$". The exact time-dependence of the scale factor $a(t)$ appearing in this expression can be obtained by

16. Black holes formed during radiation domination tend to have negligible spin, and so the formation of near-extremal Kerr black holes will likely require the introduction of new physics. While it is possible to spin up a population of PBH through accretion and mergers, the maximal spin parameter obtained in this way is $\alpha_\star \sim 0.7$. Black holes that form from the collapse of density perturbations during a period of early matter domination, on the other hand, tend to have near-extremal spins [174].

solving numerically the equations governing the background evolution

$$\begin{aligned}
\frac{\dot{a}}{a} &= \sqrt{\frac{8\pi}{3M_{\rm Pl}^2}(\rho_{\rm BH}+\rho_{\rm rad})}\,, \\
\dot{\rho}_{\rm BH} + 3\frac{\dot{a}}{a}\rho_{\rm BH} &= \rho_{\rm BH}\frac{\dot{M}}{M}\,, \\
\dot{\rho}_{\rm rad} + 4\frac{\dot{a}}{a}\rho_{\rm rad} &= -\rho_{\rm BH}\frac{\dot{M}}{M}\,,
\end{aligned} \tag{3.69}$$

where the overhead dot denotes a derivative with respect to coordinate time and the mass loss rate, given schematically in Eq. (3.43), is a `BlackHawk` output. To go from the spectrum at evaporation to that today, we redshift by four powers of the scale factor according to Eq. (3.59), with a_* again given by Eq. (3.60). Finally redshifting also the frequency to today as $\omega_0 = \omega_* a_*$ and dividing by the critical density, we arrive at the fractional energy density in gravitational waves today

$$\Omega_{\rm GW} = \frac{4n_{{\rm BH},i}a_i^3}{3M_{\rm Pl}^2H_0^2}\omega_0^2\int_{t_i}^{t_*}\frac{dt}{a(t)}Q_{\rm GW}\big(t,\frac{\omega_0}{a(t)}\big)\,. \tag{3.70}$$

In Fig. 3.9 we show the spectral shape of the energy density in gravitational waves today $\Omega_{\rm GW}h^2$ for a sampling of initial masses. In all cases we take the initial fractional energy in black holes $\Omega_{{\rm BH},i}$ sufficiently large that they come to dominate the energy density before decay, in which case the precise value of $\Omega_{{\rm BH},i}$ has no bearing on the spectrum. Solid lines correspond to the Schwarzschild case ($\alpha_\star = 0$) while dotted lines correspond to Kerr black holes with near-extremal spin, $\alpha_\star = 0.999$.

We see that the inclusion of spin has a significant impact on both the shape and amplitude of the gravitational wave spectrum today. First, note that for near-extremal black holes the peak position shifts to lower frequencies by about an order of magnitude. This is because rapidly spinning black holes have a reduced lifetime as compared with their non-spinning counterparts, and so the gravitational wave spectrum experiences a longer period

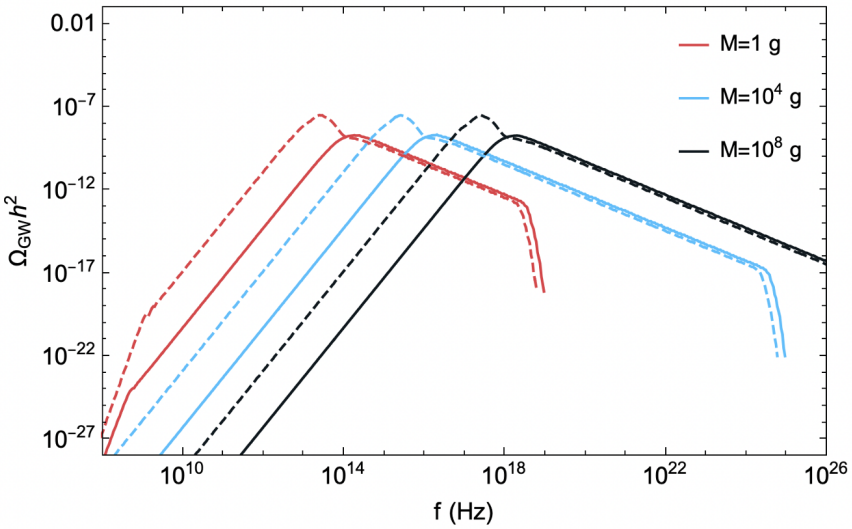

Figure 3.9: Spectral density parameter $\Omega_{\rm GW}h^2$ for a monochromatic spectrum of primordial black holes of initial mass M and mass fraction $\Omega_{{\rm BH},i}$, taken to be sufficiently large that the black holes come to dominate before decay. The peak amplitude is enhanced by several orders of magnitude for near-extremal rotating black holes ($\alpha_\star = 0.999$, dashed line) as compared with the non-rotating case ($\alpha_\star = 0$, solid line). Cosmological evolution is otherwise standard.

of cosmological redshift[17]. The amplitude of the peak is also enhanced by several orders of magnitude due to the black hole spin, which acts as an effective chemical potential biasing the emission of higher spin particles. As was the case for our analytic estimate, we see that smaller mass black holes correspond to spectra that peak at lower frequencies. This is because they can be produced at earlier times and evaporate more quickly, leading to a longer period of cosmological redshift.

We remark that the results of this section are in good agreement with [172], which also considered the gravitational wave signal from evaporating Kerr black holes presuming standard cosmological evolution. They too found that the typical peak frequencies were

17. Based on the semi-analytic estimate, one might naively conclude that this shift towards lower frequencies is due to the fact that rapidly spinning black holes are considerably cooler than those of negligible spin, as can be seen from Eq. (3.38). However for Kerr black holes, typical graviton energies are not of order the temperature, but rather the combination $T_{\rm BH} + m\Omega_{\rm BH}$, with $\Omega_{\rm BH}$ the black hole angular momentum given in Eq (3.39). The large value of $\Omega_{\rm BH}$ for rapidly spinning black holes compensates for the smaller $T_{\rm BH}$, such that the typical energies are comparable or even larger than those of gravitons from the analogous Schwarzschild black hole.

much too high for detection in current and future gravitational wave experiments. In light of this, we now go beyond this analysis to consider the kinds of gravitational signals possible in non-standard cosmologies.

3.2.4 Non-Standard Cosmologies

Early Matter Domination

Gravitational waves from evaporating primordial black holes can constitute an appreciable fraction of the energy density today. However, in most cases, the spectrum peaks at very high frequencies, far outside the reach of current and near-future detectors. Even for the best-case scenario of a population of Planck mass black holes which come to dominate the energy density before decay, the peak frequency is only as low as $\sim 10^{11}$ Hz. One way to effect a shift to lower frequencies would be to invoke a period of early matter domination (EMD). Since the universe expands to a greater extent in a fixed amount of time during matter domination than during radiation domination, this introduces extra cosmological redshift which serves to translate the spectrum to lower frequencies.

To induce this period of EMD, we introduce a heavy auxiliary field ϕ and demand that its initial energy density be greater than that in both black holes and radiation combined $\rho_{\phi,i} > \rho_{\mathrm{BH},i} + \rho_{\mathrm{rad},i}$. The universe will then remain matter dominated through black hole evaporation up until the time τ_ϕ at which ϕ decays, replenishing the SM radiation bath and reheating the universe. To distinguish this from the situation of the previous section, this should occur after black hole evaporation has completed, $\tau_\phi > t_*$. Note that while it is true that the black holes themselves will generically serve to induce a transient period of EMD (presuming they possess sufficient initial energy density and lifetime), this ends at the time of evaporation. If instead EMD comes about as a result of a heavy auxiliary field, then it can last much longer — potentially up until $\sim \mathcal{O}(1)$ s, when radiation domination must begin so as not to spoil the predictions of BBN.

An added benefit to using the auxiliary field to establish EMD is that the black holes can then form during the matter dominated epoch with appreciable spin. For the standard cosmology case of the section prior, we were agnostic as to how the black hole population acquired near-extremal spins. Given that black holes formed from the collapse of overdensities during radiation domination generally have negligible spin, one would have to rely on accretion and mergers, which generally only result in a spin parameter of $\alpha_\star \lesssim 0.7$. On the other hand, black holes formed during matter domination typically have appreciable spin, which is ideal in terms of maximizing the amplitude of the gravitational wave signal.

Recall that in the typical radiation dominated case, a density perturbation will collapse to form a black hole upon re-entering the horizon if its amplitude is greater than some threshold value $\delta_{\rm th}$. Naively applying the analytic formula for the threshold in the matter dominated case, one finds that it vanishes $\delta_{\rm th} \to 0$, which would seem to suggest that any region of overdensity could collapse to a black hole. This is of course not the case, as it turns out that non-spherical effects play a crucial role for collapse during matter domination. As an overdensity begins to collapse, its angular momentum grows significantly and prevents collapse in a majority of cases. Those horizons that do succeed in collapsing, however, form black holes which are rapidly spinning with near-extremal values of $\alpha_\star \sim 1$. See e.g. [190, 191, 174] for further detail regarding the computation of the initial mass fraction $\Omega_{{\rm BH},i}$ in the matter dominated case.

In terms of the gravitational wave signal $\Omega_{\rm GW}$, the derivation proceeds as in the case of standard cosmological evolution, and the spectral density parameter today is still given by Eq. (3.70) with $n_{{\rm BH},i}$ in Eq. (3.53). The difference[18] lies in the input for the initial conditions as well as the evolution of the scale factor $a(t)$ itself. With the addition of the

18. Technically the introduction of a new species changes the particle emission rate from the black hole, but the effect is completely negligible. We could also take ϕ to be sufficiently heavy such that its emission from the black hole is statistically suppressed.

auxiliary field ϕ, Eqs. (3.69) governing the background evolution become

$$\begin{aligned}
\frac{\dot{a}}{a} &= \sqrt{\frac{8\pi}{3M_{\rm Pl}^2}(\rho_\phi + \rho_{\rm BH} + \rho_{\rm rad})}\,, \\
\dot{\rho}_\phi + 3H\rho_\phi &= -\Gamma_\phi \rho_\phi\,, \\
\dot{\rho}_{\rm BH} + 3H\rho_{\rm BH} &= \frac{\dot{M}}{M}\rho_{\rm BH}\,, \\
\dot{\rho}_{\rm rad} + 4H\rho_{\rm rad} &= \Gamma_\phi \rho_\phi - \frac{\dot{M}}{M}\rho_{\rm BH}\,,
\end{aligned} \tag{3.71}$$

where the decay rate is approximately $\Gamma_\phi \simeq \frac{m_\phi^3}{M_{\rm Pl}^2}$ on dimensional grounds. The mass of ϕ should be appropriately chosen to ensure the corresponding lifetime $\tau_\phi = \Gamma_\phi^{-1}$ is after evaporation but before BBN.

Solving this system of equations and evaluating Eq. (3.70) for the spectrum today for various choices of decay time $t_{\rm dec} = t_i + \tau_\phi$ yields the plots in Fig. 3.10. We see that a longer period of EMD is associated with a more heavily redshifted spectrum, as expected. For the best case scenario of a population of Planck mass black holes, letting EMD persist until ~ 1 s can bring the peak frequency as low as $\sim 10^4$ Hz. However, the associated signal strength is vanishingly small at $\Omega_{\rm GW}h^2 \sim 10^{-36}$. Unfortunately, this is a generic side effect of the energy density in gravitational waves falling off as $\rho_{\rm GW} \sim a^{-4}$ while the frequency decreases as $f \sim a^{-1}$.

Generalized Equation of State

Extra cosmological redshift coming from a period of early matter domination, or any period with equation of state $w < 1/3$, will result in lower peak frequencies but also dramatically smaller amplitudes, as previously noted. Given that the peak signal is barely on the cusp of observability $\Omega_{\rm GW} \sim 10^{-7.5}$ in the standard scenario, any period of slower expansion which serves to dilute energy density will lower the signal to outside of sensitivity.

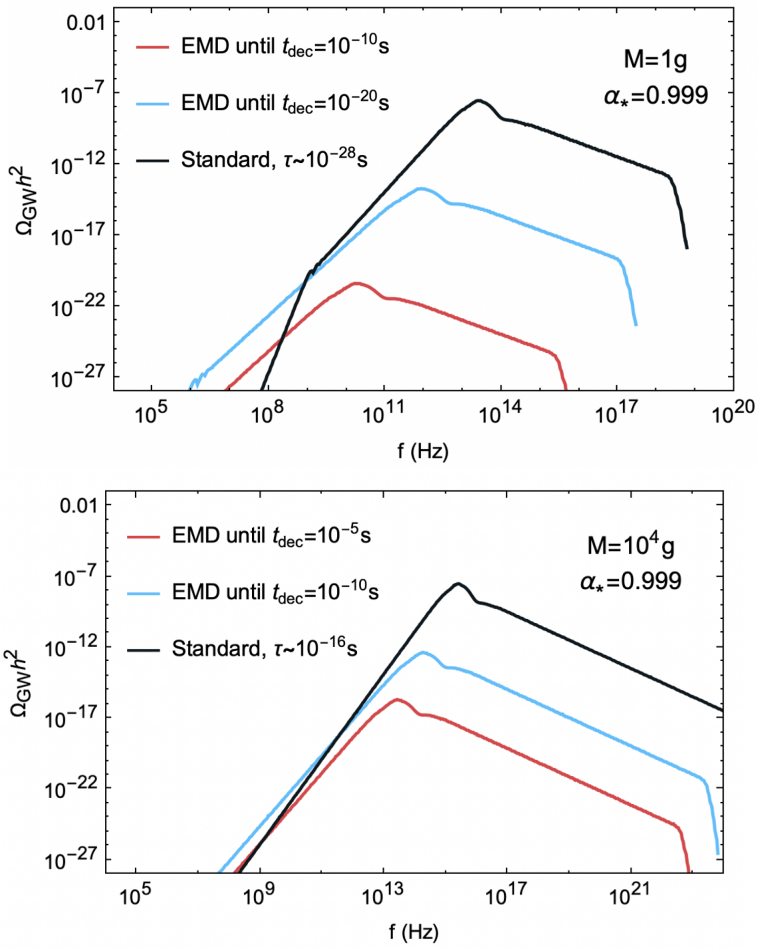

Figure 3.10: Spectral density parameter $\Omega_{\rm GW}h^2$ presuming black hole formation and evaporation during a period of early matter domination induced by the presence of a heavy auxiliary field ϕ which lasts until a time $t_{\rm dec}$.

In this section, we consider how the signal from graviton production changes upon varying the equation of state w of a species that dominates the energy density of the universe at early times. Note that Ref. [192, 193, 194, 195, 196], among others, explored the effect of non-standard cosmologies on gravitational waves produced by large density perturbations in the early universe – a complementary probe to the signal we discuss herein.

We first note that the gravitational wave signal is not the only potential observable associated with graviton emission from PBHs. High energy gravitons emitted by a population of light, evaporating PBH constitute a thermal background of dark radiation, which contributes

to the effective number of neutrino species $N_{\rm eff}$, defined as the contribution to the radiation energy density beyond that of photons

$$\rho_{\rm rad} = \rho_\gamma \left(1 + \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} N_{\rm eff}\right) . \tag{3.72}$$

It is useful to factor this as $N_{\rm eff} = N_{\rm eff}^{\rm SM} + \Delta N_{\rm eff}$, where $N_{\rm eff}^{\rm SM} = 3.046$ accounts for the contribution from SM neutrinos and $\Delta N_{\rm eff}$ parameterizes the departure from the SM prediction. This is tightly constrained by both BBN and CMB measurements. In particular, the one-tailed *Planck* TT, TE, EE+lowE+lensing+BAO constraint is $\Delta N_{\rm eff} < 0.30$ at 95% [2]. The contribution to $\Delta N_{\rm eff}$ from gravitational waves reads

$$\Delta N_{\rm eff} = \frac{8}{7}\left(\frac{11}{4}\right)^{4/3} \frac{\rho_{\rm GW}}{\rho_\gamma} . \tag{3.73}$$

The $\Delta N_{\rm eff}$ bound is an integral bound which applies to the total energy density in gravitational waves integrated over all frequencies. It can thus be interpreted as a bound on the maximum amplitude of the GW spectrum, $\Omega_{\rm GW}^{\rm max}$. Then from $\rho_{\rm GW} \simeq \Omega_{\rm GW}^{\rm max}\rho_{\rm crit}$ and $\rho_\gamma = \frac{\pi^2}{15}T_0^4$, with $T_0 = 0.235\,\text{meV}$, the CMB temperature today, we find

$$\Delta N_{\rm eff} \simeq \frac{120}{7\pi^2}\left(\frac{11}{4}\right)^{4/3} \frac{\rho_{\rm crit}}{T_0^4}\Omega_{\rm GW}^{\rm max} . \tag{3.74}$$

Demanding that $\Delta N_{\rm eff} \lesssim 0.30$ translates to a constraint on the spectral density parameter

$$\Omega_{\rm GW}^{\rm max} \lesssim 3.6 \times 10^{-6} . \tag{3.75}$$

The possibility that the early universe might be dominated by a species whose energy density redshifted differently from radiation would modify the thermal history of the universe and affect both the spectrum of gravitons emitted by evaporating PBHs and the effective

number of relativistic degrees of freedom. We shall consider here cosmological models where at early times the energy density is dominated by a species ϕ with a *generalized* equation of state $P_\phi = w_\phi \rho_\phi$ and $w_\phi > 1/3$, corresponding to a period of faster expansion relative to the standard case. If black hole evaporation occurs before or during ϕ domination, then the energy density in gravitational waves will experience less dilution, potentially giving rise to a signal which saturates the $\Delta N_{\rm eff}$ bound, and which might even be detectable with future, high-frequency gravitational wave searches, as discussed below.

The energy density in ϕ evolves with the scale factor a as

$$\rho_\phi \sim a^{-(4+n)}\,, \tag{3.76}$$

where we follow the same notation as [197, 198] and parameterize the deviation from the radiation scaling with $n = 3w_\phi - 1$. Generically, for a canonically normalized scalar field minimally coupled to gravity, the equation of state is $w_\phi = (K-V)/(K+V)$, where $K = \frac{1}{2}\dot{\phi}^2$ and $V = V(\phi)$ is the potential. Then $w_\phi \to 1$ for $K \gg V$ (a regime dubbed *kination*), and $w_\phi \to -1$ for $K \ll V$, yielding a range of $-4 \leq n \leq 2$. Models with $n > 2$ are also possible[19] for instance in the context of ekpyrotic scenario [201] or with periodic potentials and a varying w_ϕ [202, 203].

The background evolution is described by the following equations

$$\begin{aligned}
&\frac{\dot{a}}{a} = \sqrt{\frac{8\pi}{3M_{\rm Pl}^2}(\rho_\phi + \rho_{\rm BH} + \rho_{\rm rad})}\,,\\
&\dot{\rho}_\phi + (4+n)H\rho_\phi = 0\,,\\
&\dot{\rho}_{\rm BH} + 3H\rho_{\rm BH} = \frac{\dot{M}}{M}\rho_{\rm BH}\,,\\
&\dot{\rho}_{\rm rad} + 4H\rho_{\rm rad} = -\frac{\dot{M}}{M}\rho_{\rm BH}\,,
\end{aligned} \tag{3.77}$$

19. Notice that for $n > 2$ there is no causality violation, despite having $p_\phi > \rho_\phi$, since the sound speed is $c_s^2 = 1$ [199, 200].

with $n > 0$. We specify the initial fractional energy densities $\Omega_{\rm BH,}{}_i$, $\Omega_{\phi,i}$, and $\Omega_{\rm rad,}{}_i = 1 - \Omega_{\rm BH,}{}_i - \Omega_{\phi,i}$ at the time of black hole formation, where $\Omega_{x,i} = \rho_{x,i}/\rho_i$ and

$$\rho_i = \frac{3M_{\rm Pl}^2}{2(4+n)^2\pi}\frac{1}{t_i^2}\,. \tag{3.78}$$

Note that ϕ will generically dominate at early times by virtue of the way its energy density scales with redshift, Eq. (3.76), giving rise to an expansion with $a \sim t^{2/(4+n)}$. We take the PBH to form during this period at a time $t_i = \frac{4}{4+n}\frac{M_i}{M_{\rm Pl}^2}$. As the universe expands, the energy density in ϕ quickly becomes subdominant, as demonstrated in Fig. 3.11, which shows the evolution of the energy densities of the various components for different choices of initial conditions. We denote by t_ϕ the time at which the energy density in ϕ becomes subdominant to that in radiation and black holes, $\rho_\phi(t_\phi) = \rho_{\rm rad}(t_\phi) + \rho_{\rm BH}(t_\phi)$.

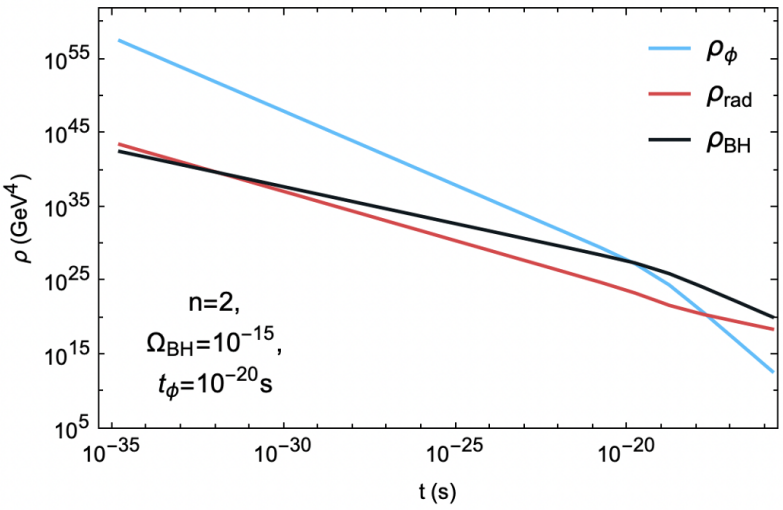

Figure 3.11: Sample evolution of $\rho_{\rm BH}$, $\rho_{\rm rad}$, and ρ_ϕ as obtained by solving Eq. (3.77) for a population of near extremal ($\alpha_\star = 0.999$) PBH with initial mass $M_i = 10^4$ g. We take $n = 2$, such that the initial ϕ domination corresponds to kination. This continues until ρ_ϕ becomes subdominant to $\rho_{\rm BH}$ at a time $t_\phi = 10^{-20}$ s. Black hole domination then continues until evaporation replenishes the radiation bath at $t_* \simeq 3 \times 10^{-15}$ s.

One might wonder about the efficiency of primordial black hole formation during a period of equation of state greater than that of radiation, $w > 1/3$. The authors of [140] have argued

that the threshold for overdensity collapse in general cosmologies with $w \geq 0$ ($n \geq -1$) is given by the following analytic formula[20]

$$\delta_{\rm th} = \left(\frac{n+4}{n+6}\right) \sin^2\left(\frac{\pi}{\sqrt{3}} \frac{\sqrt{n+1}}{(n+2)}\right) , \tag{3.79}$$

where the prefactor assumes comoving gauge. This is based on the Jeans criterion as the determining factor for primordial black hole formation and is derived by demanding that the free-fall timescale be shorter than the soundwave propagation timescale, such that gravitational collapse wins out over the pressure gradient. For radiation domination $n = 0$ it gives $\delta_{\rm th} \simeq 0.414$ while for kination $n = 2$ we have $\delta_{\rm th} \simeq 0.375$. Thus forming primordial black holes during an early period with $n > 0$ should be comparable if not marginally easier when compared with the radiation dominated case. Ref. [204] analyzed primordial black hole formation for $w > 1/3$, presenting specific results for kination, $w = 1$, and additionally studying the induced gravitational waves from large scalar modes in a kination epoch. Furthermore, Ref. [205] studied primordial black hole formation in a scenario with moduli domination and reheating after inflation (see also the recent review in Refs. [196, 206] that discusses the potential effects induced by quantum gravity corrections).

Ignoring for now more exotic scenarios with $n > 2$, the gravitational wave signal will be greatest for the case of kination $n = 2$, during with the energy density redshifts as $\rho \sim a^{-6}$. In such a scenario it is easy to violate the bound on $\Delta N_{\rm eff}$, as shown in Fig. **??**. The duration of kination is fixed by the choice of initial conditions for the fractional energy density in each sector, with a larger initial density corresponding to a longer period of kination (see Fig. **??**). For our benchmark points, we choose Ω_ϕ which gives the desired t_ϕ and set 10% of the remaining energy density to be in the form of PBH, $\Omega_{\rm BH} = 0.1(1 - \Omega_\phi)$. We see that even

20. This nominally vanishes for the matter dominated case $n = -1$, which would seem to suggest any overdensity should collapse to a black hole. However the derivation of [140] posits spherical symmetry, and as we have already argued, deviations from spherical symmetry and angular momentum play a large role in the suppression of primordial black hole formation during matter domination.

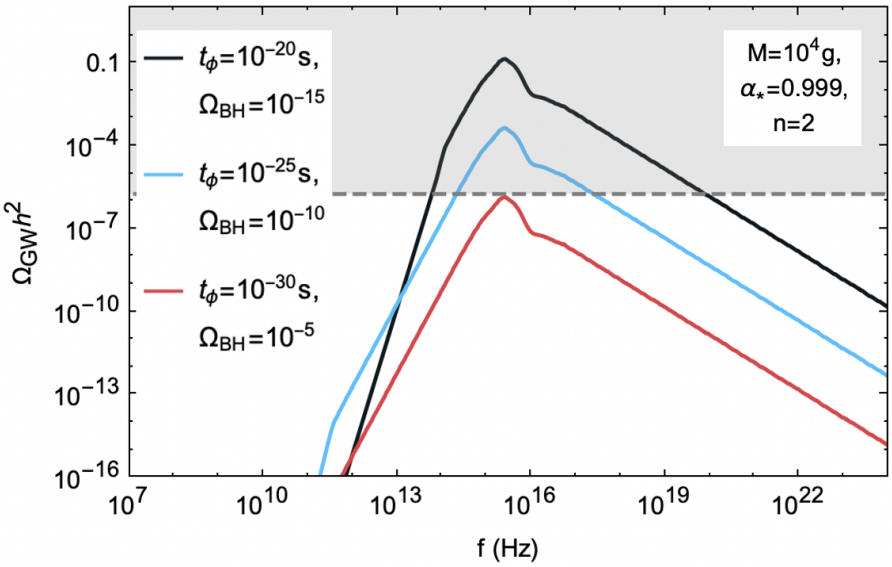

Figure 3.12: Spectral density parameter $\Omega_{\rm GW}h^2$ as compared with the $\Delta N_{\rm eff}$ bound of Eq. (3.75) (grey shaded region) presuming ϕ domination until t_ϕ. Longer periods of kination stemming from larger initial energy density in ϕ result in an amplified gravitational wave signal, potentially contributing inappropriately to $\Delta N_{\rm eff}$.

an extremely transient period of kination, completing long before black hole evaporation completes, can result in a largely boosted signal, which can be ruled out on the grounds of violation of $\Delta N_{\rm eff}$ bounds.

Fig. 3.12 studies, for a mass $M = 10^4$ grams, and quasi-maximal Kerr ($a_* = 0.999$), the spectrum of gravitational waves emitted in kination domination, with differing times t_ϕ at which the kination energy density equals the energy density of the other components, $t_\phi = 10^{-30}$ s (red line), 10^{-25} s (blue line), and 10^{-20} s (black line). The density of gravitational waves increases the longer the period of kination domination, with the peak well inside the region excluded by $\Delta N_{\rm eff}$. We also observe a steepening of the lower-frequency tail (below the peak), while the high-frequency behavior does not depend on the t_ϕ.

3.2.5 Observational Prospects

While the general upper bound on the gravitational wave spectrum from evaporating primordial black holes discussed above prevents the possibility to detect a stochastic gravitational wave signal in a standard cosmological scenario, this conclusion is affected by considering generalized early universe cosmologies, as detailed in the previous section. In particular, for $w_\phi \geq 1$ the gravitational wave amplitude can significantly exceed the $\Delta N_{\rm eff}$ bound, and potentially be detectable with future high-frequency gravitational wave searches.

The detection of ultra-high-frequency ($f \gg 1$ kHz) gravitational waves is an active area of intense experimental investigation (see the recent reviews [6] and [207]). Several experimental techniques have been proposed, ranging from table-top interferometers, holometers, optically levitated sensors, devices based on the inverse Gertsenshtein effect (the conversion of gravitational waves to photons [208]), on gravitational wave to electromagnetic wave conversion in an electric or magnetic field, bulk acoustic-wave devices, superconducting rings, and graviton-magnon resonance.

We consider the following broad classes of experimental techniques, in order of increasing frequency sensitivity, and with a proposed dimensionless sensitivity figure.[21]

- Laser interferometers (1-10) kHz [9×10^{-26}] [209]
- Optically levitated sensors (10-100) kHz [4×10^{-24}] [6]
- Enhanced electro-magnetic conversion ($\sim$10) GHz [10^{-30}] [210, 211, 212]
- Inverse Gersenshtein effect (10^{14}-10^{18}) Hz [3×10^{-30}] [213]

In particular, a promising technology repurposes axion-like particle conversion in a magnetic field to look for graviton conversion [213]; here we highlight current constraints, and the

21. Note that we present our results in terms of the dimensionless characteristic strain h_c, related to the spectral density parameter as:

$$\Omega_{\rm GW} = \frac{4\pi^2}{3H_0^2} f^2 h_c^2 \,.$$

associated relevant frequency range, for JURA (Joint Undertaking on the Research for Axion-like particles) [214], OSQAR II [215], and CAST [216].

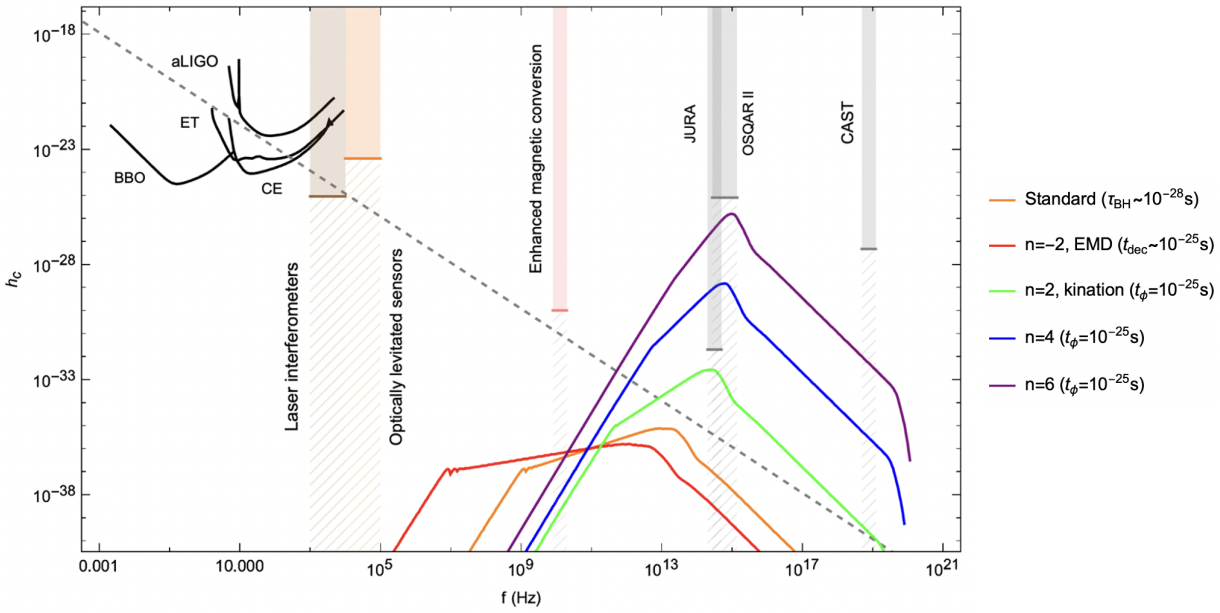

Figure 3.13: Characteristic strain h_c of $M = 1$ g, quasi-extremal ($\alpha_\star = 0.999$) PBHs for a sample of early universe cosmologies compared with the sensitivity of several proposed high-frequency gravitational wave detector technologies, as well as the $\Delta N_{\rm eff}$ bound from *Planck* [2] (grey, dashed).

Fig. 3.13 compares existing proposals for high-frequency gravity wave detectors with our predictions for the gravitational wave emission from primordial black hole evaporation with standard and non-standard early-universe cosmological histories. The latter are taken for the benchmark case of $M = 1$ g, $\alpha_\star = 0.999$ primordial black holes and shown as colored lines: The standard prediction, featuring evaporation at $\tau_{\rm BH} \sim 10^{-28}$ s, is shown in orange. The early matter domination case is shown in red, and assumes that the "extra", non-standard species (in this case, a non-relativistic matter component) decays at $t_{\rm dec} = 10^{-25}$ s. The green line corresponds to a kination scenario where the species responsible for kination becomes subdominant at $t_\phi = 10^{-25}$ s; the blue and purple lines corresponds to an even-faster redshifting species with $n = 4,\ 6$, also becoming subdominant at the same time,

$t_\phi = 10^{-25}$ s.

Our general findings are that for large n (i.e. for a "stiff" equation of state, $w_\phi \geq 1$), the peak gravitational wave emission can be quite "bright", exceeding limits from the number of relativistic species (dashed grey line), and well into the frequency range of future high-frequency gravitational wave detectors; we find that the very high and very low frequency behavior of the gravitational wave spectrum is unchanged, but at frequency around and below the peak frequency, the different redshifting in different early universe cosmologies produces different spectral shapes. Finally, albeit for larger w, the spectrum shifts to higher frequency, we find that the peak spectrum for $n = 2, 4, 6$ only mildly moves to higher frequencies.

In summary, we have considered the spectrum of gravitational waves produced by the evaporation of light, primordial black holes in the early universe in the context of generic cosmological histories. We first discussed the general features of the signal in a standard cosmological setting where the primordial black holes form and evaporate in radiation domination, potentially with a brief period of black hole domination, and highlighted how there is a general upper limit to the intensity of the ensuing gravitational wave stochastic background. We then studied the case of early matter domination by a species different from the primordial black holes themselves, and concluded that the peak of the gravity wave spectrum shifts to lower frequencies, but is also significantly suppressed in intensity, leading to bleak detection prospects. Finally, we entertained scenarios where the early universe is dominated by a species redshifting *faster* than radiation, such as kination of super-stiff fluids. In those cases, while the peak gravitational wave emission is shifted to *higher* frequencies, the intensity of the peak emission is also greatly enhanced, and possibly in conflict with constraints from the number of relativistic degrees of freedom. On a more optimistic note, however, this offers opportunities for discovery for future high-frequency gravitational wave detectors.

3.3 Constraining Extra-Dimensional Scenarios

In the previous section, we saw how the gravitational waves from evaporating primordial black holes could be used to put integral bounds on non-standard cosmologies. Even upon considering expansion histories with extra periods of cosmological redshift, however, the spectral peak remained out of reach of even the most optimistic projections for future detectors. The spectrum of gravitational waves from black hole evaporation generically peaks at frequencies of order the Hawking temperature, which is why this signal is ultra-high frequency for the hot, tiny primordial black holes of the early universe. This fact motivates us to consider small black holes in theories with large extra dimensions, for which the peak frequency can be lowered substantially, since the true bulk Planck scale M_* can be much smaller than the effective $M_{\rm Pl}$.

In this section, we study the emission of brane-localized gravitons during the Hawking evaporation of ultra-light primordial black holes in the context of theories with large extra dimensions, with the ultimate goal of computing the contribution to the stochastic gravitational wave background. To accurately model black hole evolution, we compute greybody factors for all particle species emitted on the brane and in the bulk, presuming the majority of emission proceeds during the Schwarzschild phase. We then compute the power spectrum and present day spectral density parameter for brane-localized gravitons contributing to a gravitational wave signal. We find that for an optimal choice of parameters, the peak frequency plateaus in the sub-MHz regime, within range of planned high-frequency gravitational wave detectors, making this scenario a target for detection once their sensitivity exceeds $\Delta N_{\rm eff}$ bounds. Note that the material in this section shares considerable overlap with [26], which has recently been submitted for publication by this author.

3.3.1 Overview

Theories with large extra dimensions (LED) have long been appreciated as a solution to the hierarchy problem [217, 218, 219]. In such models, the conventional 4-dimensional Planck scale $M_{\rm Pl}$ is merely an effective energy scale related to the true fundamental quantum gravity scale M_* via the relation

$$M_{\rm Pl}^2 \sim R^n M_*^{n+2}\,, \tag{3.80}$$

where R is the characteristic size of one of the n spacelike, compact extra dimensions. Clearly for large $R \gg \ell_{\rm Pl}$, M_* can be lowered significantly, such that gravity's apparent weakness can be understood as simply an artifact of the presence of these n large extra dimensions. In order that the other fundamental forces are not affected beyond phenomenological limits, the particle content of the Standard Model must be confined to a 4-dimensional "brane" embedded in this $(n+4)$-dimensional "bulk", in which only gravitons and possibly heavy Kaluza-Klein scalar modes can propagate. See e.g. [220] for a review. Presuming a fundamental Planck scale $M_* \simeq 10$ TeV, the extra dimensions must be excessively large for $n = 1$, with $R \sim 10^{12}$cm — in conflict with precision tests of Newton's law. Larger values of n are subject to constraints from various cosmological and astrophysical sources as well as collider searches, as we will review, but remain phenomenologically viable and interesting [23].

The dimensionality of spacetime generically affects solutions in general relativity, including black hole solutions. Black holes in LED scenarios are centered on the brane but extend along the extra dimensions. Because the fundamental Planck scale is lowered, black holes can also be produced in energetic particle collisions in the hot thermal plasma of the early universe, in addition to the usual formation mechanisms. Small black holes with horizon radius less than the characteristic scale of the extra dimensions $r_h \ll R$ are $(n+4)$-dimensional objects, while larger black holes with $r_h \gg R$ behave as effectively 4-dimensional. In either case, black hole properties — notably temperature and the spectrum of Hawking radiation — are affected by the presence of these extra dimensions.

In this paper, we investigate the gravitational wave signal from the evaporation of ultra-light primordial black holes in the context of theories with LED. The spectrum of gravitational waves from black hole evaporation peaks at a frequency which is an order one factor times the Hawking temperature, $\omega_{\rm peak} \sim T_H$. In the 4-dimensional case then, $\omega_{\rm peak} \sim M_{\rm Pl}^2/M$, which is generically quite high for primordial black holes with $M \lesssim 5 \times 10^8\,{\rm g}$, as required for evaporation before BBN. It was shown in Ref. [25, 221] that even after redshifting these signals to the present day, the peak frequencies tend to be outside of the range of proposed ultra-high frequency gravitational wave detection technologies. This motivates us to consider black holes in extra dimensional scenarios, which are often colder than their 4-dimensional counterparts of the same mass. As we will see, the peak frequency in this case scales with black hole mass M as

$$\omega_{\rm peak} \sim \left(\frac{M_*}{M}\right)^{\frac{1}{n+1}} M_* \,. \tag{3.81}$$

Since the true quantum gravity scale M_* can be much lower than the observed $M_{\rm Pl}$, we expect $\omega_{\rm peak}$ to be much lower — perhaps even within reach of future gravitational wave detectors.

This section is structured as follows: In Sec. 3.3.2, we briefly review black holes in the LED setup we consider, as well as general constraints on LED scenarios; Sec. 3.3.3 reviews formation mechanisms and mass evolution for extra-dimensional primordial black holes; the following two subsections lay out a detailed calculation of greybody factors for both brane (Sec. 3.3.4) and bulk (Sec. 3.3.5) degrees of freedom; finally, the results are used in Sec. 3.3.6 to produce predictions for the stochastic background of high-frequency gravitational waves from brane-localized gravitons.

3.3.2 Review of Large Extra Dimensions

Black Holes in Large Extra Dimensions

The nature of black holes in extra-dimensional scenarios depends fundamentally on their size relative to the characteristic scale R of the extra dimensions. Black holes with horizon radius $r_h > R$ are larger than the size of the extra dimensions, and so should be relatively insensitive[22] to them, behaving effectively as 4-dimensional objects. In contrast, black holes with $r_h < R$ are fully $(n+4)$-dimensional objects.

While primordial black holes can form through a variety of mechanisms in the early universe, we will primarily consider formation via energetic particle collisions in the hot primordial plasma. This mechanism is unique to the extra-dimensional scenario we will consider, since the lowering of the fundamental Planck scale implies that particle collisions with $E_{\rm CM} > M_*$ can lead to black hole formation. It can be shown that though a black hole formed in such a way generically has angular momentum; this angular momentum is quickly shed and the black hole evolves towards the spherically symmetric Schwarzschild configuration [222].

We will thus consider Schwarzschild black holes in $(n+4)$ dimensions. The line element describing this geometry is [223]

$$ds^2_{n+4} = -h(r)dt^2 + h(r)^{-1}dr^2 + r^2 d\Omega^2_{2+n}\,, \tag{3.82}$$

where $d\Omega^2_{n+2}$ is the line element of the $(n+2)$-dimensional unit sphere and

$$h(r) = 1 - \left(\frac{r_h}{r}\right)^{n+1}\,. \tag{3.83}$$

The horizon radius r_h appearing in this expression can be written as a function of n, the

22. The presence of the extra dimensions still affects the evaporation spectrum of these black holes, as well as other properties.

black hole mass M, and M_* as

$$r_h = \frac{1}{M_*}\left(\frac{M}{M_*}\right)^{\frac{1}{n+1}}\left(\frac{8\Gamma\big((n+3)/2\big)}{(n+2)\pi^{(n+1)/2}}\right)^{\frac{1}{n+1}}, \tag{3.84}$$

and enters into the expression for the Hawking temperature associated with the black hole via

$$T_H = \frac{n+1}{4\pi r_h}. \tag{3.85}$$

In the LED scenarios under consideration, black holes radiate gravitons into the bulk[23] and Standard Model particles on the brane. In the black body approximation, the power emitted in both forms scales as

$$P \sim T_H^{n+4} r_h^{n+2} \sim T_H^2, \tag{3.86}$$

regardless of the number of dimensions. The black hole mass then evolves as $-\frac{dM}{dt} \sim T_H^2$, which can be solved for the lifetime

$$\tau_{\rm BH} \sim \frac{1}{M_*}\left(\frac{M}{M_*}\right)^{\frac{n+3}{n+1}}. \tag{3.87}$$

Comparing with the strictly 4-dimensional result, which scales as $\tau_{\text{4-dim}} \sim M^3/M_{\rm Pl}^4$, we see that an extra-dimensional black hole will usually[24] be longer-lived than a 4-dimensional black hole of the same mass.

Constraints on Extra-Dimensional Scenarios

The constraints on the number of extra spatial dimensions and the bulk Planck scale M_* are coupled and broadly originate from astrophysical and cosmological observations, precision

23. The graviton zero mode is also confined to the brane.

24. This is only strictly true for black holes satisfying $r_h < R$. When the radius becomes of order the size of the extra dimensions, one needs to carry out the full calculation including the numerical prefactors, which depend on n and can be quite large.

tests of gravity at sub-mm distances, and collider searches. We summarize some of the constraints in Table 3.1 and describe them below here. Note that scenarios with $n = 1$ are ruled out and those with $n = 2$ are strongly constrained. However constraints are generally much weaker for $n > 2$ [23].

n	$M_* \geq$	Source
2	4 TeV	Tests of gravity
	27 TeV	SN1987A
	1700 TeV	NS heating
3	76 TeV	NS heating
≥ 4	~ 5 TeV	Colliders

Table 3.1: A sample of constraints on the $(n + 4)$-dimensional Planck scale M_* for various numbers of large extra spatial dimensions n. See [23] for a detailed review of the constraints.

For $n = 2$, precision tests of gravity (i.e. deviations from Newton's law) limit the size of extra dimensions for at the level of $M_* > 4.0$ TeV [224]. Energy losses to Kaluza-Klein (KK) modes in stars, including supernovae, provide very stringent constraints: Ref. [225] calculates that energy losses to KK modes constrain $M_* > 27$ TeV for $n = 2$ and $M_* > 2.4$ TeV for $n = 3$. Since, after a supernova explosion, KK gravitons are gravitationally trapped in the remnant neutron star, heating from KK decays in the neutron stars constrains $M_* > 1700$ TeV for $n = 2$ and $M_* > 76$ TeV for $n = 3$ [225]. Note that decays to non-interacting "dark sector" degrees of freedom can weaken this bound. Cosmological constraints stem chiefly from the concern that particle production in the early universe (in particular, production of relic KK gravitons) should not over-close the universe. This constrains, for instance, $M_* > 7$ TeV for $n = 2$ [226]. Stronger bounds arise from distortion of the cosmic diffuse gamma radiation, however such constraints depend on whether KK gravitons can decay into dark sector states [227].

The most stringent collider constraints hinge on the production of KK graviton modes, or other modes that eventually decay into KK gravitons, which escape collider detection and manifest as missing transverse energy. While in the standard 4D case such processes

are highly suppressed by the "standard" Planck scale, the reduced size of the true Planck scale M_* makes them, in principle, testable at colliders such as the LHC. The constraints are increasingly weak with a larger number of extra spatial dimensions and result in an overall bound on the bulk Planck scale $M_* \gtrsim 5$ TeV for $n \geq 4$ [23]. More specifically, the ATLAS analysis with 139 fb^{-1} luminosity at a 13 TeV center-of-mass energy in [228] quotes a 95% C.L. limit of $M_* > 11.2$ TeV for $n = 2$ and of $M_* > 5.9$ TeV for $n = 6$. For a similar analysis with 137 fb^{-1} luminosity, CMS quotes a 95% C.L. limit of $M_* > 10.7$ TeV for $n = 2$ and of $M_* > 5.5$ TeV for $n = 6$ [229].

3.3.3 LED Black Hole Formation and Evaporation

Formation in High-Energy Particle Collisions

Primordial black holes can be formed through a variety of mechanisms in the early universe, including first-order cosmological phase transitions, the collapse of topological defects, and the collapse of primordial overdensities seeded by inflation — which we have already discussed at length in Sec. 3.1.1. Regardless of the precise mechanism, it is suspected to be easier to produce black holes in extra-dimensional scenarios since $r_h^{(n+4)} > r_h^{(4)}$ means that matter on the brane requires less compression to form a horizon [230]. In these extra-dimensional scenarios, tiny black holes can also be produced in high-energy particle collisions in the hot thermal plasma of the early universe, provided the center-of-mass energy of the colliding particles exceeds the quantum gravity scale $E_{\rm CM} \gtrsim M_*$. For a plasma of temperature $T \lesssim M_*$, black holes produced in this way have typical masses not much larger than the Planck scale $M \gtrsim M_*$, since the production of more massive black holes is Boltzmann suppressed. It was therefore initially thought that the tiny black holes formed through collisions would be too small and hot, evaporating essentially instantaneously and leaving no observational signatures.

Refs. [231, 232] and others have since noted that this need not be the case since accretion

plays a vital role in the presence of extra dimensions. Since the accretion rate is proportional to the horizon area and since extra-dimensional black holes are generically larger than their 4-dimensional counterparts of the same mass, the effects of accretion cannot always be neglected. When accretion initially dominates over evaporation, these black holes can quickly grow to macroscopic size, asymptoting to a mass which is independent of the initial value. This asymptotic mass $M_{\rm asymp}$ depends on the nature of the bulk; for a bulk which is fully populated with gravitational radiation in thermal equilibrium with the brane, the mass will be that of a black hole with size of the extra dimensions [231]

$$M_{\rm asymp} = \frac{(n+2)\pi^{(n+1)/2}}{8\Gamma[(n+3)/2]} \left(\frac{M_{\rm Pl}}{M_*}\right)^{2(n+1)/n} M_* \,. \tag{3.88}$$

We will be more interested in the case that the bulk is not thermalized, in which case an initially microscopic ($r_h < R$) black hole will accrete matter at a rate [232]

$$\frac{dM_{\rm acc}}{dt} = f_{\rm acc}\, A_h\, \rho_{\rm rad} \,, \tag{3.89}$$

where $f_{\rm acc}$ is an $\mathcal{O}(1)$ accretion efficiency, $A_h = 4\pi r_h^2$ is the horizon area, and $\rho_{\rm rad} = \frac{\pi^2}{30} g_\star(T) T^4$ is the radiation energy density, with $g_\star$ the number of relativistic degrees of freedom on the brane. It is convenient to re-express this as

$$\frac{dM_{\rm acc}}{dt} = \left(\frac{\pi}{120}(n+1)^2 f_{\rm acc} g_\star\right) \left(\frac{T}{T_H}\right)^4 T_H^2 \equiv \chi_n \left(\frac{T}{T_H}\right)^4 T_H^2 \,. \tag{3.90}$$

Meanwhile, black hole evaporation proceeds schematically at a rate

$$\frac{dM_{\rm evap}}{dt} = -\alpha_n T_H^2 \,, \tag{3.91}$$

where α_n encodes the greybody factors and depends only weakly on T_H (see the following

sections for details). The black hole mass evolves according to the sum of these terms

$$\frac{dM}{dt} = \left[-\alpha_n + \chi_n \left(\frac{T}{T_H} \right)^4 \right] T_H^2 \,. \tag{3.92}$$

For initial black hole masses above the threshold $M_M^{\rm thresh}$, given as

$$M_{\rm thresh} = \left(\frac{n+1}{4\pi a_n} \left(\frac{\alpha_n}{\chi_n} \right)^{1/4} \frac{M_*}{T} \right)^{n+1} M_* \,, \tag{3.93}$$

the latter term dominates the former, and the black hole preferentially accretes. In this case, we can approximate the evolution as $\frac{dM}{dt} \simeq \chi_n \frac{T^4}{T_H^2}$. During radiation domination

$$H = \frac{1}{2t} = \left(\frac{4\pi^3}{45} g_\star \right)^{1/2} \frac{T^2}{M_{\rm Pl}} \,, \tag{3.94}$$

and so time and temperature can be related as

$$\frac{dT}{dt} = -\left(\frac{4\pi^3}{45} g_\star \right)^{1/2} \frac{T^3}{M_{\rm Pl}} \,. \tag{3.95}$$

Keeping in mind also that M appears in T_H according to Eqs. (3.84) and (3.85), this differential equation can be solved for the asymptotic mass

$$M_{\rm asymp} = \left[\left(\frac{\pi^3}{20} g_\star \right)^{1/2} \left(\frac{n-1}{n+1} \right) f_{\rm acc} a_n^2 \right]^{\frac{n+1}{n-1}} \left(\frac{M_{\rm Pl} T^2}{M_*^3} \right)^{\frac{n+1}{n-1}} M_* \,. \tag{3.96}$$

While the initial fractional abundance[25] $\Omega_{\rm BH}^i$ in the gravitational collapse formation scenario can essentially be treated as a free parameter, for collision-driven production it depends on the thermal distribution of particles in the plasma. Approximating the relative

25. Note that in Sec. 3.1.1 we referred to this quantity as β.

velocity between particles as relativistic, the black hole formation rate per unit mass is [231]

$$\frac{dn_{\rm BH}}{dM} \simeq \frac{a_n^2 g_\star(T)^2}{8\pi^3} M T^2 \left(\frac{M}{M_*}\right)^{\frac{2(n+2)}{n+1}} \left[\frac{M}{T} K_1\left(\frac{M}{T}\right) + 2K_2\left(\frac{M}{T}\right)\right] \Theta(M - M_*)\,, \quad (3.97)$$

where $K_\nu(x)$ are modified Bessel functions of the second kind. Note that for low temperatures $T \ll M_* \lesssim M$, production becomes Boltzmann suppressed since $K_\nu\left(\frac{M}{T}\right) \simeq \sqrt{\frac{T}{M}} \exp\left(-\frac{M}{T}\right)$ in this limit. Thus the dominant production will occur at high temperatures, with the most efficient production at the reheating temperature $T_{\rm re}$. Despite the fact that this production would seem to give rise to an extended mass function, because rapid accretion leads to the asymptotic mass of either Eq. (3.88) in the case of a thermalized bulk or Eq. (3.96) otherwise, the result is a population of tiny primordial black holes with an essentially monochromatic mass function set by n, M_*, and the reheat temperature $T_{\rm re}$ alone. The initial fractional abundance $\Omega^i_{\rm BH} = \rho^i_{\rm BH}/\rho^i_{\rm tot}$ can be obtained from $\rho_{\rm BH} = M n_{\rm BH}$, with the number density obtained by integrating Eq. (3.97) over M.

Hawking Radiation in Large Extra Dimensions

While accretion may initially dominate over evaporation for small black holes with $r_h < R$, once the temperature of the ambient plasma has fallen sufficiently relative to the Hawking temperature, evaporation will become the dominant process governing black hole evolution. Just as in the 4-dimensional case discussed in the previous section, a black hole with temperature T_H evaporates by emitting Hawking radiation with a thermal spectrum which is almost that of a perfect blackbody. Explicitly for the Standard Model particles and zero-mode gravitons emitted on the brane, the flux spectrum is [179]

$$\frac{dN^{(s)}(\omega)}{dt} = \sum_\ell \sigma^{(s)}_\ell(\omega) \frac{1}{e^{\omega/T_{\rm BH}} \mp 1} \frac{d^3k}{(2\pi)^3}\,, \quad (3.98)$$

where the spin statistics factor is ∓ 1 for bosons/fermions, respectively. Note that this expression applies for a given particle degree of freedom with spin s and angular momentum ℓ; to capture the contribution from an entire elementary particle, it is necessary to sum over the contributions from all spin or polarization degrees of freedom. The expression for the power spectrum, or energy emitted per unit time, is similarly

$$\frac{dE^{(s)}(\omega)}{dt} = \sum_{\ell} \sigma_{\ell}^{(s)}(\omega) \frac{\omega}{e^{\omega/T_{\rm BH}} \mp 1} \frac{d^3k}{(2\pi)^3} \,. \tag{3.99}$$

The degree to which this spectrum deviates from that of a perfect blackbody[26] is parameterized by the greybody factor $\sigma_{\ell}^{(s)}(\omega)$. Again, this factor appears because a particle emitted by the black hole needs to surmount the strong gravitational field about the black hole in order to reach an observer at infinity. The presence of this gravitational barrier distorts the shape of the spectrum from that of a blackbody distribution. It is conventional to relate the greybody factor to the absorption coefficient[27] $\mathcal{A}_{\ell}^{(s)}(\omega)$ as

$$\sigma_{\ell}^{(s)}(\omega) = \frac{\pi}{\omega^2} \mathcal{N}_{\ell} \left| \mathcal{A}_{\ell}^{(s)}(\omega) \right|^2 , \tag{3.100}$$

where $\mathcal{N}_{\ell} = 2\ell + 1$ is the multiplicity of states for partial wave ℓ. The absorption coefficient in turn can be defined in terms of incoming and outgoing energy fluxes at the horizon $\mathcal{F}^H$ and at infinity $\mathcal{F}^{\infty}$,

$$\left| \mathcal{A}_{\ell}^{(s)}(\omega) \right|^2 = 1 - \frac{\mathcal{F}^{\infty}_{\rm out}}{\mathcal{F}^{\infty}_{\rm in}} = \frac{\mathcal{F}^{H}_{\rm in}}{\mathcal{F}^{\infty}_{\rm in}} \,, \tag{3.101}$$

and thus can be computed by solving the equations of motion for particles propagating in this black hole background. This will be the task of the following sections.

Higher-dimensional black holes with horizons smaller than the characteristic scale of the

26. For a perfect blackbody, the greybody factor is simply the area of the emitting body.

27. The modulus square of the absorption coefficient $|\mathcal{A}_{\ell}^{(s)}(\omega)|^2$ is equivalent to the absorption probability $\Gamma_{\ell m}^{(s)}(\omega)$ of Sec. 3.2.

extra dimensions $r_h < R$ also emit massive Kaluza-Klein scalar and graviton modes which propagate in the $(n+4)$-dimensional bulk. Even though only the Hawking radiation on the brane is visible, it is essential to account for emission in the bulk, as the energy lost to the bulk determines the energy remaining for brane emission. The expressions for particle and energy flux in this case are completely analogous

$$\frac{d\hat{N}^{(s)}(\omega)}{dt} = \sum_{\ell} \hat{\sigma}_{\ell}^{(s)}(\omega) \frac{1}{e^{\omega/T_{\rm BH}} \mp 1} \frac{d^{n+3}k}{(2\pi)^{n+3}}\,, \tag{3.102}$$

$$\frac{d\hat{E}^{(s)}(\omega)}{dt} = \sum_{\ell} \hat{\sigma}_{\ell}^{(s)}(\omega) \frac{\omega}{e^{\omega/T_{\rm BH}} \mp 1} \frac{d^{n+3}k}{(2\pi)^{n+3}}\,, \tag{3.103}$$

though note that we have introduced hats to signify bulk quantities. For bulk modes, the greybody factor is related to the absorption coefficient $\hat{\mathcal{A}}_{\ell}^{(s)}(\omega)$ as [233]

$$\hat{\sigma}_{\ell}^{(s)}(\omega) = \frac{2^{n+1}\pi^{(n+1)/2}\Gamma\left(\frac{n+3}{2}\right)}{\omega^{n+2}} \hat{\mathcal{N}}_{\ell} \left|\hat{\mathcal{A}}_{\ell}^{(s)}(\omega)\right|^2\,, \tag{3.104}$$

where the combinatorial factor $\hat{\mathcal{N}}_{\ell} = \frac{(2\ell+n+1)(\ell+n)!}{\ell!(n+1)!}$ is the multiplicity of states for partial wave ℓ in $(n+4)$-dimensions. The definition of the absorption coefficient in terms of incoming and outgoing energy fluxes at the horizon and infinity is analogous to the 4-dimensional expression

$$\left|\hat{\mathcal{A}}_{\ell}^{(s)}(\omega)\right|^2 = 1 - \frac{\hat{\mathcal{F}}^{\infty}_{\rm out}}{\hat{\mathcal{F}}^{\infty}_{\rm in}} = \frac{\hat{\mathcal{F}}^{H}_{\rm in}}{\hat{\mathcal{F}}^{\infty}_{\rm in}}\,. \tag{3.105}$$

Though we are ultimately interested in the power spectrum of the zero-mode gravitons on the brane, which will give rise to the stochastic gravitational wave signal, the energy available to these gravitons will depend on that expended in emitting other brane-localized species as well as that lost to the bulk. Accurately modeling black hole evaporation thus requires a knowledge of greybody factors for all species emitted. Determining these greybody factors is the goal of the following sections; we first calculate greybody factors for brane-localized

modes in Sec. 3.3.4 before turning to bulk modes in Sec. 3.3.5. The calculation is quite involved, so we summarize the main results in Table 3.2.

3.3.4 Greybody Factors I: Brane Localized Modes

Brane-localized fields propagate in a gravitational background induced from the higher $(n+4)$-dimensional black hole solution, Eq. (3.82). If we project this solution onto the brane by fixing the angular coordinates $\theta_i = \pi/2$ for the additional compact n dimensions, the induced 4-dimensional line element takes the form

$$ds_4^2 = -h(r)dt^2 + h(r)^{-1}dr^2 + r^2 d\Omega_2^2\,, \tag{3.106}$$

with $h(r)$ given in Eq. (3.83). Greybody factors are obtained by solving master equations describing the motion of a particle of spin s on this background. These master equations can be derived by making use of the Newman-Penrose formalism [234], which is reviewed in Refs. [235, 236, 237].

We begin by factorizing the field as $\Psi_s = e^{-i\omega t} R_s(r) Y^s_{\ell m}(\theta)$, where the angular eigenfunctions $Y^s_{\ell m}(\theta) = e^{im\varphi} S^s_{\ell m}(\theta)$ are spin-weighted spherical harmonics[28]. The radial equation for $R_s(r)$ reads [236]

$$\begin{aligned}\Delta^{-s}\frac{d}{dr}\left[\Delta^{s+1}\frac{dR_s}{dr}\right] \\ + \left[\frac{\omega^2 r^2}{h} + is\omega r\left(2 - \frac{rh'}{h}\right) + s(\Delta'' - 2) - (2s-1)(s-1)rh' - \lambda_{s\ell}\right] R_s = 0\,,\end{aligned} \tag{3.107}$$

where we define $\Delta = hr^2$ and $\lambda_{s\ell} = \ell(\ell+1) - s(s+1)$ is a separation constant. This radial equation is almost identical to the usual 4-dimensional Teukolsky equation [238], but differs

28. The spin-weighted spherical harmonics $Y^s_{\ell m}(\theta)$ are generalizations of the usual spherical harmonics, and reduce to them for $s = 0$.

by the presence of the $s(\Delta'' - 2)$ term, which vanishes when $n = 0$.

We see that the radial equation and thereby also the greybody factors for brane-localized modes retain a dependence on the number of extra dimensions, which follows from the explicit appearance of n in the line element induced by the higher-dimensional theory. Note also that this expression differs from the "master equation" quoted in Refs. [239, 237, 222] by the presence of the $(2s-1)(s-1)rh'R_s$ term. This vanishes for the spin-1/2 and spin-1 species considered in these works, but is non-vanishing for gravitons. While we will ultimately be interested in $s = 2$ gravitons in the context of gravitational waves, we will also need the greybody factors for the other brane-localized species emitted by the black hole in order to faithfully model its evolution and evaporation. Thus we will leave s arbitrary and solve Eq. (3.107) in full generality.

This task will be easiest if we first make the change of variables $P_s = \Delta^s R_s$ to eliminate the $s\Delta''$ term, leading to the transformed radial equation

$$\Delta^s \frac{d}{dr}\left[\Delta^{1-s}\frac{dP_s}{dr}\right] + \left[\frac{\omega^2 r^2}{h} + is\omega r\left(2 - \frac{rh'}{h}\right) - (2s-1)(s-1)rh' - \Lambda_{s\ell}\right] P_s = 0\,, \quad (3.108)$$

where for convenience we have also defined $\Lambda_{s\ell} = \lambda_{s\ell} + 2s = \ell(\ell+1) - s(s-1)$. Despite this simplification, an exact analytic[29] solution is not possible for all r. It is possible, however, to solve in the near-horizon $r \simeq r_h$ and far-field $r \gg r_h$ regimes. This turns out to be sufficient for computing the greybody factors, which are related to the ratio of amplitudes for incoming and outgoing modes evaluated at either the horizon or infinity. The idea is to approximate the full solution of the radial equation by first solving in these limits, then stretching the solutions and matching in an intermediate regime.

More concretely, the asymptotic solution in the near-horizon limit $r \to r_h$ will turn out

29. Numerical integration of the radial master equation is also a possibility, but it is not without its own difficulties. In particular for higher spins, it becomes increasingly difficult to isolate the two asymptotic coefficients which enter into the expression for the greybody factor. Exact numerical solutions for $s = 0$, 1/2, and 1 were derived in Ref. [222], however the numerical issues are such that no analysis has yet been performed for $s = 2$.

to take the form

$$P_s^H = A_{\rm in}^H e^{-i\omega r_*} + A_{\rm out}^H \Delta^s e^{i\omega r_*}\,, \tag{3.109}$$

where r_* is a tortoise coordinate satisfying $\frac{dr^*}{dr} = h^{-1}$. The condition that there be no outgoing modes near the black hole horizon requires us to set $A_{\rm out}^H = 0$. Meanwhile the asymptotic solution in the limit $r \to \infty$ will be of the form

$$P_s^\infty = A_{\rm in}^\infty \frac{e^{-i\omega r}}{(2\omega r)^{1-2s}} + A_{\rm out}^\infty \frac{e^{i\omega r}}{2\omega r}\,. \tag{3.110}$$

In the following three subsections, we solve the radial equation in the near-horizon and far-field regimes and then match the solutions in the intermediate region, which allows us to identify the coefficients $A_{\rm in}^H$, $A_{\rm in}^\infty$, and $A_{\rm out}^\infty$. Then we will use these coefficients to construct the absorption probabilities and ultimately the greybody factors.

Near-Horizon Regime

To solve Eq. (3.108) in the near-horizon regime, we begin by making the change of variables $r \to h$, in terms of which the equation becomes

$$\begin{aligned} &h(1-h)\frac{d^2P_s}{dh^2} + \left[(1-s)(1-h) - \frac{(n+2s)}{n+1}h\right]\frac{dP_s}{dh} + \\ &\left[\frac{\omega^2 r_h^2}{(n+1)^2 h(1-h)} + \frac{2is\omega r_h - \Lambda_{s\ell}}{(n+1)^2(1-h)} - \frac{is\omega r_h}{(n+1)h} - \frac{(2s-1)(s-1)}{n+1}\right] P_s = 0\,. \end{aligned} \tag{3.111}$$

The further change of variables $P_s(h) = h^\alpha (1-h)^\beta F_s(h)$ transforms this into a hypergeometric equation of the generic form

$$h(1-h)\frac{d^2F_s}{dh^2} + [c - (1+a+b)h]\frac{dF_s}{dh} - abF_s = 0\,, \tag{3.112}$$

with hypergeometric indices

$$a = \alpha + \beta + \frac{1}{2(n+1)}\left[s + n - ns + \sqrt{\kappa_{ns}}\right] , \tag{3.113a}$$

$$b = \alpha + \beta + \frac{1}{2(n+1)}\left[s + n - ns - \sqrt{\kappa_{ns}}\right] , \tag{3.113b}$$

$$c = 1 - s + 2\alpha , \tag{3.113c}$$

where $\kappa_{ns} = n^2(s-1)^2 - 2n(5s-2)(s-1) - s(7s-12) - 4$. Substituting this Ansatz into Eq. (3.108) and solving the resultant algebraic equations, one can identify the power coefficients

$$\alpha_+ = s + \frac{i\omega r_h}{n+1} , \quad \alpha_- = -\frac{i\omega r_h}{n+1} , \tag{3.114a}$$

$$\beta_\pm = \frac{1}{2(n+1)}\left[1 - 2s \pm \sqrt{(2\ell+1)^2 - 4\omega^2 r_h^2 - 8is\omega r_h}\right] . \tag{3.114b}$$

The full solution in the near-field regime reads

$$P^{(s)}_{\rm NH} = A_- h^\alpha (1-h)^\beta F_s(a,b,c;h) + A_+ h^{-\alpha}(1-h)^\beta F_s(a-c+1, b-c+1, 2-c; h) , \tag{3.115}$$

with $A_\pm$ as-yet-undetermined constants. In the limit $r \to r_h$, or equivalently $h \to 0$, this takes the asymptotic form

$$\lim_{h\to 0} P^{(s)}_{\rm NH} \simeq A_- h^\alpha + A_+ h^{-\alpha} . \tag{3.116}$$

Evaluating on the positive root α_+, this becomes

$$\lim_{h\to 0} P^{(s)}_{\rm NH}\big|_{\alpha_+} \simeq A_- h^s e^{i\omega r_*} + A_+ h^{-s} e^{-i\omega r_*} , \tag{3.117}$$

where we have employed the tortoise coordinate r_* introduced in Eq. (3.109). This describes outgoing and incoming waves with divergent amplitudes at the horizon and so should be

discarded. Meanwhile evaluating on the negative root α_- gives

$$\lim_{h\to 0} P^{(s)}_{\rm NH}\big|_{\alpha_-} \simeq A_- e^{-i\omega r^*} + A_+ e^{i\omega r^*}\,, \tag{3.118}$$

which is perfectly regular. Thus we choose $\alpha = \alpha_-$. Recall that we must also enforce the boundary condition that there be no outgoing modes at the horizon, instructing us to set $A_+ = 0$. Then comparing with the asymptotic form Eq. (3.109), we see we should identify $A^H_{\rm in} \equiv A_-$. Finally, convergence of the hypergeometric function demands $\mathrm{Re}[c-a-b] > 0$, leading us to choose the negative root $\beta = \beta_-$. In summary, the solution in the near-field regime is

$$P^{(s)}_{\rm NH} = A^H_{\rm in} h^\alpha (1-h)^\beta F_s(a,b,c;h)\,, \tag{3.119}$$

with hypergeometric indices (a,b,c) as defined in Eq. (3.113) and power coefficients $\alpha = \alpha_-$ and $\beta = \beta_-$ as defined in Eq. (3.114). The amplitude for the incoming mode at the horizon $A^H_{\rm in}$ will be identified after matching the solutions in the intermediate regime, which requires first solving in the far-field limit.

Far-Field Regime

Returning to Eq. (3.108) and now taking the far-field limit $r \gg r_h$, or equivalently $h \to 1$, the radial equation becomes

$$\frac{d^2 P_s}{dr^2} + \frac{2(1-s)}{r}\frac{dP_s}{dr} + \left[\omega^2 + \frac{2is\omega}{r} - \frac{\Lambda_{s\ell}}{r^2}\right] P_s = 0\,. \tag{3.120}$$

Taking the Ansatz $P_s = e^{-i\omega r} r^{\ell+s} \tilde{P}_s$ and performing the change of variables $\rho = 2i\omega r$, the equation adopts the confluent hypergeometric form

$$\rho \frac{d^2 \tilde{P}_s}{d\rho^2} + (v-\rho)\frac{d\tilde{P}_s}{d\rho} - u\tilde{P}_s = 0\,, \tag{3.121}$$

with indices

$$u = \ell - s + 1\,, \;\; v = 2(\ell + 1)\,. \tag{3.122}$$

The general solution reads

$$P^{(s)}_{\rm FF} = e^{-i\omega r} r^{\ell+s} \Big(B_+ M(u,v;\rho) + B_- U(u,v;\rho) \Big)\,, \tag{3.123}$$

where M and U are Kummer functions, and we must keep both solutions. In order to fix the asymptotic normalization, consider taking the $r \to \infty$ limit, which leads to

$$\begin{aligned} \lim_{r\to\infty} P^{(s)}_{\rm FF} \simeq & \left[\frac{e^{-i\pi(\ell-s+1)/2}}{(2\omega)^{\ell+s}} \left(B_- + \frac{B_+ e^{i\pi(\ell-s+1)} \Gamma[2\ell+2]}{\Gamma[\ell+s+1]} \right) \right] \frac{e^{-i\omega r}}{(2\omega r)^{1-2s}} \\ & + \left[\frac{e^{-i\pi(\ell+s+1)/2}}{(2\omega)^{\ell+s}} \frac{B_+ \Gamma[2\ell+2]}{\Gamma[\ell-s+1]} \right] \frac{e^{i\omega r}}{2\omega r}\,, \end{aligned} \tag{3.124}$$

where the first and second terms are incoming and outgoing waves at infinity, respectively. Comparison with Eq. (3.110) leads us to identify the bracketed quantity in the first line with $A^\infty_{\rm in}$,

$$A^\infty_{\rm in} = \frac{e^{-i\pi(\ell-s+1)/2}}{(2\omega)^{\ell+s}} \left(B_- + \frac{e^{i\pi(\ell-s+1)} \Gamma[2\ell+2]}{\Gamma[\ell+s+1]} B_+ \right)\,, \tag{3.125}$$

and that in the second line with $A^\infty_{\rm out}$,

$$A^\infty_{\rm out} = \frac{e^{-i\pi(\ell+s+1)/2}}{(2\omega)^{\ell+s}} \frac{\Gamma[2\ell+2]}{\Gamma[\ell-s+1]} B_+\,. \tag{3.126}$$

Note that the incoming mode is dominant and the outgoing mode is suppressed as $1/r$ since this solution corresponds to the upper component[30] of the emitted field, with $+s$.

30. The other radiative component with $-s$ would have a suppressed incoming wave and dominant outgoing wave.

Intermediate Regime

Finally, a complete solution and identification of the integration constants A_- and $B_\pm$ requires us to match $P_{\rm NH}^{(s)}$ and $P_{\rm FF}^{(s)}$ in the intermediate region. Returning to the near horizon solution of Eq. (3.119), we want to stretch this towards large values of r. This will be easier if we first transform the argument of the hypergeometric function from h to $1-h$ via the identity

$$\begin{aligned} F(a,b,c;h) =& \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} F(a,b,1+a+b-c;1-h) \\ &+ \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)} (1-h)^{c-a-b} F(c-a,c-b,1+c-a-b;1-h)\,. \end{aligned} \tag{3.127}$$

Now taking the limit $h \to 1$ ($r \to \infty$), the near-field solution becomes

$$\begin{aligned} \lim_{h\to 1} P_{\rm NH}^{(s)} \simeq & A_{\rm in}^H \left(\frac{\Gamma[1-s+2\alpha]\Gamma[\frac{1-2s}{n+1} - 2\beta]}{\Gamma[c-a]\Gamma[c-b]} \right) \left(\frac{r_h}{r}\right)^{(1-2s-\sqrt{(2\ell+1)^2-4\omega^2 r_h^2 - 8is\omega r_h})/2} \\ & + A_{\rm in}^H \left(\frac{\Gamma[1-s+2\alpha]\Gamma[2\beta - \frac{1-2s}{n+1}]}{\Gamma[a]\Gamma[b]} \right) \left(\frac{r_h}{r}\right)^{(1-2s+\sqrt{(2\ell+1)^2-4\omega^2 r_h^2 - 8is\omega r_h})/2} . \end{aligned} \tag{3.128}$$

As for the far-field solution of Eq. (3.123), stretching towards small values of $r \to r_h$ yields

$$\lim_{r\to r_h} P_{\rm FF}^{(s)} \simeq B_+ r^{\ell+s} + \frac{B_-}{r^{\ell-s+1}} \frac{\Gamma[2\ell+1]}{(2i\omega)^{2\ell+1}\Gamma[\ell-s+1]}\,. \tag{3.129}$$

We would like to match this onto Eq. (3.128), however these two expressions have different power law scalings in r. In order to eliminate the square root in the near-horizon solution,

we must take the low energy limit $\omega r_h \ll 1$,

$$\begin{aligned}\lim_{h\to 1} P_{\rm NH}^{(s)} \simeq A_{\rm in}^H \left(\frac{\Gamma[1-s+2\alpha]\,\Gamma[\frac{1-2s}{n+1}-2\beta]}{\Gamma[c-a]\,\Gamma[c-b]}\right)\left(\frac{r}{r_h}\right)^{\ell+s} \\ + A_{\rm in}^H \left(\frac{\Gamma[1-s+2\alpha]\,\Gamma[2\beta-\frac{1-2s}{n+1}]}{\Gamma[a]\,\Gamma[b]}\right)\left(\frac{r_h}{r}\right)^{\ell-s+1}\,.\end{aligned} \tag{3.130}$$

Note that this approximation is performed for the power law scalings only; no approximation is made to the gamma functions, which increases the domain of validity. Note also that we will ultimately be interested in the part of the spectrum ranging from the low frequency tail through the peak, which is set by the Hawking temperature $\omega_{\rm peak} \sim T_H$. From the fact that $T_H \sim 1/r_h$, we have that $\omega_{\rm peak} r_h \sim 1$, and so this approximation should roughly hold for our regime of interest.

Finally we can identify

$$B_+ = \left(\frac{\Gamma[1-s+2\alpha]\,\Gamma[\frac{1-2s}{n+1}-2\beta]}{\Gamma[c-a]\,\Gamma[c-b]}\right)\frac{A_{\rm in}^H}{r_h^{\ell+s}}\,, \tag{3.131a}$$

$$B_- = \left(\frac{\Gamma[1-s+2\alpha]\,\Gamma[2\beta-\frac{1-2s}{n+1}]\,\Gamma[\ell-s+1]}{\Gamma[a]\,\Gamma[b]\,\Gamma[2\ell+1]}\right)(2i\omega r_h)^{2\ell+1}\frac{A_{\rm in}^H}{r_h^{\ell+s}}\,. \tag{3.131b}$$

Then the relation between incoming amplitudes at the horizon and infinity reads explicitly

$$\begin{aligned}A_{\rm in}^\infty = \frac{A_{\rm in}^H}{(2\omega r_h)^{\ell+s}}\Bigg[\frac{\Gamma(c)\,\Gamma\big(2\beta-\frac{1-2s}{n+1}\big)}{\Gamma(a)\,\Gamma(b)}\,\frac{(\ell-s)!}{(2\ell)!}\,(2i\omega r_h)^{2\ell+1}\,e^{-i\pi(\ell-s+1)/2} \\ + \frac{\Gamma(c)\,\Gamma\big(\frac{1-2s}{n+1}-2\beta\big)}{\Gamma(c-a)\,\Gamma(c-b)}\,\frac{(2\ell+1)!}{(\ell+s)!}\,e^{i\pi(\ell-s+1)/2}\Bigg]\,,\end{aligned} \tag{3.132}$$

with the hypergeometric indices a, b, c defined in Eq. (3.113) and coefficients α, β in Eq. (3.114).

Absorption Coefficients

Having identified $A^H_{\rm in}$, $A^\infty_{\rm in}$, and $A^\infty_{\rm out}$, we are now ready to construct absorption coefficients $|\mathcal{A}^{(s)}_\ell|^2$, which were defined in Eq. (3.101) in terms of ratios of energy fluxes of incoming and outgoing modes at the horizon and infinity. In particular, we will use the second definition involving the incoming fluxes $\mathcal{F}^\infty_{\rm in}$ and $\mathcal{F}^H_{\rm in}$. This is more convenient since, for the upper field component we will consider, the outgoing mode at infinity is suppressed.

For scalars and fermions, the incoming energy flux at the horizon or infinity can be computed by integrating the radial component of the conserved current J^μ over a 2-sphere at this location [239]. For a complex scalar Ψ, this conserved current is

$$J^\mu_{(s=0)} = hr^2(\Psi\partial^\mu\Psi^* - \Psi^*\partial^\mu\Psi)\,, \tag{3.133}$$

leading to a flux proportional to the absolute square of the radial solution

$$\mathcal{F}_{(s=0)} \sim \left|P_0\right|^2. \tag{3.134}$$

Recall from Eq. (3.109) that the radial solution for an incoming mode at the horizon takes the generic form[31]

$$P^H_s = A^H_{\rm in} h^{-i\omega r_h/(n+1)}\,, \tag{3.135}$$

while from Eq. (3.110), the radial solution for an incoming mode at infinity is

$$P^\infty_s = A^\infty_{\rm in}\frac{e^{-i\omega r}}{(2\omega r)^{1-2s}}\,. \tag{3.136}$$

31. This normalization is equivalent to that given in Eq. (3.109) upon solving for an explicit r_* satisfying $dr_*/dr = h^{-1}$.

The absorption coefficient for a scalar is then

$$\left|\mathcal{A}_\ell^{(0)}\right|^2 = (2\omega r_h)^2 \left|\frac{A_{\rm in}^H}{A_{\rm in}^\infty}\right|^2 . \tag{3.137}$$

For a fermion Ψ^A with conserved current

$$J^\mu_{(s=1/2)} = \sqrt{2}\sigma^\mu_{AB}\Psi^A\bar{\Psi}^B , \tag{3.138}$$

the flux is proportional to the difference of upper and lower field components

$$\mathcal{F}_{(s=1/2)} \sim \left|P_{1/2}\right|^2 - \left|P_{-1/2}\right|^2 . \tag{3.139}$$

Looking at Eq. (3.136), we see that for the lower field component, the incoming radial solution at infinity is suppressed. Thus the dominant contribution to the flux comes from the $s = +1/2$ helicity and the absorption coefficient is simply

$$\left|\mathcal{A}_\ell^{(1/2)}\right|^2 = \left|\frac{A_{\rm in}^H}{A_{\rm in}^\infty}\right|^2 . \tag{3.140}$$

For $s > 1/2$, there are no conserved currents; however one can still infer the flow of energy from the energy momentum tensor. For a gauge boson Ψ^{AB}, one can use the T^{rt} component of the energy momentum tensor

$$T^{\mu\nu}_{(1)} = 2\sigma^\mu_{AA'}\sigma^\nu_{BB'}\Psi^{AB}\bar{\Psi}^{A'B'} , \tag{3.141}$$

to derive the flux

$$\mathcal{F}_{s=1} \sim \frac{1}{2\omega r^2}\left(\left|P_1\right|^2 - \left|P_{-1}\right|^2\right) . \tag{3.142}$$

Only the radial wavefunctions for $s = \pm 1$ appear because only the upper and lower field

components are radiative. As with the $s = 1/2$ case, the incoming mode is suppressed for the lower field component. Thus concentrating on the contribution from $s = +1$, the absorption coefficient is

$$\left|\mathcal{A}_\ell^{(1)}\right|^2 = \frac{1}{(2\omega r_h)^2}\left|\frac{A_{\rm in}^H}{A_{\rm in}^\infty}\right|^2 . \tag{3.143}$$

For scalars, fermions, and gauge bosons, the absorption coefficient can be conveniently summarized as

$$\left|\mathcal{A}_\ell^{(s)}\right|^2 = (2\omega r_h)^{2(1-2s)}\left|\frac{A_{\rm in}^H}{A_{\rm in}^\infty}\right|^2 , \tag{3.144}$$

with the ratio $A_{\rm in}^H/A_{\rm in}^\infty$ given explicitly in Eq. (3.132). Repeating this procedure for gravitons is complicated by the fact that there is generically no conserved energy-momentum tensor for spin-2. As advocated for in [239], one option is to consider the Bel-Robinson tensor, from which one can infer an energy-momentum conserved in all static black hole spacetimes. In particular, the component T^{rttt} can be used to construct the energy flux

$$\frac{1}{4\pi}\frac{dE}{dt} = \frac{1}{16\omega^2 r^6}\left(\left|P_2\right|^2 - \left|P_{-2}\right|^2\right) . \tag{3.145}$$

Close to the horizon, we should replace the square of the frequency with $\omega_h^2 = \omega^2 + (2\pi T_{\rm H})^2$, which comes from integrating with respect to proper distance along the worldline and is necessary to properly account for gravitons close to the horizon [239]. Again the negative helicity contribution is suppressed, so from the radial solution for $s = +2$, we have the absorption coefficient

$$\left|\mathcal{A}_\ell^{(2)}\right|^2 = \frac{1}{64\omega^4\omega_h^2 r_h^6}\left|\frac{A_{\rm in}^H}{A_{\rm in}^\infty}\right|^2 , \tag{3.146}$$

with $A_{\rm in}^H/A_{\rm in}^\infty$ in Eq. (3.132). Substituting Eqs. (3.144) and (3.146) into Eq. (3.100) yields the greybody factors for all brane localized modes, and is the main result of this section.

3.3.5 Greybody Factors II: Bulk Modes

Bulk Scalars

The calculation of greybody factors for bulk modes proceeds much in the same way as the calculation for the brane-localized modes, though the starting point is the complete $(n+4)$-dimensional metric of Eq. (3.82). Since the scalar case is particularly straightforward, we will consider it first, following the treatment of [235]. Let $\hat{\Psi}$ be a bulk scalar field, which can be factorized as $\hat{\Psi}(t,r,\Omega) = e^{-i\omega t}\hat{R}^{(0)}_{\omega\ell}(r)\hat{Y}_\ell(\Omega)$, where $\Omega = (\theta_i, \varphi)$ includes the angular coordinates and $\hat{Y}_\ell(\Omega)$ are $(n+3)$-dimensional spherical harmonics. The radial wavefunction satisfies the second-order differential equation

$$\frac{h}{r^{n+2}}\frac{d}{dr}\left(hr^{n+2}\frac{d\hat{R}^{(0)}}{dr}\right) + \left(\omega^2 - \frac{h}{r^2}\ell(\ell+n+1)\right)\hat{R}^{(0)} = 0\,, \tag{3.147}$$

with $h(r)$ defined in Eq. (3.83).

To solve in the near-horizon regime, we make the change of variables $r \to h$, such that the radial equation becomes

$$h(1-h)\frac{d^2\hat{R}^{(0)}}{dh^2} + (1-h)\frac{d\hat{R}^{(0)}}{dh} + \left(\frac{\omega^2 r^2}{(n+1)^2 h(1-h)} - \frac{\ell(\ell+n+1)}{(n+1)^2(1-h)}\right)\hat{R}^{(0)} = 0\,. \tag{3.148}$$

Near the horizon we may set $\omega^2 r^2 \to \omega^2 r_h^2$; then by taking the Ansatz $\hat{R}^{(0)} = h^{\hat{\alpha}}(1-h)^{\hat{\beta}}F(h)$, this takes the form of a hypergeometric equation

$$h(1-h)\frac{d^2F}{dh^2} + [\hat{c} - (1+\hat{a}+\hat{b})h]\frac{dF}{dh} - \hat{a}\hat{b}F = 0\,, \tag{3.149}$$

where $F(\hat{a},\hat{b},\hat{c};h)$ is a hypergeometric function with indices $\hat{a} = \hat{b} = \hat{\alpha}+\hat{\beta}$ and $\hat{c} = 1+2\hat{\alpha}$, and

$$\hat{\alpha}_\pm = \pm\frac{i\omega r_h}{n+1}\,, \tag{3.150a}$$

$$\hat{\beta}_{\pm} = \frac{1}{2} \pm \frac{1}{2(n+1)} \sqrt{(2\ell + n + 1)^2 - (2\omega r_h)^2} \,. \tag{3.150b}$$

Demanding convergence instructs us to choose the solution $\hat{\beta} = \hat{\beta}_-$ and as before we set $\hat{\alpha} = \hat{\alpha}_-$. The complete solution in the near-horizon regime is

$$\hat{R}^{(0)}_{\rm NH} = \hat{A}_- h^{\hat{\alpha}} (1-h)^{\hat{\beta}} F(\hat{a}, \hat{b}, \hat{c}; h) + \hat{A}_+ h^{-\hat{\alpha}} (1-h)^{\hat{\beta}} F(\hat{a} - \hat{c} + 1, \hat{b} - \hat{c} + 1, 2 - \hat{c}; h) \,, \tag{3.151}$$

and in the near-horizon limit $h \to 0$ this becomes

$$\lim_{h \to 0} \hat{R}^{(0)}_{\rm NH} \simeq \left(\frac{r_h}{r} \right)^{(n+1)\hat{\beta}} \left(\hat{A}_- e^{-i\omega r_h^{n+2} y} + \hat{A}_+ e^{i\omega r_h^{n+2} y} \right) , \tag{3.152}$$

with y a tortoise coordinate satisfying $\frac{dy}{dr} = \frac{1}{h r^{n+2}}$. The boundary condition that there be no outgoing mode instructs us to set $\hat{A}_+ = 0$.

Returning to Eq. (3.147), to solve in the far-field regime $h \simeq 1$ we make the change of variables $\hat{R}^{(0)} = \hat{f}(r)/r^{(n+1)/2}$, in terms of which

$$\left(\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} + \left(\omega^2 - \frac{\ell(\ell + n + 1)}{r^2} - \frac{(n+1)^2}{4r^2} \right) \right) \hat{f} = 0 \,. \tag{3.153}$$

This is a Bessel equation whose solution includes Bessel functions of the first $J_\nu(x)$ and second $Y_\nu(x)$ kind,

$$\hat{R}^{(0)}_{\rm FF} = \frac{\hat{B}_+}{r^{(n+1)/2}} J_{\ell + (n+1)/2}(\omega r) + \frac{\hat{B}_-}{r^{(n+1)/2}} Y_{\ell + (n+1)/2}(\omega r) \,. \tag{3.154}$$

In the asymptotic regime $\omega r \to \infty$, this takes the form

$$\lim_{r \to \infty} \hat{R}^{(0)}_{\rm FF} \simeq \frac{\hat{B}_+ - i\hat{B}_-}{\sqrt{2\pi\omega r^{n+2}}} e^{i(\omega r - (2\ell + n + 2)\pi/4)} + \frac{\hat{B}_+ + i\hat{B}_-}{\sqrt{2\pi\omega r^{n+2}}} e^{-i(\omega r - (2\ell + n + 2)\pi/4)} \,. \tag{3.155}$$

As before, the near-horizon solution Eq. (3.151) needs to be stretched to $r \gg r_h$ and the

far-field solution Eq. (3.154) needs to be stretched to $r \ll \infty$ such that the solutions can be matched in the intermediate regime. Stretching the near-horizon solution can be achieved by first transforming the argument from h to $1-h$ using the identity in Eq. (3.127) and then taking $h \to 1$, leading to

$$\lim_{h\to 1} \hat{R}^{(0)}_{\rm NH} \simeq \hat{A}_- \left(\frac{\Gamma(1+2\hat{\alpha})\,\Gamma(1-2\hat{\beta})}{\Gamma(1+\hat{\alpha}-\hat{\beta})^2}\right)\left(\frac{r}{r_h}\right)^{\ell} + \hat{A}_- \left(\frac{\Gamma(1+2\hat{\alpha})\,\Gamma(2\hat{\beta}-1)}{\Gamma(\hat{\alpha}+\hat{\beta})^2}\right)\left(\frac{r_h}{r}\right)^{\ell+n+1} . \tag{3.156}$$

Stretching the far-field solution to smaller r and taking the low energy limit $\omega r_h \ll 1$ again to match the powers of r yields

$$\lim_{r\to r_h} \hat{R}^{(0)}_{\rm FF} \simeq \hat{B}_+ \left(\frac{(\omega/2)^{\ell+(n+1)/2}}{\Gamma\big(\ell+\frac{n+3}{2}\big)}\right) r^{\ell} - \hat{B}_- \left(\frac{\Gamma\big(\ell+\frac{n+1}{2}\big)}{\pi(\omega/2)^{\ell+(n+1)/2}}\right) r^{-(\ell+n+1)} . \tag{3.157}$$

Matching the solutions then allows us to identify $\hat{B}_\pm$ in terms of $\hat{A}_-$. For convenience we define the ratio $\hat{B} \equiv \hat{B}_+/\hat{B}_-$, which is explicitly

$$\hat{B} = -\frac{1}{\pi}\left(\frac{2}{\omega r_h}\right)^{2\ell+n+1} \Gamma\big(\ell+\frac{n+1}{2}\big)\Gamma\big(\ell+\frac{n+3}{2}\big)\frac{\Gamma(\hat{\alpha}+\hat{\beta})^2}{\Gamma(1+\hat{\alpha}-\hat{\beta})^2}\frac{\Gamma(1-2\hat{\beta})}{\Gamma(2\hat{\beta}-1)} . \tag{3.158}$$

We use the first expression[32] of Eq. (3.101) for the absorption coefficient, which depends on the ratio of incoming and outgoing fluxes at infinity

$$\big|\hat{\mathcal{A}}^{(0)}_\ell\big|^2 = 1 - \frac{|\hat{R}^{(0)}_{\rm FF,out}|^2}{|\hat{R}^{(0)}_{\rm FF,in}|^2}\bigg|_\infty . \tag{3.159}$$

In terms of the asymptotic far-field solutions of Eq. (3.155), this is

$$\big|\hat{\mathcal{A}}^{(0)}_\ell\big|^2 = 1 - \left|\frac{\hat{B}_+ - i\hat{B}_-}{\hat{B}_+ + i\hat{B}_-}\right|^2 = \frac{4\,\mathrm{Im}[\hat{B}]}{|\hat{B}|^2 + 2\,\mathrm{Im}[\hat{B}] + 1} . \tag{3.160}$$

32. In the scalar case, both incoming and outgoing modes at infinity scale as $r^{-(n+2)/2}$, as can be seen from Eq.(3.155), and so neither is suppressed.

Bulk Gravitons

The bulk graviton perturbations should be decomposed into three parts: a symmetric traceless tensor piece, a vector piece, and a scalar piece. The radial equation obeyed by each type ($\hat{R}^T, \hat{R}^V, \hat{R}^S$) is slightly different but can be solved in a manner exactly analogous to that used for the other perturbations we have considered so far. After making the change of variables to h in order to solve in the near-horizon regime, the master equation [240] is

$$h(1-h)\frac{d^2\hat{R}}{dh^2} + \left(1 - \frac{(2n+3)}{(n+1)}h\right)\frac{d\hat{R}}{dh} + \left(\frac{\omega^2 r_h^2}{(n+1)^2 h(1-h)} - \frac{\Lambda_G}{1-h} + C_G\right)\hat{R} = 0\,, \quad (3.161)$$

where

$$\Lambda_G = \begin{cases} \frac{\ell(\ell+n+1)}{(n+1)^2} + \frac{n(n+2)}{4(n+1)^2} & \text{for } G = T, V \\ \frac{z}{16(n+1)^2 m^2} & \text{for } G = S \end{cases}, \quad (3.162)$$

and

$$C_G = \begin{cases} -\frac{(n+2)^2}{4(n+1)^2} & \text{for } G = T \\ \frac{3(n+2)^2}{4(n+1)^2} & \text{for } G = V \\ -\frac{q(1-h)^2 + p(1-h) + w}{4(n+1)^2[2m+(n+2)(n+3)(1-h)]^2} & \text{for } G = S \end{cases}, \quad (3.163)$$

where for the scalar perturbations we have defined $m = \ell(\ell+n+1) - n - 2$, $q = (n+2)^4(n+3)^2$, $p = (n+2)(n+3)[4m(2n^2+5n+6) + n(n+2)(n+3)(n-2)] - \frac{z(n+2)^2(n+3)^2}{4m^2}$, $w = -12m(n+2)[m(n-2) + n(n+2)(n+3)] - \frac{z(n+2)(n+3)}{m}$, and $z = 16m^3 + 4m^2(n+2)(n+4)$. For all perturbations, the solution obeying the boundary condition of no outgoing waves at the horizon is

$$\hat{R}^{T,V,S}_{\rm NH} = \hat{A}_- h^{\hat{\alpha}}(1-h)^{\hat{\beta}} F(\hat{a}, \hat{b}, \hat{c}; h)\,, \quad (3.164)$$

where

$$\hat{\alpha} = -\frac{i\omega r_h}{n+1}\,, \quad \hat{\beta} = \frac{1}{2(n+1)}\left(-1 - \sqrt{(2\ell+n+1)^2 - 4\omega^2 r_h^2}\right)\,, \quad (3.165)$$

and the hypergeometric indices are

$$\hat{a} = \hat{\alpha} + \hat{\beta} + \frac{n+2}{2(n+1)} + \lambda_G\,, \;\; \hat{b} = \hat{\alpha} + \hat{\beta} + \frac{n+2}{2(n+1)} - \lambda_G\,, \;\; \hat{c} = 1 + 2\hat{\alpha}\,, \tag{3.166}$$

with

$$\lambda_G = \begin{cases} 0 & \text{for } G = T \\ \frac{(n+2)}{(n+1)} & \text{for } G = V \\ \frac{(n+2)}{2(n+1)}\sqrt{1 - \frac{q+p+w}{[2m(n+2)+(n+2)^2(n+3)]^2}} & \text{for } G = S \end{cases} . \tag{3.167}$$

The far-field radial equation is much simpler, and for all perturbations takes the form

$$\frac{d^2\hat{R}^{T,V,S}}{dr^2} + \left(\omega^2 - \left(\ell(\ell+n+1) + \frac{n(n+2)}{4}\right)\frac{1}{r^2}\right)\hat{R}^{T,V,S} = 0\,, \tag{3.168}$$

with solution

$$\hat{R}^{T,V,S}_{\rm FF} = \hat{B}_+\sqrt{r}J_{\ell+(n+1)/2}(\omega r) + \hat{B}_-\sqrt{r}Y_{\ell+(n+1)/2}(\omega r)\,. \tag{3.169}$$

Stretching the solutions to the intermediate regime and matching allows for the identifications

$$\frac{\hat{B}_+}{\hat{A}_-} = \left(\frac{2}{\omega r_h}\right)^{\ell+(n+1)/2}\frac{\Gamma\left(\ell+\frac{n+3}{2}\right)\Gamma(\hat{c})\Gamma(\hat{c}-\hat{a}-\hat{b})}{\Gamma(\hat{c}-\hat{a})\Gamma(\hat{c}-\hat{b})r_h^{1/2}}\,, \tag{3.170}$$

and

$$\frac{\hat{B}_-}{\hat{A}_-} = -\pi\left(\frac{\omega r_h}{2}\right)^{\ell+(n+1)/2}\frac{\Gamma(\hat{c})\Gamma(\hat{a}+\hat{b}-\hat{c})}{\Gamma\left(\ell+\frac{n+1}{2}\right)\Gamma(\hat{a})\Gamma(\hat{b})r_h^{1/2}}\,. \tag{3.171}$$

Again, we define the ratio $\hat{B} = \hat{B}_+/\hat{B}_-$, so

$$\hat{B} = -\frac{1}{\pi}\left(\frac{2}{\omega r_h}\right)^{2\ell+n+1}\frac{\Gamma\left(\ell+\frac{n+3}{2}\right)\Gamma\left(\ell+\frac{n+1}{2}\right)\Gamma(\hat{a})\Gamma(\hat{b})\Gamma(\hat{c}-\hat{a}-\hat{b})}{\Gamma(\hat{c}-\hat{a})\Gamma(\hat{c}-\hat{b})\Gamma(\hat{a}+\hat{b}-\hat{c})}\,. \tag{3.172}$$

Particle species	**Absorption coefficient** $\left\|\mathcal{A}_\ell\right\|^2$	**Relevant equations**
Scalar, $s = 0$	$(2\omega r_h)^2 \left\|A^H_{\rm in}/A^\infty_{\rm in}\right\|^2$	with $\left\|A^H_{\rm in}/A^\infty_{\rm in}\right\|^2$ in Eq. (3.132)
Fermion, $s = 1/2$	$\left\|A^H_{\rm in}/A^\infty_{\rm in}\right\|^2$	
Gauge boson, $s = 1$	$\frac{1}{(2\omega r_h)^2}\left\|A^H_{\rm in}/A^\infty_{\rm in}\right\|^2$	
Zero-mode graviton, $s = 2$	$\frac{1}{64\omega^4\omega_h^2 r_h^6}\left\|A^H_{\rm in}/A^\infty_{\rm in}\right\|^2$	
Bulk scalar	$\frac{4\,\mathrm{Im}[\hat{B}]}{\lvert\hat{B}\rvert^2+2\,\mathrm{Im}[\hat{B}]+1}$	with $\hat{B}$ in Eq. (3.158)
Graviton, scalar perturbation		with $\hat{B}$ in Eq. (3.172)
Graviton, vector perturbation		
Graviton, tensor perturbation		

Table 3.2: Summary of absorption coefficients for degrees of freedom on the brane and in the bulk.

For the absorption coefficient, we expand the far-field solution in the asymptotic limit $r \to \infty$ and use the definition in terms of incoming and outgoing fluxes at infinity. The solution is the same as in the bulk scalar case,

$$\left|\hat{\mathcal{A}}^{T,V,S}_\ell\right|^2 = 1 - \left|\frac{\hat{B}_+ - i\hat{B}_-}{\hat{B}_+ + i\hat{B}_-}\right|^2 = \frac{4\,\mathrm{Im}[\hat{B}]}{|\hat{B}|^2 + 2\,\mathrm{Im}[\hat{B}] + 1}\,, \tag{3.173}$$

but now with $\hat{B}$ given in Eq. (3.172).

3.3.6 Emission Rates and Gravitational Wave Spectra

Ultimately we would like to study the emission of brane-localized (zero-mode) gravitons during the Hawking evaporation of primordial black holes in large extra dimensions. We would like to compute the gravitational wave spectra from this source, but since the energy

available to brane-localized gravitons depends on that expended in emitting all other species, it is necessary to consider the emission rates for these as well.

Emission Rates

The emission of particles with mass greater than the Hawking temperature is exponentially suppressed, so for convenience, we consider only the emission of species[33] with $m < T_{\rm BH}$ and treat these as massless, $|\vec{k}| = \omega$. This way, the phase space integrals in the flux and power spectra of Eqs. (3.98)-(3.99) and Eqs. (3.102)-(3.103) can be evaluated explicitly. Using also Eqs. (3.100) and (3.104) to re-express the greybody factors in terms of absorption coefficients, the flux and power spectra simplify considerably to

$$\frac{dN^{(s)}}{dtd\omega} = \frac{1}{2\pi}\sum_{\ell}(2\ell+1)\left|\mathcal{A}_{\ell}^{(s)}\right|^2\frac{1}{e^{\omega/T_{\rm BH}}\mp 1}\,, \tag{3.174a}$$

$$\frac{dE^{(s)}}{dtd\omega} = \frac{1}{2\pi}\sum_{\ell}(2\ell+1)\left|\mathcal{A}_{\ell}^{(s)}\right|^2\frac{\omega}{e^{\omega/T_{\rm BH}}\mp 1}\,, \tag{3.174b}$$

on the brane and

$$\frac{d\hat{N}^{(s)}}{dtd\omega} = \frac{1}{2\pi}\sum_{\ell}\frac{(2\ell+n+1)(\ell+n)!}{\ell!(n+1)!}\left|\hat{\mathcal{A}}_{\ell}^{(s)}\right|^2\frac{1}{e^{\omega/T_{\rm BH}}\mp 1}\,, \tag{3.175a}$$

$$\frac{d\hat{E}^{(s)}}{dtd\omega} = \frac{1}{2\pi}\sum_{\ell}\frac{(2\ell+n+1)(\ell+n)!}{\ell!(n+1)!}\left|\hat{\mathcal{A}}_{\ell}^{(s)}\right|^2\frac{\omega}{e^{\omega/T_{\rm BH}}\mp 1}\,, \tag{3.175b}$$

in the bulk. Fig. 3.14 below compares emission rates for the various species for two sample benchmark points. We see that at higher frequencies, the differences in emission rate for all brane localized species become smaller, as one would expect since in the high frequency

33. The mass of the lightest KK scalar is set by the size of the extra dimensions as $m_{KK} \simeq 1/R$, which in turn is set by our choice of M_* and n since $M_{\rm Pl}^2 \sim R^n M_*^{n+2}$. In particular for $M_* = 10^3$ TeV and $n = 4$, we expect the lightest mode to have GeV-scale mass.

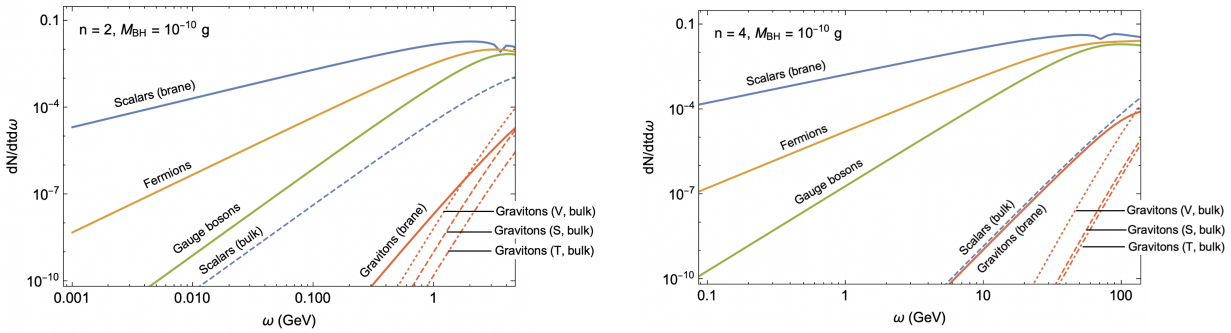

Figure 3.14: Instantaneous flux $\frac{dN}{dtd\omega}$ of species on the brane and in the bulk for a representative black hole of mass $M = 10^{-10}$ g and $n = 2$ and $n = 4$ extra dimensions, respectively. The same qualitative trends are observed for the power spectra $\frac{dE}{dtd\omega}$. The right-hand side of the x-axis is truncated where the low frequency approximation breaks down, $\omega r_h \sim 1$.

limit, the greybody factors for all species approach the geometric optics limit. Note that we truncate the x-axis at the frequency at which the low frequency approximation used in deriving the greybody factors breaks down. At low frequencies, the emission of particles with higher spin is suppressed due to the larger barrier such particles have to surmount.

In general, the emission rate of species on the brane exceeds that of bulk species, and this effect becomes more pronounced with increasing number of extra dimensions n. This is perhaps surprising, since both black hole temperature and the multiplicity of states are enhanced at higher n. However the absorption probability is suppressed with increasing n, and ultimately it is this effect that dominates. Thus for all bulk degrees of freedom, we observe a low energy emission rate which decreases with increasing n, consistent with the findings of [240]. We also confirm these authors' claim that among the gravitational perturbations in the bulk, vector-type dominate.

Gravitational Wave Signal

Now we turn to calculate the gravitational wave signal from this source, as parameterized by the spectral density parameter $\Omega_{\rm GW}$, defined as

$$\Omega_{\rm GW} = \frac{1}{\rho_{\rm crit}} \frac{d\rho_{\rm GW}}{d\ln f} \,. \tag{3.176}$$

Our starting point is the instantaneous power spectrum for a single degree of freedom for a brane-localized graviton, $\frac{dE^{(2)}}{dtd\omega}$, the general expression for which is given in Eq. (3.174b) with $|\mathcal{A}_\ell^{(2)}|^2 = \frac{1}{64\omega^4\omega_h^2 r_h^6}|A_{\rm in}^H/A_{\rm in}^\infty|^2$ and $|A_{\rm in}^H/A_{\rm in}^\infty|^2$ in Eq. (3.132). Multiplying by 2 to account for the 2 graviton polarizations we define $\frac{dE_{\rm GW}}{dtd\omega} = 2\frac{dE^{(2)}}{dtd\omega}$. For an entire population of black holes with number density $n_{\rm BH}$, the instantaneous energy density emitted in gravitational waves is then

$$\frac{d\rho_{\rm GW}}{dtd\omega} = n_{\rm BH}\frac{dE_{\rm GW}}{dtd\omega} \,. \tag{3.177}$$

To obtain the total energy density in zero-mode gravitons emitted over the black hole lifetime, we integrate this expression from black hole formation at t_i to evaporation at $t_* = t_i + \tau_{\rm BH}$, where the black hole lifetime $\tau_{\rm BH}$ can be obtained by numerically following the black hole evolution

$$\frac{dM}{dt} = -\sum_{\text{all d.o.f.}} \int d\omega \left(\frac{dE^{\,(s)}}{dtd\omega}\right) \,, \tag{3.178}$$

from $M = M_i$ to $\sim M_*$. For practical purposes when integrating $\frac{d\rho_{\rm GW}}{dtd\omega}$, it is useful to make the time dependence of certain quantities explicit. In particular recall that $\rho_{\rm GW} \sim a^{-4}$, $\omega \sim a^{-1}$, and $n_{\rm BH} \sim a^{-3}$. The integrated energy density at evaporation is then

$$\frac{d\rho_{\rm GW}^*}{d\ln\omega} = n_{\rm BH}^i \omega_* a_i^3 \int_{t_i}^{t_*} dt\, a(t)^{-3} \left(\frac{dE_{\rm GW}}{dtd\omega}\right) \,, \tag{3.179}$$

Note that we have also converted the frequency interval to a logarithmic frequency interval. We presume that black hole formation occurs during radiation domination and define $\Omega^i_{\rm BH}$ as the initial fractional energy density in black holes, in terms of which the initial number density can be expressed

$$n^i_{\rm BH} = \frac{3M^2_{\rm Pl}\Omega^i_{\rm BH}}{32\pi M_i t_i^2}\,. \tag{3.180}$$

We also allow for the possibility that the black holes come to dominate at some time $t_{\rm eq}$, which is approximately

$$t_{\rm eq} \simeq \left(\frac{1-\Omega^i_{\rm BH}}{\Omega^i_{\rm BH}}\right)^2 t_i\,. \tag{3.181}$$

The scale factor appearing in this expression then scales with time as

$$a(t) = \begin{cases} a_i \left(\frac{t}{t_i}\right)^{1/2} & t < t_{\rm eq} \\ a_i \left(\frac{t_{\rm eq}}{t_i}\right)^{1/2} \left(\frac{t}{t_{\rm eq}}\right)^{2/3} & t > t_{\rm eq}\,. \end{cases} \tag{3.182}$$

Finally to obtain the gravitational wave spectrum today, we need to account for the redshift in energy density and frequency due to the cosmological expansion between evaporation and today

$$\frac{d\rho^0_{\rm GW}}{d\ln f} = \frac{d\rho^*_{\rm GW}}{d\ln\omega}\left(\frac{a_*}{a_0}\right)^4\,. \tag{3.183}$$

We take the scale factor today to be $a_0 = 1$, which can be related to that at evaporation a_* by invoking the conservation of entropy $g_{\star,s}a^3T^3 = \text{constant}$,

$$a_* = \left(\frac{g_{\star,s}(T_0)}{g_{\star,s}(T_*)}\right)^{1/3}\frac{T_0}{T_*}\,, \tag{3.184}$$

where $T_0 = 0.235\,\text{meV}$, $g_{\star,s}(T_0) = 3.91$, and T_* is the temperature when evaporation concludes. In the event that black holes dominate, this can be estimated by equating the energy density in black holes right before decay with that in radiation immediately afterwards,

leading to

$$T_* \simeq \left(\frac{5 M_{\rm Pl}^2}{\pi^3 g_\star(T_*) t_*^2} \right)^{1/4} . \tag{3.185}$$

The energy density in gravitational waves today per logarithmic frequency interval is then

$$\frac{d\rho_{\rm GW}^0}{d\ln f} = \frac{g_{\star,s}(T_0)}{g_{\star,s}(T_*)} \left(\frac{T_0}{T_*} \right)^3 \omega_0 n_{\rm BH}^i a_i^3 \int_{t_i}^{t_*} dt\, a(t)^{-3} \left(\frac{dE_{\rm GW}}{dt d\omega} \right) , \tag{3.186}$$

and the spectral density parameter is

$$\Omega_{\rm GW} = \frac{\Omega_{\rm BH}^i \omega_0^4}{H_0^2 M^i} \mathcal{I}(\omega_0) , \tag{3.187}$$

where $H_0 = 100h \text{km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1}$ is the Hubble rate and we have pulled out the leading frequency scaling, defining the integral

$$\mathcal{I}(\omega_0) = \frac{1}{4t_i^2} \frac{g_{\star,s}(T_0)}{g_{\star,s}(T_*)} \left(\frac{T_0}{T_*} \right)^3 \frac{1}{\omega_0^3} \int_{t_i}^{t_*} dt \left(\frac{a_i}{a(t)} \right)^3 \left(\frac{dE_{\rm GW}}{dt d\omega} \right) . \tag{3.188}$$

Finally for the sake of comparing against projected sensitivities, we will sometimes present our results in terms of the dimensionless characteristic stain h_c, which is related to the spectral density parameter as

$$\Omega_{\rm GW} = \frac{4\pi^2}{3H_0^2} f^2 h_c^2 . \tag{3.189}$$

In Fig. 3.15, we plot predictions for the gravitational wave spectrum for a benchmark point with bulk Planck scale $M_* = 10^3\,\text{TeV}$, formation time $t_i = 10^{-30}\,\text{s}$, initial black hole mass $M = 1\,\text{g}$, reheating temperature $T_{\rm re} = 10^5\,\text{GeV}$, and various numbers of extra spatial dimensions n. We indicate by the dashed horizontal line the integral bound on the total energy density in gravitational waves. The observable $\Delta N_{\rm eff}$ parameterizes the additional amount of radiation energy density beyond that of photons. In particular the contribution

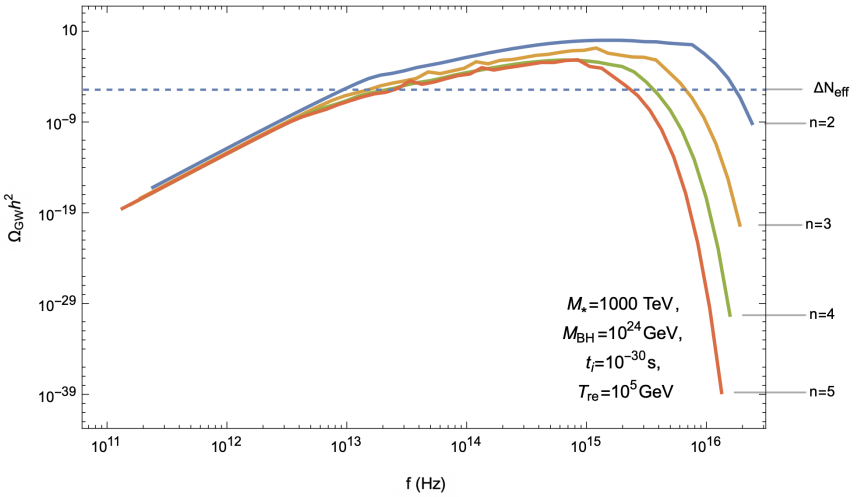

Figure 3.15: Gravitational wave spectra (in terms of the spectral density parameter $\Omega_{\rm GW}h^2$) for various numbers of extra dimensions $n = 2, 3, 4, 5$ and a benchmark set of parameters: $M_* = 10^3\,\text{TeV}$, $t_i = 10^{-30}\,\text{s}$, $M = 1\,\text{g}$, $T_{\rm re} = 10^5\,\text{GeV}$.

from gravitational waves reads

$$\Delta N_{\rm eff} = \frac{8}{7}\left(\frac{11}{4}\right)^{4/3}\frac{\rho_{\rm GW}}{\rho_\gamma}\,, \tag{3.190}$$

and so the *Planck* constraint of $\Delta N_{\rm eff} < 0.30$ [2] can be used to bound the maximum amplitude of the gravitational wave signal $\Omega_{\rm GW} < 3.6\times10^{-6}$ [25]. Finally, we normalize the initial energy density in black holes $\Omega^i_{\rm BH}$ such that prior to decay, the black holes come to dominate the energy density of the universe. Thus these plots should be interpreted as giving an *upper bound* on the maximum possible signal. As the number of extra dimensions is increased, the spectrum is qualitatively similar but increasingly suppressed at high frequencies.

In Fig. 3.16, we fix $n = 2$ and illustrate the effect of changing other parameters. In the top left panel, we vary M; in the top right, M_*; in the bottom left, $T_{\rm re}$; and in the bottom right, t_i. A few comments are in order. First, decreasing the black hole mass moves the spectrum towards lower frequencies, which makes sense as smaller black holes evaporate more promptly, leading to a longer period of cosmological redshifting of the signal. This also serves to suppress the maximum amplitude of the signal. Next, we see that decreasing the

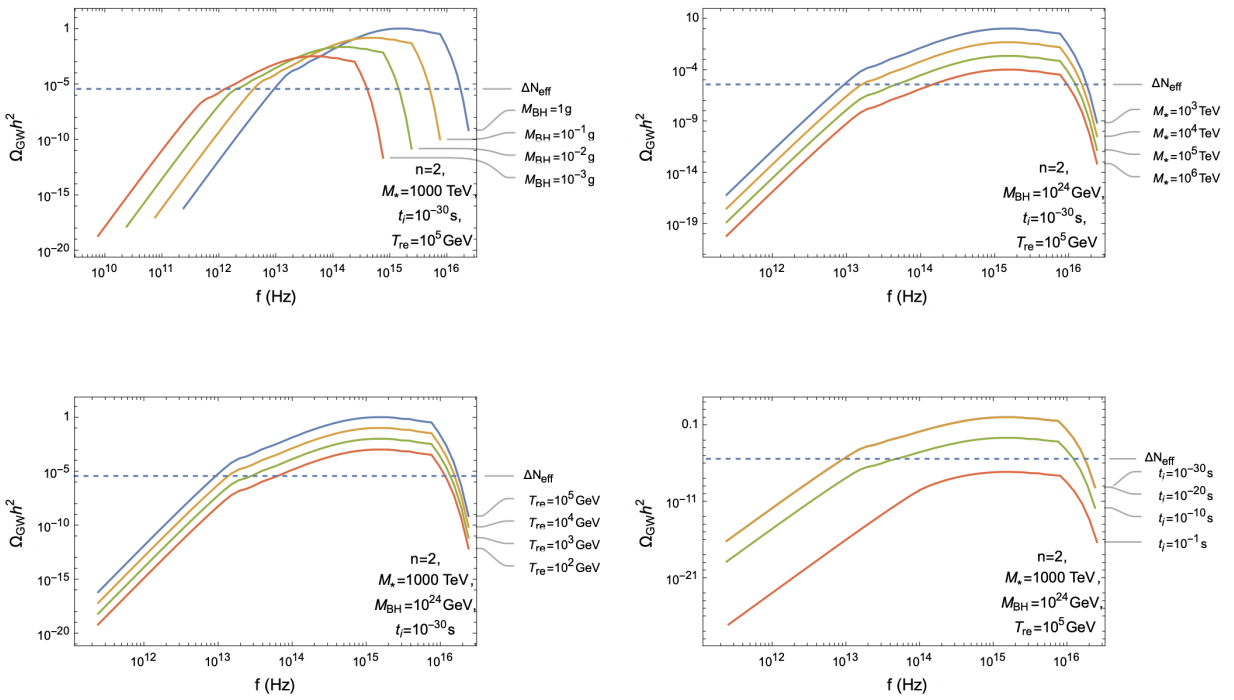

Figure 3.16: Gravitational wave spectra (in terms of the spectral density parameter $\Omega_{\rm GW}h^2$) for $n = 2$ and the base benchmark set of Fig. 3.15. In each panel we vary a single parameter. Top left: varying M; Top right: varying M_*; Bottom left: varying $T_{\rm re}$; Bottom right: varying t_i. Note that in the bottom right panel, the blue line corresponding to $t_i = 10^{-30}\,$s is essentially coincident with the yellow $t_i = 10^{-20}\,$s.

Planck scale M_* enhances the overall signal while leaving the location of the peak frequency unaffected. A similar effect is observed upon either increasing the reheating temperature $T_{\rm re}$ or decreasing the formation time t_i.

We want to identify the "optimal" scenario for experimental detection — i.e. the lowest frequency signal with maximal amplitude consistent with $\Delta N_{\rm eff}$ constraints. Clearly there are quite a few parameters that can be varied, so to determine the overall trends, in Figs. 3.17 and 3.18 we present contour plots of the peak frequency f and maximal spectral density parameter $\Omega_{\rm GW}h^2$, respectively, for various slicings of parameter space. In Fig. 3.16 we saw that changing the formation time and reheating temperature had minimal effect on the value of the peak frequency, which instead was primarily set by the choice of the reduced Planck scale M_* and the black hole mass M. Thus in Fig. 3.17 we set $t_i = 10^{-30}\,$s and

$T_{\rm re} = 10^5$ GeV and explore how peak frequency varies in the (M_*, M)-plane for $n = 2$ (left) and $n = 4$ (right). For both cases, the lowest values of the peak frequency correspond to low values of both M_* and M (bottom left corners). Increasing either[34] leads to peak gravitational wave emission at higher frequency.

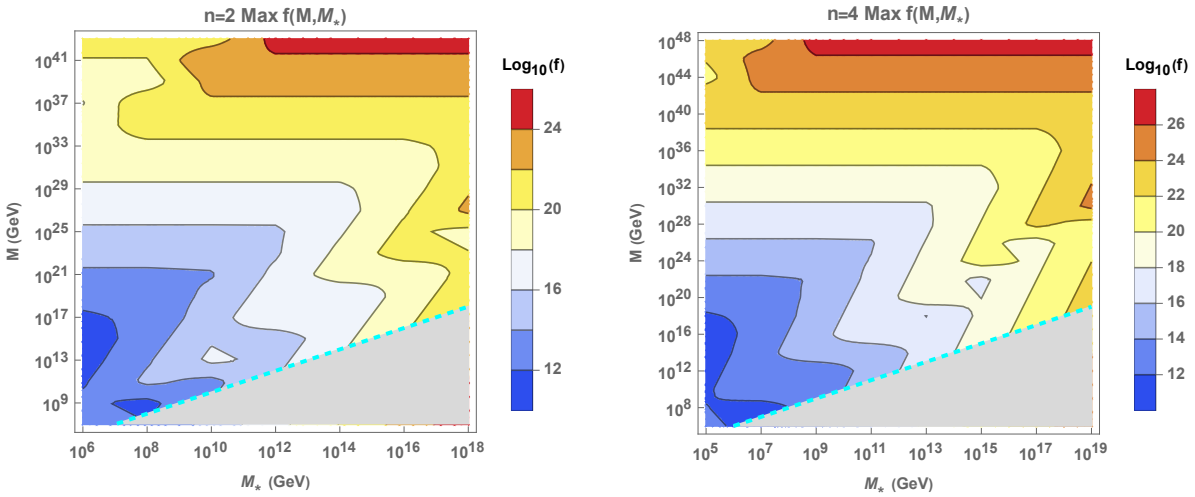

Figure 3.17: Contours of constant peak frequency in the plane defined by the reduced Planck scale M_* and the black hole mass M. Warmer colors indicate higher peak frequencies. Left panel: $n = 2$; Right panel: $n = 4$. We set the other parameters to the benchmark values of Fig. 3.15. The grey region is excluded on the basis that $M < M_*$.

Fig. 3.18 shows contour plots of the maximum gravitational wave amplitude in the (M_*, M)- and $(T_{\rm re}, t_i)$-planes for $n = 2$ (top), $n = 3$ (middle), and $n = 4$ (bottom) extra dimensions. The dashed green line and the grey region above it correspond to pairs of (M_*, M) that lead to evaporation at times larger than 1 s, violating BBN constraints. As in Fig. 3.17, the cyan dashed line and the grey region below should be excluded on the basis that $M < M_*$. The red dashed line corresponds to the $\Delta N_{\rm eff}$ bound, with larger amplitudes

34. It may appear surprising that increasing M leads to higher frequency, since from Eq. (3.85) we have that the Hawking temperature peaks at $T_H \sim (M_*/M)^{1/(n+1)} M_*$. This only sets the peak frequency at emission, though, and after redshifting the signal to today, it turns out that the peak frequency roughly scales as $f^0_{\rm peak} \sim T_0 (M/M_{\rm Pl})^{1/2}$, which is indeed increasing with increasing M. It is perhaps surprising that M_* does not appear in this expression, but this estimate was derived under many approximations (instant decay, blackbody emission), and it is expected that M_*-dependence enters at subleading order. Looking at the frequency plots, the M_*-dependence indeed appears quite weak, substantiating this reasoning.

excluded. We note that this rules out a very significant portion of the otherwise permissible parameter space, including nearly all t_i for $n = 2$ and $n = 3$ at the benchmark M and M_*.

From these trends, we are able to identify the "best-case" scenario illustrated in Fig. 3.19. The parameter values which yield the lowest peak frequency at the maximal amplitude still consistent with the $\Delta N_{\rm eff}$ bound are

$$n = 2,\ \beta = 1,\ T_{\rm re} = 16.5\,{\rm GeV},\ t_i = 10^{-30}\,{\rm s},\ M_* = 10^3\,{\rm TeV},\ M = 10M_*\,. \tag{3.191}$$

Note that the large degeneracy in parameter space means this is only one of many possible sets which could yield this result. While the tail of the signal extends down to hundreds of Hz, well within the range where Advanced LIGO (aLIGO) and Einstein Telescope (ET) are sensitive to a stochastic background [5], the amplitude is almost 10 orders of magnitude too small for detection. We note, however, that the peak frequency plateaus in the sub-MHz range accessible to planned high-frequency gravitational wave detectors, making this scenario a target for detection once their sensitivity exceeds $\Delta N_{\rm eff}$ bounds.

Summary

Relic gravitational waves from the Hawking evaporation of tiny primordial black holes are a generic prediction of many early universe scenarios beyond the standard cosmological paradigm. For light primordial black holes evaporating before BBN, however, the resultant gravitational wave spectra generically peak at ultra-high frequencies $\sim M_{\rm Pl}^2/M$. As discussed in [25], these signals remain out of reach of even the most optimistic proposed high frequency detectors, such that the only way to probe these scenarios is through integral bounds on the energy density contained in gravitational waves. This is sufficient for putting broad constraints on such scenarios, but does not provide any detailed information.

This motivated the consideration of black hole evaporation in the context of theories with

large extra dimensions, in which the reduced value of the true bulk Planck scale $M_* \ll M_{\rm Pl}$ allows for lower peak frequencies $\sim (M_*/M)^{1/(n+1)}\, M_*$. Our goal was to make predictions for gravitational wave spectra from the Hawking evaporation of ultra-light black holes in LED scenarios. To faithfully model black hole evaporation and evolution, we computed in detail the greybody factors for all particle species localized on the 4-dimensional "brane" as well as those propagating in the bulk. The resultant absorption coefficients are summarized in Table 3.2. We were then able to make predictions for the gravitational wave signal. Figs. 3.15 and 3.16 show the effect of changing the number of extra dimensions n, the bulk Planck scale M_*, the black hole mass M, the reheating temperature $T_{\rm re}$, and the formation time t_i. Figs. 3.17 and 3.18 explore this parameter space in more detail and present contour plots for the peak frequency and maximal amplitude. The trends established in these plots allowed us to identify an "optimal" scenario for experimental detection, shown in Fig. 3.19.

We find that the lowest possible peak frequencies are obtained for models with very low bulk Planck scale M_* and very light black holes with mass only slightly above M_*. The dependence on the number of extra dimensions is weaker, but increasing n weakly pushes $f_{\rm peak}$ to slightly higher frequencies. Finally, the peak frequency is relatively insensitive to formation time and reheating temperature, though these affect the amplitude of the gravitational wave signal. We find that for the sample "optimal" parameter set of Eq. (3.191), the peak frequency plateaus in the sub-MHz range. This should be accessible to planned high-frequency gravitational wave detectors, once their sensitivities are improved to exceed the $\Delta N_{\rm eff}$ bounds. The tail of this spectrum extends into the $\sim$ 100 Hz range, however in order for this tail to have an amplitude within sensitivity, the peak would violate the integral bound on the effective number of relativistic degrees of freedom $\Delta N_{\rm eff}$. Should high-frequency ($f \gtrsim$ kHz) observatories dip below the $\Delta N_{\rm eff}$ constraints, observation of evaporating extra dimensional black holes could be possible.

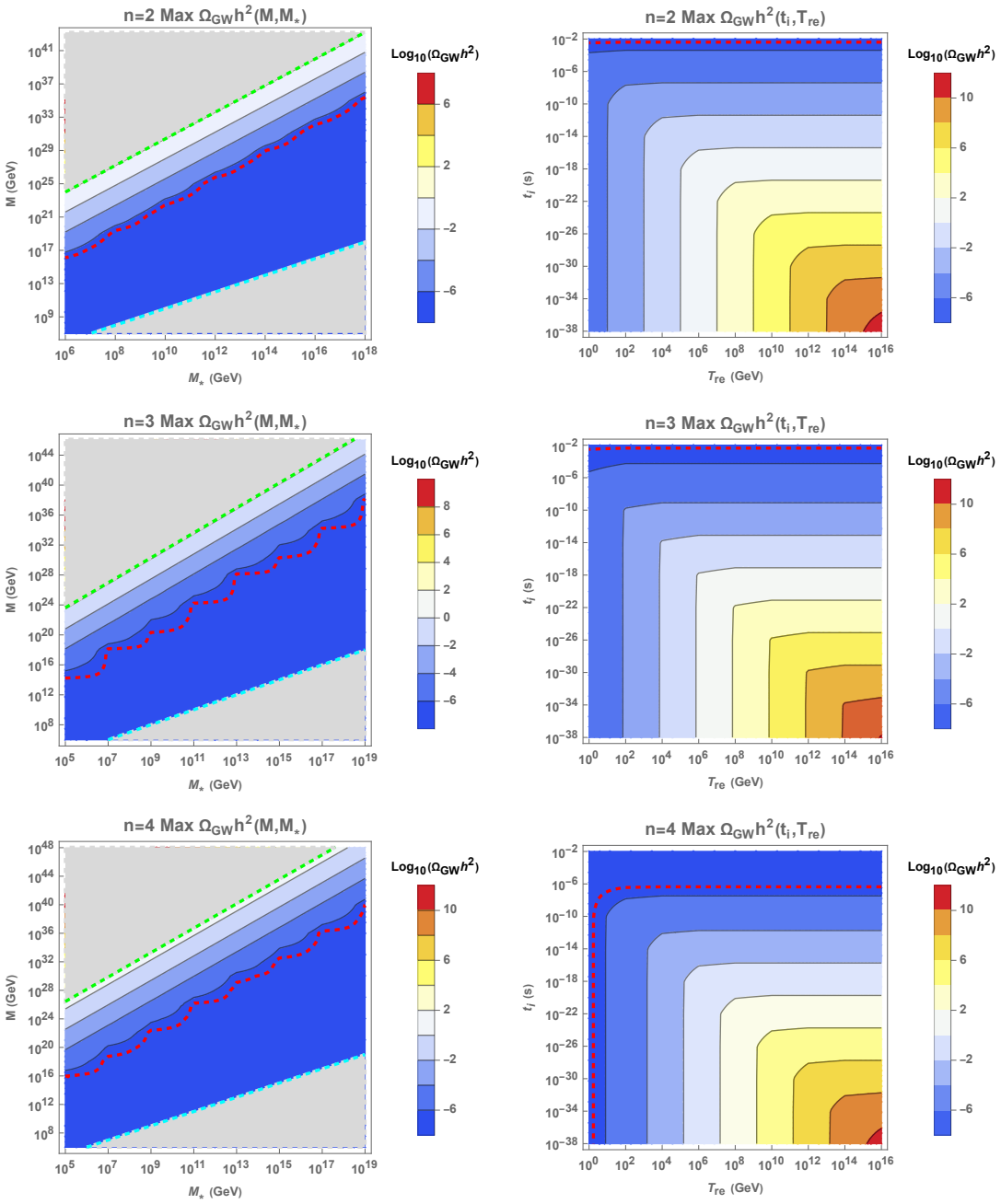

Figure 3.18: Contours of constant gravitational wave energy density $\Omega_{\rm GW}h^2$ at peak frequency in the (M_*, M)-plane (left panels) at fixed $t_i = 10^{-30}\,\mathrm{s}$ and $T_{\rm re} = 10^5\,\mathrm{GeV}$, and in the $(T_{\rm re}, t_i)$-plane (right panels) at fixed $M_* = 10^3\,\mathrm{TeV}$ and $M = 1\,\mathrm{g}$. As in the previous figure, warmer colors indicate larger amplitudes. The top panels have $n = 2$, the middle panels $n = 3$, and the bottom panels $n = 4$ extra dimensions. Amplitudes corresponding to the constraint on $\Delta N_{\rm eff}$ are shown with red dashed lines, with larger amplitudes excluded; the green dashed lines indicate parameter space where evaporation occurs during BBN, with larger amplitudes ruled out; finally, the cyan dashed lines indicate "quantum black holes" (i.e. those with $M = M_*$), with the grey area below excluded.

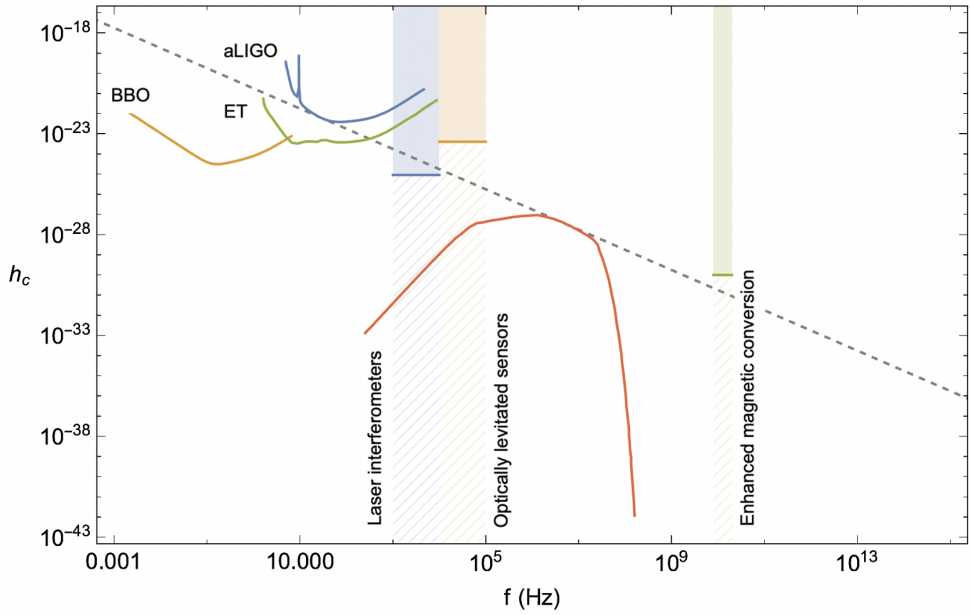

Figure 3.19: Gravitational wave prediction (in terms of the dimensionless characteristic strain h_c) for an "optimal" scenario for experimental detection — i.e. the lowest frequency signal with maximal amplitude consistent with $\Delta N_{\rm eff}$ constraints (grey dashed line). The corresponding parameter values are given in Eq. (3.191). Superposed are several current and proposed gravitational wave detectors and their projected sensitivities to a stochastic gravitational wave background. These include the ground-based observatories Advanced LIGO (aLIGO) and Einstein Telescope (ET) and the space-based Big Bang Observer (BBO) (sensitivities taken from [5]), as well as several prospective high-frequency gravitational wave detection technologies: laser interferometers, optically levitated sensors, and enhanced magnetic conversion (sensitivities taken from [6]).

CHAPTER 4

COSMOLOGICAL PHASE TRANSITIONS

As we have already seen from our discussion of the Standard Model in Sec. 2.2, the forms of the fundamental forces and the spectrum of fundamental particles looked very different in the hot, dense conditions immediately following the Big Bang. As the universe expanded and cooled, it proceeded through at least two cosmological phase transitions, whereby the forces and matter came to take the forms now seen at low energies of today. In addition to the QCD and electroweak phase transitions, there may also have been phase transitions in dark sectors, potentially detectable through their gravitational wave signatures.

Cosmological phase transitions present yet another fertile breeding ground for the symbiosis between cosmology and particle physics. Indeed, much of what we know about the QCD and electroweak phase transitions comes from laboratory experiments, collider data, and lattice studies. While much remains to be determined about the exact details of these transitions, here is what we do know:

Electroweak Phase Transition:

The electroweak phase transition occurred at the moment of electroweak symmetry breaking, which we already discussed at length in Sec. 2.2.4. Chronologically, this event is estimated to have taken place $\sim 10^{-11}\,\mathrm{s}$ after the Big Bang, at which point the universe had cooled to $\sim 200\,\mathrm{GeV}$. Based on our current understanding of the Standard Model and the Higgs sector, this is the temperature at which the minimum of the finite-temperature effective potential with non-vanishing Higgs vev $\langle H \rangle \neq 0$ became thermodynamically favored. As a consequence, the Higgs developed a non-zero vev, the value of which today (at essentially vanishing temperature $T \simeq 0$) has been measured as $v \simeq 246\,\mathrm{GeV}$. This triggered the breaking of electroweak symmetry $SU(2)_L \times U(1)_Y$ down to electromagnetism $U(1)_{\mathrm{EM}}$. The gauge bosons of the broken weak nuclear force, $W^{\pm}$ and Z, acquired masses through this

process, along with the fermions appearing in the Yukawa sector.

Presuming only the Standard Model particle content, the electroweak phase transition is expected to be *second order*, with all parameters smoothly transitioning to their final values. First-order phase transitions, on the other hand, proceed via the nucleation, growth, and mergers of bubbles of true vacuum. First-order phase transitions are of particular interest because they come with gravitational wave signals from bubble collisions, acoustic waves, and turbulence. A first-order electroweak phase transition is of even further interest because it is a prerequisite for electroweak baryogenesis, one of the leading candidates for establishing the matter-antimatter asymmetry of the universe. While the electroweak phase transition predicted in the Standard Model is not first-order, it may be made so with very minimal extensions. See e.g. [241].

QCD Phase Transition:

The QCD phase transition is the event wherein quarks and gluons first became bound in hadrons. This occurred due to the curious fact[1] that the strong force becomes stronger at lower energies (or equivalently, longer distances), as discussed in Sec. 2.2.2. Simluations on the lattice have determined the QCD scale to be roughly $\Lambda_{\text{QCD}} \sim 0.3\,\text{GeV}$, though the exact value can vary by study. For energies greater than Λ_{QCD}, the strong force can be treated perturbatively; for energies less than Λ_{QCD}, we have the non-perturbative phenomenon of *color confinement*, wherein the color-charged quarks become bound in color neutral combinations of quark-antiquark pairs (mesons) or three quarks (baryons). The phase transition, then, is estimated to have occurred when the universe had cooled to $T \sim 0.3\,\text{GeV}$, corresponding to $\sim 10^{-5}\,\text{s}$ after the Big Bang. The QCD phase transition is also a second order *smooth crossover*.

1. This follows from the running of the QCD coupling g_s, and can intuitively be understood as an "anti-screening" effect from the gluons carrying color charge.

In addition to these relatively established transitions, phase transitions may have also occurred in dark sectors coupled to the Standard Model only through gravity or portal operators. We will discuss one such transition in Sec. 5.1. If strongly first-order, such a dark sector transition could still be detectable through its gravitational wave signatures. The possibility of a first-order electroweak phase transition has received much attention of late as upcoming space-based gravitational wave interferometers like LISA [242] and DECIGO [243] will be sensitive to the mHz range, where the frequency of gravitational waves from a $T \sim 200\,\mathrm{GeV}$ electroweak phase transition is expected to peak. Meanwhile, third generation ground-based gravitational wave observatories like Einstein Telescope [244] promise increased sensitivity at higher frequencies, which is relevant for signals from a high-temperature electroweak phase transition.

In light of the coming influx of experimental data, much work is needed on the theoretical front, as perturbative calculations of first-order phase transitions suffer from large theoretical uncertainties which lead to predictions for gravitational wave spectra varying by orders of magnitude. These include uncertainties in the determination of thermal phase transition parameters — stemming from issues of renormalization group scale and gauge dependence, choice of thermal resummation method, and more — as well as in the phase transition dynamics, particularly the determination of the bubble wall velocity and details of nucleation. See e.g. [245] for a comprehensive discussion.

In this chapter, we present some efforts in this direction. We begin in Sec. 4.1 by reviewing first-order phase transitions and their gravitational wave signals. In Sec. 4.2, we address the issue of how to correctly implement thermal resummation of the effective potential, comparing the efficiency of different resummation schemes at 1- and 2-loop level and assessing the error incurred by taking the high-temperature approximation. We advocate for the so-called "partial dressing prescription" and extend this formalism to the case of mixing fields, making it suitable for phenomenological studies of beyond the Standard Model extensions

of the Higgs sector. Finally in Sec. 4.3, we look at some novel analytical bounds on 2 Higgs doublet models, which present among the simplest extensions of the Standard Model potentially capable of generating a strongly first-order phase transition and possibly even a successful electroweak baryogenesis.

4.1 First-Order Phase Transitions

In this section, we present a general overview of first-order phase transitions. We begin in Sec. 4.1.1 by reviewing phase transition thermodynamics and the hydrodynamic description of expanding bubbles in a hot thermal plasma. Then in Sec. 4.1.2, we review the sources and production of gravitational waves in first-order cosmological phase transitions. For further reading, see e.g. [246, 247].

4.1.1 Phase Transition Thermodynamics and Hydrodynamics

Free Energy

A first-order phase transition proceeding via bubble nucleation in the hot thermal plasma of the early universe can be modeled by a scalar-fluid system. The central quantity governing the thermodynamics of the phase transition is the free energy density $\mathcal{F}$. Both the fluid energy density e_f and pressure p_f are defined in terms of the fluid free energy density $\mathcal{F}_f$ as

$$e_f = \mathcal{F}_f - T\frac{\partial \mathcal{F}_f}{\partial T}\,, \quad p_f = -\mathcal{F}_f\,. \tag{4.1}$$

The free energy of the fluid is simply the sum of the free energies of its thermalized particle constituents,

$$\mathcal{F}_f = \sum_B n_B \mathcal{F}_B + \sum_F n_F \mathcal{F}_F\,, \tag{4.2}$$

with $n_{B/F}$ the number of degrees of freedom per species. To compute this explicitly, recall that the free energy density of a single bosonic degree of freedom with mass m is

$$\mathcal{F}_B = T \int \frac{d^3k}{(2\pi)^3} \ln\left(1 - e^{-\beta\omega_k}\right) \equiv T^4 J_B\left(\frac{m}{T}\right) , \tag{4.3}$$

where we have defined the bosonic thermal function J_B. Similarly for a fermionic degree of freedom,

$$\mathcal{F}_F = -T \int \frac{d^3k}{(2\pi)^3} \ln\left(1 + e^{-\beta\omega_k}\right) \equiv T^4 J_F\left(\frac{m}{T}\right) , \tag{4.4}$$

with J_F the fermionic thermal function. These admit the following high temperature expansions

$$\mathcal{F}_B \simeq -\frac{\pi^2}{90}T^4 + \frac{1}{24}m^2T^2 - \frac{1}{12\pi}m^3T - \frac{1}{32\pi^2}m^4\left[\ln\left(\frac{me^{\gamma_E}}{4\pi T}\right) - \frac{3}{4}\right] , \tag{4.5a}$$

$$\mathcal{F}_F \simeq -\frac{7}{8}\frac{\pi^2}{90}T^4 + \frac{1}{48}m^2T^2 - \frac{1}{32\pi^2}m^4\left[\ln\left(\frac{me^{\gamma_E}}{\pi T}\right) - \frac{3}{4}\right] . \tag{4.5b}$$

It is often useful to separate the field-independent term $\propto T^4$. If we define the rest to be the "thermal potential" V_T, then the total free energy density of the fluid can be expressed

$$\mathcal{F}_f = -\frac{\pi^2}{90}g_\star T^4 + V_T \, , \tag{4.6}$$

where $g_\star = \sum_B n_B + \frac{7}{8}\sum_F n_F$ is the effective number of degrees of freedom and

$$V_T = \frac{T^2}{24}\left(\sum n_B m_B^2 + \frac{1}{2}\sum_F n_F m_F^2\right) - \frac{T}{12\pi}\sum n_B m_B^3 \, , \tag{4.7}$$

neglecting higher order terms for simplicity. This quantity then defines the fluid pressure

$$p_f = \frac{\pi^2}{90}g_\star T^4 - V_T \, . \tag{4.8}$$

Meanwhile, the free energy density for the scalar field is simply the tree level potential

$$\mathcal{F}_\phi = V_0 \,. \tag{4.9}$$

The total free energy density for the scalar-fluid system is thus

$$\mathcal{F} = -\frac{\pi^2}{90} g_\star T^4 + V_{\text{eff}}(\phi, T) \,, \tag{4.10}$$

where we have defined the finite temperature effective potential $V_{\text{eff}}(\phi, T) = V_0(\phi) + V_T(\phi, T)$.

Fluid Dynamics

The total energy momentum of the combined scalar-fluid system is given by the sum

$$T^{\mu\nu} = T_f^{\mu\nu} + T_\phi^{\mu\nu} \,, \tag{4.11}$$

where the energy momentum of the scalar field with tree-level potential $V_0(\phi)$ is the usual

$$T_\phi^{\mu\nu} = \partial^\mu \phi \partial^\nu \phi - g^{\mu\nu} \left(\frac{1}{2} (\partial\phi)^2 + V_0 \right) \,. \tag{4.12}$$

When the thermal fluid is in local equilibrium, its energy momentum tensor takes the form of a perfect fluid with energy density e_f and pressure p_f

$$T_f^{\mu\nu} = (e_f + p_f) u^\mu u^\nu + p_f g^{\mu\nu} \,, \tag{4.13}$$

where u^μ is the 4-velocity of the fluid. Of course during the phase transition near the wall, we expect the system to be far from static however. Under such non-equilibrium conditions, the perfect fluid description no longer holds, and more generically the energy momentum

tensor for the fluid is given by

$$T_f^{\mu\nu} = \int \frac{d^3k}{(2\pi)^3} \frac{k^\mu k^\nu}{k^0} f(k,x)\,, \tag{4.14}$$

where $f(k,x)$ is the distribution function of the relativistic fluid. In the event that f is described by the equilibrium distribution, one can show that this more general definition reproduces Eq. (4.13).

To understand the phase transition dynamics, we need to understand how the energy momentum in either component of our scalar-fluid system evolves. Consider first the fluid component; in the absence of external forces, we expect the conservation law $\partial_\mu T_f^{\mu\nu} = 0$ to hold. We are not in this situation, however; the field-dependency of the effective masses of the particles in the plasma generates an external force as the value of the scalar changes across the wall interface.

To see this, note first that the particle masses appearing in the fluid free energy $\mathcal{F}_f$ of Eqs. (4.6) and (4.7) are actually field dependent effective masses $m^2 = \mu^2 + c\phi^2$, with μ the bare mass and c a function of dimensionless coupling constants. When the value of the scalar field changes during the phase transition, so do the masses of the particle species in the plasma. It is ultimately this field-dependency of the mass which should be interpreted as sourcing the external force which drives the bubble wall.

Consider the action for a single free particle with field-dependent mass $m(\phi)$

$$S = -\int d\tau\, m\sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}}\,. \tag{4.15}$$

Varying this action yields the equation of motion

$$\frac{d}{d\tau}\left(m\frac{dx^\mu}{d\tau}\right) + \partial^\mu\phi\frac{dm}{d\phi} = 0\,, \tag{4.16}$$

from which we can identify the force

$$F^\mu = -\partial^\mu \phi \frac{dm}{d\phi} \,. \tag{4.17}$$

F^μ enters into the relativistic Boltzmann equation describing the evolution of the particle distribution function $f(k,x)$ [248]

$$\left(k^\mu \partial_\mu + m F^\mu \frac{\partial}{\partial k^\mu} \right) \Theta(k^0) \delta(k^2 + m^2) f(k,x) = C[f] \,, \tag{4.18}$$

where the Heaviside function enforces positivity of energy and the delta function ensures the on-shell condition is satisfied. The collision function $C[f]$ on the right-hand side of this expression describes 2-body scattering and is generically a very complicated function. If we assume that particle number and momentum are conserved in these 2-body scatterings, though, $C[f]$ does have the nice property that

$$\int \frac{d^4k}{(2\pi)^4} \Big(a(x) + b_\mu(x) k^\mu \Big) C[f] = 0 \,. \tag{4.19}$$

The fact that $a(x)$ and $b_\mu(x)k^\mu$ are collision invariants will be useful in what follows.

We want to use the relativistic Boltzmann equation to determine the evolution of the particle distribution function $f(k,x)$ and thereby the fluid energy momentum tensor $T_f^{\mu\nu}$. Multiplying both sides of Eq. (4.18) by k^ν and integrating over 4-momentum should give

zero. This allows us to write

$$
\begin{aligned}
0 &= \int \frac{d^4k}{(2\pi)^4} k^\nu C[f] \\
&= \int \frac{d^4k}{(2\pi)^4} k^\nu \left(k^\mu \partial_\mu + m F^\mu \frac{\partial}{\partial k^\mu} \right) \Theta(k^0) \delta(k^2 + m^2) f(k, x) \\
&= \partial_\mu \int \frac{d^4k}{(2\pi)^4} k^\mu k^\nu \Theta(k^0) \delta(k^2 + m^2) f + m F^\mu \int \frac{d^4k}{(2\pi)^4} k^\nu \frac{\partial}{\partial k^\mu} \Theta(k^0) \delta(k^2 + m^2) f \\
&= \partial_\mu \int \frac{d^3k}{(2\pi)^3} \frac{k^\mu k^\nu}{2k^0} f - m F^\mu \delta^\nu_\mu \int \frac{d^4k}{(2\pi)^4} \Theta(k^0) \delta(k^2 + m^2) f \\
&= \frac{1}{2} \partial_\mu T^{\mu\nu}_f - m F^\nu \int \frac{d^3k}{(2\pi)^3} \frac{1}{2k^0} f \Big|_{k^0 = E_k} ,
\end{aligned}
\tag{4.20}
$$

where in the last line we have used Eq. (4.14). Finally using the explicit form for F^ν in Eq. (4.17) and generalizing from a fluid of a single species to the SM case by performing a sum over fluid degrees of freedom i, we have

$$
\partial_\mu T^{\mu\nu}_f = -\partial^\nu \phi \sum_i n_i \frac{dm_i^2}{d\phi} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2k^0} f_i(k, x) \Big|_{k^0 = E_k} . \tag{4.21}
$$

Thus we see that the fluid stress tensor is not conserved, but rather sourced by the change in mass across the wall interface. Note, however, that the total energy momentum of the combined scalar-fluid system *should* be conserved,

$$
\partial_\mu T^{\mu\nu} = 0 . \tag{4.22}
$$

Thus the energy momentum of the field should evolve as

$$
\partial_\mu T^{\mu\nu}_\phi = \partial^\nu \phi \sum_i n_i \frac{dm_i^2}{d\phi} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2k^0} f_i(k, x) \Big|_{k^0 = E_k} , \tag{4.23}
$$

and we should think of fluid contributions as changing the field's energy and momentum. The equation of motion for the scalar field can be obtained by inserting the explicit expression

for $T^{\mu\nu}_{\phi}$ of Eq. (4.12) and evaluating

$$\begin{aligned}\partial_\mu T^{\mu\nu}_{\phi} &= \partial_\mu \left[\partial^\mu\phi\partial^\nu\phi - g^{\mu\nu}\left(\frac{1}{2}(\partial\phi)^2 + V_0\right)\right] \\ &= \Box\phi\partial^\nu\phi + (\partial^\mu\phi)\partial_\mu\partial^\nu\phi - \frac{1}{2}\partial^\nu(\partial_\rho\phi\partial^\rho\phi) - \partial^\nu\phi V_0' \\ &= (\partial^\nu\phi)\Big(\Box\phi - V_0'\Big)\,.\end{aligned} \tag{4.24}$$

The equation of motion is then

$$\Box\phi - V_0'(\phi) = \sum_i n_i \frac{dm_i^2}{d\phi}\int\frac{d^3k}{(2\pi)^3}\frac{1}{2E_k}f_i(k,x)\,. \tag{4.25}$$

Following [248], we consider a distribution which is close to the equilibrium distribution and write

$$f(k,x) = f^{\rm eq}(k,x) + \delta f(k,x)\,, \tag{4.26}$$

where

$$f^{\rm eq}(k,x) = \frac{1}{e^{E_k/T}\mp 1}\,. \tag{4.27}$$

Recall that the thermal contribution to the effective potential reads

$$V_T = \frac{\pi^2}{90}g_\star T^4 + \sum_i \pm n_i T\int\frac{d^3k}{(2\pi)^3}\ln\left(1\mp e^{-E_k/T}\right)\,, \tag{4.28}$$

where $E_k = \sqrt{\vec{k}^{\,2} + m_i^2}$ and the top (bottom) sign is for bosons (fermions). Taking the

derivative with respect to ϕ yields

$$\begin{aligned} V_T' &= \sum_i \pm n_i T \int \frac{d^3k}{(2\pi)^3} \frac{d}{d\phi} \ln\left(1 \mp e^{-E_k/T}\right) \\ &= \sum_i n_i \int \frac{d^3k}{(2\pi)^3} \left(\frac{1}{e^{E_k/T} \mp 1}\right) \frac{d}{d\phi} \sqrt{\vec{k}^2 + m_i^2} \\ &= \sum_i n_i \frac{dm_i^2}{d\phi} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2E_k} \left(\frac{1}{e^{E_k/T} \mp 1}\right) \\ &= \sum_i n_i \frac{dm_i^2}{d\phi} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2E_k} f_i^{\rm eq}(k,x)\,. \end{aligned} \tag{4.29}$$

Using this result and the decomposition $f = f^{\rm eq} + \delta f$, Eq. (4.25) can be written as

$$\Box\phi - V_{\rm eff}'(\phi) = \sum_i n_i \frac{dm_i^2}{d\phi} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2E_k} \delta f_i(k,x)\,, \tag{4.30}$$

where the effective potential is the sum of tree level and one-loop thermal terms, $V_{\rm eff} = V_0 + V_T$. We see that the right-hand side is sourced by departures from the equilibrium phase space distribution $f^{\rm eq}$, which is in turn induced by gradients of the scalar field $\partial_\mu\phi$. This should be interpreted as a thermal "friction" term, since as we will see, without it the bubble wall would propagate at the speed of light.

The Expanding Bubble

We are now ready to turn to a hydrodynamic description of the expanding bubble. First, note that intuitively the condition for a bubble of true vacuum nucleated within the false vacuum to expand is that the pressure inside by greater than the pressure outside

$$p_{\rm in} > p_{\rm out}\,, \tag{4.31}$$

where the pressure of the total scalar-fluid system is given by (see Eq. (4.10))

$$p = \frac{\pi^2}{90} g_\star T^4 - V_{\rm eff}(\phi, T)\,. \tag{4.32}$$

Thus taking $\phi_{\rm EW}$ to be the value of the scalar field inside the bubble of true vacuum and 0 to be the value outside, bubble growth requires

$$V_{\rm eff}(\phi_{\rm EW}) < V_{\rm eff}(0)\,, \tag{4.33}$$

which can only be satisfied below the critical temperature T_c. When this condition is fulfilled, the bubble wall will move outward through the plasma at a speed which is governed by the dynamics of the scalar field and its interactions with the plasma, as described by Eq. (4.30). It is customary to parameterize the friction term appearing on the right-hand side of the scalar equation as

$$\sum_i n_i \frac{dm_i^2}{d\phi} \int \frac{d^3k}{(2\pi)^3} \frac{1}{2E_k} \delta f_i(k,x) \equiv \eta_T u^\mu \partial_\mu \phi\,, \tag{4.34}$$

where $\eta_T(\phi, T, v)$ is some function of the scalar field, temperature, and possibly fluid velocity. As for the motivation for this parameterization, the claim is that it makes it apparent that conservation of the plasma entropy current is violated by the deviation of f from its equilibrium distribution. To see this, we define the plasma entropy current as $S^\mu = s u^\mu$, where $s = \frac{\partial p}{\partial T}$ is the entropy of the fluid and $u^\mu = \gamma(1, \vec{v})$ the fluid 4-velocity. By taking Eq. (4.21) with the perfect fluid Ansatz of Eq. (4.13) and this parameterization of the out-of-equilibrium term and multiplying by u_ν, one can then show that

$$\partial_\mu S^\mu = \frac{\eta_T}{T} (u^\mu \partial_\mu \phi)^2\,, \tag{4.35}$$

where we have also made use of the fact that $u_\nu \partial_\mu u^\nu = 0$.

Starting from our equation for the scalar-fluid dynamics,

$$\Box\phi - V'_{\text{eff}} = \eta_T u^\mu \partial_\mu \phi \,, \tag{4.36}$$

we would like to solve for the fluid profile around an expanding bubble. Let us presume that the bubble is sufficiently large to approximate the wall as planar, and take it to move in the z-direction with a wall speed v_w which is constant in the rest frame of the universe, establishing a steady-state flow. We work in the wall frame, where the fluid 4-velocity is $u^\mu(z)$ and the scalar field profile is $\phi(z)$. Assuming that the fluid speed changes only minimally across the interface, the 4-velocity is approximately

$$u^\mu \simeq \gamma_w (1, 0, 0, v_w) \,. \tag{4.37}$$

For this particular setup, Eq. (4.36) becomes

$$\partial_z^2 \phi - V'_{\text{eff}} = \eta_T \gamma_w v_w \partial_z \phi \,. \tag{4.38}$$

We can obtain an effective equation of motion for the bubble wall expansion by integrating the scalar equation of motion across the bubble wall interface. Multiplying both sides by $\partial_z\phi$ and integrating over z, we find

$$\begin{aligned} &\int dz\, (\partial_z\phi)(\partial_z^2\phi) - \int dz\, (\partial_z\phi)(\partial_\phi V_{\text{eff}}) = \int dz\, \eta_T \gamma_w v_w (\partial_z\phi)^2 \\ \Rightarrow\ &\int dz\, \partial_z \left[\frac{1}{2}(\partial_z\phi)^2 - V_{\text{eff}}\right] = \gamma_w v_w \int dz\, \eta_T (\partial_z\phi)^2 \\ \Rightarrow\ &\Delta V_{\text{eff}} \simeq -\gamma_w v_w \int dz\, \eta_T (\partial_z\phi)^2 \,, \end{aligned} \tag{4.39}$$

where in going from the first to second line we have used the facts that $\frac{1}{2}\partial_z(\partial_z\phi)^2 = \partial_z\partial_z^2\phi$ and $\frac{d\phi}{dz}\frac{dV_{\text{eff}}}{d\phi} = \partial_z V_{\text{eff}}$, and in going to the last, we have assumed $\partial_z\phi\big|_\infty \simeq \partial_z\phi\big|_{-\infty}$ and

defined $\Delta V_{\rm eff} = V_{\rm eff}\big|_{\infty} - V_{\rm eff}\big|_{-\infty}$ as the pressure difference across the wall for the total scalar-fluid system. In principle one can solve this equation for the wall velocity v_w, though in practice this is quite difficult.

Assuming the wall velocity is known and further assuming that the phase transition completes sufficiently quickly that cosmological expansion can be neglected, one can express the volume $\mathcal{V}$ of an expanding bubble nucleated at time t' as

$$\mathcal{V}(t,t') = \frac{4\pi}{3} v_w^3 (t-t')^3 \,. \tag{4.40}$$

In order to track how the phase transition proceeds, we would like an expression for the fractional volume of the universe in the metastable phase (i.e. the volume not occupied by bubbles). We will denote this quantity $h \equiv \frac{V_{\rm false}}{V_{\rm tot}} = 1 - \frac{V_{\rm true}}{V_{\rm tot}}$. To compute h, we will need to know the rate at which bubbles nucleate, $\Gamma(t)$, defined as the number of bubbles nucleated per unit time per unit coordinate volume. In the simplified scenario where we ignore any overlaps or collisions between bubbles, the true vacuum fraction is then simply the integrated product of the nucleation rate and bubble volume, and the false vacuum fraction is 1 minus this quantity,

$$h(t) = 1 - \int_{t_n}^{t} dt' \, \mathcal{V}(t,t') \Gamma(t') \,. \tag{4.41}$$

Note that we have taken the lower limit of integration to be the time t_n corresponding to the nucleation temperature, though elsewhere in the literature it's taken to be that corresponding to the critical temperature, t_c.

In reality, bubble collisions and overlap are non-negligible effects, especially as the phase transition nears completion. To incorporate these effects, we follow the logic of Ref. [249] and consider filling a space with randomly placed spheres, which are allowed to be overlapping or nested. Let $p(V_1, V_2)$ denote the probability that a given point is **not** contained within a

sphere with volume between V_1 and V_2. The probability $p(V_1, V_2 + dV_2)$ can be factorized as

$$p(V_1, V_2 + dV_2) = p(V_1, V_2)p(V_2, V_2 + dV_2)\,, \tag{4.42}$$

where the probabilities on the right hand side are multiplied because we want the events (the point not being contained in a sphere between V_1 and V_2, and the point not being contained in a sphere between V_2 and $V_2 + dV_2$) to occur simultaneously. Now let $n(V)dV$ be the *fractional number density*[2] of spheres with volume between V and $V + dV$, such that we can write

$$p(V_2, V_2 + dV_2) = 1 - n(V_2)V_2 dV_2\,. \tag{4.43}$$

That is, the probability of the point not being contained in a sphere between V_2 and $V_2 + dV_2$ is equal to 1 minus the fractional volume occupied by these spheres. Then

$$p(V_1, V_2 + dV_2) = p(V_1, V_2)(1 - n(V_2)V_2 dV_2)\,. \tag{4.44}$$

The left hand side of this expression can also be expanded in a Taylor series as

$$p(V_1, V_2 + dV_2) \simeq p(V_1, V_2) + \frac{dp(V_1, V_2)}{dV_2} dV_2 + ... \tag{4.45}$$

Equating the right hand sides of the equations above, we arrive at the differential equation

$$\frac{dp(V_1, V_2)}{p(V_1, V_2)} = -n(V_2)V_2 dV_2\,, \tag{4.46}$$

which is solved by[3]

$$p(V_1, V_2) = \exp\left[-\int_{V_1}^{V_2} dV\, n(V)V\right]\,. \tag{4.47}$$

2. The entire quantity $n(V)dV$ has dimensionality L^{-3}.

3. The symmetry under $V_1 \leftrightarrow V_2$ allows us to change the integration variable on the right hand side.

In particular, the probability of a point not being in *any* sphere is obtained in the limit $V_1 \to 0$, $V_2 \to \infty$, and is equivalent to the false vacuum fraction

$$h \equiv p(0,\infty) = \exp\left[-\int_0^\infty dV\, n(V)V\right] . \tag{4.48}$$

Finally, the integral appearing in the exponential is the total fractional volume contained in true vacuum bubbles, which can equivalently be expressed in terms of the nucleation rate and volume of an expanding bubble

$$\int_0^\infty dV\, n(V)V = \int_{t_n}^t dt'\, \mathcal{V}(t,t')\Gamma(t') \,, \tag{4.49}$$

where $\mathcal{V}(t,t')$ is defined in Eq. (4.40). This leads to the following expression for the fractional volume remaining in the metastable phase

$$h(t) = \exp\left[-\int_{t_n}^t dt'\, \mathcal{V}(t,t')\Gamma(t')\right] , \tag{4.50}$$

which now incorporates the effect of overlaps. Since bubbles can only nucleate in the metastable phase, knowledge of h now allows us to compute the number density of bubbles as a function of time

$$n_{\text{bubble}}(t) = \int_{t_n}^t dt'\, \Gamma(t')h(t') \,. \tag{4.51}$$

4.1.2 Gravitational Wave Production

Both the scalar field and fluid are a source of metric perturbations h_{ij}, and hence gravitational waves. Recall that the entire energy momentum tensor can be decomposed into contributions from the scalar field and fluid as $T^{\mu\nu} = T^{\mu\nu}_\phi + T^{\mu\nu}_f$, with

$$T^{\mu\nu}_\phi = \partial^\mu\phi\partial^\nu\phi - \left(\frac{1}{2}(\partial\phi)^2 + V_0\right) g^{\mu\nu} \,, \tag{4.52a}$$

$$T_f^{\mu\nu} = (e_f + p_f)u^\mu u^\nu + p_f g^{\mu\nu}\,, \tag{4.52b}$$

where the fluid pressure is given by the finite temperature contribution to the effective potential $p_f = -V_T$ and we parameterize the 4-velocity u^μ as $\gamma(1, \vec{v})$. In particular, the shear stress Π_{ij} which sources the metric perturbations h_{ij} is given by projecting out the transverse traceless part of the spatial energy momentum tensor T^{kl} via $\Pi_{ij} = \Lambda_{ij,kl}T^{kl}$, where $\Lambda_{ij,kl} = \hat{P}_{ik}\hat{P}_{kl} - \frac{1}{2}\hat{P}_{ij}\hat{P}_{kl}$ and $\hat{P}_{ij} = \delta_{ij} - \hat{k}_i\hat{k}_j$. Explicitly we find

$$\Pi_{ij} = \gamma^2(e_f + p_f)v_i v_j + \partial_i\phi\partial_j\phi\,, \tag{4.53}$$

where the first term is the fluid contribution and the second comes from the scalar field.

Gravitational wave production in a first-order cosmological phase transition can be decomposed into three main stages:

- **Bubble collisions:** This initial stage of bubble collisions and mergers breaks spherical symmetry, such that the gradient energy of the scalar field (idealized as being localized on the bubble wall) can source gravitational waves. We denote this as the scalar contribution Ω_ϕ and note that it tends to be subdominant unless the transition is very strong.

- **Acoustic stage:** Each infinitesimal bubble wall is surrounded by a finite shell of fluid kinetic energy. Even after the bubbles have merged, these shells can continue to propagate as sound waves, and their collisions source gravitational waves. This acoustic stage lasts much longer than the initial stage of bubble collisions and results in a sound wave contribution $\Omega_{\rm sw}$.

- **Turbulent phase:** During the final turbulent phase, non-linearities in the fluid equations become important as the plasma develops vorticity, turbulence, and potentially shocks. The gravitational wave contribution from this source $\Omega_{\rm turb}$ is the most difficult

to estimate and depends on the amount of energy remaining in the bulk fluid motion at the conclusion of the acoustic stage.

In particular, the scalar shear stress $\Pi_{ij}^{\phi} = \partial_i\phi\partial_j\phi$ acts as a source term for the scalar contribution Ω_ϕ while the fluid shear stress $\Pi_{ij}^{f} = \gamma^2(e_f + p_f)v_i v_j$ sources the acoustic and turbulent contributions $\Omega_{\rm sw}$ and $\Omega_{\rm turb}$. The total signal is the sum of these terms[4]

$$\Omega_{\rm GW} = \Omega_\phi + \Omega_{\rm sw} + \Omega_{\rm turb}\,. \tag{4.54}$$

The relative importance of each of these terms will depend on the details of the transition. This is in turn largely determined by the following four quantities: the nucleation temperature T_n, the phase transition strength parameter α_n, the phase transition inverse duration β, and the bubble wall velocity v_w. We comment on each of these in turn.

The Nucleation Temperature

The nucleation temperature T_n indicates the start of the transition. It is defined as the temperature at which on average one bubble has nucleated in each Hubble volume,

$$N(T_n) = 1\,, \tag{4.55}$$

where the number of bubbles nucleated in a Hubble volume is given by

$$N(T) = \int_{T_c}^{T} dT' \,\frac{\Gamma(T')h(T')}{T'H^4(T')}\,. \tag{4.56}$$

4. Determining the gravitational wave signals from each source is decidedly involved and requires numerical treatment outside of the scope of this thesis. We recommend [247] for a comprehensive overview.

In this expression, T_c is the critical temperature, $h(T)$ is the false vacuum fraction of Eq. (4.50), and Γ is the nucleation rate

$$\Gamma(T) \simeq T^4 \left(\frac{S_3}{2\pi T}\right)^{3/2} e^{-S_3/T}\,, \tag{4.57}$$

where S_3 the 3-dimensional action of a critical bubble [250]. Note that T_n is slightly lower than the critical temperature T_c, since significant nucleation requires some time to pass, and it is slightly higher than the percolation temperature T_p, defined as the temperature at which one first has an infinite cluster of mutually connected bubbles,

$$T_p < T_n < T_c\,. \tag{4.58}$$

The Strength Parameter

There exist various definitions in the literature for the phase transition strength α, and while some of these are equivalent, others can be related in nontrivial, model-dependent ways. Refs. [251, 252, 246] and many others define the phase transition strength parameter in terms of the enthalpy density w and the difference in the trace anomaly $\Theta = \eta_{\mu\nu}T^{\mu\nu} = \frac{1}{4}(e - 3p)$ between the symmetric and broken phases

$$\alpha_n = \frac{4}{3}\frac{\Theta_+(T_n) - \Theta_-(T_n)}{w_+(T_n)}\,. \tag{4.59}$$

We[5] include the subscript "n" to emphasize that all quantities are evaluated at the nucleation temperature T_n. These authors argue that this is the most precise definition since the potential energy of the scalar field is captured by the trace anomaly, which is proportional to the trace of the energy momentum tensor. The trace anomaly difference can be expressed

5. Elsewhere in the literature, this definition is denoted α_Θ.

in terms of the effective potential $V_{\rm eff}$ as

$$\Theta_+(T) - \Theta_-(T) = \Delta V_{\rm eff} - \frac{T}{4}\frac{\partial \Delta V_{\rm eff}}{\partial T}. \tag{4.60}$$

Ref. [253] advocates for defining the strength analogously to Eq. (4.59) but with the trace replaced by the so-called *pseudotrace* $\bar{\Theta} = e - p/c_s^2$

$$\bar{\alpha}_n = \frac{\bar{\Theta}_+(T_n) - \bar{\Theta}_-(T_n)}{3w_+(T_n)}\,. \tag{4.61}$$

If we set $c_s^2 = 1/3$, as for a relativistic plasma, this reduces to Eq. (4.59)

$$\bar{\alpha}_n \xrightarrow{c_s^2=1/3} \frac{4}{3}\frac{\Theta_+(T_n) - \Theta_-(T_n)}{w_+(T_n)} = \alpha_n\,. \tag{4.62}$$

Ref. [245] defines the phase transition strength parameter in terms of the radiation density $\rho_{\rm rad} = \frac{\pi^2}{30} g_\star T^4$ and the difference in the trace anomaly

$$\alpha_p = \frac{\Theta_+(T_p) - \Theta_-(T_p)}{\rho_{\rm rad}(T_p)}\,, \tag{4.63}$$

with all quantities evaluated at the percolation temperature T_p. For a completely relativistic plasma, the equation of state relates the pressure and energy density $e = \rho_{\rm rad}$ as $p = \frac{1}{3}e$, which in turn means the enthalpy is $w = \frac{4}{3}e$. With this replacement,

$$\alpha_p \xrightarrow{c_s^2=1/3} \frac{4}{3}\frac{\Theta_+(T_p) - \Theta_-(T_p)}{w_+(T_p)}\,, \tag{4.64}$$

and the phase transition strength matches Eq. (4.59) save for the fact that the quantities are evaluated at the percolation temperature T_p rather than the nucleation temperature T_n. Finally, Ref. [254] and others define the strength in terms of the latent heat $\mathcal{L}(T)$ released

by the transition[6],

$$\alpha_{\mathcal{L}} = \frac{\mathcal{L}(T_n)}{\rho_{\rm rad}(T_n)}\,, \tag{4.65}$$

with $\mathcal{L}(T)$ defined as the difference between the enthalpies of the symmetric and broken phases

$$\mathcal{L}(T) = w_+(T) - w_-(T)\,. \tag{4.66}$$

Note that this is *not* equivalent to the definition in terms of the trace anomaly. The latent heat *is* proportional to the trace anomaly when evaluated at the critical temperature T_c, $\Theta_+(T_c) - \Theta_-(T_c) = \mathcal{L}(T_c)/4$. This occurs because $\Delta p = V_{\rm eff}^+ - V_{\rm eff}^- = 0$ at T_c by definition. The strength parameter of Ref. [254] is evaluated at T_n, however, and so these definitions are not equivalent, as they themselves acknowledge.

The Inverse Duration

The inverse duration of the phase transition, also known as the transition rate parameter β, is defined as

$$\beta = \frac{d}{dt} \ln\left(\frac{\Gamma(t)}{\mathcal{V}}\right)\Big|_{t=t_f}\,, \tag{4.67}$$

where t_f is defined by the condition $h(t_f) = 1/e$ and corresponds to the time at which $\sim 64\%$ of the universe has been converted to the true vacuum.

The Wall Velocity

The bubble wall velocity v_w is probably the most difficult thermal parameter to determine, though recent progress has been made in e.g. [256, 257]. Because of this, it is often either arbitrarily fixed or left as a free parameter in studies of the gravitational wave signal from

6. The convention of defining α by latent heat seems to stem from this reference [255], which claimed that the false vacuum energy density ϵ is 1/4 of the latent heat. However, they did not define latent heat explicitly. Based on the bag equation of state used by this paper, the enthalpy density $w = e + p = \frac{4}{3}aT^4$ is independent of ϵ, hence is clearly not related to the latent heat.

first-order phase transitions. A proper determination of the bubble wall velocity is however *crucial* for accurate predictions, as it affects not just the energy budget for the transition but also the type of transition itself. Depending on the wall velocity v_w, the relative velocities of the plasma inside v_- and outside v_+ the bubble wall, and the sound speed for the plasma inside c_s^- and outside c_s^+, there are three classes of velocity profiles [251]:

- **Detonations:** This solution corresponds to a supersonic wall $v_w > c_s^+$. Outside the bubble, the fluid is at rest in the plasma frame $v_+ = 0$ and equal to the wall velocity in the wall frame $\bar{v}_+ = v_w$. The fluid velocity is greater in front of the wall than behind $|\bar{v}_+| > |\bar{v}_-|$. Note that the wall velocity must also be larger than the so-called *Jouguet velocity* v_J. The phase transition front it followed be a rarefaction wave.

- **Deflagrations:** In this case, the plasma is at rest immediately behind the wall $v_- = 0$, such that in the wall frame $\bar{v}_- = v_w$. The fluid velocity is then larger behind the wall than in front, $|\bar{v}_-| > |\bar{v}_+|$. The phase transition front it preceded by a shock front.

- **Hybrid:** Hybrid profiles, also called *supersonic deflagrations*, are a superposition of these two cases. As a result, the wall velocity is not identified with either $\bar{v}_+$ or $\bar{v}_-$, and the phase transition front is both followed by a rarefaction wave and preceded by a shockfront.

4.2 Thermal Resummation

The computation of the thermal phase transition parameters described in the previous section suffers from theoretical uncertainties, many of which stem from the computation of the finite temperature effective potential. The resummation of large thermal corrections to the effective potential is mandatory for the accurate prediction of phase transitions. In this section, we discuss the accuracy of different prescriptions to perform this resummation at the one- and two-loop level and point out resulting conceptual issues if using a high-temperature expansion

appearing at the two-loop level. Moreover, we show how the partial dressing prescription consistently avoids these issues and work out a novel method to apply it to the case of mixing fields. Our approach significantly extends the range of applicability of the partial dressing prescription, making it suitable for phenomenological studies of beyond the Standard Model extensions of the Higgs sector. Note that the material in this section shares considerable overlap with material by this author soon to be submitted for publication.

4.2.1 Overview

After the discovery of the Higgs boson a little over ten years ago [258, 259], it is one of the main goals of the Large Hadron Collider (LHC) as well as potential future colliders [260, 261, 262] to work towards understanding the dynamics of electroweak (EW) symmetry breaking. Besides collider experiments, gravitational wave observatories [263, 264, 265, 266, 267, 268] will start to probe the EW phase transition within the next decades. This research could provide substantial insights into the thermal history of the Universe, which is modified in many extensions of the Standard Model (SM) of particle physics. This, in particular, includes extensions of the SM Higgs sector, which may lead to a first-order electroweak phase transition, which would enable the scenario of electroweak baryogenesis [269, 270, 271, 272] or to phenomena like vacuum trapping, inverse symmetry breaking, or EW symmetry non-restoration (EWSNR) [273, 7, 274, 275, 276, 277, 278, 279]. In addition to the electroweak sector, phase transitions can also appear outside the electroweak sector in many BSM scenarios, including hidden sectors [280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290] and high-scale models (e.g., grand-unified theories) [291, 292, 293, 294].

For all these different areas, precise theoretical predictions are crucial to fully exploit the available experimental data. In the context of phase transition, this in particular means to provide an accurate prediction for the effective potential at finite temperature using perturbation theory.

It is well known that for high temperatures large corrections occur which exacerbate the behaviour of the perturbative expansion. To resum these large corrections, various resummation schemes have been developed in the literature. These include schemes employing a diagrammatic expansion — most notably the Parwani [295] and Arnold-Espinosa [296] resummation schemes — as well as schemes which involve solving the gap equation — the full dressing (FD) and partial dressing (PD) procedures [297, 298, 299]. Moreover, the large thermal corrections can also be resummed in an effective field theory (EFT) framework reducing the spacetime dimensions from four to three. This approach is called dimensional reduction (DR) [300, 301, 302].

While the Parwani and Arnold-Espinosa schemes have the benefit of being easy to implement — explaining their widespread application in the literature —, they on the other hand suffer from several disadvantages. First, they only allow to resum the leading thermal corrections. In many scenarios, the resummation of subleading thermal corrections is, however, also important for an accurate prediction. Moreover, they intrinsically rely on the high-temperature expansion for calculating the thermal masses and the thermal counterterms (see detailed discussion below). This implies that their accuracy is questionable in the regime where the temperature is close to the order of the relevant masses, which is exactly the interesting region for phase transitions (see [298] for a more detailed discussion).

Similar issues appear in the DR approach. While DR offers a conceptionally well-defined and systematic way to resum large thermal corrections, it is technically challenging (for recent steps towards automation of the required calculation, see [303]). Moreover, as an EFT approach, it intrinsically relies on the separation of the scales. This means that it is difficult to go beyond the high-temperature expansion and that each EFT is only applicable to a pre-defined hierarchy of scales. This makes DR in particular unsuited for parameter scans in BSM models for which many different hierarchies of the masses and the temperature occur — requiring to work a different EFT for each hierarchy of scales.

Partial dressing promises to resolve the issues of the FD and DR approaches. It includes subleading thermal corrections and can be easily applied beyond the high-temperature expansion making it particularly suited for studying phase transitions. It has been applied to the singlet-extended SM [298]. Its accuracy and perturbative convergence have been studied recently in [299].

In the first part of the present work, we present a detailed comparison of partial dressing with the Arnold-Espinosa and Parwani resummation schemes. Afterwards, we discuss the application of PD to the phenomenon of EWSNR. EWSNR occurs if the thermal corrections to a particle mass dominate over the tree-level mass turning the overall thermal mass squared negative. Thus, thermal corrections are by definition large and large differences between the FD schemes have been found in the literature [276, 277]. In this paper, we extend this comparison to the two-loop level finding unphysical predictions originating from large imaginary contributions to the effective potential. We demonstrate that these issues do not occur if PD is used leading to a more reliable prediction.

In the last part of this paper, we discuss the case of mixing scalar fields. So far, the PD approach is restricted to scenarios in which only one scalar takes a non-zero value effectively forbidding the description of models with mixing scalars. This strongly limits the applicability of PD for phenomenological studies. We demonstrate a new method to extend PD for scenarios with more than one non-zero scalar field.

This section is structured as follows. In 4.2.2, we review the need for thermal resummation. Then, we discuss resummation in non-mixing one- and two-field models in 4.2.3. In 4.2.4, we compare the FD and PD models for a toy model for EWSNR. In 4.2.5, we demonstrate the application of PD to models with mixing scalar fields. We provide conclusions in 4.2.6. Finally we present some supplementary material in 4.2.7.

4.2.2 Perturbative Breakdown & Thermal Resummation

Bosonic field theories at finite temperature suffer from various issues in the infrared, principal among them is that the usual perturbation expansion breaks down. To demonstrate how this breakdown comes about, consider a simple ϕ^4 theory with quartic self-coupling λ. Working in the imaginary-time formalism [304], the 1-loop correction to the bosonic propagator corresponds to the expression

$$\bigcirc = \frac{\lambda}{2}\mathcal{I}[m] \equiv \frac{\lambda}{2}\,T\sum_{\omega_n}\int\frac{d^3k}{(2\pi)^3}\frac{1}{K^2+m^2}\,, \tag{4.68}$$

where $K=(\omega_n,\vec{k})$ is the Euclidean four-momentum and $\omega_n = 2\pi nT$ is the bosonic Matsubara frequency. There are two useful ways to decompose this expression. One option would be to split $\mathcal{I}$ into a zero-temperature piece $\mathcal{I}_0[m]$ and a finite-temperature piece $\mathcal{I}_T[m]$ [305],

$$\mathcal{I}[m] = \underbrace{\int\frac{d^4k}{(2\pi)^4}\frac{1}{k^2+m^2}}_{\mathcal{I}_0[m]} + \underbrace{\int\frac{d^3k}{(2\pi)^3}\frac{1}{E_k}\frac{1}{e^{E_k/T}-1}}_{\mathcal{I}_T[m]}\,, \tag{4.69}$$

where $E_k^2 = \vec{k}^{\,2}+m^2$. The zero-temperature $\mathcal{I}_0[m]$ is UV-divergent, and we choose to regularize it using dimensional regularization[7] and the $\overline{\text{MS}}$-scheme with renormalization scale μ_R. Meanwhile the finite-temperature piece $\mathcal{I}_T[m]$ is UV-finite, but sensitive to the IR. A more convenient decomposition to reveal this would be to split the Matsubara sum appearing in $\mathcal{I}[m]$ into a "soft" zero-mode piece with $\omega_n = 0$ and a "hard" non-zero mode piece with $\omega_n \neq 0$,

$$\mathcal{I}[m] = \underbrace{T\int\frac{d^3k}{(2\pi)^3}\frac{1}{\vec{k}^{\,2}+m^2}}_{\mathcal{I}_{\text{soft}}[m]} + \underbrace{T\sum_{n\neq 0}\int\frac{d^3k}{(2\pi)^3}\frac{1}{\omega_n^2+\vec{k}^{\,2}+m^2}}_{\mathcal{I}_{\text{hard}}[m]}\,. \tag{4.70}$$

7. Note that for convenience we display all equations with $\epsilon \to 0$, such that $D = 4-2\epsilon \to 4$.

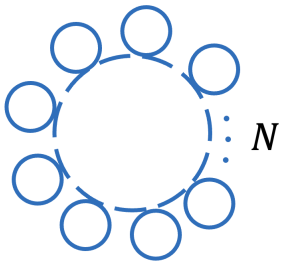

Figure 4.1: A daisy diagram featuring a zero-mode inner loop (dashed) surrounded by N hard external loops (solid).

Working in the high-temperature limit $m/T \sim \lambda \lesssim 1$, the zero-mode contribution can be evaluated explicitly as [305]

$$\mathcal{I}_{\text{soft}}[m] = -\frac{1}{4\pi} mT \,, \tag{4.71}$$

while the hard modes give a contribution

$$\mathcal{I}_{\text{hard}}[m] = \frac{T^2}{12} - \frac{m^2}{16\pi^2}\left(\frac{1}{\epsilon} + \ln\left(\frac{\mu_R^2 e^{2\gamma_E}}{16\pi^2 T^2}\right)\right) + \frac{\zeta(3)}{128\pi^4}\frac{m^4}{T^2} + \mathcal{O}\left(\frac{m^6}{T^4}\right) . \tag{4.72}$$

Thus we see that the correction to the zero-mode mass scales like $\delta m^2_{\text{soft}} \sim mT \sim \lambda T^2$ while the mass correction for the hard modes scale like $\delta m^2_{\text{hard}} \sim T^2$ in the high-T limit. Since the latter is parametrically larger, excitations of non-zero modes in the thermal plasma will act to screen the zero mode.

The IR problem manifests when considering higher loop "daisy diagrams", like that shown in Fig. 4.1. This diagram features a zero-mode inner loop (dashed line) surrounded by N hard outer loops (solid lines). Ignoring the overall symmetry factor, the contribution to the effective potential coming from such a diagram is

$$V_N^{\text{daisy}} \sim \left(T\int \frac{d^3k}{(2\pi)^3}\frac{1}{(\vec{k}^2+m^2)^N}\right)\left(\lambda T\sum_{n\neq 0}\int \frac{d^3k}{(2\pi)^3}\frac{1}{\omega_n^2+\vec{k}^2+m^2}\right)^N , \tag{4.73}$$

where the quantity in the first parenthesis comes from the N soft propagators in the inner

loop and scales as $m^{3-2N}T$ while that in the second parenthesis comes from the N hard external loops and scales as $\lambda^N T^{2N}$. The result is

$$V_N^{\text{daisy}} \sim \left(m^{3-2N}T\right)\left(\lambda^N T^{2N}\right) = m^3 T \left(\frac{\lambda T^2}{m^2}\right)^N . \tag{4.74}$$

There are two potential issues here. First comparing with the contribution from an $(N+1)$-loop daisy diagram, we see that each new hard thermal loop comes at a cost

$$\alpha \equiv \frac{V_{N+1}^{\text{daisy}}}{V_N^{\text{daisy}}} = \frac{\lambda T^2}{m^2} . \tag{4.75}$$

The issue is that this expansion parameter α is not parametrically small at all times. In particular when the system exhibits a phase transition, the expansion parameter becomes $\mathcal{O}(1)$ at the critical temperature, since here $m^2 \sim \lambda T^2$. This signals a breakdown of the perturbative expansion; diagrams which formally appear to be higher order may actually contribute with a magnitude equal to formally lower-order diagrams due to the contributions from these hard thermal loops. Intuitively, this perturbative breakdown occurs because at high temperatures, IR bosonic modes become highly occupied, leading to an enhanced expansion parameter [306].

The second issue occurs for fields with vanishing mass, for which daisy diagrams with $N \geq 2$ are IR divergent. Clearly this divergence is not physical, as the thermodynamic properties of a plasma of weakly interacting massless bosons (such as photons) are observed to be finite, and so these divergent contributions must cancel amongst one another when all terms in the expansion are summed. This is just another way to see that at finite temperature, the fixed-order perturbative expansion fails.

A natural solution[8], then, would be to reorganize the expansion using a new parameter

8. There are also alternative strategies to thermal resummation, such as dimensional reduction [307, 308], which we will not review here. Another option would be to just treat the problem non-perturbatively, using appropriate lattice techniques.

in terms of which the series is convergent — *thermal resummation*. Consider, for example, how one would resum the daisy diagrams of Fig. 4.1. Computing the contribution from an N-loop daisy diagram while more carefully keeping track of the combinatorial factors, we would find

$$V_N^{\text{daisy}} = -\frac{T}{12\pi}\frac{1}{N!}\left(\frac{\lambda T^2}{4}\right)^N \left(\frac{d}{dm^2}\right)^N m^3 \,. \tag{4.76}$$

One can check that this correctly reproduces the scaling of Eq. (4.74) by using the fact that $m^{3-2N} = \frac{4\sqrt{\pi}}{3}\frac{(-1)^N}{(N-1)!}\frac{\Gamma(N)}{\Gamma(N-3/2)}\left(\frac{d}{dm^2}\right)^N m^3$. If we now sum over all such diagrams, we find

$$\begin{aligned}\sum_{N=0}^{\infty} V_N^{\text{daisy}} &= -\frac{T}{12\pi}\sum_{N=0}^{\infty}\frac{1}{N!}\left(\frac{\lambda T^2}{4}\frac{d}{dm^2}\right)^N m^3 \\ &= -\frac{T}{12\pi}\exp\left(\frac{\lambda T^2}{4}\frac{d}{dm^2}\right) m^3 \,.\end{aligned} \tag{4.77}$$

Finally letting $x = m^2$ and noting that $\exp\left(c\frac{d}{dx}\right) f(x) = f(x+c)$, we find

$$\sum_{N=0}^{\infty} V_N^{\text{daisy}} = -\frac{T}{12\pi}\left(m^2 + \frac{\lambda}{4}T^2\right)^{3/2} \,. \tag{4.78}$$

This is a rather significant result; after summing all contributions, we find an expression where the limit $m^2 \to 0$ can meaningfully be taken without running into IR divergences. It is also significant that when $m^2 \to 0$, this correction comes in at $\mathcal{O}(\lambda^{3/2})$ rather than $\mathcal{O}(\lambda^2)$, as would naively be expected.

The quantity appearing in parenthesis is the thermally corrected mass

$$\overline{M}^2(\phi, T) \equiv m^2(\phi) + \frac{\lambda}{4}T^2 \,, \tag{4.79}$$

and to leading order, daisy resummation amounts to replacing instances of m^2 in the effective potential with the thermally corrected version $\overline{M}^2$. Of course, daisy diagrams are not the only problematic diagrams that appear in finite temperature field theory; there are also

so-called "super-daisy" diagrams as well as other sub-leading diagrams which demonstrate IR-sensitivity and so should be resummed. Given the questions of which class of diagrams to resum and how to re-order the expansion, there exist several different prescriptions for implementing thermal resummation.

Historically, the most popular methods employing a diagrammatic approach to thermal resummation are the Parwani [295] and Arnold Espinosa [296] schemes. The major conceptual difference between these methods is that in the Parwani prescription, all modes are resummed, while in the Arnold Espinosa prescription, only the problematic Matsubara zero-modes are resummed. On a technical level, this is equivalent to whether the thermally corrected mass $\overline{M}^2$ is substituted into all terms of the effective potential or only those non-analytic in m^2, which can be shown to correspond to zero-modes. Because the thermally corrected mass is included in different terms, when working to fixed order in perturbation theory these methods feature different higher-order terms, resulting in different degrees of convergence.

In addition to these diagrammatic approaches to thermal resummation, which in complicated theories quickly become impractical at higher-loop order, there is also a non-diagrammatic method which we will refer to as *gap resummation* [309, 310, 311]. Rather than computing the contributions from many higher loop diagrams, which quickly becomes impractical in complicated theories, in gap resummation one need simply compute the one-loop effective potential $V_{\text{eff}}^{(1)}$ and then solve the gap equation for the thermally corrected mass. This gap equation includes the dominant contributions from many[9] higher-order diagrams, in particular daisy diagrams to all orders in the effective potential. In a theory with $i = 1...n$ bosonic species $\{\phi_i\}$, the gap equation for the thermal mass of species i reads

$$M_i^2 = \frac{\partial^2}{\partial \phi_i^2} V_{\text{eff}}^{(1)}(M_j^2)\,, \tag{4.80}$$

9. Notably, contributions from superdaisy diagrams and other sub-leading diagrams like the bosonic sunset are not automatically included.

where masses appearing in the effective potential on the right-hand side are the thermally corrected masses for all species in the plasma M_j^2. Because the thermal mass appears on both sides of this equation, it must generally be solved numerically. For convenience, it is common to solve the truncated gap equation

$$M_i^2\Big|_{\text{trunc.}} = \frac{\partial^2}{\partial\phi_i^2} V_{\text{eff}}^{(1)}(m_j^2)\,, \tag{4.81}$$

where now the right hand side is evaluated on the field dependent effective masses m_j^2. When combined with the high temperature expansion, this truncated treatment only resums the leading order hard thermal loops, and results in a solution of the form $M_i^2|_{\text{trunc.}} \overset{\text{high-}T}{=} m_i^2 + cT^2 \equiv \overline{M}_i^2$, with c some constant function of the couplings.

After solving the gap equation, one usually proceeds by replacing the background field dependent masses $m_i^2(\phi_j)$ with the thermally corrected versions $M_i^2(\phi_j, T)$ in the effective potential $V_{\text{eff}}^{(1)}$. This prescription is called *full dressing* (FD), or *truncated full dressing* (TFD) if one uses the thermal mass obtained by solving the truncated gap equation, and diagrammatically it corresponds to dressing both the propagator and vertex in 1-loop tadpole diagrams.

Interestingly at the 1-loop level, the Arnold-Espinosa and Parwani prescriptions coincide with special cases of truncated gap resummation. As we will soon see, the 1-loop potential can be factorized into a zero-temperature Coleman-Weinberg (CW) piece V_{CW} and a finite temperature piece V_T. In the Parwani prescription, we replace $m_i^2 \to \overline{M}_i^2$ in both V_{CW} and V_T while in the Arnold-Espinosa prescription we replace $m_i^2 \to \overline{M}_i^2$ only in the non-analytic term appearing in the thermal piece V_T, corresponding to resumming only the Matsubara zero-modes. The former then coincides with TFD at 1-loop while the latter corresponds to a special case of TFD.

While the FD prescription has the obvious benefit of not needing to evaluate leading order diagrams analytically, it also faces several difficulties. Beginning at 2-loop order,

certain higher-order diagrams such as the sunset diagram are not automatically included and need to be added by hand. More concerningly, the FD prescription has been shown to miscount daisy and superdaisy diagrams starting at 2-loop order [312, 313]. An alternative procedure which reliably resums the dominant contributions to higher order is the *partial dressing* (PD) prescription, first introduced in [297] under the name of *tadpole resummation*. Rather than substituting $m_i^2 \to M_i^2$ directly in the effective potential, the PD prescription instructs us to perform the replacement in the first derivative of the effective potential V'_{eff}, which can then be integrated to obtain V_{eff} as

$$V_{\text{eff}} = \int d\phi \, \frac{\partial V_{\text{eff}}}{\partial \phi}\bigg|_{m^2 \to M^2}. \tag{4.82}$$

This scheme corresponds to dressing just the propagator and has been demonstrated by explicit calculation to 4-loop order to give the right counting of daisy and superdaisy diagrams [297]. A variant of PD resummation proposed in [298], *optimized partial dressing*, has been shown to yield an even better degree of convergence.

Despite its promise as a resummation candidate, the PD prescription is not without its challenges. PD omits a class of subleading diagrams starting at 2-loops. These are the lollipop diagrams (obtained from the vacuum sunset diagram by attaching one external leg to one of the vertices). Moreover, sunset-type tadpole diagrams are miscounted (obtained from the vacuum sunset diagram by attaching one external leg to one of the propagators). These issues can be resolved by adding the lollipop diagrams by hand and adjusting the gap equation to fix the miscounting of the sunset diagrams (see [313, 298] as well as **??**).

Another more pressing issue pointed out in [298] is that it is unclear how to implement PD in multi-field scenarios where field excursions can occur along more than one direction. Given that this is the case in a variety of beyond the Standard Model (BSM) extensions capable of yielding a strongly first-order electroweak phase transition, it is crucial that the formalism be extended to accommodate this situation. We propose a multi-field generalization of PD

resummation which can be applied in scenarios where the Higgs mixes with BSM scalars and then go on to compare the convergence of this scheme with that of other resummation techniques.

4.2.3 Resummation in ϕ^4 theory

Before getting into these technicalities, we will review the various resummation prescriptions in the context of a simple ϕ^4 theory. We begin by computing the effective potential at 1- and 2-loop order in the context of a single field ϕ^4 theory, to later generalize to the multi-field case (without mixing). Next, we resum the effective potential using the Parwani, Arnold-Espinosa, full dressing, and partial dressing schemes, respectively. Finally, we compare for a few benchmark points and comment on the differences.

Unresummed $V_{\text{eff}}(\phi)$ at 1- and 2-loops

We consider a single-field ϕ^4 theory with tree-level potential

$$V_0 = \frac{\mu^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4 \,. \tag{4.83}$$

At 1-loop, the effective potential receives radiative and finite temperature corrections captured by the sum-integral [314]

$$V_{\text{1-loop}} = \mathcal{J}[m] \equiv \frac{1}{2}\sum\!\!\!\!\!\!\int_K \ln\Big(K^2 + m^2\Big)\,, \tag{4.84}$$

where $K = (\omega_n, \vec{k})$ are Euclidean 4-momenta, $\omega_n = 2\pi n T$ are bosonic Matsubara modes, and the symbol $\sum\!\!\!\!\!\!\int$ is shorthand for Euclidean integration

$$\sum\!\!\!\!\!\!\int_K \equiv T\sum_{\omega_n}\int \frac{d^3k}{(2\pi)^3}\,. \tag{4.85}$$

The mass entering into this 1-loop correction is the field-dependent effective mass

$$m^2 = \frac{\partial^2 V_0}{\partial \phi^2} = \mu^2 + 3\lambda\phi^2 \,, \tag{4.86}$$

where ϕ here is understood to take its background field value. In many contexts, it is convenient to decompose the bosonic $\mathcal{J}$-function into a zero temperature Coleman-Weinberg piece[10]

$$V_{\rm CW} = \frac{m^4}{64\pi^2}\left(\ln\left(\frac{m^2}{\mu_R^2}\right) - \frac{3}{2}\right) \,, \tag{4.87}$$

and a finite temperature piece

$$V_T = \frac{T^4}{2\pi^2} J_B\left(\frac{m^2}{T^2}\right) \,, \quad \text{with} \quad J_B(y^2) = \int_0^\infty dx\, x^2 \ln\left(1 - e^{-\sqrt{x^2+y^2}}\right) \,, \tag{4.88}$$

such that $V_{\text{1-loop}} = V_{\rm CW} + V_T$. In the high- and low-temperature limits, $J_B(y^2)$ admits expansions given in Eqs. (4.179) and (4.180) of Appendix A. We will often have cause to work in the high-temperature regime, in which the full 1-loop correction reads:

$$V_{\text{1-loop}} = \mathcal{J}[m] \simeq \frac{1}{24} m^2 T^2 - \frac{1}{12\pi} m^3 T - \frac{L_R}{64\pi^2} m^4 + \mathcal{O}\left(\frac{m^6}{T^2}\right) \,, \tag{4.89}$$

where $L_R = \ln\left(\mu_R^2/T^2\right) + 2(\gamma_E - \ln 4\pi)$ and we have ignored the field-independent constant. Note that the logarithmic terms have cancelled between the Coleman-Weinberg and finite temperature contributions, such that $-\frac{1}{12\pi} m^3 T$ is the only term originating from the Matsubara zero-mode, as evidenced by the fact that it is non-analytic in m^2. Combining with the tree level piece, the effective potential at 1-loop $V_{\rm eff}^{(1)} = V_0 + V_{\text{1-loop}}$ in the high-temperature

10. We work in the $\overline{\rm MS}$ scheme with renormalization scale μ_R. Note that this prescription at finite temperature does not eliminate all factors of 4π and γ_E, as it does at zero temperature.

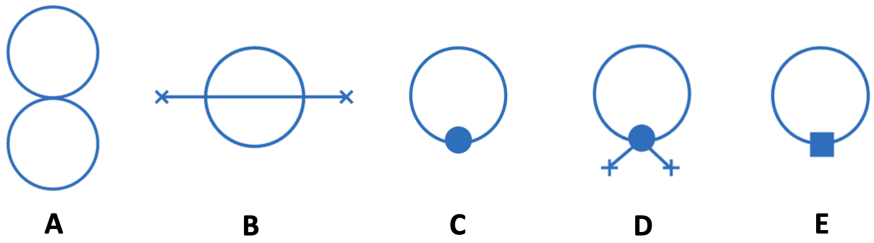

Figure 4.2: 2-loop corrections to the effective potential in ϕ^4 theory, including (A) the figure-8 diagram, (B) the sunset diagram, (C) the 1-loop mass counterterm diagram, and (D) the 1-loop vertex counterterm diagram. Solid circles denote 1-loop counterterms and x's indicate explicit field insertions. Diagram (E) should only be added to the effective potential upon performing thermal resummation in the Parwani scheme and features a 1-loop thermal "counterterm" (solid square).

expansion is

$$V_{\text{eff}}^{(1)} = \frac{1}{2}(\mu^2 + c_\phi T^2)\phi^2 + \frac{\lambda}{4}\phi^4 - \frac{1}{12\pi}m^3 T - \frac{L_R}{64\pi^2}m^4 + \mathcal{O}\left(\frac{m^6}{T^2}\right), \tag{4.90}$$

where we have defined $c_\phi = \lambda/4$.

At 2-loops, the corrections to the effective potential are summarized diagramatically in 4.2. The figure-8 diagram (A) corresponds to the contribution

$$V_{\text{2-loop}}^{A} = \frac{3\lambda}{4}\mathcal{I}[m]^2\,, \tag{4.91}$$

where

$$\mathcal{I}[m] \equiv \sum\!\!\!\!\!\!\int_K \frac{1}{K^2 + m^2}\,. \tag{4.92}$$

This sum-integral is related to the $\mathcal{J}$ of Eq. (4.84) as $\mathcal{I} = m^{-1}\frac{d\mathcal{J}}{dm}$, and so admits the high-temperature expansion given in Eq. (4.182) of Appendix A. Using this expansion, $V_{\text{2-loop}}^{A}$ explicitly evaluates to

$$V_{\text{2-loop}}^{A} = -\frac{\lambda}{32\pi}mT^3 + \frac{(3 - L_R)\lambda}{64\pi^2}m^2T^2 + \frac{3L_R\lambda}{128\pi^3}m^3T + \frac{(\zeta(3) + 3L_R^2)\lambda}{1024\pi^4}m^4\,. \tag{4.93}$$

The sunset diagram (B) corresponds to the contribution

$$V^B_{\text{2-loop}} = -3\lambda^2\phi^2\mathcal{H}[m,m,m]\,, \tag{4.94}$$

where, for three arbitrary masses m_1, m_2, m_3

$$\mathcal{H}[m_1,m_2,m_3] = \sum\!\!\!\!\!\!\int_P \sum\!\!\!\!\!\!\int_Q \frac{1}{(P^2+m_1^2)(Q^2+m_2^2)((P+Q)^2+m_3^2)}\,. \tag{4.95}$$

Using the high-temperature expansion of $\mathcal{H}$ in Eq. (4.190), the contribution $V^B_{\text{2-loop}}$ is explicitly

$$\begin{aligned} V^B_{\text{2-loop}} = &-\frac{3\lambda^2}{32\pi^2}\left(\ln\left(\frac{\mu_R^2}{m^2}\right) - 2\ln 3 + 1\right)T^2\phi^2 \\ &+ \frac{9\lambda^2}{64\pi^3}\left(\ln\left(\frac{\mu_R^2}{m^2}\right) + L_R - 2\ln 2 + 2\right)mT\phi^2 \\ &+ \frac{9\lambda^2}{256\pi^4}\left(L_R^2 + L_R - 2\gamma_E^2 - 4\gamma_1 + \frac{\pi^2}{4} + \frac{3}{2}\right)m^2\phi^2\,. \end{aligned} \tag{4.96}$$

Next we have diagrams (C) and (D), which feature 1-loop mass and vertex counterterms[11], respectively. The former corresponds to an expression of the form

$$V^C_{\text{2-loop}} = \frac{3\lambda}{32\pi^2}\mu^2\mathcal{I}[m]\frac{1}{\epsilon}\,. \tag{4.97}$$

The leading order field-dependent contribution first comes in at $\mathcal{O}(\lambda^{5/2}T^4)$, which is higher than the order to which we work, and so can be neglected. The vertex counterterm diagram (D) corresponds to

$$V^D_{\text{2-loop}} = \frac{27\lambda^2}{32\pi^2}\phi^2\mathcal{I}[m]\frac{1}{\epsilon}\,. \tag{4.98}$$

11. Note that due to the $\mathcal{O}(\epsilon)$ pieces in $\mathcal{I}[m]$, not only do these diagrams feature divergent $\mathcal{O}(1/\epsilon)$ pieces that cancel against the divergences in the 2-loop diagrams, but they can also give a finite contribution to the effective potential.

The leading order finite contribution comes from the $\mathcal{O}(\epsilon)$ piece of $\mathcal{I}[m] \supset \frac{\epsilon}{12} L_R T^2$, leading to

$$V^D_{\text{2-loop}} = \frac{9 L_R \lambda^2}{128\pi^2} T^2 \phi^2 \,. \tag{4.99}$$

The 2-loop correction to the effective potential is the sum of Eqs. (4.93), (4.96), and (4.99),

$$\begin{aligned} V_{\text{2-loop}} = &-\frac{\lambda}{32\pi} m T^3 + \frac{(12 - 3L_R)\lambda}{256\pi^2} m^2 T^2 + \frac{3 L_R \lambda}{128\pi^3} m^3 T + \frac{(\zeta(3) + 3L_R^2)\lambda}{1024\pi^4} m^4 \\ &- \frac{3\lambda^2}{128\pi^2} \left(4 \ln\left(\frac{\mu_R^2}{m^2}\right) - 3L_R - 8\ln 3 + 4 \right) T^2 \phi^2 \\ &+ \frac{9\lambda^2}{64\pi^3} \left(\ln\left(\frac{\mu_R^2}{m^2}\right) + L_R - 2\ln 2 + 2 \right) m T \phi^2 \\ &+ \frac{9\lambda^2}{256\pi^4} \left(L_R^2 + L_R - 2\gamma_E^2 - 4\gamma_1 + \frac{\pi^2}{4} + \frac{3}{2} \right) m^2 \phi^2 \,. \end{aligned} \tag{4.100}$$

Combining with the 1-loop correction of Eq. (4.89) and the tree-level potential of Eq. (4.83) yields the 2-loop effective potential V_{eff} in the high-temperature approximation.

Unresummed $V_{\text{eff}}(\phi_1, \phi_2)$ at 1- and 2-loops

Interesting finite-temperature effects can arise upon the addition of a second scalar field. Consider the tree-level potential

$$V_0 = \frac{\mu_1^2}{2}\phi_1^2 + \frac{\mu_2^2}{2}\phi_2^2 + \frac{\lambda_1}{4}\phi_1^4 + \frac{\lambda_2}{4}\phi_2^4 + \frac{\lambda_{12}}{4}\phi_1^2\phi_2^2 \,. \tag{4.101}$$

For now, we will restrict to the case where only one of the fields develops a vacuum expectation value, such that we need not worry about mixing. In this case there will be no off-diagonal terms in the field-dependent mass matrix, and the effective masses for ϕ_1 and ϕ_2 are

$$m_1^2 = \mu_1^2 + 3\lambda_1 \phi_1^2 + \frac{\lambda_{12}}{2}\phi_2^2 \,, \tag{4.102a}$$

$$m_2^2 = \mu_2^2 + 3\lambda_2\phi_2^2 + \frac{\lambda_{12}}{2}\phi_1^2\,. \tag{4.102b}$$

The 1-loop contribution to the effective potential is

$$V_{\text{1-loop}} = \mathcal{J}[m_1] + \mathcal{J}[m_2]\,, \tag{4.103}$$

with the bosonic $\mathcal{J}$-function defined in Eq. (4.84). In the high-temperature approximation, the effective potential at 1-loop is then

$$\begin{aligned} V_{\text{eff}}^{(1)} =& \frac{1}{2}(\mu_1^2 + c_1T^2)\phi_1^2 + \frac{1}{2}(\mu_2^2 + c_2T^2)\phi_2^2 + \frac{\lambda_1}{4}\phi_1^4 + \frac{\lambda_2}{4}\phi_2^4 + \frac{\lambda_{12}}{4}\phi_1^2\phi_2^2 \\ & - \frac{1}{12\pi}(m_1^3 + m_2^3)T - \frac{L_R}{64\pi^2}(m_1^4 + m_2^4)\,, \end{aligned} \tag{4.104}$$

where we have defined the thermal mass parameters

$$c_1 = \frac{1}{24}(6\lambda_1 + \lambda_{12})\,,\;\; c_2 = \frac{1}{24}(6\lambda_2 + \lambda_{12})\,. \tag{4.105}$$

When the mixed quartic coupling is negative, $\lambda_{12} < 0$, one[12] of these thermal mass parameters c_i can be negative. Then at high temperatures when the thermal contribution dominates the bare mass, one of the thermal masses

$$\overline{M}_i^2 = m_i^2 + c_iT^2 \tag{4.106}$$

may become negative — indicative of spontaneous symmetry breaking. Occasionally in the literature (see e.g. [7]), the thermal masses are modified by inserting an additional Boltzmann factor,

$$\overline{M}_{\text{Boltzmann},i}^2 = m_i^2 + c_iT^2e^{-m_i/T}\,, \tag{4.107}$$

12. The boundedness-from-below condition on λ_1 and λ_2 prevents both c_1 and c_2 from being negative.

which allows for a better approximation of the full thermal loop function for low temperatures.

The 2-loop contributions to the effective potential are summarized diagrammatically in 4.3. Now in addition to figure-8, sunset, and counterterm diagrams for each species, we

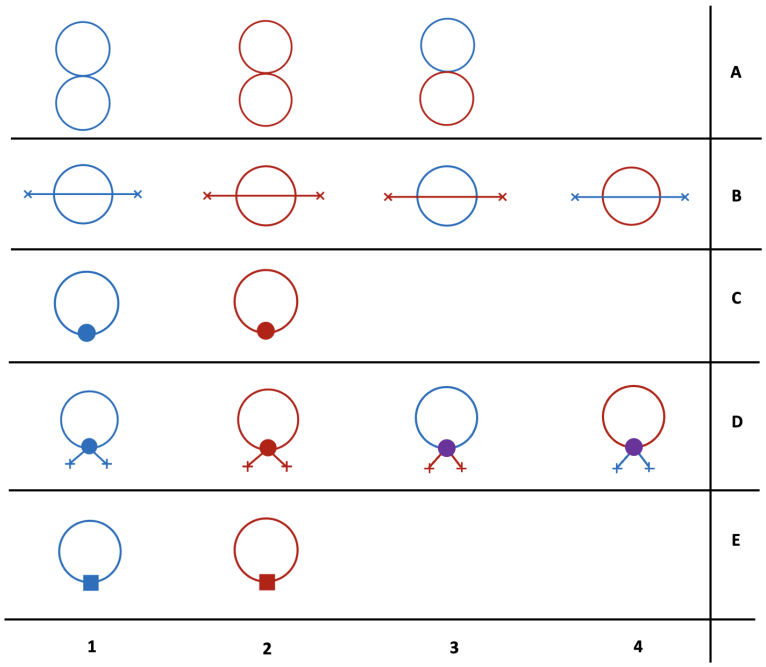

Figure 4.3: 2-loop corrections to the effective potential, including (A) figure-8 diagrams, (B) the sunset diagrams, (C) 1-loop mass counterterm diagrams, and (D) 1-loop vertex counterterm diagrams. Blue propagators correspond to ϕ_1 while red correspond to ϕ_2. Solid circles indicate 1-loop counterterms and x's indicate explicit field insertions. In particular, the blue, red, and purple circles of row (D) are λ_1, λ_2, and λ_{12} vertex counterterms, respectively. The 1-loop thermal "counterterm" (solid square) diagrams of row (E) should be included upon performing thermal resummation in the Parwani scheme.

must also consider mixed diagrams. The first row of figure-8 diagrams corresponds to the expressions

$$V_{\text{2-loop}}^{A1} = \frac{3\lambda_1}{4}\mathcal{I}[m_1]^2\,, \tag{4.108a}$$

$$V_{\text{2-loop}}^{A2} = \frac{3\lambda_2}{4}\mathcal{I}[m_2]^2\,, \tag{4.108b}$$

$$V_{\text{2-loop}}^{A3} = \frac{\lambda_{12}}{4}\mathcal{I}[m_1]\,\mathcal{I}[m_2]\,, \tag{4.108c}$$

which, using the high-temperature expansion of $\mathcal{I}$ in Eq. (4.182), may be written explicitly in this limit as

$$V_{\text{2-loop}}^{A1} = \frac{\lambda_1}{32\pi}\bigg[- m_1 T^3 + \frac{(6 - L_R)}{4\pi} m_1^2 T^2 + \frac{3L_R}{4\pi^2} m_1^3 T + \frac{(\zeta(3) + 3L_R^2)}{32\pi^3} m_1^4 \bigg], \tag{4.109a}$$

$$V_{\text{2-loop}}^{A2} = \frac{\lambda_2}{32\pi}\bigg[- m_2 T^3 + \frac{(6 - L_R)}{4\pi} m_2^2 T^2 + \frac{3L_R}{4\pi^2} m_2^3 T + \frac{(\zeta(3) + 3L_R^2)}{32\pi^3} m_2^4 \bigg], \tag{4.109b}$$

$$\begin{aligned} V_{\text{2-loop}}^{A3} = \frac{\lambda_{12}}{64\pi^2}\bigg[&- \frac{\pi}{3}(m_1 + m_2) T^3 + m_1 m_2 T^2 - \frac{L_R}{12}(m_1^2 + m_2^2) T^2 \\ &+ \frac{L_R}{4\pi}(m_1 + m_2) m_1 m_2 T + \frac{\zeta(3)}{96\pi^2}(m_1^4 + m_2^4) + \frac{L_R^2}{16\pi^2} m_1^2 m_2^2 \bigg]. \end{aligned} \tag{4.109c}$$

The second row of sunset diagrams corresponds to expressions

$$V_{\text{2-loop}}^{B1} = -3\lambda_1^2 \phi_1^2 \mathcal{H}[m_1, m_1, m_1], \tag{4.110a}$$

$$V_{\text{2-loop}}^{B2} = -3\lambda_2^2 \phi_2^2 \mathcal{H}[m_2, m_2, m_2], \tag{4.110b}$$

$$V_{\text{2-loop}}^{B3} = -\frac{\lambda_{12}^2}{4} \phi_2^2 \mathcal{H}[m_1, m_1, m_2], \tag{4.110c}$$

$$V_{\text{2-loop}}^{B4} = -\frac{\lambda_{12}^2}{4} \phi_1^2 \mathcal{H}[m_1, m_2, m_2]. \tag{4.110d}$$

Note that since these expressions feature external ϕ_i insertions, they vanish unless ϕ_i has non-zero vacuum expectation value. Suppose we take ϕ_1 to be the field which develops a vacuum expectation value; then $V_{\text{2-loop}}^{B2} = V_{\text{2-loop}}^{B3} = 0$. Using the high-temperature expansion of $\mathcal{H}$

in Eq. (4.190), the remaining $V^{B1}_{\text{2-loop}}$ and $V^{B4}_{\text{2-loop}}$ are explicitly

$$\begin{aligned} V^{B1}_{\text{2-loop}} = &- \frac{3\lambda_1^2}{32\pi^2}\left(\ln\left(\frac{\mu_R^2}{9m_1^2}\right)+1\right)\phi_1^2T^2 \\ &+ \frac{9\lambda_1^2}{64\pi^3}\left(\ln\left(\frac{\mu_R^2}{m_1^2}\right)+L_R-2\ln 2+2\right)m_1\phi_1^2T \\ &+ \frac{9\lambda_1^2}{256\pi^4}\left(L_R^2+L_R-2\gamma_E^2-4\gamma_1+\frac{\pi^2}{4}+\frac{3}{2}\right)m_1^2\phi_1^2\,, \end{aligned} \tag{4.111a}$$

$$\begin{aligned} V^{B4}_{\text{2-loop}} = &- \frac{\lambda_{12}^2}{64\pi^2}\left(\ln\left(\frac{\mu_R}{m_1+2m_2}\right)+\frac{1}{2}\right)\phi_1^2T^2 \\ &+ \frac{\lambda_{12}^2}{256\pi^3}\left(\ln\left(\frac{\mu_R^2}{4m_1^2}\right)+L_R+2\right)m_1\phi_1^2T \\ &+ \frac{\lambda_{12}^2}{128\pi^3}\left(\ln\left(\frac{\mu_R^2}{4m_2^2}\right)+L_R+2\right)m_2\phi_1^2T \\ &+ \frac{\lambda_{12}^2}{1024\pi^4}\left(L_R^2+L_R-2\gamma_E^2-4\gamma_1+\frac{\pi^2}{4}+\frac{3}{2}\right)(m_1^2+2m_2^2)\phi_1^2\,. \end{aligned} \tag{4.111b}$$

As in the single ϕ^4 case, the leading order, field-dependent contribution from the diagrams of row (C) are higher order than that to which we work, and so can be neglected. Of the diagrams of row (D), corresponding to expressions

$$V^{D1}_{\text{2-loop}} = \frac{3}{128\pi^2}(36\lambda_1^2+\lambda_{12}^2)\phi_1^2\,\mathcal{I}[m_1]\frac{1}{\epsilon}\,, \tag{4.112a}$$

$$V^{D2}_{\text{2-loop}} = \frac{3}{128\pi^2}(36\lambda_2^2+\lambda_{12}^2)\phi_2^2\,\mathcal{I}[m_2]\frac{1}{\epsilon}\,, \tag{4.112b}$$

$$V^{D3}_{\text{2-loop}} = \frac{1}{64\pi^2}\left(3(\lambda_1+\lambda_2)\lambda_{12}+2\lambda_{12}^2\right)\phi_2^2\,\mathcal{I}[m_1]\frac{1}{\epsilon}\,, \tag{4.112c}$$

$$V^{D4}_{\text{2-loop}} = \frac{1}{64\pi^2}\left(3(\lambda_1+\lambda_2)\lambda_{12}+2\lambda_{12}^2\right)\phi_1^2\,\mathcal{I}[m_2]\frac{1}{\epsilon}\,, \tag{4.112d}$$

we need only consider $V^{D1}_{\text{2-loop}}$ and $V^{D4}_{\text{2-loop}}$, since ϕ_2's vanishing background value means $V^{D2}_{\text{2-loop}} = V^{D3}_{\text{2-loop}} = 0$. The explicit contributions from the non-vanishing diagrams in the

high-temperature approximation are

$$V_{\text{2-loop}}^{D1} = \frac{L_R}{512\pi^2}(36\lambda_1^2 + \lambda_{12}^2)\phi_1^2 T^2\,, \tag{4.113a}$$

$$V_{\text{2-loop}}^{D4} = \frac{L_R}{768\pi^2}(3(\lambda_1 + \lambda_2)\lambda_{12} + 2\lambda_{12}^2)\phi_1^2 T^2\,. \tag{4.113b}$$

Gathering all these terms, the 2-loop contribution to the effective potential is

$$V_{\text{2-loop}} = V_{\text{2-loop}}^{A1} + V_{\text{2-loop}}^{A2} + V_{\text{2-loop}}^{A3} + V_{\text{2-loop}}^{B1} + V_{\text{2-loop}}^{B4} + V_{\text{2-loop}}^{D1} + V_{\text{2-loop}}^{D4}\,, \tag{4.114}$$

and the effective potential at 2-loop order is

$$V_{\text{eff}}^{(2)} = V_{\text{eff}}^{(1)} + V_{\text{2-loop}}\,, \tag{4.115}$$

with $V_{\text{eff}}^{(1)}$ in Eq. (4.104).

Parwani

The Parwani prescription [295] is a diagramatic approach to resummation in which all modes are resummed. On a technical level, it simply amounts to replacing m_i^2 with the high-temperature expanded thermal mass $\overline{M}_i^2 = m_i^2 + c_i T^2$ everywhere and adding to the 2-loop effective potential the "thermal counterterm" diagrams shown in row E of 4.3. This has the effect of resumming the dominant parts of ring diagrams.

To see why, consider for the moment a simpler single field ϕ^4 theory with tree-level potential $V_0 = \frac{\mu^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4$. As discussed in Sec. 4.2.2, resumming daisy diagrams amounts to replacing the m^2 appearing in propagators with $\overline{M}^2$. In order to do this consistently, we can add and subtract the thermal contribution to the mass $c_\phi T^2 = \frac{\lambda}{4}T^2$ in a clever way,

such that the tree-level potential becomes

$$V_0 = \frac{1}{2}(\mu^2 + c_\phi T^2)\phi^2 + \frac{\lambda}{4}\phi^4 - \frac{c_\phi}{2}T^2\phi^2 \,. \tag{4.116}$$

This is equivalent to the original potential, but now the idea is to treat the first two terms as defining the unperturbed theory, and the last term as a perturbation — a "thermal counterterm". Now order-by-order, the $c_\phi T^2$ pieces of quadratically divergent sub-loops will cancel against new diagrams involving thermal counterterms, resulting in a new convergent loop expansion parameter $\lambda T/\overline{M}$ [296]. Returning to the full theory, we see that to implement Parwani resummation we should:

1. Replace all field dependent effective masses m_i^2 with leading order thermal masses $\overline{M}_i^2 = m_i^2 + c_i T^2$, with c_i given in Eq. (4.105).

2. Include thermal counterterm diagrams in calculating V_{eff}.

These thermal counterm diagrams do not enter until 2-loop order, so at 1-loop the Parwani-resummed effective potential is simply $V^{(1)}_{\text{eff,P}} = V^{(1)}_{\text{eff}}|_{m_i^2 \to M_i^2}$, or explicitly

$$\begin{aligned} V^{(1)}_{\text{eff,P}} =& \frac{1}{2}(\mu_1^2 + c_1 T^2)\phi_1^2 + \frac{1}{2}(\mu_2^2 + c_2 T^2)\phi_2^2 + \frac{\lambda_1}{4}\phi_1^4 + \frac{\lambda_2}{4}\phi_2^4 + \frac{\lambda_{12}}{4}\phi_1^2\phi_2^2 \\ & - \frac{1}{12\pi}(\overline{M}_1^3 + \overline{M}_2^3)T - \frac{L_R}{64\pi^2}(\overline{M}_1^4 + \overline{M}_2^4) \,. \end{aligned} \tag{4.117}$$

The 2-loop diagrams featuring "thermal counterterms" are shown in row E of Fig. 4.3 and correspond to expressions

$$V^{E1}_{\text{2-loop}} = -\frac{1}{2}c_1 T^2 \mathcal{I}[\overline{M}_1] \,, \tag{4.118a}$$

$$V^{E2}_{\text{2-loop}} = -\frac{1}{2}c_2 T^2 \mathcal{I}[\overline{M}_2] \,. \tag{4.118b}$$

Using the high-temperature expansion of $\mathcal{I}[m]$ in Eq. (4.182), these evaluate to

$$V^{E1}_{\text{2-loop}} = \left(\frac{\lambda_1}{4} + \frac{\lambda_{12}}{24}\right)\left[\frac{1}{8\pi}\overline{M}_1 T^3 + \frac{L_R}{32\pi^2}\overline{M}_1^2 T^2 - \frac{\zeta(3)}{256\pi^4}\overline{M}_1^4\right], \tag{4.119a}$$

$$V^{E2}_{\text{2-loop}} = \left(\frac{\lambda_2}{4} + \frac{\lambda_{12}}{24}\right)\left[\frac{1}{8\pi}\overline{M}_2 T^3 + \frac{L_R}{32\pi^2}\overline{M}_2^2 T^2 - \frac{\zeta(3)}{256\pi^4}\overline{M}_2^4\right]. \tag{4.119b}$$

We see that these expressions have pieces which exactly cancel the IR sensitive pieces coming from $V^{A}_{\text{2-loop}}|_{m_i^2 \to M_i^2}$,

$$V^{A}_{\text{2-loop}}|_{m_i^2 \to M_i^2} \supset -\frac{\lambda_1}{32\pi}\overline{M}_1 T^3 - \frac{\lambda_2}{32\pi}\overline{M}_2 T^2 - \frac{\lambda_{12}}{192\pi}(\overline{M}_1 + \overline{M}_2)T^3\,. \tag{4.120}$$

Upon adding these contributions to $V_{\text{2-loop}}$, as given in Eq. (4.114), and replacing $m_i^2 \to]\overline{M}_i^2$ everywhere in $V_{\text{2-loop}}$, we arrive at the Parwani-resummed 2-loop effective potential

$$V^{(2)}_{\text{eff,P}} = V^{(1)}_{\text{eff,P}} + V_{\text{2-loop}}(\overline{M}_i^2) + V^{E1}_{\text{2-loop}}(\overline{M}_i^2) + V^{E2}_{\text{2-loop}}(\overline{M}_i^2)\,. \tag{4.121}$$

As demonstrated in the discussion above, Parwani resummation intrinsically depends on the high-temperature expansion and also only resum the leading contributions in the high-temperature limit.

Arnold-Espinosa

Alternatively because only the Matsubara zero mode ω_0 demonstrates problematic behavior in the IR, we could just resum these "soft" modes whilst leaving the hard non-zero modes untouched. This is the basic premise behind the Arnold-Espinosa prescription [296], another popular diagrammatic approach to resummation. Because it requires splitting calculations into soft and hard modes, it is in principle a bit more cumbersome to implement. Consider for example the $\mathcal{J}$-function defined in Eq. (4.84). Previously we had taken the customary

approach of separating this into a zero-temperature Coleman-Weinberg piece and a finite temperature piece [311],

$$\mathcal{J}[m] = \underbrace{\frac{1}{2}\int \frac{d^4k}{(2\pi)^4}\ln\Big(k^2+m^2\Big)}_{\mathcal{J}_{\rm CW}[m]} - \underbrace{T\int \frac{d^3k}{(2\pi)^3}\ln\Big(1\mp n_{\rm B/F}(E_k,T)\Big)}_{\mathcal{J}_T[m]}\,, \tag{4.122}$$

where $n_{\rm B/F}(E_k,T) = 1/\left(e^{E_k/T}\mp 1\right)$. A more useful decomposition for the purposes of Arnold-Espinosa resummation would be to isolate the zero mode

$$\mathcal{J}[m] = \underbrace{\frac{T}{2}\int \frac{d^3k}{(2\pi)^3}\ln\Big(k^2+m^2\Big)}_{\mathcal{J}_{\rm soft}[m]} + \underbrace{\frac{T}{2}\sum_{n\neq 0}\int \frac{d^3k}{(2\pi)^3}\ln\Big(\omega_n^2+k^2+m^2\Big)}_{\mathcal{J}_{\rm hard}[m]}\,, \tag{4.123}$$

where $\omega_n = 2\pi nT$. Working in the high-temperature expansion, one can show that this first zero-mode piece evaluates to

$$\mathcal{J}_{\rm soft}[m] \simeq -\frac{1}{12\pi}m^3T\,. \tag{4.124}$$

Comparing against the full expression in Eq. (4.89), we note that the zero-mode contribution is just the term non-analytic in m^2. This is a more generic phenomenon; it will turn out to be the case that all terms non-analytic in m^2 contain zero-mode contributions. This makes implementing Arnold-Espinosa resummation surprisingly simple in practice when working in the high-temperature expansion, since the terms requiring resummation are readily identifiable. At 1-loop there is only one such term (per scalar field ϕ_i), and we resum it by replacing $m_i^3 \to \overline{M}_i^3$. The 1-loop effective potential in the Arnold-Espinosa scheme then reads

$$\begin{aligned} V^{(1)}_{\rm eff,AE} =& \frac{1}{2}(\mu_1^2+c_1T^2)\phi_1^2 + \frac{1}{2}(\mu_2^2+c_2T^2)\phi_2^2 + \frac{\lambda_1}{4}\phi_1^4 + \frac{\lambda_2}{4}\phi_2^4 + \frac{\lambda_{12}}{4}\phi_1^2\phi_2^2 \\ & - \frac{1}{12\pi}(\overline{M}_1^3+\overline{M}_2^3)T - \frac{L_R}{64\pi^2}(m_1^4+m_2^4)\,. \end{aligned} \tag{4.125}$$

This expression is similar to the Parwani-resummed $V^{(1)}_{\text{eff,P}}$ but differs in the higher order terms $\propto m^4$. Explicitly, the difference between the two is

$$V^{(1)}_{\text{eff,AE}} - V^{(1)}_{\text{eff,P}} = \frac{L_R}{64\pi^2}\left[(c_1^2+c_2^2)T^4 + 2(c_1 m_1^2 + c_2 m_2^2)T^2\right]. \tag{4.126}$$

Before moving on to the 2-loop effective potential, we note that often in the literature one speaks of resumming $V^{(1)}_{\text{eff}}$ by adding the daisy "ring improvement" term

$$V_{\text{daisy}} = -\frac{1}{12\pi}(\overline{M}^3 - m^3)T\,. \tag{4.127}$$

We see that the result is completely identical to that of the procedure described above.

At 2-loop order, there are many more terms with non-analytic m^2 dependence. It will be useful to define the Arnold-Espinosa resummed $\mathcal{I}$-function, whose high-temperature expansion reads

$$\mathcal{I}_{\text{AE}}[m] \simeq \frac{1}{12}T^2 - \frac{1}{4\pi}\overline{M}T - \frac{L_R}{16\pi^2}m^2 + \frac{\zeta(3)}{128\pi^4}\frac{m^4}{T^2}\,. \tag{4.128}$$

Then from Eqs. (4.108), the contributions to the resummed $V_{\text{2-loop}}$ coming from the figure-8 diagrams are

$$V^{A1}_{\text{2-loop,AE}} = \frac{\lambda_1}{32\pi}\left[-\overline{M}_1 T^3 - \frac{L_R}{4\pi}m_1^2T^2 + \frac{3}{2\pi}\overline{M}_1^2T^2 + \frac{3L_R}{4\pi^2}m_1^2\overline{M}_1T + \frac{(\zeta(3)+3L_R^2)}{32\pi^3}m_1^4\right], \tag{4.129a}$$

$$V^{A2}_{\text{2-loop,AE}} = \frac{\lambda_2}{32\pi}\left[-\overline{M}_2 T^3 - \frac{L_R}{4\pi}m_2^2T^2 + \frac{3}{2\pi}\overline{M}_2^2T^2 + \frac{3L_R}{4\pi^2}m_2^2\overline{M}_2T + \frac{(\zeta(3)+3L_R^2)}{32\pi^3}m_2^4\right], \tag{4.129b}$$

$$\begin{aligned} V^{A3}_{\text{2-loop,AE}} = \frac{\lambda_{12}}{64\pi^2}\Bigg[&-\frac{\pi}{3}(\overline{M}_1+\overline{M}_2)T + \overline{M}_1\overline{M}_2T^2 - \frac{L_R}{12}(m_1^2+m_2^2)T^2 \\ &+ \frac{L_R}{4\pi}(\overline{M}_1 m_2^2 + \overline{M}_2 m_1^2)T + \frac{L_R^2}{16\pi^2}m_1^2m_2^2 + \frac{\zeta(3)}{96\pi^2}(m_1^4+m_2^4)\Bigg]. \end{aligned} \tag{4.129c}$$

From Eq. (4.111), the contribution from the sunset diagrams are

$$\begin{aligned} V^{B1}_{\text{2-loop,AE}} =& -\frac{3\lambda_1^2}{32\pi^2}\left(\ln\left(\frac{\mu_R^2}{9\overline{M}_1^2}\right)+1\right)\phi_1^2T^2 \\ &+\frac{9\lambda_1^2}{64\pi^3}\left(\ln\left(\frac{\mu_R^2}{\overline{M}_1^2}\right)+L_R-2\ln 2+2\right)\overline{M}_1\phi_1^2T \\ &+\frac{9\lambda_1^2}{256\pi^4}\left(L_R^2+L_R-2\gamma_E^2-4\gamma_1+\frac{\pi^2}{4}+\frac{3}{2}\right)m_1^2\phi_1^2\,, \end{aligned} \tag{4.130a}$$

$$\begin{aligned} V^{B4}_{\text{2-loop,AE}} =& -\frac{\lambda_{12}^2}{64\pi^2}\left(\ln\left(\frac{\mu_R}{\overline{M}_1+2\overline{M}_2}\right)+\frac{1}{2}\right)\phi_1^2T^2 \\ &+\frac{\lambda_{12}^2}{256\pi^3}\left(\ln\left(\frac{\mu_R^2}{4\overline{M}_1^2}\right)+L_R+2\right)\overline{M}_1\phi_1^2T \\ &+\frac{\lambda_{12}^2}{128\pi^3}\left(\ln\left(\frac{\mu_R^2}{4\overline{M}_2^2}\right)+L_R+2\right)M_2\phi_1^2T \\ &+\frac{\lambda_{12}^2}{1024\pi^4}\left(L_R^2+L_R-2\gamma_E^2-4\gamma_1+\frac{\pi^2}{4}+\frac{3}{2}\right)(m_1^2+2m_2^2)\phi_1^2\,, \end{aligned} \tag{4.130b}$$

where again we are taking ϕ_1 as the field which develops a vacuum expectation value. Finally from Eq. (4.113), we see that the contributions from vertex counterterm diagrams are unaffected by the resummation. The 2-loop effective potential resummed in the Arnold-Espinosa scheme is then

$$V^{(2)}_{\text{eff,AE}} = V^{(1)}_{\text{eff,AE}}+V^{A1}_{\text{2-loop,AE}}+V^{A2}_{\text{2-loop,AE}}+V^{A3}_{\text{2-loop,AE}}+V^{B1}_{\text{2-loop,AE}}+V^{B4}_{\text{2-loop,AE}}+V^{D1}_{\text{2-loop}}+V^{D4}_{\text{2-loop}}\,. \tag{4.131}$$

As for Parwani resummation, the Arnold-Espinosa resummation scheme intrinsically depends on the high-temperature expansion and resums only the leading contributions in the high-temperature limit.

Gap Resummation

Gap resummation offers an alternative to the diagrammatic approaches to resummation described above, which quickly become cumbersome at higher loop order. Rather than evaluating such diagrams analytically, in gap resummation one need merely compute $V_{\text{eff}}^{(1)}$ and then solve the so-called "gap equation" for the thermal mass. This gap equation includes the dominant contributions from many higher-order diagrams, though admittedly it does not include contributions from certain sub-leading diagrams at each order (for example, parts of the 2-loop sunset diagram). After solving the gap equation for the thermal mass, this is substituted into either $V_{\text{eff}}^{(1)}$ in the full dressing (FD) prescription or into $\partial_\phi V_{\text{eff}}^{(1)}$ in the partial dressing (PD) prescription. The latter is sometimes also called tadpole resummation, and has been demonstrated to count daisy and superdaisy diagrams more faithfully to higher order.

The first step in either procedure is solving the gap equation for the thermal mass,

$$M_i^2 = \frac{\partial^2}{\partial \phi_i^2} V_{\text{eff}}^{(1)}(M_j^2)\,. \tag{4.132}$$

Note that the thermal mass M^2 appears on both the left- and right-hand sides of this equation, which must be solved numerically. Considering for the moment single-field ϕ^4 theory in the high temperature expansion — the effective potential for which given in Eq. (4.90) — this gap equation is explicitly

$$M^2 \overset{\text{high-}T}{=} m^2 + \frac{\lambda T^2}{4} - \frac{3\lambda M T}{4\pi} - \frac{3\lambda L_R M^2}{16\pi^2} - \frac{9\lambda^2 \phi^2 T}{4\pi M} - \frac{9\lambda^2 L_R \phi^2}{8\pi^2}\,. \tag{4.133}$$

The set of gap equations in the 2-field case is even more complicated given their coupled

nature

$$M_1^2 \overset{\text{high-}T}{=} \mu_1^2 + 3\lambda_1\phi_1^2 + \frac{\lambda_{12}}{2}\phi_2^2 + c_1T^2 - \frac{3\lambda_1}{4\pi}M_1T - \frac{\lambda_{12}}{8\pi}M_2T \\ - \frac{3\lambda_1}{16\pi^2}L_RM_1^2 - \frac{\lambda_{12}}{32\pi^2}L_RM_2^2 - \frac{L_R}{32\pi^2}(36\lambda_1^2 + \lambda_{12}^2)\phi_1^2\,, \tag{4.134a}$$

$$M_2^2 \overset{\text{high-}T}{=} \mu_2^2 + 3\lambda_2\phi_2^2 + \frac{\lambda_{12}}{2}\phi_1^2 + c_2T^2 - \frac{3\lambda_2}{4\pi}M_2T - \frac{\lambda_{12}}{8\pi}M_1T \\ - \frac{3\lambda_2}{16\pi^2}L_RM_2^2 - \frac{\lambda_{12}}{32\pi^2}L_RM_1^2 - \frac{L_R}{32\pi^2}(36\lambda_2^2 + \lambda_{12}^2)\phi_2^2\,. \tag{4.134b}$$

Full Dressing

In the full dressing (FD) prescription, the thermal masses M_i^2 obtained by solving the gap equations are substituted directly into the effective potential, $V_{\text{eff}}^{\text{FD}} = V_{\text{eff}}^{(1)}|_{m_i^2\to M_i^2}$, with the result

$$V_{\text{eff}}^{\text{FD}} = \frac{1}{2}(\mu_1^2 + c_1T^2)\phi_1^2 + \frac{1}{2}(\mu_2^2 + c_2T^2)\phi_2^2 + \frac{\lambda_1}{4}\phi_1^4 + \frac{\lambda_2}{4}\phi_2^4 + \frac{\lambda_{12}}{4}\phi_1^2\phi_2^2 \\ - \frac{1}{12\pi}(M_1^3 + M_2^3)T - \frac{L_R}{64\pi^2}(M_1^4 + M_2^4)\,. \tag{4.135}$$

When one uses the high-temperature expanded truncated thermal masses, this is identical to the 1-loop effective potential in the Parwani scheme $V_{\text{eff,P}}^{(1)}$, given in Eq. (4.117). More generally when using the full solutions of the gap equation, however, they differ.

Partial Dressing

One issue with the FD prescription is that it miscounts certain diagrams starting at 2-loop order [313, 312]. This shortcoming led to the introduction of the partial dressing (PD) prescription, in which one replaces $m^2 \to M^2$ on the level of the first derivative of the

effective potential $\partial_\phi V_{\text{eff}}$ and then integrates to obtain the resummed effective potential

$$V_{\text{eff}}^{\text{PD}} = \int d\phi \left(\frac{\partial V_{\text{eff}}^{(1)}(m_i^2)}{\partial \phi} \right)_{m_i^2 \to M_i^2}, \tag{4.136}$$

where the M_i's are the full solution of the gap equations. The procedure can be understood by thinking about Dyson resummation in zero-temperature field theory. There, self-energy corrections are resummed into the propagator

$$\frac{i}{p^2 - m^2} \to \frac{i}{p^2 - m^2 + \Sigma(p^2)}, \tag{4.137}$$

where $\Sigma(p^2)$ is the self energy and p^2 is the momentum. If Σ is not dependent on p^2, all corrections to the propagator can be absorbed into the mass via

$$m^2 \to M^2 = m^2 - \Sigma\,. \tag{4.138}$$

This resummation can also be used to absorb thermal corrections since their leading part does not depend on the momentum. Moreover, we can identify M^2 as the solution of the gap equation. To ensure proper resummation, it just needs to be ensured that this replacement is done only for all propagators and not for vertices, which in the effective potential are, however, expressed in terms of the mass. To avoid also dressing the vertex, the replacement is done on the level of the tadpole for which the coupling is explicit (see below).

The PD procedure has been demonstrated to correctly count the most relevant diagrams up to 4-loop order [297]. To compare the two and explicitly understand why FD leads to a miscounting while PD does not, we will consider the 1-loop tadpole diagrams which compute $\partial_\phi V_{\text{1-loop}}$ (in contrast to the vacuum diagrams which compute $V_{\text{1-loop}}$). These tadpoles can be formed from the corresponding vacuum diagrams by attaching a zero-momentum truncated external leg to each part of the vacuum diagram. By shifting $m^2 \to M^2$ at the

level of vacuum diagrams, FD is equivalent to dressing both the propagator and 3-point vertex $c_3 \equiv \partial_\phi m^2$ of the corresponding 1-loop tadpole diagrams. In contrast, by shifting $m^2 \to M^2$ at the tadpole level, PD dresses only the propagator.

To see explicitly that the former dressing of both vertex and propagator leads to a miscounting, let us return for the moment to the single-field ϕ^4 theory. Presuming the hierarchy $\lambda T^2/m^2 \simeq 1$, $\lambda T/m < 1$, and $\phi^2/T^2 \ll 1$, the leading order gap equation is[13]

$$M^2 \overset{\text{high-}T}{=} m^2 + \frac{\lambda T^2}{4} - \frac{3\lambda M T}{4\pi} - \frac{9\lambda^2\phi^2 T}{4\pi M}\,. \tag{4.139}$$

The solution gives the dressed propagator, and to $\mathcal{O}(\lambda^3)$ reads

$$M \overset{\text{high-}T}{\simeq} m + \frac{\lambda T^2}{8m} - \frac{3\lambda T}{8\pi} - \frac{\lambda^2 T^4}{128 m^3} + \frac{9\lambda^2 T^2}{128\pi^2 m} + \frac{\lambda^3 T^6}{1024 m^5} - \frac{9\lambda^3 T^4}{1024\pi^2 m^3}\,. \tag{4.140}$$

Meanwhile, the dressed 3-point vertex C_3 is obtained by differentiating the gap equation, $C_3 = \partial_\phi M^2$. Working to the same order, the solution is

$$C_3 \overset{\text{high-}T}{=} 6\lambda\phi\left(1 - \frac{9\lambda T}{8\pi m} + \frac{9\lambda^2 T^3}{64\pi m^3} - \frac{27\lambda^3 T^5}{1024\pi m^5} + \frac{81\lambda^3 T^3}{1024\pi^3 m^3}\right)\,, \tag{4.141}$$

In FD, both the propagator and vertex of the tadpole are improved, and so starting from Eq. (4.89), the contribution to the derivative of the 1-loop effective potential reads

$$\partial_\phi V_{\text{1-loop}}^{\text{FD}} \overset{\text{high-}T}{=} \frac{T^2}{24}(6\lambda\phi) - \frac{MT}{8\pi} C_3 + ... \tag{4.142}$$

where the first leading order term comes from the hard thermal loop, the second term comes from the zero mode, and we suppress higher order terms. This is in contrast to PD, for

13. The last term proportional to ϕ^2 is formally subleading in this equation for M^2. It, however, becomes important when taking the derivative with respect to ϕ for computing the 3-point vertex.

which only the propagator is dressed, leading to

$$\partial_\phi V_{\text{1-loop}}^{\text{PD}} \overset{\text{high-}T}{=} \frac{T^2}{24}(6\lambda\phi) - \frac{MT}{8\pi}c_3 + ... \tag{4.143}$$

where $c_3 = \partial_\phi m^2 = 6\lambda\phi$ is the undressed 3-point vertex. Explicitly using the forms of M and C_3 above, the zero-mode piece in either case reads

$$\begin{aligned} -\frac{MT}{8\pi}C_3 \overset{\text{high-}T}{=} (6\lambda\phi)\Bigg(&- \frac{mT}{8\pi} - \frac{\lambda T^3}{64\pi m} + \frac{3\lambda T^2}{16\pi^2} + \frac{\lambda^2 T^5}{1024\pi m^3} - \frac{63\lambda^2 T^3}{1024\pi^3 m} \\ &- \frac{\lambda^3 T^7}{8192\pi m^5} + \frac{63\lambda^3 T^5}{8192\pi^3 m^3}\Bigg), \end{aligned} \tag{4.144}$$

$$\begin{aligned} -\frac{MT}{8\pi}c_3 \overset{\text{high-}T}{=} (6\lambda\phi)\Bigg(&- \frac{mT}{8\pi} - \frac{\lambda T^3}{64\pi m} + \frac{3\lambda T^2}{64\pi^2} + \frac{\lambda^2 T^5}{1024\pi m^3} - \frac{9\lambda^2 T^3}{1024\pi^3 m} \\ &- \frac{\lambda^3 T^7}{8192\pi m^5} + \frac{9\lambda^3 T^5}{8192\pi^3 m^3}\Bigg). \end{aligned} \tag{4.145}$$

The difference between the fully dressed and partially dressed 1-loop tadpole, defined as $\Delta \equiv \partial_\phi V_{\text{1-loop}}^{\text{FD}} - \partial_\phi V_{\text{1-loop}}^{\text{PD}}$, is then

$$\Delta \overset{\text{high-}T}{=} (6\lambda\phi)\left(\frac{9\lambda T^2}{64\pi^2} - \frac{27\lambda^2 T^3}{512\pi^3 m} + \frac{27\lambda^3 T^5}{4096\pi^3 m^3}\right). \tag{4.146}$$

By computing contributions from the relevant Feynman diagrams up to 4-loop order, Ref. [297] finds an expression for $\partial_\phi V_{\text{1-loop}}$ which precisely matches that of $\partial_\phi V_{\text{1-loop}}^{\text{PD}}$ above, and so Δ quantifies the extraneous contribution due to miscounting in the FD procedure. These extra terms are not present for PD, which automatically includes subleading thermal corrections of super-daisy order. This is reflected in Fig. 4.7, which compares the PD procedure with FD, Parwani, and Arnold-Espinosa resummation schemes at 1- and 2-loop level.

One shortcoming of PD is that an ambiguity arises when field excursions can proceed along multiple directions. Namely it is unclear which field to take the derivative of V_{eff} with respect to, since in general $V_{\text{eff}}^{\text{PD},1} \neq V_{\text{eff}}^{\text{PD},2}$, where $V_{\text{eff}}^{\text{PD},i} = \int d\phi_i\, \partial_i V_{\text{eff}}\big|_{M^2}$. Shortly, we

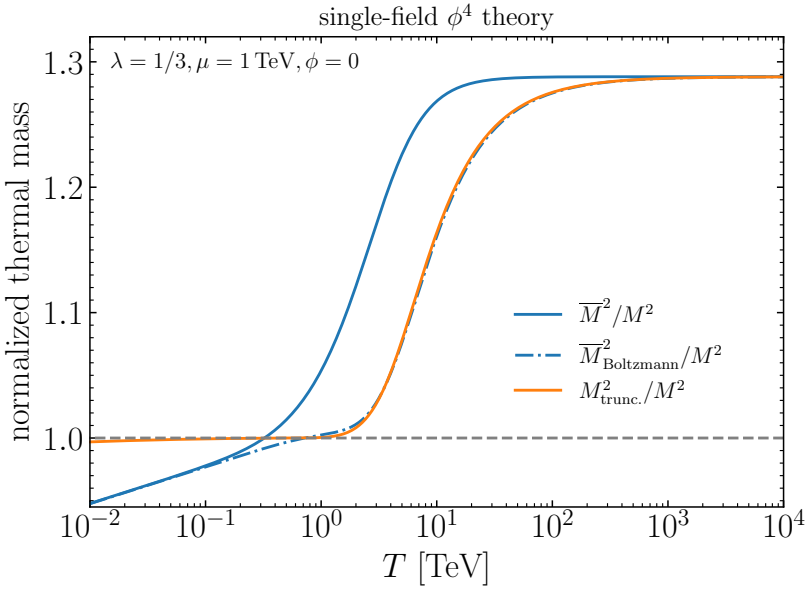

Figure 4.4: Different approximations of the thermal masses (blue: high-temperature thermal mass, $\overline{M}^2$, see Eq. (4.106); blue dot-dashed: high-temperature thermal mass with additional Boltzmann factor, $\overline{M}^2_{\text{Boltzmann}}$, see Eq. (4.107); orange: tree-level mass plus full one-loop correction, $M^2|_{\text{trunc.}}$, see Eq. (4.81) normalized to the thermal mass obtained by solving the gap equation M^2 as a function of the temperature.

will propose a multi-field generalization which holds even in the case of mixing scalar fields, making PD suitable for a wider range of applications, in particular BSM extensions of the Higgs sector.

Numerical Comparison

Single-Field ϕ^4 Theory

For our numerical discussion of the single-field ϕ^4 theory, we focus on the benchmark point

$$\lambda = 1/3, \quad \mu = 1\,\text{TeV}\,. \tag{4.147}$$

We set the renormalization scale equal to the temperature.

We start our numerical discussion by comparing different approximations for the thermal masses as a function of the temperature in Fig. 4.4. The thermal masses are normalized to the full thermal mass computed by solving the gap equation.

As expected, the high-temperature expansion of the thermal mass $\overline{M}^2$ is close to the full solution of the gap equation M^2 for low temperatures since the overall thermal corrections are negligible (see blue curve and 4.106)). The small $\sim 5\%$ deviation for temperatures close to zero is explained by loop corrections induced by the Coleman-Weinberg potential, which are not taken into account in the high-temperature expansion. For temperatures above 1TeV, the ratio of the high-temperature thermal mass to the full thermal mass increases quickly until the curve converges at ~ 1.3 for $T \gtrsim 10$TeV and then stays constant for higher temperatures. The constant off-shift is caused by temperature-dependent higher-loop contributions which are generated by solving the gap equation but not taken into account in the high-temperature thermal mass.[14]

If an additional Boltzmann factor is included in the equation for the high-temperature mass (see Eq. (4.107), blue dot-dashed curve), the ratio stays close to one for a slightly larger temperature range than without the Boltzmann factor. The overall agreement with the full thermal mass is, however, not substantially improved.

If instead the full one-loop correction (without any high-temperature expansion) is used to calculate the thermal mass, $M^2|_{\text{trunc.}}$, (see orange curve and Eq. (4.81)), the low-temperature behaviour of the full thermal mass is very well captured (since now also the one-loop corrections from the Coleman-Weinberg potential are taken into account). At $T \gtrsim 4$TeV, where temperature-dependent higher-loop order corrections start to become relevant, however, also this approximation fails to capture the temperature dependence of the full thermal mass.

The real parts of the effective potential are compared in Fig. 4.5. For low temperatures,

14. This can for example easily be seen by solving Eq. (4.139) for $\phi = 0$ without expanding in λ (as done for Eq. (4.140)).

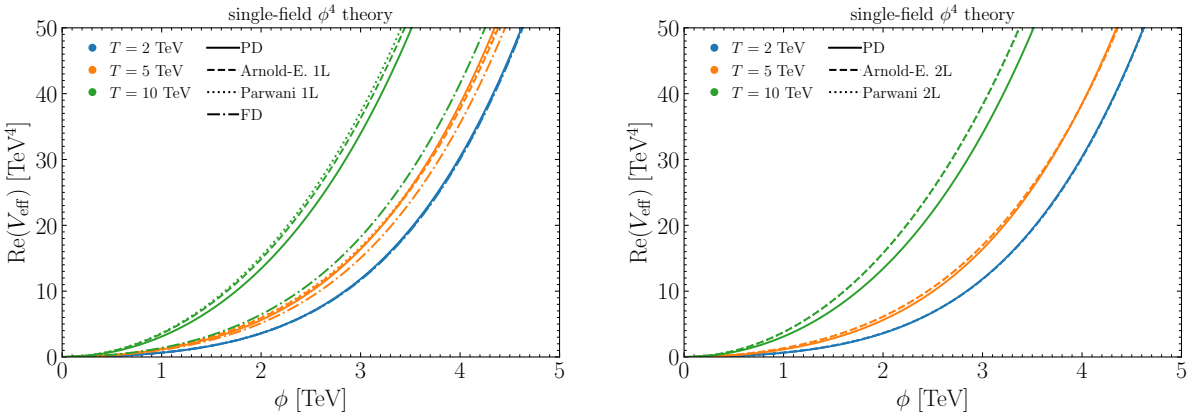

Figure 4.5: Left: Real part of the one-loop effective potential in the one-field ϕ^4 model evaluated using different resummation methods. Right: Same as left, but two-loop corrections are included for the full-dressing approaches.

the thermal corrections are small and, consequently, the different resummation methods yield almost identical results. Only for higher temperatures of $T \gtrsim 10$TeV, differences between the three methods become visible.

In the left panel of Fig. 4.5, showing the real part of the effective potential evaluated using various resummation methods at the one-loop level, very small differences between Parwani and Arnold-Espinosa resummation are visible originating from subleading thermal corrections which are partially included for Parwani resummation but not for Arnold-Espinosa resummation. This discrepancy is completely gone in the right panel of Fig. 4.5, for which the effective potential is evaluated at the two-loop level for the Parwani and Arnold-Espinosa curves. As a consequence of explicitly including the subleading thermal two-loop corrections, whose leading contribution is $\propto \lambda^2 m^2 T^2$, the curves using Parwani and Arnold-Espinosa resummation lie on top of each other. This seemingly signals a well-behaved perturbative convergence of the subleading thermal corrections.

The result using partial dressing lies below the Parwani and Arnold-Espinoa results for high temperatures. Interestingly, the difference is increased when comparing the partial-dressing result to the two-loop full-dressing results in comparison to the one-loop full-dressing

results. While in principle a difference is expected since partial dressing correctly includes subleading thermal corrections, one would naively expect this difference to shrink down once these subleading thermal corrections are explicitly included at the two-loop order for the full-dressing methods. To understand why this increases the difference between full and partial dressing, it is instructive to understand the proportionalities of the formally leading terms missed by the full dressing method. While the formally leading missed term is $\propto \lambda^2 m^2 T^2$ if computing the effective potential at the one-loop level, it is $\propto \lambda^3 m T^3$ if computing the effective potential at the two-loop level. For high temperatures, this three-loop term is larger than the respective two-loop term demonstrating that including the full two-loop corrections in the full dressing approach can worsen the result. Also explicitly including the term $\propto \lambda^3 m T^3$ might not improve the result since for high temperatures the four-loop term $\propto \lambda^4 T^5/m$ could be even larger. This demonstrates the necessity of correctly resuming also subleading thermal corrections and is in direct correspondence to the behaviour of the thermal masses at high temperatures (see Fig. 4.4).

Moreover, we show in the left panel of Fig. 4.5 the result using full dressing. As discussed previously, this resummation scheme miscounts diagrams starting at the two-loop level. This is clearly visible by the large difference to the other resummation methods for $T = 10\text{TeV}$.

Two-Field ϕ^4 Theory

Next, we compare the different resummation methods in the two-field ϕ^4 theory without mixing. We focus on the benchmark point

$$\lambda_1 = \lambda_2 = 1/3, \quad \lambda_{12} = 2, \quad \mu_{11}^2 = -4\text{TeV}^2, \quad \mu_{22}^2 = -1\text{TeV}^2\,. \tag{4.148}$$

The renormalisation scale is again set equal to the temperature.

Fig. 4.6 shows the thermal masses computed either by numerically solving the gap equa-

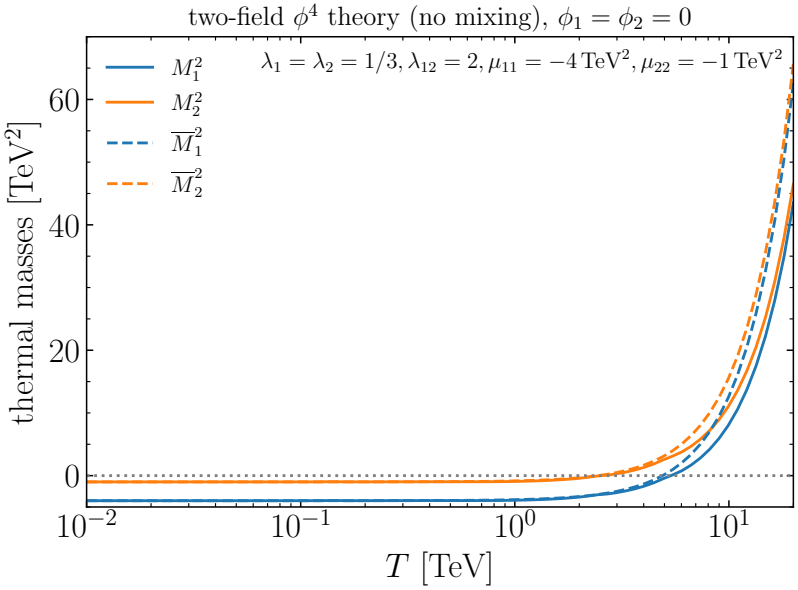

Figure 4.6: Thermal masses of ϕ_1 (blue) and ϕ_2 (orange) in the two-field ϕ^4 model. The thermal masses are either evaluated by numerically solving the gap equations (solid) or in the high-temperature expansion (dashed).

tion (solid lines) or by keeping only the leading term in the high-temperature expansion (dashed lines). As expected, the full and high-temperature versions of the thermal masses agree well for low temperatures since thermal effects are small in general. For very high temperatures, small differences are visible originating from high-order corrections induced by numerically solving the gap equations (see discussion of Fig. 4.4). We observe the largest absolute differences for intermediary temperatures, for which the temperature is similar to the tree-level masses. Here, thermal corrections are important but the high-temperature expansion is not yet a good approximation.

The real part of the effective potential is shown in Fig. 4.7 as a function of ϕ_1 setting $\phi_2 = 0$, implying that there is no mixing between the fields. As for the one-field ϕ^4 theory, the Parwani, Arnold-Espinosa, and partial dressing approaches agree well for low temperatures. For higher temperatures, larger differences are visible. In contrast to the single-field case (see Fig. 4.5), the difference between the Arnold-Espinosa and Parwani resummation methods is

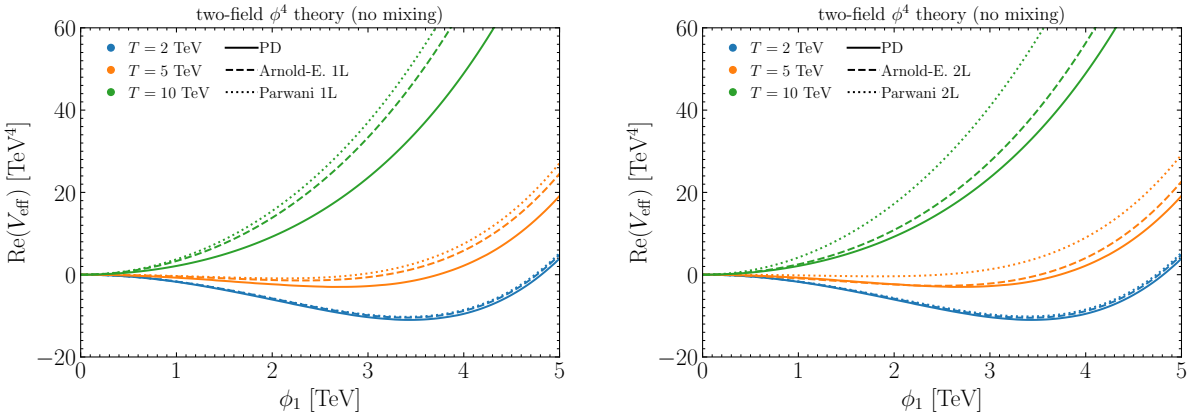

Figure 4.7: Left: Real part of the one-loop effective potential in the two-field ϕ^4 model without mixing evaluated using different resummation methods. Right: Same as left, but two-loop corrections are included for the full-dressing approaches.

increasing from the one- (left panel of Fig. 4.7) to the two-loop level (right panel of Fig. 4.7). This signals that the dominant difference between the two approaches in the given scenario is not of two-loop order but induced by higher-order effects. This is due to the comparably large numerical values for the λ's as well as the larger number of fields which enhance the significance of higher-order corrections. This again shows the importance of resuming also subleading effects as achieved in the partial dressing approach.

4.2.4 Toy Model for Symmetry Non-Restoration

Next, we discuss the phenomenon of electroweak symmetry non-restoration (EWSNR). EWSNR refers to situations in which the EW symmetry is not only broken at low temperatures but also not restored at high temperatures (or the restoration is delayed up to very high energies). This is particularly interesting from the point of resuming thermal corrections. In order for EWSNR to occur, the thermal corrections need to dominate over the tree-level mass turning the squared thermal mass negative and thereby ensuring that the EW symmetry is broken.

We base our discussion on a toy model for symmetry non-restoration which was presented

in [7]. Its potential is given by

$$V^{(0)}(\phi,\chi,S) = \frac{1}{2}\mu_S^2 S^2 + \frac{1}{2}\mu_\chi^2 \sum_i \chi_i^2 + \frac{1}{2}\mu_\phi^2\phi^2 + \frac{1}{4}\lambda_\phi\phi^4 + \frac{1}{4}\lambda_\chi \sum_i \chi_i^4 + \frac{1}{4}\lambda_S S^4 + \frac{1}{4}\lambda_{\phi\chi}\phi^2 \sum_i \chi_i^2 + \frac{1}{4}\lambda_{\phi S}\phi^2 S^2\,,$$

where S and χ_i are vectors of dimension N_S and $N_{\chi i}$. The index i is a generation index which runs from 1 to N_{gen}. We will only evaluate the potential at zero field values for S and the χ_i. Therefore, all χ_i have the same mass m_χ and we will drop the sum over i and instead use $N_\chi = N_{\chi_i} N_{\text{gen}}$. The parameters are chosen such that only ϕ develops a non-zero vacuum expectation value. Therefore, we are only interested in the ϕ direction and set χ_i and S to zero for the evaluation of the effective potential.

In the high-temperature limit, the thermal masses are given

$$\overline{M}_\chi^2 = m_\chi^2 + T^2 c_\chi = m_{\chi i}^2 + T^2\left[\frac{1}{12}(N_\chi+2)\lambda_\chi + \frac{1}{24}N_\phi\lambda_{\phi\chi}\right], \tag{4.149a}$$

$$\overline{M}_{S,\text{Boltzmann}}^2 = m_S^2 + T^2 c_S = m_S^2 + T^2\left[\frac{1}{12}(N_S+2)\lambda_S e^{-m_S/T} + \frac{1}{24}N_\phi\lambda_{\phi S}\right], \tag{4.149b}$$

$$\overline{M}_{\phi,\text{Boltzmann}}^2 = m_\phi^2 + T^2 c_\phi = m_\phi^2 + T^2\left[\frac{1}{12}(N_\phi+2)\lambda_\phi + \frac{1}{24}N_\chi\lambda_{\phi\chi} + \frac{1}{24}N_S\lambda_{\phi S}e^{-m_S/T}\right].$$

This also defines the coefficients $\{c_\chi, c_S, c_\phi\}$, which are used later. Since the parameters are chosen such that $m_\phi \sim m_\chi \ll m_S$, the thermal contributions of S are multiplied by the Boltzmann factor $e^{-m_S/T}$ to better approximate the full thermal loop function for $T \lesssim m_S$ (see also Eq. (4.107).

To achieve symmetry non-restoration, $\lambda_{\phi\chi}$ is chosen to be negative such that the thermal mass of ϕ becomes negative. For $T \sim m_S$, the thermal contribution of S compensates for the negative contribution of the χ_i resulting in the eventual symmetry restoration at $T \gtrsim m_S$.

Since the stability of the potential at the tree level requires

$$\lambda_{\phi\chi} > -2\sqrt{\frac{\lambda_\phi \lambda_\chi}{N_{\rm gen}}}\,, \tag{4.150}$$

a large number of generations is required to ensure symmetry non-restoration,[15] while still satisfying perturbative unitarity bounds.

One-Loop Effective Potential

The one-loop effective potential is given by

$$V^{(1)}(\phi, \chi, S) = \mathcal{J}(m_\phi) + N_\chi \mathcal{J}(m_\chi) + N_S \mathcal{J}(m_S)\,. \tag{4.151}$$

The counterterm contributions are

$$V^{(1,\rm CT)}(\phi, \chi_i = 0, S = 0) = \frac{1}{2}\delta^{(1)}\mu_\phi^2\phi^2 + \frac{1}{4}\delta^{(1)}\lambda_\phi\phi^4\,. \tag{4.152}$$

We choose to renormalize λ_ϕ in the $\overline{\rm MS}$ scheme. For the renormalization of μ_ϕ, we include a finite piece to the counterm

$$\delta^{(1)}\mu_\phi^2\Big|_{\rm fin} = -\frac{1}{2}\phi^2\left[\frac{\partial^2}{\partial\phi^2}V^{(1)}\right]^{\mathcal{O}(\mu_S^2)}_{\phi=\chi_i=S=0,T=0} = \frac{1}{64\pi^2}N_S\lambda_{\phi S}\phi^2\mu_S^2\left(1 - \ln\frac{\mu_S^2}{\mu_R^2}\right)\,, \tag{4.153}$$

where the superscript $\mathcal{O}(\mu_S^2)$ denotes that only the leading contribution proportional to μ_S^2 is considered. This counterterm is chosen to absorb the very large zero-temperature loop corrections to m_ϕ from S (due to $\mu_S \gg |\mu_\phi|$) into the definition of μ_ϕ.

15. In more realistic models, the negative BSM contributions to for instance the thermal mass of the Higgs boson must also overcome the positive contributions of other SM particles.

Two-Loop Effective Potential

The genuine two-loop corrections to the potential are given by

$$\begin{aligned} V^{(2,\text{gen})}(\phi, \chi_i = 0, S = 0) =& \frac{3}{4}\left[\lambda_\phi \mathcal{I}(m_\phi)^2 + N_S \lambda_S \mathcal{I}(m_S)^2 + N_\chi \mathcal{I}(m_{\chi_i})^2\right] \\ &+ \frac{1}{4}\Bigg[(N_S^2 - N_S)\lambda_S \mathcal{I}(m_S)^2 + N_\chi (N_{\chi_i} - 1)\lambda_\chi \mathcal{I}(m_\chi)^2 \\ &\quad + N_S \lambda_{\phi S} \mathcal{I}(m_\phi)\mathcal{I}(m_S) + N_\chi \lambda_{\phi\chi} \mathcal{I}(m_\phi)\mathcal{I}(m_\chi)\Bigg]. \end{aligned}$$

The counterterm contributions at the two-loop level are

$$\begin{aligned} V^{(2,\text{CT})}(\phi, \chi_i = 0, S = 0) =& \frac{1}{2}\delta^{(2)}\mu_\phi^2\phi^2 + \frac{1}{4}\delta^{(2)}\lambda_\phi\phi^4 \\ &+ \frac{1}{2}\delta^{(1)}\mu_\phi^2 \mathcal{I}(m_\phi) + \frac{1}{2}N_\chi \delta^{(1)}\mu_\chi^2 \mathcal{I}(m_\chi) + \frac{1}{2}N_S \delta^{(1)}\mu_S^2 \mathcal{I}(m_S) \\ &+ \frac{3}{2}\delta^{(1)}\lambda_\phi \phi^2 \mathcal{I}(m_\phi) + \frac{1}{4}\delta^{(1)}\lambda_{\phi S}\phi^2 \mathcal{I}(m_S) \\ &+ \frac{1}{4}\delta^{(1)}\lambda_{\phi\chi}\phi^2 \mathcal{I}(m_\chi), \end{aligned}$$

where here we already set $\chi_i = S = 0$. The first line contains the needed two-loop counterterm; the two last lines represent the subloop renormalization. We checked analytically that all UV divergencies and ϵ^1 pieces of the loop integrals (with ϵ being the UV regulator) cancel.

The thermal mass counterterms, relevant for Parwani and Arnold-Espinosa approaches, give additional two-loop contributions:

$$V^{(2,\text{thermal-CT})}(\phi, \chi, S) = -\frac{1}{2}T^2\left(c_\phi \mathcal{I}(m_\phi^2) + N_\chi c_\chi \mathcal{I}(m_\chi^2) + N_S c_S \mathcal{I}(m_S^2)\right). \tag{4.154}$$

We checked explicitly that, in the high-temperature expansion, all T^3 terms cancel. This cross-check was performed separately for Parwani and Arnold/Espinosa resummation in the limits $m_S \ll T$ and $m_S \gg T$. The two-loop counterterm for μ_ϕ is again chosen such that the very large zero-temperature loop corrections to m_ϕ from S are absorbed into the definition

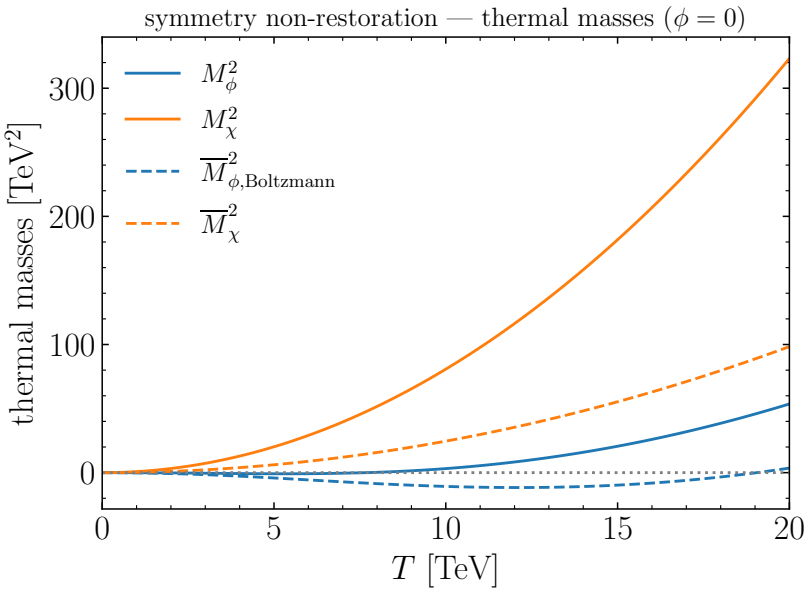

Figure 4.8: Thermal masses of ϕ (blue) and χ_i (orange) in the symmetry non-restoration toy model of [7]. The thermal masses are either evaluated by numerically solving the gap equations (solid) or in the high-temperature expansion (dashed).

of μ_ϕ.

Numerical Comparison

For our numerical comparison, we choose the benchmark point already used in [7]:

$$\begin{aligned} &N_{\text{gen}} = 12, \quad N_{\chi_i} = 4, \quad N_S = 12, \\ &\mu_\phi^2 = -0.01\,\text{TeV}^2, \quad \mu_\chi^2 = 0.01\,\text{TeV}^2, \quad \mu_S^2 = 400\,\text{TeV}^2, \\ &\lambda_\phi = 0.1\,\text{TeV}^2, \quad \lambda_\chi = 0.5, \quad \lambda_S = 1, \quad \lambda_{\phi\chi} = -0.1, \quad \lambda_{\phi S} = 1\,. \end{aligned} \tag{4.155}$$

This benchmark is chosen to realize symmetry non-restoration for $T \lesssim 10\,\text{TeV}$. At higher temperatures, the S field eventually ensures symmetry restoration.

Thermal Masses

We start with a comparison of the thermal masses M_ϕ and M_χ in Fig. 4.8. For the scalar ϕ, the thermal mass calculated by solving the gap equation M_ϕ^2 is negative only for small temperatures ($T \lesssim 7\,\text{TeV}$). If instead the high-temperature expansion is used (see Eq. (4.149)), the squared thermal mass stays negative until much higher masses ($T \sim 19\,\text{TeV}$). This difference has two origins: 1) the thermal loop functions appearing in the gap equation is not expanded in the high-temperature limit; 2) solving the gap equation numerically effectively includes higher-order corrections. For the scalars χ, the differences between the thermal mass obtained by solving the gap equation and the high-temperature expansion of Eq. (4.149) is even more pronounced due to the relatively large number of fields $N_\chi = 48$ coupled to each other.

Effective Potential

Next, we study the effective potential itself. Fig. 4.9 shows the dependence of the effective potential on the value of ϕ for various temperature values.

In the upper left panel, we show the real part of V_{eff} calculated at the one-loop level using Arnold-Espinosa and Parwani. For both Parwani resummation (solid lines) and Arnold-Espinosa reummation (dashed lines), electroweak symmetry non-restoration is clearly visible for $T \lesssim 6\,\text{TeV}$. For higher temperatures, the thermal contribution of the S triggers the eventual symmetry restoration. While for low temperatures both methods yield very similar results, there is an increasing difference for higher temperatures. As discussed previously, this difference arises from subleading super daisy-like contributions which are partially included in the Parwani approach.

One natural way to reduce the difference between both methods and thereby the theoretical uncertainty is to explicitly include the full two-loop corrections as outlined above. The results are shown in the upper right panel of Fig. 4.9. Remarkably, the difference between

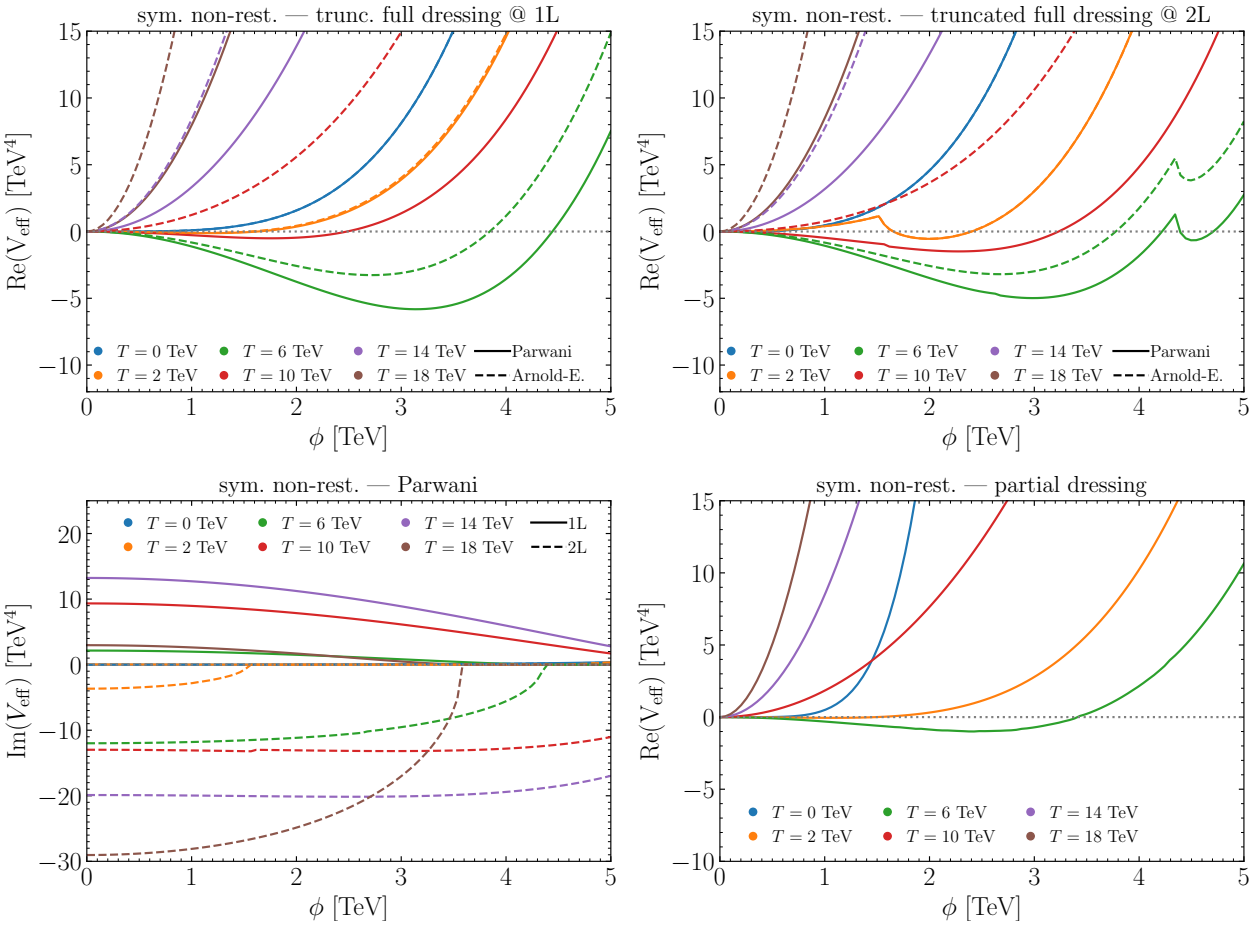

Figure 4.9: Upper left: Real part of the one-loop effective potential for the symmetry non-restoration model of [7] evaluated using Arnold-Espinosa and Parawni resummation. For the thermal resummation, we employ either the Parwani (solid) or Arnold-Espinsa (dashed) methods. Upper right: Same as upper left, but the effective potential is evaluated at the two-loop level. Lower left: Same as upper right, but the imaginary part of the effective potential is shown. Lower right: Same as upper plots, but the effective potential is evaluated using partial dressing.

Parawani and Arnold/Espinosa resummation is not reduced if including the full two-loop corrections but of similar size as at one-loop level. After including the full two-loop corrections, the difference between the two resummation schemes is of three-loop order. The fact that the difference is not decreased when going from the one- to the two-loop level signals that also the three-loop difference is sizeable. This is a consequence of the large multiplicity of fields and demonstrates that a resummation of subleading super-daisy corrections is needed.

It is even more astonishing that some of the two-loop results feature unphysical kinks. These kinks appear if one of the squared thermal masses crosses zero (see e.g. the scenario of Fig. 4.8). For a negative mass squared, the loop functions develop an imaginary part. While the imaginary parts of a negative tree-level mass squared cancel once large thermal effects are resumed as shown in [315], an imaginary part remains if one of the squared thermal masses becomes negative. We show this remaining imaginary part of the effective potential (without normalizing the potential to zero a the origin) in the lower left panel of Fig. 4.8. We see that imaginary parts already occur at the one-loop level. The size of the imaginary parts is, however, enhanced at the two-loop level. This is due to a mismatch between the figure-eight diagrams and the thermal counterterm contributions. While the full thermal loop functions (without any high-temperature expansion) are considered for the former contribution, the thermal counterterms appearing in the latter contribution are by definition derived in the high-temperature expansion. This (unavoidably) different treatment of the two contributions artificially enhances the imaginary part of the effective potential and also induces kinks in the real part via products of two imaginary parts. In general, large imaginary parts also question the validity of the perturbative calculation of the effective potential. As discussed in [316], the imaginary part of the effective potential corresponds to a decay width of the localised ground state into a non-localised ground state. Therefore, calculations based on a perturbatively calculated effective potential should only be trusted if the imaginary part is

small against the real part.

The problems with the perturbative convergence and large imaginary parts are completely avoided if we use partial dressing. We show the corresponding result in the lower right panel of Fig. 4.9. As discussed previously, partial dressing does not only resum daisy but also super-daisy contributions. Since no high-temperature expansion is applied and all contributions are treated on the same footing, no large imaginary parts appear. Consequently, no kinks appear in the result and calculations based on the pertubatively calculated effective potential are trustworthy.

4.2.5 Resummation in Multi-Field ϕ^4 Theory with Mixing

After discussing resummation in multi-field theories without mixing, we now turn to the case with mixing between the scalar fields. For simplicity, we focus on a simple toy model consisting of two real scalar fields ϕ_1 and ϕ_2 with tree-level potential

$$V_0 = -\frac{\mu_1^2}{2}\phi_1^2 - \frac{\mu_2^2}{2}\phi_2^2 + \frac{\lambda_1}{4}\phi_1^4 + \frac{\lambda_2}{4}\phi_2^4 + \frac{\lambda_{12}}{4}\phi_1^2\phi_2^2\,. \tag{4.156}$$

We allow both fields to potentially develop a zero-temperature vacuum expectation value. While this significantly complicates several formal aspects of the resummation procedure, it is nonetheless important that we allow them to mix in anticipation of concrete BSM applications.

At the tree level, the scalar mass matrix is given by

$$\mathcal{M}^2(\phi_1,\phi_2) \equiv \begin{pmatrix} m_{11}^2 & m_{12}^2 \\ m_{12}^2 & m_{22}^2 \end{pmatrix} = \begin{pmatrix} -\mu_1^2 + 3\lambda_1\phi_1^2 + \frac{\lambda_{12}}{2}\phi_2^2 & \lambda_{12}\phi_1\phi_2 \\ \lambda_{12}\phi_1\phi_2 & -\mu_2^2 + 3\lambda_2\phi_2^2 + \frac{\lambda_{12}}{2}\phi_1^2 \end{pmatrix}, \tag{4.157}$$

where the fields $\phi_{1,2}$ take on their background values. This matrix can be diagonalized as

$\mathcal{R}_\theta^{-1}\mathcal{M}^2\mathcal{R}_\theta = \mathcal{M}^2_{\rm diag} \equiv {\rm diag}(m_+^2, m_-^2)$, with mass eigenvalues

$$m_\pm^2 = \frac{1}{2}\left(m_{11}^2 + m_{22}^2 \pm D\right)\,, \;\; \text{with} \;\; D = \sqrt{(m_{11}^2 - m_{22}^2)^2 + 4m_{12}^4}\,. \tag{4.158}$$

It is convenient to parameterize $\mathcal{S}$ in terms of the tree level mixing angle θ

$$\mathcal{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\,, \;\; \sin 2\theta = \frac{2m_{12}^2}{\sqrt{(m_{11}^2 - m_{22}^2)^2 + 4m_{12}^4}}\,, \tag{4.159}$$

which can then be used to relate $\phi_{1,2}$ to the mass eigenstates $\phi_\pm$

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \mathcal{R}_\theta \begin{pmatrix} \phi_+ \\ \phi_- \end{pmatrix}\,. \tag{4.160}$$

The background field-dependent mass eigenstates enter into the 1-loop contribution to the effective potential as

$$V_{\text{1-loop}} = \mathcal{J}[m_+] + \mathcal{J}[m_-]\,, \tag{4.161}$$

where the $\mathcal{J}$ function is defined in 4.2.7.

High-Temperature Expansion and Truncated Full Dressing

To gain some intuition, we will first consider the high-temperature limit, in which the field-dependent part of the one-loop effective potential reads

$$V_{\text{1-loop}} \simeq \frac{T^2}{24}\left(m_+^2 + m_-^2\right) - \frac{T}{12\pi}\left(m_+^3 + m_-^3\right) - \frac{L}{64\pi^2}\left(m_+^4 + m_-^4\right)\,, \tag{4.162}$$

where $L = \log\left(\frac{\mu_R^2}{T^2}\right) + 2(\gamma_E - \ln\pi)$ is field independent. Considering just the leading contribution $\sim T^2$, the one-loop corrected effective potential is

$$V_{\text{eff}}^{(1)} \xrightarrow{\text{high } T} \frac{1}{2}\left(-\mu_1^2 + c_1 T^2\right)\phi_1^2 + \frac{\lambda_1}{4}\phi_1^4 + \frac{1}{2}\left(-\mu_2^2 + c_2 T^2\right)\phi_2^2 + \frac{\lambda_2}{4}\phi_2^4 + \frac{\lambda_{12}}{4}\phi_1^2\phi_2^2\,, \quad (4.163)$$

where we have defined the coefficients

$$c_1 = \frac{1}{24}(6\lambda_1 + \lambda_{12})\,, \;\; c_2 = \frac{1}{24}(6\lambda_2 + \lambda_{12})\,. \quad (4.164)$$

Letting $M_i^2(\Phi, T) = m_i^2(\Phi) + \delta m_i^2(\Phi, T)$ be the thermally corrected mass, the truncated gap equation is simply

$$M_i^2 = \frac{\partial^2}{\partial \phi_i^2} V_{\text{eff}}^{(1)}\,, \quad (4.165)$$

which leads to the finite-temperature mass matrix

$$\begin{aligned}\mathcal{M}_T^2(\phi_1, \phi_2, T) &\equiv \begin{pmatrix} M_{11}^2 & M_{12}^2 \\ M_{12}^2 & M_{22}^2 \end{pmatrix} \\ &= \begin{pmatrix} -\mu_1^2 + c_1 T^2 + 3\lambda_1\phi_1^2 + \frac{\lambda_{12}}{2}\phi_2^2 & \lambda_{12}\phi_1\phi_2 \\ \lambda_{12}\phi_1\phi_2 & -\mu_2^2 + c_2 T^2 + 3\lambda_2\phi_2^2 + \frac{\lambda_{12}}{2}\phi_1^2 \end{pmatrix}. \end{aligned} \quad (4.166)$$

Diagonalizing $\mathcal{M}_T^2$ yields the finite-temperature mass eigenstates $M_\pm^2(\Phi, T)$,

$$M_\pm^2 = \frac{1}{2}(M_{11}^2 + M_{22}^2 \pm \mathcal{D})\,, \;\; \text{with } \mathcal{D} = \sqrt{(M_{11}^2 - M_{22}^2)^2 + 4M_{12}^2}\,, \quad (4.167)$$

as well as the finite-temperature mixing angle $\Theta(\phi_1, \phi_2, T)$,

$$\sin 2\Theta = 2M_{12}^2/\mathcal{D}\,. \quad (4.168)$$

With these preliminaries out of the way, we now turn to the resummation of the 1-loop effective potential.

In the truncated full dressing (TFD) prescription, resummation amounts to simply replacing $m_i^2 \to M_i^2|_{\text{trunc.}}$ on the level of the effective potential: $V_{\text{eff}}^{\text{TFD}} = V_{\text{eff}}\big|_{M_i^2|_{\text{trunc.}}}$, with $i = \pm$ labeling the finite-temperature mass eigenstates. In the high-temperature expansion, the resummed potential reduces to the one obtained with the Parwani prescription, namely

$$V_{\text{eff}}^{\text{TFD,Parwani}} = V_0 + \frac{T^2}{24}\left(\overline{M}_+^2 + \overline{M}_-^2\right) - \frac{T}{12\pi}\left(\overline{M}_+^3 + \overline{M}_-^3\right) - \frac{L}{64\pi^2}\left(\overline{M}_+^4 + \overline{M}_-^4\right) , \quad (4.169)$$

or equivalently

$$\begin{aligned} V_{\text{eff}}^{\text{TFD,Parwani}} =& \frac{1}{2}\left(-\mu_1^2 + c_1 T^2\right)\phi_1^2 + \frac{\lambda_1}{4}\phi_1^4 + \frac{1}{2}\left(-\mu_2^2 + c_2 T^2\right)\phi_2^2 + \frac{\lambda_2}{4}\phi_2^4 + \frac{\lambda_{12}}{4}\phi_1^2\phi_1^2 \\ & - \frac{T}{12\pi}\left(\overline{M}_+^3 + \overline{M}_-^3\right) - \frac{L}{64\pi^2}\left(\overline{M}_+^4 + \overline{M}_-^4\right) . \end{aligned}$$

Using the Arnold-Espinosa prescription, the thermal mass is only inserted in the $T\overline{M}^3$ terms.

Partial Dressing

As discussed previously, truncated dressing suffers from various issues. First, it does not resum subleading super-daisy corrections. Second, it unavoidably relies on a high-temperature expansion of the thermal masses, which often is not justified. Third, mismatches between the treatment of various two-loop contributions lead to unphysical kinks in the effective potential. As we have discussed, partial dressing avoids these issues. One shortcoming, however, is that prior to this work it was unknown how to apply the partial dressing prescription in the case where multiple scalar fields acquire non-zero vacuum expectation values and mix. Here, we will demonstrate how partial dressing can be applied to the case of mixing scalar fields.

We start with the gap equations. If we go beyond the leading term in the high-temperature

expansion, the gap equations are promoted to a matrix equation

$$\mathcal{M}_T^2 = \left[\begin{pmatrix} \frac{\partial^2}{\partial\phi_1^2} & \frac{\partial^2}{\partial\phi_1\partial\phi_2} \\ \frac{\partial^2}{\partial\phi_1\partial\phi_2} & \frac{\partial^2}{\partial\phi_2^2} \end{pmatrix} V_{\text{eff}} \right]_{(m_\pm, s_{2\theta})\to(M_\pm, s_{2\Theta})} , \tag{4.170}$$

where $M_\pm$ and $s_{2\Theta} = \sin 2\Theta$ are determined by diagonalizing $\mathcal{M}_T^2$. The second derivatives $\frac{\partial^2}{\partial\phi_i\partial\phi_j} V_{\text{eff}}$ directly correspond to the $\phi_i\phi_j$ two-point functions at zero momentum. The resulting mixing angle relates the original fields $\phi_{1,2}$ to the loop-corrected fields $\Phi_\pm$,

$$\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \mathcal{R}_\Theta \begin{pmatrix} \Phi_+ \\ \Phi_- \end{pmatrix} . \tag{4.171}$$

These loop corrections include both zero-temperature as well as finite-temperature effects.

To solve Eq. (4.170) iteratively, it is important to express the right-hand side completely in terms of the masses and the mixing angle. This can be done either by calculating the second derivatives of the effective potential diagrammatically (i.e., in terms of self-energy Feynman diagrams) or by expressing the first and second derivatives of the (field-dependent) masses in terms of the masses and the mixing angle. For example,

$$\frac{\partial m_+^2}{\partial\phi_1} = \Big[6\lambda_1 \sin^2\theta(\phi_1,\phi_2) + \lambda_{12}\cos^2\theta(\phi_1,\phi_2) \Big]\, \phi_1 + \lambda_{12} \sin 2\theta(\phi_1,\phi_2)\, \phi_2 . \tag{4.172}$$

In the Feynman-diagrammatic approach, this angular dependence follows directly from the Feynman rules. For example, the coupling of ϕ_1 to two Φ_-, which appears in the ϕ_1 tadpole corrections, is given by

$$\begin{aligned} c(\phi_1,\Phi_-,\Phi_-) =& c_\Theta^2 c(\phi_1,\phi_1,\phi_1) - 2c_\Theta s_\Theta c(\phi_1,\phi_1,\phi_2) + s_\Theta^2 c(\phi_1,\phi_2,\phi_2) = \\ =& \phi_1\lambda_1 c_\Theta^2 - 2\phi_2\lambda_{12} c_\Theta s_\Theta + \phi_1\lambda_{12} s_\Theta^2 , \end{aligned}$$

In the diagrammatic approach, it is furthermore straightforward to also add the dependence on the external momentum. In this case, the matrix $\mathcal{R}_\theta$ becomes non-unitary and can not be parameterized by a single mixing angle. We leave this for future work.

After the determination of the thermal masses and the thermal mixing angle, we insert them into the first derivatives of the effective potential $\partial_i V_{\text{eff}}$ (tadpoles). While for the case of vanishing mixing, we only need to consider one tadpole (the one for the non-vanishing field), which is then integrated to obtain the effective potential, an ambiguity arises in the case of non-vanishing mixing since in general $V_{\text{eff}}^{\text{TPD},1} \neq V_{\text{eff}}^{\text{TPD},2}$ (with $V_{\text{eff}}^{\text{TPD},i} = \int d\phi_i \, \partial_i V_{\text{eff}}|_{M_j^2}$).[16]

A reasonable solution would be to replace the derivative with a gradient $V'_{\text{eff}} \to \nabla V_{\text{eff}}$ and the integral over ϕ with a line integral to the position in field space (ϕ_1^*, ϕ_2^*) where we intend to evaluate the potential $\int d\phi \to \int_{\mathcal{C}} d\vec{s}$. We therefore propose the following multi-field generalization

$$V_{\text{eff}}^{\text{TPD}} = \int_{\mathcal{C}} d\vec{s} \cdot \nabla V_{\text{eff}}^{(1)}\big|_{(m_\pm,\theta)\to(M_\pm,\Theta)} \,, \tag{4.173}$$

where $\nabla V_{\text{eff}} = \hat{\phi}_1 \frac{\partial V_{\text{eff}}}{\partial \phi_1} + \hat{\phi}_2 \frac{\partial V_{\text{eff}}}{\partial \phi_2}$ and the curve $\mathcal{C}$ connects the origin to (ϕ_1^*, ϕ_2^*). As a consequence of Green's theorem and the fact that the curl of a gradient is zero, the exact form of $\mathcal{C}$ does not matter. For simplicity, we choose $\mathcal{C}$ to be a straight line, which we parameterize as $\vec{s}(t) = (\phi_1^* t, \, \phi_2^* t)$ with $t \in [0, 1]$. The expression for the effective potential becomes

$$V_{\text{eff}}^{\text{TPD}} = \int_0^1 dt \, \left(\phi_1^* \, \frac{\partial V_{\text{eff}}^{(1)}}{\partial \phi_1}\bigg|_{(\phi_1^* t, \phi_2^* t)} + \phi_2^* \, \frac{\partial V_{\text{eff}}^{(1)}}{\partial \phi_2}\bigg|_{(\phi_1^* t, \phi_2^* t)} \right) \Bigg|_{(M_\pm, s_{2\Theta})} \,, \tag{4.174}$$

where the unresummed effective potential appearing on the right-hand side is $V_{\text{eff}}^{(1)} = V_0 + V_{\text{1-loop}}$, with V_0 in Eq. (4.156) and $V_{\text{1-loop}}$ in Eq. (4.162).

16. $V_{\text{eff}}^{\text{TPD},1}$ correctly captures the ϕ_1-dependent part of the effective potential but not the ϕ_2-dependent part. In contrast, $V_{\text{eff}}^{\text{TPD},2}$ correctly captures the ϕ_2-dependent part of the effective potential but not the ϕ_1-dependent part.

In the high-temperature expansion, the first derivative with respect to ϕ_1 is

$$\begin{aligned}\left.\frac{\partial V_{\text{eff}}}{\partial \phi_1}\right|_{\text{high-}T} &= -\mu_1^2\phi_1 + \lambda_1\phi_1^3 + \frac{\lambda_{12}}{2}\phi_1\phi_2^2 \\ &+ \left(\frac{T^2}{24} - \frac{Tm_+}{8\pi} - \frac{Lm_+^2}{32\pi^2}\right)\frac{dm_+^2}{d\phi_1} + \left(\frac{T^2}{24} - \frac{Tm_-}{8\pi} - \frac{Lm_-^2}{32\pi^2}\right)\frac{dm_-^2}{d\phi_1}\,,\end{aligned} \tag{4.175}$$

where

$$\frac{dm_\pm^2}{d\phi_1} = \left[6\lambda_1 \sin^2\theta + \lambda_{12}\cos^2\theta\right]\phi_1 \pm \lambda_{12}\sin 2\theta\phi_2\,. \tag{4.176}$$

A similar expression holds for $\partial V_{\text{eff}}/\partial\phi_2$ with $1 \leftrightarrow 2$. Next we should replace any instances of $m_\pm^2$ and θ with the thermal quantities $M_\pm^2$ and Θ. Finally, we evaluate at $(\phi_1^* t, \phi_2^* t)$, multiply by $\phi_{1,2}^*$ respectively, take the sum, and integrate over $0 \leq t \leq 1$, in accordance with Eq. (4.174).

Due to neglecting the momentum dependence of the self-energy insertions, partial dressing fails to correctly reproduce two-loop sunset diagrams since it can not account for the case of overlapping loop momenta. While for the non-mixing case, one can easily correct for this by multiplying the $\mathcal{I}_2$-contribution to the gap equations by a factor 2/3, the correction is more subtle in the case of mixing scalar fields. We discuss this in detail in **??**. As already known for the non-mixing case in the literature, we also find this correction to be numerically of minor importance for the case of mixing scalars.

Numerical Comparison

For our numerical comparison, we choose the following parameter point:

$$\lambda_1 = \lambda_2 = 1/3,\;\; \lambda_{12} = 2,\;\; \mu_1^2 = -4\,\text{TeV}^2,\;\; \mu_2^2 = -1\,\text{TeV}^2\,. \tag{4.177}$$

These parameters are chosen such that both fields can develop a non-zero vacuum expectation value.

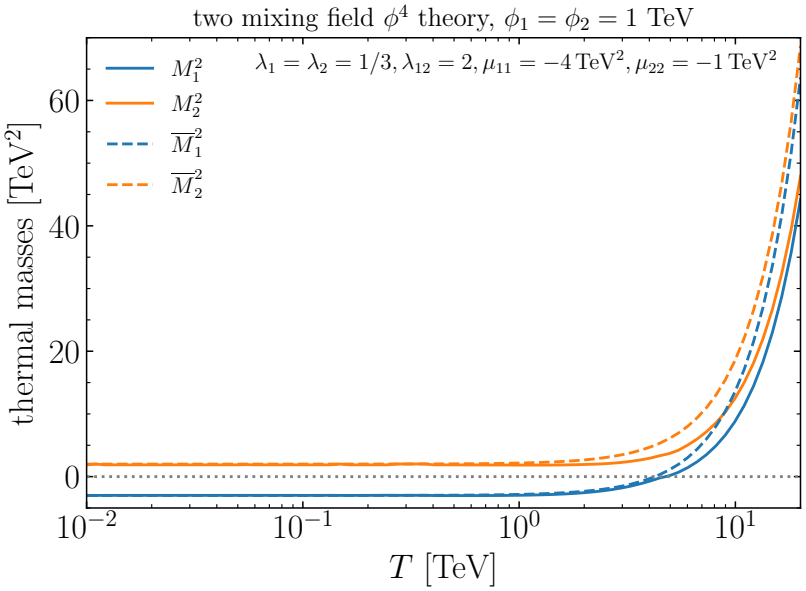

Figure 4.10: Thermal masses of ϕ_1 and ϕ_2 in the two-field ϕ^2 theory as a function of the temperature calculated by solving the gap equations (solid) and the high-temperature expansion (dashed).

Thermal Masses

We start by investigating the thermal masses as a function of the temperature (see Fig. (4.10) for $\phi_1 = \phi_2 = 1\,\mathrm{TeV}$. While for low temperatures, the solutions of the gap equation agree well with the thermal masses in the high-temperature masses, a sizeable difference arises for temperatures close to the zero-temperature masses ($T \sim 5\,\mathrm{TeV}$). For larger temperatures, the differences shrink again even though a visible difference remains. This behaviour is very similar to the results obtained in the EWSNR toy model (see 4.2.4).

Effective Potential

Next, we investigate the real part of the effective potential itself in Fig. (4.11. For low temperatures (see upper left panel), partial dressing and Arnold-Espinosa/Parwani resummation yield very similar results regardless of the values for ϕ_1 and ϕ_2. This is expected

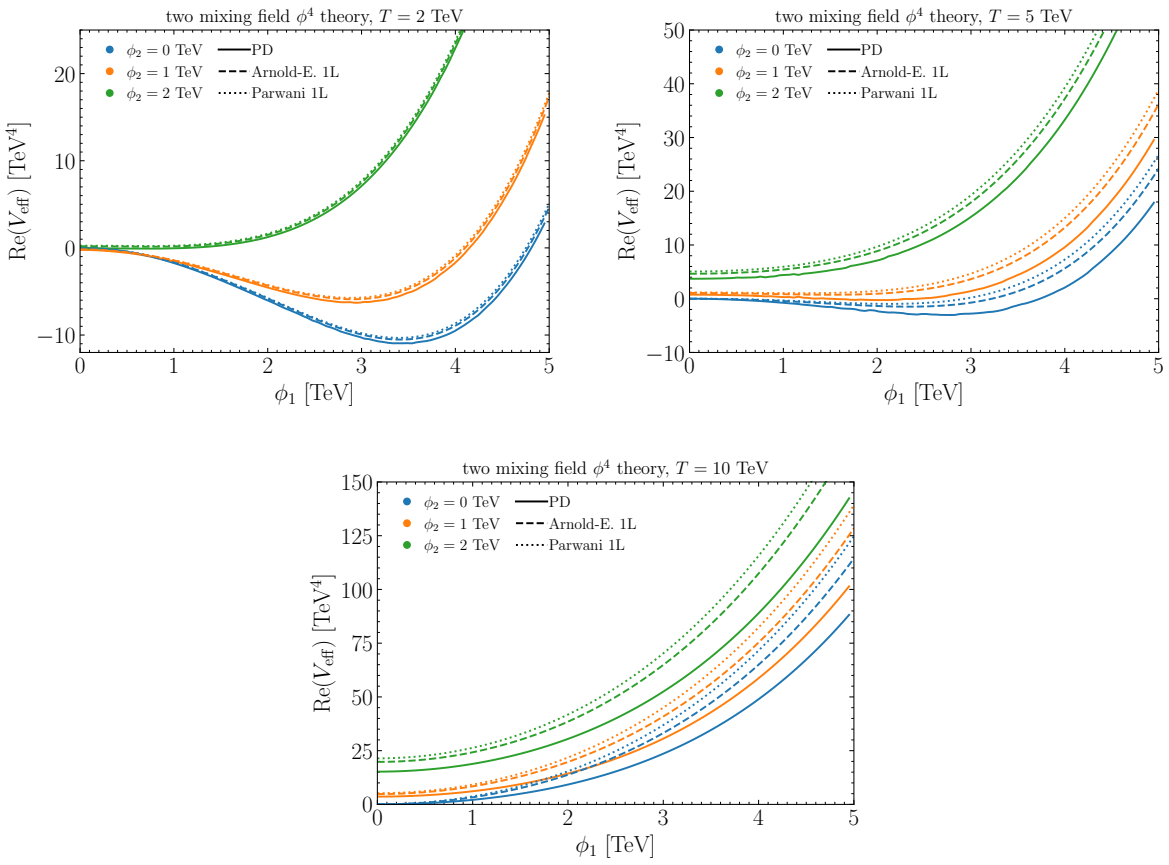

Figure 4.11: Real part of the effective potential in the two-field ϕ^4 theory with mixing as a function of ϕ_1 for different values of ϕ_2. The temperature is chosen to be 2 TeV (upper left), 5 GeV (upper right), and 10 GeV (bottom). Three different resummation methods are compared: partial dressing (solid), Arnold-Espinosa (dashed), and Parwani (dotted).

since for low temperatures, the difference in the thermal masses and also the thermal corrections to the effective potential are small. For higher temperatures comparable to the zero-temperature masses of ϕ_1 and ϕ_2, visible differences between Arnold-Espinosa/Parwani resummation and partial dressing arise. In this regime, the high-temperature expansion for the calculation of the thermal masses is not a good approximation (see discussion above). Moreover, partial dressing includes the resummation of subleading thermal corrections, which give a sizeable contribution at the considered parameter point. This is evident from the fact that the two different Parwani and Arnold-Espinosa resummation show visible differences indicating the importance of subleading thermal corrections. For even larger temperatures (see bottom panel), the differences between the various resummation methods are further increased. While the thermal masses are in slightly better agreement, the subleading thermal corrections have a bigger impact resulting in an overall larger difference between the three resummation methods.

We finally note again that partial dressing fails to correctly reproduce two-loop sunset contributions. However, the numerical impact of this effect is significantly smaller than the difference between partial dressing and Arnold-Espinosa/Parwani resummation.

4.2.6 Summary

Accurate predictions for phase transitions are very important for the investigation of a wide range of physics phenomena. This necessitates a precise calculation of the effective potential at finite temperatures implying the need to resum large thermal corrections. In this section, we reviewed various resummation methods focusing on partial dressing and truncated full dressing (Arnold-Espinosa/Parwani resummation). Using a scalar toy model, we discussed at the one- and two-loop level that partial dressing is advantageous since it does not rely on the high-temperature expansion and also resums subleading thermal corrections.

We then turned to a toy model for EWSNR, for which large thermal corrections are

expected implying the need to include subleading corrections. While these are automatically included if using partial dressing, higher-loop corrections have to be explicitly calculated if using Arnold-Espinosa/Parwani resummation. We, moreover, demonstrated that the inclusion of two-loop corrections in the Arnold-Espinosa/Parwani approaches leads to unphysical kinks in the prediction for the effective potential. These kinks originate from the occurrence of large imaginary contributions to the effective potential caused by negative mass squares. While these contributions largely cancel in the partial dressing approach, the cancellation is incomplete in the Arnold-Espinosa/Parwani resummation approaches due to the unavoidable high-temperature expansion in parts of the calculation.

While the discussion of EWSNR concentrates on the case in which only one of the fields takes a non-zero value (implying the absence of mixing between the scalar fields), we then focused on the case of mixing fields. So far, it has not been known how to apply partial dressing for mixing fields severely constraining its applicability. We presented a novel method which lifts this restriction. Promoting the gap equation to a matrix equation and performing a path integration in the multi-dimensional field space allowed us to consistently implement partial dressing even for mixing fields. This novel approach will allow the application of partial dressing for many BSM extensions of the SM Higgs sector resulting in more reliable predictions for the effective potential.

4.2.7 Supplementary Materials: Loop Functions

In this Appendix, we collect the various thermal loop functions used in this paper. We start with the bosonic thermal loop function appearing in the one-loop effective potential, which is given by

$$J_B(y^2) = \int_0^\infty dx\, x^2 \ln\left(1 - e^{-\sqrt{x^2+y^2}}\right) . \tag{4.178}$$

In the limits of small and large argument, this admits expansions

$$J_B(y^2 \ll 1) \simeq -\frac{\pi^4}{45} + \frac{\pi^2}{12}y^2 - \frac{\pi}{6}y^3 - \frac{1}{32}y^4 \log\Big(y^2/a_B\Big) + \mathcal{O}(y^6)\,, \tag{4.179}$$

$$J_B(y^2 \gg 1) \simeq -\sum_{n=1} \frac{1}{n^2} y^2 K_2(yn)\,, \tag{4.180}$$

where $a_B = 16\pi^2 e^{3/2-2\gamma_E}$ and K_2 is the modified Bessel function of the second kind.

In addition to the J integral, we also need the one-loop vacuum integrals with up to three vertices. The one-vertex integral is defined by

$$\mathcal{I}[m] \equiv \mathcal{I}_1[m] \equiv \sum\!\!\!\!\!\!\int_K \frac{1}{K^2 + m^2}\,. \tag{4.181}$$

Its high-temperature expansion is given by

$$\mathcal{I}[m] \simeq \frac{1}{12}T^2 - \frac{1}{4\pi}mT - \frac{L_R}{16\pi^2}m^2 + \frac{\zeta(3)}{128\pi^4}\frac{m^4}{T^2}\,. \tag{4.182}$$

Note that we have included the $\mathcal{O}(T^2)$ constant as well as kept terms up to $\mathcal{O}(1/T^2)$, since these lead to field-dependent contributions in $V^A_{\text{2-loop}} \sim \mathcal{I}[m]^2$ of $\mathcal{O}(T^0)$. There are also terms $\mathcal{O}(1/\epsilon)$ and $\mathcal{O}(\epsilon)$ in this expansion, which we omit for simplicity but which will nevertheless give a finite contribution to $V^A_{\text{2-loop}} \sim \mathcal{I}[m]^2$. See Ref. [296] for the complete expression.

The vacuum integral with two propagators is defined via

$$\mathcal{I}_2[m_1, m_2] \equiv \sum\!\!\!\!\!\!\int_K \frac{1}{K^2 + m_1^2}\frac{1}{K^2 + m_2^2}\,. \tag{4.183}$$

If both masses are different, we can write $\mathcal{I}_2$ in terms of $\mathcal{I}$ integrals

$$\mathcal{I}_2[m_1, m_2] = \frac{\mathcal{I}[m_2] - \mathcal{I}[m_1]}{m_1^2 - m_2^2}\,. \tag{4.184}$$

The high-temperature expansion follows immediately,

$$\mathcal{I}_2[m_1, m_2] \simeq \frac{1}{4\pi} \frac{T}{m_1 + m_2} + \dots \,. \tag{4.185}$$

We define the three-propagator vacuum integral via

$$\mathcal{I}_3[m_1, m_2, m_3] \equiv \sum\!\!\!\!\!\!\int_K \frac{1}{K^2 + m_1^2} \frac{1}{K^2 + m_2^2} \frac{1}{K^2 + m_3^2} \,. \tag{4.186}$$

If all masses (or a subset) are different, we can write it in terms of $\mathcal{I}$ and $\mathcal{I}_2$ functions

$$\begin{aligned} \mathcal{I}_3[m_1, m_2, m_3] = & \frac{\mathcal{I}[m_1]}{(m_1^2 - m_2^2)(m_1^2 - m_3^2)} - \frac{\mathcal{I}[m_2]}{(m_1^2 - m_2^2)(m_2^2 - m_3^2)} \\ & + \frac{\mathcal{I}[m_3]}{(m_1^2 - m_3^2)(m_2^2 - m_3^2)} \,, \end{aligned} \tag{4.187}$$

$$\mathcal{I}_3[m_1, m_1, m_2] = - \frac{\mathcal{I}[m_1]}{(m_1^2 - m_2^2)^2} + \frac{\mathcal{I}[m_2]}{(m_1^2 - m_2^2)^2} - \frac{\mathcal{I}_2[m_1, m_1]}{m_1^2 - m_2^2} \,. \tag{4.188}$$

In addition to the one-loop integrals, also the two-loop bosonic sunset diagram appears. For three arbitrary masses — m_1, m_2, m_3 —, it is defined by

$$\mathcal{H}[m_1, m_2, m_3] = \sum\!\!\!\!\!\!\int_P \sum\!\!\!\!\!\!\int_Q \frac{1}{(P^2 + m_1^2)(Q^2 + m_2^2)((P+Q)^2 + m_3^2)} \,. \tag{4.189}$$

The high-temperature expansion of the bosonic sunset is rather involved and has been eval-

uated in [317],

$$\begin{aligned}\mathcal{H}[m_1,m_2,m_3] \simeq{}& \frac{T^2}{16\pi^2}\left[\ln\left(\frac{\mu_R}{m_1+m_2+m_3}\right)+\frac{1}{2}\right]\\ &-\frac{T}{64\pi^3}\left[\sum_{i=1,2,3} m_i\left(\ln\left(\frac{\mu_R^2}{4m_i^2}\right)+L_R+2\right)\right]\\ &-\frac{1}{256\pi^4}\left[\sum_{i=1,2,3} m_i^2\left(L_R^2+L_R-2\gamma_E^2-4\gamma_1+\frac{\pi^2}{4}+\frac{3}{2}\right)\right]+\dots\,,\end{aligned} \tag{4.190}$$

with $\gamma_1 \simeq -0.0728$ the first Stieltjes constant.

We also need the bosonic sunset integral with one additional propagator, which we denote by $\widetilde{\mathcal{H}}$,

$$\widetilde{\mathcal{H}}[m_1,m_2,m_3,m_4] = \sum\!\!\!\!\!\!\int_P \sum\!\!\!\!\!\!\int_Q \frac{1}{(P^2+m_1^2)(P^2+m_2^2)(Q^2+m_3^2)((P+Q)^2+m_4^2)}\,. \tag{4.191}$$

If $m_1 \neq m_2$, it can be related to the normal sunset integral via

$$\widetilde{\mathcal{H}}[m_1,m_2,m_3,m_4] = \frac{1}{m_1^2-m_2^2}\left(\mathcal{H}[m_2,m_3,m_4]-\mathcal{H}[m_1,m_3,m_4]\right) \tag{4.192}$$

For $m_1 = m_2$, it can be derived by a derivative of the normal sunset integral

$$\widetilde{\mathcal{H}}[m_1,m_1,m_3,m_4] = -\frac{\partial}{\partial m_1^2}\mathcal{H}[m_1,m_3,m_4] \tag{4.193}$$

The high-temperature expansion of $\widetilde{\mathcal{H}}$ is then given by

$$\widetilde{\mathcal{H}}[m_1,m_2,m_3,m_4] = \frac{T^2}{16\pi^2}\frac{1}{m_1^2-m_2^2}\ln\left(\frac{m_1+m_3+m_4}{m_2+m_3+m_4}\right)+\dots\,, \tag{4.194}$$

$$\widetilde{\mathcal{H}}[m_1,m_1,m_3,m_4] = \frac{T^2}{32\pi^2}\frac{1}{m_1}\frac{1}{m_1+m_3+m_4}+\dots\,. \tag{4.195}$$

4.3 Constraints on 2 Higgs Doublet Models

Two Higgs doublet models (2HDM) provide the low energy effective theory description in many well motivated extensions of the Standard Model. Further, this extension may lead to a strongly first-order electroweak phase transition and possibly even successful baryogenesis [318, 319]. It is therefore relevant to study their properties, as well as the theoretical constraints on these models. In this article we concentrate on three relevant requirements for the validity of the 2HDM framework, namely the perturbative unitarity bounds, the bounded from below constraints, and the vacuum stability constraints. In this section, we concentrate on the most general renormalizable version of the 2HDM — without imposing any parity symmetry, which may be violated in many UV extensions. We derive novel analytical expressions that generalize those previously obtained in more restrictive scenarios to the most general case. We also discuss the phenomenological implications of these bounds, focusing on CP violation. Note that much of the material in this section was previously published by this author in [27].

4.3.1 Overview

The Standard Model (SM) [320] relies on the introduction of a Higgs doublet, whose vacuum expectation value breaks the electroweak symmetry [321, 322, 323]. This mechanism generates masses for the weak gauge bosons and charged fermions, as well as potentially the neutrinos (although there may be other mass sources for the latter). The Standard Model Higgs sector is the simplest way of implementing the Higgs mechanism for generating the masses of the known elementary particles. However, it is not the only possibility, and may be easily extended to the case of more than one Higgs doublet without violating any of the important properties of the SM. Moreover, one of the simplest of these extensions — two Higgs doublet models (2HDMs) [324] — appears as a low-energy effective theory of many well motivated extensions of the Standard Model (e.g. those based on supersymme-

try [325, 326, 327, 328, 329, 330, 331, 332, 333, 334] or little Higgs [335]).

Two Higgs doublet models may differ in the mechanism of generation of fermion masses. If both Higgs doublets couple to fermions of a given charge, their couplings will be associated to two different, complex sets of Yukawa couplings, which would form two different matrices in flavor space. The fermion mass matrices would be the sum of these, each multiplied by the corresponding Higgs vacuum expectation value. So diagonalization of the fermion mass matrices does not lead to the diagonalization of the fermion Yukawa matrices. Such theories are then associated with large flavor violating couplings of the Higgs bosons at low energies — a situation which is experimentally strongly disfavored. Hence, it is usually assumed that each charged fermion species couples only to one of the two Higgs doublets. In most works related to 2HDM, this is accomplished by implementing a suitable $\mathbb{Z}_2$ symmetry. The different possible charge assignments for this $\mathbb{Z}_2$ symmetry then fix the Higgs–fermion coupling choices and define different types of 2HDMs.

This $\mathbb{Z}_2$ symmetry not only fixes the Higgs–fermion couplings but also forbids certain terms in the Higgs potential that are far less problematic with respect to flavor violation. As a starting point for an investigation of the phenomenological implications of these terms, we will in this work discuss the theoretical bounds on the boson sector of the theory (without any need to specify the nature of the Higgs-fermion couplings). We will concentrate on the constraints that come from the perturbative unitarity of the theory, the stability of the physical vacuum, and the requirement that the effective potential is bounded from below. Existing works [336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 324, 348] focus either on the $\mathbb{Z}_2$-symmetric case or only provide a numerical procedure to test these constraints in the general 2HDM (see [349] for a recent work on analytic conditions for boundedness-from-below). We will go beyond current studies by deriving analytic bounds that apply to the most general, renormalizable realization of 2HDMs. Our conditions will be given in terms of the mass parameters and dimensionless couplings of the 2HDM tree-level potential. At

the quantum level, however, these parameters are scale dependent; although we will refrain from doing so here, one can apply these conditions at arbitrarily high energy scales by using the renormalization group evolution of these parameters.

Our article is organized as follows. In Sec. 4.3.2 we introduce the scalar sector of the most general 2HDM that defines the framework for most of the work presented in this article. Sec. 4.3.3 reviews three theorems from linear algebra which will allow us to derive analytic bounds in the coming sections. In Sec. 4.3.4, we concentrate on the requirement of perturbative unitarity. Sec. 4.3.5 presents the bounds coming from the requirement that the tree-level potential be bounded from below. In Sec. **??**, we discuss the vacuum stability. Finally, we reserve Sec. 4.3.7 for a brief analysis of the phenomenological implications (focusing on CP violation) and Sec. 4.3.8 for our conclusions. A Table listing the most relevant findings of our work may be found at the beginning of Sec. 4.3.8.

4.3.2 The General 2HDM

As emphasized above, we focus on the scalar sector of the theory. In general, gauge invariance implies that the potential can only include bilinear and quartic terms. Each of the three bilinear terms has a corresponding mass parameter, of which two (m_{11}^2 and m_{22}^2) are real while the third, m_{12}^2, is associated with a bilinear mixing of both Higgs doublets and may be complex.

Regarding the quartic couplings in the scalar potential, the two associated with self interactions of each of the Higgs fields, λ_1 and λ_2, must be real and, due to vacuum stability, positive. There are two couplings associated with Hermitian combinations of the Higgs fields, λ_3 and λ_4, which must be real, though not necessarily positive. The coupling λ_5 is associated with the square of the gauge invariant bilinear of both Higgs fields, and it may therefore be complex. The couplings λ_6 and λ_7 are associated with the product of Hermitian bilinears of each of the Higgs fields with the gauge invariant bilinear of the two Higgs fields, and, as

with λ_5, they may be complex. The most general scalar potential for a complex 2HDM is therefore:

$$\begin{aligned} V = & m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - (m_{12}^2 \Phi_1^\dagger \Phi_2 + h.c.) \\ & + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 (\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2) + \lambda_4 (\Phi_1^\dagger \Phi_2)(\Phi_2^\dagger \Phi_1) \\ & + \left[\frac{\lambda_5}{2} (\Phi_1^\dagger \Phi_2)^2 + \lambda_6 (\Phi_1^\dagger \Phi_1)(\Phi_1^\dagger \Phi_2) + \lambda_7 (\Phi_2^\dagger \Phi_2)(\Phi_1^\dagger \Phi_2) + h.c. \right] , \end{aligned} \tag{4.196}$$

with $\Phi_{1,2} = (\Phi_{1,2}^+, \Phi_{1,2}^0)^T$ being complex $SU(2)$ doublets with hypercharge $+1$.

One way to prevent Higgs-induced flavor violation in the fermion sector is to introduce a $\mathbb{Z}_2$ parity symmetry under which each charged fermion species transforms as even or odd. The Higgs doublets are assigned opposite parities and couple only to those charged fermions that carry their own parity. In such a scenario, the terms accompanying the couplings λ_6 and λ_7 would violate parity symmetry and hence should vanish. The mass parameter m_{12}^2 is also odd under the parity symmetry but induces only a soft breaking of this symmetry, which does not affect the ultraviolet properties of the theory. Thus it is usually allowed.

There are alternative ways of suppressing flavor violating couplings of the Higgs to fermions which do not rely on a simple parity symmetry and hence allow for the presence of λ_6 and λ_7 terms. One example is the flavor-aligned 2HDM [350]. Alternatively, one can impose a parity symmetry in the ultraviolet but allow the effective low energy field theory to be affected by operators generated by the decoupling of a sector where this symmetry is broken softly by dimensionful couplings which do not respect the parity symmetry properties. One example of such a theory is the NMSSM in the presence of heavy singlets, as discussed in Ref. [351]. In this case, the presence of the couplings λ_6 and λ_7 is essential to allow for the alignment of the light Higgs boson with a SM-like Higgs, leading to a good agreement with precision Higgs physics even in the case of large Higgs self couplings.

So we see that it is not necessary to restrict to the $\mathbb{Z}_2$-symmetric 2HDM with vanishing

λ_6 and λ_7 to avoid Higgs induced flavor violation in the fermion sector.[17] Further, it is phenomenologically interesting to study the 2HDM in full generality with these terms present. One consequence would be the possibility of having charge-parity (CP) violation in the bosonic sector. Indeed, to keep good agreement with Higgs precision data [352, 353], one is normally interested in studying 2HDM in (or close to) the exact alignment limit — the limit in which one of the neutral scalars carries the full vacuum expectation value and has SM-like tree-level couplings [354, 355, 356, 357, 358, 359]. If one imposes exact alignment in the $\mathbb{Z}_2$-symmetric 2HDM, however, CP is necessarily conserved, as will be explained in detail in Sec. 4.3.7. In the full 2HDM, on the other hand, one can have CP violation whilst remaining in exact alignment thanks to the presence of λ_6 and λ_7 terms. This CP violation could manifest in the neutral scalar mass eigenstates as well as bosonic couplings (see also [360]), providing many potential experimental signatures. With this motivation in mind, we keep λ_6 and λ_7 non-zero throughout this work.

4.3.3 Methods for Bounding Matrix Eigenvalues

In this work, much of the analysis of perturbative unitarity and vacuum stability involves placing bounds on matrix eigenvalues. In the most general 2HDM, analytic expressions for these constraints are either very complicated or simply can not be formulated. To obtain some analytic insight, we derive conditions which are either necessary or sufficient. Their derivation is based on three linear algebra theorems which we briefly review here.

Frobenius Norm

One may derive a bound on the magnitude of the eigenvalues of a matrix using the matrix norm. The following definition and theorem are needed:

17. Note that non-vanishing $\lambda_{6,7}$ induces flavor violation via Higgs mixing. This effect is, however, loop suppressed.

Theorem: *The magnitude of the eigenvalues e_i of a square matrix A are bounded from above by the matrix norm:* $|e_i| \leq ||A||$.

where a matrix norm is defined as:

Definition: *Given two $m \times n$ matrices A and B, the matrix norm $||A||$ satisfies the following properties:*

- $||A|| \geq 0$,
- $||A|| = 0 \Leftrightarrow A = 0_{m,n}$,
- $||\alpha A|| = |\alpha| ||A||$,
- $||A + B|| \leq ||A|| + ||B||$.

The above theorem holds for any choice of matrix norm, and thus one may employ the Frobenius norm [361], $||A|| = \sqrt{\text{Tr}(A^\dagger A)}$, to find the following result:

$$|e_i| \leq \sqrt{\text{Tr}(A^\dagger A)} \tag{4.197}$$

This bound on the eigenvalues will be used to derive sufficient bounds in the following sections.

Gershgorin Disk Theorem

We will use the Gershgorin disk theorem [362] in upcoming sections to derive sufficient conditions for perturbative unitarity and vacuum stability of the 2HDM potential. The theorem is typically used to constrain the spectra of complex square matrices. The basic idea is that

one identifies each of the diagonal elements with a point in the complex plane and then constructs a disk around this central point, with the radius given by the sum of the magnitudes of the other $n-1$ entries of the corresponding row.[18] The theorem says that all eigenvalues must lie within the union of these disks. Formally, we have the following definition and theorem:

Definition: *Let A be a complex $n \times n$ matrix with entries a_{ij}, and let R_i be the sum of the magnitudes of the non-diagonal entries of the i^{th} row, $R_i = \sum_{j \neq i} |a_{ij}|$. Then the Gershgorin disk $D(a_{ii}, R_i)$ is defined as the closed disk in the complex plane centered on a_{ii} with radius R_i.*

Theorem: *Every eigenvalue of A lies within at least one such Gershgorin disk $D(a_{ii}, R_i)$.* This theorem can be used to derive an upper bound on the magnitude of the eigenvalues of a matrix. We will use this technique below when discussing perturbative unitarity and vacuum stability. Since all the matrices we will consider in the subsequent sections on perturbative unitarity and boundedness from below are Hermitian matrices, each eigenvalue will lie within the intervals formed by the intersection of the Gershgorin disks with the real axis.

We shall proceed in the following manner: We will first construct the intervals containing the eigenvalues of each matrix A. For each interval, the rightmost and leftmost endpoints $x_i^{\pm}$ will be given by the sum and difference, respectively, of the center and the radius,

$$x_i^{\pm} \equiv a_{ii} \pm R_i \,, \quad \text{with} \quad R_i = \sum_{j \neq i} |a_{ij}| \,. \tag{4.198}$$

We then identify which $x_i^{\pm}$ extends furthest in the positive or negative direction. We know

18. One can also construct the radius by summing the magnitudes of the $n-1$ column entries.

that every eigenvalue e_k must lie within the endpoints of the largest possible total interval,

$$\min(x_i^-) \leq e_k \leq \max(x_i^+)\,. \tag{4.199}$$

This may be rephrased into an upper bound on $|e_k|$ as:

$$|e_k| \leq \max_i \left(\sum_j |a_{ij}| \right), \tag{4.200}$$

where the left-hand side of Eq. (4.200) represents the absolute value of any given eigenvalue e_k and the right-hand side represents the maximum value of $\sum_j |a_{ij}|$ over all rows i of the matrix A. In fact, this condensed statement of Gershgorin circle theorem is an application of the matrix norm theorem, employing the norm $||A|| = \max_i(\sum_j |a_{ij}|)$.

Principal Minors

In order to derive necessary conditions, one may employ Sylvester's criterion in a clever way, as proposed in [363]. Sylvester's criterion involves the principal minors D_k of a matrix, where D_k is the determinant of the upper-left $k \times k$ sub-matrix. The statement of Sylvester's criterion is the following:

Theorem: *Let M be a Hermitian $n \times n$ matrix. M is positive definite if and only if all of the principal minors $D_k(M)$ are positive.*

We further need the following result about Hermitian matrices[19]:

Theorem: *Let M be a Hermitian matrix. Then M is positive definite if and only if all of*

19. This theorem is in fact used in the proof of Sylvester's criterion

its eigenvalues are positive.

One can apply this to derive an upper bound on the eigenvalues of a diagonalizable matrix in the following way:

Theorem: *Let M be an $n \times n$ diagonalizable, Hermitian (symmetric) matrix and let c be a positive real number. The eigenvalues e_i of M are bounded as $|e_i| < c$ if and only if all principal minors $D_k(c\mathbb{I} - M)$ and $D_k(c\mathbb{I} + M)$ are positive for all $k = 1...n$.*

To see this, consider applying a unitary transformation which diagonalizes M to the matrix $c\mathbb{I} \pm \text{diag}(M)$. Then for symmetric or Hermitian matrices, the statement that $c\mathbb{I} \pm M$ is positive definite becomes a statement on the relative values of e_i and c. In this manner, the application of Sylvester's criterion to these specific matrices allows one to put an upper bound on the magnitude of the eigenvalues without diagonalizing the matrix M. Note that for the absolute value $|e_i|$ to be bounded by c, we require the use of both $c\mathbb{I} \pm M$ matrices. On the other hand, if one has only an upper bound on e_i, i.e. $e_i < c$, as we will have in the case of vacuum stability, then one only requires the principal minors of the matrix $c\mathbb{I} - M$ to be positive.

We note that the use specifically of the upper-left sub-matrices in Sylvester's criterion is an arbitrary choice, and basis-dependent. One could instead consider the lower-right sub-matrices, or any matrices along the diagonal. As such, it is possible to derive further conditions using this criteria by further considering, for example, the upper-left, lower-right, and central 2×2 sub-matrices of a 4×4 matrix. We will do so in later analyses to strengthen the lower-k bounds.

This use of sub-determinants has been proposed in [363] as a method to increase the efficiency of parameter scans in models with large scattering matrices. For such theories (e.g.

the model with N Higgs doublets, NHDM, considered in [363] for higher N), the numerical calculation of the scattering matrix eigenvalues is computationally expensive. We note that the use of the Gershgorin disk theorem proposed in Eq. (4.3.3) provides an additional complementary method to speed-up parameter scans.

4.3.4 Perturbative Unitarity

Tree-level constraints for perturbative unitarity in the most general 2HDM have already been investigated in the literature [341, 364, 348]. However, for a non-zero λ_6 and λ_7, no exact analytic conditions have been obtained yet. Here, we will first review the existing literature and then derive analytic expressions for the case of non-vanishing λ_6 and λ_7.

Numerical Bound

Perturbative unitarity is usually imposed by demanding that the eigenvalues e_i of the scalar scattering matrix at high energy be less than the unitarity limit, $|e_i| < 8\pi$. Thus to derive the constraints on the quartic couplings, one must construct the scattering matrix for all physical scalar states.

We are interested in all processes $AB \to CD$, where the fields $A...D$ represent any combination of the physical[20] scalars $(H_1, H_2, H_3, H^\pm, W_L^\pm, Z_L)$. The interactions and hence S-matrix take a complicated form in terms of the physical states. However since we are only interested in the eigenvalues of the S-matrix, we may choose any basis related to the physical basis by a unitary transformation. The derivation is simplest in the basis of the original Higgs

20. Technically since we are working in the high energy limit, the equivalence theorem allows us to replace $W_L^\pm$ and Z_L by their corresponding Goldstone bosons.

fields $(w_i^{\pm}, h_i, z_i)$, appearing as

$$\Phi_i = \begin{pmatrix} w_i^+ \\ \frac{1}{\sqrt{2}}(v_i + h_i + iz_i) \end{pmatrix}, \tag{4.201}$$

with $v = \sqrt{\sum_i v_i^2} = 246$ GeV.

Out of these fields, we can construct 14 neutral two-body states: $\left|w_i^+ w_i^-\right\rangle$, $\frac{1}{\sqrt{2}}\left|z_i z_i\right\rangle$, $\frac{1}{\sqrt{2}}\left|h_i h_i\right\rangle$, $\left|h_i z_i\right\rangle$, $\left|w_1^+ w_2^-\right\rangle$, $\left|w_2^+ w_1^-\right\rangle$, $\left|z_1 z_2\right\rangle$, $\left|h_1 h_2\right\rangle$, $\left|z_1 h_2\right\rangle$, and $\left|h_1 z_2\right\rangle$. By constructing states which are linear combinations with definite hypercharge and total weak isospin, denoted by (Y, I), and grouping the ones with the same set of quantum numbers, the matrix of S-wave amplitudes a_0 takes a block diagonal form (for more details see [341, 364, 348]). For the neutral scattering channels, this is:

$$a_0^{(0)} = \frac{1}{16\pi} \begin{pmatrix} X_{(0,0)} & & & \\ & X_{(0,1)} & & \\ & & X_{(1,1)} & \\ & & & X_{(1,1)} \end{pmatrix}, \tag{4.202}$$

where the subscript of each submatrix denotes the quantum numbers (Y, I) of the corresponding states. The entries are:

$$X_{(0,0)} = \begin{pmatrix} 3\lambda_1 & 2\lambda_3 + \lambda_4 & 3\lambda_6 & 3\lambda_6^* \\ 2\lambda_3 + \lambda_4 & 3\lambda_2 & 3\lambda_7 & 3\lambda_7^* \\ 3\lambda_6^* & 3\lambda_7^* & \lambda_3 + 2\lambda_4 & 3\lambda_5^* \\ 3\lambda_6 & 3\lambda_7 & 3\lambda_5 & \lambda_3 + 2\lambda_4 \end{pmatrix}, \tag{4.203a}$$

$$X_{(0,1)} = \begin{pmatrix} \lambda_1 & \lambda_4 & \lambda_6 & \lambda_6^* \\ \lambda_4 & \lambda_2 & \lambda_7 & \lambda_7^* \\ \lambda_6^* & \lambda_7^* & \lambda_3 & \lambda_5^* \\ \lambda_6 & \lambda_7 & \lambda_5 & \lambda_3 \end{pmatrix}, \tag{4.203b}$$

$$X_{(1,1)} = \begin{pmatrix} \lambda_1 & \lambda_5 & \sqrt{2}\lambda_6 \\ \lambda_5^* & \lambda_2 & \sqrt{2}\lambda_7^* \\ \sqrt{2}\lambda_6^* & \sqrt{2}\lambda_7 & \lambda_3 + \lambda_4 \end{pmatrix}. \tag{4.203c}$$

For the 8 singly-charged two-body states $\left|w_i^+ z_i\right\rangle$, $\left|w_i^+ h_i\right\rangle$, $\left|w_1^+ z_2\right\rangle$, $\left|w_1^+ h_2\right\rangle$, $\left|w_2^+ z_1\right\rangle$, $\left|w_2^+ h_1\right\rangle$, the block diagonal 8×8 singly-charged S-matrix is given by:

$$a_0^{(+)} = \frac{1}{16\pi} \begin{pmatrix} X_{(0,1)} & & & \\ & X_{(1,0)} & & \\ & & X_{(1,1)} & \end{pmatrix}, \tag{4.204}$$

where the new entry $X_{(1,0)}$ is just the one-dimensional eigenvalue:

$$X_{(1,0)} = \lambda_3 - \lambda_4 \,. \tag{4.205}$$

Finally, the 3×3 S-matrix for the three doubly-charged 2-body states $\left|w_i^+ w_i^+\right\rangle$, $\left|w_1^+ w_2^+\right\rangle$ is given by:

$$a_0^{(++)} = \frac{1}{16\pi} X_{(1,1)} \,. \tag{4.206}$$

We impose perturbative unitarity by demanding that the eigenvalues of the scattering matrix are smaller than 8π implying that $|a_0| < \frac{1}{2}$. Indeed, the eigenvalues of the submatrices $X_{(0,0)}$, $X_{(0,1)}$, $X_{(1,0)}$, and $X_{(1,1)}$, which we denote as e_i, must all satisfy

$$|e_i| < 8\pi \,. \tag{4.207}$$

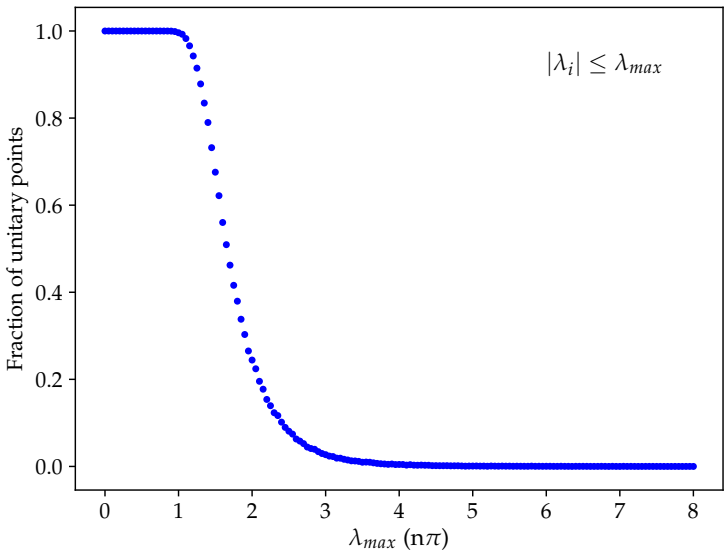

Figure 4.12: Plot showing the fraction of points that pass the unitarity bound $|e_i| < 8\pi$ for different choices of λ_{max}, in units of multiples of π. The values of the λ_i are each chosen randomly such that $|\lambda_i| < \lambda_{max}$. We test 20,000 random sets of λ_i for each λ_{max}.

Obtaining analytic expressions for the eigenvalues requires solving cubic and quartic equations, and the result is complicated and not very useful. Given a choice of input parameters, however, it is easy to check this condition numerically.

Assuming all $\lambda_{1...7}$ to be equal, the strongest constraint arises from the 4×4 matrix X_{00}. If we set all $\lambda_i \equiv \lambda$ and solve for the eigenvalues, we find the bound:

$$\lambda < \frac{2\pi}{3}\,. \tag{4.208}$$

This value is an order of magnitude smaller than 8π, implying that if all quartic couplings are sizable (i.e. of $\mathcal{O}(1)$), perturbative unitarity may be lost even at values of the couplings much smaller than 4π, which is a bound often encountered in the literature to ensure perturbativity.

We investigate the validity of such an upper bound further in Fig. 4.12. For this figure, we randomly choose each λ_i within the range $|\lambda_i| < \lambda_{\rm max}$ and then show the fraction of test points which pass the numerical unitarity constraint, as a function of $\lambda_{\rm max}$. For $\lambda_{\rm max} \lesssim \pi$ almost all points survive the perturbative unitarity constraint. For larger $\lambda_{\rm max}$ values the

survival rate quickly drops to almost zero for $\lambda_{\max} \gtrsim 4\pi$. This highlights again that the simplified perturbativity bound of $|\lambda_i| < 4\pi$, which is often encountered in the literature, is too loose. Based on the results in Fig. 4.12, a better choice of bound might be $|\lambda_i| \lesssim \pi$ or $|\lambda_i| \lesssim 3\pi/2$.

A Necessary Condition for Perturbative Unitarity

To gain some intuition for the perturbative unitarity constraint, we now turn to derive some simplified analytic conditions which are either necessary or sufficient, though not both. In this section we will focus on the former, which can be used to quickly rule out invalid parameter sets which violate perturbative unitarity. One can derive necessary conditions by invoking the method of principal minors, which we reviewed previously and can be used to give an upper bound on the maximal value of the eigenvalues of a Hermitian matrix. Since the scattering matrices are Hermitian, demanding the eigenvalues to be bounded as $|e_i| < 8\pi$, as required by perturbative unitarity, amounts to requiring

$$D_k(8\pi\mathbb{I} + X) > 0 \text{ and } D_k(8\pi\mathbb{I} - X) > 0\,, \tag{4.209}$$

for $k = 1, 2, 3, 4$. Since satisfying both criteria for all $k = 1, 2, 3, 4$ is a necessary and sufficient condition, any single k condition provides a necessary condition.

Since the eigenvalues of $X_{(0,0)}$ are generically the largest and therefore the most constraining, we will focus on bounds coming from this matrix. We begin with the upper left 2×2 submatrices. Taking the determinant, we have

$$D_2^L(8\pi\mathbb{I} + X_{(0,0)}) > 0 \;\Rightarrow\; 64\pi^2 + 24\pi(\lambda_1 + \lambda_2) + 9\lambda_1\lambda_2 - (2\lambda_3 + \lambda_4)^2 > 0\,, \tag{4.210a}$$

$$D_2^L(8\pi\mathbb{I} - X_{(0,0)}) > 0 \;\Rightarrow\; 64\pi^2 - 24\pi(\lambda_1 + \lambda_2) + 9\lambda_1\lambda_2 - (2\lambda_3 + \lambda_4)^2 > 0\,. \tag{4.210b}$$

Clearly the latter constraint coming from $D_2^L(8\pi\mathbb{I} - X)$ will be the stronger of the two, since $\lambda_1, \lambda_2 > 0$ if boundedness from below is imposed. Thus the necessary $k = 2$ condition reduces to Eq. (4.210b). We also examine the constraints that arise from using the lower-right and center 2×2 sub-matrices, as proposed in Eq. (4.3.3). The analytic conditions from the lower-right D^R and center D^C sub-matrices are, respectively,

$$D_2^R(8\pi\mathbb{I} - X_{(0,0)}) = 64\pi^2 + (\lambda_3 - 16\pi)\lambda_3 + (4(\lambda_3 + \lambda_4) - 32\pi)\lambda_4 - 9|\lambda_5|^2 \quad > 0\,, \quad (4.211a)$$

$$D_2^C(8\pi\mathbb{I} - X_{(0,0)}) = 64\pi^2 - 8\pi(\lambda_3 + 3\lambda_2 + 2\lambda_4) + 3\lambda_3\lambda_2 + 6\lambda_2\lambda_4 - 9|\lambda_7|^2 \quad > 0\,. \quad (4.211b)$$

The combination of these three expressions provides a stronger constraint than just the upper-left minor constraint alone.

While it is immediately clear that $D_2^L(8\pi\mathbb{I} - X)$ is stronger than the addition-based bound for the upper-left matrix, it is not clear for the center and lower-right matrices; in fact, including these bounds provides a slightly more constraining result. We thus employ the $D_2^{C,R}(8\pi\mathbb{I} + X)$ constraints in our analysis of the $k = 2$ bound as well:

$$D_2^R(8\pi\mathbb{I} + X_{(0,0)}) = 64\pi^2 + (\lambda_3 + 16\pi)\lambda_3 + (4(\lambda_3 + \lambda_4) + 32\pi)\lambda_4 - 9|\lambda_5|^2 \quad > 0\,, \quad (4.212a)$$

$$D_2^C(8\pi\mathbb{I} + X_{(0,0)}) = 64\pi^2 + 8\pi(\lambda_3 + 3\lambda_2 + 2\lambda_4) + 3\lambda_3\lambda_2 + 6\lambda_2\lambda_4 - 9|\lambda_7|^2 \quad > 0\,. \quad (4.212b)$$

Next, we look at the upper left 3×3 submatrices. Unlike the $k = 2$ case, it is not clear that one of these is generically more constraining than the other. To be consistent with the $k = 2$ case, we will examine the $8\pi\mathbb{I} - X_{(0,0)}$ matrix. We additionally consider the lower-right 3×3 sub-matrix. These give the following bounds:

$$\begin{aligned} D_3^L(8\pi\mathbb{I} - X_{(0,0)}) =& (8\pi - \lambda_3 - 2\lambda_4)((8\pi - 3\lambda_1)(8\pi - 3\lambda_2) - (2\lambda_3 + \lambda_4)^2) \\ & - 9(8\pi - 3\lambda_2)|\lambda_6|^2 - 9(8\pi - 3\lambda_1)|\lambda_7|^2 \\ & - 9(2\lambda_3 + \lambda_4)(\lambda_6\lambda_7^* + \lambda_6^*\lambda_7) > 0\,, \end{aligned} \quad (4.213a)$$

$$
\begin{aligned}
D_3^R(8\pi\mathbb{I} - X_{(0,0)}) =&(8\pi - \lambda_3 - 2\lambda_4)^2(8\pi - 3\lambda_2) - 9|\lambda_5|^2(8\pi - 3\lambda_2) \\
&- 27(\lambda_5^*\lambda_7^2 + \lambda_5(\lambda_7^*)^2) - 18(8\pi - \lambda_3 - 2\lambda_4)|\lambda_7|^2 > 0\,.
\end{aligned} \tag{4.213b}
$$

Meanwhile, considering $8\pi\mathbb{I} + X_{(0,0)}$ gives

$$
\begin{aligned}
D_3^L(8\pi\mathbb{I} + X_{(0,0)}) =&(8\pi + \lambda_3 + 2\lambda_4)((8\pi + 3\lambda_1)(8\pi + 3\lambda_2) - (2\lambda_3 + \lambda_4)^2) \\
&- 9(8\pi + 3\lambda_2)|\lambda_6|^2 - 9(8\pi + 3\lambda_1)|\lambda_7|^2 \\
&+ 9(2\lambda_3 + \lambda_4)(\lambda_6\lambda_7^* + \lambda_6^*\lambda_7) > 0\,,
\end{aligned} \tag{4.214a}
$$

$$
\begin{aligned}
D_3^R(8\pi\mathbb{I} + X_{(0,0)}) =&(8\pi + \lambda_3 + 2\lambda_4)^2(8\pi + 3\lambda_2) - 9|\lambda_5|^2(8\pi + 3\lambda_2) \\
&+ 27(\lambda_5^*\lambda_7^2 + \lambda_5(\lambda_7^*)^2) - 18(8\pi + \lambda_3 + 2\lambda_4)|\lambda_7|^2 > 0\,.
\end{aligned} \tag{4.214b}
$$

While the $D_3(8\pi\mathbb{I} - X_{(0,0)})$ provide the strongest constraints for values of $|\lambda_i| \lesssim 4\pi$, the inclusion of the $D_3(8\pi\mathbb{I} + X_{(0,0)})$ and D_2 bounds improves the performance of the $k = 3$ bounds at higher $|\lambda_i|$. We omit analytic expressions for the $k = 4$ case, since they cannot be simplified to a useful form. Moreover, the $k = 3$ expressions already provide constraints very close to the full numerical bound (see Fig. 4.13).

Sufficient Conditions for Perturbative Unitarity

Next, we turn to derive sufficient conditions for perturbative unitarity by applying the Gershgorin disk theorem, which gives an upper bound on the maximal value of the eigenvalues. By demanding that this upper bound is less than 8π, we obtain a sufficient condition for perturbative unitarity.

We first construct the intervals $x_i^{(Y,I)}$ containing the eigenvalues of each of the scattering matrices, $X_{(0,0)}$, $X_{(0,1)}$, $X_{(1,0)}$, and $X_{(1,1)}$. We know that in order to uphold perturbative unitarity, we must have $|e_i| < 8\pi$. Thus we arrive at the sufficient condition:

$$
\max(x_i^{(Y,I)}) < 8\pi\,. \tag{4.215}
$$

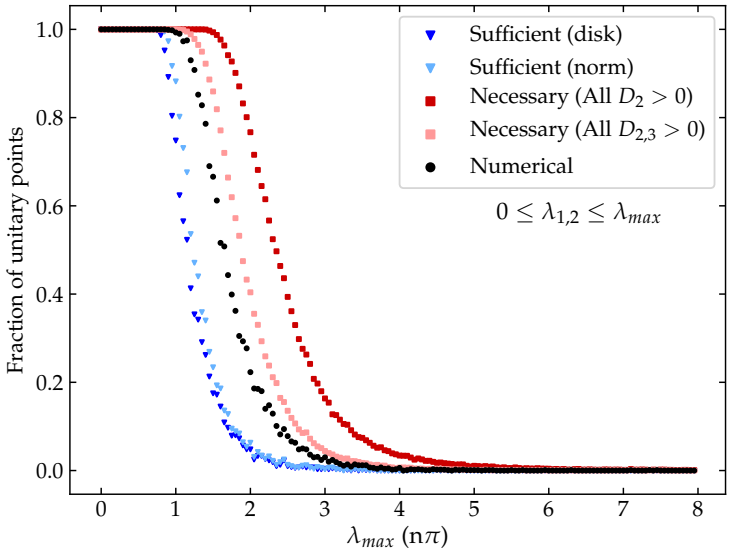

Figure 4.13: Plot comparing the number of points that pass the exact numerical bound $|e_i| < 8\pi$ (black), the sufficient bound from the Gershgorin disk theorem Eq. (4.220) (dark blue), the sufficient bound from the Frobenius norm Eq. (4.221) (light blue), the necessary condition $D_2(8\pi\mathbb{I} \pm X_{(0,0)}) > 0$ (dark red), and the necessary condition $D_{2,3}(8\pi\mathbb{I} \pm X_{(0,0)}) > 0$ (light red). The λ_i values are randomly chosen from the range of values satisfying $|\lambda_i| < \lambda_{max}$, where λ_{max} is given by the x-axis in units of multiples of π. The minimal bounded from below condition $\lambda_{1,2} \geq 0$ is enforced. The $\lambda_{5,6,7}$ values are allowed to be complex. The total number of points checked for each λ_{max} is 10,000.

For each of the X matrices, we can work out the $x_i^{(Y,I)}$ explicitly. For $X_{(0,0)}$, we obtain:

$$x_1^{(0,0)} = 3|\lambda_1| + (|2\lambda_3 + \lambda_4| + 6|\lambda_6|)\,, \tag{4.216a}$$

$$x_2^{(0,0)} = 3|\lambda_2| + (|2\lambda_3 + \lambda_4| + 6|\lambda_7|)\,, \tag{4.216b}$$

$$x_3^{(0,0)} = x_4^{(0,0)} = |\lambda_3 + 2\lambda_4| + 3(|\lambda_5| + |\lambda_6| + |\lambda_7|)\,. \tag{4.216c}$$

For $X_{(0,1)}$, they are:

$$x_1^{(0,1)} = |\lambda_1| + (|\lambda_4| + 2|\lambda_6|)\,, \tag{4.217a}$$

$$x_2^{(0,1)} = |\lambda_2| + (|\lambda_4| + 2|\lambda_7|)\,, \tag{4.217b}$$

$$x_3^{(0,1)} = x_4^{(0,1)} = |\lambda_3| + (|\lambda_5| + |\lambda_6| + |\lambda_7|)\,. \tag{4.217c}$$

For $X_{(1,1)}$, we have:

$$x_1^{(1,1)} = |\lambda_1| + (|\lambda_5| + \sqrt{2}|\lambda_6|)\,, \tag{4.218a}$$

$$x_2^{(1,1)} = |\lambda_2| + (|\lambda_5| + \sqrt{2}|\lambda_7|)\,, \tag{4.218b}$$

$$x_3^{(1,1)} = |\lambda_3 + \lambda_4| + \sqrt{2}(|\lambda_6| + |\lambda_7|)\,. \tag{4.218c}$$

Finally, $X_{(1,0)}$, we have:

$$x_1^{(1,0)} = |\lambda_3 - \lambda_4|\,. \tag{4.219}$$

In examining these conditions, the leading coefficient of 3 in the first set suggests that the $x_i^{(0,0)}$ corresponding to $X_{(0,0)}$ will generically be larger than those corresponding to $X_{(0,1)}$, $X_{(1,0)}$, and $X_{(1,1)}$; a numerical check confirms this intuition. Thus the sufficient condition

for perturbative unitarity simplifies slightly to

$$\max\left(x_i^{(0,0)}\right) < 8\pi\,. \tag{4.220}$$

One may alternatively employ the bound arising from the Frobenius norm. Taking the dominant $X_{(0,0)}$ matrix, one finds the condition:

$$\sqrt{9(\lambda_1^2+\lambda_2^2)+10(\lambda_3^2+\lambda_4^2)+16\lambda_3\lambda_4+18(|\lambda_5|^2+2|\lambda_6|^2+2|\lambda_7|^2)} \leq 8\pi \tag{4.221}$$

The dependence here on the signs of the λ_i is similar to the dependence seen in the Gershgorin disk conditions: the bound is not sensitive to the signs of any λ_i except for the relative sign between λ_3 and λ_4.

Numerical Comparison

In order to compare the various bounds derived in this section, in Fig. 4.13 we plot the number of points which pass the exact, sufficient, and necessary conditions for different values of λ_{max}. For each λ_{max}, we consider 10,000 randomly-drawn values for the λ_i within the range $|\lambda_i| < \lambda_{max}$. For the necessary conditions, the results are derived from the combination of all possible 2×2 (3×3) sub-matrices along the diagonal for the $D_{2(3)}$ bound. We enforce the minimal bounded from below condition $\lambda_{1,2} > 0$, which has been derived e.g. in [324]. We find that the necessary $D_3(8\pi\mathbb{I} - X_{(0,0)})$ condition lies very close to the exact condition and is effective at ruling out parameter sets which fail perturbative unitarity, while for $\lambda_{max} \lesssim \pi$ all tested points satisfy perturbative unitarity.

4.3.5 Boundedness from Below

Next, we seek to determine the conditions on the parameters such that the potential of Eq. (4.196) is bounded from below (BFB). For this, it is necessary to ensure that the quartic

part of the potential does not acquire negative values. If negative values were present, one could easily find indefinite negative values of the potential by rescaling all fields to infinity in the same direction as the one in which the negative value was found. We remark that analytic expressions have been formulated previously in [349], though for the case of explicit CP conservation. We will make no such assumption. There are also previous analyses of the BFB condition using the eigenvalues of a 4×4 matrix ([343]); however, these analyses do not lead to analytical expressions, and we will follow an alternative approach.

We begin by reparameterizing the potential via $\Phi_1^\dagger\Phi_1 = \frac{1}{2}h_1^2$, $\Phi_2^\dagger\Phi_2 = \frac{1}{2}h_2^2$, $\Phi_1^\dagger\Phi_2 = \frac{1}{2}h_1h_2\rho e^{i\eta}$, with $\rho \in [0,1]$. Moreover, we decompose the complex couplings λ_5, λ_6, λ_7 into real and imaginary parts as: $\lambda_i e^{i\eta} + \lambda_i^* e^{-i\eta} = 2\mathrm{Re}[\lambda_i]\cos\eta - 2\mathrm{Im}[\lambda_i]\sin\eta$. The quartic part of the potential then becomes:

$$\begin{aligned} V_{\text{quartic}} = \frac{1}{4}\Big\{ & \frac{\lambda_1}{2}h_1^4 + \frac{\lambda_2}{2}h_2^4 + \big[\lambda_3 + \big(\lambda_4 + \mathrm{Re}[\lambda_5]\cos 2\eta - \mathrm{Im}[\lambda_5]\sin 2\eta\big)\rho^2\big]h_1^2h_2^2 \\ & + 2\big(\mathrm{Re}[\lambda_6]\cos\eta - \mathrm{Im}[\lambda_6]\sin\eta\big)\rho\, h_1^3h_2 \\ & + 2\big(\mathrm{Re}[\lambda_7]\cos\eta - \mathrm{Im}[\lambda_7]\sin\eta\big)\rho\, h_1h_2^3\Big\}. \end{aligned} \tag{4.222}$$

We can then cast V_{quartic} into the form

$$V_{\text{quartic}} = \frac{1}{4}h_2^4\left[a\left(\frac{h_1}{h_2}\right)^4 + b\left(\frac{h_1}{h_2}\right)^3 + c\left(\frac{h_1}{h_2}\right)^2 + d\left(\frac{h_1}{h_2}\right) + e\right], \tag{4.223}$$

with

$$a = \frac{\lambda_1}{2}, \quad e = \frac{\lambda_2}{2}, \tag{4.224a}$$

$$b = 2\big(\mathrm{Re}[\lambda_6]\cos\eta - \mathrm{Im}[\lambda_6]\sin\eta\big)\rho, \tag{4.224b}$$

$$c = \Big[\lambda_3 + \big(\lambda_4 + \mathrm{Re}[\lambda_5]\cos 2\eta - \mathrm{Im}[\lambda_5]\sin 2\eta\big)\rho^2\Big], \tag{4.224c}$$

$$d = 2\big(\mathrm{Re}[\lambda_7]\cos\eta - \mathrm{Im}[\lambda_7]\sin\eta\big)\rho. \tag{4.224d}$$

Clearly $a > 0$ and $e > 0$ has to be fulfilled as a minimum condition for BFB. We can then divide out by $e\, h_2^4$ and define the simplified polynomial:

$$f(x) = x^4 + \alpha x^3 + \beta x^2 + \gamma x + 1\,, \tag{4.225}$$

with $x = \frac{a^{1/4}}{e^{1/4}} \frac{h_1}{h_2}$ and:

$$\alpha = b a^{-3/4} e^{-1/4}, \quad \beta = c a^{-1/2} e^{-1/2}, \quad \gamma = d a^{-1/4} e^{-3/4}\,. \tag{4.226}$$

We then define the following quantities:

$$\Delta = 4[\beta^2 - 3\alpha\gamma + 12]^3 - [72\beta + 9\alpha\beta\gamma - 2\beta^3 - 27\alpha^2 - 27\gamma^2]^2\,, \tag{4.227a}$$

$$\chi_1 = (\alpha - \gamma)^2 - 16(\alpha + \beta + \gamma + 2)\,, \tag{4.227b}$$

$$\chi_2 = (\alpha - \gamma)^2 - \frac{4(\beta + 2)}{\sqrt{\beta - 2}} \left(\alpha + \gamma + 4\sqrt{\beta - 2}\right)\,. \tag{4.227c}$$

The positivity of V_{quartic} is ensured if and only if one of the following conditions holds [365]:

$$\begin{aligned}
&(1)\ \ \beta < -2 \text{ and } \Delta \leq 0 \text{ and } \alpha + \gamma > 0\,,\\
&(2)\ \ -2 \leq \beta \leq 6 \text{ and } \Delta \leq 0 \text{ and } \alpha + \gamma > 0\,,\\
&(3)\ \ -2 \leq \beta \leq 6 \text{ and } \Delta \geq 0 \text{ and } \chi_1 \leq 0\,,\\
&(4)\ \ \beta > 6 \text{ and } \Delta \leq 0 \text{ and } \alpha + \gamma > 0\,,\\
&(5)\ \ \beta > 6 \text{ and } \alpha > 0 \text{ and } \gamma > 0\,,\\
&(6)\ \ \beta > 6 \text{ and } \Delta \geq 0 \text{ and } \chi_2 \leq 0\,.
\end{aligned} \tag{4.228}$$

If any of these conditions is true for a given set of input parameters $\lambda_{1\ldots7}$ and for all possible values of $\rho \in [0, 1]$, $\eta \in [0, 2\pi)$, then the potential is BFB.

Note that under the transformation $\eta \to \eta + \pi$, both α and γ are anti-symmetric ($\alpha \to$

$-\alpha$ and $\gamma \to -\gamma$). This in turn implies that conditions (1), (2), (4), and (5) are always violated for some value of η, and therefore can never guarantee the positivity of V_{quartic}. Consequently, we are left with only two conditions under which the potential is BFB:

$$\text{The potential is BFB if and only if}: \ \Delta \geq 0 \text{ and } \begin{cases} -2 \leq \beta \leq 6 \text{ and } \chi_1 \leq 0\,, \text{or} \\ \beta > 6 \text{ and } \chi_2 \leq 0\,. \end{cases} \tag{4.229}$$

Note that upon setting $\lambda_6 = \lambda_7 = 0$, and after extremizing with respect to η, Eq. (4.229) becomes

$$\beta + 2 \geq 0 \ \Rightarrow \ \lambda_3 + \rho^2(\lambda_4 - |\lambda_5|) \geq -\sqrt{\lambda_1 \lambda_2}\,. \tag{4.230}$$

This is a monotonic function of ρ, and hence the strongest constraints are derived for either $\rho = 1$ or $\rho = 0$, namely

$$\lambda_3 + \lambda_4 - |\lambda_5| \geq -\sqrt{\lambda_1 \lambda_2}\,, \ \text{ and} \tag{4.231}$$

$$\lambda_3 \geq -\sqrt{\lambda_1 \lambda_2}\,, \tag{4.232}$$

These conditions reproduce the well-known conditions for BFB in the $\mathbb{Z}_2$-symmetric 2HDM [324]. Let us also stress that, for $\rho = 0$, Eq. (4.229) leads to Eq. (4.232) independently of the value of the other quartic couplings and hence this equation is a necessary condition for the potential stability even in the generic 2HDM case.

Necessary Conditions for Boundedness from Below

The two options of Eq. (4.229) present a necessary and sufficient condition for BFB. In order to implement this bound, one should scan over all possible values of ρ and η, which can be computationally expensive for large parameter spaces. Thus we present here two simplified

necessary (though not sufficient) conditions which can be used to quickly rule out invalid parameter sets and speed up scans.

We can first derive generalized versions of the existing literature bounds [324] by setting $x = 1$ and taking $\rho = 1$ and $\eta = \frac{n\pi}{4}$, with $n = \{0, ..., 7\}$, in Eq. (4.222) and Eq. (4.225). Applying this procedure to Eq. (4.222) leads to the following conditions

$$\frac{\lambda_1 + \lambda_2}{2} + \lambda_3 + \lambda_4 + \lambda_5^R - 2|\lambda_6^R + \lambda_7^R| > 0\,, \tag{4.233a}$$

$$\frac{\lambda_1 + \lambda_2}{2} + \lambda_3 + \lambda_4 - \lambda_5^R - 2|\lambda_6^I + \lambda_7^I| > 0\,, \tag{4.233b}$$

$$\frac{\lambda_1 + \lambda_2}{2} + \lambda_3 + \lambda_4 + \lambda_5^I - \sqrt{2}\left|(\lambda_6^R + \lambda_7^R) + (\lambda_6^I + \lambda_7^I)\right| > 0\,, \tag{4.233c}$$

$$\frac{\lambda_1 + \lambda_2}{2} + \lambda_3 + \lambda_4 - \lambda_5^I - \sqrt{2}\left|(\lambda_6^R + \lambda_7^R) - (\lambda_6^I + \lambda_7^I)\right| > 0\,, \tag{4.233d}$$

while applying the same procedure to Eq. (4.225) leads to the conditions

$$\sqrt{\lambda_1\lambda_2} + \lambda_3 + \lambda_4 + \lambda_5^R - 2\left|\tilde{\lambda}_6^R + \tilde{\lambda}_7^R\right| > 0\,, \tag{4.234a}$$

$$\sqrt{\lambda_1\lambda_2} + \lambda_3 + \lambda_4 - \lambda_5^R - 2\left|\tilde{\lambda}_6^I + \tilde{\lambda}_7^I\right| > 0\,, \tag{4.234b}$$

$$\sqrt{\lambda_1\lambda_2} + \lambda_3 + \lambda_4 + \lambda_5^I - \sqrt{2}\left|(\tilde{\lambda}_6^R + \tilde{\lambda}_7^R) + (\tilde{\lambda}_6^I + \tilde{\lambda}_7^I)\right| > 0\,, \tag{4.234c}$$

$$\sqrt{\lambda_1\lambda_2} + \lambda_3 + \lambda_4 - \lambda_5^I - \sqrt{2}\left|(\tilde{\lambda}_6^R + \tilde{\lambda}_7^R) - (\tilde{\lambda}_6^I + \tilde{\lambda}_7^I)\right| > 0\,. \tag{4.234d}$$

Note that we have combined the $\eta, \eta + \pi$ conditions in each set to obtain four conditions instead of eight.

Alternatively, we can collapse the two conditions of Eq. (4.229) into a single necessary condition as follows. Consider the two different branches with $\chi_{1,2} < 0$. Under the transformation $\eta \to \eta + \pi$, $\chi_1 \leq 0$ produces two conditions that must be satisfied simultaneously: $(\alpha - \gamma)^2 - 16(\alpha + \beta + \gamma + 2) \leq 0$ and $(\alpha - \gamma)^2 - 16(-\alpha + \beta - \gamma + 2) \leq 0$. We can add these together to obtain the simplified condition: $(\alpha - \gamma)^2 \leq 16(\beta + 2)$. Similarly, demanding that $\chi_2 \leq 0$ for both η and $\eta + \pi$ gives us the simplified condition $(\alpha - \gamma)^2 \leq 16(\beta + 2)$. So, we

see that demanding $\chi_1 \leq 0$ and $\chi_2 \leq 0$ are equivalent, and both translate to the constraint:

$$\chi_1 \leq 0\,, \chi_2 \leq 0 \;\Rightarrow\; (\alpha-\gamma)^2 \leq 16(\beta+2)\,. \tag{4.235}$$

In this way, the condition for the potential to be BFB can be reduced to the form:

$$\Delta \geq 0 \text{ and } \beta \geq -2 \text{ and } (\alpha-\gamma)^2 \leq 16(\beta+2)\,. \tag{4.236}$$

Note that both $\beta \geq -2$ and $(\alpha-\gamma)^2 \leq 16(\beta+2)$ restrict β, but that the latter will always be a stronger condition since $(\alpha-\gamma)^2 \geq 0$. Then this necessary but not sufficient BFB condition simplifies further to

$$\Delta \geq 0 \text{ and } \beta \geq \frac{1}{16}(\alpha-\gamma)^2 - 2\,. \tag{4.237}$$

This condition still depends on ρ and η. Without loss of generality, we set[21] $\rho = 1$. As for η, we need to find the value which extremizes the expression for each condition. Take for instance the latter condition of Eq. (4.237) and define

$$f(\eta) \equiv \beta - \frac{1}{16}(\alpha-\gamma)^2 + 2 \geq 0\,. \tag{4.238}$$

We can recast everything in terms of $\cos 2\eta$ and $\sin 2\eta$ such that $f(\eta)$ only depends on these quantities. We can then easily determine the extremal value of $\eta_{\min}$ which gives the minimal $f_{\min}$. After some algebra, the positivity condition $f_{\min} \geq 0$ reads

$$2(\lambda_1\lambda_2 + \sqrt{\lambda_1\lambda_2}(\lambda_3+\lambda_4)) - \frac{1}{2}\big|\tilde{\lambda}_6 - \tilde{\lambda}_7\big|^2 - \big|2\sqrt{\lambda_1\lambda_2}\,\lambda_5 - \frac{1}{2}(\tilde{\lambda}_6-\tilde{\lambda}_7)^2\big| \geq 0\,, \tag{4.239}$$

21. This gives us one necessary condition. We could obtain others by choosing $\rho < 1$, but these tend to be less constraining in most cases.

where we have defined the rescaled couplings

$$\tilde{\lambda}_6 \equiv \left(\frac{\lambda_2}{\lambda_1}\right)^{1/4} \lambda_6\,, \quad \tilde{\lambda}_7 \equiv \left(\frac{\lambda_1}{\lambda_2}\right)^{1/4} \lambda_7\,. \tag{4.240}$$

Eqs. (4.233), (4.234), and (4.239) present simplified necessary conditions for BFB and are the main result of this section.

Sufficient Conditions for Boundedness from Below

It is also useful to have simplified sufficient conditions which allow one to quickly determine if the potential is BFB for a given parameter set. Consider the top branch of Eq. (4.229) — i.e. $\Delta \geq 0$ and $-2 \leq \beta \leq 6$ and $\chi_1 \leq 0$. One can show[22] that a stronger condition (which will lead to a sufficient condition) is $\beta \leq 6$ *and* $\alpha + \frac{1}{2}(\beta+2) > 0$ and $\gamma + \frac{1}{2}(\beta+2) > 0$. In terms of the λ's, this translates to the sufficient condition

$$\begin{aligned} &3\sqrt{\lambda_1\lambda_2} - (\lambda_3 + |\lambda_4| + |\lambda_5|) \geq 0\,, \\ &\text{and } \sqrt{\lambda_1\lambda_2} + \lambda_3 - (|\lambda_4| + |\lambda_5| + 4|\tilde{\lambda}_6|) > 0\,, \\ &\text{and } \sqrt{\lambda_1\lambda_2} + \lambda_3 - (|\lambda_4| + |\lambda_5| + 4|\tilde{\lambda}_7|) > 0\,. \end{aligned} \tag{4.241}$$

Now consider the bottom branch of Eq. (4.229) — i.e. $\Delta \geq 0$ *and* $\beta > 6$ *and* $\chi_2 \leq 0$. To arrive at an analytic sufficient condition, consider the stronger bound $\beta > 6$ *and* $\alpha + 2\sqrt{\beta-2} > 0$

22. Note that $\alpha = -\frac{1}{2}(\beta+2)$ and $\gamma = -\frac{1}{2}(\beta+2)$ are the directions along which $\Delta = 0$. In order to satisfy the $\chi_1 \leq 0$ condition, relevant for $\beta \leq 6$, we must have $\alpha + \frac{1}{2}(\beta+2) \geq 0$ and $\gamma + \frac{1}{2}(\beta+2) \geq 0$. So long as $\beta \geq -2$, these conditions combined will always yield positive Δ. Then Eq. (4.241) is a sufficient condition that follows from the top branch of Eq. (4.229).

and $\gamma + 2\sqrt{\beta - 2} > 0$. In terms of the potential parameters, this condition reads:

$$\begin{aligned} &\lambda_3 - (3\sqrt{\lambda_1\lambda_2} + |\lambda_4| + |\lambda_5|) \geq 0\,, \\ &\text{and } \sqrt{\lambda_3 - (\sqrt{\lambda_1\lambda_2} + |\lambda_4| + |\lambda_5|)} - \frac{\sqrt{2}}{(\lambda_1\lambda_2)^{1/4}}|\tilde{\lambda}_6| > 0\,, \\ &\text{and } \sqrt{\lambda_3 - (\sqrt{\lambda_1\lambda_2} + |\lambda_4| + |\lambda_5|)} - \frac{\sqrt{2}}{(\lambda_1\lambda_2)^{1/4}}|\tilde{\lambda}_7| > 0\,. \end{aligned} \tag{4.242}$$

Eqs. (4.241) and (4.242) are the main result of this section.

Numerical Analysis

To compare the performance of our analytic conditions with the numerical BFB condition, we perform a scan over 10,000 randomly chosen points in the 7-dimensional parameter space of $\{\lambda_1, ...\lambda_7\}$. We take as allowed ranges $\lambda_{1,2} \in [0, 2\pi]$ and $|\lambda_{3,4,5,6,7}| \leq \frac{\pi}{2}$, with $\lambda_{5,6,7}$ complex, as this choice yields about half of the points BFB. Fig. 4.14 shows the number of points which pass the numerical condition Eq. (4.229) as well as the number which pass the combination of our necessary conditions Eqs. (4.233), (4.234), and (4.239) and the number which pass the combination of our sufficient conditions Eqs. (4.241) and (4.242). While we display in the figure the result of combining all necessary conditions derived in Sec. 4.3.5, we note that Eq. (4.234) provides the strongest necessary condition, with the combination of all conditions improving the results by a few percent.

We see that our necessary conditions are very effective at eliminating points which are not BFB, with only approximately 51% of points passing this condition, as compared with the 45% of points which actually satisfy BFB. Meanwhile, our analytic sufficient conditions guarantee approximately 11% of points are BFB. Of these points, essentially all are obtained from Eq. (4.241), which was derived from the upper branch of Eq. (4.229). The second condition Eq. (4.242), derived from the lower branch, is too strong and admits almost no points, which also reflects the fact that most of the points sampled fall within the regime of

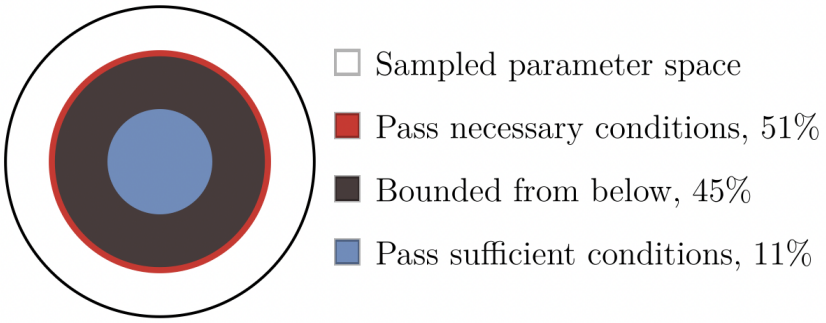

Figure 4.14: The white circle represents 10,000 randomly chosen points in the 7-dimensional parameter space of couplings $\{\lambda_1, ...\lambda_7\}$. We take as priors $\lambda_{1,2} \in [0, 2\pi]$ and $|\lambda_{3,4,5,6,7}| \leq \frac{\pi}{2}$, with $\lambda_{5,6,7}$ allowed to be complex. The red circle encompasses the points which pass our analytic necessary conditions of Eqs. (4.233), (4.234), and (4.239); the black circle contains the points which pass the necessary and sufficient BFB condition of Eq. (4.229); and the innermost blue circle contains the points which pass our sufficient condition of Eq. (4.241).

the first branch.

An examination of the analytical expressions indicates that the quantity $\sqrt{\lambda_1 \lambda_2}$ may play an important role in the determination of BFB. To examine whether the analytical form of our bounds indeed captures the primary underlying behavior of BFB with respect to the λ_i, we plot a histogram in $\sqrt{\lambda_1 \lambda_2}$ of the fraction of tested points which pass the numerical, necessary, and sufficient bounds of Eqs. (4.229), (4.233), (4.234), (4.239) and (4.241), respectively. As in Fig. 4.14, we choose parameters in the range $\lambda_{1,2} \in [0, 2\pi]$ and $|\lambda_{3,4,5,6,7}| \leq \pi/2$ with $\lambda_{5,6,7}$ allowed to be complex. The resulting figure is shown in Fig. 4.15. As can be seen from the figure, more points pass the BFB condition for higher $\sqrt{\lambda_1 \lambda_2}$, as indicated by the forms of the necessary and sufficient conditions. We find that both the necessary and sufficient bounds follow the same behavior as the exact numerical results, indicating that the analytic bounds do indeed capture the relevant behavior in $\sqrt{\lambda_1 \lambda_2}$.

Finally, we note that within the existing literature, some simplified analytic BFB constraints for the most general 2HDM (i.e. involving $\lambda_{6,7} \neq 0$) do exist. For example, the

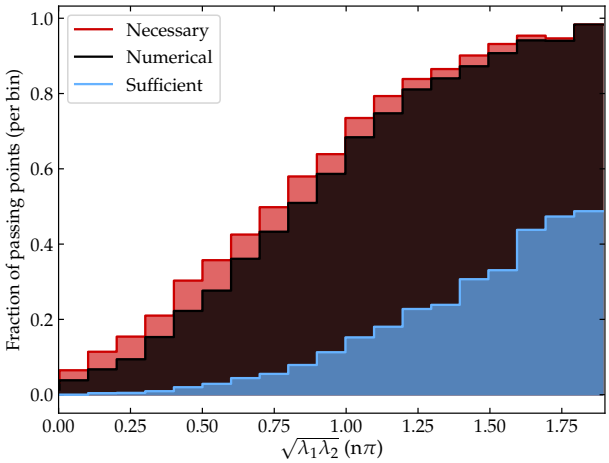

Figure 4.15: Histogram of $\sqrt{\lambda_1\lambda_2}$ displaying the fraction of tested points per bin which pass the necessary conditions Eqs. (4.239, 4.233, 4.234) (red), numerical test Eq. (4.229) (black), and sufficient condition Eq. (4.241) (blue). As in Fig. 4.14, we take as priors $\lambda_{1,2} \in [0, 2\pi]$ and $|\lambda_{3,4,5,6,7}| \in [0, \frac{\pi}{2}]$, with $\lambda_{5,6,7}$ allowed to be complex.

authors of [324] find as a necessary condition

$$\frac{1}{2}(\lambda_1 + \lambda_2) + \lambda_3 + \lambda_4 + \lambda_5 - 2|\lambda_6 + \lambda_7| > 0\,. \tag{4.243}$$

This expression, that agrees with Eq. (4.233) in the appropriate limit, is derived by assuming that the Higgs doublets are aligned in field space, and is limited to the case that all λ_i are taken to be real. Restricting ourselves to this regime, we find that the literature expression excludes approximately 17% of points while ours excludes approximately 49%, making our condition the stronger of the two by a large margin.

4.3.6 Vacuum Stability

We can also place constraints on the allowed 2HDM potential parameters by demanding the existence of a stable neutral vacuum. Strictly speaking, this not a necessary requirement: it is only necessary that the vacuum is meta-stable, with a lifetime longer than the age of the Universe. Here, we just derive the conditions for absolute stability, more precisely the

absence of deeper minima at scales of the order of the TeV scale.

The discriminant D introduced in [366, 367] offers a prescription for distinguishing the nature of a solution obtained by extremizing the potential. We summarize the method here, beginning by writing the potential as:

$$V = -M_\mu \mathsf{r}^\mu + \frac{1}{2}\Lambda_{\mu\nu}\mathsf{r}^\mu \mathsf{r}^\nu - \frac{1}{2}\zeta \mathsf{r}^\mu \mathsf{r}_\mu \,, \tag{4.244}$$

where M_μ encodes the mass terms:

$$M_\mu = \left(-\tfrac{1}{2}(m_{11}^2 + m_{22}^2),\;\; (m_{12}^2)^R,\;\; -(m_{12}^2)^I,\;\; -\tfrac{1}{2}(m_{11}^2 - m_{22}^2) \right) , \tag{4.245}$$

r^μ is a vector of field bilinears:

$$\mathsf{r}^\mu = \left(|\Phi_1|^2 + |\Phi_2|^2,\;\; 2\mathrm{Re}[\Phi_1^\dagger \Phi_2],\;\; 2\mathrm{Im}[\Phi_1^\dagger \Phi_2],\;\; |\Phi_1|^2 - |\Phi_2|^2 \right) , \tag{4.246}$$

and $\Lambda_{\mu\nu}$ encodes the quartic terms:

$$\Lambda_{\mu\nu} = \frac{1}{2} \begin{pmatrix} \frac{1}{2}(\lambda_1 + \lambda_2) + \lambda_3 & \lambda_6^R + \lambda_7^R & -(\lambda_6^I + \lambda_7^I) & \frac{1}{2}(\lambda_1 - \lambda_2) \\ (\lambda_6^R + \lambda_7^R) & (\lambda_4 + \lambda_5^R) & -\lambda_5^I & \lambda_6^R - \lambda_7^R \\ -(\lambda_6^I + \lambda_7^I) & -\lambda_5^I & \lambda_4 - \lambda_5^R & -(\lambda_6^I - \lambda_7^I) \\ \frac{1}{2}(\lambda_1 - \lambda_2) & \lambda_6^R - \lambda_7^R & -(\lambda_6^I - \lambda_7^I) & \frac{1}{2}(\lambda_1 + \lambda_2) - \lambda_3 \end{pmatrix} . \tag{4.247}$$

The last term in Eq. (4.244) is a Lagrange multiplier we have introduced to enforce the condition $\mathsf{r}^\mu \mathsf{r}_\mu = 0$, which ensures we are in a charge-neutral minimum; we enforce this condition since charge-breaking and normal minima cannot coexist in the 2HDM (see [366, 339] for more details). In the above equations, indices are raised and lowered using a Minkowski metric.

Provided the matrix $\Lambda_{\mu\nu}$, which contains the coefficients of the quartic terms in the

potential, is positive definite, corresponding to a potential which is BFB, it can be brought into a diagonal form by an $SO(1,3)$ transformation:

$$\Lambda_{\mu\nu}^{\text{diag}} = \begin{pmatrix} \Lambda_0 & 0 & 0 & 0 \\ 0 & -\Lambda_1 & 0 & 0 \\ 0 & 0 & -\Lambda_2 & 0 \\ 0 & 0 & 0 & -\Lambda_3 \end{pmatrix}, \tag{4.248}$$

with Λ_0 the "timelike" eigenvalue and Λ_i "spacelike". Let us define the "signature matrix" S as $S \equiv \Lambda_{\mu\nu} - \zeta g_{\mu\nu}$. In diagonal form, it looks like:

$$S = \begin{pmatrix} \Lambda_0 - \zeta & 0 & 0 & 0 \\ 0 & \zeta - \Lambda_1 & 0 & 0 \\ 0 & 0 & \zeta - \Lambda_2 & 0 \\ 0 & 0 & 0 & \zeta - \Lambda_3 \end{pmatrix}. \tag{4.249}$$

The discriminant is generically given by the determinant of the signature matrix:

$$D = \det S\,. \tag{4.250}$$

By using the diagonal form above, we can write this as:

$$D = (\Lambda_0 - \zeta)(\zeta - \Lambda_1)(\zeta - \Lambda_2)(\zeta - \Lambda_3)\,. \tag{4.251}$$

We finally come to the vacuum stability condition. Suppose we have already verified that our potential is BFB and calculated the discriminant, time-like eigenvalue Λ_0, and Lagrange

multiplier ζ.

$$\text{We are in a global minimum if and only if : } \begin{cases} D > 0\,, \text{ or} \\ \\ D < 0 \text{ and } \zeta > \Lambda_0\,. \end{cases} \tag{4.252}$$

For our purposes, it is more useful to work with the "Euclideanized" version of $\Lambda_{\mu\nu}$ obtained by lowering one of the indices with the Minkowski metric, $\Lambda_E \equiv \Lambda^\mu_\nu$. Explicitly:

$$\Lambda_E = \frac{1}{2}\begin{pmatrix} \frac{1}{2}(\lambda_1+\lambda_2)+\lambda_3 & \lambda_6^R+\lambda_7^R & -(\lambda_6^I+\lambda_7^I) & \frac{1}{2}(\lambda_1-\lambda_2) \\ -(\lambda_6^R+\lambda_7^R) & -(\lambda_4+\lambda_5^R) & \lambda_5^I & -(\lambda_6^R-\lambda_7^R) \\ \lambda_6^I+\lambda_7^I & \lambda_5^I & -(\lambda_4-\lambda_5^R) & \lambda_6^I-\lambda_7^I \\ -\frac{1}{2}(\lambda_1-\lambda_2) & -(\lambda_6^R-\lambda_7^R) & \lambda_6^I-\lambda_7^I & -\frac{1}{2}(\lambda_1+\lambda_2)+\lambda_3 \end{pmatrix}. \tag{4.253}$$

In terms of Λ_E, the discriminant is:

$$D = -\det[\Lambda_E - \mathbb{I}\zeta]\,. \tag{4.254}$$

The other quantity necessary for formulating the discriminant is the Lagrange multiplier ζ. This may be obtained by looking at any component of the minimization condition:

$$\Lambda^\mu_\nu r^\nu - M^\mu = \zeta r^\mu\,. \tag{4.255}$$

We parameterize the vacuum expectation values (vevs) of the doublets as:

$$\langle\Phi_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v_1 \end{pmatrix}\,,\quad \langle\Phi_2\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v_2 e^{i\eta} \end{pmatrix}\,. \tag{4.256}$$

Then the expectation value of field bilinears $r^\mu \equiv \langle \mathsf{r}^\mu \rangle$ is:

$$r^\mu = \left(\tfrac{1}{2}(v_1^2 + v_2^2), \quad v_1 v_2 \cos\eta, \quad v_1 v_2 \sin\eta, \quad \tfrac{1}{2}(v_1^2 - v_2^2) \right) . \tag{4.257}$$

The expression for ζ is particularly simple if we choose the "1" component. In particular if we take $\eta = 0$, then:

$$\zeta = \frac{(m_{12}^2)^R}{v_1 v_2} - \frac{1}{2}\left(\frac{v_1}{v_2}\lambda_6^R + \frac{v_2}{v_1}\lambda_7^R + (\lambda_4 + \lambda_5^R) \right) . \tag{4.258}$$

Note that this has the interpretation of the charged Higgs mass over the vev v squared,

$$\zeta = \frac{M_{H^\pm}^2}{v^2} , \tag{4.259}$$

as first demonstrated in [343, 344].

If $D > 0$, then the physical minimum is the global one, implying absolute stability. If, instead, $D < 0$, we need to compare the timelike eigenvalue Λ_0 with ζ: we are in a global minimum if $\zeta > \Lambda_0$; otherwise, the minimum is metastable. Provided we have already verified that the potential is BFB, however, there is an even simpler way to assess the nature of the extremum.

As an aside, working at the level of eigenvalues the two options of Eq. (4.252) for an extremum to be the global minimum can actually be collapsed into one. Recall that when the potential is BFB, $\Lambda^{\mu\nu}$ is positive definite and $\Lambda_0 > \Lambda_{1,2,3}$. Then from Eq. (4.251), $D > 0$ necessarily implies that we have the ordering $\Lambda_0 > \zeta > \Lambda_{1,2,3}$. Similarly $D < 0$ and $\zeta > \Lambda_0$ necessarily implies the ordering $\zeta > \Lambda_0 > \Lambda_{1,2,3}$. So we see that the relative ordering of Λ_0 and ζ does not actually matter−all that matters for a potential which has been verified to be BFB is that ζ be larger than the spatial eigenvalues, $\zeta > \Lambda_{1,2,3}$.

Sufficient Conditions for Stability

Gershgorin Bounds

We can bound the eigenvalues of Λ_E using the Gershgorin disk theorem in order to derive a sufficient condition for a given vacuum solution to be stable. We first construct the intervals containing the eigenvalues of Λ_E and define the endpoint of each interval as $\Gamma_i \equiv a_{ii} + R_i$, with $R_i = \sum_{j\neq i} |a_{ij}|$:

$$\Gamma_1 = \frac{1}{4}(\lambda_1+\lambda_2) + \frac{\lambda_3}{2} + \frac{1}{2}\left(|\lambda_6^R + \lambda_7^R| + |\lambda_6^I + \lambda_7^I| + \frac{1}{2}|\lambda_1 - \lambda_2|\right), \tag{4.260a}$$

$$\Gamma_2 = -\frac{1}{4}(\lambda_1+\lambda_2) + \frac{\lambda_3}{2} + \frac{1}{2}\left(|\lambda_6^R - \lambda_7^R| + |\lambda_6^I - \lambda_7^I| + \frac{1}{2}|\lambda_1 - \lambda_2|\right), \tag{4.260b}$$

$$\Gamma_3 = -\frac{1}{2}(\lambda_4+\lambda_5^R) + \frac{1}{2}\Big(|\lambda_5^I| + |\lambda_6^R + \lambda_7^R| + |\lambda_6^R - \lambda_7^R|\Big), \tag{4.260c}$$

$$\Gamma_4 = -\frac{1}{2}(\lambda_4-\lambda_5^R) + \frac{1}{2}\Big(|\lambda_5^I| + |\lambda_6^I + \lambda_7^I| + |\lambda_6^I - \lambda_7^I|\Big). \tag{4.260d}$$

We know that all eigenvalues must be less than the endpoint of the interval extending the furthest in the $+\hat{x}$ direction,

$$\max[\Gamma_i] \geq \Lambda_{0,1,2,3}\,. \tag{4.261}$$

Meanwhile, an extremum will be the global minimum if $\zeta > \Lambda_{1,2,3}$. Thus, it is sufficient to demand:

$$\zeta > \max[\Gamma_i]\,. \tag{4.262}$$

Frobenius Bounds

One may also bound the eigenvalues using the Frobenius norm to obtain a single-equation condition. We require the maximum eigenvalue be less than ζ, which gives the constraint:

$$\zeta > \frac{1}{2}\sqrt{\lambda_1^2 + \lambda_2^2 + 2(\lambda_3^2 + \lambda_4^2 + |\lambda_5|^2) + 4(|\lambda_6|^2 + |\lambda_7|^2)} \tag{4.263}$$

Note that in this case, the Frobenius bound is insensitive to the signs of the λ_i, while the Gershgorin condition is sensitive to the signs of λ_3, λ_4, and λ_5^R. We thus expect the Gershgorin bound to be the stronger of the two.

Principal Minors

In the case of a non-symmetric matrix such as Λ_E, Sylvester's criterion no longer holds and so cannot be applied in a straightforward manner. However, the following statement does hold: if the symmetric part of a matrix M is positive-definite, then the real parts of the eigenvalues of M are positive. This statement does not hold in the other direction, and therefore cannot be used to derive necessary conditions. However, we can apply Sylvester's criterion to the symmetric part of Λ_E to obtain a sufficient condition.

The symmetric part of Λ_E is given by

$$\begin{aligned}\Lambda_E^S &= \frac{1}{2}(\Lambda_E^T + \Lambda_E) \\ &= \frac{1}{2}\begin{pmatrix} \frac{1}{2}(\lambda_1+\lambda_2)+\lambda_3 & 0 & 0 & 0 \\ 0 & -(\lambda_4+\lambda_5^R) & \lambda_5^I & -(\lambda_6^R-\lambda_7^R) \\ 0 & \lambda_5^I & -(\lambda_4-\lambda_5^R) & \lambda_6^I-\lambda_7^I \\ 0 & -(\lambda_6^R-\lambda_7^R) & \lambda_6^I-\lambda_7^I & -\frac{1}{2}(\lambda_1+\lambda_2)+\lambda_3 \end{pmatrix}\end{aligned} \tag{4.264}$$

We require the matrix $\zeta\mathbb{I} - \Lambda_E^S$ to be positive-definite. Since the lower-right 3×3 matrix decouples from the "11" element, we can analyze them separately when considering positive-definiteness. We require the "11" element to be positive, and apply Sylvester's criterion to the lower-right 3×3 submatrix. This gives the following set of conditions:

$$\zeta - \frac{1}{4}(\lambda_1+\lambda_2) - \frac{1}{2}\lambda_3 > 0\,, \tag{4.265a}$$

$$\zeta + \frac{1}{2}\lambda_4 + \frac{1}{2}\lambda_5^R > 0\,, \tag{4.265b}$$

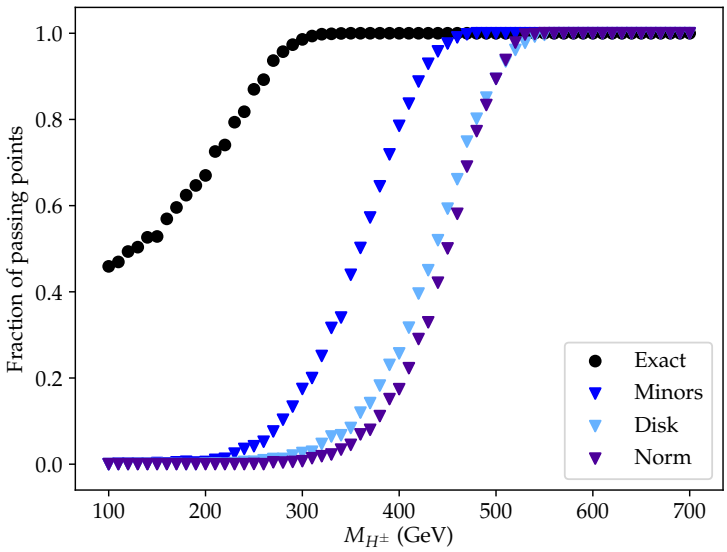

Figure 4.16: Comparison of the fraction of points that pass the exact stability conditions, Eq. (4.252) (black dots), with respect to the fraction passing the three sufficient conditions for vacuum stability: principal minors Eq. (4.265) (blue), Gershgorin disk theorem Eq. (4.262) (light blue), and Frobenius norm Eq. (4.263) (purple). We plot the fraction of points that pass each condition as a function of $M_{H^\pm}$. We take $\lambda_{1,2} \in [0, 2\pi]$ and $|\lambda_{3,4,5,6,7}| \leq \frac{\pi}{2}$, with $\lambda_{5,6,7}$ allowed to be complex, and restrict to examining points which are BFB.

$$\left(\zeta + \frac{1}{2}\lambda_4\right)^2 - \frac{1}{4}|\lambda_5|^2 > 0\,, \tag{4.265c}$$

$$\begin{aligned}(4\zeta + \lambda_1 + \lambda_2 - 2\lambda_3)((2\zeta + \lambda_4)^2 - |\lambda_5|^2) \\ + \lambda_5(\lambda_6^* - \lambda_7^*)^2 + \lambda_5^*(\lambda_6 - \lambda_7)^2 - 2(2\zeta + \lambda_4)|\lambda_6 - \lambda_7|^2 > 0\,.\end{aligned} \tag{4.265d}$$

Taken together, the Eqs. (4.265) provide a sufficient condition for vacuum stability.

Numerical Comparison

In Fig. 4.16 we plot the performance of the three sufficient conditions for vacuum stability, Eqs. (4.262), (4.263), (4.265), as a function of the charged Higgs mass $M_{H^\pm}$. We compare these results with the fraction that pass the exact stability condition, Eq. (4.252). As in previous sections, we choose the λ_i randomly with $\lambda_{1,2} \in [0, 2\pi]$ and $|\lambda_{3,4,5,6,7}| \leq \frac{\pi}{2}$, with $\lambda_{5,6,7}$ allowed to be complex. The y-axis shows the fraction of tested points which pass

the respective stability condition; we restrict to testing points which are BFB, to ensure the validity of the stability conditions implemented in Fig. 4.16. We find that the set of conditions arising from the application of Sylvester's criterion capture the most stable points, while all three bounds capture more stable points when the λ_i are small compared to the ratio $M^2_{H^\pm}/v^2$.

Vacuum Stability in the Higgs Basis

It is particularly interesting to study vacuum stability in the Higgs basis, in which only one of the doublets possesses a vev (see Sec. 4.3.9 for a review of the conversion to the Higgs basis as well as our conventions). One advantage of this basis is that the potential parameters are closely related to physical observables: for example, Z_1 controls the trilinear coupling of three SM-like Higgs bosons hhh, Z_6 controls the trilinear coupling of two SM-like and one non-SM-like CP-even Higgs bosons hhH, etc. (see e.g. [358] for an exhaustive list of couplings). Since none of the bounds obtained in this article have relied on the choice of basis, they can equally well be applied to Higgs basis parameters. Using the close relationship between the Higgs basis parameters and physical quantities, we here aim at obtaining approximate bounds on the physical observables of the model.

In our notation, the scalar which obtains a vev is denoted by ϕ_1^0. The mass matrix for the neutral scalars $\vec{\phi} = (\phi_1^0, \phi_2^0, a_0)^{\mathrm{T}}$ reads:

$$\mathcal{M}^2 = v^2 \begin{pmatrix} Z_1 & Z_6^R & -Z_6^I \\ Z_6^R & \frac{M^2_{H^\pm}}{v^2} + \frac{1}{2}(Z_4 + Z_5^R) & -\frac{1}{2}Z_5^I \\ -Z_6^I & -\frac{1}{2}Z_5^I & \frac{M^2_{H^\pm}}{v^2} + \frac{1}{2}(Z_4 - Z_5^R) \end{pmatrix}, \tag{4.266}$$

where $M^2_{H^\pm}$ is the charged Higgs mass:

$$M^2_{H^\pm} = M^2_{22} + \frac{1}{2}Z_3 v^2 \,. \tag{4.267}$$

We will restrict ourselves to the alignment limit, which is the limit in which ϕ_1^0 is aligned with the 125 GeV mass eigenstate. In this case, the 125 GeV Higgs couples to the electroweak gauge bosons and all fermions with SM strength, and the alignment limit is therefore phenomenologically well-motivated by precision Higgs results from the LHC [352, 353].

Examining the above matrix, it appears that there are two ways in which one may obtain alignment. The first option, known as the decoupling limit, corresponds to taking $M_{H^\pm}^2 + \frac{1}{2}(Z_4 \pm Z_5^R)v^2 \gg Z_1 v^2$. Under this limit, the heavy mass eigenstates h_2 and h_3 and the heavy charged Higgs $H^\pm$ decouple from the light mass eigenstate, leaving h_1 aligned with ϕ_1^0. More interesting from a phenomenological standpoint is the approximate *alignment without decoupling* limit, as it leaves the non-standard Higgs states potentially within collider reach. This corresponds to taking $|Z_6| \ll 1$, for which the mixing between ϕ_1^0 and the other neutral scalars vanishes, leading to the identification of ϕ_1^0 with the mass eigenstate h_1. For the following discussion we will take $|Z_6| \ll 1$ and work in the alignment without decoupling limit.

We define $h_1 \equiv h$ to be the SM-like Higgs boson, which has a mass given by

$$M_h^2 = Z_1 v^2 \,. \tag{4.268}$$

To obtain a physical Higgs mass close to the experimental value of 125 GeV, it is required that we fix $Z_1 \approx 0.25$. The remaining 2×2 mass matrix can be diagonalized to obtain the masses of the remaining scalars h_2 and h_3:

$$M_{h_3,h_2}^2 = M_{H^\pm}^2 + \frac{1}{2}(Z_4 \pm |Z_5|)v^2 \,. \tag{4.269}$$

There are two possibilities for the CP properties of these states. So long as $Z_5^I \neq 0$, h_2 and h_3 have mixed CP properties. In the limit of $Z_5^I = 0$, meanwhile, the non-standard Higgs mass matrix becomes diagonal, and we obtain mass eigenstates H and A with definite CP

character,

$$M^2_{H,A} = M^2_{H^\pm} + \frac{1}{2}(Z_4 \pm Z_5^R)v^2\,. \tag{4.270}$$

The masses of the general mass eigenstates and the states of definite CP character can be related by

$$\begin{aligned} M^2_H &= M^2_{h_3,h_2} + \frac{1}{2}(Z_5^R \mp |Z_5|)v^2 \\ M^2_A &= M^2_{h_3,h_2} - \frac{1}{2}(Z_5^R \pm |Z_5|)v^2. \end{aligned} \tag{4.271}$$

In the following analysis we will make no assumptions about the CP character of the mass eigenstates, and will work with the generic physical masses M_{h_3,h_2}.

With the above definitions, we can rephrase our sufficient vacuum stability conditions into constraints on physical quantities. We start with the Gershgorin condition of Eq. (4.262). Expressing the Γ's in terms of physical masses, a sufficient condition for vacuum stability becomes:

$$\begin{aligned} &M^2_{h_2} > \frac{1}{2}(|Z_5^I| - Z_5^R - |Z_5|)v^2 + \frac{1}{2}|Z_7^R|v^2 \ \text{ and} \\ &M^2_{h_3} > \frac{1}{2}(|Z_5^I| + Z_5^R + |Z_5|)v^2 + \frac{1}{2}|Z_7^I|v^2 \ \text{ and} \\ &M^2_{H^\pm} > \frac{1}{2}\max[M_h^2,\, Z_2 v^2] + \frac{1}{2}(Z_3 + |Z_7^R| + |Z_7^I|)v^2\,. \end{aligned} \tag{4.272}$$

Next, we can recast the Frobenius sufficient condition, Eq. (4.263), in terms of the physical masses; doing so results in the following condition:

$$2M^2_{H^\pm}\big(M^2_{h_2} + M^2_{h_3}\big) - \big(M^4_{H^\pm} + M^4_{h_2} + M^4_{h_3} + \frac{1}{4}M^4_h\big) > \frac{1}{4}(Z_2^2 + 2Z_3^2 + 4|Z_7|^2)v^4\,. \tag{4.273}$$

Finally, Sylvester's criterion provides an additional set of sufficient conditions. A sample set

of sufficient conditions for vacuum stability in the alignment limit based on Eq. (4.265) is:

$$
\begin{aligned}
&4M_{H^\pm}^2 - M_h^2 > (Z_2 + 2Z_3)v^2, \\
&M_{h_3}^2 > \frac{1}{2}(|Z_5| - Z_5^R), \\
&M_{h_2}^2 M_{h_3}^2 > 0, \\
4M_{h_2}^2 M_{h_3}^2 &(4M_{H^\pm}^2 + M_h^2 + Z_2 v^2 - 2Z_3 v^2) - 2(M_{h_2}^2 + M_{h_3}^2)|Z_7|^2 + Z_5 Z_7^{*2} + Z_5^* Z_7^2 > 0\,.
\end{aligned}
\tag{4.274}
$$

4.3.7 CP Violation in the General 2HDM

The bounds we have derived in this work have implications for the allowed values of physical parameters in a given 2HDM. This can be seen in a straightforward way in the previous section, where the conditions for vacuum stability were recast into expressions that restrict the physical masses of the bosonic sector. One particularly interesting question to which our bounds can be applied is that of the amount of CP violation permitted in the alignment limit. This possibility has largely been neglected in the many previous studies which restrict themselves to the $\mathbb{Z}_2$-symmetric 2HDM. This is understandable since exact alignment implies CP conservation in the $\mathbb{Z}_2$-symmetric case. When working in the fully general 2HDM, however, it is possible to have CP violation whilst still keeping the SM-like Higgs boson fully aligned.

To justify this claim, recall that there are four complex parameters in the 2HDM: $\{M_{12}^2, Z_5, Z_6, Z_7\}$. One of these is fixed by the minimization condition $M_{12}^2 = -\frac{1}{2}Z_6 v^2$, leaving just three independent parameters, which we take to be the couplings $\{Z_5, Z_6, Z_7\}$. These complex parameters enter into the three basic CP violating invariants of the 2HDM scalar sector J_1, J_2, and J_3, which can be thought of as analogous to the Jarlskog invariant J of the SM quark sector. They are worked out explicitly in [368, 369]; the important fact is that they scale as $J_1 \sim \mathrm{Im}[Z_5^* Z_6^2]$, $J_2 \sim \mathrm{Im}[Z_5^* Z_7^2]$, $J_3 \sim \mathrm{Im}[Z_6^* Z_7]$. It is then clear that

the condition for the Higgs sector to be CP invariant is:

$$\mathrm{Im}[Z_5^* Z_6^2] = \mathrm{Im}[Z_5^* Z_7^2] = \mathrm{Im}[Z_6^* Z_7] = 0\,. \tag{4.275}$$

There are two ways in which this can be satisfied [370]: either $Z_5^I = Z_6^R = Z_7^R = 0$ or $Z_5^I = Z_6^I = Z_7^I = 0$. Note that in the limit of exact alignment we have $Z_6 = 0$, so this reduces to demanding either $Z_5^I = Z_7^R = 0$ or $Z_5^I = Z_7^I = 0$.

Meanwhile in the 2HDM with a (softly broken) $\mathbb{Z}_2$ symmetry, the fact that $\lambda_6 = \lambda_7 = 0$ implies the following two relations between the parameters in the Higgs basis (see Eq. (4.283) in Appendix A) [371, 372]

$$Z_6 + Z_7 = \frac{1}{2}\tan 2\beta (Z_2 - Z_1), \tag{4.276}$$

$$Z_6 - Z_7^* = \frac{1}{\tan 2\beta}(Z_1 + 2Z_6 \cot 2\beta - Z_3 - Z_4 - Z_5). \tag{4.277}$$

It immediately follows that:

$$Z_7^I = -Z_6^I, \tag{4.278}$$

$$Z_5^I = 2\,\frac{1 - \tan^2 2\beta}{\tan 2\beta} Z_6^I. \tag{4.279}$$

These conditions imply that in the exact alignment limit (i.e. $Z_6 = 0$), it will necessarily be the case that $Z_5^I = Z_7^I = 0$. Thus, exact alignment directly leads to CP conservation in the $\mathbb{Z}_2$-symmetric or softly broken $\mathbb{Z}_2$-symmetric 2HDM. This need not be the case in the fully general 2HDM, where the above relations no longer hold. If we allow for a small misalignment (i.e. $|Z_6| \gtrsim 0$), $|Z_5^I|$, which controls the mixing between the H and A bosons, can still be large for large $\tan\beta$.

The physical consequences of the difference in the CP properties in the alignment limit between the (softly broken) $\mathbb{Z}_2$-symmetric and the general 2HDMs are illustrated in Fig. 4.17,

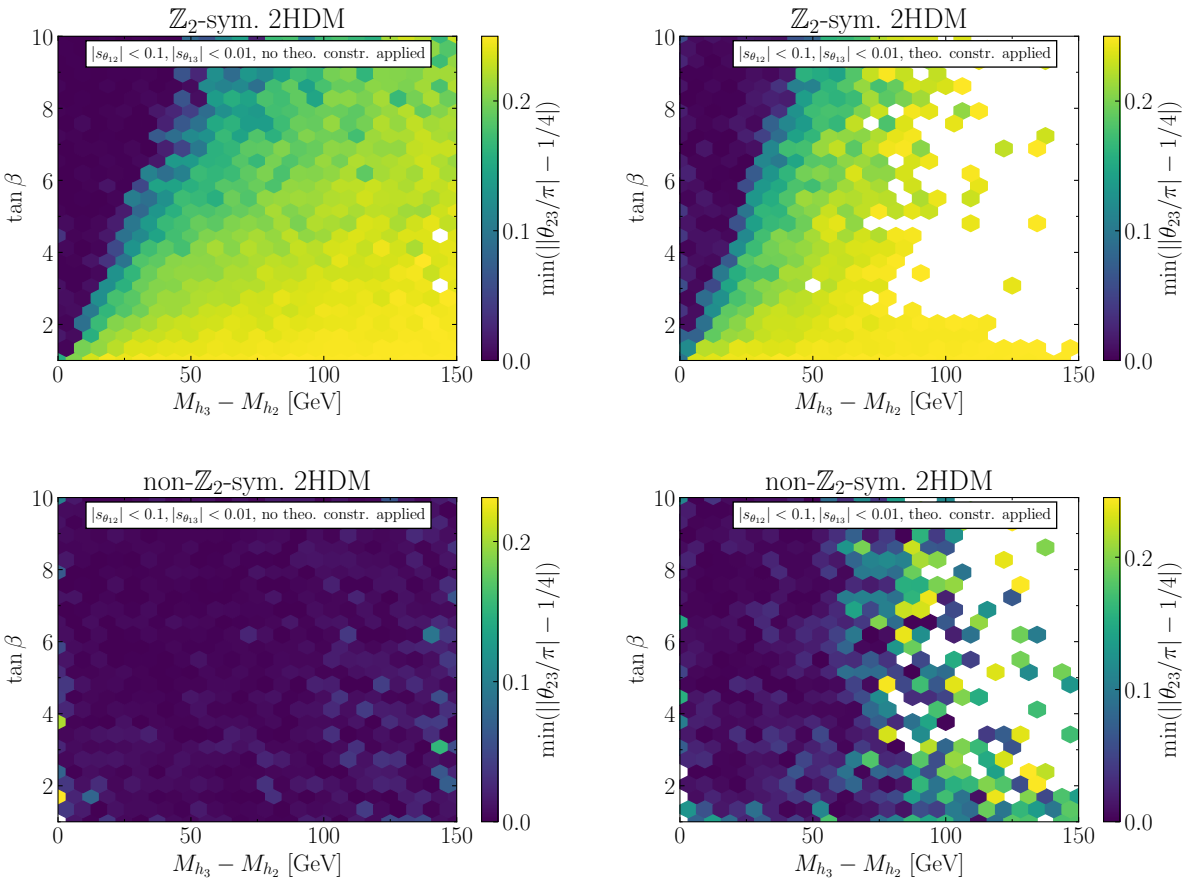

Figure 4.17: *Upper left:* Parameter scan of the 2HDM with a softly broken $\mathbb{Z}_2$ symmetry in the $(M_{h_3} - M_{h_2}, \tan\beta)$ parameter plane with parameter range $\lambda_{1,2} \in [0, 2\pi]$ and $|\lambda_{3,4,5}| \in [0, \pi]$. The conditions $|\sin\theta_{12}| < 0.1$ and $|\sin\theta_{13}| < 0.01$ are imposed. The colour indicates the minimal value of $|\theta_{23}/\pi - 1/4|$ in each hexagonal patch. *Upper right:* Same as upper left, but constraints from perturbative unitary, BFB, and vacuum stability are applied in addition. *Lower left:* Same as upper left, but the scan is performed in the general 2HDM without a (softly broken) $\mathbb{Z}_2$ symmetry ($|\lambda_{6,7}| \in [0, \pi]$). *Lower right:* Same as lower left, but constraints from perturbative unitary, BFB, and vacuum stability are applied in addition.

in which we show the results of several parameter scans. Motivated by Higgs precision [352, 353] and electric dipole moment bounds (see e.g. [373, 374]), we demand a small mixing of the h and the H states and an even smaller mixing of the h and the A states; we impose these constraints on the mixing in an approximate manner by demanding $|\sin\theta_{12}| < 0.1$ and $|\sin\theta_{13}| < 0.01$, where θ_{12} and θ_{13} are the respective mixing angles.[23] The colour code in the figure indicates the minimum value of $|\theta_{23}/\pi - 1/4|$ in each hexagonal patch, where θ_{23} is the mixing angle between the H and A states. This variable is chosen such that the maximal mixing case corresponds to a value of zero. Correspondingly, a dark blue color indicates that a large CP-violating mixing between the H and A states can be realized; a bright yellow color instead signals that no large mixing can be realized.

In the upper left panel of Fig. 4.17 we present a parameter scan of the 2HDM with a softly broken $\mathbb{Z}_2$ symmetry in the $(M_{h_3} - M_{h_2}, \tan\beta)$ parameter plane. We observe that for large $\tan\beta$, a large mixing between the neutral BSM Higgs bosons can be realized if their mass difference is below $\sim 70\,\text{GeV}$. Larger mass differences originate from differences between the diagonal terms of the Higgs mass matrix, suppressing possible mixing effects induced by the off-diagonal $\frac{1}{2}Z_5^I v^2$ term. For lower $\tan\beta$, the condition $|\sin\theta_{13}| < 0.01$ directly implies that Z_5^I is small, resulting in substantial mixing only when the mass difference is close to zero.

In the upper right panel of Fig. 4.17, we again consider a softy broken $\mathbb{Z}_2$ symmetry but additionally impose perturbative unitarity, boundedness-from-below, and vacuum stability constraints following the discussions in the previous Sections. Aside from lowering the maximal possible mass difference between h_3 and h_2, the region in which large mixing between the BSM Higgs bosons can be realized is also reduced.

If we instead investigate the general 2HDM without a (softly broken) $\mathbb{Z}_2$ symmetry (see lower panels of Fig. 4.17), large mixing between the H and A states can be realized throughout the shown parameter plane. Applying the bounds derived in the previous sections

23. The stated bounds on θ_{12} and θ_{13} are just illustrative. We checked that our conclusions do not depend strongly on the values inserted.

Conditions	Perturbative unitarity	Bounded from below	Vacuum stability
Exact	Eq. (4.207)	Eq. (4.229)	Eq. (4.252)
Necessary	Eqs. (4.210), (4.213)	Eq. (4.234)	—
Sufficient	Eq. (4.221)	Eq. (4.241)	Eq. (4.265)

Table 4.1: Overview of the primary results of this paper; further constraints and their analysis may be found in the main text.

only excludes the region with $M_{h_3} - M_{h_2} \gtrsim 100\,\text{GeV}$.

Finally, we want to remark that CP violation can become manifest not only in the neutral mass matrix but also in the bosonic couplings. This occurs if either $\tilde{Z}_7^R \neq 0$ or $\tilde{Z}_7^I \neq 0$, since these couplings enter into couplings like $g_{h_1 h_2 h_3}$ [370]. Exotic decays like $h_3 \to h_1 h_2 \to 3h_1$ would then be indicative of CP violation in the bosonic sector (see e.g. [360]).

4.3.8 Summary

Two Higgs doublet models (2HDMs) present a natural extension of the Standard Model description. In spite of the simplicity of this SM extension, many new parameters appear in this theory, and it is very important to understand the constraints on these parameters which will impact in a relevant way the 2HDM phenomenology. Most existing studies concentrate on the case in which a $\mathbb{Z}_2$ symmetry is imposed on the 2HDM potential and Yukawa sector. While this symmetry is an easy way to avoid flavor-changing neutral currents, it also forbids certain terms in the Higgs potential which do not induce flavor-changing neutral currents at tree level. In fact, in many scenarios in which the 2HDM is the low-energy effective field theory of a more complete high-scale model, these couplings are predicted to be non-zero.

Based on this motivation, in this work we present a step towards a systematic exploration of the non-$\mathbb{Z}_2$-symmetric 2HDM. We studied three of the most important theoretical constraints on the scalar potential parameters: perturbative unitarity, boundedness from below, and vacuum stability. In all three cases, we concentrated on the most general renormalizable potential (not restricted by any discrete symmetry) extending previous works by

deriving analytic necessary and sufficient conditions for these constraints. For convenience, our main results (i.e., those conditions which approximate the exact conditions the best) are summarized in Table 4.1.

The derivation of our constraints makes use of several relevant mathematical properties, of which many have not been exploited in the literature before. These properties are not only applicable to the 2HDM but are also useful for the exploration of other models with extended Higgs sectors. As a first phenomenological application of our bounds, we studied how much CP-violating mixing between the BSM Higgs bosons can be realized in the general 2HDM in comparison to a 2HDM with a (softly broken) $\mathbb{Z}_2$ symmetry. While we found that large CP-violating mixing can only be realized for large $\tan\beta$ in the 2HDM with a softly broken $\mathbb{Z}_2$ symmetry, no such theoretical constraints exist for the general 2HDM. We leave a comprehensive study of the phenomenological consequences of not imposing a $\mathbb{Z}_2$ symmetry for future work.

4.3.9 Supplementary Material: Higgs Basis Conversion

The phenomenological properties of the Higgs sector are more easily analyzed in the Higgs basis, in which only one of the doublets possesses a vev[24]. We parameterize the doublets as:

$$H_1 = \begin{pmatrix} G^+ \\ \frac{1}{\sqrt{2}}(v + \phi_1^0 + iG^0) \end{pmatrix} , \quad H_2 = \begin{pmatrix} H^+ \\ \frac{1}{\sqrt{2}}(\phi_2^0 + ia^0) \end{pmatrix} , \tag{4.280}$$

where $G^\pm$ and G^0 are the Goldstones that become the longitudinal components of $W^\pm$ and Z, $H^\pm$ is the physical singly charged scalar state, and $(\phi_1^0, \phi_2^0, a^0)$ are the neutral scalars.

24. This is technically not enough to uniquely define the Higgs basis. The $U(1)$ diagonal subgroup of the $SU(2)$ symmetry in Higgs flavor space remains intact following SSB. This corresponds to transformations $\Phi_1 \to e^{i\chi}\Phi_1$, $\Phi_2 \to e^{-i\chi}\Phi_2$. As a result, we have a one-dimensional family of Higgs bases parameterized by χ: $\{e^{-i\chi}H_1, e^{i\chi}H_2\}$.

The potential in the Higgs basis reads:

$$\begin{aligned} V = {} & M_{11}^2 H_1^\dagger H_1 + M_{22}^2 H_2^\dagger H_2 - (M_{12}^2 H_1^\dagger H_2 + h.c.) \\ & + \frac{1}{2} Z_1 (H_1^\dagger H_1)^2 + \frac{1}{2} Z_2 (H_2^\dagger H_2)^2 + Z_3 (H_1^\dagger H_1)(H_2^\dagger H_2) + Z_4 (H_1^\dagger H_2)(H_2^\dagger H_1) \\ & + \left[\frac{1}{2} Z_5 (H_1^\dagger H_2)^2 + Z_6 (H_1^\dagger H_1)(H_1^\dagger H_2) + Z_7 (H_2^\dagger H_2)(H_1^\dagger H_2) + h.c. \right] . \end{aligned} \tag{4.281}$$

The conversion between the potential parameters in the weak eigenstate basis and those in the Higgs basis have been worked out in [369]; to be self-contained, we reproduce them here. They are obtained by a rotation by an angle β in field space of the original two Higgs doublets. The mass terms in the two bases are related as:

$$m_{11}^2 = M_{11}^2 c_\beta^2 + M_{22}^2 s_\beta^2 + \mathrm{Re}[M_{12}^2 e^{i\eta}] s_{2\beta} \,, \tag{4.282a}$$

$$m_{22}^2 = M_{11}^2 s_\beta^2 + M_{22}^2 c_\beta^2 - \mathrm{Re}[M_{12}^2 e^{i\eta}] s_{2\beta} \,, \tag{4.282b}$$

$$m_{12}^2 e^{i\eta} = \frac{1}{2}(M_{22}^2 - M_{11}^2) s_{2\beta} + \mathrm{Re}[M_{12}^2 e^{i\eta}] c_{2\beta} + i\, \mathrm{Im}[M_{12}^2 e^{i\eta}] \,, \tag{4.282c}$$

where $\tan\beta = v_2/v_1$ with range $0 \leq \beta \leq \frac{\pi}{2}$, and η is the phase accompanying v_2 in the general basis parameterization of the doublets in Eq. (4.256). The relations between the

quartic couplings are:

$$\lambda_1 = Z_1 c_\beta^4 + Z_2 s_\beta^4 + \frac{1}{2} Z_{345} s_{2\beta}^2 - 2 s_{2\beta} \left(\mathrm{Re}[Z_6 e^{i\eta}] c_\beta^2 + \mathrm{Re}[Z_7 e^{i\eta}] s_\beta^2 \right) , \tag{4.283a}$$

$$\lambda_2 = Z_1 s_\beta^4 + Z_2 c_\beta^4 + \frac{1}{2} Z_{345} s_{2\beta}^2 + 2 s_{2\beta} \left(\mathrm{Re}[Z_6 e^{i\eta}] s_\beta^2 + \mathrm{Re}[Z_7 e^{i\eta}] c_\beta^2 \right) , \tag{4.283b}$$

$$\lambda_3 = \frac{1}{4} \left(Z_1 + Z_2 - 2 Z_{345} \right) s_{2\beta}^2 + Z_3 + \mathrm{Re}[(Z_6 - Z_7) e^{i\eta}] s_{2\beta} c_{2\beta} , \tag{4.283c}$$

$$\lambda_4 = \frac{1}{4} \left(Z_1 + Z_2 - 2 Z_{345} \right) s_{2\beta}^2 + Z_4 + \mathrm{Re}[(Z_6 - Z_7) e^{i\eta}] s_{2\beta} c_{2\beta} , \tag{4.283d}$$

$$\begin{aligned} \lambda_5 e^{2i\eta} = {} & \frac{1}{4} (Z_1 + Z_2 - 2 Z_{345}) s_{2\beta}^2 + \mathrm{Re}[Z_5 e^{2i\eta}] + i \, \mathrm{Im}[Z_5 e^{2i\eta}] c_{2\beta} \\ & + \mathrm{Re}[(Z_6 - Z_7) e^{i\eta}] s_{2\beta} c_{2\beta} + i \, \mathrm{Im}[(Z_6 - Z_7) e^{i\eta}] s_{2\beta} , \end{aligned} \tag{4.283e}$$

$$\begin{aligned} \lambda_6 e^{i\eta} = {} & \frac{1}{2} (Z_1 c_\beta^2 - Z_2 s_\beta^2 - Z_{345} c_{2\beta} - i \, \mathrm{Im}[Z_5 e^{2i\eta}]) s_{2\beta} \\ & + \mathrm{Re}[Z_6 e^{i\eta}] c_\beta c_{3\beta} + i \, \mathrm{Im}[Z_6 e^{i\eta}] c_\beta^2 + \mathrm{Re}[Z_7 e^{i\eta}] s_\beta s_{3\beta} + i \, \mathrm{Im}[Z_7 e^{i\eta}] s_\beta^2 , \end{aligned} \tag{4.283f}$$

$$\begin{aligned} \lambda_7 e^{i\eta} = {} & \frac{1}{2} (Z_1 s_\beta^2 - Z_2 c_\beta^2 + Z_{345} c_{2\beta} + i \, \mathrm{Im}[Z_5 e^{2i\eta}]) s_{2\beta} \\ & + \mathrm{Re}[Z_6 e^{i\eta}] s_\beta s_{3\beta} + i \, \mathrm{Im}[Z_6 e^{i\eta}] s_\beta^2 + \mathrm{Re}[Z_7 e^{i\eta}] c_\beta c_{3\beta} + i \, \mathrm{Im}[Z_7 e^{i\eta}] c_\beta^2 , \end{aligned} \tag{4.283g}$$

where we have defined $Z_{345} \equiv (Z_3 + Z_4 + \mathrm{Re}[Z_5 e^{2i\eta}])$. For the reverse conversion from the Higgs basis to the general basis, one can perform the same series of identifications, but substituting $\lambda_i \leftrightarrow Z_i$ and $\beta \leftrightarrow -\beta$.

CHAPTER 5

MYSTERIES OF LATE-TIME COSMOLOGY

Much of this thesis has focused on model building in the early universe. In this chapter, we now turn to several open questions of "late-time" cosmology. Note that this division is somewhat superficial, as these questions may also find their solutions in the early universe. By "late-time", we simply mean that these are discrepancies informed by present day observations and data, as opposed to more speculative questions of what may have happened in the early universe. Many of these open questions left unaddressed by ΛCDM and the Standard Model were already discussed at length in Secs. 2.3.1 and 2.3.2. We present here model building efforts to address three of them.

One of the most blatant of these questions is that of the identity of the dark matter constituting $\sim 26\%$ of the universe's energy. Many compelling particle dark matter candidates have been put forward, including several with light masses. Model building in light dark sectors is complicated by constraints from precision measurements of cosmological parameters, and generically requires some mechanism of *asymmetric reheating* in order to populate the dark sector at a lower temperature. In Sec. 5.1, we highlight a novel scheme of asymmetric reheating that can reconcile models of light dark species with observation. Next, we turn to the question of the origin of the supermassive black holes observed at very high redshifts. These challenge the standard picture of supermassive black hole growth from stellar-mass seeds through Eddington-limited accretion. In order to alleviate this timing problem, we put explore the possibility of supermassive black holes of a primordial origin in Sec. 5.2. Finally, another long-standing mystery concerns the origins of the tiny primordial magnetic fields needed to seed the galactic fields observed today. It is generically very difficult to produce magnetic fields with sufficient correlation lengths to survive the diffusive processes of the hot primordial plasma. We put forth a novel mechanism ulilitzing a population of charged, rotating black holes in Sec. 5.3.

5.1 Asymmetric Reheating in Light Dark Sectors

Asymmetric reheating is a generic requirement for models of dark sectors with light species, but its implementation is usually in tension with unique phenomenologies otherwise possible in compelling theories containing dark copies of the Standard Model. In this section, we present a simple module to implement asymmetric reheating during a $\mathbb{Z}_2$-breaking phase *above* some critical temperature. This reinvigorates the possibility of an exactly degenerate mirror sector and the striking phenomenology of composite particles oscillating into their mirror counterparts. Note that this section shares significant text overlap with this author's previous work [28], which was recently submitted for publication.

5.1.1 Overview

The phenomenology of extended dark sectors is powerfully constrained by early universe data, particularly as dark sectors with light degrees of freedom may be probed through purely gravitational effects. Precision measurements of cosmological parameters such as N_{eff} significantly circumscribe particle physics models and have led to the generic need for some mechanism of 'asymmetric reheating', whereby the dark sector is populated at a lower temperature [375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385].

Particularly affected are mirror models, which introduce a $\mathbb{Z}_2$-symmetric copy of the Standard Model (SM) fields and gauge groups [386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396]. Known schemes for implementing asymmetric reheating in mirror models generally require a broken $\mathbb{Z}_2$ in the late universe [397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408][1], which limits perhaps their most interesting observational signature: The oscillations of neutral SM particles into their mirror counterparts. This includes oscillations between SM and mirror neutrinos [411, 412, 413, 414, 415, 416, 417, 418] and photons [419, 420, 421, 422,

1. See [409, 410] for recent, complementary work on asymmetric reheating with minimal symmetry breaking.

423, 424, 425, 426, 427, 428, 429, 430, 431], as well as oscillations between entire composite particles in the limit of an exact $\mathbb{Z}_2$ symmetry. Oscillations between neutrons and mirror neutrons have seen much study in light of neutron lifetime anomalies [432, 433, 410, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447]. More exotically, oscillation of entire hydrogen atoms into their mirror forms has recently been shown to have interesting effects in late-time cosmology [448, 449].

In the related Twin Higgs literature, the cosmological concerns are often dealt with by simply abandoning the full $\mathbb{Z}_2$ symmetry at the level of the spectrum [450, 402, 451]. However, this explicit breaking dramatically restricts the phenomenology of these models, and consequently mirror worlds have received less attention of late. With the aim of reviving these interesting phenomenological possibilities, we seek to have a mirror sector which is exactly degenerate with our own.

Our tool is the richness of phase structures allowed in finite temperature quantum field theory, as first clearly demonstrated by Weinberg [273]. Counter-intuitively, it is possible to have a mirror symmetry which is broken only *above* a critical temperature—a phenomenon known as 'inverse symmetry breaking'. Scalar fields receive corrections to their mass from interactions with other particles in the thermal plasma, and negative cross-quartic interactions with other scalars yield negative contributions to the finite temperature mass. A scalar may then develop a vacuum expectation value (vev) at high temperature when the thermal contribution to its mass dominates [452, 453, 454, 455]. Early concerns that such phenomena might be artifacts of fixed-order perturbation theory have been alleviated by follow-up work on the lattice, robustly evincing high-temperature symmetry-breaking phases [456, 457, 458].

We present here a minimal module to implement the asymmetric reheating of a degenerate mirror sector via inverse symmetry breaking. The idea is to use the high-temperature $\mathbb{Z}_2$-breaking phase to set up an initial asymmetry in the energy densities of the SM and mirror sectors. The immediate model-building challenge is that we are asking for effects derived

from thermal equilibrium to result in a far-out-of-equilibrium configuration. Further, this asymmetry in abundances must persist at late times once symmetry is restored and the sectors become exactly degenerate at low temperatures.

Our strategy will be to use a non-thermal production mechanism—freeze-in [459, 460, 461]—to populate the SM and mirror sectors. We introduce an auxiliary $\mathbb{Z}_2$-breaking sector which is feebly-coupled to heavy right-handed SM and mirror neutrinos N, N', where primes denote mirror species. Annihilations of scalars yield asymmetric abundances due to asymmetric couplings in the $\mathbb{Z}_2$-broken phase. The heavy, non-relativistic N and N' are long-lived and act as reheatons, with the asymmetric number density leading to an asymmetry in reheating temperatures $T_{\rm SM} > T_{\rm mirror}$. See Fig. 5.1 for a schematic timeline.

We consider a theory of three sectors: the SM supplemented with heavy right-handed neutrinos N, a mirror copy (whose species are denoted by primes), and a thermal sector of two real scalar singlets: ϕ_+ and ϕ_-. Under the $\mathbb{Z}_2$ symmetry which exchanges the particles of the SM and mirror sectors, ϕ_- is odd while ϕ_+ is even. After inflation, the scalar sector is reheated to high temperatures and ϕ_- develops a negative thermal mass, breaking the $\mathbb{Z}_2$ symmetry. During this broken phase, ϕ_+ serves to populate the N and N'.

We first review inverse symmetry breaking in a sector of just two scalars, but note that we can have more freedom in realizing this scenario with more fields. We then demonstrate how freeze-in production of heavy right-handed neutrinos during the broken phase can translate to an asymmetry in reheating temperatures.

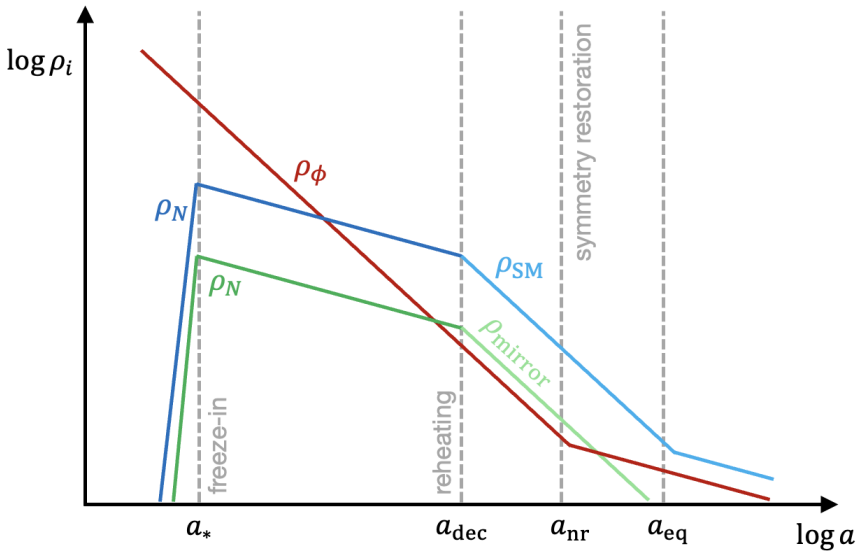

Figure 5.1: A schematic overview of our cosmological timeline. At early times and high temperatures, the scalar sector dominates and the $\mathbb{Z}_2$ is spontaneously broken. At a value of the scale factor a_* corresponding to a temperature $T_* \sim M_N$, freeze-in occurs, resulting in an asymmetric yield of heavy right-handed neutrinos $\rho_N \gg \rho_{N'}$. These come to dominate the universe's energy budget before decaying to asymmetrically reheat the SM and mirror sectors at a time $a_{\rm dec}$ corresponding to $T_{\rm dec} \sim y_\nu \sqrt{M_N M_{\rm Pl}}$. The $\mathbb{Z}_2$ is restored and the ϕ's become non-relativistic at a time $a_{\rm nr}$ corresponding to $T \sim \mu_\pm$, after which they remain a component of the dark matter. See text for definitions and futher details.

5.1.2 *Inverse Symmetry Breaking*

At tree-level, the potential for the scalar sector reads[2]

$$V_0 = \frac{\mu_+^2}{2}\phi_+^2 + \frac{\mu_-^2}{2}\phi_-^2 + \frac{\lambda_+}{4}\phi_+^4 + \frac{\lambda_-}{4}\phi_-^4 + \frac{\lambda_\pm}{4}\phi_+^2\phi_-^2 \,. \tag{5.1}$$

While the quartic self-couplings must be positive for the potential to be bounded from below, $\lambda_\pm$ may be negative provided

$$\lambda_\pm > -2\sqrt{\lambda_+\lambda_-} \,. \tag{5.2}$$

At 1-loop level, the potential receives radiative corrections described by the zero-temperature Coleman-Weinberg potential $V_{\rm CW}$ and the 1-loop thermal potential $V_T^{\text{1-loop}}$,

$$V_{\rm eff}(\phi_i, T) = V_0(\phi_i) + V_{\rm CW}(\phi_i) + V_T^{\text{1-loop}}(\phi_i, T) \,, \tag{5.3}$$

where $i = \pm$ and T denotes the temperature of the scalar sector. See e.g. [462, 298] for a review. The thermal potential dominates for our high temperature regime of interest, at least until new degrees of freedom come in at a scale Λ. Working to leading order in the high-temperature expansion, the quadratic terms from which we find the leading order contributions to the masses are

$$V_T^{\text{1-loop}} \simeq \frac{T^2}{48}(6\lambda_+ + \lambda_\pm)\phi_+^2 + \frac{T^2}{48}(6\lambda_- + \lambda_\pm)\phi_-^2 + ... \tag{5.4}$$

Defining the coefficients

$$c_+ = \frac{1}{24}(6\lambda_+ + \lambda_\pm) \,, \;\; c_- = \frac{1}{24}(6\lambda_- + \lambda_\pm) \,, \tag{5.5}$$

2. Note that for simplicity we have ignored the cubic couplings, since these do not qualitatively affect the phase structure at high temperatures. Further, the assumption that they are negligibly small will be consistent with the naturalness expectations in our freeze-in model.

the masses for ϕ_+ and ϕ_- are

$$M_+^2(T) = \mu_+^2 + c_+ T^2\,, \;\; M_-^2(T) = \mu_-^2 + c_- T^2\,. \tag{5.6}$$

Examining these, it is apparent how phenomena[3] like inverse symmetry breaking can arise from a negative cross-quartic coupling. If $\lambda_\pm < 0$ and $|\lambda_\pm| > 6\lambda_-$, then c_- becomes negative and ϕ_- develops a negative thermal mass $M_-^2(T) < 0$ for sufficiently high temperatures—the hallmark of spontaneous symmetry breaking. Thus, at high temperatures the theory will be in the $\mathbb{Z}_2$-broken phase, while at zero temperature the symmetry will be intact—an instance of inverse symmetry breaking.

The zero-temperature vacuum located at $(\phi_+, \phi_-) = (0, 0)$ is $\mathbb{Z}_2$-symmetric, but as the temperature is increased there is a phase transition at the critical point

$$T_c = \sqrt{\frac{\mu_-^2}{|c_-|}}\,. \tag{5.7}$$

Above this temperature, the theory enters into the broken phase as ϕ_- develops the temperature-dependent vacuum expectation value $\langle \phi_-(T) \rangle \equiv v_-(T)$, given at leading order in the high temperature expansion by

$$v_-(T) = \sqrt{-\frac{1}{\lambda_-}(\mu_-^2 + c_- T^2)} \simeq \sqrt{\frac{|c_-|}{\lambda_-}}\, T\,. \tag{5.8}$$

Note that while the high temperature expansion suffices for our purposes, a more detailed study of the phase transition and precise predictions for quantities like the critical temperature would require going beyond the 1-loop approximation and performing thermal resummation of the effective potential, as the perturbative expansion breaks down in the

3. There has recently been much interest in the related phenomenon of 'symmetry non-restoration'—where the zero-temperature mass is also negative—which may have applications for the electroweak phase transition and baryogenesis [463, 7, 464, 465, 466, 467, 468, 276, 469, 470].

infrared [295, 471, 462, 298]. These technicalities will not concern us here, as our purpose is not to study this sector in detail. See Sec. 4.2 for an in-depth discussion of resummation of the finite temperature effective potential.

5.1.3 Freeze-In Production

We wish to take advantage of this high-temperature $\mathbb{Z}_2$-broken phase to establish an asymmetry in the energy densities of the SM and mirror sectors. The freeze-in mechanism is a natural candidate to accomplish this since it populates states which are never in equilibrium with the thermal sector.

We will focus on the following two portal operators between the auxiliary scalar sector and heavy right-handed neutrinos N and N' of mass M_N

$$-\mathcal{L}_N = \lambda\,\phi_+\,(NN + N'N') + \frac{C}{\Lambda}\,\phi_+\,\phi_-\,(NN - N'N')\,, \tag{5.9}$$

where the dimension-5 operator may be generated by integrating out heavier fields at the scale Λ. In the high-temperature phase, it is convenient to define the effective couplings

$$\lambda_N(T) \equiv \lambda\left(1 + \frac{v_-(T)}{\Lambda_{\text{eff}}}\right)\,,\;\; \lambda_{N'}(T) \equiv \lambda\left(1 - \frac{v_-(T)}{\Lambda_{\text{eff}}}\right)\,, \tag{5.10}$$

with $\Lambda_{\text{eff}} \equiv \frac{\lambda}{C}\Lambda$, in terms of which

$$-\mathcal{L}_N = \lambda_N(T)\phi_+ NN + \lambda_{N'}(T)\phi_+ N'N'\,. \tag{5.11}$$

Note that λ now controls the overall size of the freeze-in production while Λ_{eff} controls the timing.

In order to prevent equilibration of the SM and mirror sectors with the thermal sector, we require $\lambda_N \ll 1$, which restricts λ as well as the maximum temperature at which this

effective theory remains sensible. During the broken phase $\lambda_N(T) > \lambda_{N'}(T)$, such that ϕ will couple preferentially to N.

Taking the initial abundance of N to be vanishing, the Boltzmann equation governing the evolution of the number density n_N is given by

$$\dot{n}_N + 3Hn_N \simeq (n_+^{\rm eq})^2 \langle \sigma v \rangle \ , \tag{5.12}$$

where $n_+^{\rm eq} = \frac{T}{2\pi^2} M_+^2 K_2(\frac{M_+}{T})$ is the equilibrium number density for ϕ_+ and $\langle \sigma v \rangle$ is the thermally-averaged annihilation cross section for the production of N. We presume the neutrinos are very heavy $M_+ < 2M_N$ and that their direct couplings to ϕ_- are somewhat smaller than those to ϕ_+, such that the dominant process contributing to their production will be the $2 \to 2$ annihilation $\phi_+\phi_+ \to NN$. This occurs at tree level via t- and u-channel diagrams with an amplitude $|\mathcal{M}_{++\to NN}|^2$ and cross section $\sigma_{++\to NN}$. The thermal average appearing in Eq. (5.12) is then obtained by performing an integral over the squared center of mass energy s, in the manner described in [472]. Crucially, the freeze-in rate will be proportional to $\lambda_N(T)^4$, enhancing the effect of the asymmetry in couplings.

To solve the Boltzmann equation, it is more convenient to work with the yield $Y_N = \frac{n_N}{S}$ and reparameterize in terms of temperature via $\frac{d}{dt} \simeq -HT\frac{d}{dT}$, valid when the number of relativistic degrees of freedom in the bath remains roughly constant. Then the left-hand side becomes $\dot{n}_N + 3Hn_N = -HTS\frac{dY_N}{dT}$. Integrating, the yield as a function of temperature is

$$\begin{aligned} Y_N(T) = & \frac{2}{(4\pi)^5} \int_T^{T_{\rm max}} dT' \frac{1}{H(T')S(T')} \\ & \times \int_{4M_\phi^2}^\infty ds \, \frac{\sqrt{s-4M_N^2}\sqrt{s-4M_+^2}}{4\sqrt{s}} K_1\left(\frac{\sqrt{s}}{T'}\right) \\ & \times \int_{-1}^1 d\cos\theta \, |\mathcal{M}_{++\to NN}(s,T',\theta)|^2 \, , \end{aligned} \tag{5.13}$$

where K_1 is a modified Bessel function of the second kind, $H = \sqrt{\frac{4\pi^3}{45} g_\star} \frac{T^2}{M_{\rm Pl}}$, $S = \frac{2\pi^2}{45} g_{\star,S} T^3$, and $g_\star \simeq g_{\star,S} \simeq 2$ for our scalar sector. An analogous expression holds for the N' yield. The crucial difference is that the coupling $\lambda_{N'}$ becomes vanishingly small at a temperature

$$T_\star \equiv \sqrt{\frac{\lambda_-}{|c_-|}}\, \Lambda_{\rm eff}\,, \tag{5.14}$$

presuming that production takes place at temperatures far greater than μ_-. Thus if the dominant freeze-in production takes place around $T_\star$, the result will be a much smaller abundance of N'.

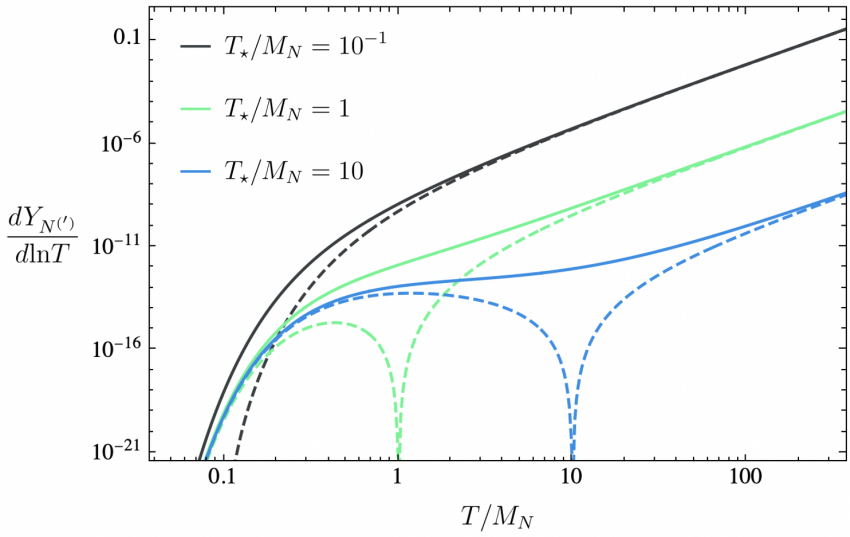

Figure 5.2: Logarithmic differential yield of N (solid) and N' (dashed) as a function of temperature. Parameters fixed as $M_N = 10^{14}$ GeV and $\lambda = 4 \times 10^{-5}$; changing either just results in an overall vertical translation.

What should we expect for the yield curve as a function of temperature? If ϕ_-'s vev were temperature independent, this would be a $2 \to 2$ freeze-in through marginal operators, and so infrared dominated. The production rate would be largest at $T \sim M_N$, after which the process would become Boltzmann suppressed. This suggests $M_N \sim T_\star$ should generate appreciably asymmetric abundances. Indeed, this is observed in Fig. 5.2.

However, $T_\star$ is also the temperature above which the couplings $\lambda_{N^{(\prime)}}(T)$ become dom-

inated by the term linear in the vev. This additional temperature-dependence results in a yield which is sensitive to the high-temperature initial conditions, as in ultraviolet freeze-in. In order for the symmetric production at high temperatures to not exceed the infrared contribution[4], spoiling our mechanism, we require the process shut off at some $T_{\max}$ not much larger than $T_\star$. Note that the consideration $M_+(T) < 2M_N$ also restricts $T_{\max}$ to lie not too far above M_N, or more precisely $T_{\max} \lesssim \frac{2}{c_+} M_N$. These requirements are reflected in Figs. 5.3 and 5.4.

One possibility is for the scalar sector to have only begun at $T_{\max}$ following inflationary reheating. Alternatively, since our calculations must anyway have some $T_{\max} \lesssim \Lambda$ where the EFT breaks down, it is possible that the degrees of freedom at Λ which have generated the dimension-5 operator also contribute to the effective potential at this scale. The resultant modification to the quartic couplings could cause c_- to flip signs, such that at higher energies we are once again in the $\mathbb{Z}_2$-symmetric phase. With zero vev, the freeze-in contribution from higher temperatures becomes negligible. In any case, our analysis will stay agnostic to the physics of $T_{\max}$.

5.1.4 Asymmetric Reheating

Reheating of the SM and mirror sectors occurs via the out-of-equilibrium decays of N and N', respectively. As a prerequisite, we should first ensure that the massive neutrinos are sufficiently long-lived that they come to dominate over the radiation energy density in the thermal scalar bath before they decay (see Fig. 5.1). Let $R = \rho_N/\rho_\phi$ be the ratio of energy density in N to that in the scalar bath. The initial value R_* is set by freeze-in and quantitatively ranges from 10^{-8} to 10^{-12} for the parameter space in which we can have a successful asymmetric reheating in our toy model. Since the heavy neutrinos are non-

4. Another potentially problematic symmetric contribution comes from the gravitational production of N/N' during inflationary reheating. Graviton-mediated scattering leads to a freeze-in yield with a rate $R_{1/2} \sim T^8/M_{\rm Pl}^4$ [473, 474], which is easily subdominant so long as $T_{\max}$ is not too close to the Planck scale.

relativistic, their energy density dilutes as $\rho_N \propto a^{-3}$ while that in the scalar bath falls as $\rho_\phi \propto a^{-4}$, meaning R grows as $a \sim 1/T$. By the time of N decay, $R_{\rm dec} = T_* R_* / T_{\rm dec} \simeq M_N R_* / T_{\rm dec}$. The temperature $T_{\rm dec}$ at which N decays is roughly set by $\Gamma_N \sim y_\nu^2 M_N \simeq H(T_{\rm dec})$, allowing us to identify $T_{\rm dec} \sim y_\nu \sqrt{M_N M_{\rm Pl}}$. For N to dominate at decay, we require $R_{\rm dec} > 1$, corresponding to the upper bound on the Yukawa coupling

$$y_\nu < \sqrt{\frac{M_N}{M_{\rm Pl}}} R_* \,. \tag{5.15}$$

Thus N can easily be made to dominate at decay by turning down the Yukawa coupling $y_\nu \tilde{H} L N$, which is technically natural.

To determine the parameter space corresponding to a successful reheating, we should calculate the final ratio of temperatures $x_{\rm RH} = T_{\rm mirror}/T_{\rm SM}$, which must be sufficiently small, as well as the absolute scale of the SM reheating temperature, which should be at least $T_{\rm RH} \gtrsim 10\,{\rm MeV}$ to ensure the predictions of big bang nucleosynthesis (BBN) are unaffected. In the instantaneous decay approximation, we can estimate the SM reheating temperature $T_{\rm RH}$ as

$$T_{\rm RH} = \left(\frac{30}{\pi^2 g_\star} \rho_N(t_{\rm dec}) \right)^{1/4} , \tag{5.16}$$

where $g_\star$ now counts the SM degrees of freedom at $T_{\rm RH}$, and $\rho_N(t_{\rm dec})$ is the energy density in N at their decay.

We define the ratio of energy densities

$$x_{\rm RH} \equiv \left(\frac{\rho_{N'}(t_{\rm dec})}{\rho_N(t_{\rm dec})} \right)^{1/4} \simeq \left(\frac{Y_{N'}}{Y_N} \right)^{1/4} , \tag{5.17}$$

and note that $x_{\rm RH}$ coincides with the final ratio of temperatures $T_{\rm mirror}/T_{\rm SM}$ at late times once the only remaining light degrees of freedom in each sector are the photon and active neutrinos, provided the asymmetry is not erased by processes which bring the SM and mirror sectors into thermal equilibrium. We have verified that the rate for scalar mediated N-N'

scattering satisfies $\Gamma = n\langle\sigma v\rangle \lesssim H$ and so is inefficient for all parameter space of interest. In Fig. 5.3 we plot the values of $x_{\rm RH}$ that can be realized in this toy model.

The light species of the mirror sector contribute to the excess radiation energy density, as parameterized by the change in the effective number of neutrino species $\Delta N_{\rm eff}$,

$$\Delta N_{\rm eff} \simeq \frac{29}{7}\left(\frac{11}{4}\right)^{4/3} x_{\rm RH}^4 \,. \tag{5.18}$$

Demanding $\Delta N_{\rm eff} \lesssim 0.5$, corresponding to the 2σ constraint from Planck [475], requires $x_{\rm RH} \lesssim 0.42$. Comparing with Fig. 5.3, we see that this is indeed achievable provided T_* is not too far from $T_{\rm max}$.

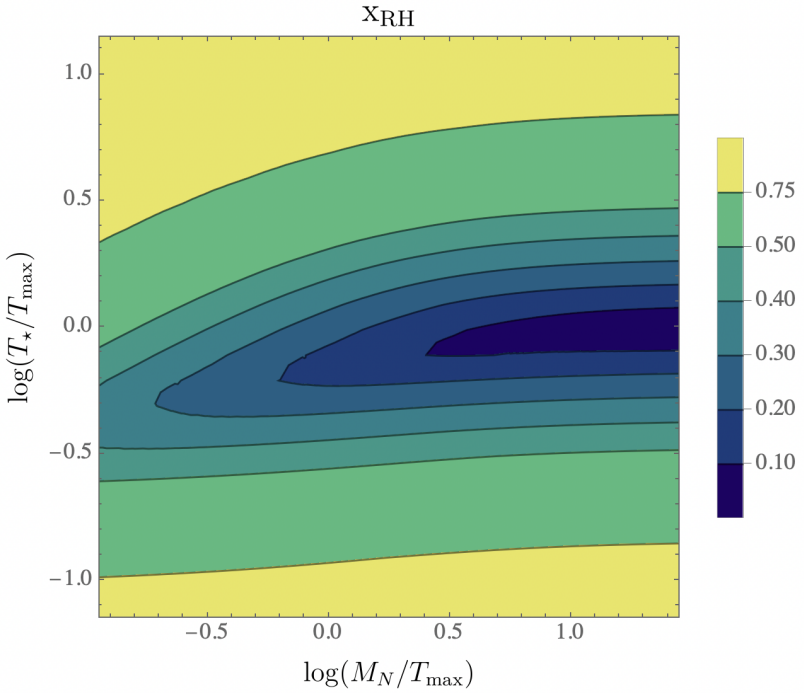

Figure 5.3: The ratio of energies injected into the mirror and SM sectors, as a function of ratios of important scales. The overall energy scale is arbitrary so long as the scalar sector is in the high-temperature regime.

To obtain the absolute scale of the reheating temperature, we must track the evolution of the energy densities from freeze-in to decay. Since we choose values for the couplings such that the sectors remain decoupled, the dominant effect governing the evolution is simply

dilution due to cosmic expansion. For heavy neutrinos which decay not too long after coming to dominate the universe, we have the following approximate expression for the SM reheating temperature,

$$T_{\rm RH} \simeq \frac{4}{3}\left(\frac{2}{g_\star}\right)^{\frac{1}{4}}(1+x_{\rm RH}^4)^{\frac{3}{4}}Y_N M_N\,. \tag{5.19}$$

Note the product $Y_N M_N$ is insensitive to the overall scale of freeze-in, since the only other scale in Eq. (5.13) is a factor $M_{\rm Pl}$ from Hubble. The very rough estimate $T_{\rm RH} \sim \lambda^4 M_{\rm Pl}$ works surprising well, as observed in Fig. 5.4. We require that the SM is reheated to at least $T_{\rm RH} \gtrsim 10$ MeV to ensure BBN is not substantially affected. This limits the absolute scale of the yield and prevents realizing $T_{\rm max} \ll M_N$ and freezing in solely during the Boltzmann tail, despite this still producing a large asymmetry of yields.

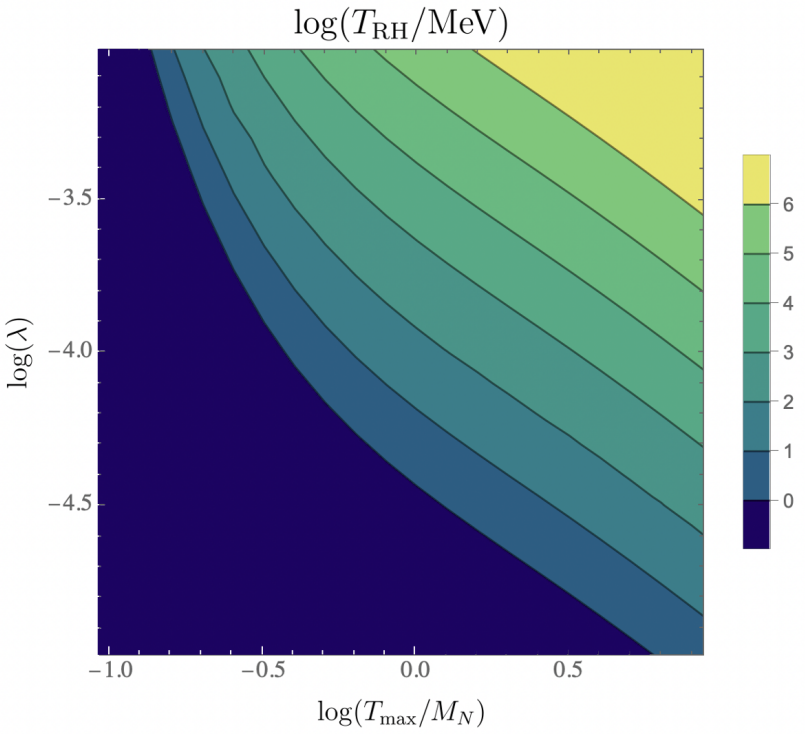

Figure 5.4: Maximum reheating temperature of the SM sector, fixing $T_\star = M_N$. $T_{\rm RH}$ may be turned down by moving to $T_\star \ll M_N$ or by further increasing the neutrino lifetime.

Finally in order to ensure a consistent late-time cosmology, we turn to the fate of the scalar sector. The leading decay channel for both ϕ's is to active neutrinos and proceeds through off-shell heavy neutrinos with a heavily suppressed rate $\Gamma_\phi \sim \lambda^2 y_\nu^4 T^5/M_N^4$ at early times and $\Gamma_\phi \sim \lambda^2 y_\nu^4 (v_h/M_N)^4 \mu_\pm$ at late times once the scalars have become non-relativistic, where v_h is the Higgs vev. Crucially these rates go as y_ν^4, and given the tiny values of y_ν required for the massive N's to dominate the energy density prior to decay, the corresponding scalar lifetimes can easily be made significantly longer than the age of the universe. It will thus generically be the case that the scalars are stable on cosmological time scales.

The scalars must be non-relativistic by BBN so as not to contribute to $\Delta N_{\rm eff}$, which restricts the bare masses $\mu_\pm \gtrsim 10\,\text{MeV}$. Being non-relativistic and stable at late times, the scalars then constitute some component of the dark matter (DM), and are harmless as long as their relic abundance is not too large.

To check this, we define the new ratio $\tilde{R} = \rho_{\rm SM}/\rho_\phi$, and demand $\tilde{R} \gtrsim 1$ from the time the heavy neutrinos decay up through shortly after matter-radiation equality, such that ρ_ϕ does not come to dominate appreciably over the SM radiation bath. The assumption of instantaneous decay $\rho_N \simeq \rho_{\rm SM}$ leads to the initial condition $\tilde{R}_{\rm dec} \simeq R_{\rm dec}$. Neglecting SM entropy dumps, the ratio remains roughly fixed until the scalars become non-relativistic at some time $t_{\rm nr}$ corresponding to $T_{\rm nr} \sim \mu_\pm$, leading to $\tilde{R}_{\rm nr} \simeq \tilde{R}_{\rm dec}$. Afterwards, ρ_ϕ will begin to grow relative to the still-relativistic $\rho_{\rm SM}$, leading $\tilde{R}$ to decrease as $1/a \sim T$. Demanding that $\tilde{R} \gtrsim 1$ through matter-radiation equality of the SM sector imposes the constraint $R_{\rm dec} \gtrsim 10^7 (\mu_\pm/10\,\text{MeV})$, implying

$$y_\nu \lesssim 10^{-7} R_* \sqrt{\frac{M_N}{M_{\rm Pl}}} \left(\frac{10\,\text{MeV}}{\mu_\pm}\right) . \tag{5.20}$$

This is a stronger condition than Eq. (5.15), but can still be compatible with technically natural values for y_ν. For example, a benchmark point for successful asymmetric reheating with $R_* = 10^{-8}$ and $M_N = 10^{17}\,\text{GeV}$ would correspond to at maximum $y_\nu \sim 10^{-16}$.

5.1.5 Summary

In this section, we have constructed a model of asymmetric reheating using the finite temperature phenomenon of inverse symmetry breaking. Our focus has been on constructing a minimal realization of this mechanism, which has the benefit of providing a module which may be annexed onto a variety of theories. This reinvigorates the well-motivated scenario of degenerate mirror models and the rich phenomenology that accompanies them. A clear direction for future work is to further integrate this into such models or other theories of the early universe—perhaps exploring connections to leptogenesis or further developing the connection to dark matter.

Finally, this mechanism does require a rough confluence of scales to produce an appreciable temperature asymmetry. We emphasize that this is not an instance of fine-tuning — to ask that the dimensionful scales in a new sector be of the same order of magnitude is exactly what one expects in a natural theory where there is some underlying scale Λ and the relevant physics is controlled by this scale and order-one couplings. Still, it would be pleasing to study concrete models where, for example, $T_\star$ and $T_{\rm max}$ arise from the same additional degrees of freedom interacting with ϕ_+ and ϕ_-.

5.2 High-Redshift Supermassive Black Holes

There is controversy surrounding the origin and evolution of our universe's largest supermassive black holes (SMBHs). In this study, we consider the possibility that some of these black holes formed from the direct collapse of primordial density perturbations. Since the mass of a primordial black hole is limited by the size of the cosmological horizon at the time of collapse, these SMBHs must form rather late, and are naively in conflict with constraints from CMB spectral distortions. These limits can be avoided, however, if the distribution of primordial curvature perturbations is highly non-Gaussian. After quantifying the departure from Gaussianity needed to evade these bounds, we explore a model of multi-field inflation —

a non-minimal, self-interacting curvaton model — which has all the necessary ingredients to yield such dramatic non-Gaussianities. We leave the detailed model building and numerics to a future study, however, as our goal is to highlight the challenges associated with forming SMBHs from direct collapse and to identify features that a successful model would need to have. This study is particularly timely in light of recent observations of high-redshift massive galaxy candidates by the James Webb Space Telescope as well as evidence from the NANOGrav experiment for a stochastic gravitational wave background consistent with SMBH mergers. Note that much of the material in this section was previously published by this author in [24].

5.2.1 Overview

Supermassive black holes (SMBHs) are ubiquitous in our universe, being present in the centers of nearly all massive galaxies. Quasars,[5] powered by black holes with with masses $M \sim 10^{8-10}\, M_\odot$, are also found in large numbers in the high-redshift universe. At present, over 170 quasars have been observed at $z > 6$, with the most distant at $z = 7.54$, and several hundred others at $z = 5 - 6$ [476, 477, 478, 479, 480, 481, 482, 483, 484]. Fig. 5.5 shows the quasar abundance as a function of redshift over several ranges of black hole mass. While this magnitude limited quasar catalog is believed to be nearly complete out to $z \sim 5$, only a small fraction of SMBHs are, in fact, quasars. A more complete census of SMBHs is possible in the local universe where one finds $\Omega_{\rm SMBH}(z = 0) \sim 10^{-6}$ [485]. This contrasts with the peak quasar mass density of $\Omega_{\rm quasar}(z = 2) \sim 10^{-8}$. While specific models for SMBH population evolution have been proposed [486, 487, 488], the limited available data leaves a great deal of uncertainty.

5. Quasars are the most luminous active galactic nuclei (AGN). In this paper, quasar refers to an AGN that is sufficiently luminous to appear in a quasar catalog such as SDSS DR7 [8]. While this definition is redshift-dependent, the most luminous quasars should be consistently present in the catalog up to the DR7 redshift limit of 5. In the text, the phrase "quasar mass" refers to the mass of the black hole that powers the quasar.

It is usually assumed that SMBHs grow over time from relatively low-mass seeds (possibly the remnants of Population III stars [489]) through the process of accretion. The rate of mass accretion is Eddington limited to

$$\dot{M}_{\rm Edd} \lesssim \frac{M_{\rm BH}}{\tau_S}, \tag{5.21}$$

where

$$\tau_{\rm S} = \frac{\varepsilon\,\sigma_{\rm T}}{4\pi\,G\,m_{\rm p}} \approx 45\ {\rm Myr}\left(\frac{\varepsilon}{0.1}\right) \tag{5.22}$$

is the Salpeter time [490], $\sigma_{\rm T} = 8\pi\,\alpha^2/(3m_e^2)$ is the Thomson cross section, m_p is the proton mass, and ε is the radiative efficiency. At the Eddington-limited rate, a $10^2\,M_\odot$ black hole seed would require $\sim 0.8\,{\rm Gyr}$ to grow to $M \sim 10^{10}\,M_\odot$. Since $z \sim 6-7$ corresponds to only $\sim 0.7-0.9\,{\rm Gyr}$ after the Big Bang, such a scenario would require these large high-redshift black holes to have grown at high accretion rates almost continuously throughout the first Gyr of our universe's history. Interestingly, there seems to be quite a significant population of quasars at $z \sim 6$ to 7 in this mass range [491, 478, 492, 493].

The continuous accretion required to explain this rapid growth contrasts with the intermittent accretion observed of SMBHs at lower redshifts. Further, from Fig. 5.5 we see that the most massive $M_{\rm BH} \gtrsim 10^{10}\,M_\odot$ population has remained approximately constant since at least $z \sim 5$ [494]. This would require any growth in the number of SMBHs to be balanced by a decrease in the fraction of SMBHs that are actively quasars. From this perspective, it is surprising that so many highly-massive quasars have been observed at such high redshifts [495, 496, 497, 498, 499]. These observations prompt two intriguing questions:

1. If these quasars grew from small black hole seeds, how did they come to be so massive on such a short timescale?

2. Why did their growth rate dramatically slow down during the subsequent 13 Gyr?

Various solutions to the first question have been put forth, including an enhanced role

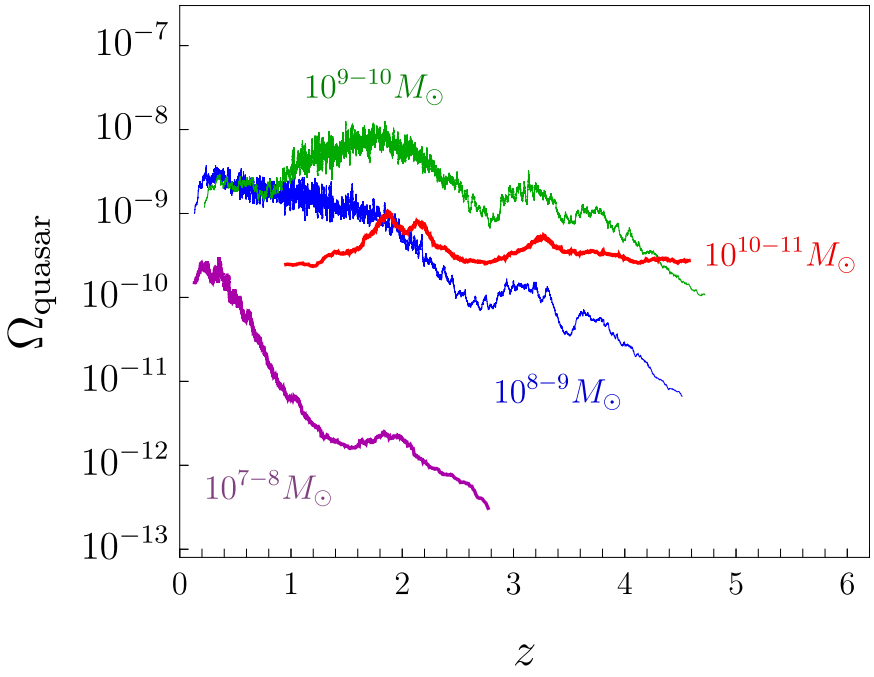

Figure 5.5: Estimates of the quasar black hole comoving mass density in units of the $z = 0$ critical density as a function of redshift, plotted for the four different mass classes indicated. Black hole masses and redshifts are taken from the DR7 SDSS quasar survey [8], which covers 20% of the sky and the redshift range $0 < z < 5$. The results shown represent the moving average of 100 quasars sorted by redshift.

of mergers as well as the possibility of super-Eddington accretion [118]. However frequent mergers would require more heavily clustered populations in the early universe, and further can have both positive and negative impacts on SMBH growth, since they can also kick SMBHs out of the material-rich centers of galaxies. A certain degree of super-Eddington growth is expected to occur in high-redshift galaxies containing large reservoirs of gas and efficient angular momentum transport due to turbulence. However as increased accretion produces powerful jets and outflows which drive material away, it is uncertain how long it can be sustained. Feedback effects from transient periods of super-Eddington growth are actually expected to impede SMBH growth within a few Myr [119].

Regarding the second question, it has been suggested that the suppressed growth of the most massive black holes after the first Gyr could perhaps be attributed to galaxy-scale

feedback. This, however, would require the M-σ relation[6] to evolve with redshift and for the quasar luminosity function to steepen at the highest values [500, 501]. Alternatively, it has been proposed that there might be a maximum mass that black holes can reach through accretion, resulting from the fragmentation of the accretion disks that could have otherwise facilitated rapid black hole growth [502]. Despite these suggestions, there remain many open questions concerning the origin and evolution of our universe's most massive black holes.

In this study, we take these questions to motivate another possibility: that our universe's most massive black holes did not acquire most of their mass through accretion, but are instead predominantly primordial in origin[7]. Unlike the small primordial black holes (PBH) typically considered as dark matter candidates, these massive objects would have formed at "late" times, as governed by the size of the cosmological horizon. During the radiation dominated era, the horizon contains the following amount of energy

$$M_H = \frac{4\pi}{3}\rho R_H^3 \approx 3 \times 10^9 \, M_\odot \left(\frac{10\,\mathrm{keV}}{T}\right)^2 \left(\frac{3.36}{g_\star(T)}\right)^{1/2}, \tag{5.23}$$

where $R_H = H^{-1}$ is the size of the horizon, $\rho = \pi^2 g_\star(T) T^4/30$ is the radiation density, and $g_\star$ is the number of relativistic species at temperature T. When a sufficiently large density fluctuation collapses to form a PBH,[8] the mass of the resulting black hole is typically an order one fraction of the horizon mass, $M_{\rm BH} \simeq \gamma \, M_H$, where $\gamma \sim 0.2$ quantifies the efficiency of collapse [509]. From Eq. (5.23), we conclude that the PBHs in the mass range of interest here

6. The M-σ relation is the observed correlation between the velocity dispersion of a stellar bulge σ and the mass of the SMBH at its center.

7. Primordial black holes (PBHs) have previously been considered as SMBH seeds [120]. Unlike population III stars, which do not form until $z \sim 30 - 20$, PBH seeds form much earlier, and so the timing problem is allayed. Refs. [503, 504] have argued that establishing a population of PBHs with mass $10^2 - 10^6 M_\odot$ by $z \sim 20$ would be sufficient to seed even the most massive SMBHs. Here we consider the more exotic possibility of SMBHs from direct collapse, with little growth through accretion required.

8. There exist other PBH formation mechanisms beyond the collapse of overdensities seeded by inflation. For example, PBH could form during first-order cosmological phase transitions [505] or due to the collapse of supercritical vacuum bubbles nucleated during inflation [506]. The prospect of primordial SMBHs from these mechanisms is discussed in Ref. [507] and Ref. [508], respectively.

form after the start of Big Bang nucleosynthesis ($T \sim$ MeV corresponds to $M_H \sim 10^5\, M_\odot$) but well before the onset of matter domination or recombination ($T \sim$ eV corresponds to $M_H \sim 3 \times 10^{17}\, M_\odot$).

Many inflationary models enable PBHs to form efficiently [510], including those in which the inflaton undergoes a period of ultra-slow-roll [511, 161, 162, 163], or whose potential features localized bumps, dips, or steps on small scales [164, 165, 512]. More generically, a local enhancement of the power spectrum $\mathcal{P}_\zeta$ requires a deviation from slow-roll evolution (see for example, Refs. [513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 153, 526]). Regardless of the mechanism, PBH formation requires a significant enhancement in the amplitude of the primordial power spectrum. Assuming Gaussian statistics, this amplitude must be $\mathcal{P}_\zeta(k_{\rm BH}) \sim 10^{-2}$, seven orders of magnitude larger than observed on large scales, $\mathcal{P}_\zeta(k_{\rm CMB}) \simeq 2.1 \times 10^{-9}$ [2]. Such an amplification inevitably leads to large CMB spectral distortions. In particular, for scales $k_{\rm BH} \lesssim 10^5\,{\rm Mpc}^{-1}$, corresponding to black holes in the mass range $M_{\rm BH} \gtrsim 10^3 M_\odot$, the predicted spectral distortions are in strong conflict with COBE/FIRAS measurements [12, 13].

In principle, PBHs can form from smaller peaks in the power spectrum if the tail of the ζ distribution is sufficiently non-Gaussian. Of course, observations on large scales and measurements of the non-linearity parameters $f_{\rm NL}$ and $g_{\rm NL}$ seem to indicate that ζ is very nearly Gaussian on CMB scales [527, 10]. However the scales on which PBHs are formed can be disconnected from CMB scales, and so these bounds need not apply.

In this section, we quantify the degree of non-Gaussianity that would be required to viably produce primordial SMBHs and present a model of inflation that contains the necessary features. We refer the reader to Secs. 3.1.1 and 3.1.2 for reviews of primordial black hole formation and achieving enhanced small scale power in inflationary models, respectively. In Sec. 5.2.2, we examine how measurements of spectral distortions constrain the primordial power spectrum, and demonstrate that the appreciable formation of SMBHs from Gaussian

density fluctuations is excluded on the basis of these constraints. Sec. 5.2.3 considers departures from Gaussianity, and quantifies the heaviness of the distribution needed to evade these bounds. In Sec. 5.2.4, we review the calculation of the curvature perturbation and its statistics in the standard curvaton scenario, which has been shown to be capable of producing appreciable non-Gaussianity. This model, however, cannot viably produce a significant abundance of primordial SMBHs, so in Sec. 5.2.5 we introduce self-interactions to augment the non-Gaussianity and speculate on the maximum PBH abundance in this model. We conclude in Sec. 5.2.6 with a discussion of our results and some comments on directions for future investigations.

5.2.2 Spectral Distortions

The small scales relevant for primordial SMBH formation are well below the angular resolution probed by current CMB measurements. Nevertheless, inhomogeneities on these scales will generate isotropic deviations from the usual blackbody spectrum [528, 529, 530, 531, 532]. These deviations are known as spectral distortions and in this context it is useful to distinguish between three characteristic redshift intervals:

- **Thermalization era ($z > 2 \times 10^6$)**: At high redshifts, Compton scattering $\gamma e \to \gamma e$, double Compton scattering $\gamma e \to \gamma\gamma e$, and Bremsstrahlung $ep \to ep\gamma$ maintain a blackbody spectrum for the photons, and spectral distortions are exponentially suppressed.

- **μ-era ($2 \times 10^5 < z < 2 \times 10^6$)**: During this era, photon number changing processes, double Compton scattering and Bremsstrahlung, become ineffective at maintaining a blackbody spectrum. Compton scattering, however, continues to redistribute photon energies to maintain a Bose-Einstein distribution, parameterized by both a temperature, T, and a chemical potential, μ. A μ-distortion refers to a Bose-Einstein distribution with $\mu \neq 0$.

- ***y*-era ($z < 2\times10^5$)**: Compton scattering becomes ineffective at redistributing photon energies during this era, so there are no processes to maintain a Bose-Einstein distribution. Spectral distortions that are generated at these redshifts are characterized by a departure from an equilibrium distribution, and often yield so-called y-distortions.[9]

A spectrum with a (positive) y-distortion can be expressed as an average of blackbodies with slightly different temperatures [534, 535, 536, 537]. An average of blackbodies with a mean temperature $\bar{T}$ and variance $\bar{T}^2\,\Delta$ will (for $\Delta \ll 1$) be a y-distorted blackbody characterized by the temperature $T = \bar{T}\,(1+\Delta^2)$ and the parameter $y = \Delta^2/2$. y-type distortions can be generated through a variety of mechanisms, including the Compton scattering of CMB photons with a population of electrons with a different temperature. In this paper, we will be interested in y-distortions that are generated through photon diffusion. Limits on such spectral distortions allow us to constrain the amplitude of inhomogeneities on very small scales [538].

μ- and y-type spectral distortions are traditionally quantified in terms of the parameters μ and y, which are related to the fractional increase in energy per photon (relative to a blackbody spectrum with the same number density of photons) [528, 532]:

$$\mu \simeq 1.4\,\frac{\Delta\rho_\gamma}{\rho_\gamma} \qquad \text{and} \qquad y \simeq 0.25\,\frac{\Delta\rho_\gamma}{\rho_\gamma}\ . \tag{5.24}$$

By introducing k-space window functions accounting for the effects of thermalization and dissipation, μ and y can be approximately calculated from the spectrum of density pertur-

9. Technically, the division between the two types of spectral distortions is not entirely unambiguous, and inhomogeneities dissipating around $z \sim 5\times10^4$ can give rise to distortions of an intermediate type, whose shape is not simply the sum of μ- and y-type distortions [533].

bations, P_ζ [536, 532]:

$$\mu \simeq 2.2 \int_{k_{\rm min}}^{\infty} \frac{dk}{k} \mathcal{P}_\zeta(k) \left[\exp\left(-\frac{k}{5400\,{\rm Mpc}^{-1}} \right) - \exp\left(-\left[\frac{k}{31.6\,{\rm Mpc}^{-1}} \right]^2 \right) \right] , \tag{5.25}$$

$$y \simeq 0.4 \int_{k_{\rm min}}^{\infty} \frac{dk}{k} \mathcal{P}_\zeta(k) \exp\left(-\left[\frac{k}{31.6\,{\rm Mpc}^{-1}} \right]^2 \right) , \tag{5.26}$$

where $k_{\rm min} = 1\,{\rm Mpc}^{-1}$.

The strongest existing constraints on spectral distortions come from the COBE/FIRAS instrument, which restricts $|\mu| \lesssim 9.0\times10^{-5}$ and $|y| \lesssim 1.5\times10^{-5}$ at the 95% C.L. [12, 13]. The standard cosmological model predicts spectral distortions of $\mu \sim 2\times10^{-8}$ and $y \sim 10^{-6}$ [539], consistent with current limits. The models of interest in this study have enhanced P_ζ, and thus enhanced μ and y, which can be constrained by CMB measurements.

For concreteness, consider the case of a power spectrum that is sharply peaked at a single scale, $k_{\rm BH}$:

$$\mathcal{P}_\zeta(k) = \sigma_\zeta^2 \, k \, \delta(k - k_{\rm BH}). \tag{5.27}$$

Using the delta-function to perform the integral, μ and y become functions of $k_{\rm BH}$ and σ_ζ^2 alone. From the horizon crossing condition $k = aH$, and fact that $H = 1.66\sqrt{g_\star(T)}T^2/M_{\rm Pl}$ during radiation domination, we can relate the wavenumber to the temperature,

$$k_{\rm BH} = 92\,{\rm Mpc}^{-1} \left(\frac{T}{10\,{\rm keV}} \right) \left(\frac{g_\star(T)}{3.36} \right)^{1/2} \left(\frac{3.91}{g_{\star,S}(T)} \right)^{1/3} , \tag{5.28}$$

which can then be related to the horizon mass using Eq. (5.23) to estimate the mass of the resulting black hole:

$$M_{\rm BH} \simeq 5 \times 10^8 \, M_\odot \left(\frac{92\,{\rm Mpc}^{-1}}{k_{\rm BH}} \right)^2 \left(\frac{\gamma}{0.2} \right) \left(\frac{g_\star(T)}{3.36} \right)^{1/2} \left(\frac{3.91}{g_{\star,S}(T)} \right)^{2/3} . \tag{5.29}$$

For the sharply peaked spectrum of Eq. (5.27), adopting $\gamma = 0.2$, and for black holes in the

mass range of interest, the μ- and y-parameters can be written as

$$\mu \simeq 2.2\sigma_\zeta^2 \left\{ \exp\left[-\left(\frac{1.5\times 10^5\, M_\odot}{M_{\rm BH}} \right)^{1/2} \right] - \exp\left[-\left(\frac{4.5\times 10^9\, M_\odot}{M_{\rm BH}} \right) \right] \right\}, \tag{5.30}$$

$$y \simeq 0.4\sigma_\zeta^2 \exp\left[-\left(\frac{4.5\times 10^9\, M_\odot}{M_{\rm BH}} \right) \right]. \tag{5.31}$$

Note that these results should also hold, for example, in the case of a log-normal spectrum of sufficiently narrow width.

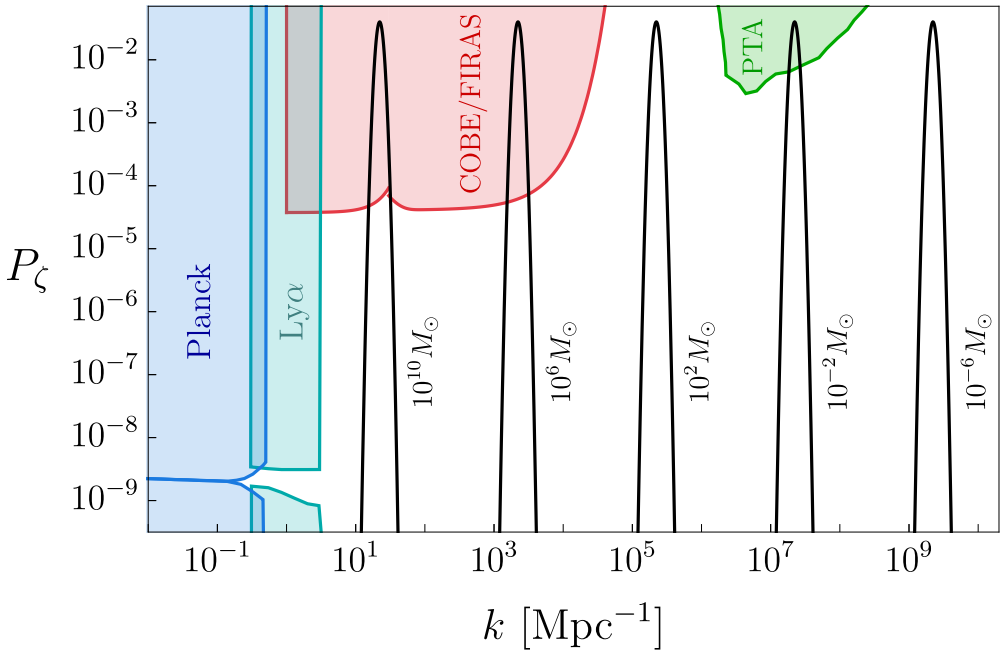

Figure 5.6: Constraints on the primordial power spectrum $\mathcal{P}_\zeta$ [9] coming from CMB temperature anisotropies (dark blue) [10], Lyman-α forest (light blue) [11], CMB spectral distortions (red) [12, 13], and pulsar timing arrays (green) [14]. The cusp in the COBE/FIRAS excluded region signifies the wavenumber where constraints from μ- and y-type distortions are equally restrictive. Overlaid are illustrative sharply peaked log-normal power spectra resulting in the formation of PBHs with $M_{\rm BH} = 10^{10}, 10^6, 10^2, 10^{-2}$, and $10^{-6} M_\odot$ and an initial abundance of $\beta = 10^{-20}$, assuming Gaussian statistics for ζ.

In Eqs. (5.30) and (5.31), we can identify the impact of the various eras described earlier in this section. In particular, for $M_{\rm BH} \ll 1.5\times 10^5\, M_\odot$ (corresponding to $k_{\rm BH} \gg 5400/\text{Mpc}$), the black holes are formed during the thermalization era, and both μ- and y-type spectral

distortions are suppressed. For $1.5 \times 10^5\, M_\odot \ll M_{\rm BH} \ll 4.5 \times 10^9\, M_\odot$ ($31.6/{\rm Mpc} \lesssim k_{\rm BH} \lesssim 5400/{\rm Mpc}$), the black holes are forming during the μ-era, leading primarily to μ-type spectral distortions. Larger black holes form later, yielding primarily y-type spectral distortions.

We can use Eqs. (5.30) and (5.31) to quickly estimate whether a scenario featuring primordial SMBHs is consistent with spectral distortion constraints. Recall that, for the case of Gaussian statistics, a value of $\sigma_\zeta^2 \gtrsim 10^{-2}$ is required in order to obtain a non-negligible abundance of PBHs. For such large values of σ_ζ^2, spectral distortions exclude all black holes masses $M_{\rm BH} \gtrsim$ few $\times\, 10^3 M_\odot$. This conclusion is consistent with Fig. 5.6, where we compare the bounds from COBE/FIRAS with the power spectra predicted for several values of $M_{\rm BH}$. Therefore, the existence of a non-negligible abundance of primordial SMBHs requires the presence of significant non-Gaussianities in the distribution of the primordial curvature perturbations.

5.2.3 Departures from Gaussianity

For Gaussian density perturbations, constraints from spectral distortions severely limit the abundance of primordial SMBHs that could have formed in the early universe. In this case, the variance must be small in order to limit spectral distortions, while a large variance is required to generate a non-negligible abundance of PBHs. This tension could be resolved, however, if the distribution of curvature perturbations features a heavier tail than that of a Gaussian. Such non-Gaussianities thus potentially allow primordial SMBHs to form[10] without necessarily violating spectral distortion constraints.[11]

10. As noted in [540], large curvature perturbations in models with heavy-tailed distributions are often type II. Such perturbations have been demonstrated to form PBHs [541], however collapse in the type II case — in particular the mass of the resultant PBH — is not completely understood.

11. Ref. [542] proposes an alternative way to obtain the extreme statistics needed to form primordial SMBHs compatible with spectral distortion bounds. Their multi-field model results in effectively two different inflationary histories for the casually disconnected Hubble patches, with a subdominant fraction experiencing more expansion. By the δN formalism this equates to large curvature perturbations, such that these patches collapse to form PBH. Since this occurs only in a tiny fraction of patches, there is no observable generation of spectral distortions.

Fortunately, the distribution of primordial density perturbations is generically predicted to be non-Gaussian. Firstly, there is intrinsic non-Gaussianity that arises from the non-linear mapping between the curvature perturbation ζ and the density contrast δ, as can be seen in Eq. (**??**). Thus, even if the probability distribution function for ζ were exactly Gaussian, the distribution in δ would not be. Secondly, and more crucially, large departures from Gaussianity are generically found in models which produce a local enhancement in the primordial power spectrum [146].

To quantify the degree of non-Gaussianity that would be required to generate primordial SMBHs without violating spectral distortion constraints, it is instructive to consider a class of probability distribution functions of the form [542]

$$P_\delta^{(n)} = \frac{1}{2\sqrt{2}\,\sigma_0\,\Gamma\left(1+\frac{1}{n}\right)} \exp\left[-\left(\frac{|\delta|}{\sqrt{2}\sigma_0}\right)^n\right] , \tag{5.32}$$

where n parameterizes the heaviness of the distribution's tail. The variance of the density contrast is set by the second moment of the distribution:

$$\sigma_\delta^2(\sigma_0) \equiv \int_{-\infty}^{\infty} d\delta\,\delta^2\,P_\delta^{(n)} = \frac{\sigma_0^2\,\Gamma\left(1+\frac{3}{n}\right)}{3\Gamma\left(1+\frac{1}{n}\right)} . \tag{5.33}$$

Note that for $n = 2$, this reduces to the Gaussian form of Eq. (**??**) with $\sigma_0^2 = \sigma_\delta^2$. For $n = 1$ the tail falls off exponentially, while for $0 < n < 1$ it falls off even more slowly; we will refer to this class of distributions with $n < 1$ as "heavy-tailed."[12]

In Fig. 5.7, we show the shape of this class of probability distributions for various choices of n. As expected, we see that smaller n gives rise to heavier tails. This raises the question of how small n must be in order to efficiently form primordial SMBHs while keeping the peak of the power spectrum within bounds of spectral distortions, $\mathcal{P}_\zeta \lesssim 10^{-4}$. In Fig. 5.8,

12. Formally, the probability distribution P_x of a random variable x is said to be "heavy" if its tail is not exponentially bounded, $\lim_{x\to\infty} e^{\lambda x}\bar{F}(x) = \infty\,\forall\,\lambda > 0$, where $\bar{F}(x) = \int_x^\infty dx'\,P_x(x')$.

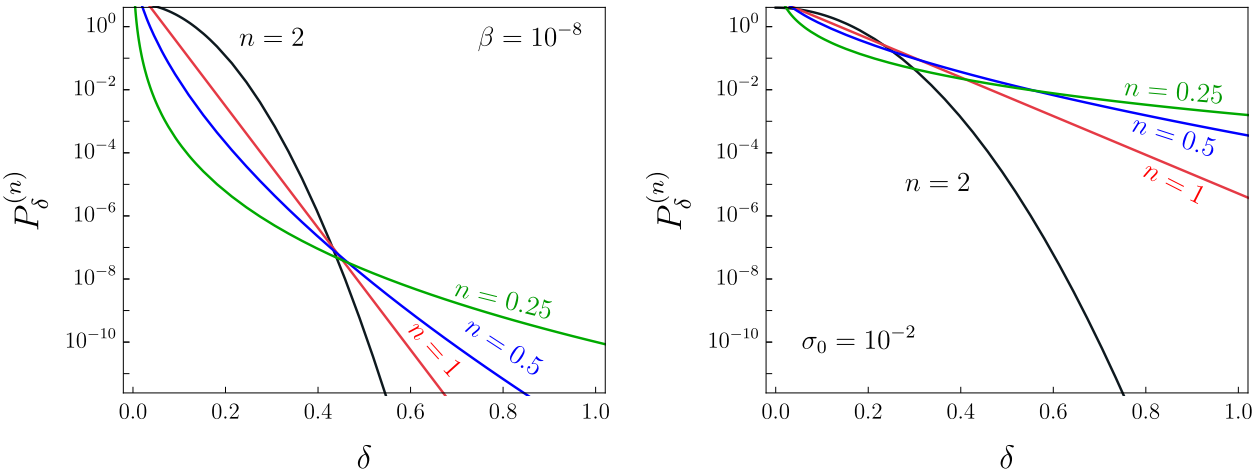

Figure 5.7: Probability distribution functions Eq. 5.32 for various choices of n at fixed β (left) and fixed variance σ^2 (right). We see that smaller n corresponds to heavier-tailed distributions.

we plot the maximum PBH mass fraction at formation $\beta_{\rm max}$, for a variance that saturates the spectral distortion constraints from COBE/FIRAS. Note that to generate a present day abundance of $\Omega_{\rm BH} \gtrsim 10^{-20}$ with $M_{\rm BH} \lesssim 10^{11} M_\odot$, we need a heavy tail with $n \lesssim 0.6$.

We now turn to the question of what inflationary models could yield such dramatic departures from Gaussianity. In the context of single-field inflation, Refs. [543, 544, 545, 546] perturbatively studied the local non-Gaussianity that arises in models which deviate from the slow-roll attractor, as in ultra-slow-roll inflation. Going beyond perturbation theory, Refs. [517, 512] used the δN formalism [547, 65, 548, 549, 550, 551] to compute the non-perturbative distribution of curvature perturbations.[13] For inflaton potentials with a small step or bump-like feature that induces a period of off-attractor behavior, these studies found that the tail of the distribution can become exponential, without inducing any significant non-Gaussianities in the perturbative regime [512]. This result highlights the fact that perturbative measures of non-Gaussianity are not generally adequate to describe the large, rare fluctuations that lead to PBH formation. Finally, many of the mechanisms for enhancing

13. Ref. [552] presents another non-perturbative method for calculating the PBH abundance in models where ζ is related to a Gaussian reference variable ζ_G by a "generalized local transformation".

local curvature perturbations rely on a temporary reduction in the inflaton's velocity. When the slow-roll classical drift vanishes, the field dynamics can receive large corrections from quantum diffusion, and the stochastic inflationary formalism [65, 553] may be necessary for a proper description of the dynamics. Combining this with the δN formalism [554], a number of studies [513, 515, 516, 555, 556, 557] have found that prominent exponential tails arise generically from quantum diffusion.

While many single-field models have been found to yield exponential tails [558], there is currently no known model which generates a heavier-tailed distribution.[14] However, as shown in Fig. 5.8, a heavy-tailed $P_\delta \sim \exp\left(-|\delta|^n\right)$ with $n \lesssim 0.6$ is needed to yield a non-negligible population of primordial SMBHs while satisfying bounds from CMB spectral distortions. While it is unclear whether primordial SMBHs can appreciably form in any viable single-field models, it is plausible that the necessary heavy tails can arise in certain multi-field scenarios. In particular, curvaton models have long been known to generate density perturbations with sizeable non-Gaussianities in the case that the curvaton remains subdominant at the time of its decay [560]. While the standard curvaton with quadratic potential is incapable of producing the dramatic non-Gaussianities needed for our scenario, there is good reason to believe that these can be augmented for a self-interacting curvaton model. We explore these possibilities in the following sections.

5.2.4 Standard Curvaton Scenario

Curvaton models introduce a second light, unstable spectator field that is present during inflation and that is responsible for generating the dominant contribution to the primordial curvature perturbations [561, 562, 563, 564]. The perturbations of the curvaton are initially isocurvature, but become adiabatic upon curvaton decay sometime after inflation ends [565].

14. Ref. [559] interprets the NANOGrav signal as evidence of PBH mergers with $M_{\rm BH} \sim 10^{11} - 10^{12} M_\odot$, and claims that μ-distortion constraints can be overcome for sufficiently non-Gaussian single-field models. However they make no reference to y-type distortions, which are more constraining for this mass range, and which we have verified rule out this single-field scenario for any non-negligible abundance.

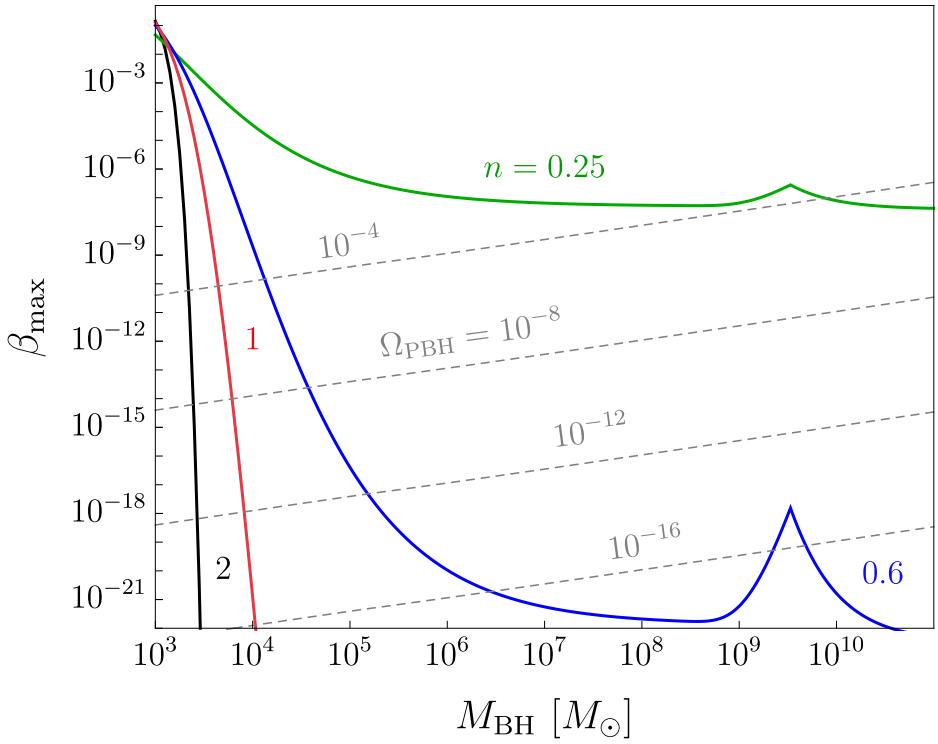

Figure 5.8: The maximum primordial black hole mass fraction at formation $\beta_{\rm max}$ as a function of mass $M_{\rm BH}$ for a value of the variance σ_ζ^2 that saturates the spectral distortion constraints from COBE/FIRAS, as estimated according to Eq. (5.30). We assume the distribution function given in Eq. (5.32) and consider Gaussian ($n = 2$, black), exponential ($n = 1$, red), and power law ($n = 0.6$ blue, $n = 0.25$ green) behavior in the tail. Note that the cusps which appear near $M_{\rm BH} \sim 3 \times 10^9\, M_\odot$ correspond to the value of $M_{\rm BH}$ at which μ- and y-type spectral distortions are equally restrictive. Contours of constant $\Omega_{\rm BH}$ are shown in dashed gray.

Due to the non-linearity inherent in this transfer, the full perturbation ζ can become quite non-Gaussian. In particular, when the curvaton is still very subdominant at decay, the inefficient conversion can yield a very heavy-tailed distribution for ζ.

Non-Gaussianity in the curvaton model was first investigated using the δN formalism in Ref. [560], with implications for PBH formation analyzed in Refs. [566, 524, 540]. As we will see, the standard curvaton model with only a quadratic potential cannot produce sufficient non-Gaussianity to generate a non-negligible abundance of SMBHs without violating spectral distortions constraints.

Curvaton Cosmology

We begin by reviewing the calculation of the curvature perturbation ζ and its statistics in the standard curvaton scenario [560], for which the total potential is

$$V(\phi, \chi) = V_\phi(\phi) + V_\chi(\chi)\,, \tag{5.34}$$

where V_ϕ is the unspecified potential of the inflaton ϕ, and $V_\chi = m_\chi^2 \chi^2/2$ is the quadratic potential of the curvaton χ. The curvaton mass is required to be light, satisfying $m_\chi \ll H$ throughout inflation, such that quantum perturbations are the dominant influence on its evolution. This also implies that the background field $\bar{\chi}$ will remain effectively fixed at some initial value $\bar{\chi}_*$ during inflation, where a star denotes the value of quantities at horizon exit. For the curvaton energy density to remain subdominant throughout inflation, we demand $\bar{\chi}_* \ll \sqrt{2V_\phi/m_\chi^2}$. Just like the inflaton, the curvaton receives perturbations $\delta\chi_* \simeq H_*/2\pi$ set by the Hubble rate at horizon exit H_*. Since the curvaton is a weakly coupled field, we expect the perturbations $\delta\chi_*$ to be described by a Gaussian random field. Thus we can write the curvaton at horizon exit as the sum of a background field and a linear perturbation, with no higher order terms:

$$\chi_* = \bar{\chi}_* + \delta\chi_*\,. \tag{5.35}$$

The goal of this section will be to relate these initial Gaussian field perturbations to the total curvature perturbation ζ via some mapping $\zeta = f(\delta\chi_*)$. This will be the key to constructing the probability distribution function for ζ, since the statistics of a non-Gaussian variable are completely determined by the statistics of a Gaussian reference variable when the mapping between them is specified.

When inflation ends and the inflaton decays, the universe enters into an era of radiation domination, with $\rho_R \sim a^{-4}$. At this point, the curvaton energy density is subdominant and its fluctuations are still isocurvature in nature. As the Hubble rate decreases, it eventually

drops below m_χ, causing the curvaton to start oscillating about the minimum of its potential. We denote the field value at which this occurs by $\bar{\chi}_0$. During this oscillating phase, the curvaton redshifts like matter with $\rho_\chi \propto a^{-3}$ and its energy density grows linearly relative to radiation. Finally, when $H \sim \tau^{-1}$, where τ is the χ lifetime, the curvaton decays to radiation and its isocurvature perturbations become adiabatic perturbations, assuming the decay products thermalize with the existing radiation.

The δN Formalism

To calculate the distribution of the curvature perturbations in this model, we employ the δN formalism [547, 65, 548, 549, 550, 551, 567], which we review in Appendix A. This technique identifies ζ on large scales ($k \ll aH$) with the variation of inflationary e-folds across Hubble patches and non-perturbatively captures its non-Gaussianities. The δN formalism was first used to study non-Gaussianity in curvaton models in Refs. [568, 560]. On a general hypersurface of uniform curvaton density, the conserved curvaton curvature perturbation ζ_χ is [568, 550]

$$\zeta_\chi(t, \vec{x}) = \delta N(t, \vec{x}) + \frac{1}{3} \ln \left(\frac{\rho_\chi(t, \vec{x})}{\bar{\rho}_\chi(t)} \right) , \tag{5.36}$$

where $\delta N(t, \vec{x})$ is the perturbed number of e-folds, $\rho_\chi(t, \vec{x})$ is the χ energy density, and $\bar{\rho}_\chi(t)$ is its background value. In spatially flat slicing, this becomes

$$\zeta_\chi(t, \vec{x}) = \frac{1}{3} \ln \left(\frac{\rho_\chi(t, \vec{x})}{\bar{\rho}_\chi(t)} \right) , \tag{5.37}$$

and the curvaton energy density can be written as

$$\rho_\chi(t, \vec{x}) = e^{3\zeta_\chi(t, \vec{x})} \, \bar{\rho}_\chi(t) \, . \tag{5.38}$$

In uniform total density slicing, Eq. (5.36) becomes

$$\zeta_\chi(t,\vec{x}) = \zeta + \frac{1}{3}\ln\left(\frac{\rho_\chi^{(u)}(t,\vec{x})}{\bar{\rho}_\chi^{(u)}(t)}\right) , \tag{5.39}$$

where ζ is the total curvature perturbation.[15] Since ζ and ζ_χ are gauge invariant quantities, Eq. (5.38) can be equated with Eq. (5.39) to yield

$$\rho_\chi^{(u)}(t,\vec{x}) = e^{3(\zeta_\chi-\zeta)}\bar{\rho}_\chi^{(u)}(t) . \tag{5.40}$$

A similar treatment for the radiation energy density in uniform total density slicing gives

$$\rho_R^{(u)}(t,\vec{x}) = e^{4(\zeta_R-\zeta)}\bar{\rho}_R^{(u)}(t) \simeq e^{-4\zeta}\bar{\rho}_R^{(u)}(t) , \tag{5.41}$$

where we have assumed for simplicity that the main contribution to the curvature perturbation comes from the curvaton. In order to derive analytic results, we work in the instantaneous decay approximation such that the curvaton decays when $H = \tau^{-1}$, where τ is the curvaton lifetime. On a uniform total density slice at $t = \tau$, the energy densities satisfy

$$\rho_R^{(u)}(\tau,\vec{x}) + \rho_\chi^{(u)}(\tau,\vec{x}) = \bar{\rho}^{(u)}(\tau) , \tag{5.42}$$

where $\bar{\rho}^{(u)} = \bar{\rho}_R^{(u)} + \bar{\rho}_\chi^{(u)}$ is the total homogeneous energy density. Substituting Eqs. (5.40) and (5.41), this becomes a 4^{th} degree algebraic equation for ζ at τ:

$$e^{4\zeta} - \left(e^{3\zeta_\chi}\,\Omega_{\chi,\tau}\right)e^{\zeta} + \left(\Omega_{\chi,\tau} - 1\right) = 0 . \tag{5.43}$$

15. Note that this generically has non-vanishing mean, $\langle\zeta\rangle \neq 0$, and so when we later consider PBH formation, we will have to define a physical $\zeta_{\rm phys} \equiv \zeta - \langle\zeta\rangle$ [524]. The expectation value $\langle\zeta\rangle$ can be computed using the Gaussian $P_{\delta\chi}$ as $\langle\zeta\rangle = \int d\delta_\chi\, \zeta\, P_{\delta_\chi}$, with ζ expressed as a function of δ_χ given in Eq. (5.46).

Alternatively, it is customary to introduce the parameter r_τ, defined as [561]:

$$r_\tau = \frac{3\Omega_{\chi,\tau}}{4-\Omega_{\chi,\tau}} = \left.\frac{3\bar{\rho}_\chi^{(u)}}{3\bar{\rho}_\chi^{(u)}+4\bar{\rho}_R^{(u)}}\right|_\tau , \tag{5.44}$$

in terms of which the equation for ζ becomes:

$$e^{4\zeta} - \frac{4r_\tau}{3+r_\tau}\left(e^{3\zeta_\chi}\right)e^{\zeta} + \frac{3r_\tau-3}{3+r_\tau} = 0\,. \tag{5.45}$$

The general solution is

$$\zeta = \ln X\,,\;\; X = \frac{B^{1/2}+\sqrt{Ar_\tau B^{-1/2}-B}}{(3+r_\tau)^{1/3}}\,, \tag{5.46}$$

where $A = e^{3\zeta_\chi}$ and we have defined

$$B = \frac{1}{2}\left[C^{1/3} + (r_\tau-1)(3+r_\tau)^{1/3}C^{-1/3}\right]\,, \tag{5.47a}$$

$$C = (Ar_\tau)^2 + \sqrt{(Ar_\tau)^4 + (3+r_\tau)(1-r_\tau)^3}\,. \tag{5.47b}$$

This gives the mapping $\zeta = \ln[X(\zeta_\chi)]$ between ζ and ζ_χ.

Calculating the Probability Distribution

Finally, to obtain ζ_χ in terms of $\delta\chi_*$, we return to Eq. (5.37), which gave the curvature perturbation in spatially flat slicing as a function of perturbed and background energy densities. For the simple quadratic potential of this model, we have $\rho_\chi(t,\vec{x}) = m_\chi^2\chi^2/2$. Expanding $\chi(t,\vec{x}) = \bar{\chi}(t) + \delta\chi(t,\vec{x})$, we can write this as

$$\rho_\chi(t,\vec{x}) = \frac{1}{2}m_\chi^2\bar{\chi}^2\left(1+\frac{\delta\chi}{\bar{\chi}}\right)^2 = \bar{\rho}_\chi(t)(1+\delta_\chi)^2\,, \tag{5.48}$$

where $\delta_\chi = \delta\chi/\bar{\chi}$ is the curvaton contrast in spatially flat slicing. Comparing with $\rho_\chi = e^{3\zeta_\chi}\bar{\rho}_\chi$, we see we should identify $e^{3\zeta_\chi} = (1+\delta_\chi)^2$. Finally in order to relate the contrast δ_χ to its value at horizon exit, consider the equations of motion for χ and the perturbation $\delta\chi$:

$$\frac{d^2\bar{\chi}}{dt^2} + 3H\frac{d\bar{\chi}}{dt} + m_\chi^2\bar{\chi} = 0\,, \tag{5.49a}$$

$$\frac{d^2}{dt^2}(\delta\chi) + 3H\frac{d}{dt}(\delta\chi) + \left(\frac{k^2}{a^2} + m_\chi^2\right)\delta\chi = 0\,. \tag{5.49b}$$

On superhorizon scales $k \ll aH$ is negligible and so these reduce to the same equation. This implies $\delta\chi \sim \bar{\chi}$, such that $\delta\chi/\bar{\chi} = \delta\chi_*/\bar{\chi}_*$ and

$$A = e^{3\zeta_\chi} = (1+\delta_\chi)^2 = \left(1 + \frac{\delta\chi_*}{\bar{\chi}_*}\right)^2\,. \tag{5.50}$$

Combining with Eq. (5.46), we obtain the desired mapping between ζ and the Gaussian initial field perturbations $\delta\chi_*$. We can now use probability conservation to write

$$P_\zeta[\zeta] = P_{\delta_\chi}\left[\delta_\chi^+(\zeta)\right]\left|\frac{d\delta_\chi^+}{d\zeta}\right| + P_{\delta_\chi}\left[\delta_\chi^-(\zeta)\right]\left|\frac{d\delta_\chi^-}{d\zeta}\right|, \tag{5.51}$$

where P_{δ_χ} is fully determined by the Gaussian variance σ_0^2 and the roots $\delta_\chi^\pm(\zeta)$ satisfy

$$\delta_\chi^\pm = -1 \pm \sqrt{\left(\frac{3+r_\tau}{4r_\tau}\right)e^{3\zeta} + \left(\frac{3r_\tau - 3}{4r_\tau}\right)e^{-\zeta}}\,, \tag{5.52}$$

which arise from solving Eq. (5.45) and substituting Eq. (5.50). In Fig. (5.9), we plot the probability distribution for the curvature perturbation as given by Eq. (5.51) for a few choices of r_τ, which controls the heaviness of the tail.

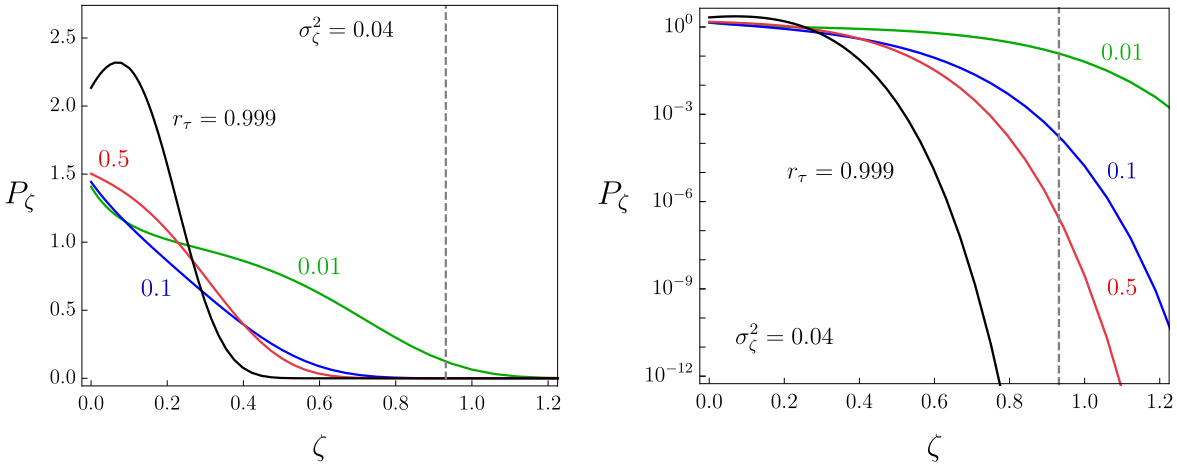

Figure 5.9: The probability distribution function for the curvature perturbation ζ as given by Eq. (5.51) for various choices of r_τ. The vertical dashed line corresponds to the threshold value for collapse ζ_c, and the PBH mass fraction β is obtained by integrating beyond this threshold. We see that a smaller r_τ corresponds to a heavier-tailed distribution, leading to a larger PBH abundance. Note that the reference value $\sigma_\zeta^2 = 0.04$ is chosen to illustrate the heaviness of the tail as r_τ is varied, but the minimal scenario presented in Sec. 5.2.4 cannot realize such a large value without violating the observational limits on the power spectrum shown in Fig. 5.6.

5.2.5 Heavy Tails and Primordial Black Holes

Viably forming an appreciable number of primordial SMBHs requires both amplified power on small scales and a departure from Gaussianity. Thus we require the following two additional ingredients:

- **Enhanced Power:** One mechanism for enhancing the power spectrum is to introduce a non-canonical kinetic term for the curvaton, which depends on the inflaton's field value [524]. In Sec. 5.2.5, we review this scenario and calculate the power spectrum resulting from such a kinetic term.

- **Large Non-Gaussianity:** The non-Gaussianity in a curvaton model can be amplified through self-interactions, which lead to non-linear growth of χ perturbations between horizon exit and the onset of curvaton oscillations [569, 570, 571, 572, 573, 574, 575, 576, 577]. In Sec. 5.2.5, we review the evolution of the non-quadratic curvaton and

calculate the resulting probability distribution function.

Non-Minimal Curvaton Scenario

The minimal curvaton model provides non-Gaussian statistics, but does not amplify $\mathcal{P}_\zeta$. This deficiency can be remedied with a non-canonical kinetic term[16] for the curvaton [524]:

$$\mathcal{L} \supset \frac{1}{2} f(\phi)^2 (\partial\chi)^2 \,. \tag{5.53}$$

If $f(\phi)$ is chosen such that this kinetic term is suppressed at field values $\phi = \phi_*$, the power spectrum will be enhanced on scales corresponding to the horizon size at ϕ_*. For concreteness, consider the evolution of the inflaton and curvaton governed by the following system of equations

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = f f' \dot{\bar{\chi}}^2 \simeq 0 \,, \tag{5.54a}$$

$$\ddot{\bar{\chi}} + \left(3H + \frac{2f'}{f}\dot{\phi} \right) \dot{\bar{\chi}} + \frac{m_\chi^2}{f^2} \bar{\chi} = 0 \,, \tag{5.54b}$$

where $f' \equiv \partial_\phi f$ and the source term of the first equation is negligible until the curvaton begins to oscillate. Similarly, the curvaton perturbation evolves according to

$$\frac{d^2}{dt^2}(\delta\chi) + \left(3H + \frac{2f'}{f}\dot{\phi} \right) \frac{d}{dt}(\delta\chi) + \left(\frac{k^2}{a^2} + \frac{m_\chi^2}{f^2} \right) \delta\chi \simeq 0 \,, \tag{5.55}$$

which, to leading order, can be simplified inside the horizon, $k \gg aH$, to yield

$$\frac{d^2}{dt^2}\left(f\delta\chi\right) + \frac{k^2}{a^2}(f\delta\chi) \simeq 0 \,, \tag{5.56}$$

16. Such a term naturally arises in many dilatonic and axionic models of inflation [578].

whose solution can be written as

$$\delta\chi \simeq \frac{1}{\sqrt{2k}\,af} \exp\left(-ik\int \frac{dt}{a}\right), \tag{5.57}$$

which establishes the initial conditions in the adiabatic vacuum. As in Sec. **??**, on super-horizon scales, $k \ll aH$, $\bar{\chi}$ and $\delta\chi$ evolve according to the same equation, so their solutions have the same functional form for all $t > t_*$:

$$\frac{\delta\chi}{\bar{\chi}} = \frac{\delta\chi_*}{\bar{\chi}_*} \simeq \frac{H_*}{\sqrt{2k^3} f(\phi_*)\bar{\chi}_*}, \tag{5.58}$$

where $k = a(t_*)H(t_\star)$. Using Eq. (5.58), the power spectrum for δ_χ becomes

$$\mathcal{P}_{\delta_\chi}(k) = \frac{k^3}{2\pi^2}\left|\frac{\delta\chi_k}{\bar{\chi}}\right|^2 = \frac{1}{\bar{\chi}_*^2}\left(\frac{H_*}{2\pi f(\phi_*)}\right)^2, \tag{5.59}$$

so, if $f(\phi)$ is chosen to have a dip at ϕ_*, corresponding to $k_* = k_{\rm BH}$, $\mathcal{P}_{\delta_\chi}$ will exhibit a peak at $k_{\rm BH}$. However, for modes far away from $k_{\rm BH}$, $f(\phi_*) \approx 1$, recovering the nearly scale-invariant spectrum required for consistency with CMB observations on larger scales. Assuming $f(\phi)$ has such a localized feature, combining Eqs. (5.51), (5.59), and (**??**), the PBH abundance becomes

$$\begin{aligned}\beta &= 2\int_{\delta^+_{\chi,c}}^{\infty} d\delta_\chi\, P_{\delta_\chi}[\delta_\chi] + 2\int_{-\infty}^{\delta^-_{\chi,c}} d\delta_\chi\, P_{\delta_\chi}[\delta_\chi] \\ &= \mathrm{erfc}\left(\frac{\delta^+_{\chi,c}}{\sqrt{2}\sigma_0}\right) + \mathrm{erfc}\left(\frac{|\delta^-_{\chi,c}|}{\sqrt{2}\sigma_0}\right),\end{aligned} \tag{5.60}$$

where $\delta^\pm_{\chi,c} = \delta^\pm_\chi(\zeta_c)$ are the roots of Eq. (5.52) evaluated at the threshold.

Using Eq. (5.30), we obtain the maximum value of β consistent with spectral distortion constraints. Note that the variance σ^2_ζ corresponds to the physical curvature perturbation, $\sigma^2_\zeta = \langle\zeta^2_{\rm phys}\rangle = \langle\zeta^2\rangle - \langle\zeta\rangle^2$, which can be computed from $\zeta(\delta_\chi)$ in Eq. (5.46) as $\sigma^2_\zeta =$

$\int d\delta_\chi\, \zeta^2\, P_{\delta_\chi} - \left(\int d\delta_\chi\, \zeta\, P_{\delta_\chi}\right)^2$. The degree of non-Gaussianity in this scenario is governed by

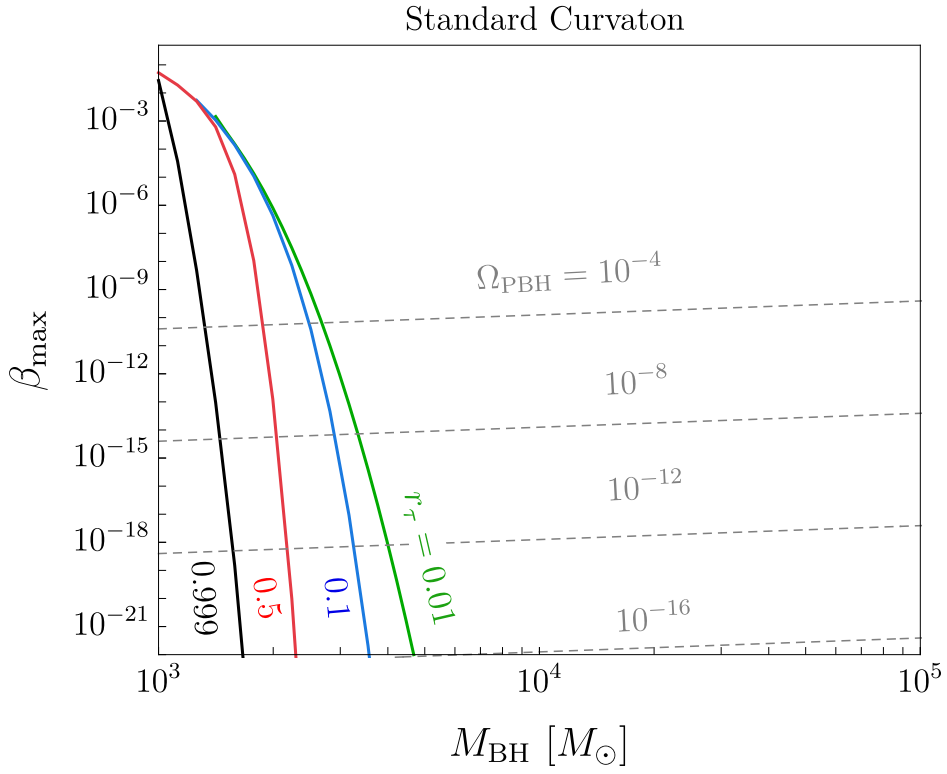

Figure 5.10: The maximal PBH mass fraction at formation $\beta_{\rm max}$ in the standard curvaton scenario for σ_ζ^2 subject to spectral distortion constraints for various choices of r_τ from Eq. (5.44).

r_τ, as defined in Eq. (5.44). In the $r_\tau \to 1$ limit, the curvaton dominates the energy density prior to its decay, so the relation between ζ and ζ_χ from Eq. (5.45) is approximately linear. Thus, in this regime, the non-Gaussianity comes entirely from the non-linear relationship between ζ_χ and δ_χ; see Eq. (5.50). In the opposite regime[17] that the curvaton is still very subdominant when it decays ($r_\tau \ll 1$), the relation between ζ and ζ_χ is highly non-linear, and so the degree of non-Gaussianity is large. This is reflected in Figs. 5.9 and 5.10. Clearly this scenario is incapable of producing primordial SMBHs while satisfying spectral distortion

17. It may seem counterintuitive that a very subdominant curvaton can still generate the curvature perturbation. The key is that because the curvaton is subdominant during inflation, its perturbations from quantum fluctuations are large relative to the background field value. Ref. [579] finds that a subdominant curvaton can still viably produce the curvature perturbation provided $r_\tau \gtrsim 10^{-3}$.

constraints.

Self-Interacting Curvaton Scenario

Introducing curvaton self-interactions allows for non-linear evolution of δ_χ between horizon exit and the onset of oscillations, which can lead to even more dramatic departures from Gaussianity. This scenario is also physically well-motivated since the curvaton needs to decay for the isocurvature perturbations to be converted to adiabatic perturbations. The effects of self-interactions in curvaton models were first investigated in Refs. [569, 570, 571, 572, 573, 574, 575, 576, 577]. However, these studies restricted themselves to computing the non-linearity parameters $f_{\rm NL}$ and $g_{\rm NL}$, which do not capture the non-perturbative statistics in the tail of the distribution. Unfortunately following the curvaton's evolution in the non-quadratic regime and computing the exact resulting curvature distribution require extensive numerics beyond the scope of this study. We offer here a schematic picture, but leave the detailed model building to future studies.

We now allow the curvaton potential $V(\chi)$ to be an arbitrary well-defined function of χ. The cosmological evolution of the curvaton proceeds similarly to the case of the quadratic potential, but with a few crucial differences. At a time $t = t_{\rm int}$ corresponding to $V'(\chi_*) \sim H$, χ begins rolling towards the minimum of its potential. We choose parameters such that this occurs while the curvaton's energy density is dominated by the interaction terms. At a later time $t = t_0 > t_{\rm int}$, these interaction terms become subdominant and the curvaton mass term drives field evolution, resulting in matter-like scaling $\rho_\chi \propto a^{-3}$. Note that with self-interactions, the curvaton energy density generically falls off faster than in the quadratic case, resulting in a smaller r_τ at the time of decay.

Recall that in the case of the quadratic potential, the curvaton density contrast δ_χ remained constant after horizon exit since the background field $\bar{\chi}$ and perturbation $\delta\chi$ obeyed the same equation of motion, shown in Eq. (5.49). This led to $\delta\chi/\bar{\chi} = \delta\chi_*/\bar{\chi}_*$, which al-

lowed δ_χ to be used as a Gaussian reference variable. Upon introducing interactions this is no longer the case, as δ_χ evolves non-trivially between $t_{\rm int}$ and the onset of quadratic oscillations at t_0. In this regime, the equation of motion for the perturbation is:

$$\frac{d^2}{dt^2}(\delta\chi) + 3H\frac{d}{dt}(\delta\chi) + V''_\chi \delta\chi = 0\,, \tag{5.61}$$

where it is understood that the second derivative of the potential should be evaluated at the background field value.

A natural choice of Gaussian reference variable is the initial curvaton perturbation $\delta\chi_*$, which can be related to the ζ_χ by first solving Eq. (5.61) along with the equation for the background field $\bar{\chi}$, computing the total and background energy densities at decay, and finally applying Eq. (5.37). The resulting $\delta\chi^i_* = g_i(\zeta_\chi)$ can then be mapped onto the total curvature perturbation ζ via Eq. (5.45), since the relationship between ζ_χ and ζ is unchanged in the presence of χ self-interactions. Although an exact solution requires the use of numerical techniques, an approximate relation can be derived in the limit of weak interactions.

We are interested in the relation between $\delta\chi_*$ and ζ_χ at the time of curvaton decay. For χ values sufficiently close to the minimum of its potential, the potential is approximately quadratic, and the energy density is

$$\rho_\chi \simeq \frac{1}{2}m_\chi^2\chi_0^2\,, \tag{5.62}$$

where χ_0 is the amplitude at the onset of oscillations. Since non-linear evolution takes place between horizon exit and oscillation, this initial amplitude is a function of initial field value χ_*). In terms of background field values $\bar{\chi}_0 \equiv \chi_0(\bar{\chi}_*)$, this can be expanded as:

$$\chi_0 = \bar{\chi}_0 + \sum_{n=1}\frac{1}{n!}\bar{\chi}_0^{(n)}\delta\chi_*^n \quad , \quad \bar{\chi}_0^{(n)} \equiv \frac{\partial^n\chi_0}{\partial\chi_*^n}\Big|_{\chi=\bar{\chi}_*}\,, \tag{5.63}$$

and the energy density can then be written as:

$$\rho_\chi \simeq \bar{\rho}_\chi \left[1 + \frac{1}{\bar{\chi}_0} \sum_{n=1} \frac{1}{n!} \bar{\chi}_0^{(n)} \delta\chi_*^n\right]^2 , \tag{5.64}$$

where $\bar{\rho}_\chi = \frac{1}{2} m_\chi^2 \bar{\chi}_0^2$. Comparing with Eq. (5.37), the bracketed quantity is identified with $e^{3\zeta_\chi}$. Then also expanding

$$\zeta_\chi = \zeta_{\chi,1} + \sum_{n=2}^{\infty} \frac{1}{n!} \zeta_{\chi,n} \,, \tag{5.65}$$

we can write to leading order:

$$\begin{aligned} \zeta_\chi(\delta\chi_*) = \frac{2}{3}\left(\frac{\bar{\chi}_0'}{\bar{\chi}_0}\right)\delta\chi_* + \frac{1}{3}\left(\frac{\bar{\chi}_0\bar{\chi}_0''}{\bar{\chi}_0'^2} - 1\right)\left(\frac{\bar{\chi}_0'}{\bar{\chi}_0}\right)^2 \delta\chi_*^2 \\ + \frac{1}{9}\left(\frac{\bar{\chi}_0^2\bar{\chi}_0'''}{\bar{\chi}_0'^3} - 3\frac{\bar{\chi}_0\bar{\chi}_0''}{\bar{\chi}_0'^2} + 2\right)\left(\frac{\bar{\chi}_0'}{\bar{\chi}_0}\right)^3 \delta\chi_*^3 \,, \end{aligned} \tag{5.66}$$

which can then be substituted into Eq. (5.46) to obtain ζ as a function of the Gaussian reference variable $\delta\chi_*$. Note that though we have written this expression to fixed order, higher order terms can become significant in the presence of large non-Gaussianities. It is also possible to invert this mapping to obtain the roots $g_i(\zeta) = \delta\chi_*^i(\zeta)$. The probability distribution is then:

$$P_\zeta[\zeta] = \sum_j \left|\frac{g_j(\zeta)}{d\zeta}\right| P_{\delta\chi_*}[g_j(\zeta)] \,, \tag{5.67}$$

where the sum runs over all real roots.

In Fig. 5.11, we plot the maximal PBH abundance consistent with spectral distortion constraints for a sample set of parameters. Comparison with $\beta_{\max}$ in the standard curvaton scenario (see Fig. 5.10) reveals the dramatic amplification of non-Gaussianity that self-interactions can afford. We leave the determination of a potential that could actually yield these parameter values to future investigation.

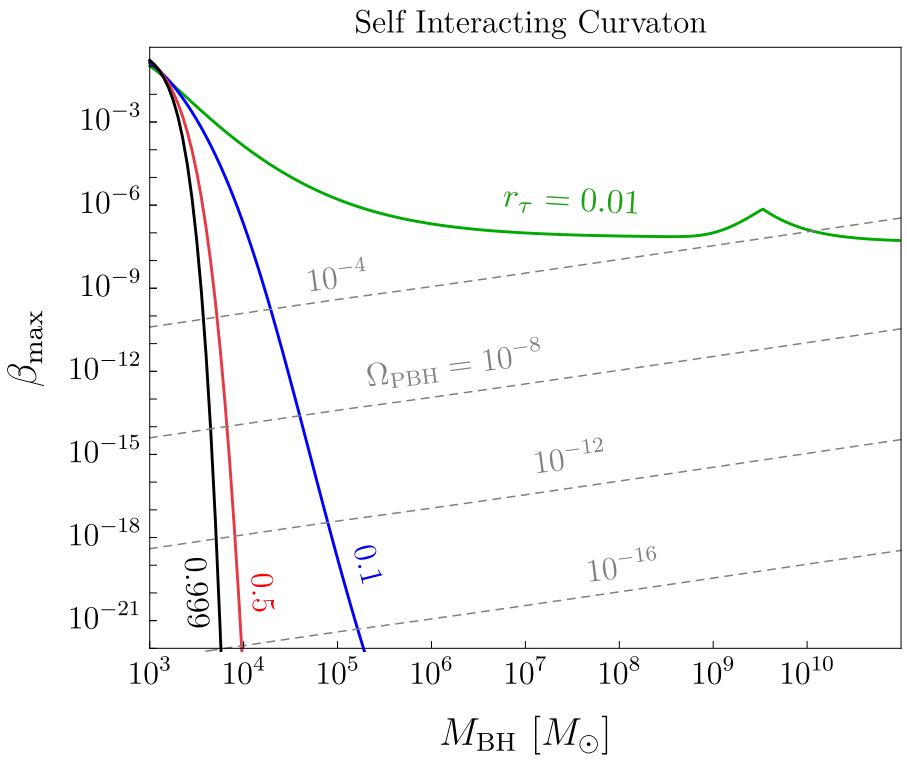

Figure 5.11: Maximal PBH mass fraction at formation $\beta_{\rm max}$ as a function of PBH mass in the self-interacting curvaton model from Sec. 5.2.5, where σ_ζ^2 saturates spectral distortion bounds for different values of r_τ, as defined in Eq. (5.44). We take $\zeta_\chi = c_1\,\delta\chi_* + c_2\,\delta\chi_*^2 + c_3\,\delta\chi_*^3$ to take the form of Eq. (5.66), with sample parameters fixed as $(c_1, c_2, c_3) = (1, 0.5, 0.1)$. Contours of constant $\Omega_{\rm BH}$ today are shown in dashed gray.

5.2.6 Discussion

Much remains to be understood about the origin and evolution of our universe's most massive black holes. The inferred population of supermassive black holes with $M_{\rm BH} \sim 10^6 - 10^{11} M_\odot$ in the high-redshift universe is perhaps surprising, and challenges the standard assumption that these objects formed from low mass seeds which grew through the processes of accretion and mergers. In this study, we have taken this as motivation to consider the possibility that some of our universe's supermassive black holes may be primordial in origin, having formed from the direct collapse of overdensities seeded by inflation.

Forming primordial SMBH from direct collapse requires an enhanced power spectrum on small scales ($k \sim 10 - 10^4\,{\rm Mpc}^{-1}$), which results in dangerous CMB spectral distortions.

Since the CMB exhibits a nearly perfect blackbody spectrum, such distortions exclude the possibility that a population of primordial SMBH could have formed from Gaussian density perturbations. However, if the distribution of primordial curvature perturbations were highly non-Gaussian, it is possible that primordial black holes may have formed from smaller peaks in the power spectrum. To evade limits from spectral distortions, the tail of the probability distribution must be very heavy, falling off as $P_\zeta \sim \exp\left(-|\zeta|^n\right)$ with $n \lesssim 0.6$; we are not aware of any single-field inflationary model that can realize this behavior.

In this section, we have explored the possibility of generating curvature perturbations with heavy-tailed probability distributions in the event that a self-interacting curvaton field is also present as a spectator during inflation. In the standard curvaton scenario, non-Gaussianity arises from the inefficient conversion of isocurvature perturbations into adiabatic perturbations when the curvaton decays. However, the degree of non-Gaussianity in this minimal realization is insufficient to yield an appreciable primordial SMBH population. Introducing curvaton self-interactions results in non-linear evolution of the curvaton contrast δ_χ between horizon exit and the onset of quadratic oscillations, potentially resulting in a heavier-tailed distribution, as seen in Fig. 5.11. We leave the detailed model building and determination of a potential that can realize this benchmark point to future investigation. Provided such a potential can be found, the non-minimal self-interacting curvaton could viably yield SMBHs from the direct collapse of primordial perturbations, without violating spectral distortion constraints.

We note that beyond the bounds from CMB spectral distortions, there are two further more speculative constraints that potentially need to be addressed in a full model. First, as pointed out by Ref. [580], the small scale curvature perturbation impacts structure evolution as traced by dark matter halos. Presuming a bump in the small scale power spectrum, they investigated host halo and subhalo evolution. By comparing the results of their simulations with the observed number of dwarf spheroidal galaxies and stellar streams, they

were able to place bounds on $\mathcal{P}_\zeta$ to greater sensitivity than COBE/FIRAS for a region of wavenumbers. The result would exclude some, but not all, of the SMBH mass range we consider. These bounds are less robust than those coming from CMB spectral distortion and are subject to modeling uncertainties, but nevertheless should be noted. Second, the large non-Gaussianities described in this model would likely lead to heavy clustering of quasars. Ref. [581] calculated the observed angular correlation function using the 92 quasars at $z \sim 6$ reported by the Subaru High-z Exploration of Low-Luminosity Quasars project and compared this with that predicted in a PBH model which used a light spectator field to source non-Gaussian primordial perturbations. They found the amplitude of the angular correlation function predicted by the model to be much larger than that observed, a result which would seem to preclude PBHs as the sole progenitors of SMBHs in this redshift range. More concretely, they were able to restrict the fraction of PBH originating SMBHs to $\lesssim 10^{-4}$.

These potential constraints notwithstanding, this work is especially timely in light of two distinct recent observations. The James Webb Space Telescope (JWST) has reported a number of surprisingly luminous high-redshift galaxy candidates [130, 131, 132, 133] whose existence poses a challenge to the standard ΛCDM paradigm. These massive early galaxies could conceivably be explained if primordial black holes accelerated galaxy formation in the early universe [582]. Meanwhile, the NANOGrav collaboration and several other pulsar timing array experiments have just announced evidence of a signal consistent with the stochastic gravitational wave background in the nHz frequency range [583, 123, 124, 125]. The leading astrophysical interpretation of this signal is that it consists of gravitational waves from supermassive black hole binary mergers. However, some aspects of this data, such as the frequency scaling of the spectral density parameter, are not particularly well-fit by this interpretation [584, 585]. Given that the distribution of SMBH binaries would be different if these objects were of a primordial origin, one avenue for future investigation would be to compute the gravitational wave signal predicted in this scenario. While for homogenously distributed

primordial SMBHs this possibility is in conflict with constraints on total abundance from large scale structure and the UV galaxy luminosity function [586], a highly clustered population could still viably source the gravitational wave signal [587]. Given that a clustered population can arise in the presence of non-Gaussianities, it would be interesting to see what our model predicts in this context; see Refs. [588, 581] for closely related work. There are also other signals, such as scalar-induced secondary gravitational waves, which could offer complimentary evidence of this scenario and deserve further study; see Refs. [589, 590] for a potential connection between scalar-induced gravitational waves and the PTA signal. Regardless, these recent observations provide us with strong motivation to better understand the cosmic origin of our universe's supermassive black holes.

5.2.7 Supplementary Material: The δN Formalism

The δN formalism [547, 65, 548, 549, 550, 551, 567] is a technique for computing the nonlinear curvature perturbation ζ on large scales by identifying it with the perturbed logarithmic expansion from some initial state to a final state of fixed energy density. In a homogenous background, the number of e-folds elapsed between two moments of times t_1 and t_2 is

$$\bar{N}(t_2, t_1) = \int_{t_1}^{t_2} dt\, H\,. \tag{5.68}$$

Meanwhile, the amount of expansion in a perturbed universe is [550]

$$N(t_2, t_1; \vec{x}) = \int_{t_1}^{t_2} dt\, (H + \dot{\psi})\,, \tag{5.69}$$

where ψ is the curvature perturbation appearing in the decomposition of the spatial 3-metric $g_{ij} = a^2(t)e^{2\psi(t,\vec{x})}\delta_{ij}$.[18] Note that this expression holds on superhorizon scales, where spatial gradients can be neglected. We define $\delta N \equiv N(t_2, t_1; x) - \bar{N}(t_2, t_1)$ to be the difference

18. We ignore tensor perturbations.

between the perturbed and unperturbed expansion, and see that it equates to the change in the curvature perturbation from an initial hypersurface at t_1 to a final hypersurface at t_2

$$\delta N \equiv N(t_2, t_1; \vec{x}) - \bar{N}(t_2, t_1) = \psi(t_2, \vec{x}) - \psi(t_1, \vec{x})\,. \tag{5.70}$$

Equivalently, the change in ψ going from one choice of slicing to another is the difference between the actual number of e-folds N and the homogenous background value $\bar{N}$.

Since the curvature perturbation ψ is a gauge-dependent quantity, whose value depends on the choice of slicing, it is convenient to combine this with the gauge-dependent density perturbation $\delta\rho$ to form the gauge-invariant curvature perturbation [591]:

$$\zeta(t, \vec{x}) = \psi(t, \vec{x}) + \frac{1}{3}\int_{\bar{\rho}(t)}^{\rho(t,\vec{x})} \frac{d\rho}{(1+w)\rho}\,, \tag{5.71}$$

where w is the equation of state of the cosmological fluid, $\rho(t, \vec{x})$ is the inhomogenous local energy density, and $\bar{\rho}(t)$ is the homogenous background energy density. It was demonstrated in Ref. [550] that this quantity is conserved on superhorizon scales.

We would like to equate this gauge-invariant, conserved curvature perturbation ζ with the perturbed logarithmic expansion δN. By choosing the initial hypersurface at t_1 to be spatially flat, such that $\psi(t_1, \vec{x}) = 0$, from Eq. (5.70) we can equate $\delta N(t_2, t_1; \vec{x} = \psi(t_2, \vec{x})$. By choosing the finial hypersurface at t_2 to be uniform density, such that $\rho(t_2, \vec{x}) = \bar{\rho}(t_2)$, from Eq. (5.71) we have $\zeta(t_2, \vec{x}) = \psi(t_2, \vec{x})$. Combining these expressions gives the δN formula

$$\zeta(t_2, \vec{x}) = \delta N(t_2, t_1; \vec{x})\,, \tag{5.72}$$

relating the curvature perturbation with the perturbed expansion between a spatially flat and uniform energy density hypersurface.

It is also possible to implement the δN formalism numerically. Following inflaton decay,

the exact system of equations describing the evolution of the curvaton and radiation bath is

$$\ddot{\chi} + \left(3H + \frac{1}{\tau}\right)\dot{\chi} + V'_\chi = 0 \quad , \quad \dot{\rho}_R + 4H\rho_R = \frac{\dot{\chi}^2}{\tau} \quad , \quad H^2 = \frac{8\pi}{3M_{\rm Pl}^2}(\rho_R + \rho_\chi)\,. \tag{5.73}$$

We set the initial conditions at the end of inflation by specifying $\chi_i = \chi_*$, $\dot{\chi}_i = 0$, and $H_i = H_*$, which in turn determines $\rho_{R,i} = \frac{3M_{\rm Pl}^2}{8\pi}H_i^2 - V_\chi(\chi_i)$. The system of equations should be evolved until a final time t_f satisfying $H_f \ll \tau^{-1}$, such that the curvaton has decayed completely. The number of e-folds elapsed is then computed as $N = \ln\big(a_f/a_i\big)$. This procedure should then be repeated for the perturbed field value $\chi_* + \delta\chi_*$, with the fluctuation determined by the size of Hubble at horizon exit $\delta\chi_* = H_*/2\pi$. Evolving until the same final hypersurface of fixed energy density, the curvature perturbation is computed as:

$$\zeta = N(\chi_* + \delta\chi_*) - N(\chi_*)\,. \tag{5.74}$$

Repeating for many different $\delta\chi_*$ gives a functional relation between ζ and the Gaussian $\delta\chi_*$. We leave an in-depth numerical study to future investigation.

5.3 Cosmological Magnetic Fields

The origin of our universe's cosmological magnetic fields remains a mystery. In this section, we consider whether these magnetic fields could have been generated in the early universe by a population of charged, spinning primordial black holes. To this end, we calculate the strength and correlation length of the magnetic fields generated by this population, and describe their evolution up to the current epoch. We find that near-extremal black holes in the mass range $M \sim 10^{28} - 10^{36}\,\mathrm{g}$ could potentially generate magnetic fields with present day values as large as $B \sim 10^{-20} - 10^{-15}\,\mathrm{G}$; those with $M \gtrsim 10^{38}\,\mathrm{g}$ could have produced even larger fields $B \gtrsim 10^{-14}\,\mathrm{G}$. To motivate this scenario, we briefly discuss how new physics may have induced a chemical potential which could have briefly maintained the

black holes in an electrically charged state in the early universe. Finally, we comment on a correlation between the parameters of the cosmological magnetic field and the stochastic gravitational wave background coming from the merger of primordial black hole binaries as the primary observable signature of this scenario. Note that much of the text in this section was previously published by this author in [29].

5.3.1 Overview

According to the standard paradigm[19], the magnetic fields present within galaxies and galaxy clusters were generated through the amplification of preexisting, but much weaker, magnetic fields through the dynamo mechanism [594, 595, 596, 597, 598]. This process is effective, however, only if a non-zero magnetic field is present for the dynamos to amplify. The origin of these magnetic field "seeds," which were present at the onset of structure formation, remains an open question and has generated a great deal of speculation [599, 600, 601, 597, 602, 603]. It has been proposed that primordial magnetic fields could arise within the context of inflation [604, 605, 606, 607, 602, 608, 609] or during phase transitions that took place in the early universe [610, 611, 612, 613, 614, 255, 615]. None of these scenarios is completely satisfactory, though, and each faces its own challenges. In particular, it is difficult for these mechanisms to produce fields of sufficiently large correlation length so as to survive until today.

The origin of the primordial magnetic field is somewhat obscured by the complicated plasma and magnetohydrodynamics processes that have taken place over cosmic time. One can attempt, however, to constrain the properties of the seed field by studying the magnetic fields found within the voids of the intergalactic medium, where primordial fields could exist in a relatively pristine state. In such environments, the evolution of the magnetic field would

19. Certain inflationary mechanisms are capable of producing magnetic fields of sufficient strength ($\gtrsim 0.1$ nG on Mpc scales) that they can be adiabatically compressed to explain the μG strength fields in galaxies today without invoking the dynamo mechanism [592, 593].

be largely driven by the expansion of the universe, leading to the dilution of the field strength as $B \propto a^{-2}$ (corresponding to $\rho_B \propto a^{-4}$), and to the growth of the field's correlation length as $\xi \propto a$.

In this section, we consider the possibility that primordial magnetic fields may have been generated in the early universe by a subdominant population of primordial black holes. In order to produce a non-zero magnetic field, these black holes must have been both spinning and electrically charged, corresponding to the Kerr-Newman solution. In our scenario, this population is temporarily charged in the early universe due to a nonzero chemical potential, which eventually relaxes to zero, at which point the black holes discharge. Afterwards, the magnetic fields evolve according to Hubble expansion and the (now neutral) black holes constitute a present day dark matter abundance. While such a scenario is admittedly quite speculative and involves some rather exotic elements, we find that astrophysically interesting magnetic fields could have potentially been generated by such objects.

5.3.2 Kerr-Newman Black Holes

Generating a magnetic field requires both an electromagnetic current and a departure from spherical symmetry. For this reason, we are interested here in black holes that are both charged and rotating. Such Kerr-Newman black holes are entirely characterized by their mass, M, angular momentum, J, and charge, Q. In Boyer-Lindquist coordinates, the geometry associated with such an object is described by the following line element [178, 616, 617]:

$$ds^2 = -\frac{\Delta}{\rho^2}(dt - \alpha \sin^2\theta \, d\phi)^2 + \frac{\rho^2}{\Delta} dr^2 + \rho^2 d\theta^2 + \frac{\sin^2\theta}{\rho^2}\left[(r^2+\alpha^2)d\phi - \alpha dt\right]^2,$$

where $\alpha = J/M$, and we have defined

$$\rho^2 = r^2 + \alpha^2 \cos^2\theta, \;\; \Delta = r^2 + \alpha^2 - \frac{2Mr}{M_{\rm Pl}^2} + \frac{Q^2}{M_{\rm Pl}^2}, \tag{5.75}$$

and $M_{\rm Pl} = 1.22 \times 10^{19}$ GeV is the Planck mass. The charge and angular momentum of a black hole are constrained to lie within the following domain:

$$\alpha^2 \, M_{\rm Pl}^2 + Q^2 \leq \frac{M^2}{M_{\rm Pl}^2} \, . \tag{5.76}$$

From the metric, we see that the Kerr-Newman black hole has two horizons located at

$$r_\pm = \frac{1}{M_{\rm Pl}^2} \left(M \pm \sqrt{M^2 - \alpha^2 \, M_{\rm Pl}^4 - Q^2 \, M_{\rm Pl}^2} \right) . \tag{5.77}$$

Integrating over the angular volume element evaluated on the $r = r_+$ hypersurface yields the area of the event horizon

$$A = 4\pi \, (r_+^2 + \alpha^2). \tag{5.78}$$

From the Killing vector associated with the event horizon, the surface gravity can be written as [617]

$$\kappa = \frac{2\pi}{A}(r_+ - r_-). \tag{5.79}$$

These two quantities are related to a black hole's temperature and entropy as follows [618]:

$$T_{\rm BH} = \frac{\kappa}{2\pi} = \frac{r_+ - r_-}{4\pi(r_+^2 + \alpha^2)} \tag{5.80}$$

$$S_{\rm BH} = \frac{A}{4} = \pi(r_+^2 + \alpha^2). \tag{5.81}$$

These expressions, in conjunction with the fact that the mass of a black hole can be identified with energy, yields the first law of black hole thermodynamics:

$$dM = \frac{M_{\rm Pl}^2}{8\pi} \, \kappa \, dA + \Omega \, dJ + \Phi \, dQ, \tag{5.82}$$

where Ω and Φ are the angular velocity and the electrostatic potential of the black hole. Note that the quantities κ (and hence $T_{\rm BH}$), Ω, and Φ are constant over the horizon. In order to obtain explicit forms for Ω and Φ in the context of a Kerr-Newman black hole, we need to take the differential of the area given in Eq. (5.78). After some algebra, we can write

$$\frac{M_{\rm Pl}^2}{8\pi}\,\kappa\,dA = \frac{r_+^2\,dM}{r_+^2+\alpha^2} - \frac{r_+ Q dQ}{r_+^2+\alpha^2} - \frac{M\alpha d\alpha}{r_+^2+\alpha^2}. \tag{5.83}$$

Substituting $-\alpha M d\alpha = -\alpha dJ + \alpha^2 dM$ and inserting the explicit form for κ, we arrive at the following expression:

$$dM = \frac{M_{\rm Pl}^2}{4}\,T_{\rm BH}\,dA + \frac{\alpha dJ}{r_+^2+\alpha^2} + \frac{r_+ Q dQ}{r_+^2+\alpha^2}. \tag{5.84}$$

Comparing this to Eq. (5.82), we can determine the black hole's angular velocity and electrostatic potential:

$$\Omega = \frac{\alpha}{r_+^2+\alpha^2}, \quad \Phi = \frac{r_+ Q}{r_+^2+\alpha^2}. \tag{5.85}$$

5.3.3 Generating Cosmological Magnetic Fields

We begin by considering an isolated black hole whose mass, angular momentum, and charge are not appreciably evolving with time, hence neglecting the possible effects of Hawking evaporation and accretion. This stationary geometry is described by the Kerr-Newman metric given in Eq. (5.75). Technically this is just one half of the complete solution to the coupled Einstein-Maxwell equations, which describe the interplay between the dynamical metric and electromagnetic field. For a full solution, we must also specify the vector potential A_μ [617]

$$A_\mu dx^\mu = -\frac{Qr}{r^2+\alpha^2\cos^2\theta}\left(dt - \alpha\sin^2\theta\,d\phi\right). \tag{5.86}$$

Using the field strength, $F_{\mu\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu$, the E and B fields are given by

$$\vec{E} = \frac{Q(r^2 - \alpha^2 \cos^2\theta)}{\rho^4}\,\hat{r} - \frac{2Q\alpha^2 \cos\theta \sin\theta}{\rho^4}\,\hat{\theta}\,, \tag{5.87a}$$

$$\vec{B} = \frac{Q\alpha}{r}\left[\frac{2(\alpha^2 + r^2)\cos\theta}{\rho^4}\,\hat{r} + \frac{(r^2 - \alpha^2\cos^2\theta)\sin\theta}{\rho^4}\,\hat{\theta}\right]. \tag{5.87b}$$

Note that the E_ϕ and B_ϕ components are both vanishing, since we've taken the black hole to be rotating in the $\hat{\phi}$ direction. Also note that in the $r \to \infty$ limit, these fields have the expected asymptotic forms

$$\lim_{r\to\infty} \vec{E} = \frac{Q}{r^2}\,\hat{r} + \mathcal{O}\left(\frac{1}{r^3}\right), \tag{5.88a}$$

$$\lim_{r\to\infty} \vec{B} = \frac{Q\alpha}{r^3}\left(2\cos\theta\,\hat{r} + \sin\theta\,\hat{\theta}\right) + \mathcal{O}\left(\frac{1}{r^4}\right).$$

In considering the case of an isotropic population of black holes,[20] it will be useful to have an expression for the magnetic field of a single black hole averaged over a sphere of radius $R > r_+$. We adopt the volume-averaged convention[21]

$$\langle \vec{B} \rangle = \frac{1}{V}\int_V d^3x\,\vec{B}\,, \tag{5.89}$$

where $V = 4\pi R^3/3$ is the volume of the sphere over which we are averaging. Starting from Eq. (5.87) and omitting the algebraic details, the volume-averaged magnetic field magnitude can be written as

$$\langle B \rangle = \frac{3Q}{R^2}\left[\left(1 + \frac{R^2}{\alpha^2}\right)\tan^{-1}\left(\frac{\alpha}{R}\right) - \frac{R}{\alpha}\right]. \tag{5.90}$$

In the $\alpha \ll R$ limit, the average magnetic field reduces to $\langle B \rangle \approx 2Q\alpha/R^3$. This limit will

20. A possible objection to this scenario is that the black holes might act as an ensemble of magnetic dipoles which interact to form domains of some characteristic scale. This will not be applicable in this case, however, as we will consider black hole number densities which are sufficiently small such that no more than one black hole will be present in a given Hubble radius at early times.

21. We have confirmed numerically that our definition coincides with the RMS average value, $B^2_{\rm RMS} = \frac{1}{V}\int d^3x\,\vec{B}^2$, up to an $\mathcal{O}(1)$ factor.

be applicable throughout our entire parameter space of interest.

The primordial magnetic field is also characterized by a correlation length ξ, which governs the extent to which diffusion and damping will suppress any magnetic fields that are generated by black holes in the early universe. On scales greater than the magnetic diffusion length $\ell_{\rm diff}$, diffusive effects can be neglected, so the comoving field is said to be "frozen in" and ξ grows linearly with the scale factor of the universe.

Formally ξ_B is defined as the length scale after which correlations, as quantified by the two-point correlation function, fall off exponentially. Intuitively we expect this to coincide approximately with the average distance between neighboring primordial black holes, as it is at this length scale that magnetic field lines will begin to interfere with and wash out one another

$$\xi \sim \left(\frac{3}{4\pi n_{\rm BH}}\right)^{1/3} = \left(\frac{45}{2\pi^3 g_\star(T)}\frac{M}{f_{\rm BH}T^4}\right)^{1/3}, \tag{5.91}$$

where $f_{\rm BH} = Mn_{\rm BH}/\rho_R(T)$ is the energy fraction in black holes relative to that in radiation at the time of magnetogenesis and $g_\star(T)$ is the number of effective relativistic degrees of freedom at temperature T.

Once a magnetic field is generated at some initial temperature, T_i, there are several processes which can affect its evolution, including small scale damping, diffusion, and the expansion of the universe [598, 600]. We will make the simplifying assumption that the initial correlation length is sufficiently large that we do not need to account for the former effects $\xi_i > \ell_{\rm diff}$, and focus solely on the impact of Hubble expansion. We will later verify that this assumption is self-consistent for all parameter space of interest. In an expanding universe, the magnetic field redshifts as $B \propto a^{-2}$, while the correlation length grows as $\xi \propto a$. These scalings are manifest when writing B and ξ in terms of temperature

$$B(T) = B_i\left(\frac{T}{T_i}\right)^2\left[\frac{g_{\star,S}(T)}{g_{\star,S}(T_i)}\right]^{2/3}, \tag{5.92a}$$

$$\xi(T) = \xi_i \left(\frac{T_i}{T}\right) \left[\frac{g_{\star,S}(T_i)}{g_{\star,S}(T)}\right]^{1/3} , \tag{5.92b}$$

where $g_{\star,S}(T)$ is the effective number of degrees-of-freedom in entropy, and the initial values at magnetogenesis, B_i and ξ_i, can be related to black hole parameters using Eqs. (5.90) and (5.91), with $R = \xi_i$. Defining the following dimensionless parameters

$$\alpha_\star \equiv \alpha \frac{M_{\rm Pl}^2}{M} = J \frac{M_{\rm Pl}^2}{M^2} , \;\; Q_\star \equiv Q \frac{M_{\rm Pl}}{M}, \tag{5.93}$$

the magnetic field from Eq. (5.92a) can be written as

$$\langle B_0 \rangle = \frac{4\pi^3 \alpha_\star Q_\star f_{{\rm BH},i} g_\star(T_i) T_i^2 T_0^2 M}{45\, M_{\rm Pl}^3} \left[\frac{g_{\star,S}(T_0)}{g_{\star,S}(T_i)}\right]^{2/3} , \tag{5.94}$$

where present day values are denoted by a "0" subscript, $T_0 = 2.725$ K is the CMB temperature, and $f_{{\rm BH},i}$ is the black hole energy fraction at T_i. In order to express this in terms of current observables, we apply the conservation of entropy

$$\frac{g_{\star,S}(T_0)}{g_{\star,S}(T_i)} = \left(\frac{a_i T_i}{a_0 T_0}\right)^3 , \tag{5.95}$$

where $a_{i,0}$ is the scale factor at the corresponding epoch. Noting also that the initial black hole energy density at magnetogenesis satisfies

$$\rho_{\rm BH}(T_i) = f_{{\rm BH},i} \left(\frac{\pi^2 g_\star(T_i) T_i^4}{30}\right) = \Omega_{\rm BH} \rho_c \left(\frac{a_0}{a_i}\right)^3 , \tag{5.96}$$

where $\Omega_{\rm BH} \equiv \rho_{\rm BH}/\rho_c$ is the present day energy density in black holes relative to the critical

density $\rho_c \approx 4 \times 10^{-47}$ GeV^4, we can rewrite Eq. (eq:B) as

$$\begin{aligned}\langle B_0 \rangle &= \frac{8\pi\alpha_\star Q_\star \Omega_{\rm BH}\rho_c M}{3M_{\rm Pl}^3}\frac{T_i}{T_0}\left[\frac{g_{\star,S}(T_i)}{g_{\star,S}(T_0)}\right]^{1/3} \\ &\approx 6\times 10^{-16}\,{\rm G}\left(\frac{Q_\star\alpha_\star}{0.5}\right)\left(\frac{\Omega_{\rm BH}}{0.01}\right)\frac{M}{M_\odot}\left(\frac{T_i}{\rm GeV}\right),\end{aligned} \tag{5.97}$$

where in the last line we have used $g_\star = g_{\star,S} = 73$ at $T_i = 1$ GeV. Note that in terms of $\alpha_\star$ and $Q_\star$, the extremality condition is $\alpha_\star^2 + Q_\star^2 \leq 1$, which implies $\alpha_\star Q_\star \leq 0.5$. Similarly, combining Eqs. (5.91) and (5.92b), the present day correlation length can be written as

$$\xi_0 = \frac{1}{T_0}\frac{45M}{2\pi^3 g_\star(T_i)\, f_{{\rm BH},i}\, T_i}^{1/3}\left[\frac{g_{\star,S}(T_i)}{g_{\star,S}(T_0)}\right]^{1/3}. \tag{5.98}$$

Using Eqs. (5.95) and (5.96), we obtain

$$\xi_0 = \left(\frac{3M}{4\pi\,\Omega_{\rm BH}\rho_c}\right)^{1/3} \approx 0.6\,{\rm kpc}\,\left(\frac{0.01}{\Omega_{\rm BH}}\right)^{1/3}\left(\frac{M}{M_\odot}\right)^{1/3}. \tag{5.99}$$

Naively applying Eq. 5.97, it might appear that arbitrarily strong magnetic fields could be generated by black holes at sufficiently high temperatures, $T_i \gg$ GeV. Black holes of a given mass, however, can only be formed once $M > M_H$, where M_H is the mass contained within the horizon

$$M_H = \frac{M_{\rm Pl}^2}{2H_i} \approx 0.06 M_\odot\,\left(\frac{\rm GeV}{T_i}\right)^2\left(\frac{73}{g_\star(T_i)}\right)^{1/2}. \tag{5.100}$$

By evaluating Eq. (5.97) at M_H, we find the following upper limit for the magnetic field strength that could be generated by spinning, charged black holes

$$\langle B_0\rangle_{\rm max} \approx 4\times 10^{-17}\,{\rm G}\,\left(\frac{Q_\star\alpha_\star}{0.5}\right)\left(\frac{\Omega_{\rm BH}}{0.01}\right)\left(\frac{\rm GeV}{T_i}\right). \tag{5.101}$$

Alternatively, in terms of the horizon mass, this maximum magnetic field can be written as

$$\langle B_0 \rangle_{\rm max} \approx 1.5 \times 10^{-16}\,{\rm G}\,\left(\frac{Q_\star \alpha_\star}{0.5}\right)\left(\frac{\Omega_{\rm BH}}{0.01}\right)\left(\frac{M_H}{M_\odot}\right)^{1/2}. \tag{5.102}$$

5.3.4 Potentially Viable Parameter Space

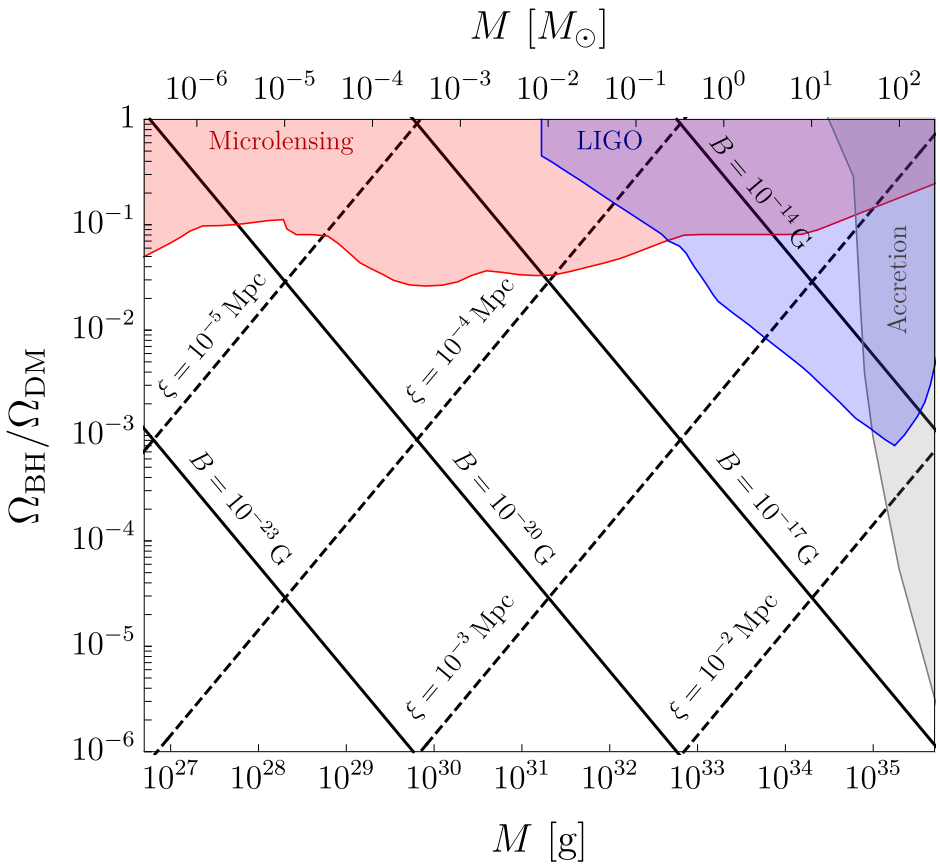

Figure 5.12: The present day strength and correlation length of the magnetic fields generated by primordial black holes, for the optimal case of $Q_\star \alpha_\star = 0.5$. Also shown are the constraints on this parameter space from gravitational microlensing surveys [15, 16, 17, 18], gravitational wave observations [19, 20, 21], and from the impact of accretion [22]. Astrophysically relevant magnetic fields ($B \gtrsim 10^{-20}\,{\rm G}$) could be generated by primordial black holes in the mass range of $M \sim 10^{28} - 10^{36}\,{\rm g}$ without violating existing constraints. For non-extremal black holes, the strength of the resulting magnetic fields would be smaller than those shown by a factor of $Q_\star \alpha_\star / 0.5$.

In Fig. 5.12, we plot the strength and correlation length of the magnetic fields gen-

erated by primordial black holes, for the optimal case of $Q_\star \alpha_\star = 0.5$. Also shown in this figure are the constraints on this parameter space from gravitational microlensing surveys [15, 16, 17, 18], gravitational wave observations [19, 20, 21], and from the impact of accretion on the CMB [22] (for reviews, see Refs. [619, 620]). From this figure, we see that astrophysically relevant magnetic fields ($B \gtrsim 10^{-20}\,\mathrm{G}$) could potentially have been generated by primordial black holes with masses in the range of $M \sim 10^{28} - 10^{36}\,\mathrm{g}$, without violating any existing constraints. The smallest mass in this range corresponds to the horizon mass at ~ 100 GeV, so this should be interpreted as an upper bound on the range of viable production temperatures. A lower bound on T_i comes from the requirement that the black holes discharge prior to BBN, $T_i \gtrsim 10$ MeV. Throughout this mass range, once the black holes discharge, their Hawking radiation is negligible, so this population constitutes a fraction of the dark matter today [621, 622].

In this region of parameter space, the correlation length of the present day magnetic field falls in the range of $\xi \sim 10^{-6} - 10^{-1}\,\mathrm{Mpc}$. Across this range of values, the magnetic fields are predicted to have survived the effects of magnetic dissipation and diffusion [598, 623, 624, 599]. More explicitly, in order to avoid early magnetic dissipation the present day field should satisfy [625]

$$\xi_0 \gtrsim 10^{-7}\,\mathrm{Mpc}\left(\frac{\langle B_0 \rangle}{10^{-15}\,\mathrm{G}}\right) . \tag{5.103}$$

This condition is easily fulfilled for the relevant parameter space in Fig. 5.12. Thus, in this regime we are justified in considering only Hubble expansion in translating the early universe field to its present day value.

Though too heavy and subdominant to be of interest as dark matter candidates, and so not included in figure, there is additional viable parameter space at small $\Omega_{\rm BH}/\Omega_{\rm DM} \lesssim 10^{-5}$ and to the right of the accretion bound, which truncates at $M \sim 5 \times 10^{37}\,\mathrm{g}$ (see e.g. [620]). The existence of this additional parameter space is important nevertheless in light of updated lower bounds on the intergalactic magnetic field (IGMF) strength based on observations of

electromagnetic cascades from blazars. The non-detection of GeV-scale gamma-ray emission from electromagnetic cascades initiated from extragalatic TeV sources can be used to place lower bounds on the IGMF, as first argued by [626], who claimed a lower bound of $B \geq 3 \times 10^{-16}\,\mathrm{G}$. This bound has since been improved upon by analysis of VERITAS [627] and *Fermi*/LAT [628] data. Most recently, incorporating data from MAGIC [629] has improved the bound to $B \gtrsim 10^{-14}\,\mathrm{G}$. This is much more robust than previous estimates as it does away with the uncertainty in variability of the TeV band source flux. However, it still suffers from some uncertainty in the fraction of power carried away by plasma instabilities which develop from interactions between the generated particle stream and mimic the effect of the IGMF in reducing the expected cascade flux.

5.3.5 Charged Black Holes and Chemical Potentials

Thus far, we have remained agnostic regarding the origin of the Kerr-Newman black holes. We present here one concrete model capable of endowing an existing population of primordial black holes with electric charge. Of course it's very difficult to create black holes with geometrically significant charge in the early universe. In a cosmological setting, any net charge would be quickly neutralized by the surrounding plasma, which must[22] have an opposite compensating charge to maintain the charge neutrality of the universe. Even if one were to consider a charged black hole in vacuum, its charge is expelled exponentially quickly through Hawking radiation or Schwinger pair production [630]. A population of charged black holes thus requires the introduction of new physics.

The Hawking radiation of electrically neutral black holes is symmetric with respect to the production of particles and anti-particles. By contrast, charged black holes preferentially radiate particles with the same sign charge as the black hole. To understand why, consider the

22. Black holes may violate global symmetries, but not gauge symmetries, and so any charge taken up by the black hole must be lost by the plasma.

flux spectrum for a Kerr-Newman black hole, which follows the thermal distribution [621, 631]

$$dN \sim \frac{d\omega}{\exp\left[(\omega - m\Omega - q\Phi)/T_{\rm BH}\right] \mp 1}\,, \tag{5.104}$$

where ω and m are the energy and angular momentum of an emitted particle, Ω and Φ are the angular velocity and electrostatic potential of the black hole, and $\mp$ refers to bosons and fermions, respectively. We can identify $\mu_q \equiv q\Phi$ with a chemical potential, biasing[23] the emission of particles whose charge is aligned with the black hole.

This chemical potential is actually sourced by the electromagnetic potential A_μ of the Kerr-Newman black hole itself. A particle of charge q at the horizon couples to A_μ as $\mathcal{L} \supset -qA_\mu J^\mu_{\rm EM}$. Since the time-like component couples to the charge density $J^0_{\rm EM}$, we can identify the combination $-qA_0|_{r_+}$ with a chemical potential, μ_q

$$\mathcal{L} \supset -qA_0 J^0_{\rm EM} \equiv \mu_q J^0_{\rm EM}\,. \tag{5.105}$$

Using Eqs. (5.85) and (5.86), one can verify that this combination $\mu_q = -qA_0|_{r_+}$ is identical to $\mu_q = q\Phi$, and so we see this is self-consistent.

Just as the intrinsic chemical potential of the Kerr-Newman black hole allows it to shed its charge, one can imagine charging up a black hole (or maintaining a black hole in a charged state) by means of an external chemical potential. If such a chemical potential is greater than that of the black hole itself, then the black hole will build up charge until it reaches an extremal state.

One possible mechanism for realizing such a chemical potential involves a new scalar field

23. From this expression, we also see that $m\Omega$ acts in a similar manner, leading the black hole to preferentially expel particles whose angular momentum is aligned with the black hole. Thus, the black hole will shed both quantities as it evaporates, evolving towards a neutral, non-rotating state.

ϕ derivatively coupled to the electromagnetic current via:

$$\mathcal{L} \supset \frac{1}{\Lambda}\partial_\mu \phi J^\mu_{\rm EM}\,. \tag{5.106}$$

Such an operator generically arises in the effective Lagrangian description as the most relevant coupling to the Standard Model if ϕ has an approximate shift symmetry. In this case Λ would be interpreted as the symmetry breaking scale. If ϕ is initially displaced from the origin and begins rolling in the early universe, its time derivative will source an effective chemical potential for charged particles, $\mu_\phi \equiv \dot{\phi}$, leading the black hole to preferentially absorb particles with charge of a particular sign. Magnetic field generation will occur during the period in which the external chemical potential is active because the scalar field is rolling. Once ϕ stops rolling at temperature T_i, the chemical potential will vanish and the black hole will quickly expel its charge, thereby returning to a neutral state.

5.3.6 Summary

In this section, we have studied the possibility that cosmological magnetic fields may have been generated in the early universe by a population of primordial Kerr-Newman black holes. We find that black holes near extremality ($\alpha_\star Q_\star \sim 0.5$) in the mass range of $M \sim 10^{28} - 10^{36}\,$g would have been capable of producing present day magnetic fields that are as large as $B \sim 10^{-15}\,$G. Black holes at larger masses $M \gtrsim 10^{38}\,$g and smaller fractional abundance could have produced even larger fields $B \gtrsim 10^{-14}\,$G. The corresponding correlation lengths are sufficiently large that these fields would have survived the effects of early magnetic dissipation and diffusion. Thus these fields could have seeded larger galactic and intergalactic fields through the dynamo mechanism.

In order to generate a magnetic field, the black holes in this scenario must be both spinning and electrically charged. Throughout most of our analysis, we have remained agnostic

as to the origin of these Kerr-Newman black holes. While it is straightforward to create spinning black holes through the mergers of an initial population of Schwarzschild black holes [632], it is more challenging to explain how these black holes acquire an appreciable net electric charge in the early universe. One possibility involves a rolling scalar field which dynamically generates a chemical potential for charged particle species, thereby biasing the charge distribution of Hawking radiation and the net flow of charge into the black holes. We leave the model building that concretely realizes such a scenario for future work.

Our analysis has also largely omitted the complicated magnetohydrodynamics (MHD) processes that govern the evolution of magnetic fields in a hot thermal plasma. While a proper treatment depends on the spectral shape and is beyond the scope of this letter, we note that large scale MHD decay will generically serve to increase the correlation length and power on large scales [633]. This occurs because power on small scales dissipates more efficiently and because there exists a weak inverse cascade which shifts energy away from the dissipation scale, even in the case of non-helical freely decaying MHD turbulence [634]. Consequently, our estimates for the correlation length, which took into account only the passive expansion of the universe, are conservative, and we can all the more safely neglect the effects of diffusion and dissipation. Note though that there may still be phenomena such as turbulence which could serve to tangle magnetic field lines and excite dissipative modes. Turbulent decay will generically serve to reduce the total magnetic energy density [635], such that the figures for the field strength in the previous section should be seen as an upper estimate.

Another feature omitted from our treatment is the helicity of the magnetic field, $\mathcal{H} = \int d^3x \, \vec{A} \cdot \vec{B}$. Magnetic helicity is conserved during evolution and can result in interesting effects. One can see from the form of the vector potential in Eq. (5.86) and magnetic field in Eq. (5.87) that there is no intrinsic magnetic helicity associated with the Kerr-Newman solution. If the initial velocity field has a non-vanishing kinetic helicity, however,

it is possible for the magnetic field on large scales to develop a corresponding helicity, with an opposite helicity developing on small scales by conservation of helicity [636]. A more thorough investigation of these effects, including numerical simulations, is left to future investigation.

Finally, we note that this scenario predicts a nontrivial relationship between the primordial magnetic field parameters and the merger rate for the progenitor black hole population. This suggests that the primary observable signature of the model would be a correlation between the parameters of the cosmological magnetic field and the stochastic gravitational wave background coming from the merger of primordial black hole binaries. Presuming a monochromatic mass function, the present day merger rate of binaries which formed in the early universe can be roughly estimated as [637]

$$R_3(t_0) \sim 80 \left(\frac{\Omega_{\rm BH}}{10^{-3}}\right)^{53/37} \left(\frac{M_{\rm BH}}{10^{34}\,{\rm g}}\right)^{-32/37} {\rm Gpc}^{-3}{\rm yr}^{-1}\,, \tag{5.107}$$

where we have conservatively set the local density contrast at decoupling $\delta_{\rm dc}$, which quantifies the effect of clustering, to ~ 1. This can then be compared with the rate indicated by LIGO observations, $12 - 213\,{\rm Gpc}^{-3}{\rm yr}^{-1}$ [638]. A more thorough analysis depends sensitively on the clustering dynamics and is left to future investigation.

CHAPTER 6

CONCLUSIONS

The Standard Model and the ΛCDM cosmological paradigm are both tremendously successful theories, and their combined power has allowed us to understand much of the current state of our universe and how is came to be. However, as discussed in Sec. 2.3, there are a number of open questions left unaddressed by either. This thesis has presented several examples of model building efforts towards a more complete theory of particle cosmology.

We began in Chapter 2 by reviewing the current state of the Standard Model of particle physics and the ΛCDM cosmological model, including both their successes and their failings. We then turned to the early universe, about which much remains to be understood. In Chapter 3, we saw that primordial black holes and gravitational waves show exceptional promise as probes of the early universe. In particular, we saw how constraints on primordial black hole abundances could inform us about the shape of the small scale primordial power spectrum and thereby inflation. We also saw how gravitational waves from evaporating primordial black holes allows us to put constraints on non-standard expansion histories and extra-dimensional scenarios.

In Chapter 4, we studied another source of gravitational waves in the early universe — strongly first-order cosmological phase transitions. After reviewing the thermodynamic and hydrodynamic descriptions of these transitions and the ingredients that go into computing the gravitational wave signal, we turned to one of the major sources of uncertainties in calculations of first-order phase transitions — thermal resummation of the effective potential. We compared the efficiency and accuracy of different resummation prescriptions, in addition to commenting on the errors incurred by certain approximations in the literature. We then turned to study 2 Higgs doublet models, which are well-motivated extensions of the Standard Model in principle capable of making the electroweak phase transition strongly first-order.

Finally, Chapter 5 was dedicated to progress towards resolving three persistent cosmolog-

ical puzzles. In the context of dark matter model building, we introduced a new scheme of asymmetric reheating capable of reconciling dark sectors with light degrees of freedom and exact symmetries with precision constraints on cosmological observables. Next, we examined the questions raised by recent observations of supermassive black holes at high redshifts, and posited that the timing problem could be mollified with supermassive black holes of a primordial origin. Finally, we suggested a novel mechanism for producing the tiny primordial magnetic fields required to seed the galactic fields observed today, another longstanding mystery.

Of course, the model building efforts that have been presented here are ultimately just small steps towards solutions for a tiny subset of problems. The bulk of the work lies before us. Thankfully, we have much experimental data on the horizon to guide us. Now, at the onset of the age of multi-messenger cosmology and gravitational wave astronomy, it is perhaps the most exciting time to be a particle cosmologist. With the influx of coming data, we are now in a better position than ever before to unravel the mysteries of the early universe and their echoes today. And hopefully, we will prompt a few new questions as well.

REFERENCES

[1] N. Aghanim et al. Planck 2018 results. V. CMB power spectra and likelihoods. *Astron. Astrophys.*, 641:A5, 2020.

[2] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[3] Julien Lesgourgues. The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *arXiv e-prints*, page arXiv:1104.2932, April 2011.

[4] Diego Blas, Julien Lesgourgues, and Thomas Tram. The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. *jcap*, 2011(7):034, July 2011.

[5] C. J. Moore, R. H. Cole, and C. P. L. Berry. Gravitational-wave sensitivity curves. *Class. Quant. Grav.*, 32(1):015014, 2015.

[6] Nancy Aggarwal et al. Challenges and opportunities of gravitational-wave searches at MHz to GHz frequencies. *Living Rev. Rel.*, 24(1):4, 2021.

[7] Iason Baldes and Géraldine Servant. High scale electroweak phase transition: baryogenesis & symmetry non-restoration. *JHEP*, 10:053, 2018.

[8] Yue Shen, Gordon T. Richards, Michael A. Strauss, Patrick B. Hall, Donald P. Schneider, Stephanie Snedden, Dmitry Bizyaev, Howard Brewington, Viktor Malanushenko, Elena Malanushenko, Dan Oravetz, Kaike Pan, and Audrey Simmons. A catalog of quasar properties from sloan digital sky survey data release 7. *The Astrophysical Journal Supplement Series*, 194(2):45, jun 2011.

[9] Anne M. Green and Bradley J. Kavanagh. Primordial Black Holes as a dark matter candidate. *J. Phys. G*, 48(4):043001, 2021.

[10] Y. Akrami et al. Planck 2018 results. X. Constraints on inflation. *Astron. Astrophys.*, 641:A10, 2020.

[11] Simeon Bird, Hiranya V. Peiris, Matteo Viel, and Licia Verde. Minimally Parametric Power Spectrum Reconstruction from the Lyman-alpha Forest. *Mon. Not. Roy. Astron. Soc.*, 413:1717–1728, 2011.

[12] John C. Mather et al. Measurement of the Cosmic Microwave Background spectrum by the COBE FIRAS instrument. *Astrophys. J.*, 420:439–444, 1994.

[13] D. J. Fixsen, E. S. Cheng, J. M. Gales, John C. Mather, R. A. Shafer, and E. L. Wright. The Cosmic Microwave Background spectrum from the full COBE FIRAS data set. *Astrophys. J.*, 473:576, 1996.

[14] Christian T. Byrnes, Philippa S. Cole, and Subodh P. Patil. Steepest growth of the power spectrum and primordial black holes. *JCAP*, 06:028, 2019.

[15] R. A. Allsman et al. MACHO project limits on black hole dark matter in the 1-30 solar mass range. *Astrophys. J. Lett.*, 550:L169, 2001.

[16] P. Tisserand et al. Limits on the Macho Content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds. *Astron. Astrophys.*, 469:387–404, 2007.

[17] Hiroko Niikura, Masahiro Takada, Shuichiro Yokoyama, Takahiro Sumi, and Shogo Masaki. Constraints on Earth-mass primordial black holes from OGLE 5-year microlensing events. *Phys. Rev. D*, 99(8):083503, 2019.

[18] Masamune Oguri, Jose M. Diego, Nick Kaiser, Patrick L. Kelly, and Tom Broadhurst. Understanding caustic crossings in giant arcs: characteristic scales, event rates, and constraints on compact dark matter. *Phys. Rev. D*, 97(2):023518, 2018.

[19] Bradley J. Kavanagh, Daniele Gaggero, and Gianfranco Bertone. Merger rate of a subdominant population of primordial black holes. *Phys. Rev. D*, 98(2):023536, 2018.

[20] B. P. Abbott et al. Search for Subsolar Mass Ultracompact Binaries in Advanced LIGO's Second Observing Run. *Phys. Rev. Lett.*, 123(16):161102, 2019.

[21] Zu-Cheng Chen and Qing-Guo Huang. Distinguishing Primordial Black Holes from Astrophysical Black Holes by Einstein Telescope and Cosmic Explorer. *JCAP*, 08:039, 2020.

[22] Pasquale D. Serpico, Vivian Poulin, Derek Inman, and Kazunori Kohri. Cosmic microwave background bounds on primordial black holes including dark matter halo accretion. *Phys. Rev. Res.*, 2(2):023204, 2020.

[23] R. L. Workman and Others. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[24] Dan Hooper, Aurora Ireland, Gordan Krnjaic, and Albert Stebbins. Supermassive Primordial Black Holes From Inflation. 8 2023.

[25] Aurora Ireland, Stefano Profumo, and Jordan Scharnhorst. Primordial gravitational waves from black hole evaporation in standard and nonstandard cosmologies. *Phys. Rev. D*, 107(10):104021, 2023.

[26] Aurora Ireland, Stefano Profumo, and Jordan Scharnhorst. Gravitational Waves from Primordial Black Hole Evaporation with Large Extra Dimensions. 12 2023.

[27] Henning Bahl, Marcela Carena, Nina M. Coyle, Aurora Ireland, and Carlos E. M. Wagner. New tools for dissecting the general 2HDM. *JHEP*, 03:165, 2023.

[28] Aurora Ireland and Seth Koren. Asymmetric Reheating via Inverse Symmetry Breaking. 11 2022.

[29] Dan Hooper, Aurora Ireland, and Gordan Krnjaic. Cosmological magnetic fields from primordial Kerr-Newman black holes. *Phys. Rev. D*, 107(10):103524, 2023.

[30] Aurora Ireland and Vasudev Shyam. $T\overline{T}$ deformed YM_2 on general backgrounds from an integral transformation. *JHEP*, 07:058, 2020.

[31] George F. R. Ellis and Roy Maartens. The emergent universe: Inflationary cosmology with no singularity. *Class. Quant. Grav.*, 21:223–232, 2004.

[32] Robert Brandenberger and Patrick Peter. Bouncing Cosmologies: Progress and Problems. *Found. Phys.*, 47(6):797–850, 2017.

[33] Paul J. Steinhardt and Neil Turok. Cosmic evolution in a cyclic universe. *Phys. Rev. D*, 65:126003, 2002.

[34] G. Lemaître. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Societe Scientifique de Bruxelles*, 47:49–59, January 1927.

[35] Georges Lemaître. L'Univers en expansion. *Annales de la Société Scientifique de Bruxelles*, 53:51, January 1933.

[36] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *apj*, 142:419–421, July 1965.

[37] Robert V. Wagoner, William A. Fowler, and F. Hoyle. On the Synthesis of Elements at Very High Temperatures. *apj*, 148:3, April 1967.

[38] P. J. E. Peebles. Primordial Helium Abundance and the Primordial Fireball. 2. *Astrophys. J.*, 146:542–552, 1966.

[39] P. J. E. Peebles. Recombination of the Primeval Plasma. *Astrophys. J.*, 153:1, 1968.

[40] Michael S. Smith, Lawrence H. Kawano, and Robert A. Malaney. Experimental, Computational, and Observational Analysis of Primordial Nucleosynthesis. *apjs*, 85:219, April 1993.

[41] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helv. Phys. Acta*, 6:110–127, 1933.

[42] Vera C. Rubin and Jr. Ford, W. Kent. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *apj*, 159:379, February 1970.

[43] S. Perlmutter et al. Measurements of Ω and Λ from 42 High Redshift Supernovae. *Astrophys. J.*, 517:565–586, 1999.

[44] B.P. Schmidt et al. The High-Z Supernova Search: Measuring Cosmic Deceleration and Global Curvature of the Universe Using Type IA Supernovae. *apj*, 507(1):46–63, November 1998.

[45] Elcio Abdalla et al. Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies. *JHEAp*, 34:49–211, 2022.

[46] Eleonora Di Valentino, Olga Mena, Supriya Pan, Luca Visinelli, Weiqiang Yang, Alessandro Melchiorri, David F. Mota, Adam G. Riess, and Joseph Silk. In the realm of the Hubble tension—a review of solutions. *Class. Quant. Grav.*, 38(15):153001, 2021.

[47] Dan Hooper. *Particle Cosmology and Astrophysics*. Princeton University Press, 5 2024.

[48] R. A. Alpher, H. Bethe, and G. Gamow. The origin of chemical elements. *Phys. Rev.*, 73:803–804, 1948.

[49] Keith A. Olive, David N. Schramm, Gary Steigman, and Terry P. Walker. BIG BANG NUCLEOSYNTHESIS REVISITED. *Phys. Lett. B*, 236:454–460, 1990.

[50] Brian D. Fields, Paolo Molaro, and Subir Sarkar. Big-Bang Nucleosynthesis. *Chin. Phys. C*, 38:339–344, 2014.

[51] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *apj*, 142:419–421, July 1965.

[52] D. J. Fixsen. The Temperature of the Cosmic Microwave Background. *apj*, 707(2):916–920, December 2009.

[53] G. Efstathiou, J. R. Bond, and Simon D. M. White. COBE Background radiation anisotropies and large scale structure in the universe. *Mon. Not. Roy. Astron. Soc.*, 258:1–6, 1992.

[54] R. K. Sachs and A. M. Wolfe. Perturbations of a cosmological model and angular variations of the microwave background. *Astrophys. J.*, 147:73–90, 1967.

[55] Donald G. York et al. The Sloan Digital Sky Survey: Technical Summary. *Astron. J.*, 120:1579–1587, 2000.

[56] Daniel J. Eisenstein et al. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *Astrophys. J.*, 633:560–574, 2005.

[57] Alexia M. Lopez, Roger G. Clowes, and Gerard M. Williger. A Big Ring on the Sky. 2 2024.

[58] Alexia M. Lopez, Roger G. Clowes, and Gerard M. Williger. A Giant Arc on the Sky. *Mon. Not. Roy. Astron. Soc.*, 516(2):1557–1572, 2022.

[59] Istvan Horvath, Zsolt Bagoly, Jon Hakkila, and L. Viktor Toth. New data support the existence of the Hercules-Corona Borealis Great Wall. *Astron. Astrophys.*, 584:A48, 2015.

[60] Adam G. Riess, Stefano Casertano, Wenlong Yuan, J. Bradley Bowers, Lucas Macri, Joel C. Zinn, and Dan Scolnic. Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with ΛCDM. *Astrophys. J. Lett.*, 908(1):L6, 2021.

[61] Alan H. Guth. The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. *Phys. Rev. D*, 23:347–356, 1981.

[62] Alexei A. Starobinsky. A New Type of Isotropic Cosmological Models Without Singularity. *Phys. Lett. B*, 91:99–102, 1980.

[63] Lev Kofman, Andrei D. Linde, and Alexei A. Starobinsky. Towards the theory of reheating after inflation. *Phys. Rev. D*, 56:3258–3295, 1997.

[64] Viatcheslav F. Mukhanov and G. V. Chibisov. Quantum Fluctuations and a Nonsingular Universe. *JETP Lett.*, 33:532–535, 1981.

[65] Alexei A. Starobinsky. Dynamics of Phase Transition in the New Inflationary Universe Scenario and Generation of Perturbations. *Phys. Lett. B*, 117:175–178, 1982.

[66] Alan H. Guth and S. Y. Pi. Fluctuations in the New Inflationary Universe. *Phys. Rev. Lett.*, 49:1110–1113, 1982.

[67] James M. Bardeen, Paul J. Steinhardt, and Michael S. Turner. Spontaneous Creation of Almost Scale - Free Density Perturbations in an Inflationary Universe. *Phys. Rev. D*, 28:679, 1983.

[68] Joseph John Thomson. Cathode Rays. *Philosophical Magazine*, 44:293–316, 1897.

[69] Ernst Rutherford. The Scattering of α and β Particles by Matter and the Structure of the Atom. *Philosophical Magazine*, 21:669–688, 1911.

[70] Niels Bohr. On the Constitution of Atoms and Molecules. *Philosophical Magazine*, 26:1–25, 1913.

[71] Niels Bohr. The Spectra of Helium and Hydrogen. *Nature*, 92:231—-232, 1913.

[72] James Sir Chadwick. The existence of a neutron. *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences*, 136:692–708, 1932.

[73] Max Planck. Zur theorie des gesetzes der energieverteilung im normalspektrum. *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 2:237, 1900.

[74] A. Einstein. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 322(6):132–148, January 1905.

[75] Arthur H. Compton. A quantum theory of the scattering of x-rays by light elements. *Phys. Rev.*, 21:483–502, May 1923.

[76] Hideki Yukawa. On the interaction of elementary particles. *Proc. Phys. Math. Soc. Japan*, 17:48–57, 1935.

[77] P.A.M. Dirac. The quantum theory of the electron. *Proc. R. Soc. Lond. A*, 117:610—-624, 1918.

[78] Carl Anderson. The positron. *Nature*, 133:313—-316, 1934.

[79] Murray Gell-Mann. The Eightfold Way: A Theory of strong interaction symmetry. 3 1961.

[80] Murray Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8:214–215, 1964.

[81] G. Zweig. An SU(3) model for strong interaction symmetry and its breaking. Version 1. 1964.

[82] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

[83] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968.

[84] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[85] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV. *Phys. Lett. B*, 122:103–116, 1983.

[86] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider. *Phys. Lett. B*, 126:398–410, 1983.

[87] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys. Lett. B*, 122:476–485, 1983.

[88] P. Bagnaia et al. Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider. *Phys. Lett. B*, 129:130–140, 1983.

[89] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[90] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[91] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.

[92] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.

[93] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[94] Ta-Pei Cheng and Ling-Fong Li. *Gauge Theory of Elementary Particle Physics*. Oxford University Press, Oxford, UK, 1984.

[95] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3 2014.

[96] Y. Aharonov and D. Bohm. Significance of electromagnetic potentials in the quantum theory. *Phys. Rev.*, 115:485–491, Aug 1959.

[97] Steven Weinberg. *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, 6 2005.

[98] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[99] Léo Morel, Zhibin Yao, Pierre Cladé, and Saïda Guellati-Khélifa. Determination of the fine-structure constant with an accuracy of 81 parts per trillion. *Nature*, 588(7836):61–65, 2020.

[100] Jonathan L. Feng. Dark Matter Candidates from Particle Physics and Methods of Detection. *Ann. Rev. Astron. Astrophys.*, 48:495–545, 2010.

[101] Bernard Carr, Florian Kuhnel, and Marit Sandstad. Primordial Black Holes as Dark Matter. *Phys. Rev. D*, 94(8):083504, 2016.

[102] Thomas Buchert and Syksy Räsänen. Backreaction in late-time cosmology. *Ann. Rev. Nucl. Part. Sci.*, 62:57–79, 2012.

[103] Ulf H. Danielsson and Thomas Van Riet. What if string theory has no de Sitter vacua? *Int. J. Mod. Phys. D*, 27(12):1830007, 2018.

[104] C. Armendariz-Picon, Viatcheslav F. Mukhanov, and Paul J. Steinhardt. A Dynamical solution to the problem of a small cosmological constant and late time cosmic acceleration. *Phys. Rev. Lett.*, 85:4438–4441, 2000.

[105] A. D. Sakharov. Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967.

[106] R. Acciarri et al. Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE): Conceptual Design Report, Volume 2: The Physics Program for DUNE at LBNF. 12 2015.

[107] Julien Lesgourgues and Sergio Pastor. Massive neutrinos and cosmology. *Phys. Rept.*, 429:307–379, 2006.

[108] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. The Hierarchy problem and new dimensions at a millimeter. *Phys. Lett. B*, 429:263–272, 1998.

[109] Nathaniel Craig and Seth Koren. IR Dynamics from UV Divergences: UV/IR Mixing, NCFT, and the Hierarchy Problem. *JHEP*, 03:037, 2020.

[110] J.M. Pendlebury *et al.* Revised experimental upper limit on the electric dipole moment of the neutron. *Phys. Rev. D*, 92:092003, Nov 2015.

[111] R. D. Peccei and Helen R. Quinn. CP Conservation in the Presence of Instantons. *Phys. Rev. Lett.*, 38:1440–1443, 1977.

[112] Craig J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman. On the large-angle anomalies of the microwave sky. *Mon. Not. Roy. Astron. Soc.*, 367:79–102, 2006.

[113] Ruth Durrer and Andrii Neronov. Cosmological Magnetic Fields: Their Generation, Evolution and Observation. *Astron. Astrophys. Rev.*, 21:62, 2013.

[114] Alejandra Kandus, Kerstin E. Kunze, and Christos G. Tsagas. Primordial magnetogenesis. *Phys. Rept.*, 505:1–58, 2011.

[115] Eduardo Banados et al. An 800-million-solar-mass black hole in a significantly neutral Universe at redshift 7.5. *Nature*, 553(7689):473–476, 2018.

[116] F. Wang *et al.* A Luminous Quasar at Redshift 7.642. *apjl*, 907(1):L1, January 2021.

[117] Roberto Maiolino et al. A small and vigorous black hole in the early Universe. *Nature*, 627(8002):59–63, 2024.

[118] Marta Volonteri, Mélanie Habouzit, and Monica Colpi. The origins of massive black holes. *Nature Reviews Physics*, 3(11):732–743, September 2021.

[119] Warren Massonneau, Marta Volonteri, Yohan Dubois, and Ricarda S. Beckmann. How the super-Eddington regime regulates black hole growth in high-redshift galaxies. *Astron. Astrophys.*, 670:A180, 2023.

[120] Masahiro Kawasaki, Alexander Kusenko, and Tsutomu T. Yanagida. Primordial seeds of supermassive black holes. *Phys. Lett. B*, 711:1–5, 2012.

[121] B. P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102, 2016.

[122] Gabriella Agazie et al. The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background. *Astrophys. J. Lett.*, 951(1):L8, 2023.

[123] J. Antoniadis et al. The second data release from the European Pulsar Timing Array - III. Search for gravitational wave signals. *Astron. Astrophys.*, 678:A50, 2023.

[124] Daniel J. Reardon et al. Search for an Isotropic Gravitational-wave Background with the Parkes Pulsar Timing Array. *Astrophys. J. Lett.*, 951(1):L6, 2023.

[125] Heng Xu et al. Searching for the Nano-Hertz Stochastic Gravitational Wave Background with the Chinese Pulsar Timing Array Data Release I. *Res. Astron. Astrophys.*, 23(7):075024, 2023.

[126] Bernard Carr and Florian Kuhnel. Primordial black holes as dark matter candidates. *SciPost Phys. Lect. Notes*, 48:1, 2022.

[127] Albert Escrivà, Florian Kuhnel, and Yuichiro Tada. Primordial Black Holes. 11 2022.

[128] S. W. Hawking, I. G. Moss, and J. M. Stewart. Bubble collisions in the very early universe. *Phys. Rev. D*, 26:2681–2693, Nov 1982.

[129] Alexander Polnarev and Robert Zembowicz. Formation of Primordial Black Holes by Cosmic Strings. *Phys. Rev. D*, 43:1106–1109, 1991.

[130] Rohan P Naidu, Pascal A Oesch, Pieter van Dokkum, Erica J Nelson, Katherine A Suess, Gabriel Brammer, Katherine E Whitaker, Garth Illingworth, Rychard Bouwens, Sandro Tacchella, et al. Two remarkably luminous galaxy candidates at $z \approx 10 - -12$ revealed by jwst. *The Astrophysical Journal Letters*, 940(1):L14, 2022.

[131] NJ Adams, CJ Conselice, L Ferreira, D Austin, JAA Trussler, I Juodžbalis, SM Wilkins, J Caruana, P Dayal, A Verma, et al. Discovery and properties of ultra-high redshift galaxies ($9 < z < 12$) in the jwst ero smacs 0723 field. *Monthly Notices of the Royal Astronomical Society*, 518(3):4755–4766, 2023.

[132] Ivo Labbe, Pieter van Dokkum, Erica Nelson, Rachel Bezanson, Katherine Suess, Joel Leja, Gabriel Brammer, Katherine Whitaker, Elijah Mathews, and Mauro Stefanon. A very early onset of massive galaxy formation. *arXiv preprint arXiv:2207.12446*, 2022.

[133] Steven L Finkelstein, Micaela B Bagley, Henry C Ferguson, Stephen M Wilkins, Jeyhan S Kartaltepe, Casey Papovich, LY Aaron Yung, Pablo Arrabal Haro, Peter Behroozi, Mark Dickinson, et al. Ceers key paper. i. an early look into the first 500 myr of galaxy formation with jwst. *The Astrophysical Journal Letters*, 946(1):L13, 2023.

[134] James M. Bardeen, J. R. Bond, Nick Kaiser, and A. S. Szalay. The Statistics of Peaks of Gaussian Random Fields. *Astrophys. J.*, 304:15–61, 1986.

[135] Anne M. Green, Andrew R. Liddle, Karim A. Malik, and Misao Sasaki. A New calculation of the mass fraction of primordial black holes. *Phys. Rev. D*, 70:041502, 2004.

[136] Sam Young and Marcello Musso. Application of peaks theory to the abundance of primordial black holes. *JCAP*, 11:022, 2020.

[137] Chul-Moon Yoo, Tomohiro Harada, Shin'ichi Hirano, and Kazunori Kohri. Abundance of Primordial Black Holes in Peak Theory for an Arbitrary Power Spectrum. *PTEP*, 2021(1):013E02, 2021.

[138] William H. Press and Paul Schechter. Formation of galaxies and clusters of galaxies by selfsimilar gravitational condensation. *Astrophys. J.*, 187:425–438, 1974.

[139] Bernard J. Carr. The Primordial black hole mass spectrum. *Astrophys. J.*, 201:1–19, 1975.

[140] Tomohiro Harada, Chul-Moon Yoo, and Kazunori Kohri. Threshold of primordial black hole formation. *Phys. Rev. D*, 88(8):084051, 2013. [Erratum: Phys.Rev.D 89, 029903 (2014)].

[141] Masaru Shibata and Misao Sasaki. Black hole formation in the Friedmann universe: Formulation and computation in numerical relativity. *Phys. Rev. D*, 60:084002, 1999.

[142] Albert Escrivà, Cristiano Germani, and Ravi K. Sheth. Universal threshold for primordial black hole formation. *Phys. Rev. D*, 101(4):044022, 2020.

[143] Vicente Atal, Judith Cid, Albert Escrivà, and Jaume Garriga. PBH in single field inflation: the effect of shape dispersion and non-Gaussianities. *JCAP*, 05:022, 2020.

[144] Vicente Atal, Jaume Garriga, and Airam Marcos-Caballero. Primordial black hole formation with non-Gaussian curvature perturbations. *JCAP*, 09:073, 2019.

[145] Alex Kehagias, Ilia Musco, and Antonio Riotto. Non-Gaussian Formation of Primordial Black Holes: Effects on the Threshold. *JCAP*, 12:029, 2019.

[146] Giacomo Ferrante, Gabriele Franciolini, Antonio Iovino, Junior., and Alfredo Urbano. Primordial non-Gaussianity up to all orders: Theoretical aspects and implications for primordial black hole models. *Phys. Rev. D*, 107(4):043520, 2023.

[147] Valerio De Luca, Alex Kehagias, and Antonio Riotto. How Well Do We Know the Primordial Black Hole Abundance? The Crucial Role of Non-Linearities when Approaching the Horizon. 7 2023.

[148] Sam Young, Christian T. Byrnes, and Misao Sasaki. Calculating the mass fraction of primordial black holes. *JCAP*, 07:045, 2014.

[149] Kenta Ando, Keisuke Inomata, and Masahiro Kawasaki. Primordial black holes and uncertainties in the choice of the window function. *Phys. Rev. D*, 97(10):103528, 2018.

[150] Sam Young. The primordial black hole formation criterion re-examined: Parametrisation, timing and the choice of window function. *Int. J. Mod. Phys. D*, 29(02):2030002, 2019.

[151] Christian T. Byrnes and Philippa S. Cole. Lecture notes on inflation and primordial black holes. 12 2021.

[152] Gabriele Franciolini. *Primordial Black Holes: from Theory to Gravitational Wave Observations*. PhD thesis, Geneva U., Dept. Theor. Phys., 2021.

[153] Ogan Özsoy and Gianmassimo Tasinato. Inflation and Primordial Black Holes. *Universe*, 9(5):203, 2023.

[154] Daniel Baumann. Inflation. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small*, pages 523–686, 2011.

[155] Misao Sasaki. Large Scale Quantum Fluctuations in the Inflationary Universe. *Prog. Theor. Phys.*, 76:1036, 1986.

[156] Viatcheslav F. Mukhanov. Quantum Theory of Gauge Invariant Cosmological Perturbations. *Sov. Phys. JETP*, 67:1297–1302, 1988.

[157] Hayato Motohashi, Alexei A. Starobinsky, and Jun'ichi Yokoyama. Inflation with a constant rate of roll. *JCAP*, 09:018, 2015.

[158] T. S. Bunch and P. C. W. Davies. Quantum Field Theory in de Sitter Space: Renormalization by Point Splitting. *Proc. Roy. Soc. Lond. A*, 360:117–134, 1978.

[159] Shamit Kachru, Renata Kallosh, Andrei D. Linde, Juan Martin Maldacena, Liam P. McAllister, and Sandip P. Trivedi. Towards inflation in string theory. *JCAP*, 10:013, 2003.

[160] Juan Garcia-Bellido and Ester Ruiz Morales. Primordial black holes from single field models of inflation. *Phys. Dark Univ.*, 18:47–54, 2017.

[161] Cristiano Germani and Tomislav Prokopec. On primordial black holes from an inflection point. *Phys. Dark Univ.*, 18:6–10, 2017.

[162] Guillermo Ballesteros and Marco Taoso. Primordial black hole dark matter from single field inflation. *Phys. Rev. D*, 97(2):023501, 2018.

[163] Mark P. Hertzberg and Masaki Yamada. Primordial Black Holes from Polynomial Potentials in Single Field Inflation. *Phys. Rev. D*, 97(8):083509, 2018.

[164] Swagat S. Mishra and Varun Sahni. Primordial Black Holes from a tiny bump/dip in the Inflaton potential. *JCAP*, 04:007, 2020.

[165] Keisuke Inomata, Evan McDonough, and Wayne Hu. Amplification of primordial perturbations from the rise or fall of the inflaton. *JCAP*, 02(02):031, 2022.

[166] S. W. Hawking. Black hole explosions. *Nature*, 248:30–31, 1974.

[167] S. W. Hawking. Particle Creation by Black Holes. *Commun. Math. Phys.*, 43:199–220, 1975. [Erratum: Commun.Math.Phys. 46, 206 (1976)].

[168] Don N. Page. Particle Emission Rates from a Black Hole. 2. Massless Particles from a Rotating Hole. *Phys. Rev. D*, 14:3260–3273, 1976.

[169] A. D. Dolgov, P. D. Naselsky, and I. D. Novikov. Gravitational waves, baryogenesis, and dark matter from primordial black holes. 9 2000.

[170] Richard Anantua, Richard Easther, and John T. Giblin. GUT-Scale Primordial Black Holes: Consequences and Constraints. *Phys. Rev. Lett.*, 103:111303, 2009.

[171] Alexander D. Dolgov and Damian Ejlli. Relic gravitational waves from light primordial black holes. *Phys. Rev. D*, 84:024028, 2011.

[172] Ruifeng Dong, William H Kinney, and Dejan Stojkovic. Gravitational wave production by Hawking radiation from rotating primordial black holes. *JCAP*, 10:034, 2016.

[173] Dan Hooper, Gordan Krnjaic, John March-Russell, Samuel D. McDermott, and Rudin Petrossian-Byrne. Hot Gravitons and Gravitational Waves From Kerr Black Holes in the Early Universe. 4 2020.

[174] Alexandre Arbey, Jérémy Auffinger, Pearl Sandick, Barmak Shams Es Haghi, and Kuver Sinha. Precision calculation of dark radiation from spinning primordial black holes and early matter-dominated eras. *Phys. Rev. D*, 103:123549, 2021.

[175] Andrew Cheek, Lucien Heurtier, Yuber F. Perez-Gonzalez, and Jessica Turner. Redshift effects in particle production from Kerr primordial black holes. *Phys. Rev. D*, 106(10):103012, 2022.

[176] Andrew Cheek, Lucien Heurtier, Yuber F. Perez-Gonzalez, and Jessica Turner. Evaporation of Primordial Black Holes in the Early Universe: Mass and Spin Distributions. 12 2022.

[177] Don N. Page. Particle emission rates from a black hole: Massless particles from an uncharged, nonrotating hole. *Phys. Rev. D*, 13:198–206, Jan 1976.

[178] Roy P. Kerr. Gravitational field of a spinning mass as an example of algebraically special metrics. *Phys. Rev. Lett.*, 11:237–238, Sep 1963.

[179] S. W. Hawking. Particle Creation by Black Holes. *Commun. Math. Phys.*, 43:199–220, 1975. [Erratum: Commun.Math.Phys. 46, 206 (1976)].

[180] Don N. Page. Particle emission rates from a black hole. ii. massless particles from a rotating hole. *Phys. Rev. D*, 14:3260–3273, Dec 1976.

[181] Saul A. Teukolsky. Rotating black holes: Separable wave equations for gravitational and electromagnetic perturbations. *Phys. Rev. Lett.*, 29:1114–1118, Oct 1972.

[182] S. A. Teukolsky and W. H. Press. Perturbations of a rotating black hole. III - Interaction of the hole with gravitational and electromagnet ic radiation. *Astrophys. J.*, 193:443–461, 1974.

[183] S. Chandrasekhar and Steven L. Detweiler. Equations governing axisymmetric perturbations of the Kerr black-hole. *Proc. Roy. Soc. Lond. A*, 345:145–167, 1975.

[184] S. Chandrasekhar. On a Transformation of Teukolsky's Equation and the Electromagnetic Perturbations of the Kerr Black Hole. *Proc. Roy. Soc. Lond. A*, 348:39–55, 1976.

[185] S. Chandrasekhar and Steven L. Detweiler. Equations governing gravitational perturbations of the Kerr black-hole. *Proc. Roy. Soc. Lond. A*, 350:165–174, 1976.

[186] Subrahmanyan Chandrasekhar and Steven L. Detweiler. On the Reflection and Transmission of Neutrino Waves by a Kerr Black Hole. *Proc. Roy. Soc. Lond. A*, 352:325–338, 1977.

[187] Alexei A. Starobinsky and S. M. Churilov. Amplification of electromagnetic and gravitational waves scattered by a rotating "black hole". *Sov. Phys. JETP*, 65(1):1–5, 1974.

[188] Alexandre Arbey and Jérémy Auffinger. BlackHawk: A public code for calculating the Hawking evaporation spectra of any black hole distribution. *Eur. Phys. J. C*, 79(8):693, 2019.

[189] Alexandre Arbey and Jérémy Auffinger. Physics Beyond the Standard Model with BlackHawk v2.0. *Eur. Phys. J. C*, 81:910, 2021.

[190] Tomohiro Harada, Chul-Moon Yoo, Kazunori Kohri, Ken-ichi Nakao, and Sanjay Jhingan. Primordial black hole formation in the matter-dominated phase of the Universe. *Astrophys. J.*, 833(1):61, 2016.

[191] Tomohiro Harada, Chul-Moon Yoo, Kazunori Kohri, and Ken-Ichi Nakao. Spins of primordial black holes formed in the matter-dominated phase of the universe. *Phys. Rev. D*, 96:083517, Oct 2017.

[192] Guillem Domènech. Induced gravitational waves in a general cosmological background. *Int. J. Mod. Phys. D*, 29(03):2050028, 2020.

[193] Guillem Domènech, Shi Pi, and Misao Sasaki. Induced gravitational waves as a probe of thermal history of the universe. *JCAP*, 08:017, 2020.

[194] Theodoros Papanikolaou. Gravitational waves induced from primordial black hole fluctuations: the effect of an extended mass function. *JCAP*, 10:089, 2022.

[195] Guillem Domènech. Scalar Induced Gravitational Waves Review. *Universe*, 7(11):398, 2021.

[196] Theodoros Papanikolaou. Toward the primordial black hole formation threshold in a time-dependent equation-of-state background. *Phys. Rev. D*, 105(12):124055, 2022.

[197] Francesco D'Eramo, Nicolas Fernandez, and Stefano Profumo. When the Universe Expands Too Fast: Relentless Dark Matter. *JCAP*, 05:012, 2017.

[198] Francesco D'Eramo, Nicolas Fernandez, and Stefano Profumo. Dark matter freeze-in production in fast-expanding universes. *Journal of Cosmology and Astroparticle Physics*, 2018:046 – 046, 2017.

[199] C. Armendariz-Picon, T. Damour, and Viatcheslav F. Mukhanov. k - inflation. *Phys. Lett. B*, 458:209–218, 1999.

[200] Adam J. Christopherson and Karim A. Malik. The non-adiabatic pressure in general scalar field systems. *Phys. Lett. B*, 675:159–163, 2009.

[201] Justin Khoury, Burt A. Ovrut, Paul J. Steinhardt, and Neil Turok. The Ekpyrotic universe: Colliding branes and the origin of the hot big bang. *Phys. Rev. D*, 64:123522, 2001.

[202] Kiwoon Choi. String or M theory axion as a quintessence. *Phys. Rev. D*, 62:043509, 2000.

[203] Carl L. Gardner. Quintessence and the transition to an accelerating universe. *Nucl. Phys. B*, 707:278–300, 2005.

[204] Sukannya Bhattacharya, Subhendra Mohanty, and Priyank Parashari. Primordial black holes and gravitational waves in nonstandard cosmologies. *Phys. Rev. D*, 102(4):043522, 2020.

[205] Sukannya Bhattacharya, Anirban Das, and Koushik Dutta. Solar mass primordial black holes in moduli dominated universe. *JCAP*, 10:071, 2021.

[206] Theodoros Papanikolaou. Primordial black holes in loop quantum gravity: The effect on the threshold. 1 2023.

[207] Andreas Ringwald and Carlos Tamarit. Revealing the cosmic history with gravitational waves. *Phys. Rev. D*, 106(6):063027, 2022.

[208] M. E. Gertsenshtein and V. I. Pustovoit. On the Detection of Low Frequency Gravitational Waves. *Sov. Phys. JETP*, 16:433, 1962.

[209] K. Ackley et al. Neutron Star Extreme Matter Observatory: A kilohertz-band gravitational-wave detector in the global network. *Publ. Astron. Soc. Austral.*, 37:e047, 2020.

[210] Fangyu Li, Nan Yang, Zhenyun Fang, Robert M. L. Baker, Jr., Gary V. Stephenson, and Hao Wen. Signal Photon Flux and Background Noise in a Coupling Electromagnetic Detecting System for High Frequency Gravitational Waves. *Phys. Rev. D*, 80:064013, 2009.

[211] Asuka Ito, Tomonori Ikeda, Kentaro Miuchi, and Jiro Soda. Probing GHz gravitational waves with graviton–magnon resonance. *Eur. Phys. J. C*, 80(3):179, 2020.

[212] Andreas Ringwald, Jan Schütte-Engel, and Carlos Tamarit. Gravitational Waves as a Big Bang Thermometer. *JCAP*, 03:054, 2021.

[213] Aldo Ejlli, Damian Ejlli, Adrian Mike Cruise, Giampaolo Pisano, and Hartmut Grote. Upper limits on the amplitude of ultra-high-frequency gravitational waves from graviton to photon conversion. *Eur. Phys. J. C*, 79(12):1032, 2019.

[214] J. Beacham et al. Physics Beyond Colliders at CERN: Beyond the Standard Model Working Group Report. *J. Phys. G*, 47(1):010501, 2020.

[215] R. Ballou et al. New exclusion limits on scalar and pseudoscalar axionlike particles from light shining through a wall. *Phys. Rev. D*, 92(9):092002, 2015.

[216] V. Anastassopoulos et al. New CAST Limit on the Axion-Photon Interaction. *Nature Phys.*, 13:584–590, 2017.

[217] Philip C. Argyres, Savas Dimopoulos, and John March-Russell. Black holes and submillimeter dimensions. *Phys. Lett. B*, 441:96–104, 1998.

[218] Nima Arkani-Hamed, Savas Dimopoulos, and John March-Russell. Stabilization of submillimeter dimensions: The New guise of the hierarchy problem. *Phys. Rev. D*, 63:064020, 2001.

[219] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. The Hierarchy problem and new dimensions at a millimeter. *Phys. Lett. B*, 429:263–272, 1998.

[220] R. Rattazzi. Cargese lectures on extra-dimensions. In *Cargese School of Particle Physics and Cosmology: the Interface*, pages 461–517, 8 2003.

[221] Ruifeng Dong, William H. Kinney, and Dejan Stojkovic. Gravitational wave production by hawking radiation from rotating primordial black holes. *Journal of Cosmology and Astroparticle Physics*, 2016(10):034–034, October 2016.

[222] C. M. Harris and P. Kanti. Hawking radiation from a (4+n)-dimensional rotating black hole. *Phys. Lett. B*, 633:106–110, 2006.

[223] Robert C. Myers and M. J. Perry. Black Holes in Higher Dimensional Space-Times. *Annals Phys.*, 172:304, 1986.

[224] Wen-Hai Tan, Shan-Qing Yang, Cheng-Gang Shao, Jia Li, An-Bin Du, Bi-Fu Zhan, Qing-Lan Wang, Peng-Shun Luo, Liang-Cheng Tu, and Jun Luo. New test of the gravitational inverse-square law at the submillimeter range with dual modulation and compensation. *Phys. Rev. Lett.*, 116:131101, Mar 2016.

[225] Christoph Hanhart, José A. Pons, Daniel R. Phillips, and Sanjay Reddy. The likelihood of gods' existence: improving the sn 1987a constraint on the size of large compact dimensions. *Physics Letters B*, 509(1–2):1–9, June 2001.

[226] Lawrence J. Hall and David Smith. Cosmological constraints on theories with large extra dimensions. *Physical Review D*, 60(8), September 1999.

[227] Nima Arkani-Hamed, Savas Dimopoulos, and Gia Dvali. Phenomenology, astrophysics, and cosmology of theories with submillimeter dimensions and tev scale quantum gravity. *Physical Review D*, 59(8), March 1999.

[228] Georges Aad et al. Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector. *Phys. Rev. D*, 103(11):112006, 2021.

[229] Armen Tumasyan et al. Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:153, 2021.

[230] Philip C. Argyres, Savas Dimopoulos, and John March-Russell. Black holes and submillimeter dimensions. *Phys. Lett. B*, 441:96–104, 1998.

[231] John A. Conley and Tommer Wizansky. Microscopic Primordial Black Holes and Extra Dimensions. *Phys. Rev. D*, 75:044006, 2007.

[232] Avi Friedlander, Katherine J. Mack, Sarah Schon, Ningqiang Song, and Aaron C. Vincent. Primordial black hole dark matter in the context of extra dimensions. *Phys. Rev. D*, 105(10):103508, 2022.

[233] Steven S. Gubser, Igor R. Klebanov, and Arkady A. Tseytlin. String theory and classical absorption by three-branes. *Nucl. Phys. B*, 499:217–240, 1997.

[234] Ezra Newman and Roger Penrose. An Approach to gravitational radiation by a method of spin coefficients. *J. Math. Phys.*, 3:566–578, 1962.

[235] Panagiota Kanti and John March-Russell. Calculable corrections to brane black hole decay. I. The scalar case. *Phys. Rev. D*, 66:024023, 2002.

[236] Panagiota Kanti and John March-Russell. Calculable corrections to brane black hole decay. II. Greybody factors for spin 1/2 and 1. *Phys. Rev. D*, 67:104019, May 2003.

[237] Panagiota Kanti. Black Holes in Theories with Large Extra Dimensions: A Review. *International Journal of Modern Physics A*, 19:4899–4951, 2004.

[238] Saul A. Teukolsky. Perturbations of a rotating black hole. 1. Fundamental equations for gravitational electromagnetic and neutrino field perturbations. *Astrophys. J.*, 185:635–647, 1973.

[239] Mirjam Cvetic and Finn Larsen. Greybody factors for black holes in four-dimensions: Particles with spin. *Phys. Rev. D*, 57:6297–6310, 1998.

[240] S. Creek, O. Efthimiou, P. Kanti, and K. Tamvakis. Graviton emission in the bulk from a higher-dimensional Schwarzschild black hole. *Phys. Lett. B*, 635:39–49, 2006.

[241] Marcela Carena, Zhen Liu, and Yikun Wang. Electroweak phase transition with spontaneous Z_2-breaking. *JHEP*, 08:107, 2020.

[242] Pau Amaro-Seoane et al. Laser Interferometer Space Antenna. 2 2017.

[243] Seiji Kawamura et al. Current status of space gravitational wave antenna DECIGO and B-DECIGO. *PTEP*, 2021(5):05A105, 2021.

[244] M. Punturo et al. The Einstein Telescope: A third-generation gravitational wave observatory. *Class. Quant. Grav.*, 27:194002, 2010.

[245] Djuna Croon, Oliver Gould, Philipp Schicho, Tuomas V. I. Tenkanen, and Graham White. Theoretical uncertainties for cosmological first-order phase transitions. *JHEP*, 04:055, 2021.

[246] Mark B. Hindmarsh, Marvin Lüben, Johannes Lumma, and Martin Pauly. Phase transitions in the early universe. *SciPost Phys. Lect. Notes*, 24:1, 2021.

[247] Peter Athron, Csaba Balázs, Andrew Fowlie, Lachlan Morris, and Lei Wu. Cosmological phase transitions: From perturbative particle physics to gravitational waves. *Prog. Part. Nucl. Phys.*, 135:104094, 2024.

[248] Guy D. Moore and Tomislav Prokopec. How fast can the wall move? A Study of the electroweak phase transition dynamics. *Phys. Rev. D*, 52:7182–7204, 1995.

[249] Alan H. Guth and Erick J. Weinberg. Cosmological Consequences of a First Order Phase Transition in the SU(5) Grand Unified Model. *Phys. Rev. D*, 23:876, 1981.

[250] Andrei D. Linde. Decay of the False Vacuum at Finite Temperature. *Nucl. Phys. B*, 216:421, 1983. [Erratum: Nucl.Phys.B 223, 544 (1983)].

[251] Jose R. Espinosa, Thomas Konstandin, Jose M. No, and Geraldine Servant. Energy Budget of Cosmological First-order Phase Transitions. *JCAP*, 06:028, 2010.

[252] Mark Hindmarsh and Mulham Hijazi. Gravitational waves from first order cosmological phase transitions in the Sound Shell Model. *JCAP*, 12:062, 2019.

[253] Felix Giese, Thomas Konstandin, and Jorinde van de Vis. Model-independent energy budget of cosmological first-order phase transitions—A sound argument to go beyond the bag model. *JCAP*, 07(07):057, 2020.

[254] Mark Hindmarsh, Stephan J. Huber, Kari Rummukainen, and David J. Weir. Shape of the acoustic gravitational wave power spectrum from a first order phase transition. *Phys. Rev. D*, 96(10):103520, 2017. [Erratum: Phys.Rev.D 101, 089902 (2020)].

[255] Marc Kamionkowski, Arthur Kosowsky, and Michael S. Turner. Gravitational radiation from first order phase transitions. *Phys. Rev. D*, 49:2837–2851, 1994.

[256] Stefan Höche, Jonathan Kozaczuk, Andrew J. Long, Jessica Turner, and Yikun Wang. Towards an all-orders calculation of the electroweak bubble wall velocity. *JCAP*, 03:009, 2021.

[257] Benoit Laurent and James M. Cline. First principles determination of bubble wall velocity. *Phys. Rev. D*, 106(2):023501, 2022.

[258] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.

[259] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.

[260] Andreas Papaefstathiou and Graham White. The electro-weak phase transition at colliders: confronting theoretical uncertainties and complementary channels. *JHEP*, 05:099, 2021.

[261] Michael J. Ramsey-Musolf. The electroweak phase transition: a collider target. *JHEP*, 09:179, 2020.

[262] David Curtin, Patrick Meade, and Chiu-Tien Yu. Testing Electroweak Baryogenesis with Future Colliders. *JHEP*, 11:127, 2014.

[263] M Punturo *et al.* The einstein telescope: a third-generation gravitational wave observatory. *Classical and Quantum Gravity*, 27(19):194002, 9 2010.

[264] Kent Yagi and Naoki Seto. Detector configuration of DECIGO/BBO and identification of cosmological neutron-star binaries. *Phys. Rev. D*, 83:044011, 2011. [Erratum: Phys.Rev.D 95, 109901 (2017)].

[265] Chiara Caprini et al. Detecting gravitational waves from cosmological phase transitions with LISA: an update. *JCAP*, 03:024, 2020.

[266] Alberto Sesana et al. Unveiling the gravitational universe at μ-Hz frequencies. *Exper. Astron.*, 51(3):1333–1383, 2021.

[267] S. Hild et al. Sensitivity Studies for Third-Generation Gravitational Wave Observatories. *Class. Quant. Grav.*, 28:094013, 2011.

[268] Yousef Abou El-Neaj et al. AEDGE: Atomic Experiment for Dark Matter and Gravity Exploration in Space. *EPJ Quant. Technol.*, 7:6, 2020.

[269] Andrew G. Cohen, D. B. Kaplan, and A. E. Nelson. Progress in electroweak baryogenesis. *Ann. Rev. Nucl. Part. Sci.*, 43:27–70, 1993.

[270] Antonio Riotto and Mark Trodden. Recent progress in baryogenesis. *Ann. Rev. Nucl. Part. Sci.*, 49:35–75, 1999.

[271] David E. Morrissey and Michael J. Ramsey-Musolf. Electroweak baryogenesis. *New J. Phys.*, 14:125003, 2012.

[272] Carlos E. M. Wagner. Electroweak Baryogenesis and Higgs Physics. *LHEP*, 2023:466, 2023.

[273] Steven Weinberg. Gauge and Global Symmetries at High Temperature. *Phys. Rev. D*, 9:3357–3378, 1974.

[274] James M. Cline, Guy D. Moore, and Geraldine Servant. Was the electroweak phase transition preceded by a color broken phase? *Phys. Rev. D*, 60:105035, 1999.

[275] Sebastian Baum, Marcela Carena, Nausheen R. Shah, Carlos E. M. Wagner, and Yikun Wang. Nucleation is more than critical: A case study of the electroweak phase transition in the NMSSM. *JHEP*, 03:055, 2021.

[276] Thomas Biekötter, Sven Heinemeyer, José Miguel No, María Olalla Olea, and Georg Weiglein. Fate of electroweak symmetry in the early Universe: Non-restoration and trapped vacua in the N2HDM. *JCAP*, 06:018, 2021.

[277] Thomas Biekötter, Sven Heinemeyer, José Miguel No, María Olalla Olea-Romacho, and Georg Weiglein. The trap in the early Universe: impact on the interplay between gravitational waves and LHC physics in the 2HDM. *JCAP*, 03:031, 2023.

[278] Aurora Ireland and Seth Koren. Asymmetric Reheating via Inverse Symmetry Breaking. 11 2022.

[279] Jae Hyeok Chang, María Olalla Olea-Romacho, and Erwin H. Tanin. Electroweak asymmetric early Universe via a scalar condensate. *Phys. Rev. D*, 106(11):113003, 2022.

[280] Pedro Schwaller. Gravitational Waves from a Dark Phase Transition. *Phys. Rev. Lett.*, 115(18):181101, 2015.

[281] Iason Baldes and Camilo Garcia-Cely. Strong gravitational radiation from a simple dark matter model. *JHEP*, 05:190, 2019.

[282] Moritz Breitbach, Joachim Kopp, Eric Madge, Toby Opferkuch, and Pedro Schwaller. Dark, Cold, and Noisy: Constraining Secluded Hidden Sectors with Gravitational Waves. *JCAP*, 07:007, 2019.

[283] Djuna Croon, Verónica Sanz, and Graham White. Model Discrimination in Gravitational Wave spectra from Dark Phase Transitions. *JHEP*, 08:203, 2018.

[284] Eleanor Hall, Thomas Konstandin, Robert McGehee, Hitoshi Murayama, and Géraldine Servant. Baryogenesis From a Dark First-Order Phase Transition. *JHEP*, 04:042, 2020.

[285] Iason Baldes. Gravitational waves from the asymmetric-dark-matter generating phase transition. *JCAP*, 05:028, 2017.

[286] Michael Geller, Anson Hook, Raman Sundrum, and Yuhsin Tsai. Primordial Anisotropies in the Gravitational Wave Background from Cosmological Phase Transitions. *Phys. Rev. Lett.*, 121(20):201303, 2018.

[287] Djuna Croon, Alexander Kusenko, Anupam Mazumdar, and Graham White. Solitosynthesis and Gravitational Waves. *Phys. Rev. D*, 101(8):085010, 2020.

[288] Eleanor Hall, Thomas Konstandin, Robert McGehee, and Hitoshi Murayama. Asymmetric matter from a dark first-order phase transition. *Phys. Rev. D*, 107(5):055011, 2023.

[289] Wei Chao, Xiu-Fei Li, and Lei Wang. Filtered pseudo-scalar dark matter and gravitational waves from first order phase transition. *JCAP*, 06:038, 2021.

[290] James B. Dent, Bhaskar Dutta, Sumit Ghosh, Jason Kumar, and Jack Runburg. Sensitivity to dark sector scales from gravitational wave signatures. *JHEP*, 08:300, 2022.

[291] Fa Peng Huang and Xinmin Zhang. Probing the gauge symmetry breaking of the early universe in 3-3-1 models and beyond by gravitational waves. *Phys. Lett. B*, 788:288–294, 2019.

[292] Djuna Croon, Tomás E. Gonzalo, and Graham White. Gravitational Waves from a Pati-Salam Phase Transition. *JHEP*, 02:083, 2019.

[293] Katsuya Hashino, Mitsuru Kakizaki, Shinya Kanemura, Pyungwon Ko, and Toshinori Matsui. Gravitational waves from first order electroweak phase transition in models with the $\mathrm{U}(1)_X$ gauge symmetry. *JHEP*, 06:088, 2018.

[294] Vedran Brdar, Lukas Graf, Alexander J. Helmboldt, and Xun-Jie Xu. Gravitational Waves as a Probe of Left-Right Symmetry Breaking. *JCAP*, 12:027, 2019.

[295] Rajesh R. Parwani. Resummation in a hot scalar field theory. *Phys. Rev. D*, 45:4695–4705, 6 1992.

[296] Peter Arnold and Olivier Espinosa. Effective potential and first-order phase transitions: Beyond leading order. *Phys. Rev. D*, 47:3546–3579, 4 1993.

[297] C. Glenn Boyd, David E. Brahm, and Stephen D. H. Hsu. Resummation methods at finite temperature: The Tadpole way. *Phys. Rev. D*, 48:4963–4973, 1993.

[298] David Curtin, Patrick Meade, and Harikrishnan Ramani. Thermal Resummation and Phase Transitions. *Eur. Phys. J. C*, 78(9):787, 2018.

[299] David Curtin, Jyotirmoy Roy, and Graham White. Gravitational waves and tadpole resummation: Efficient and easy convergence of finite temperature QFT. 11 2022.

[300] K. Farakos, K. Kajantie, K. Rummukainen, and Mikhail E. Shaposhnikov. 3-D physics and the electroweak phase transition: Perturbation theory. *Nucl. Phys. B*, 425:67–109, 1994.

[301] K. Kajantie, M. Laine, K. Rummukainen, and Mikhail E. Shaposhnikov. Generic rules for high temperature dimensional reduction and their application to the standard model. *Nucl. Phys. B*, 458:90–136, 1996.

[302] Eric Braaten and Agustin Nieto. Effective field theory approach to high temperature thermodynamics. *Phys. Rev. D*, 51:6990–7006, 1995.

[303] Andreas Ekstedt, Philipp Schicho, and Tuomas V. I. Tenkanen. DRalgo: A package for effective field theory approach for thermal phase transitions. *Comput. Phys. Commun.*, 288:108725, 2023.

[304] J. I. Kapusta and Charles Gale. *Finite-temperature field theory: Principles and applications*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2011.

[305] Mikko Laine and Aleksi Vuorinen. *Basics of Thermal Field Theory*, volume 925. Springer, 2016.

[306] Andrei D. Linde. Infrared Problem in Thermodynamics of the Yang-Mills Gas. *Phys. Lett. B*, 96:289–292, 1980.

[307] K. Kajantie, M. Laine, K. Rummukainen, and Mikhail E. Shaposhnikov. Generic rules for high temperature dimensional reduction and their application to the standard model. *Nucl. Phys. B*, 458:90–136, 1996.

[308] Philipp M. Schicho, Tuomas V. I. Tenkanen, and Juuso Österman. Robust approach to thermal resummation: Standard Model meets a singlet. *JHEP*, 06:130, 2021.

[309] J. R. Espinosa, M. Quiros, and F. Zwirner. On the phase transition in the scalar theory. *Phys. Lett. B*, 291:115–124, 1992.

[310] J. R. Espinosa, M. Quiros, and F. Zwirner. On the nature of the electroweak phase transition. *Phys. Lett. B*, 314:206–216, 1993.

[311] Mariano Quiros. On daisy and superdaisy resummation of the effective potential at finite temperature. In *4th Hellenic School on Elementary Particle Physics*, pages 502–511, 9 1992.

[312] Michael Dine, Robert G. Leigh, Patrick Y. Huet, Andrei D. Linde, and Dmitri A. Linde. Towards the theory of the electroweak phase transition. *Phys. Rev. D*, 46:550–571, 1992.

[313] C. Glenn Boyd, David E. Brahm, and Stephen D. H. Hsu. Corrections to the electroweak effective action at finite temperature. *Phys. Rev. D*, 48:4952–4962, 1993.

[314] L. Dolan and R. Jackiw. Symmetry Behavior at Finite Temperature. *Phys. Rev. D*, 9:3320–3341, 1974.

[315] Cedric Delaunay, Christophe Grojean, and James D. Wells. Dynamics of Non-renormalizable Electroweak Symmetry Breaking. *JHEP*, 04:029, 2008.

[316] Erick J. Weinberg and Ai-qun Wu. UNDERSTANDING COMPLEX PERTURBATIVE EFFECTIVE POTENTIALS. *Phys. Rev. D*, 36:2474, 1987.

[317] Andreas Ekstedt and Johan Löfgren. The High-Temperature Expansion of the Thermal Sunset. 6 2020.

[318] James M. Cline, Kimmo Kainulainen, and Axel P. Vischer. Dynamics of two Higgs doublet CP violation and baryogenesis at the electroweak phase transition. *Phys. Rev. D*, 54:2451–2472, 1996.

[319] James M. Cline, Kimmo Kainulainen, and Michael Trott. Electroweak Baryogenesis in Two Higgs Doublet Models and B meson anomalies. *JHEP*, 11:089, 2011.

[320] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev. D*, 98(3):030001, 2018.

[321] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.

[322] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.

[323] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, 1964.

[324] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1–102, 2012.

[325] Howard E. Haber and Ralf Hempfling. The Renormalization group improved Higgs sector of the minimal supersymmetric model. *Phys. Rev. D*, 48:4280–4309, 1993.

[326] Apostolos Pilaftsis and Carlos E. M. Wagner. Higgs bosons in the minimal supersymmetric standard model with explicit CP violation. *Nucl. Phys. B*, 553:3–42, 1999.

[327] Marcela Carena, J. R. Espinosa, M. Quiros, and C. E. M. Wagner. Analytical expressions for radiatively corrected Higgs masses and couplings in the MSSM. *Phys. Lett. B*, 355:209–221, 1995.

[328] Gabriel Lee and Carlos E. M. Wagner. Higgs bosons in heavy supersymmetry with an intermediate m_A. *Phys. Rev. D*, 92(7):075032, 2015.

[329] M. Carena, J. Ellis, J. S. Lee, A. Pilaftsis, and C. E. M. Wagner. CP Violation in Heavy MSSM Higgs Scenarios. *JHEP*, 02:123, 2016.

[330] Henning Bahl and Wolfgang Hollik. Precise prediction of the MSSM Higgs boson masses for low M_A. *JHEP*, 07:182, 2018.

[331] Henning Bahl, Stefan Liebler, and Tim Stefaniak. MSSM Higgs benchmark scenarios for Run 2 and beyond: the low $\tan\beta$ region. *Eur. Phys. J. C*, 79(3):279, 2019.

[332] Nick Murphy and Heidi Rzehak. Higgs-Boson Masses and Mixings in the MSSM with CP Violation and Heavy SUSY Particles. 9 2019.

[333] Henning Bahl, Nick Murphy, and Heidi Rzehak. Hybrid calculation of the MSSM Higgs boson masses using the complex THDM as EFT. *Eur. Phys. J. C*, 81(2):128, 2021.

[334] Henning Bahl and Ivan Sobolev. Two-loop matching of renormalizable operators: general considerations and applications. *JHEP*, 03:286, 2021.

[335] Ian Low, Witold Skiba, and David Tucker-Smith. Little Higgses from an antisymmetric condensate. *Phys. Rev. D*, 66:072001, 2002.

[336] K. G. Klimenko. On Necessary and Sufficient Conditions for Some Higgs Potentials to Be Bounded From Below. *Theor. Math. Phys.*, 62:58–65, 1985.

[337] Shinya Kanemura, Takahiro Kubota, and Eiichi Takasugi. Lee-Quigg-Thacker bounds for Higgs boson masses in a two doublet model. *Phys. Lett. B*, 313:155–160, 1993.

[338] Andrew G. Akeroyd, Abdesslam Arhrib, and El-Mokhtar Naimi. Note on tree level unitarity in the general two Higgs doublet model. *Phys. Lett. B*, 490:119–124, 2000.

[339] P. M. Ferreira, R. Santos, and A. Barroso. Stability of the tree-level vacuum in two Higgs doublet models against charge or CP spontaneous violation. *Phys. Lett. B*, 603:219–229, 2004. [Erratum: Phys.Lett.B 629, 114–114 (2005)].

[340] J. Horejsi and M. Kladiva. Tree-unitarity bounds for THDM Higgs masses revisited. *Eur. Phys. J. C*, 46:81–91, 2006.

[341] I. F. Ginzburg and I. P. Ivanov. Tree-level unitarity constraints in the most general 2HDM. *Phys. Rev. D*, 72:115010, 2005.

[342] A. Barroso, P. M. Ferreira, and R. Santos. Charge and CP symmetry breaking in two Higgs doublet models. *Phys. Lett. B*, 632:684–687, 2006.

[343] I. P. Ivanov. Minkowski space structure of the Higgs potential in 2HDM. *Phys. Rev. D*, 75:035001, 2007. [Erratum: Phys.Rev.D 76, 039902 (2007)].

[344] M. Maniatis, A. von Manteuffel, O. Nachtmann, and F. Nagel. Stability and symmetry breaking in the general two-Higgs-doublet model. *Eur. Phys. J. C*, 48:805–823, 2006.

[345] A. Barroso, P. M. Ferreira, and R. Santos. Neutral minima in two-Higgs doublet models. *Phys. Lett. B*, 652:181–193, 2007.

[346] Igor P. Ivanov. Minkowski space structure of the Higgs potential in 2HDM. II. Minima, symmetries, and topology. *Phys. Rev. D*, 77:015017, 2008.

[347] P. M. Ferreira and D. R. T. Jones. Bounds on scalar masses in two Higgs doublet models. *JHEP*, 08:069, 2009.

[348] D. Jurčiukonis and L. Lavoura. The three- and four-Higgs couplings in the general two-Higgs-doublet model. *JHEP*, 12:004, 2018.

[349] Yisheng Song. Co-positivity of tensors and Stability conditions of two Higgs potentials. 3 2022.

[350] Antonio Pich and Paula Tuzon. Yukawa Alignment in the Two-Higgs-Doublet Model. *Phys. Rev. D*, 80:091702, 2009.

[351] Nina M. Coyle, Bing Li, and Carlos E. M. Wagner. Wrong sign bottom Yukawa coupling in low energy supersymmetry. *Phys. Rev. D*, 97(11):115028, 2018.

[352] Albert M Sirunyan et al. Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C*, 79(5):421, 2019.

[353] Georges Aad et al. Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment. *Phys. Rev. D*, 101(1):012002, 2020.

[354] John F. Gunion and Howard E. Haber. The CP conserving two Higgs doublet model: The Approach to the decoupling limit. *Phys. Rev. D*, 67:075019, 2003.

[355] Marcela Carena, Ian Low, Nausheen R. Shah, and Carlos E. M. Wagner. Impersonating the Standard Model Higgs Boson: Alignment without Decoupling. *JHEP*, 04:015, 2014.

[356] P. S. Bhupal Dev and Apostolos Pilaftsis. Maximally Symmetric Two Higgs Doublet Model with Natural Standard Model Alignment. *JHEP*, 12:024, 2014. [Erratum: JHEP 11, 147 (2015)].

[357] Marcela Carena, Howard E. Haber, Ian Low, Nausheen R. Shah, and Carlos E. M. Wagner. Complementarity between Nonstandard Higgs Boson Searches and Precision Higgs Boson Measurements in the MSSM. *Phys. Rev. D*, 91(3):035003, 2015.

[358] Marcela Carena, Howard E. Haber, Ian Low, Nausheen R. Shah, and Carlos E. M. Wagner. Alignment limit of the NMSSM Higgs sector. *Phys. Rev. D*, 93(3):035013, 2016.

[359] Jérémy Bernon, John F. Gunion, Howard E. Haber, Yun Jiang, and Sabine Kraml. Scrutinizing the alignment limit in two-Higgs-doublet models: m_h=125 GeV. *Phys. Rev. D*, 92(7):075004, 2015.

[360] Howard E. Haber, Venus Keus, and Rui Santos. P-even, CP-violating Signals in Scalar-Mediated Processes. 6 2022.

[361] Kenneth Ross Garren. *Bounds for the Eigenvalues of a Matrix*. PhD thesis, William and Mary, 1965.

[362] S. Gershgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Bulletin de l'Academie des Sciences de l'URSS*, 6(3):749–754, 1931.

[363] Miguel P. Bento, Jorge C. Romão, and João P. Silva. Unitarity bounds for all symmetry-constrained 3HDMs. 4 2022.

[364] Shinya Kanemura and Kei Yagyu. Unitarity bound in the most general two Higgs doublet model. *Phys. Lett. B*, 751:289–296, 2015.

[365] Gary Ulrich and Layne T. Watson. Positivity conditions for quartic polynomials. *SIAM Journal on Scientific Computing*, 15(3):528–544, 1994.

[366] A. Barroso, P. M. Ferreira, I. P. Ivanov, and Rui Santos. Metastability bounds on the two Higgs doublet model. *JHEP*, 06:045, 2013.

[367] I. P. Ivanov and Joao P. Silva. Tree-level metastability bounds for the most general two Higgs doublet model. *Phys. Rev. D*, 92(5):055017, 2015.

[368] L. Lavoura and Joao P. Silva. Fundamental CP violating quantities in a SU(2) x U(1) model with many Higgs doublets. *Phys. Rev. D*, 50:4619–4624, 1994.

[369] Sacha Davidson and Howard E. Haber. Basis-independent methods for the two-Higgs-doublet model. *Phys. Rev. D*, 72:035004, 2005. [Erratum: Phys.Rev.D 72, 099902 (2005)].

[370] Ian Low, Nausheen R. Shah, and Xiao-Ping Wang. Higgs alignment and novel CP-violating observables in two-Higgs-doublet models. *Phys. Rev. D*, 105(3):035009, 2022.

[371] Howard E. Haber and Oscar Stål. New LHC benchmarks for the $\mathcal{CP}$ -conserving two-Higgs-doublet model. *Eur. Phys. J. C*, 75(10):491, 2015. [Erratum: Eur.Phys.J.C 76, 312 (2016)].

[372] Rafael Boto, Tiago V. Fernandes, Howard E. Haber, Jorge C. Romão, and João P. Silva. Basis-independent treatment of the complex 2HDM. *Phys. Rev. D*, 101(5):055023, 2020.

[373] V. Andreev et al. Improved limit on the electric dipole moment of the electron. *Nature*, 562(7727):355–360, 2018.

[374] Wolfgang Altmannshofer, Stefania Gori, Nick Hamer, and Hiren H. Patel. Electron EDM in the complex two-Higgs doublet model. *Phys. Rev. D*, 102(11):115042, 2020.

[375] Jonathan L. Feng, Huitzu Tu, and Hai-Bo Yu. Thermal Relics in Hidden Sectors. *JCAP*, 10:043, 2008.

[376] Lotty Ackerman, Matthew R. Buckley, Sean M. Carroll, and Marc Kamionkowski. Dark Matter and Dark Radiation. *Phys. Rev. D*, 79:023519, 2009.

[377] R. Foot and S. Vagnozzi. Dissipative hidden sector dark matter. *Phys. Rev. D*, 91:023512, 2015.

[378] Matthew Reece and Thomas Roxlo. Nonthermal production of dark radiation and dark matter. *JHEP*, 09:096, 2016.

[379] Peter Adshead, Yanou Cui, and Jessie Shelton. Chilly Dark Sectors and Asymmetric Reheating. *JHEP*, 06:016, 2016.

[380] Edward Hardy and James Unwin. Symmetric and Asymmetric Reheating. *JHEP*, 09:113, 2017.

[381] Tommi Tenkanen and Ville Vaskonen. Reheating the Standard Model from a hidden sector. *Phys. Rev. D*, 94(8):083516, 2016.

[382] Peter Adshead, Pranjal Ralegankar, and Jessie Shelton. Reheating in two-sector cosmology. *JHEP*, 08:151, 2019.

[383] Kaloian D. Lozanov. Lectures on Reheating after Inflation. 7 2019.

[384] Pearl Sandick, Barmak Shams Es Haghi, and Kuver Sinha. Asymmetric reheating by primordial black holes. *Phys. Rev. D*, 104(8):083523, 2021.

[385] Wen Han Chiu, Sungwoo Hong, and Lian-Tao Wang. Conformal Freeze-In, Composite Dark Photon, and Asymmetric Reheating. 9 2022.

[386] T. D. Lee and Chen-Ning Yang. Question of Parity Conservation in Weak Interactions. *Phys. Rev.*, 104:254–258, 1956.

[387] I. Yu. Kobzarev, L. B. Okun, and I. Ya. Pomeranchuk. On the possibility of experimental observation of mirror particles. *Sov. J. Nucl. Phys.*, 3(6):837–841, 1966.

[388] Matej Pavsic. External inversion, internal inversion, and reflection invariance. *Int. J. Theor. Phys.*, 9:229–244, 1974.

[389] S. I. Blinnikov and M. Yu. Khlopov. ON POSSIBLE EFFECTS OF 'MIRROR' PARTICLES. *Sov. J. Nucl. Phys.*, 36:472, 1982.

[390] Robert Foot, H. Lew, and R. R. Volkas. A Model with fundamental improper space-time symmetries. *Phys. Lett. B*, 272:67–70, 1991.

[391] Robert Foot, H. Lew, and R. R. Volkas. Possible consequences of parity conservation. *Mod. Phys. Lett. A*, 7:2567–2574, 1992.

[392] Edward W. Kolb, D. Seckel, and Michael S. Turner. The Shadow World. *Nature*, 314:415–419, 1985.

[393] H. M. Hodges. Mirror baryons as the dark matter. *Phys. Rev. D*, 47:456–459, 1993.

[394] Lawrence M. Krauss, Alan H. Guth, David N. Spergel, George B. Field, and William H. Press. Inflation and Shadow Matter. *Nature*, 319:748–751, 1986.

[395] Z. G. Berezhiani, A. D. Dolgov, and R. N. Mohapatra. Asymmetric inflationary reheating and the nature of mirror universe. *Phys. Lett. B*, 375:26–36, 1996.

[396] L. B. Okun. Mirror particles and mirror matter: 50 years of speculations and search. *Phys. Usp.*, 50:380–389, 2007.

[397] Robert Foot, H. Lew, and R. R. Volkas. Unbroken versus broken mirror world: A Tale of two vacua. *JHEP*, 07:032, 2000.

[398] Zurab Berezhiani, Denis Comelli, and Francesco L. Villante. The Early mirror universe: Inflation, baryogenesis, nucleosynthesis and dark matter. *Phys. Lett. B*, 503:362–375, 2001.

[399] A. Yu. Ignatiev and R. R. Volkas. Discovering mirror particles at the large hadron collider and the implied cold universe. *Phys. Lett. B*, 487:294–298, 2000.

[400] Zurab Berezhiani. Mirror world and its cosmological consequences. *Int. J. Mod. Phys. A*, 19:3775–3806, 2004.

[401] Riccardo Barbieri, Thomas Gregoire, and Lawrence J. Hall. Mirror world at the large hadron collider. 9 2005.

[402] Riccardo Barbieri, Lawrence J. Hall, and Keisuke Harigaya. Minimal Mirror Twin Higgs. *JHEP*, 11:172, 2016.

[403] Nathaniel Craig, Seth Koren, and Timothy Trott. Cosmological Signals of a Mirror Twin Higgs. *JHEP*, 05:038, 2017.

[404] Zackaria Chacko, Nathaniel Craig, Patrick J. Fox, and Roni Harnik. Cosmology in Mirror Twin Higgs and Neutrino Masses. *JHEP*, 07:023, 2017.

[405] Csaba Csaki, Eric Kuflik, and Salvator Lombardo. Viable Twin Cosmology from Neutrino Mixing. *Phys. Rev. D*, 96(5):055013, 2017.

[406] Zackaria Chacko, David Curtin, Michael Geller, and Yuhsin Tsai. Cosmological Signatures of a Mirror Twin Higgs. *JHEP*, 09:163, 2018.

[407] David Curtin and Shayne Gryba. Twin Higgs portal dark matter. *JHEP*, 08:009, 2021.

[408] Hugues Beauchesne and Yevgeny Kats. Cosmology of the Twin Higgs without explicit $\mathbb{Z}_2$ breaking. *JHEP*, 12:160, 2021.

[409] James M. Cline and Jean-Samuel Roux. Asymmetric reheating from a symmetric inflationary potential. *Phys. Rev. D*, 105(4):043506, 2022.

[410] Kaladi S. Babu and Rabindra N. Mohapatra. Theoretical Constraints on Neutron-Mirror-Neutron Oscillation. *Symmetry*, 14(4):731, 2022.

[411] Robert Foot. Neutrino oscillations and the exact parity model. *Mod. Phys. Lett. A*, 9:169–180, 1994.

[412] Robert Foot and R. R. Volkas. Neutrino physics and the mirror world: How exact parity symmetry explains the solar neutrino deficit, the atmospheric neutrino anomaly and the LSND experiment. *Phys. Rev. D*, 52:6595–6606, 1995.

[413] Zurab G. Berezhiani and Rabindra N. Mohapatra. Reconciling present neutrino puzzles: Sterile neutrinos as mirror neutrinos. *Phys. Rev. D*, 52:6607–6611, 1995.

[414] Z. K. Silagadze. Neutrino mass and the mirror universe. *Phys. Atom. Nucl.*, 60:272–275, 1997.

[415] Robert Foot and R. R. Volkas. How neutrino oscillations can induce an effective neutrino number of less than three during big bang nucleosynthesis. *Phys. Rev. D*, 56:6653–6664, 1997. [Erratum: Phys.Rev.D 59, 029901 (1999)].

[416] M. Collie and Robert Foot. Neutrino masses in the SU(5) x SU(5)-prime mirror symmetric model. *Phys. Lett. B*, 432:134–138, 1998.

[417] Robert Foot and R. R. Volkas. Implications of mirror neutrinos for early universe cosmology. *Phys. Rev. D*, 61:043507, 2000.

[418] S. Blinnikov. Gamma-ray bursts produced by mirror stars. 2 1999.

[419] Robert Foot, A. Yu. Ignatiev, and R. R. Volkas. Physics of mirror photons. *Phys. Lett. B*, 503:355–361, 2001.

[420] Robert Foot and Sergei N. Gninenko. Can the mirror world explain the orthopositronium lifetime puzzle? *Phys. Lett. B*, 480:171–175, 2000.

[421] Robert Foot and S. Mitra. Ordinary atom mirror atom bound states: A New window on the mirror world. *Phys. Rev. D*, 66:061301, 2002.

[422] A. Badertscher et al. An Apparatus to search for mirror dark matter via the invisible decay of orthopositronium in vacuum. *Int. J. Mod. Phys. A*, 19:3833–3848, 2004.

[423] Zurab Berezhiani. Through the looking-glass: Alice's adventures in mirror world. In *From Fields to Strings: Circumnavigating Theoretical Physics: A Conference in Tribute to Ian Kogan*, pages 2147–2195, 2005.

[424] Zurab Berezhiani and Angela Lepidi. Cosmological bounds on the 'millicharges' of mirror particles. *Phys. Lett. B*, 681:276–281, 2009.

[425] R. Foot. Implications of mirror dark matter kinetic mixing for CMB anisotropies. *Phys. Lett. B*, 718:745–751, 2013.

[426] Zackaria Chacko, Can Kilic, Saereh Najjari, and Christopher B. Verhaaren. Collider signals of the Mirror Twin Higgs boson through the hypercharge portal. *Phys. Rev. D*, 100(3):035037, 2019.

[427] David Curtin and Jack Setford. How To Discover Mirror Stars. *Phys. Lett. B*, 804:135391, 2020.

[428] David Curtin and Jack Setford. Signatures of Mirror Stars. *JHEP*, 03:041, 2020.

[429] Seth Koren and Robert McGehee. Freezing-in twin dark matter. *Phys. Rev. D*, 101(5):055024, 2020.

[430] Aaron Howe, Jack Setford, David Curtin, and Christopher D. Matzner. How to search for mirror stars with Gaia. *JHEP*, 07:059, 2022.

[431] Abdaljalel Alizzi and Z. K. Silagadze. Dark photon portal into mirror world. *Mod. Phys. Lett. A*, 36(30):2150215, 2021.

[432] R. N. Mohapatra, S. Nasri, and S. Nussinov. Some implications of neutron mirror neutron oscillation. *Phys. Lett. B*, 627:124–130, 2005.

[433] Rabindra N. Mohapatra and Shmuel Nussinov. Constraints on Mirror Models of Dark Matter from Observable Neutron-Mirror Neutron Oscillation. *Phys. Lett. B*, 776:22–25, 2018.

[434] Zurab Berezhiani and Luis Bento. Neutron - mirror neutron oscillations: How fast might they be? *Phys. Rev. Lett.*, 96:081801, 2006.

[435] Zurab Berezhiani and Luis Bento. Fast neutron: Mirror neutron oscillation and ultra high energy cosmic rays. *Phys. Lett. B*, 635:253–259, 2006.

[436] A. P. Serebrov et al. Experimental search for neutron: Mirror neutron oscillations using storage of ultracold neutrons. *Phys. Lett. B*, 663:181–185, 2008.

[437] Fredrik Sandin and Paolo Ciarcelluti. Effects of mirror dark matter on neutron stars. *Astropart. Phys.*, 32:278–284, 2009.

[438] A. P. Serebrov et al. Search for neutron–mirror neutron oscillations in a laboratory experiment with ultracold neutrons. *Nucl. Instrum. Meth. A*, 611:137–140, 2009.

[439] Zurab Berezhiani. More about neutron - mirror neutron oscillation. *Eur. Phys. J. C*, 64:421–431, 2009.

[440] Zurab Berezhiani and Askhat Gazizov. Neutron Oscillations to Parallel World: Earlier End to the Cosmic Ray Spectrum? *Eur. Phys. J. C*, 72:2111, 2012.

[441] Z. Berezhiani and F. Nesti. Magnetic anomaly in UCN trapping: signal for neutron oscillations to parallel world? *Eur. Phys. J. C*, 72:1974, 2012.

[442] Zurab Berezhiani, Matthew Frost, Yuri Kamyshkov, Ben Rybolt, and Louis Varriano. Neutron Disappearance and Regeneration from Mirror State. *Phys. Rev. D*, 96(3):035039, 2017.

[443] Zurab Berezhiani. Neutron lifetime puzzle and neutron–mirror neutron oscillation. *Eur. Phys. J. C*, 79(6):484, 2019.

[444] Zurab Berezhiani. Neutron lifetime and dark decay of the neutron and hydrogen. *LHEP*, 2(1):118, 2019.

[445] Revaz Beradze, Merab Gogberashvili, and Alexander S. Sakharov. Binary Neutron Star Mergers with Missing Electromagnetic Counterparts as Manifestations of Mirror World. *Phys. Lett. B*, 804:135402, 2020.

[446] Zurab Berezhiani. A possible shortcut for neutron–antineutron oscillation through mirror world. *Eur. Phys. J. C*, 81(1):33, 2021.

[447] Yuri Kamyshkov, James Ternullo, Louis Varriano, and Zurab Berezhiani. Neutron-Mirror Neutron Oscillations in Absorbing Matter. *Symmetry*, 14(2):230, 2022.

[448] Lucas Johns and Seth Koren. Hydrogen Mixing as a Novel Mechanism for Colder Baryons in 21 cm Cosmology. 12 2020.

[449] Lucas Johns and Seth Koren. The Hydrogen Mixing Portal, Its Origins, and Its Cosmological Effects. 12 2020.

[450] Nathaniel Craig, Andrey Katz, Matt Strassler, and Raman Sundrum. Naturalness in the Dark at the LHC. *JHEP*, 07:105, 2015.

[451] Keisuke Harigaya, Robert Mcgehee, Hitoshi Murayama, and Katelin Schutz. A predictive mirror twin Higgs with small Z_2 breaking. *JHEP*, 05:155, 2020.

[452] Thomas G. Roos. Wilson renormalization group study of inverse symmetry breaking. *Phys. Rev. D*, 54:2944–2959, 1996.

[453] Jean Orloff. The UV price for symmetry nonrestoration. *Phys. Lett. B*, 403:309–315, 1997.

[454] Borut Bajc. High temperature symmetry nonrestoration. In *3rd International Conference on Particle Physics and the Early Universe*, pages 247–253, 2000.

[455] Marcus Benghi Pinto and Rudnei O. Ramos. Inverse symmetry breaking in multi-scalar field theories. *J. Phys. A*, 39:6649–6655, 2006.

[456] K. Jansen and M. Laine. Inverse symmetry breaking with 4-D lattice simulations. *Phys. Lett. B*, 435:166–174, 1998.

[457] G. Bimonte, D. Iniguez, A. Tarancon, and C. L. Ullod. Inverse symmetry breaking on the lattice: An Accurate MC study. *Nucl. Phys. B*, 559:103–122, 1999.

[458] Marcus B. Pinto and Rudnei O. Ramos. A Nonperturbative study of inverse symmetry breaking at high temperatures. *Phys. Rev. D*, 61:125016, 2000.

[459] John McDonald. Thermally generated gauge singlet scalars as selfinteracting dark matter. *Phys. Rev. Lett.*, 88:091304, 2002.

[460] Lawrence J. Hall, Karsten Jedamzik, John March-Russell, and Stephen M. West. Freeze-In Production of FIMP Dark Matter. *JHEP*, 03:080, 2010.

[461] Nicolás Bernal, Matti Heikinheimo, Tommi Tenkanen, Kimmo Tuominen, and Ville Vaskonen. The Dawn of FIMP Dark Matter: A Review of Models and Constraints. *Int. J. Mod. Phys. A*, 32(27):1730023, 2017.

[462] Mariano Quiros. Finite temperature field theory and phase transitions. In *ICTP Summer School in High-Energy Physics and Cosmology*, pages 187–259, 1 1999.

[463] Patrick Meade and Harikrishnan Ramani. Unrestored Electroweak Symmetry. *Phys. Rev. Lett.*, 122(4):041802, 2019.

[464] Alfredo Glioti, Riccardo Rattazzi, and Luca Vecchi. Electroweak Baryogenesis above the Electroweak Scale. *JHEP*, 04:027, 2019.

[465] Borut Bajc, Adrián Lugo, and Francesco Sannino. Asymptotically free and safe fate of symmetry nonrestoration. *Phys. Rev. D*, 103:096014, 2021.

[466] Oleksii Matsedonskyi and Geraldine Servant. High-Temperature Electroweak Symmetry Non-Restoration from New Fermions and Implications for Baryogenesis. *JHEP*, 09:012, 2020.

[467] Oleksii Matsedonskyi. High-Temperature Electroweak Symmetry Breaking by SM Twins. *JHEP*, 04:036, 2021.

[468] Wei Chao, Huai-Ke Guo, and Xiu-Fei Li. First Order Color Symmetry Breaking and Restoration Triggered by Electroweak Symmetry Non-restoration. 12 2021.

[469] M. Carena, Claudius Krause, Zhen Liu, and Yikun Wang. New approach to electroweak symmetry nonrestoration. *Physical Review D*, 104, 09 2021.

[470] Yang Bai, Seung J. Lee, Minho Son, and Fang Ye. Global electroweak symmetric vacuum. *JHEP*, 07:225, 2021.

[471] Peter Arnold and Olivier Espinosa. Effective potential and first-order phase transitions: Beyond leading order. *Phys. Rev. D*, 47:3546–3579, Apr 1993.

[472] Paolo Gondolo and Graciela Gelmini. Cosmic abundances of stable particles: Improved analysis. *Nuclear Physics B - NUCL PHYS B*, 360:145–179, 08 1991.

[473] Yann Mambrini and Keith A. Olive. Gravitational production of dark matter during reheating. *Phys. Rev. D*, 103:115009, Jun 2021.

[474] Simon Cléry, Yann Mambrini, Keith A. Olive, and Sarunas Verner. Gravitational portals in the early universe. *Phys. Rev. D*, 105:075005, Apr 2022.

[475] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.*, 641:A6, 2020. [Erratum: Astron.Astrophys. 652, C4 (2021)].

[476] E. Bañados *et al.* An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5. *Nature*, 553(7689):473–476, January 2018.

[477] Nicholas P. Ross and Nicholas J. G. Cross. The near and mid-infrared photometric properties of known redshift $z \geq 5$ quasars. *MNRAS*, 494(1):789–803, May 2020.

[478] Chris J. Willott et al. The Canada-France High-z Quasar Survey: nine new quasars and the luminosity function at redshift 6. *Astron. J.*, 139:906, 2010.

[479] E. Bañados et al. The Pan-STARRS1 distant $z > 5.6$ quasar survey: more than 100 quasars within the first Gyr of the universe. *Astrophys. J. Suppl.*, 227(1):11, 2016.

[480] Linhua Jiang, Ian D. McGreer, Xiaohui Fan, Michael A. Strauss, Eduardo Bañados, Robert H. Becker, Fuyan Bian, Kara Farnsworth, Yue Shen, Feige Wang, Ran Wang, Shu Wang, Richard L. White, Jin Wu, Xue-Bing Wu, Jinyi Yang, and Qian Yang. The Final SDSS High-redshift Quasar Sample of 52 Quasars at $z > 5.7$. *APJ*, 833(2):222, December 2016.

[481] S. L. Reed et al. Eight new luminous z $\geq$ 6 quasars discovered via SED model fitting of VISTA, WISE and Dark Energy Survey Year 1 observations. *Mon. Not. Roy. Astron. Soc.*, 468(4):4702–4718, 2017.

[482] Y. Matsuoka *et al.* Subaru High-z Exploration of Low-luminosity Quasars (SHELLQs). X. Discovery of 35 Quasars and Luminous Galaxies at $5.7 \leq z \leq 7.0$. *APJ*, 883(2):183, October 2019.

[483] J. Yang *et al.* Filling in the Quasar Redshift Gap at $z \sim 5.5$ II: A Complete Survey of Luminous Quasars in the Post-Reionization Universe. *arXiv e-prints*, page arXiv:1810.11927, October 2018.

[484] F. Wang *et al.* Exploring Reionization-era Quasars. III. Discovery of 16 Quasars at $6.4 \lesssim z \lesssim 6.9$ with DESI Legacy Imaging Surveys and the UKIRT Hemisphere Survey and Quasar Luminosity Function at $z \sim 6.7$. *APJ*, 884(1):30, October 2019.

[485] Burç in Mutlu-Pakdil, Marc S. Seigar, and Benjamin L. Davis. The local black holes mass function derived from the $m_{BH} - p$ and the $m_{BH} - n$ relations. *The Astrophysical Journal*, 830(2):117, oct 2016.

[486] Francesco Shankar, David H. Weinberg, and Jordi Miralda-Escudé. Accretion-driven evolution of black holes: Eddington ratios, duty cycles and active galaxy fractions. *Monthly Notices of the Royal Astronomical Society*, 428(1):421–446, October 2012.

[487] Xuejian Shen, Philip F Hopkins, Claude-André Faucher-Giguère, D M Alexander, Gordon T Richards, Nicholas P Ross, and R C Hickox. The bolometric quasar luminosity function at $z = 0 - 7$. *Monthly Notices of the Royal Astronomical Society*, 495(3):3252–3275, January 2020.

[488] Alex Sicilia, Andrea Lapi, Lumen Boco, Francesco Shankar, David M. Alexander, Viola Allevato, Carolin Villforth, Marcella Massardi, Mario Spera, Alessandro Bressan, and Luigi Danese. The black hole mass function across cosmic time. II. heavy seeds and (super)massive black holes. *The Astrophysical Journal*, 934(1):66, July 2022.

[489] Nilanjan Banik, Jonathan C Tan, and Pierluigi Monaco. The formation of supermassive black holes from Population III.1 seeds. I. Cosmic formation histories and clustering properties. *Monthly Notices of the Royal Astronomical Society*, 483(3):3592–3606, 12 2018.

[490] E. E. Salpeter. Accretion of Interstellar Matter by Massive Objects. *apj*, 140:796–800, August 1964.

[491] Xiaohui Fan et al. A Survey of z $>$ 5.8 quasars in the Sloan Digital Sky Survey I: Discovery of three new quasars and the spatial density of luminous quasars at $z \sim 6$. *Astron. J.*, 122:2833, 2001.

[492] Daniel J. Mortlock et al. A luminous quasar at a redshift of z = 7.085. *Nature*, 474:616, 2011.

[493] Xue-Bing Wu, Feige Wang, Xiaohui Fan, Weimin Yi, Wenwen Zuo, Fuyan Bian, Linhua Jiang, Ian D. McGreer, Ran Wang, Jinyi Yang, Qian Yang, David Thompson, and Yuri Beletsky. An ultraluminous quasar with a twelve-billion-solar-mass black hole at redshift 6.30. *Nature*, 518(7540):512–515, February 2015.

[494] Benny Trakhtenbrot, Hagai Netzer, Paulina Lira, and Ohad Shemmer. Black-Hole Mass and Growth Rate at $z \simeq 4.8$: A Short Episode of Fast Growth Followed by Short Duty Cycle Activity. *Astrophys. J.*, 730:7, 2011.

[495] Marta Volonteri. Formation of Supermassive Black Holes. *Astron. Astrophys. Rev.*, 18:279–315, 2010.

[496] Zoltan Haiman. Constraints from gravitational recoil on the growth of supermassive black holes at high redshift. *Astrophys. J.*, 613:36–40, 2004.

[497] Stuart L. Shapiro. Spin, accretion and the cosmological growth of supermassive black holes. *Astrophys. J.*, 620:59–68, 2005.

[498] Marta Volonteri and M. J. Rees. Quasars at z=6: The survival of the fittest. *Astrophys. J.*, 650:669–678, 2006.

[499] Takamitsu Tanaka and Zoltan Haiman. The Assembly of Supermassive Black Holes at High Redshifts. *Astrophys. J.*, 696:1798–1822, 2009.

[500] Hagai Netzer. The Largest black holes and the most luminous galaxies. *Astrophys. J. Lett.*, 583:L5–L8, 2003.

[501] Priyamvada Natarajan and Ezequiel Treister. Is there an upper limit to black hole masses? *Mon. Not. Roy. Astron. Soc.*, 393:838, 2009.

[502] Kohei Inayoshi and Zoltán Haiman. Is There a Maximum Mass for Black Holes in Galactic Nuclei? *APJ*, 828(2):110, September 2016.

[503] José Luis Bernal, Alvise Raccanelli, Licia Verde, and Joseph Silk. Signatures of primordial black holes as seeds of supermassive black holes. *JCAP*, 05:017, 2018. [Erratum: JCAP 01, E01 (2020)].

[504] Aaron Smith, Volker Bromm, and Abraham Loeb. The first supermassive black holes. *Astronomy & Geophysics*, 58(3):3–22, 2017.

[505] Jing Liu, Ligong Bian, Rong-Gen Cai, Zong-Kuan Guo, and Shao-Jiang Wang. Primordial black hole production during first-order phase transitions. *Phys. Rev. D*, 105(2):L021303, 2022.

[506] Jaume Garriga, Alexander Vilenkin, and Jun Zhang. Black holes and the multiverse. *JCAP*, 02:064, 2016.

[507] Hooman Davoudiasl, Peter B. Denton, and Julia Gehrlein. Supermassive Black Holes, Ultralight Dark Matter, and Gravitational Waves from a First Order Phase Transition. *Phys. Rev. Lett.*, 128(8):081101, 2022.

[508] Hai-Long Huang, Yong Cai, Jun-Qian Jiang, Jun Zhang, and Yun-Song Piao. Supermassive primordial black holes in multiverse: for nano-Hertz gravitational wave and high-redshift JWST galaxies. 6 2023.

[509] Bernard J. Carr. The Primordial black hole mass spectrum. *Astrophys. J.*, 201:1–19, 1975.

[510] Samuel M Leach, Misao Sasaki, David Wands, and Andrew R Liddle. Enhancement of superhorizon scale inflationary curvature perturbations. *Phys. Rev. D*, 64:023512, 2001.

[511] Juan Garcia-Bellido and Ester Ruiz Morales. Primordial black holes from single field models of inflation. *Phys. Dark Univ.*, 18:47–54, 2017.

[512] Yi-Fu Cai, Xiaocui Ma, Misao Sasaki, Dong-Gang Wang, and Zihan Zhou. Highly non-Gaussian tails and primordial black holes from single-field inflation. *JCAP*, 12:034, 2022.

[513] Chris Pattison, Vincent Vennin, Hooshyar Assadullahi, and David Wands. Quantum diffusion during inflation and primordial black holes. *JCAP*, 10:046, 2017.

[514] Matteo Biagetti, Gabriele Franciolini, Alex Kehagias, and Antonio Riotto. Primordial Black Holes from Inflation and Quantum Diffusion. *JCAP*, 07:032, 2018.

[515] Jose Maria Ezquiaga, Juan Garcia-Bellido, and Vincent Vennin. The exponential tail of inflationary fluctuations: consequences for primordial black holes. *JCAP*, 03:029, 2020.

[516] Christopher A. Pattison, Vincent Vennin, David Wands, and Hooshyar Assadullahi. Ultra-slow-roll inflation with quantum diffusion. *Journal of Cosmology and Astroparticle Physics*, 2021, 2021.

[517] Yi-Fu Cai, Xiao-Han Ma, Misao Sasaki, Dong-Gang Wang, and Zihan Zhou. One small step for an inflaton, one giant leap for inflation: A novel non-Gaussian tail and primordial black holes. *Phys. Lett. B*, 834:137461, 2022.

[518] Juan Garcia-Bellido, Andrei D. Linde, and David Wands. Density perturbations and black hole formation in hybrid inflation. *Phys. Rev. D*, 54:6040–6058, 1996.

[519] Sebastien Clesse and Juan Garcia-Bellido. Massive primordial black holes from hybrid inflation as dark matter and the seeds of galaxies. *Phys. Rev. D*, 92:023524, Jul 2015.

[520] Adam R. Brown. Hyperbolic Inflation. *Phys. Rev. Lett.*, 121(25):251601, 2018.

[521] Gonzalo A. Palma, Spyros Sypsas, and Cristobal Zenteno. Seeding primordial black holes in multifield inflation. *Phys. Rev. Lett.*, 125(12):121301, 2020.

[522] Jacopo Fumagalli, Sébastien Renaux-Petel, John W. Ronayne, and Lukas T. Witkowski. Turning in the landscape: A new mechanism for generating primordial black holes. *Phys. Lett. B*, 841:137921, 2023.

[523] Matteo Braglia, Dhiraj Kumar Hazra, Fabio Finelli, George F. Smoot, L. Sriramkumar, and Alexei A. Starobinsky. Generating PBHs and small-scale GWs in two-field models of inflation. *JCAP*, 08:001, 2020.

[524] Shi Pi and Misao Sasaki. Primordial Black Hole Formation in Non-Minimal Curvaton Scenario. 12 2021.

[525] Sarah R. Geller, Wenzer Qin, Evan McDonough, and David I. Kaiser. Primordial black holes from multifield inflation with nonminimal couplings. *Phys. Rev. D*, 106(6):063535, 2022.

[526] Yong Cai, Mian Zhu, and Yun-Song Piao. Primordial black holes from null energy condition violation during inflation. 5 2023.

[527] Y. Akrami et al. Planck 2018 results. IX. Constraints on primordial non-Gaussianity. *Astron. Astrophys.*, 641:A9, 2020.

[528] R. A. Sunyaev and Ya. B. Zeldovich. Small scale fluctuations of relic radiation. *Astrophys. Space Sci.*, 7:3–19, 1970.

[529] Ya. B. Zeldovich and R. A. Sunyaev. The Interaction of Matter and Radiation in a Hot-Model Universe. *Astrophys. Space Sci.*, 4:301–316, 1969.

[530] A. F. Illarionov and R. A. Sunyaev. Why the number of galactic X-ray stars is so small? *Astron. Astrophys.*, 39:185–196, 1975.

[531] Wayne Hu and Joseph Silk. Thermalization and spectral distortions of the cosmic background radiation. *Phys. Rev. D*, 48:485–502, 1993.

[532] Jens Chluba, Adrienne L. Erickcek, and Ido Ben-Dayan. Probing the inflaton: Small-scale power spectrum constraints from measurements of the CMB energy spectrum. *Astrophys. J.*, 758:76, 2012.

[533] Rishi Khatri and Rashid A. Sunyaev. Beyond y and μ: the shape of the cmb spectral distortions in the intermediate epoch, $1.5 \times 10^4 < z < 2 \times 10^5$. *JCAP*, 09:016, 2012.

[534] J. Chluba and R. A. Sunyaev. Superposition of blackbodies and the dipole anisotropy: A possibility to calibrate CMB experiments. *Astronomy & Astrophysics*, 424(2):389–408, August 2004.

[535] Albert Stebbins. Cmb spectral distortions from the scattering of temperature anisotropies, 2007.

[536] J. Chluba and R. A. Sunyaev. The evolution of CMB spectral distortions in the early Universe. *Mon. Not. Roy. Astron. Soc.*, 419:1294–1314, 2012.

[537] Cyril Pitrou and Albert Stebbins. Parameterization of temperature and spectral distortions in future CMB experiments. *General Relativity and Gravitation*, 46(11), October 2014.

[538] B. J. Carr and James E. Lidsey. Primordial black holes and generalized constraints on chaotic inflation. *Physical Review D*, 48(2):543–553, July 1993.

[539] Jens Chluba. Which spectral distortions does ΛCDM actually predict? *Mon. Not. Roy. Astron. Soc.*, 460(1):227–239, 2016.

[540] Andrew D. Gow, Tays Miranda, and Sami Nurmi. Primordial black holes from a curvaton scenario with strongly non-Gaussian perturbations. 7 2023.

[541] Michael Kopp, Stefan Hofmann, and Jochen Weller. Separate Universes Do Not Constrain Primordial Black Hole Formation. *Phys. Rev. D*, 83:124025, 2011.

[542] Tomohiro Nakama, Teruaki Suyama, and Jun'ichi Yokoyama. Supermassive black holes formed by direct collapse of inflationary perturbations. *Phys. Rev. D*, 94(10):103522, 2016.

[543] Mohammad Hossein Namjoo, Hassan Firouzjahi, and Misao Sasaki. Violation of non-Gaussianity consistency relation in a single field inflationary model. *EPL*, 101(3):39001, 2013.

[544] Xingang Chen, Hassan Firouzjahi, Mohammad Hossein Namjoo, and Misao Sasaki. A Single Field Inflation Model with Large Local Non-Gaussianity. *EPL*, 102(5):59001, 2013.

[545] Xingang Chen, Hassan Firouzjahi, Eiichiro Komatsu, Mohammad Hossein Namjoo, and Misao Sasaki. In-in and δN calculations of the bispectrum from non-attractor single-field inflation. *JCAP*, 12:039, 2013.

[546] Yi-Fu Cai, Xingang Chen, Mohammad Hossein Namjoo, Misao Sasaki, Dong-Gang Wang, and Ziwei Wang. Revisiting non-Gaussianity from non-attractor inflation models. *JCAP*, 05:012, 2018.

[547] D. S. Salopek and J. R. Bond. Nonlinear evolution of long wavelength metric fluctuations in inflationary models. *Phys. Rev. D*, 42:3936–3962, 1990.

[548] Misao Sasaki and Ewan D. Stewart. A General analytic formula for the spectral index of the density perturbations produced during inflation. *Prog. Theor. Phys.*, 95:71–78, 1996.

[549] Misao Sasaki and Takahiro Tanaka. Superhorizon scale dynamics of multiscalar inflation. *Prog. Theor. Phys.*, 99:763–782, 1998.

[550] David H. Lyth, Karim A. Malik, and Misao Sasaki. A General proof of the conservation of the curvature perturbation. *JCAP*, 05:004, 2005.

[551] David H. Lyth and Yeinzon Rodriguez. The Inflationary prediction for primordial non-Gaussianity. *Phys. Rev. Lett.*, 95:121302, 2005.

[552] Andrew D. Gow, Hooshyar Assadullahi, Joseph H. P. Jackson, Kazuya Koyama, Vincent Vennin, and David Wands. Non-perturbative non-Gaussianity and primordial black holes. *EPL*, 142(4):49001, 2023.

[553] Alexei A. Starobinsky. Stochastic de Sitter (Inflationary) Stage in the Early Universe. *Lect. Notes Phys.*, 246:107–126, 1986.

[554] Vincent Vennin and Alexei A. Starobinsky. Correlation Functions in Stochastic Inflation. *Eur. Phys. J. C*, 75:413, 2015.

[555] Yuichiro Tada and Vincent Vennin. Statistics of coarse-grained cosmological fields in stochastic inflation. *JCAP*, 02(02):021, 2022.

[556] Hassan Firouzjahi, Amin Nassiri-Rad, and Mahdiyar Noorbala. Stochastic Ultra Slow Roll Inflation. *JCAP*, 01:040, 2019.

[557] Ana Achucarro, Sebastian Cespedes, Anne-Christine Davis, and Gonzalo A. Palma. The hand-made tail: non-perturbative tails from multifield inflation. *JHEP*, 05:052, 2022.

[558] Shi Pi and Misao Sasaki. Logarithmic Duality of the Curvature Perturbation. *Phys. Rev. Lett.*, 131(1):011002, 2023.

[559] Vicente Atal, Albert Sanglas, and Nikolaos Triantafyllou. NANOGrav signal as mergers of Stupendously Large Primordial Black Holes. *JCAP*, 06:022, 2021.

[560] Misao Sasaki, Jussi Valiviita, and David Wands. Non-Gaussianity of the primordial perturbation in the curvaton model. *Phys. Rev. D*, 74:103003, 2006.

[561] David H. Lyth and David Wands. Generating the curvature perturbation without an inflaton. *Phys. Lett. B*, 524:5–14, 2002.

[562] Takeo Moroi and Tomo Takahashi. Effects of cosmological moduli fields on cosmic microwave background. *Phys. Lett. B*, 522:215–221, 2001. [Erratum: Phys.Lett.B 539, 303–303 (2002)].

[563] Kari Enqvist and Martin S. Sloth. Adiabatic CMB perturbations in pre - big bang string cosmology. *Nucl. Phys. B*, 626:395–409, 2002.

[564] N Bartolo and Andrew R Liddle. The Simplest curvaton model. *Phys. Rev. D*, 65:121301, 2002.

[565] Silvia Mollerach. Isocurvature Baryon Perturbations and Inflation. *Phys. Rev. D*, 42:313–325, 1990.

[566] Sam Young and Christian T. Byrnes. Primordial black holes in non-Gaussian regimes. *JCAP*, 08:052, 2013.

[567] David Langlois, Filippo Vernizzi, and David Wands. Non-linear isocurvature perturbations and non-Gaussianities. *JCAP*, 12:004, 2008.

[568] N. Bartolo, S. Matarrese, and A. Riotto. On nonGaussianity in the curvaton scenario. *Phys. Rev. D*, 69:043503, 2004.

[569] Kari Enqvist and Tomo Takahashi. Signatures of Non-Gaussianity in the Curvaton Model. *JCAP*, 09:012, 2008.

[570] Kari Enqvist, Sami Nurmi, Olli Taanila, and Tomo Takahashi. Non-Gaussian Fingerprints of Self-Interacting Curvaton. *JCAP*, 04:009, 2010.

[571] Kari Enqvist, Sami Nurmi, Gerasimos Rigopoulos, Olli Taanila, and Tomo Takahashi. The Subdominant Curvaton. *JCAP*, 11:003, 2009.

[572] Olli Taanila. *Primordial Perturbations from a Self-interacting Curvaton.* PhD thesis, Helsinki U., 2010.

[573] Kari Enqvist. The self-interacting curvaton. *Prog. Theor. Phys. Suppl.*, 190:62–74, 2011.

[574] Christian T. Byrnes, Kari Enqvist, and Tomo Takahashi. Scale-dependence of Non-Gaussianity in the Curvaton Model. *JCAP*, 09:026, 2010.

[575] Jose Fonseca and David Wands. Non-Gaussianity and Gravitational Waves from Quadratic and Self-interacting Curvaton. *Phys. Rev. D*, 83:064025, 2011.

[576] Christian T. Byrnes, Kari Enqvist, Sami Nurmi, and Tomo Takahashi. Strongly scale-dependent polyspectra from curvaton self-interactions. *JCAP*, 11:011, 2011.

[577] Takeshi Kobayashi and Tomo Takahashi. Runnings in the Curvaton. *JCAP*, 06:004, 2012.

[578] Guillem Domènech and Misao Sasaki. Conformal Frame Dependence of Inflation. *JCAP*, 04:022, 2015.

[579] Kari Enqvist, Sami Nurmi, Gerasimos Rigopoulos, Olli Taanila, and Tomo Takahashi. The Subdominant Curvaton. *JCAP*, 11:003, 2009.

[580] Shin'ichiro Ando, Nagisa Hiroshima, and Koji Ishiwata. Constraining the primordial curvature perturbation using dark matter substructure. *Phys. Rev. D*, 106(10):103014, 2022.

[581] Takumi Shinohara, Wanqiu He, Yoshiki Matsuoka, Tohru Nagao, Teruaki Suyama, and Tomo Takahashi. Supermassive primordial black holes: a view from clustering of quasars at $z \sim 6$. 4 2023.

[582] Gert Hütsi, Martti Raidal, Juan Urrutia, Ville Vaskonen, and Hardi Veermäe. Did JWST observe imprints of axion miniclusters or primordial black holes? *Phys. Rev. D*, 107(4):043502, 2023.

[583] Gabriella Agazie et al. The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background. *Astrophys. J. Lett.*, 951(1):L8, 2023.

[584] Gabriella Agazie et al. The NANOGrav 15-year Data Set: Bayesian Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries. 6 2023.

[585] Gabriella Agazie et al. The NANOGrav 15-year Data Set: Constraints on Supermassive Black Hole Binaries from the Gravitational Wave Background. 6 2023.

[586] Yann Gouttenoire, Sokratis Trifinopoulos, Georgios Valogiannis, and Miguel Vanvlasselaer. Scrutinizing the Primordial Black Holes Interpretation of PTA Gravitational Waves and JWST Early Galaxies. 7 2023.

[587] Paul Frederik Depta, Kai Schmidt-Hoberg, and Carlo Tasillo. Do pulsar timing arrays observe merging primordial black holes? 6 2023.

[588] Takumi Shinohara, Teruaki Suyama, and Tomo Takahashi. Angular correlation as a novel probe of supermassive primordial black holes. *Phys. Rev. D*, 104(2):023526, 2021.

[589] Gabriele Franciolini, Antonio Iovino, Junior., Ville Vaskonen, and Hardi Veermae. The recent gravitational wave observation by pulsar timing arrays and primordial black holes: the importance of non-gaussianities. 6 2023.

[590] Sai Wang, Zhi-Chao Zhao, Jun-Peng Li, and Qing-Hua Zhu. Exploring the Implications of 2023 Pulsar Timing Array Datasets for Scalar-Induced Gravitational Waves and Primordial Black Holes. 7 2023.

[591] James M. Bardeen. Gauge Invariant Cosmological Perturbations. *Phys. Rev. D*, 22:1882–1905, 1980.

[592] Sayan Mandal, Neelima Sehgal, and Toshiya Namikawa. Finding evidence for inflation and the origin of galactic magnetic fields with cmb surveys. *Phys. Rev. D*, 105(6):063537, 2022.

[593] Levon Pogosian, Meir Shimon, Matthew Mewes, and Brian Keating. Future cmb constraints on cosmic birefringence and implications for fundamental physics. *Phys. Rev. D*, 100(2):023507, 2019.

[594] Eugene N. Parker. Hydromagnetic Dynamo Models. *Astrophys. J.*, 122:293, 1955.

[595] Russell M. Kulsrud. A critical review of galactic dynamos. *Ann. Rev. Astron. Astrophys.*, 37:37–64, 1999.

[596] Axel Brandenburg and Kandaswamy Subramanian. Astrophysical magnetic fields and nonlinear dynamo theory. *Phys. Rept.*, 417:1–209, 2005.

[597] Russell M. Kulsrud and Ellen G. Zweibel. The Origin of Astrophysical Magnetic Fields. *Rept. Prog. Phys.*, 71:0046091, 2008.

[598] Ruth Durrer and Andrii Neronov. Cosmological Magnetic Fields: Their Generation, Evolution and Observation. *Astron. Astrophys. Rev.*, 21:62, 2013.

[599] Philipp P. Kronberg. Extragalactic magnetic fields. *Rept. Prog. Phys.*, 57:325–382, 1994.

[600] Dario Grasso and Hector R. Rubinstein. Magnetic fields in the early universe. *Phys. Rept.*, 348:163–266, 2001.

[601] Lawrence M. Widrow. Origin of galactic and extragalactic magnetic fields. *Rev. Mod. Phys.*, 74:775–823, 2002.

[602] Alejandra Kandus, Kerstin E. Kunze, and Christos G. Tsagas. Primordial magnetogenesis. *Phys. Rept.*, 505:1–58, 2011.

[603] Lawrence M. Widrow, Dongsu Ryu, Dominik R. G. Schleicher, Kandaswamy Subramanian, Christos G. Tsagas, and Rudolf A. Treumann. The First Magnetic Fields. *Space Sci. Rev.*, 166:37–70, 2012.

[604] Michael S. Turner and Lawrence M. Widrow. Inflation Produced, Large Scale Magnetic Fields. *Phys. Rev. D*, 37:2743, 1988.

[605] Bharat Ratra. Cosmological 'seed' magnetic field from inflation. *Astrophys. J. Lett.*, 391:L1–L4, 1992.

[606] Jerome Martin and Jun'ichi Yokoyama. Generation of Large-Scale Magnetic Fields in Single-Field Inflation. *JCAP*, 01:025, 2008.

[607] Kandaswamy Subramanian. Magnetic fields in the early universe. *Astron. Nachr.*, 331:110–120, 2010.

[608] Leonardo Motta and Robert R. Caldwell. Non-Gaussian features of primordial magnetic fields in power-law inflation. *Phys. Rev. D*, 85:103532, 2012.

[609] Rajeev Kumar Jain and Martin S. Sloth. Consistency relation for cosmic magnetic fields. *Phys. Rev. D*, 86:123528, 2012.

[610] T. Vachaspati. Magnetic fields from cosmological phase transitions. *Phys. Lett. B*, 265:258–261, 1991.

[611] K. Enqvist and P. Olesen. On primordial magnetic fields of electroweak origin. *Phys. Lett. B*, 319:178–185, 1993.

[612] Dario Grasso and Antonio Riotto. On the nature of the magnetic fields generated during the electroweak phase transition. *Phys. Lett. B*, 418:258–265, 1998.

[613] M. Joyce and Mikhail E. Shaposhnikov. Primordial magnetic fields, right-handed electrons, and the Abelian anomaly. *Phys. Rev. Lett.*, 79:1193–1196, 1997.

[614] Chiara Caprini, Ruth Durrer, and Geraldine Servant. The stochastic gravitational wave background from turbulence and magnetic fields generated by a first-order phase transition. *JCAP*, 12:024, 2009.

[615] Stephan J. Huber and Thomas Konstandin. Gravitational Wave Production by Collisions: More Bubbles. *JCAP*, 09:022, 2008.

[616] 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.

[617] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. W. H. Freeman, San Francisco, 1973.

[618] J. M. Bardeen, B. Carter, and S. W. Hawking. The four laws of black hole mechanics. *Commun. Math. Phys.*, 31(2):161, 1973.

[619] Anne M. Green and Bradley J. Kavanagh. Primordial Black Holes as a dark matter candidate. *J. Phys. G*, 48(4):043001, 2021.

[620] Bernard Carr, Kazunori Kohri, Yuuiti Sendouda, and Jun'ichi Yokoyama. Constraints on primordial black holes. *Rept. Prog. Phys.*, 84(11):116902, 2021.

[621] S. W. Hawking. Particle Creation by Black Holes. *Commun. Math. Phys.*, 43:199–220, 1975. [Erratum: Commun.Math.Phys. 46, 206 (1976)].

[622] G. W. Gibbons and S. W. Hawking. Cosmological Event Horizons, Thermodynamics, and Particle Creation. *Phys. Rev. D*, 15:2738–2751, 1977.

[623] Robi Banerjee and Karsten Jedamzik. The Evolution of cosmic magnetic fields: From the very early universe, to recombination, to the present. *Phys. Rev. D*, 70:123003, 2004.

[624] A. Neronov and D. V. Semikoz. Sensitivity of γ-ray telescopes for detection of magnetic fields in the intergalactic medium. *prd*, 80(12):123012, December 2009.

[625] Rafael Alves Batista and Andrey Saveliev. The gamma-ray window to intergalactic magnetism. *Universe*, 7(7):223, 2021.

[626] A. Neronov and I. Vovk. Evidence for strong extragalactic magnetic fields from Fermi observations of TeV blazars. *Science*, 328:73–75, 2010.

[627] Elisa Pueschel. VERITAS Search for Magnetically-Broadened Emission From Blazars. *PoS*, ICRC2015:817, 2016.

[628] M. Ackermann et al. The Search for Spatial Extension in High-latitude Sources Detected by the *Fermi* Large Area Telescope. *Astrophys. J. Suppl.*, 237(2):32, 2018.

[629] V. A. Acciari et al. A lower bound on intergalactic magnetic fields from time variability of 1ES 0229+200 from MAGIC and Fermi/LAT observations. *Astron. Astrophys.*, 670:A145, 2023.

[630] William A. Hiscock and Lance D. Weems. Evolution of charged evaporating black holes. *Phys. Rev. D*, 41(4):1142–1151, 1990.

[631] Don N. Page. Particle Emission Rates from a Black Hole: Massless Particles from an Uncharged, Nonrotating Hole. *Phys. Rev. D*, 13:198–206, 1976.

[632] Dan Hooper, Gordan Krnjaic, John March-Russell, Samuel D. McDermott, and Rudin Petrossian-Byrne. Hot Gravitons and Gravitational Waves From Kerr Black Holes in the Early Universe. 4 2020.

[633] D. T. Son. Magnetohydrodynamics of the early universe and the evolution of primordial magnetic fields. *Phys. Rev. D*, 59(6):063008, 1999.

[634] Tanmay Vachaspati. Progress on cosmological magnetic fields. *Rept. Prog. Phys.*, 84(7):074901, 2021.

[635] Axel Brandenburg and Tina Kahniashvili. Classes of hydrodynamic and magnetohydrodynamic turbulent decay. *Phys. Rev. Lett.*, 118(5):055102, 2017.

[636] Axel Brandenburg, Tina Kahniashvili, Sayan Mandal, Alberto Roper Pol, Alexander G. Tevzadze, and Tanmay Vachaspati. The dynamo effect in decaying helical turbulence. *Phys. Rev. Fluids*, 4:024608, 2019.

[637] Martti Raidal, Ville Vaskonen, and Hardi Veermäe. Gravitational Waves from Primordial Black Hole Mergers. *JCAP*, 09:037, 2017.

[638] Benjamin P. Abbott et al. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Phys. Rev. Lett.*, 118(22):221101, 2017. [Erratum: Phys.Rev.Lett. 121, 129901 (2018)].

[639] Robert V Wagoner. Big-bang nucleosynthesis revisited. *Astrophysical Journal, Vol. 179, pp. 343-360 (1973)*, 179:343–360, 1973.

[640] G. Gamow. The Origin of Elements and the Separation of Galaxies. *Physical Review*, 74(4):505–506, August 1948.

[641] Andrew Pontzen and Fabio Governato. How supernova feedback turns dark matter cusps into cores. *mnras*, 421(4):3464–3471, April 2012.

[642] James S. Bullock. Notes on the Missing Satellites Problem. *arXiv e-prints*, page arXiv:1009.4505, September 2010.

[643] Theodoros Papanikolaou, Vincent Vennin, and David Langlois. Gravitational waves from a universe filled with primordial black holes. *JCAP*, 03:053, 2021.

[644] Raymond Y. Chiao. Conceptual tensions between quantum mechanics and general relativity: Are there experimental consequences, e.g., superconducting transducers between electromagnetic and gravitational radiation? 8 2002.

[645] F. Pegoraro, L. A. Radicati, P. Bernard, and E. Picasso. Electromagnetic Detector for Gravitational Waves. *Phys. Lett. A*, 68:165–168, 1978.

[646] Aaron S. Chou et al. MHz Gravitational Wave Constraints with Decameter Michelson Interferometers. *Phys. Rev. D*, 95(6):063002, 2017.

[647] Asher Berlin, Diego Blas, Raffaele Tito D'Agnolo, Sebastian A. R. Ellis, Roni Harnik, Yonatan Kahn, and Jan Schütte-Engel. Detecting high-frequency gravitational waves with microwave cavities. *Phys. Rev. D*, 105(11):116011, 2022.

[648] Alexandre Arbey and Jérémy Auffinger. Physics Beyond the Standard Model with BlackHawk v2.0. *Eur. Phys. J. C*, 81:910, 2021.

[649] Francesco D'Eramo, Nicolas Fernandez, and Stefano Profumo. When the Universe Expands Too Fast: Relentless Dark Matter. *JCAP*, 05:012, 2017.

[650] Francesco D'Eramo, Nicolas Fernandez, and Stefano Profumo. Dark Matter Freeze-in Production in Fast-Expanding Universes. *JCAP*, 02:046, 2018.

[651] Stefano Profumo and Piero Ullio. SUSY dark matter and quintessence. *JCAP*, 11:006, 2003.

[652] Bernard Carr and Florian Kühnel. Primordial black holes as dark matter candidates. *SciPost Physics Lecture Notes*, may 2022.

[653] Roberto Emparan and Harvey S. Reall. Black holes in higher dimensions. *Living Reviews in Relativity*, 11(1), sep 2008.

[654] Hideo Kodama and Akihiro Ishibashi. A Master equation for gravitational perturbations of maximally symmetric black holes in higher dimensions. *Prog. Theor. Phys.*, 110:701–722, 2003.

[655] George Johnson. Primordial black hole constraints with large extra dimensions. *Journal of Cosmology and Astroparticle Physics*, 2020(09):046–046, sep 2020.

[656] Torsten Bringmann, Martin Eriksson, and Michael Gustafsson. Cosmological evolution of homogeneous universal extra dimensions. *Phys. Rev. D*, 68:063516, 2003.

[657] U. Gunther and A. Zhuk. Stabilization of internal spaces in multidimensional cosmology. *Phys. Rev. D*, 61:124001, 2000.

[658] U. Gunther and A. Zhuk. Multidimensional perfect fluid cosmology with stable compactified internal dimensions. *Class. Quant. Grav.*, 15:2025–2035, 1998.

[659] Sean M. Carroll, James Geddes, Mark B. Hoffman, and Robert M. Wald. Classical stabilization of homogeneous extra dimensions. *Phys. Rev. D*, 66:024036, 2002.

[660] N. Mohammedi. Dynamical compactification, standard cosmology and the accelerating universe. *Phys. Rev. D*, 65:104018, 2002.

[661] M. Bailes *et al.* Gravitational-wave physics and astronomy in the 2020s and 2030s. *Nature Reviews Physics*, 3(5):344–366, April 2021.

[662] Jay Marx et al. The Gravitational Wave International Committee Roadmap: The future of gravitational wave astronomy. 11 2011.

[663] J. Aasi et al. Advanced LIGO. *Class. Quant. Grav.*, 32:074001, 2015.

[664] Marcela Carena, Ying-Ying Li, Tong Ou, and Yikun Wang. Anatomy of the electroweak phase transition for dark sector induced baryogenesis. *JHEP*, 02:139, 2023.

[665] Daniel Cutting, Mark Hindmarsh, and David J. Weir. Vorticity, kinetic energy, and suppressed gravitational wave production in strong first order phase transitions. *Phys. Rev. Lett.*, 125(2):021302, 2020.

[666] Thomas Konstandin, Germano Nardini, and Ingo Rues. From Boltzmann equations to steady wall velocities. *JCAP*, 09:028, 2014.

[667] Chiara Caprini, Ruth Durrer, and Geraldine Servant. Gravitational wave generation from bubble collisions in first-order phase transitions: An analytic approach. *Phys. Rev. D*, 77:124015, 2008.

[668] Ryusuke Jinno and Masahiro Takimoto. Gravitational waves from bubble collisions: An analytic derivation. *Phys. Rev. D*, 95(2):024009, 2017.

[669] Peter Athron, Csaba Balázs, Andrew Fowlie, Lachlan Morris, and Lei Wu. Cosmological phase transitions: from perturbative particle physics to gravitational waves. 5 2023.

[670] T. P. Cheng and Marc Sher. Mass Matrix Ansatz and Flavor Nonconservation in Models with Multiple Higgs Doublets. *Phys. Rev. D*, 35:3484, 1987.

[671] Carlos A. Argüelles, Diyaselis Delgado, Avi Friedlander, Ali Kheirandish, Ibrahim Safa, Aaron C. Vincent, and Henry White. Dark matter decay to neutrinos. *Phys. Rev. D*, 108(12):123021, 2023.

[672] Gianpiero Mangano, Gennaro Miele, Sergio Pastor, Teguayco Pinto, Ofelia Pisanti, and Pasquale D. Serpico. Relic neutrino decoupling including flavor oscillations. *Nucl. Phys. B*, 729:221–234, 2005.

[673] R. N. Mohapatra and Vigdor L. Teplitz. Structures in the mirror universe. *Astrophys. J.*, 478:29–38, 1997.

[674] Robert Foot. Seven (and a half) reasons to believe in mirror matter: from neutrino puzzles to the inferred dark matter in the universe. *Acta Phys. Polon. B*, 32:2253–2270, 2001.

[675] Robert Foot. Mirror matter-type dark matter. *Int. J. Mod. Phys. D*, 13:2161–2192, 2004.

[676] Zurab Berezhiani, Santi Cassisi, Paolo Ciarcelluti, and Adriano Pietrinferni. Evolutionary and structural properties of mirror star MACHOs. *Astropart. Phys.*, 24:495–510, 2006.

[677] Paolo Ciarcelluti. Cosmology with mirror dark matter. *Int. J. Mod. Phys. D*, 19:2151–2230, 2010.

[678] Itzhak Goldman, Rabindra N. Mohapatra, and Shmuel Nussinov. Bounds on neutron-mirror neutron mixing from pulsar timing. *Phys. Rev. D*, 100(12):123021, 2019.

[679] Jean-Samuel Roux and James M. Cline. Constraining galactic structures of mirror dark matter. *Phys. Rev. D*, 102(6):063518, 2020.

[680] Maurício Hippert, Jack Setford, Hung Tan, David Curtin, Jacquelyn Noronha-Hostler, and Nicolas Yunes. Mirror neutron stars. *Phys. Rev. D*, 106(3):035025, 2022.

[681] Csaba Csáki, Cong-Sen Guan, Teng Ma, and Jing Shu. Twin Higgs with exact Z_2. *JHEP*, 12:005, 2020.

[682] Csaba Csáki, Teng Ma, and Jing Shu. Trigonometric Parity for Composite Higgs Models. *Phys. Rev. Lett.*, 121(23):231801, 2018.

[683] Torsten Bringmann, Saniya Heeba, Felix Kahlhoefer, and Kristian Vangsnes. Freezing-in a hot bath: resonances, medium effects and phase transitions. *JHEP*, 02:110, 2022.

[684] David Curtin, Patrick Meade, and Harikrishnan Ramani. Thermal Resummation and Phase Transitions. *Eur. Phys. J. C*, 78(9):787, 2018.

[685] Sungwoo Hong, Gowri Kurup, and Maxim Perelstein. Conformal Freeze-In of Dark Matter. *Phys. Rev. D*, 101(9):095037, 2020.

[686] Sungwoo Hong, Gowri Kurup, and Maxim Perelstein. Dark Matter from a Conformal Dark Sector. 7 2022.

[687] Michele Redi and Andrea Tesi. General freeze-in and freeze-out. *JHEP*, 12:060, 2021.

[688] Sidney Coleman and Erick Weinberg. Radiative corrections as the origin of spontaneous symmetry breaking. *Phys. Rev. D*, 7:1888–1910, Mar 1973.

[689] L. Dolan and R. Jackiw. Symmetry behavior at finite temperature. *Phys. Rev. D*, 9:3320–3341, Jun 1974.

[690] Anne M. Green. Primordial Black Holes: sirens of the early Universe. *Fundam. Theor. Phys.*, 178:129–149, 2015.

[691] J. A. Peacock and A. F. Heavens. Alternatives to the Press-Schechter cosmological mass function. *Mon. Not. Roy. Astron. Soc.*, 243:133–143, 1990.

[692] Richard G. Bower. The Evolution of groups of galaxies in the Press-Schechter formalism. *Mon. Not. Roy. Astron. Soc.*, 248:332, 1991.

[693] J. R. Bond, S. Cole, G. Efstathiou, and Nick Kaiser. Excursion set mass functions for hierarchical Gaussian fluctuations. *Astrophys. J.*, 379:440, 1991.

[694] Tomohiro Fujita, Masahiro Kawasaki, Yuichiro Tada, and Tomohiro Takesako. A new algorithm for calculating the curvature perturbations in stochastic inflation. *JCAP*, 12:036, 2013.

[695] Tomohiro Fujita, Masahiro Kawasaki, and Yuichiro Tada. Non-perturbative approach for curvature perturbations in stochastic δN formalism. *JCAP*, 10:030, 2014.

[696] Masahiro Kawasaki and Yuichiro Tada. Can massive primordial black holes be produced in mild waterfall hybrid inflation? *JCAP*, 08:041, 2016.

[697] Andrei D. Linde, Dmitri A. Linde, and Arthur Mezhlumian. From the Big Bang theory to the theory of a stationary universe. *Phys. Rev. D*, 49:1783–1826, 1994.

[698] Alexei A. Starobinsky and Junichi Yokoyama. Equilibrium state of a selfinteracting scalar field in the De Sitter background. *Phys. Rev. D*, 50:6357–6368, 1994.

[699] Kenta Ando and Vincent Vennin. Power spectrum in stochastic inflation. *JCAP*, 04:057, 2021.

[700] Sandeep Kumar Acharya and Jens Chluba. CMB spectral distortions from continuous large energy release. *Mon. Not. Roy. Astron. Soc.*, 515(4):5775–5789, 2022.

[701] Joaquim Iguaz Juan, Pasquale D. Serpico, and Guillermo Franco Abellán. The QCD phase transition behind a PBH origin of LIGO/Virgo events? *JCAP*, 07(07):009, 2022.

[702] Tullia Sbarrato. Big and young supermassive black holes in the early Universe. 7 2021.

[703] Chul-Moon Yoo, Jinn-Ouk Gong, and Shuichiro Yokoyama. Abundance of primordial black holes with local non-Gaussianity in peak theory. *JCAP*, 09:033, 2019.

[704] Chul-Moon Yoo, Tomohiro Harada, Jaume Garriga, and Kazunori Kohri. Primordial black hole abundance from random Gaussian curvature perturbations and a local density threshold. *PTEP*, 2018(12):123E01, 2018.

[705] J. C. Mather *et al.* Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. *The Astrophysical Journal*, 420:439, January 1994.

[706] Valerio De Luca, Gabriele Franciolini, and Antonio Riotto. Heavy Primordial Black Holes from Strongly Clustered Light Black Holes. *Phys. Rev. Lett.*, 130(17):171401, 2023.

[707] Andrew D. Gow, Christian T. Byrnes, Philippa S. Cole, and Sam Young. The power spectrum on small scales: Robust constraints and comparing PBH methodologies. *JCAP*, 02:002, 2021.

[708] Ilia Musco, Karsten Jedamzik, and Sam Young. Primordial black hole formation during the QCD phase transition: threshold, mass distribution and abundance. 3 2023.

[709] Bernard Carr, Sebastien Clesse, Juan García-Bellido, and Florian Kühnel. Cosmic conundra explained by thermal history and primordial black holes. *Phys. Dark Univ.*, 31:100755, 2021.

[710]

[711] Tomohiro Harada, Chul-Moon Yoo, Kazunori Kohri, Yasutaka Koga, and Takeru Monobe. Spins of primordial black holes formed in the radiation-dominated phase of the universe: First-order effect. *The Astrophysical Journal*, 908(2):140, 2021.

[712] Kandaswamy Subramanian. The origin, evolution and signatures of primordial magnetic fields. *Rept. Prog. Phys.*, 79(7):076901, 2016.

[713] Julien Manshanden, Daniele Gaggero, Gianfranco Bertone, Riley M. T. Connors, and Massimo Ricotti. Multi-wavelength astronomical searches for primordial black holes. *JCAP*, 06:026, 2019.

[714] Philip Lu, Volodymyr Takhistov, Graciela B. Gelmini, Kohei Hayashi, Yoshiyuki Inoue, and Alexander Kusenko. Constraining Primordial Black Holes with Dwarf Galaxy Heating. *Astrophys. J. Lett.*, 908(2):L23, 2021.

[715] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975.

[716] J. Buchner, A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni, M. Salvato, J. Donley, and D. Kocevski. X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *Astron. Astrophys.*, 564:A125, 2014.

[717] Georg Weidenspointner et al. An asymmetric distribution of positrons in the Galactic disk revealed by gamma-rays. *Nature*, 451:159–162, 2008.

[718] William DeRocco and Peter W. Graham. Constraining Primordial Black Hole Abundance with the Galactic 511 keV Line. *Phys. Rev. Lett.*, 123(25):251102, 2019.

[719] Georg Weidenspointner et al. The sky distribution of 511-keV positron annihilation line emission as measured with INTEGRAL/SPI. *ESA Spec. Publ.*, 622:25, 2007.

[720] Pierre Jean, J. Knodlseder, W. Gillard, N. Guessoum, K. Ferriere, A. Marcowith, V. Lonjou, and J. P. Roques. Spectral analysis of the galactic e+ e- annihilation emission. *Astron. Astrophys.*, 445:579–589, 2006.

[721] Georg Weidenspointner et al. SPI observations of positron annihilation radiation from the 4th galactic quadrant: Sky distribution. In *5th INTEGRAL Workshop: The INTEGRAL Universe*, 6 2004.

[722] Dan Hooper and Lisa Goodenough. Dark Matter Annihilation in The Galactic Center As Seen by the Fermi Gamma Ray Space Telescope. *Phys. Lett. B*, 697:412–428, 2011.

[723] Lisa Goodenough and Dan Hooper. Possible Evidence For Dark Matter Annihilation In The Inner Milky Way From The Fermi Gamma Ray Space Telescope. 10 2009.

[724] Dan Hooper and Tim Linden. On The Origin Of The Gamma Rays From The Galactic Center. *Phys. Rev. D*, 84:123005, 2011.

[725] Nicolás Bernal, Lina Necib, and Tracy R. Slatyer. Spherical Cows in Dark Matter Indirect Detection. *JCAP*, 12:030, 2016.

[726] Matthieu Schaller et al. Dark matter annihilation radiation in hydrodynamic simulations of Milky Way haloes. *Mon. Not. Roy. Astron. Soc.*, 455(4):4442–4451, 2016.

[727] Francesca Calore, Nassim Bozorgnia, Mark Lovell, Gianfranco Bertone, Matthieu Schaller, Carlos S. Frenk, Robert A. Crain, Joop Schaye, Tom Theuns, and James W. Trayford. Simulated Milky Way analogues: implications for dark matter indirect searches. *JCAP*, 12:053, 2015.

[728] Matthieu Schaller, Carlos S. Frenk, Richard G. Bower, Tom Theuns, Adrian Jenkins, Joop Schaye, Robert A. Crain, Michelle Furlong, Claudio Dalla Vecchia, and I. G. McCarthy. Baryon effects on the internal structure of ΛCDM haloes in the EAGLE simulations. *Mon. Not. Roy. Astron. Soc.*, 451(2):1247–1267, 2015.

[729] Arianna Di Cintio, Chris B. Brook, Aaron A. Dutton, Andrea V. Macciò, Greg S. Stinson, and Alexander Knebe. A mass-dependent density profile for dark matter haloes including the influence of galaxy formation. *Mon. Not. Roy. Astron. Soc.*, 441(4):2986–2995, 2014.

[730] Arianna Di Cintio, Chris B. Brook, Andrea V. Macciò, Greg S. Stinson, Alexander Knebe, Aaron A. Dutton, and James Wadsley. The dependence of dark matter profiles on the stellar-to-halo mass ratio: a prediction for cusps versus cores. *Mon. Not. Roy. Astron. Soc.*, 437(1):415–423, 2014.

[731] Oleg Y. Gnedin, Daniel Ceverino, Nickolay Y. Gnedin, Anatoly A. Klypin, Andrey V. Kravtsov, Robyn Levine, Daisuke Nagai, and Gustavo Yepes. Halo Contraction Effect in Hydrodynamic Simulations of Galaxy Formation. 8 2011.

[732] Tansu Daylan, Douglas P. Finkbeiner, Dan Hooper, Tim Linden, Stephen K. N. Portillo, Nicholas L. Rodd, and Tracy R. Slatyer. The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter. *Phys. Dark Univ.*, 12:1–23, 2016.

[733] M. Ackermann et al. The Fermi Galactic Center GeV Excess and Implications for Dark Matter. *Astrophys. J.*, 840(1):43, 2017.

[734] A. W. Strong, H. Bloemen, R. Diehl, W. Hermsen, and V. Schoenfelder. Comptel skymapping: A New approach using parallel computing. *Astrophys. Lett. Commun.*, 39:209, 1999.

[735] A. De Angelis et al. The e-ASTROGAM mission. *Exper. Astron.*, 44(1):25–82, 2017.

[736] Regina Caputo et al. All-sky Medium Energy Gamma-ray Observatory: Exploring the Extreme Multimessenger Universe. 7 2019.

[737] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. The Structure of cold dark matter halos. *Astrophys. J.*, 462:563–575, 1996.

[738] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. A Universal density profile from hierarchical clustering. *Astrophys. J.*, 490:493–508, 1997.

[739] J. H. MacGibbon and B. R. Webber. Quark and gluon jet emission from primordial black holes: The instantaneous spectra. *Phys. Rev. D*, 41:3052–3079, 1990.

[740] Jane H. MacGibbon. Quark and gluon jet emission from primordial black holes. 2. The Lifetime emission. *Phys. Rev. D*, 44:376–392, 1991.

[741] John F. Beacom, Nicole F. Bell, and Gianfranco Bertone. Gamma-ray constraint on Galactic positron production by MeV dark matter. *Phys. Rev. Lett.*, 94:171301, 2005.

[742] M. Ackermann et al. Search for Gamma-Ray Emission from Local Primordial Black Holes with the Fermi Large Area Telescope. *Astrophys. J.*, 857(1):49, 2018.

[743] Ranjan Laha, Julian B. Muñoz, and Tracy R. Slatyer. INTEGRAL constraints on primordial black holes and particle dark matter. *Phys. Rev. D*, 101(12):123514, 2020.

[744] Adam Coogan, Logan Morrison, and Stefano Profumo. Direct Detection of Hawking Radiation from Asteroid-Mass Primordial Black Holes. 10 2020.

[745] Ranjan Laha. Primordial Black Holes as a Dark Matter Candidate Are Severely Constrained by the Galactic Center 511 keV γ -Ray Line. *Phys. Rev. Lett.*, 123(25):251101, 2019.

[746] Dan Hooper and Lian-Tao Wang. Possible evidence for axino dark matter in the galactic bulge. *Phys. Rev. D*, 70:063506, 2004.

[747] Celeste Keith and Dan Hooper. 511 keV excess and primordial black holes. *Phys. Rev. D*, 104(6):063033, 2021.

[748] Eugene Churazov, R. Sunyaev, S. Sazonov, M. Revnivtsev, and D. Varshalovich. Positron annihilation spectrum from the Galactic Center region observed by SPI/INTEGRAL. *Mon. Not. Roy. Astron. Soc.*, 357:1377–1386, 2005.

[749] C. A. Kierans et al. Detection of the 511keV Galactic Positron Annihilation Line with COSI. *Astrophys. J.*, 895(1):44, 2020.

[750] Basudeb Dasgupta, Ranjan Laha, and Anupam Ray. Neutrino and positron constraints on spinning primordial black hole dark matter. *Phys. Rev. Lett.*, 125(10):101101, 2020.

[751] Laurent Bouchet, Andrew W. Strong, Troy A. Porter, Igor V. Moskalenko, Elisabeth Jourdain, and Jean-Pierre Roques. Diffuse emission measurement with INTEGRAL/SPI as indirect probe of cosmic-ray electrons and positrons. *Astrophys. J.*, 739:29, 2011.

[752] Mathieu Boudaud and Marco Cirelli. Voyager 1 $e^{\pm}$ Further Constrain Primordial Black Holes as Dark Matter. *Phys. Rev. Lett.*, 122(4):041104, 2019.

[753] Laurent Bouchet, Jean-Pierre Roques, and Elisabeth Jourdain. On the morphology of the electron-positron annihilation emission as seen by SPI/INTEGRAL. *Astrophys. J.*, 720:1772–1780, 2010.

[754] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *apj*, 622(2):759–771, April 2005.

[755] A. E. Vladimirov, S. W. Digel, G. Jóhannesson, P. F. Michelson, I. V. Moskalenko, P. L. Nolan, E. Orlando, T. A. Porter, and A. W. Strong. GALPROP WebRun: An internet-based service for calculating galactic cosmic ray propagation and associated photon emissions. *Computer Physics Communications*, 182(5):1156–1161, May 2011.

[756] Dan Hooper. TASI Lectures on Indirect Searches For Dark Matter. *PoS*, TASI2018:010, 2019.

[757] M. Ackermann et al. The spectrum of isotropic diffuse gamma-ray emission between 100 MeV and 820 GeV. *Astrophys. J.*, 799:86, 2015.

[758] Ilias Cholis, Yi-Ming Zhong, Samuel D. McDermott, and Joseph P. Surdutovich. The Return of the Templates: Revisiting the Galactic Center Excess with Multi-Messenger Observations. 12 2021.

[759] John F. Beacom and Hasan Yuksel. Stringent constraint on galactic positron production. *Phys. Rev. Lett.*, 97:071102, 2006.

[760] Rong-Gen Cai, Yu-Chen Ding, Xing-Yu Yang, and Yu-Feng Zhou. Constraints on a mixed model of dark matter particles and primordial black holes from the Galactic 511 keV line. 7 2020.

[761] Rouven Essig, Eric Kuflik, Samuel D. McDermott, Tomer Volansky, and Kathryn M. Zurek. Constraining Light Dark Matter with Diffuse X-Ray and Gamma-Ray Observations. *JHEP*, 11:193, 2013.

[762] Celine Boehm, Dan Hooper, Joseph Silk, Michel Casse, and Jacques Paul. MeV dark matter: Has it been detected? *Phys. Rev. Lett.*, 92:101301, 2004.

[763] Shaaban Khalil and Osamu Seto. Sterile neutrino dark matter in B - L extension of the standard model and galactic 511-keV line. *JCAP*, 10:024, 2008.

[764] G. W. Gibbons and S. W. Hawking, editors. *Euclidean quantum gravity.* 1994.

[765] Dan Hooper and Kathryn M. Zurek. A Natural Supersymmetric Model with MeV Dark Matter. *Phys. Rev. D*, 77:087302, 2008.

[766] Emrah Kalemci, S. E. Boggs, P. A. Milne, and Stephen P. Reynolds. Searching for annihilation radiation from sn 1006 with spi on integral. *Astrophys. J. Lett.*, 640:L55–L58, 2006.

[767] M. Casse, B. Cordier, J. Paul, and Stephane Schanne. Hypernovae / GRB in the Galactic center as possible sources of Galactic positrons. *Astrophys. J. Lett.*, 602:L17–L20, 2004.

[768] Gianfranco Bertone, Alexander Kusenko, Sergio Palomares-Ruiz, Silvia Pascoli, and Dmitry Semikoz. Gamma ray bursts and the origin of galactic positrons. *Phys. Lett. B*, 636:20–24, 2006.

[769] N. Guessoum, Pierre Jean, and N. Prantzos. Microquasars as sources of positron annihilation radiation. *Astron. Astrophys.*, 457:753, 2006.

[770] Kyle Lawson and Ariel Zhitnitsky. Quark nugget dark matter: no contradiction with 511 keV line emission from dwarf galaxies. *JCAP*, 02:049, 2017.

[771] Yasaman Farzan and Meshkat Rajaee. Pico-charged intermediate particles rescue dark matter interpretation of 511 keV signal. *JHEP*, 12:083, 2017.

[772] Yasaman Farzan and M. Rajaee. Pico-charged particles explaining 511 keV line and XENON1T signal. *Phys. Rev. D*, 102(10):103532, 2020.

[773] Marco Cirelli, Nicolao Fornengo, Bradley J. Kavanagh, and Elena Pinetti. Integral X-ray constraints on sub-GeV Dark Matter. 7 2020.

[774] Volodymyr Takhistov. Novel Signals from Neutron Star Mergers at 511 keV. *PoS*, ICRC2019:803, 2020.

[775] George M. Fuller, Alexander Kusenko, David Radice, and Volodymyr Takhistov. Positrons and 511 keV Radiation as Tracers of Recent Binary Neutron Star Mergers. *Phys. Rev. Lett.*, 122(12):121101, 2019.

[776] Richard Bartels, Francesca Calore, Emma Storm, and Christoph Weniger. Galactic binaries can explain the Fermi Galactic centre excess and 511 keV emission. *Mon. Not. Roy. Astron. Soc.*, 480(3):3826–3841, 2018.

[777] Lian-Bao Jia. Explanation of the 511 keV line: cascade annihilating dark matter with the^8 Be anomaly. *Eur. Phys. J. C*, 78(2):112, 2018.

[778] R. Lehoucq, M. Casse, J. M. Casandjian, and I. Grenier. New constraints on the primordial black hole number density from Galactic gamma-ray astronomy. *Astron. Astrophys.*, 502:37, 2009.

[779] A. De Angelis et al. The e-ASTROGAM mission. *Exper. Astron.*, 44(1):25–82, 2017.

[780] Thomas Siegert, Roland Diehl, Aaron C. Vincent, Fabrizia Guglielmetti, Martin G. H. Krause, and Celine Boehm. Search for 511 keV Emission in Satellite Galaxies of the Milky Way with INTEGRAL/SPI. *Astron. Astrophys.*, 595:A25, 2016.

[781] Ryan J. Wilkinson, Aaron C. Vincent, Céline Bœhm, and Christopher McCabe. Ruling out the light weakly interacting massive particle explanation of the Galactic 511 keV line. *Phys. Rev. D*, 94(10):103525, 2016.

[782] Man Ho Chan. Electron–positron pair production near the Galactic Centre and the 511 keV emission line. *Mon. Not. Roy. Astron. Soc.*, 456(1):L113–L116, 2016.

[783] Aaron C. Vincent, Antonio Vecchio, Jordi Miralda-Escudé, and Carlos Peña Garay. The Galactic 511 keV Line and the Intergalactic Positron Density. *Phys. Procedia*, 61:796–801, 2015.

[784] A. Vecchio, A. C. Vincent, J. Miralda-Escude, and C. Pena-Garay. The positron density in the intergalactic medium and the galactic 511 keV line. 4 2013.

[785] Andrew R. Frey and Nicholas B. Reid. Cosmic microwave background constraints on dark matter models of the Galactic center 511 keV signal. *Phys. Rev. D*, 87(10):103508, 2013.

[786] Aaron C. Vincent, Pierrick Martin, and James M. Cline. Interacting dark matter contribution to the Galactic 511 keV gamma ray emission: constraining the morphology with INTEGRAL/SPI observations. *JCAP*, 04:022, 2012.

[787] N. Prantzos et al. The 511 keV emission from positron annihilation in the Galaxy. *Rev. Mod. Phys.*, 83:1001–1056, 2011.

[788] James M. Cline, Andrew R. Frey, and Fang Chen. Metastable dark matter mechanisms for INTEGRAL 511 keV γ rays and DAMA/CoGeNT events. *Phys. Rev. D*, 83:083511, 2011.

[789] Nicolas Bernal and Sergio Palomares-Ruiz. Constraining Dark Matter Properties with Gamma-Rays from the Galactic Center with Fermi-LAT. *Nucl. Phys. B*, 857:380–410, 2012.

[790] Shinta Kasuya and Fuminobu Takahashi. 511-kev line from q balls in the galactic center. *Phys. Rev. D*, 72:085015, 2005.

[791] J. M. Cline and A. R. Frey. Abelian dark matter models for 511 keV gamma rays and direct detection. *Annalen Phys.*, 524:579–590, 2012.

[792] Maxim Pospelov and Adam Ritz. The galactic 511 keV line from electroweak scale WIMPs. *Phys. Lett. B*, 651:208–215, 2007.

[793] Hooman Davoudiasl and Gilad Perez. The INTEGRAL/SPI 511 keV Signal from Hidden Valleys in Type Ia and Core Collapse Supernova Explosions. *JHEP*, 04:058, 2010.

[794] R. E. Lingenfelter, J. C. Higdon, and R. E. Rothschild. Is There a Dark Matter Signal in the Galactic Positron Annihilation Radiation? *Phys. Rev. Lett.*, 103:031301, 2009.

[795] Cosimo Bambi, Alexander D. Dolgov, and Alexey A. Petrov. Primordial black holes and the observed Galactic 511-keV line. *Phys. Lett. B*, 670:174–178, 2008. [Erratum: Phys.Lett.B 681, 504–504 (2009)].

[796] Paul H. Frampton and Thomas W. Kephart. Primordial black holes, Hawking radiation and the early universe. *Mod. Phys. Lett. A*, 20:1573–1576, 2005.

[797] Nathaniel J. Craig and Stuart Raby. Modulino Dark Matter and the INTEGRAL 511-keV Line. 8 2009.

[798] Jose A. R. Cembranos and Louis E. Strigari. Diffuse MeV Gamma-rays and Galactic 511 keV Line from Decaying WIMP Dark Matter. *Phys. Rev. D*, 77:123519, 2008.

[799] Ji-Haeng Huh, Jihn E. Kim, Jong-Chul Park, and Seong Chan Park. Galactic 511 keV line from MeV milli-charged dark matter. *Phys. Rev. D*, 77:123503, 2008.

[800] Douglas P. Finkbeiner and Neal Weiner. Exciting Dark Matter and the INTEGRAL/SPI 511 keV signal. *Phys. Rev. D*, 76:083519, 2007.

[801] Nikos Prantzos. On the intensity and spatial morphology of the 511-kev emission in the milky way. *Astron. Astrophys.*, 449:869–878, 2006.

[802] Yago Ascasibar, P. Jean, C. Boehm, and J. Knoedlseder. Constraints on dark matter and the shape of the Milky Way dark halo from the 511-keV line. *Mon. Not. Roy. Astron. Soc.*, 368:1695–1705, 2006.

[803] B. J. Carr, Kazunori Kohri, Yuuiti Sendouda, and Jun'ichi Yokoyama. New cosmological constraints on primordial black holes. *Phys. Rev. D*, 81:104019, 2010.

[804] Yohei Ema, Filippo Sala, and Ryosuke Sato. Dark matter models for the 511 keV galactic line predict keV electron recoils on Earth. *Eur. Phys. J. C*, 81(2):129, 2021.

[805] J. L. G. Sobrinho and P. Augusto. Direct detection of Black Holes via electromagnetic radiation. *Mon. Not. Roy. Astron. Soc.*, 441(4):2878–2884, 2014.

[806] S. W. Hawking, I. G. Moss, and J. M. Stewart. Bubble Collisions in the Very Early Universe. *Phys. Rev. D*, 26:2681, 1982.

[807] 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 ? cold dark matter galaxies. *Monthly Notices of the Royal Astronomical Society*, 422(2):1231?1240, Mar 2012.

[808] Michael Kuhlen, Javiera Guedes, Annalisa Pillepich, Piero Madau, and Lucio Mayer. An Off-center Density Peak in the Milky Way's Dark Matter Halo? *Astrophys. J.*, 765:10, 2013.

[809] Oleg Y. Gnedin, Andrey V. Kravtsov, Anatoly A. Klypin, and Daisuke Nagai. Response of dark matter halos to condensation of baryons: Cosmological simulations and improved adiabatic contraction model. *Astrophys. J.*, 616:16–26, 2004.

[810] Martin D. Weinberg and Neal Katz. Bar-driven dark halo evolution: a resolution of the cusp-core controversy. *Astrophys. J.*, 580:627–633, 2002.

[811] Martin D. Weinberg and Neal Katz. The bar-halo interaction. 2. secular evolution and the religion of n-body simulations. *Mon. Not. Roy. Astron. Soc.*, 375:460–476, 2007.

[812] J. A. Sellwood. Bars and dark matter halo cores. *Astrophys. J.*, 587:638–648, 2003.

[813] Octavio Valenzuela and Anatoly Klypin. Secular bar formation in galaxies with significant amount of dark matter. *Mon. Not. Roy. Astron. Soc.*, 345:406, 2003.

[814] Pedro Colin, O. Valenzuela, and A. Klypin. Bars and cold dark matter halos. *Astrophys. J.*, 644:687–700, 2006.

[815] Bernard Carr, Florian Kuhnel, and Luca Visinelli. Black Holes and WIMPs: All or Nothing or Something Else. 11 2020.

[816] Julian Adamek, Christian T. Byrnes, Mateja Gosenca, and Shaun Hotchkiss. WIMPs and stellar-mass primordial black holes are incompatible. *Phys. Rev. D*, 100(2):023506, 2019.

[817] S. Abdollahi et al. *Fermi* Large Area Telescope Fourth Source Catalog. *Astrophys. J. Suppl.*, 247(1):33, 2020.