

# Probing Models of Dark Matter and the Early Universe

by

Nicholas David Orlofsky

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Physics)  
in the University of Michigan  
2017

Doctoral Committee:

Professor Aaron T. Pierce, Chair  
Associate Professor Henriette Elvang  
Associate Professor Oleg Y. Gnedin  
Professor Gordon L. Kane  
Professor Roberto D. Merlin

Nicholas David Orlofsky

[norlofs@umich.edu](mailto:norlofs@umich.edu)

ORCID iD: [0000-0002-8848-1843](https://orcid.org/0000-0002-8848-1843)

To my parents, who encouraged my curiosity and supported my goals,  
and to Lauren, my partner in life.

# Acknowledgements

I would like to thank my advisor, Aaron Pierce, for his guidance and support throughout my graduate studies. He has been an encouraging mentor and has shaped my view on particle physics.

I would also like to thank my other collaborators Jim Halverson, Jack Kearney, and James Wells, without whom very little of this thesis would have been written.

My thanks also go to Henriette Elvang, Oleg Gnedin, Gordon Kane, and Roberto Merlin for serving on my committee and for all of the positive interactions and feedback they have given me.

My professors Ratin Akhoury, Finn Larsen, Jim Liu, Leo Pando-Zayas, and Kathryn Zurek at UM have taught me a great deal in and out of the classroom. I have also enjoyed many stimulating conversations with a number of the students and postdocs in the group, including Sebastian Ellis, Sam Roland, Bibhushan Shakya, Kevin Zhang, Yue Zhao, and Bob Zheng. In addition, my undergraduate mentor Francesc Ferrer played a pivotal role in my early research inclinations.

# Table of Contents

Dedication	ii
Acknowledgements	iii
List of Tables	vi
List of Figures	vii
Abstract	x
<b>Chapter 1: Introduction</b>	<b>1</b>
1.1 Cosmology Review . . . . .	2
1.2 WIMP Dark Matter . . . . .	11
1.3 Axions . . . . .	14
1.4 Primordial Black Holes . . . . .	21
1.5 Motivation and Outline . . . . .	23
<b>Chapter 2: Models of Dark Matter that Couple to Electroweak Mediators</b>	<b>25</b>
2.1 $Z$ -Mediated Dark Matter . . . . .	28
2.2 Embedding $Z$ -Mediated Dark Matter in the Singlet-Doublet Model . . . . .	33
2.3 Vectorlike Leptons as the Tip of the Dark Matter Iceberg . . . . .	40
2.4 Conclusions . . . . .	53
<b>Chapter 3: High-Scale Axions without Isocurvature from Inflationary Dynamics</b>	<b>55</b>
3.1 Suppressing Isocurvature via Wave Function Renormalization . . . . .	59
3.2 Axion Mass from Enhanced Explicit $U(1)_{PQ}$ Breaking During Inflation . . . . .	67
3.3 Additional Constraints on Radial Field Displacement . . . . .	87
3.4 Can the Radial Field Evolve Adiabatically? . . . . .	90

3.5	Conclusions . . . . .	94
<b>Chapter 4: Inflationary Theory and Pulsar Timing Investigations of Primordial Black Holes and Gravitational Waves</b>		<b>97</b>
4.1	PBH Binary Merger Rate . . . . .	99
4.2	PBH and GW Spectra from a Primordial Scalar Spectrum . . . . .	102
4.3	Idealized Delta Function Spectrum . . . . .	106
4.4	Models of Primordial Scalar Spectra . . . . .	107
4.5	Discussion . . . . .	115
4.6	Conclusions . . . . .	117
<b>Chapter 5: Summary and Outlook</b>		<b>118</b>
<b>Bibliography</b>		<b>124</b>

# List of Tables

2.1	Cross sections of backgrounds and signal (following cuts (i) – (vi) in the text) and signal significances for the monojet + $\cancel{E}_T$ channel with $\sqrt{s} = 14$ TeV, $\mathcal{L} = 3000 \text{ fb}^{-1}$ , and $p_T^{\text{cut}} = 500$ GeV. . . . .	48
2.2	Number of the disappearing track background and signal events for $\sqrt{s} = 14$ TeV, $\mathcal{L} = 3000 \text{ fb}^{-1}$ , and $500 \text{ GeV} < p_T^{\text{track}} < 1500$ GeV. . . . .	52

# List of Figures

- 2.1 Constraints on  $g_A \equiv cv^2/(2\Lambda^2)$  as a function of  $m_\chi$  for  $Z$ -mediated DM. The black solid line indicates where the thermal relic abundance matches the observed abundance [1]. The orange dotted (dot-dashed) line corresponds to omitting higher dimensional couplings to  $Zhh$  (and  $Zh$ ), in violation of gauge invariance. Shaded (dashed) blue is excluded by PandaX [2] (LZ [3]), with bounds obtained as described in the text. The purple shaded region is excluded by IceCube (IC) [4]. Invisible  $Z$  decay limits ( $\Gamma_Z$ ) [5] are shown in gray. Pink shaded regions indicate where  $T < -0.09$  (upper, darker) and  $T < -0.01$  (lower, lighter), corresponding to  $2\sigma$  excluded regions depending if  $S$  is nonzero or zero, respectively. Green shaded regions indicate where the EFT is precarious, taking  $c_{\max} = 4\pi$  (upper, darker) or 1 (lower, lighter). . . . . 30
- 2.2 Constraints on the singlet-doublet model at the Higgs blind spot. The blind spot condition fixes  $y^c$  via (2.11), while  $M_D$  (contours in blue) is set by demanding the thermal relic abundance matches the observed DM abundance. Shown in red are current bounds from PandaX [2] (solid, shaded) and prospective bounds from Xenon1T [6] (thick dashed) and LZ [3] (thin dashed) on  $\sigma_{\text{SD}}$ . In the shaded gray region, the DM- $Z$  is at least 95% of that in the  $Z$ -mediated model of the previous section. . . . . 36
- 2.3 Similar to Fig. 2.2, but with  $y^c$  deviating from the Higgs blind spot by  $\delta_{y^c} = -0.3$ , see Eq. (2.14). Mirroring the previous figure, contours of  $M_D$  are shown in blue. Current and projected limits on  $\sigma_{\text{SD}}$  ( $\sigma_{\text{SI}}$ ) are shown in red (orange). 38
- 2.4 Relic density as a function of  $T_{\text{RH}}$  for various values of  $\eta$  for  $\mu_X = 100$  GeV. The solid lines correspond to, from top to bottom,  $\eta = 10^{-10}$  (orange),  $10^{-12}$  (blue),  $10^{-14}$  (purple), and  $10^{-16}$  (gray). The horizontal dashed lines correspond to the maximum allowed relic density to evade LUX [7] (top) and prospective Xe1T [8] (bottom) bounds. The vertical dashed line represents the cutoff of allowable  $T_{\text{RH}}$  due to BBN. . . . . 45

2.5	Allowed regions (shaded) of $c_X$ for $\mu_X = 100$ GeV from LUX [7] (blue solid) and prospective Xe1T [8] (red dashed) bounds and for $\mu_X = 1$ TeV from LUX (yellow dotted) and prospective Xe1T (green dot-dashed) bounds. . . . .	46
2.6	The $p_T^{\text{track}}$ distribution of the background (black, dashed) and signal (solid) with $p_T(j_1) > 300$ GeV and $P_{\text{mis}} = 1$ at $\sqrt{s} = 14$ TeV and $\mathcal{L} = 3000$ fb $^{-1}$ . Signal spectra correspond to, from top to bottom, $\mu_X = 100$ GeV (red), 130 GeV (purple), 150 GeV (blue), and 170 GeV (green). . . . .	51
2.7	The significance $\chi$ as a function of $\mu_X$ with $\sqrt{s} = 14$ TeV, $500$ GeV $< p_T^{\text{track}} < 1500$ GeV, and $\sigma_B = \beta_{\text{tot}}B$ , $\beta_{\text{tot}} = .25$ . The dashed curves are for $\mathcal{L} = 300$ fb $^{-1}$ assuming $P_{\text{mis}} = 10$ and $p_T(j_1) > 300$ GeV (green, lower) or $P_{\text{mis}} = 1$ and $p_T(j_1) > 200$ GeV (blue, upper). The solid curves are for $\mathcal{L} = 3000$ fb $^{-1}$ assuming $P_{\text{mis}} = 10$ and $p_T(j_1) > 500$ GeV (purple, lower) or $P_{\text{mis}} = 1$ and $p_T(j_1) > 300$ GeV (red, upper). All $p_T(j_1)$ cuts have been chosen to optimize the significance as described in the text. Dotted lines indicate $2\sigma$ exclusion and $5\sigma$ discovery thresholds. . . . .	52
3.1	Constraints on the radial part of the PQ field acting as the inflaton. The red curves denote isocurvature constraints after taking into account $f_I > f_a$ assuming $R_a = 1$ with an abundance determined by either Eq. (3.4) (solid, shaded upwards) or Eq. (3.11) with $T_{\text{RH}} = 6$ MeV (dot-dashed). The two blue curves show two prescriptions for determining where the PQ symmetry may be restored by parametric resonance (see text for details); the solid shaded bound should be viewed as definitively excluded. The black vertical region on the right shows the current BICEP2/Keck/Planck bounds on the tensor to scalar ratio [9,10], and the gray vertical region on the left is inaccessible in this model. . . . .	66
3.2	Bounds on the $S^N$ model in the parameter space of $f_a$ and $N$ for $ k  = 1$ . The right axis shows the required value for $f_I$ to give the axion a large enough inflationary mass to suppress isocurvature. Blue corresponds to the strong CP constraint. Red indicates where the contributions to the slow-roll parameters from the $S$ potential are greater than present bounds. The hatched region indicates where $\langle S \rangle_e > 10^4 \frac{f_a}{\sqrt{2}}$ so PQ symmetry restoration via parametric resonance may occur (see text for caveats). Vertical lines indicate the CASPER Phase 2 (right) and ideal (left) reach (prospective bounds extend to the right) [11]. . . . .	73

3.3	Bounds on $IS^N$ model with $m_I = 10^{-5}M_P$ and $ k  = 1$ . All curves are the same as in Fig. 3.2. . . . .	78
3.4	Bounds on multiple PQ field model for $ k  =  k'  =  k''  = 1$ , $m_{\bar{S}}^2 = 10^{-11}M_P^2$ . Contours of the maximal $\frac{f_a}{M_P}$ allowed by strong CP constraints are solid black. Regions where the contributions to the slow-roll parameters are greater than 1 (10) times current bounds are shaded light (dark) red. The region where $v_{\bar{S},e} > 10^4 v_{\bar{S},0}$ so that symmetry restoration via parametric resonance may be a concern is crosshatched. Within the dashed black contour denotes where our assumption that the $\lambda_{\bar{S}}$ term is negligible today breaks down. In the horizontal-hatched region, $a_{\text{QCD}} \simeq a_{\bar{S}}$ always. . . . .	85
3.5	Bounds on multiple PQ field model for $m_{\bar{S}}^2 = 10^{-12}M_P^2$ . All curves are the same as in Fig. 3.4. . . . .	86
4.1	Uncertainty in the merger rate of binary PBHs today from varying $10^{-1} < \gamma, \delta < 10$ as described in the text (blue band). The band outlined by gray dashed lines is the LIGO observed merger rate. For comparison, the red line shows a similar calculation in an older version of [12]. . . . .	102
4.2	Gravitational wave abundance for an idealized delta function scalar spectrum peaked at the scale $k_f$ corresponding to PBH mass $30M_{\odot}$ according to Eq. (4.9) and normalized so that $\Omega_{\text{PBH}} = \Omega_{\text{DM}}$ . . . . .	107
4.3	Gravitational wave abundance (envelopes) as a function of frequency assuming $\Omega_{\text{PBH}} = \Omega_{\text{DM}}$ . The PBH abundance spectrum is peaked at 30 (top) or 10 (bottom) $M_{\odot}$ . We display SGW for a top-hat spectrum with width set by expectations from parametric resonance (green “PR” curve), a red-tilted scalar spectrum with spectral index $n_s = -1$ supplemented by a cutoff at a minimum frequency (red “No PR” curve), and the spectrum from the running mass model (purple curve). Black solid lines are current spectrum-independent bounds from EPTA (upper) [13], NANOGrav (middle) [14], and PPTA (lower) [15]. The black dashed line is a projection for bounds from SKA [16]. The top axis indicates the approximate observing time $T$ to be sensitive to a given minimum frequency $f_{\text{min}} \sim 1/T$ . . . . .	109

# Abstract

This thesis discusses models for dark matter (DM) and their behavior in the early universe. An important question is how phenomenological probes can directly search for signals of DM today. Another topic of investigation is how the DM and other processes in the early universe must evolve. Then, astrophysical bounds on early universe dynamics can constrain DM. We will consider these questions in the context of three classes of DM models—weakly interacting massive particles (WIMPs), axions, and primordial black holes (PBHs).

Starting with WIMPs, we consider models where the DM is charged under the electroweak gauge group of the Standard Model. Such WIMPs, if generated by a thermal cosmological history, are constrained by direct detection experiments. To avoid present or near-future bounds, the WIMP model or cosmological history must be altered in some way. This may be accomplished by the inclusion of new states that coannihilate with the WIMP or a period of non-thermal evolution in the early universe. Future experiments are likely to probe some of these altered scenarios, and a non-observation would require a high degree of tuning in some of the model parameters in these scenarios.

Next, axions, as light pseudo-Nambu-Goldstone bosons, are susceptible to quantum fluctuations in the early universe that lead to isocurvature perturbations, which are constrained by observations of the cosmic microwave background (CMB). We ask what it would take to allow axion models in the face of these strong CMB bounds. We revisit models where inflationary dynamics modify the axion potential and discuss how isocurvature bounds can be relaxed, elucidating the difficulties in these constructions. Avoiding disruption of inflationary

dynamics provides important limits on the parameter space.

Finally, PBHs have received interest in part due to observations by LIGO of merging black hole binaries. We ask how these PBHs could arise through inflationary models and investigate the opportunity for corroboration through experimental probes of gravitational waves at pulsar timing arrays. We provide examples of theories that are already ruled out, theories that will soon be probed, and theories that will not be tested in the foreseeable future. The models that are most strongly constrained are those with relatively broad primordial power spectra.

# Chapter 1

## Introduction

The well-established existence of dark matter (DM) offers some of the best evidence for physics beyond the Standard Model (BSM). The first indication for DM came from observations of velocity dispersions of galaxies in galaxy clusters [17] (and later from stars in galaxies [18]), where it was observed that galaxies were moving so quickly that they should not remain gravitationally bound to their clusters if all of the mass of the cluster lay in visible matter. Later evidence came from measurements of the cosmic microwave background (CMB) [19, 20], simulations of large scale structure [21], and gravitational lensing observations [22], including those of colliding galaxy clusters like the Bullet Cluster [23, 24]. All of these observations have shown that DM interacts gravitationally, however no other interactions of DM with itself or with the Standard Model (SM) particles have been observed. Based on these observations, no SM particle can be identified with the DM [25]. Thus, while the evidence for dark matter (DM) is firmly established, its composition remains a mystery.

An understanding of the theoretical structure for DM may hold clues to uncover a fuller picture of all of BSM physics. Thus, a wealth of models have been proposed to explain what DM is composed of. Sections 1.2-1.4 will introduce three broad paradigms of DM that have generated some of the most research interest—weakly interacting massive particles (WIMPs), axions, and primordial black holes (PBHs) as massive compact halo objects (MACHOs)—

and we will revisit specific models for each in greater detail in later chapters.

However, before introducing these DM paradigms, we provide a brief introduction to cosmology and inflation in Section 1.1. The reader may find a fuller treatment of these topics in, *e.g.*, [26–28].

Throughout, we will use natural units,  $c = \hbar = k_B = 1$ , where  $c$  is the speed of light,  $\hbar$  is the reduced Planck constant, and  $k_B$  is the Boltzmann constant.

## 1.1 Cosmology Review

Assuming the universe is homogeneous and isotropic, it can be modeled by the Friedmann-Robertson-Walker (FRW) metric

$$ds^2 = -dt^2 + a^2(t) \left( \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right), \quad (1.1)$$

where  $k$  describes the curvature of the universe and  $a(t)$  is the scale factor normalized so that today, when  $t = t_0$ ,  $a(t_0) = 1$ . It will be convenient later to define the conformal time

$$\eta = \int_0^t \frac{dt'}{a(t')}, \quad (1.2)$$

under which the FRW metric becomes

$$ds^2 = a^2(t) \left( -d\eta^2 + \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right). \quad (1.3)$$

Using the Einstein equations, the Friedmann equation can be derived from the FRW metric as

$$\left( \frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} = \frac{1}{3M_P^2} \rho. \quad (1.4)$$

Here,  $M_P \equiv (8\pi G)^{-1/2} \simeq 2.435 \times 10^{18}$  GeV is the reduced Planck mass,  $G$  is Newton's constant,  $\rho$  is the energy density, and the dot represents a derivative with respect to time  $t$ .

In the first term,

$$H \equiv \frac{\dot{a}}{a} \tag{1.5}$$

describes the expansion rate of the universe and is called the Hubble parameter. The Hubble parameter today  $H_0$  is often written using the dimensionless Hubble parameter  $h$  as  $H_0 = h \times 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Additionally, the acceleration of the expansion of the universe can be derived from the Einstein equations as

$$\frac{\ddot{a}}{a} = -\frac{1}{6M_P^2}(\rho + 3p), \tag{1.6}$$

where  $p$  is the pressure.

The Friedmann equation can be recast as

$$\frac{k}{a^2 H^2} = \frac{\rho}{3H^2 M_P^2} - 1 \equiv \Omega - 1. \tag{1.7}$$

Thus, if the universe is flat ( $k = 0$ ), then  $\Omega = 1$  and the energy density of the universe is equal to the “critical density”

$$\rho_c \equiv 3H^2 M_P^2. \tag{1.8}$$

The energy density  $\rho$ , number density  $n$ , and pressure  $p$  of particle species can be determined from thermodynamics. They are calculated by integrating the phase space distribution weighted by the appropriate factors, assuming kinetic equilibrium. For relativistic particles (“radiation”) with masses much less than the temperature  $m \ll T$  and small chemical po-

tential  $\mu \ll T$ ,

$$\rho = \left\{ \begin{array}{c} 1 \\ 7/8 \end{array} \right\} \frac{\pi^2}{30} g T^4, \quad (1.9)$$

$$n = \left\{ \begin{array}{c} 1 \\ 3/4 \end{array} \right\} \frac{\zeta(3)}{\pi^2} g T^3, \quad (1.10)$$

$$p = \rho/3, \quad (1.11)$$

where the upper (lower) numbers in braces are for bosons (fermions) and  $g$  is the number of particle degrees of freedom. The differing numerical coefficients result from the use of Bose-Einstein or Fermi-Dirac phase space distributions. Thus, the total energy density in relativistic degrees of freedom is

$$\rho_r = \frac{\pi^2}{30} g_* T^4, \quad (1.12)$$

where  $g_*$  has been helpfully defined,

$$g_* \equiv \sum_{i=\text{bosons}} g_i \left( \frac{T_i}{T} \right)^4 + \frac{7}{8} \sum_{i=\text{fermions}} g_i \left( \frac{T_i}{T} \right)^4. \quad (1.13)$$

Here,  $T$  is the photon temperature and the summations are only over relativistic species.

In the non-relativistic (“matter”) limit  $m \gg T$ ,

$$n = g \left( \frac{mT}{2\pi} \right)^{3/2} \exp \left( -\frac{m - \mu}{T} \right), \quad (1.14)$$

$$\rho = mn, \quad (1.15)$$

$$p = nT \ll \rho, \quad (1.16)$$

valid for fermions and bosons.

During periods of thermal equilibrium, the comoving entropy density

$$S = sa^3 = \frac{\rho + p}{T} a^3 \quad (1.17)$$

is conserved. The entropy density is given by

$$s = \frac{2\pi^2}{45} g_{*S} T^3, \quad (1.18)$$

where

$$g_{*S} \equiv \sum_{i=\text{bosons}} g_i \left(\frac{T_i}{T}\right)^3 + \frac{7}{8} \sum_{i=\text{fermions}} g_i \left(\frac{T_i}{T}\right)^3, \quad (1.19)$$

where again the summations are only over relativistic particles.

The continuity equation can also be derived from Eqs. (1.4) and (1.6) as

$$\dot{\rho} + 3H(\rho + p) = 0. \quad (1.20)$$

Using this and the relationship  $p = \rho/3$  (0) for radiation (matter) given above, the scaling of the energy density of radiation and matter, respectively, can be derived as

$$\rho_r \propto a^{-4}, \quad (1.21)$$

$$\rho_m \propto a^{-3}. \quad (1.22)$$

These scalings have an intuitive explanation. For matter, the number of particles in a comoving volume ought to be conserved:  $N_m = n_m a^3 = \text{constant}$ , and  $\rho_m = mn_m$ . The comoving number density for radiation also ought to be constant, however its wavelength should scale as  $\lambda \propto a$ , introducing the additional power of  $a^{-1}$  in the energy density compared to matter.

### 1.1.1 Inflation

From observations of the cosmic microwave background (CMB) [20] and baryon acoustic oscillations (BAO) [29–32], the universe is known to be flat to within experimental precision:  $\Omega = 1.000 \pm 0.005$  at 95% confidence level [20]. However, rewriting the Friedman equation (1.4), for any nonzero  $k$  the value for  $\Omega$  should evolve away from unity as the universe expands according to:

$$\Omega^{-1} = 1 - 3M_P^2 k^2 \frac{1}{\rho a^2}. \quad (1.23)$$

During radiation (matter) domination,  $\rho \simeq \rho_r$  ( $\rho_m$ )  $\propto a^{-4}$  ( $a^{-3}$ ), so the effect of the last term grows as the universe expands in both cases. Thus,  $\Omega$  needed to be exceedingly close to unity in the early universe for the universe to still appear flat today. Indeed, the radius of curvature during the early universe must have been many orders of magnitude greater than the horizon size. This raises the question of why the initial conditions of the universe were such that the universe began so nearly flat, which is known as the “flatness problem.”

The CMB temperature is also homogeneous and isotropic over the whole sky down to the level of  $\mathcal{O}(10^{-5})$ . However, the distance scales for this temperature correlation are much larger than the horizon size at the time when the CMB was formed. This would imply that these seemingly causally-disconnected regions of the early universe somehow began with the same initial conditions despite being unable to communicate with each other. This is another initial conditions problem known as the “horizon problem.”

Both problems can be explained with a period of rapid expansion in the universe, called inflation [33]. As we will see, this rapid expansion leads to the universe’s apparent flatness and to regions that were initially causally connected becoming isolated.

In addition to solving these problems, inflation provides a mechanism by which the  $\mathcal{O}(10^{-5})$  fluctuations can be explained. Further, inflation can dilute the abundance of unwanted relics that may arise in new physics models to explain why these relics are not detected today.

The key to inflation is to achieve a period of time when the energy density of the universe is approximately constant with respect to time. Then, the Friedmann equation becomes

$$H_I = \frac{\dot{a}}{a} = \sqrt{\frac{\rho}{3M_P^2}} \simeq \text{const}, \quad (1.24)$$

where  $H_I$  denotes the Hubble parameter during inflation. This is solved by an exponential

$$a(t) \sim e^{H_I t}. \quad (1.25)$$

This exponential expansion solves the flatness and horizon problems described above. The flatness problem is solved because  $\rho \simeq \text{constant}$  and  $a$  is increasing exponentially, so the last term in Eq. (1.23) becomes exponentially suppressed and  $\Omega$  is driven very close to unity before matter or radiation domination. The horizon problem is solved because previously causally-connected patches of the universe are exponentially enlarged to well beyond the size of the horizon.

To demonstrate how inflation works, we will take the example of a single scalar field  $I$ , called the inflaton, that drives inflation. The equation of motion (EOM) for  $I$  is

$$\ddot{I} + 3H\dot{I} + \Gamma_I \dot{I} + V'(I) = 0, \quad (1.26)$$

where  $V(I)$  is the potential for  $I$ ,  $\Gamma_I$  is its decay width, and the prime denotes a derivative with respect to  $I$ .

To achieve a constant energy density, it is assumed that for some value of  $I$ ,  $V(I)$  is nonzero and greater than the energy density from all other fields. Further,  $I$  must evolve slowly. This “slow roll” condition can be made more precise using the EOM. If  $I$  is evolving slowly, then  $\ddot{I}$  and  $\Gamma_I \dot{I} \ll 3H^2 \dot{I}$  and  $V'(I)$ . The condition on  $\Gamma_I$  is simply the statement that  $I$  should not decay until after inflation. The smallness of  $\ddot{I}$ —necessary so the field value does not evolve quickly in accordance with (1.24)—leads to the condition that the slow roll

parameters, defined by

$$\epsilon \equiv \frac{M_P^2}{2} \left( \frac{V'}{V} \right)^2, \quad (1.27)$$

$$\eta \equiv M_P^2 \frac{V''}{V}, \quad (1.28)$$

have absolute values  $\ll 1$ . Slow roll parameters involving higher derivatives on  $V$  can also be derived.

Once the slow roll condition is violated, the inflaton will evolve rapidly and inflation will cease. Just as inflation is ending, the universe is very cold and sparsely populated because of the rapid expansion it has just undergone. It is commonly assumed that  $V(I)$  has some local minimum, and the energy in  $I$  converts from purely vacuum energy to energy in coherent oscillations, behaving as matter. To repopulate the universe with other particles, it is then assumed that  $I$  is unstable and decays to other particles, giving rise the  $\Gamma_I$  term in Eq. (1.26). This time period is known as “reheating,” and it continues until  $\Gamma_I \simeq H$  when the inflaton decays quickly and ceases to dominate the universe.

As alluded to above, inflation also seeds initial perturbations in the universe. During inflation, the inflaton is approximately massless and is susceptible to quantum fluctuations, which are of the order

$$\delta I = \frac{H_I}{2\pi}. \quad (1.29)$$

Thus, different patches of the universe will have slightly different values for the inflaton field. But the value of the inflaton field determines its energy density, so the quantum fluctuations lead directly to energy density perturbations. These perturbations could, for example, explain the CMB temperature anisotropies observed at the level of  $\mathcal{O}(10^{-5})$ , and they may provide the seeds for structure formation. While the CMB perturbations need not necessarily be seeded during inflation, inflationary perturbations are important in many contexts, some of which will be discussed in Chapters 3 and 4.

The inflationary perturbations can be written as perturbations on the FRW metric. At

leading order in perturbation theory,

$$ds^2 = -(1 + 2\Phi)dt^2 + 2aB_i dx^i dt + a^2 [(1 - 2\Psi)\delta_{ij} + E_{ij}] dx^i dx^j. \quad (1.30)$$

Using the scalar-vector-tensor (SVT) decomposition for the metric perturbations,

$$B_i \equiv \partial_i B - S_i, \quad \partial_i S_i = 0, \quad (1.31)$$

$$E_{ij} \equiv 2\partial_{ij} E + 2\partial_{(i} F_{j)} + h_{ij}, \quad \partial^i F_i = 0, \quad h_i^i = \partial^i h_{ij} = 0. \quad (1.32)$$

Inflation will not produce vector perturbations  $S_i$  and  $F_i$  because massless vector fields are conformally invariant (and they would anyways decay as the universe expands), so we will not consider them further. The tensor perturbations  $h_{ij}$  are often decomposed into polarization states using the eigenmodes of the spacial Lagrangian  $\nabla^2 e_{ij} = -k^2 e_{ij}$ :

$$h_{ij} = h(t) e_{ij}^{+, \times}(x). \quad (1.33)$$

The scalar perturbations  $\Phi, \Psi, E, B$  are not invariant under coordinate transformation. Two useful gauge invariant quantities are the curvature perturbation on uniform-density hypersurfaces and the comoving curvature perturbation, respectively,

$$-\zeta \equiv \Psi + \frac{H}{\bar{\rho}} \delta\rho, \quad (1.34)$$

$$\mathcal{R} \equiv \Psi - \frac{H}{\bar{\rho} + \bar{p}} \delta q. \quad (1.35)$$

Here, barred quantities denote background values, quantities with  $\delta$  denote fluctuations about the background, and  $\delta q$  is the scalar part of the 3-momentum density  $T_i^0 = \partial_i \delta q$ . During slow-roll inflation and on superhorizon scales  $k \ll aH$ ,  $\zeta = \mathcal{R}$ .

The power spectra of  $X = \mathcal{R}, h^+, h^\times$ , where the latter two are the polarization modes of

$h_{ij}$ , are defined using the ensemble average

$$\langle X_{\mathbf{k}} X_{\mathbf{k}'} \rangle \equiv (2\pi)^3 \delta(\mathbf{k} + \mathbf{k}') \mathcal{P}_X(k), \quad \mathcal{P}_X(k) \equiv \frac{2\pi^2}{k^3} \Delta_X^2(k). \quad (1.36)$$

It can be shown that inflation leads to scalar and tensor perturbations satisfying, respectively,

$$\Delta_{\mathcal{R}}^2(k) = \frac{1}{8\pi^2} \frac{H^2}{M_P^2} \frac{1}{\epsilon} \Big|_{k=aH}, \quad (1.37)$$

$$\Delta_t^2(k) = 2\Delta_h^2(k) = \frac{2}{\pi^2} \frac{H^2}{M_P^2} \Big|_{k=aH}. \quad (1.38)$$

The quantity  $\Delta_{\mathcal{R}}^2$  gives the scalar perturbations and is measured by Planck as  $\Delta_{\mathcal{R}}^2 = 2.142 \times 10^{-9}$  at scale  $k_0 = 0.05 \text{ Mpc}^{-1}$  [20]. Defining the tensor to scalar ratio,

$$r \equiv \frac{\Delta_t^2}{\Delta_{\mathcal{R}}^2} = 16\epsilon, \quad (1.39)$$

current bounds from measurements of the CMB by BICEP2/Keck/Planck place a constraint from non-observation of tensor modes at  $r < 0.07$  at the same scale  $k_0$  [9, 10].

These measurements and bounds provide information about the inflationary energy and potential. For example, the small measured value for the scalar perturbations gives information about how the inflaton potential must be constructed. Additionally, the bound on the tensor perturbations limits the energy scale of inflation. Using the above equations,

$$H_I = \frac{\pi}{\sqrt{2}} M_P \Delta_{\mathcal{R}} \sqrt{r} \lesssim 10^{14} \text{ GeV}. \quad (1.40)$$

Future CMB experiments [34] will probe as low as  $H_I \sim 10^{13} \text{ GeV}$ . However, probing many orders of magnitude lower than this will pose a challenge because the tensor perturbations would be difficult to distinguish from the scalar perturbations. Thus, only high scales of inflation are likely to be probed by measurements of the CMB.

The evolution of the inflaton field during the course of inflation can be estimated as

$$\frac{\Delta I}{M_P} = \mathcal{O}(1) \times \left( \frac{r}{0.01} \right)^{1/2}. \quad (1.41)$$

This is known as the Lyth bound [35], an important consequence of which is that the inflaton must take super-Planckian values for detectably-large values of  $r$ .

## 1.2 WIMP Dark Matter

We begin by using cosmology to understand thermally-produced DM, which provides the motivation for WIMP DM from the so-called ‘‘WIMP miracle.’’ In the very early universe, the DM  $\chi$  may annihilate to other particles  $X$ , and  $\chi$  may be populated by other particles annihilating to it, leading to a thermal equilibrium. We will schematically say

$$\chi\bar{\chi} \longleftrightarrow X\bar{X}, \quad (1.42)$$

although the interactions need not always be  $2 \longleftrightarrow 2$  nor always involve a particle-antiparticle pair on either side of the interaction. As the universe expands and cools to temperatures below the mass of  $\chi$ , the rate of the backwards interaction in (1.42) decreases as it becomes kinematically inaccessible, so the abundance of  $\chi$  decreases exponentially with temperature as it annihilates without being replenished. However, these annihilations will cease when the density of  $\chi$  becomes so low that they are unable to find each other to annihilate due to the expansion of the universe. At that point, the annihilations of  $\chi$  ‘‘freeze out,’’ and the comoving number density of  $\chi$  becomes constant.

This qualitative picture is governed by the Boltzmann equation written in the form

$$\frac{dn_\chi}{dt} + 3Hn_\chi = -\langle\sigma|v|\rangle \left( n_\chi^2 - (n_\chi^{\text{EQ}})^2 \right), \quad (1.43)$$

where  $n_\chi$  is the number density of  $\chi$ ,  $n_\chi^{\text{EQ}}$  is the equilibrium number density of  $\chi$  (see Section 1.1), and  $\langle\sigma|v|\rangle$  is the total thermally averaged annihilation cross section of  $\chi$  (see, *e.g.*, [26] for a formal definition).

Let us reinterpret the qualitative story of freeze out using the Boltzmann equation. At early times,  $n_\chi \simeq n_\chi^{\text{EQ}}$ . Once  $T < m$ ,  $n_\chi^{\text{EQ}}$  begins to exponentially decrease as it annihilates without being replenished, see Eq. (1.14). However, when the annihilation rate  $\Gamma \equiv \langle\sigma|v|\rangle n_\chi$  falls below the Hubble expansion rate ( $\Gamma < H$ ), the  $\chi$  annihilation rate cannot keep up with the expansion rate of the universe, so  $\chi$  effectively ceases to annihilate. At this point, the right hand side of (1.43) is negligible, so the number density of  $\chi$  is governed by the expansion of the universe. Accordingly, the comoving number density  $N_\chi = n_\chi a^3$  is constant. This is freeze out.

Using the Boltzmann equation and assuming that (i) the universe is radiation-dominated (RD) from the time of  $\chi$  freeze-out until the time that DM dominates and (ii) comoving entropy density  $S = sa^3$  is conserved following freeze-out, the relic density of  $\chi$  today can be estimated as

$$\Omega_\chi \equiv \frac{\rho_\chi}{\rho_c} \simeq \frac{2 \times 10^{-10} \text{ GeV}^{-2}}{\langle\sigma|v|\rangle}. \quad (1.44)$$

Assuming an  $s$ -wave annihilation,  $\langle\sigma|v|\rangle \sim \alpha^2/M^2$ , where  $\alpha$  and  $M$  denote the coupling and mass scale for the annihilation, respectively. Then, if  $\chi$  has a weak scale coupling and mass of  $\alpha \sim \mathcal{O}(10^{-2})$  and  $M \sim \mathcal{O}(100 \text{ GeV})$ , it may have a relic abundance today matching the observed DM abundance today  $\Omega_{\text{DM}} \simeq 0.26$  [20]. This is the WIMP miracle—that a mass and interaction scale that is important in other areas of particle physics happens to give the right abundance for thermally-produced DM.

That DM could be related to the weak scale is appealing theoretically due to the idea of naturalness [36] of the Higgs boson mass, which predicts new physics near the weak scale in order to explain why the Higgs is light despite quadratic divergences in the Higgs mass. Indeed, many concrete models to explain naturalness contain WIMP candidates. An example of this is supersymmetry (SUSY). If  $R$ -parity is conserved, then the lightest

supersymmetry particle (LSP) is stable and can be DM. The neutralino of SUSY is predicted to be electrically neutral with weak-scale mass and interaction strength, making it an ideal WIMP DM candidate if it is the LSP.

The WIMP miracle may even indicate that the DM interactions are not only weak-scale, but that the DM is actually charged under the SM electroweak symmetry group. We will discuss models where this is the case in Chapter 2.

Searches for WIMPs are performed at a variety of different types of experiments, including direct detection, indirect detection, and colliders. These experiments rely on variations of the process in (1.42).

- Direct detection searches look for relic WIMPs in the Milky Way halo that scatter off of materials in a laboratory experiment via the interaction  $\chi X \rightarrow \chi X$ . This collision can be observed via the energy it deposits in the detector [37, 38]. Examples of such experiments discussed in this thesis include the liquid-xenon-based experiments XENON [6, 8, 39], LUX [7, 40–42], PandaX [2], and LZ [3].
- Indirect detection experiments search for the annihilation products of relic DM that annihilate today via the interaction  $\chi\bar{\chi} \rightarrow X\bar{X}$  [38]. These interactions are most likely to occur where the DM density is highest today. This thesis will discuss searches for DM that is captured in the Sun or Earth via a collision of the form  $\chi X \rightarrow \chi X$  [43–45]. If the captured DM annihilates into neutrinos, the neutrinos could be detected with large-volume ice or water experiments like IceCube [4] or Super-Kamiokande [46]. Searches are also performed for DM annihilating to photons—either directly or through intermediate states. These tend to focus on the Milky Way galactic center [47, 48], Milky Way dwarf galaxies [49], other nearby galaxies [50, 51], or galactic substructures [52, 53]. Another possibility is that DM annihilates to stable charged particles, however those particles will not point back to their source due to galactic magnetic fields. Nevertheless, searches for cosmic rays can probe such a scenario—indeed, the AMS-02 experiment measured an excess of high-energy positrons [54], which may give a hint

for new physics.

- Collider searches seek to produce DM in collisions of SM particles via the interaction  $X\bar{X} \rightarrow \chi\bar{\chi}$ . The DM itself will not be measured directly by the detector because it is electrically neutral. Rather, it will appear as missing energy, and its presence must be inferred by the measurement of other states. The detailed search strategy depends on the specifics of the DM model, including whether there are other new states predicted by the model that the collider could produce.

### 1.3 Axions

Another dark matter candidate is the axion. The axion [55, 56] was first proposed as a solution to the strong CP problem [57], which we review now. The SM Lagrangian should contain the term

$$\mathcal{L} \supset \theta \frac{g^2}{32\pi^2} G^{a\mu\nu} \tilde{G}_{\mu\nu}^a. \quad (1.45)$$

Such a term is not forbidden by any symmetry—in fact it is generated by quark masses which have the effect of changing  $\theta \rightarrow \bar{\theta} = \theta + \text{Arg}(\det M)$ , where  $M$  is the quark mass matrix which is in general complex-valued with one physical phase [58, 59]. Its presence also solves the  $U(1)_A$  problem [60–62].

As a consequence of this Lagrangian term, the neutron is predicted to have an electric dipole moment [63]:

$$d_n \simeq (5 \times 10^{-16} \text{ e cm}) \bar{\theta}. \quad (1.46)$$

However, experimentally [64],

$$d_n < 2.9 \times 10^{-26} \text{ e cm} \implies \bar{\theta} \lesssim \text{few} \times 10^{-11}. \quad (1.47)$$

This incredibly fine-tuned cancellation between  $\theta$  and  $\text{Arg}(\det M)$ , which in principle have

nothing to do with each other, is known as the strong CP problem.

The axion is the most elegant solution to the strong CP problem (for a review, see, *e.g.*, [65, 66]). It arises as a pseudo-Nambu-Goldstone boson (pNGB) of a spontaneously broken Peccei-Quinn (PQ) symmetry at a scale  $f_a$ , and it couples to quantum chromodynamics (QCD) as:

$$\mathcal{L} \supset C \frac{a}{f_a} \frac{g^2}{32\pi^2} G^{a\mu\nu} \tilde{G}_{\mu\nu}^a, \quad (1.48)$$

where  $a$  is the pseudoscalar axion field and  $C$  is a constant (constructions of the renormalizable theory underlying this non-renormalizable term include the KSVZ [67, 68] and DSFZ [69, 70] models). If its potential is dominated by the QCD instanton effects arising from this term, the axion settles into an approximately CP conserving minimum:

$$\langle a \rangle = -\frac{f_a \bar{\theta}}{C}. \quad (1.49)$$

This exactly cancels the  $\bar{\theta}$  term, solving the strong CP problem.

While it was first proposed to solve the strong CP problem as outlined above, in addition, the axion is an attractive candidate for the dark matter (DM) in our universe [71–73]. Axions are predominantly produced via the misalignment mechanism or topological defects in the early universe. A small population of axions can also be produced thermally, but this is generally subdominant to the other effects [26] and will not be further discussed here.

Both the misalignment mechanism and topological defect production rely on the idea that the axion can take a random field value in the early universe. This occurs when the PQ symmetry is spontaneously broken, at which point different patches of the universe will select from the continuum of degenerate vacua.

If PQ symmetry breaking occurs after the end of inflation, our observable universe will initially consist of many patches with different initial values for the axion field. This will lead to topological defects like cosmic strings and domain walls, both of which produce axions when they decay.

On the other hand, topological defects cannot form if PQ symmetry breaking occurs before the end of inflation. This is because a tiny patch where the axion takes on one initial value before inflation will be exponentially inflated to encompass our entire visible universe. If the axion field is approximately single-valued, no topological defects can exist.

Whether or not topological defects form, axions are (also) formed via the misalignment mechanism. The axion field value in the early universe is not expected to match the CP-conserving minimum of Eq. (1.49) generated by QCD. This is certainly true when PQ symmetry breaks after inflation because the axion field takes on many values randomly throughout the observable universe. It is also true for the other case when PQ symmetry breaks before inflation because there is not reason to expect the single initial field value to randomly coincide with the QCD minimum. Were it not for QCD effects, the axion would remain a Nambu-Goldstone boson (NGB). QCD effects lift the flat axion potential so that

$$V(a) = \left( m_a \frac{f_a}{N_{\text{DW}}} \right)^2 \left( 1 - \cos \left( \frac{a}{f_a/N_{\text{DW}}} \right) \right), \quad (1.50)$$

where  $N_{\text{DW}}$  is the model-dependent integer domain wall number and the axion field has been redefined so that it is minimized at zero. The axion mass is

$$m_a = \begin{cases} m_{\text{QCD}}, & T \lesssim \Lambda_{\text{QCD}}, \\ m_{\text{QCD}} b \left( \frac{\Lambda_{\text{QCD}}}{T} \right)^4, & T \gtrsim \Lambda_{\text{QCD}}, \end{cases} \quad (1.51)$$

with  $b = 0.018$  [74],  $\Lambda_{\text{QCD}} = 200$  MeV the scale of the QCD phase transition, and the zero-temperature QCD axion mass [55]

$$m_{\text{QCD}} = 6.2 \mu\text{eV} \left( \frac{10^{12} \text{ GeV}}{f_a/N_{\text{DW}}} \right). \quad (1.52)$$

For the remainder of this section we will set  $N_{\text{DW}} = 1$ , though it can be recovered by replacing  $f_a$  with  $f_a/N_{\text{DW}}$ .

After the QCD phase transition, the initially displaced axion field oscillates around its

QCD minimum, and an axion energy density results from these coherent oscillations. The energy density in the axion field before it begins coherent oscillations can be approximated by the quadratic term in Eq. (1.50),  $V(\theta) \simeq \frac{1}{2}m_a^2(T)f_a^2\theta^2$ , where  $\theta \equiv a/f_a$ . Thus, once the axion begins oscillating at temperature  $T_f$ , the axion number density due to the misalignment mechanism will be

$$n_a(T_f) = \frac{1}{2}\chi m_a(T_f)f_a^2 \langle \theta_i^2 f(\theta_i) \rangle. \quad (1.53)$$

Here,  $\chi \simeq 1.44$  is a correction for the temperature-dependence of the axion mass [75],  $\theta_i$  is the initial displacement of  $\theta$  away from the CP-conserving minimum, and the brackets denote a spacial average. The anharmonicity function  $f(\theta_i)$  accounts for the fact that the axion potential is not truly quadratic farther away from  $\theta_i = 0$ . Thus, for  $\theta_i$  near zero,  $f(\theta_i) \sim 1$ , while for  $\theta_i$  near  $\pm\pi$ ,  $f(\theta_i)$  diverges. It is given by [76]

$$f(\theta_i) = \left[ \ln \left( \frac{e}{1 - \theta_i^2/\pi^2} \right) \right]^{7/6}. \quad (1.54)$$

The temperature at which the axion field begins oscillating can be calculated using the equation of motion for  $\theta$ . Approximating  $V(\theta)$  as quadratic as above,

$$\ddot{\theta} + 3H(T)\dot{\theta} + m_a^2(T)\theta = 0, \quad (1.55)$$

At large  $T \gg \Lambda_{\text{QCD}}$ ,  $m_a(T) \simeq 0$  so  $\theta = \theta_i$  is constant in time (though it may depend on spatial position). Define the point at which oscillations begin as

$$3H(T_f) = m_a(T_f). \quad (1.56)$$

Assuming the universe is radiation dominated, this yields the result

$$T_f = \begin{cases} 618 \text{ MeV} \left( \frac{10^{12} \text{ GeV}}{f_a} \right)^{1/6}, & T \gtrsim \Lambda_{\text{QCD}}, \\ 68.1 \text{ MeV} \left( \frac{10^{18} \text{ GeV}}{f_a} \right)^{1/2}, & T \lesssim \Lambda_{\text{QCD}}. \end{cases} \quad (1.57)$$

Finally, the comoving number density is conserved following the beginning of coherent oscillations. Thus, assuming no additional entropy injection, the present relic density of axions produced by coherent oscillations can be computed as [76]

$$\Omega_a h^2 = \begin{cases} 0.236 \langle \theta_i^2 f(\theta_i) \rangle \left( \frac{f_a}{10^{12} \text{ GeV}} \right)^{7/6}, & f_a \lesssim \hat{f}_a \\ 0.0051 \langle \theta_i^2 f(\theta_i) \rangle \left( \frac{f_a}{10^{12} \text{ GeV}} \right)^{3/2}, & f_a \gtrsim \hat{f}_a, \end{cases} \quad (1.58)$$

where  $\hat{f}_a = 0.991 \times 10^{17} \text{ GeV}$ .

As described above, the value for  $\langle \theta_i^2 f(\theta_i) \rangle$  depends on whether PQ symmetry breaking occurs before or after the end of inflation. If it occurs after, the value for  $\theta_i$  takes on many random values throughout our Hubble volume. Thus, one should simply average all possible values for  $\theta_i$  weighted evenly, resulting in  $\langle \theta_i^2 f(\theta_i) \rangle = 2.67 \frac{\pi^2}{3}$ .

If PQ symmetry breaking occurs before inflation ends,  $\theta_i$  is nearly uniform in our observable universe. This is true up to any quantum fluctuations in  $\theta$  that arise during inflation. Calling the variance in these fluctuations  $\sigma_\theta^2$ , the spatial average will be  $\langle \theta_i^2 f(\theta_i) \rangle = \theta_i^2 + \sigma_\theta^2$ . Note that if the axion is massless during inflation, which would be the case if only QCD breaks the PQ symmetry to give the axion a mass, the fluctuations are the usual size for a massless field [33, 77–83]:  $\sigma_\theta = \frac{H_I}{2\pi f_I}$ , where  $H_I$  and  $f_I$  are the Hubble parameter and PQ-breaking scale during inflation, respectively.

Unlike the misalignment mechanism, the estimation of the relic abundance of axions due to topological defects must be performed numerically. First, if topological defects form,  $N_{\text{DW}} = 1$  is required so that the domain walls are unstable; otherwise, stable domain walls would overclose the universe. Assuming that  $N_{\text{DW}} = 1$ , the contribution to the axion relic density from topological defects has been estimated by various authors [84–89] in the range of  $\mathcal{O}(0.1)$  to  $\mathcal{O}(100)$  times that of axions produced via misalignment mechanism. To reiterate, topological defects only form if PQ symmetry breaking occurs after inflation.

From these calculations of the axion relic abundance, we see that while the axion may have been originally suggested to solve the strong CP problem, the axion is also capable of

explaining the observed relic abundance of DM  $\Omega_{\text{DM}}h^2 = 0.1199$  [20]. For example, if PQ symmetry breaks before the end of inflation, it is probable that  $\theta_i = \mathcal{O}(1)$ , in which case an axion with PQ breaking scale  $f_a \simeq 10^{12}$  GeV would reproduce the observed DM abundance under the assumptions made above. If PQ symmetry breaks after inflation,  $f_a \simeq 10^9$  to  $10^{11}$  GeV gives  $\Omega_a = \Omega_{\text{DM}}$ , depending on the axion abundance produced by topological defects.

It should also be noted that string theories predict an axion with a PQ breaking scale of order the grand unified theory (GUT) scale,  $f_a \simeq 10^{16}$  GeV [90, 91]. This would require that  $\theta_i \ll 1$ —a tuning in initial conditions, though not as severe as the tuning required by the strong CP problem without the axion—or that some mechanism dilutes the axion abundance. For example, a late period of reheating would dilute the axion abundance and thus allow  $\theta_i = \mathcal{O}(1)$  with  $f_a \simeq 10^{15}$  GeV [92], close to the GUT scale. Alternatively, the tuning in  $\theta_i$  can be explained by an anthropic argument [93, 94]. We will return to these possibilities in Chapter 3.

It is in principle possible to have several axion-like particles (ALPs). These share many of the properties of the axion, but they do not couple to  $G\tilde{G}$  as in (1.48) or they receive another contribution to their mass and thus do not solve the strong CP problem. Thus, the mass and symmetry-breaking scale of an ALP are arbitrary, unlike the QCD axion for which the two are related by Eq. (1.52). ALPs are predicted by some BSM models (see, *e.g.*, [95] and references therein).

There are various astrophysical bounds on the axion and ALPs. Observations of white dwarf cooling and supernovae limit the production of QCD axions in these systems, resulting in the bound  $f_a \gtrsim 4 \times 10^8$  GeV [96]. Bounds from black hole superradiance require  $f_a \lesssim 2 \times 10^{17}$  GeV [97].<sup>1</sup> Additionally, if PQ symmetry breaking occurs before inflation ends, the quantum fluctuations of the axion discussed above will result in isocurvature perturbations in the CMB, which are bounded by Planck [20] and may pose a severe constraint on axion

---

<sup>1</sup>In applying either of these bounds to ALPs, one must take into account independent changes to the ALP mass and PQ-breaking scale.

models depending on the scale of inflation. These isocurvature perturbations will be explored further in Chapter 3.

Some astrophysical observations have also provided “hints” for the existence of axions or ALPs. The universe appears anomalously transparent to gamma-rays at large optical depth  $\tau \gtrsim 2$  [98–102], which may be explained by photon-ALP oscillations [103–108]. An X-ray excess in the Coma cluster could be explained by ALP background radiation, with the hint pointing towards a region of parameter space that overlaps with the gamma-ray transparency hint [109, 110]. There are also hints from the cooling of white dwarfs and red giants that indicate another stellar energy-loss mechanism may be necessary, possibly indicating an ALP or low- $f_a$  QCD axion near the bounds presently set by stellar cooling [111–114].

In addition to astrophysical probes, there are several existing and proposed experiments to search for the axion or ALPs directly. Some experiments look for axions or ALPs as relic DM. The ADMX experiment searches for axion DM using a resonant cavity. The data it has taken so far exclude axion DM within a narrow band around  $f_a \sim 10^{12}$  GeV with future plans to broaden the exclusion band [115, 116]. Other experiments have been proposed or are under development to look for axion DM with larger  $f_a$  by probing the axion coherent oscillations using nuclear spins [11], SQUIDs [117], atomic transitions [118], or electromagnetism [119, 120].

Other current or future experiments do not rely on the axion making up all of the observed DM abundance. However, these experiments are sensitive only to ALPs and not the QCD axion. The experiments include IAXO, which will search for solar axions [121], ALPS-II, which is a “light-shining-through-wall” experiment searching for photon-ALP-photon oscillations [122], and CMB observatories PIXIE [123] and PRISM [124].

## 1.4 Primordial Black Holes

It is possible that some or all of the observed DM abundance is not particulate in nature but rather comes from massive compact halo objects (MACHOs). A MACHO candidate is the primordial black hole (PBH) [125, 126]. Initially formed in the early universe, PBHs must have masses  $M_{\text{PBH}} \gtrsim 5 \times 10^{-19} M_{\odot}$ , where  $M_{\odot}$  is a solar mass, in order to have not evaporated before today except for the possibility of Planck-mass PBH relics. While not the focus here, PBHs may also seed supermassive black holes in galactic nuclei.

PBHs are formed from local matter overdensities that are seeded during the early universe. These overdensities may arise during inflation resulting from quantum fluctuations, or they may come about sometime after inflation, for example due to preheating or a curvaton. When inflationary overdensities reenter the horizon or when later overdensities are formed, if the overdensities are large enough, their gravitational attraction can overcome Hubble expansion and collapse, forming a black hole. The threshold value in the overdensity  $\delta \equiv \delta\rho/\rho$  for PBH formation is often quoted as  $\delta_c = 1/3$  [127], though higher values of *e.g.*  $\delta_c \simeq 0.45$  have also been indicated [128–130], and non-sphericity effects [130–133] can make  $\delta_c$  higher still. The mass of the resulting PBH is roughly given by the energy within the Hubble volume containing the overdensity. For an overdensity reentering the horizon at scale  $k_f$ , this horizon mass is

$$M_{\text{PBH}} = \frac{4\pi}{3} \rho_r H^{-3} \simeq 10 M_{\odot} \left( \frac{g_*}{100} \right)^{-1/6} \left( \frac{\text{pc}^{-1}}{k_f} \right)^2, \quad (1.59)$$

where the universe is assumed to be radiation dominated with energy density  $\rho_r$  (see Eq. (1.12)).

Interest in PBHs has been renewed by the LIGO detection of gravitational waves (GWs) resulting from the merger of black holes with masses  $\sim 10$  to  $30 M_{\odot}$  [134, 135]. The merger rate observed by LIGO may be consistent with primordial black holes (PBHs) making up all [136] or a fraction  $\sim 10^{-3}$  [12, 137] of the observed dark matter (DM) density.

Observational constraints on PBH DM come from gamma-rays associated with BH evap-

oration, X-rays associated with BH accretion, gravitational lensing, and their effects on stellar, small-scale, and large-scale structure. See [126] and references therein for bounds across the entire PBH mass range. Since we will be interested predominantly in PBHs at masses relevant to LIGO, we will here summarize the bounds on PBHs at these scales. PBHs within this mass range are subject to several constraints, especially if the fractional DM relic abundance  $f = \Omega_{\text{PBH}}/\Omega_{\text{DM}} \gtrsim 0.1$ . Microlensing measurements [138–140] constrain black hole masses  $M_{\text{PBH}} \lesssim 30M_{\odot}$ , with the strongest constraints coming at  $M_{\text{PBH}} \lesssim M_{\odot}$ . Constraints from the ultra-faint dwarf Eridanus II [141] give complementary bounds for  $M_{\text{PBH}} \gtrsim 5$  to  $30M_{\odot}$ , depending on assumptions about the density and velocity dispersion of DM. These two constraints, when taken together, may allow a monochromatic PBH mass spectrum at  $M_{\text{PBH}} \simeq 30M_{\odot}$  with  $f = \Omega_{\text{PBH}}/\Omega_{\text{DM}}$  close to one or allow an extended distribution with  $f \lesssim 0.1$  over a wide range of masses around this. Bounds from these experiments on extended mass spectra are discussed in [126, 142]. Bounds may also be placed on PBHs with mass  $M_{\text{PBH}} \sim 1$  to  $1000M_{\odot}$  moving relative to pulsar lines of sight [143]. In addition to these, there are strong bounds from WMAP and FIRAS [144] and Planck [145] that may limit  $f \lesssim 10^{-2}$  to  $10^{-4}$  around these PBH masses, though these bounds have been disputed [136]. Such a low abundance, nevertheless, may be consistent with the observed rate at LIGO if the higher merger rate estimates such as those given in Refs. [12, 137] obtain.

If the black holes detected by LIGO were produced primordially, they would have resulted from the collapse of large density perturbations as described above. While these would be sourced as scalar perturbations, because of their size they can lead to non-trivial tensor perturbations at second order in cosmological perturbation theory [146–151]. A key observation is PBHs in the mass range detected by LIGO generate tensor perturbations that may be detected as gravitational waves at pulsar timing arrays (PTAs) [152]. This will be the topic of Chapter 4.

## 1.5 Motivation and Outline

The central idea for this thesis is the interplay between cosmology and models of DM. An important question is how the early universe must behave in order for certain models of DM to explain present-day observations. In many cases, the history of the universe can be constrained via astrophysical observation, leading to constraints on DM models. Also relevant is how present and future experiments can constrain DM models; examples of this were discussed in the previous subsections. In this thesis, these topics are considered within the context of WIMP, axion, and PBH models.

WIMP models are the subject of Chapter 2, which is based on [153] and [154], written in collaboration with James Halverson, John Kearney, and Aaron Pierce. In particular, we consider WIMPs that have SM electroweak charges. These models communicate directly with the SM, so direct detection searches can be quite constraining, necessitating changes to the structure of the models or to cosmological history. As we discuss, the presence of additional states other than the DM particle is often important for the cosmology and the experimental signatures of such models.

We address these issues in the context of DM interacting with the SM primarily through an axial-vector coupling to the  $Z$ . We show that the simplest EFT realization of this coupling is already constrained, but a UV completion like the singlet-doublet model can alleviate present bounds and even evade future searches due to the presence of additional particles. We also discuss how a particle interacting via a vector-coupling to the  $Z$  could have interesting experimental signatures even if it makes up only a small portion of the total DM abundance. In this case, the small abundance can be achieved by modified dynamics due to new particles present in the early universe.

Chapter 3 focuses on QCD axion models of DM. Of particular interest are constraints on the quantum fluctuations of the axion field that arise during inflation and lead to isocurvature perturbations in the CMB. In order to evade these constraints and allow for both a large scale of inflation and axion DM with  $f_a > H_I/2\pi$ , the axion must have a non-trivial

cosmological history that affects its dynamics during inflation. Our work revisits models that have previously been proposed to accomplish this. We find that models that rely on a larger scale of PQ symmetry breaking during inflation generally do not work. We also show that while models that rely on explicit PQ-symmetry breaking can be constructed, they require finely tuned Lagrangian parameters. This analysis is from [155], written in collaboration with John Kearney and Aaron Pierce.

In Chapter 4, based on work with Aaron Pierce and James Wells [156], models of PBH production are discussed to explain the LIGO detections of black hole binary mergers. For PBH production to occur, large perturbations must form during or shortly after inflation that will subsequently collapse to black holes. These large perturbations can lead to gravitational waves at second order in cosmological perturbation theory (SGWs). Pulsar timing arrays could be sensitive to the SGWs associated with the production of PBHs with masses relevant to LIGO. We study how the SGW signature depends on the model that gave rise to the PBH-forming perturbations, focusing on models that can be explicitly constructed. We show that inflationary models with relatively broad primordial perturbation spectra are already constrained by non-observation of SGWs at PTAs, while models with narrow—but still physically realizable—spectra will soon be probed. We also give an example of a model for PBH formation for which PTAs will not be sensitive.

Finally, Chapter 5 contains concluding remarks and an outlook for future study.

## Chapter 2

# Models of Dark Matter that Couple to Electroweak Mediators

*This chapter was completed in collaboration with James Halverson, John Kearney, and Aaron Pierce [153, 154].*

Weakly-interacting massive particles (WIMPs) remain an attractive thermal dark matter (DM) candidate. However, while WIMPs exhibit weak scale interactions, the precise mechanism through which the DM interacts with visible matter (beyond its gravitational interactions) is unknown. One possibility is to take the “W” in WIMP seriously. That is, the interactions with the Standard Model (SM) are mediated not just by particles with masses near the weak scale, but by the carriers of the weak force: the  $W$ ,  $Z$ , and Higgs ( $h$ ) bosons. It is of interest to understand the current experimental status of such models, as they represent minimal set-ups and give insight into the extent to which the WIMP paradigm is being probed.

Direct detection experiments place bounds on the spin-independent (SI) couplings of such WIMPs, which at tree-level arise from exchange of the  $h$  or  $Z$ , and their spin-dependent (SD) couplings, which at tree-level arise from exchange of the  $Z$ . The latest bounds on SI scattering arise from PandaX [157] and LUX [42]. DM that interacts with the  $Z$  boson via

vectorial couplings,

$$\mathcal{L} \supset g_V(\bar{\chi}\gamma_\mu\chi)Z^\mu, \quad (2.1)$$

is very strongly constrained, see, *e.g.*, [158]. For  $g_V \sim g_Z \equiv g_2/(2\cos\theta_W)$ , such a  $\chi$  can comprise only  $\lesssim 10^{-6}$  of the DM. This may be dealt with in two ways.

The first will be the topic of Sections 2.1 and 2.2, based on [154]. The dangerous interaction in (2.1) can be neatly forbidden by positing that the DM is a Majorana fermion, for which  $\bar{\chi}\gamma^\mu\chi$  vanishes identically. This is the case, for example, for the neutralino of the Minimal Supersymmetric Standard Model (MSSM). Majorana fermions can retain direct detection cross sections that appear at an interesting level either via SD couplings to the  $Z$  boson and/or SI interactions with the Higgs boson.

We pay special attention to regions of parameter space where the DM thermal relic abundance matches the value measured by the Planck Experiment  $\Omega_{DM}h^2 = 0.1198(26)$  [1]. Direct detection experiments place stringent bounds on the DM-Higgs coupling  $\mathcal{L} \ni y_{\chi\chi h}(\bar{\chi}\chi h)$ ,  $y_{\chi\chi h} \lesssim 7 \times 10^{-3}(m_\chi/50 \text{ GeV})^{1/2}$  for  $m_\chi \gtrsim 50 \text{ GeV}$  [42, 159]. These bounds make it difficult to realize the thermal abundance solely via a Higgs boson coupling, *i.e.*, “Higgs portal” DM is constrained—see, *e.g.*, [160]. It is therefore natural to consider the possibility where the thermal abundance is obtained absent a large coupling to the Higgs boson—perhaps solely via coupling to the  $Z$ , which induces only SD scattering. However, as the xenon (Xe) nuclei of LUX and PandaX have spin, direct detection experiments also probe this scenario.<sup>1</sup> As we will show, Majorana DM with thermal history primarily determined by  $Z$  couplings is being probed now.

In Section 2.1, we first discuss the simplest, gauge-invariant DM model wherein a Majorana fermion interacts with a  $Z$  boson. We will see this coupling generically induces a large contribution to the  $\rho$  parameter. We then discuss in Section 2.2 how  $Z$ -mediated dark matter may be realized as a limit of the singlet-doublet model [161]. This model has no

---

<sup>1</sup>Additional bounds result from the lack of observation of neutrinos in the IceCube detector [4], as produced via solar DM capture and subsequent annihilation, though these are generally weaker.

problem with the  $\rho$  parameter. However, while direct detection is primarily mediated via  $Z$  boson exchange, other couplings may be important for the dark matter’s thermal history. We will demonstrate that a large region of parameter space in this model is close to being probed, even if the couplings to the Higgs boson that determine the SI interactions vanish.

The second possibility to deal with the interaction in (2.1) is to suppose that the particle makes up only a tiny fraction of the observed DM abundance. This is discussed in Section 2.3, based on [153]. Since it is only necessary to forbid this operator if the relic makes up the entirety of the dark matter, this relic may comprise a minuscule fraction of the dark matter, but its enormous direct detection scattering cross section can lead to an interesting signal (see [162] for related work on detecting a subdominant component of the DM in the context of the MSSM). We will see dilution by the necessary amount is possible in simple cosmologies. We briefly discuss ways in which such a particle—which could be the first discovery at direct detection experiments—might be disentangled from the dominant dark matter using information from both colliders and direct detection.

For both of these possibilities, a natural question is what mass scale might we expect for these particles. This question can be approached from a phenomenological or top-down model building perspective. Starting with the phenomenological, as discussed in Section 1.2, the WIMP miracle would lead us to expect weak scale masses of order hundreds of GeV to TeV because the interactions of the DM are weak-scale, by virtue of their coupling to the electroweak gauge groups. Indeed, this scale is further supported by experimental data. As we will see in Section 2.1, masses smaller than about half the  $Z$  mass are experimentally excluded by measurements of the  $Z$  invisible decay width. Thus, only DM masses at or above the weak scale need be considered.

Now consider the top-down perspective. For concreteness, consider the addition of a vectorlike pair of doublets  $X, \bar{X}$  to the minimal supersymmetric standard model (MSSM). Some mechanism must generate a mass  $\mu$  for the Higgsinos of the MSSM. Whether this is the Giudice-Masiero mechanism [163], the vacuum expectation value of a singlet (as in the next-

to minimal supersymmetric standard model [164–166]), or a D-brane instanton [167–169], it is plausible that whatever generates  $\mu$  would also generate a mass  $\mu_X$  at the same scale. Naturalness arguments then indicate a  $\mu \sim \mu_X \sim 100$  GeV – few TeV. This mass range will thus be our main focus in this chapter.

## 2.1 $Z$ -Mediated Dark Matter

A gauge invariant coupling of a Majorana fermion  $\chi$  to the  $Z$  may be generated via a higher-dimension operator involving the Higgs doublet  $H$

$$\mathcal{L} \supset \frac{c}{2\Lambda^2} (iH^\dagger D_\mu H + \text{h.c.}) \bar{\chi} \gamma^\mu \gamma^5 \chi, \quad (2.2)$$

where  $c$  is a coupling constant and  $\Lambda$  is the effective scale for new physics. We have implemented this model in `Micromegas v4.3.1` [170], which we use for calculations of relic density and direct detection processes. This operator induces a coupling to the  $Z$  boson (using  $\langle H \rangle = \frac{v}{\sqrt{2}}$ ,  $v = 246$  GeV)

$$\mathcal{L} \supset -\frac{g_2}{4c_W} \frac{cv^2}{\Lambda^2} Z_\mu \bar{\chi} \gamma^\mu \gamma^5 \chi \equiv -\frac{g_2}{2c_W} g_A Z_\mu \bar{\chi} \gamma^\mu \gamma^5 \chi, \quad (2.3)$$

in addition to four- and five-point interactions between the DM and the  $Z$  and  $h$  bosons with related strength. In terms of  $m_\chi$  and  $g_A$ , one can calculate the relic density and direct detection rate. The dominant direct detection signal is spin dependent through the nucleon effective operator  $\bar{\chi} \gamma_\mu \gamma^5 \chi N \gamma^\mu \gamma^5 N$ . Other operators are velocity-suppressed [171, 172].

Fermionic DM with purely axial vector coupling to the  $Z$  was studied in [173], whose results for the relic density we have reproduced, and more recently in [174], whose results are in agreement with ours.<sup>2</sup> However, the model considered in these papers is not gauge invariant. The gauge invariant version was studied in [175], with closely related work in [176].

---

<sup>2</sup>A discrepancy with the relic density calculation appearing in the original version of [174] has since been resolved in a later version.

Our results appear consistent with [176], but our relic density calculation and direct detection limits differ from [175].

In Fig. 2.1, we show (black line) the value of  $g_A$  that reproduces the observed thermal relic density as a function of DM mass  $m_\chi$ . Also shown (dot-dashed orange curve) is the required value if one were to introduce the coupling in (2.3) without the attendant  $\chi\chi Zh$  or  $\chi\chi Zhh$  couplings, thereby violating gauge invariance (analytic results for this case are given in [173]). Without the  $\chi\chi Zh$  coupling, the annihilation to  $Zh$  grows more rapidly as a function of  $\sqrt{s}$ . Similarly, shown in dotted orange is the relic density calculation if only  $2 \rightarrow 2$  annihilations are considered. This neglects the  $2 \rightarrow 3$  annihilation to  $Zhh$ , which becomes important at large  $m_\chi$  where the cross section’s mass dependence outweighs the phase space suppression.<sup>3</sup>

The blue shaded region is excluded by PandaX bounds on SD scattering [2], assuming  $\chi$  makes up all of the DM. This excludes thermal relics with  $m_\chi \lesssim 200$  GeV from making up all of the DM aside from the resonance at  $m_\chi = m_Z/2$ .<sup>4</sup> Incidentally, the region above this curve with  $m_\chi \lesssim 200$  GeV is excluded even for thermally-produced  $\chi$  making up only a portion of dark matter because the bound scales as  $\sigma_{\text{SD}}\Omega \propto \sigma_{\text{SD}}/\langle\sigma v\rangle \propto g_A^2/g_A^2$ . Projected bounds from LZ [3] are also shown (blue dashed). Since projections for SD scattering bounds are not given by LZ, we estimate them by rescaling LUX results from [41] by the same factor as SI bounds improve from [40]—which has the same exposure as [41]—to the projected LZ SI bounds [3]. LZ will probe thermal relics up to  $m_\chi \lesssim 2$  TeV.

IceCube bounds [4] on annihilations of DM captured in the sun by the  $\chi$ -proton SD cross section are shown in purple. To account for multiple annihilation final states, we estimate

$$\sigma_{\text{SD,p,bound}} = \left( \sum_{i=\text{channels}} \frac{\text{Br}_i}{\sigma_i} \right)^{-1}, \quad (2.4)$$

<sup>3</sup>To our knowledge, no other studies have considered these  $2 \rightarrow 3$  annihilations.

<sup>4</sup>Currently published SD bounds from LUX [41] are slightly less sensitive than PandaX. However, Ref. [41] does not include the full exposure of the most recent bound on SI scattering published by LUX [42]. A rescaling of the bounds in [41] including the longer exposure of [42] would exclude thermal relics with  $m_\chi \lesssim 240$  GeV, slightly stronger than the PandaX bound.

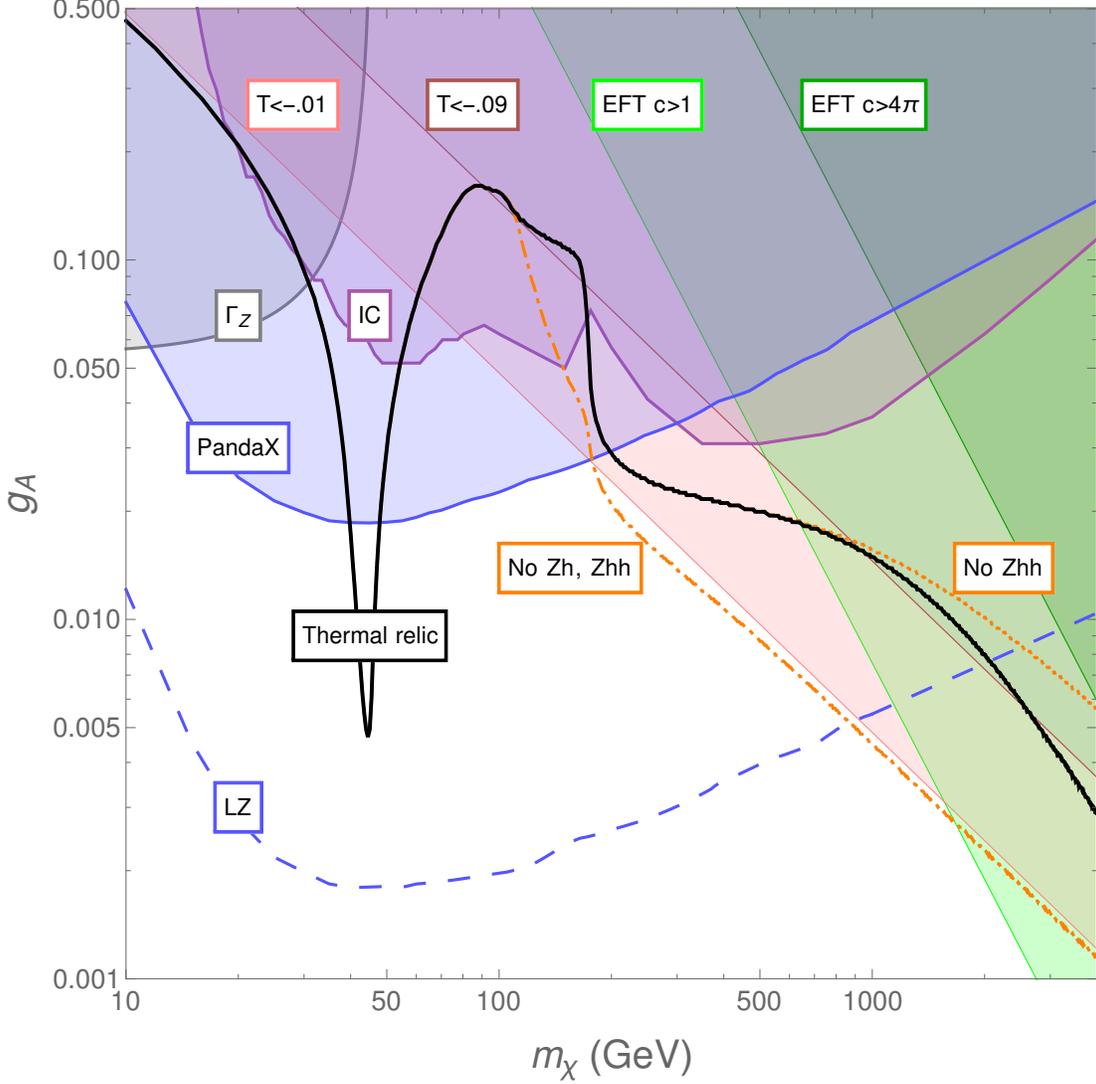


Figure 2.1: Constraints on  $g_A \equiv cv^2/(2\Lambda^2)$  as a function of  $m_\chi$  for  $Z$ -mediated DM. The black solid line indicates where the thermal relic abundance matches the observed abundance [1]. The orange dotted (dot-dashed) line corresponds to omitting higher dimensional couplings to  $Zhh$  (and  $Zh$ ), in violation of gauge invariance. Shaded (dashed) blue is excluded by PandaX [2] (LZ [3]), with bounds obtained as described in the text. The purple shaded region is excluded by IceCube (IC) [4]. Invisible  $Z$  decay limits ( $\Gamma_Z$ ) [5] are shown in gray. Pink shaded regions indicate where  $T < -0.09$  (upper, darker) and  $T < -0.01$  (lower, lighter), corresponding to  $2\sigma$  excluded regions depending if  $S$  is nonzero or zero, respectively. Green shaded regions indicate where the EFT is precarious, taking  $c_{\max} = 4\pi$  (upper, darker) or 1 (lower, lighter).

where  $\text{Br}_i$  is the branching ratio to channel  $i$  and  $\sigma_i$  is the IceCube bound assuming 100% branching ratio to channel  $i$ . At DM masses below the  $W$  mass, annihilations to neutri-

nos and taus give the strongest bounds, while intermediate masses are most constrained by annihilations to tops and higher masses  $\gtrsim$  TeV by  $Zhh$  annihilations.<sup>5</sup> Although the approximation (2.4) is not precise, we view it as suitable, particularly as sensitivities of Xe-based experiments and IceCube are currently only competitive at higher  $m_\chi \gtrsim$  few hundred GeV, and Xe-based limits will soon dominate. Constraints from searches for gamma-ray signals from DM annihilation [177] are similarly subdominant.

For  $m_\chi < m_Z/2$ ,  $g_A$  is bounded by LEP measurements of the invisible  $Z$  width, which limits  $\Gamma(Z \rightarrow \chi\chi) < 2$  MeV at 95% confidence level [5]. This bound is shown in gray using

$$\Gamma(Z \rightarrow \chi\chi) = \frac{m_Z}{6\pi} \left( g_A \frac{g_2}{2c_W} \right)^2 \left( 1 - \frac{4m_\chi^2}{m_Z^2} \right)^{3/2}. \quad (2.5)$$

The LHC can probe larger DM masses than LEP with monojet-type searches, however both present and future sensitivities will be subdominant to LEP or direct detection bounds [175], with the background ( $Z \rightarrow \nu\nu$ ) + jet representing an important irreducible background.

While pure  $Z$ -mediated DM currently evades the above experimental constraints at larger masses, there are other considerations that should be taken into account when evaluating whether this is a reasonable benchmark model. First, the coupling of Eq. (2.3) will generate a large contribution to the  $\rho$  parameter. At loop level, two insertions of the operator in Eq. (2.2) generate a contribution to the self-energy of the  $Z$  boson

$$\delta\mathcal{L} \supset \frac{c^2 m_\chi^2}{\pi^2 \Lambda^4} \log\left(\frac{\Lambda}{m_\chi}\right) |H^\dagger D_\mu H|^2, \quad (2.6)$$

without a corresponding contribution to the  $W$  boson self-energy. This gives

$$\delta\rho = -\frac{c^2 m_\chi^2 v^2}{2\pi^2 \Lambda^4} \log\left(\frac{\Lambda}{m_\chi}\right). \quad (2.7)$$

---

<sup>5</sup>We take limits from  $Zhh$  to be roughly comparable to those from  $ZZ$  and  $hh$ . While the neutrinos from a three-body final state will be less energetic, for large  $m_\chi$  the energies are still expected to be above threshold. As such, the presence of additional neutrinos should lead to comparable (if not stronger) limits.

Under the strong assumption that no other operators affecting electroweak precision physics are generated,  $T = \delta\rho/\alpha(M_Z) > -0.01$  at  $2\sigma$  [178]. The corresponding constraint is shown in Fig. 2.1 as a light pink shaded region, taking the logarithm in Eq. (2.7) to be unity. If, however, non-trivial contributions to  $S$  are simultaneously allowed (but  $U = 0$ ),  $T > -0.09$  at  $2\sigma$ . This constraint is shown in dark pink. It is clear there is tension between precision electroweak constraints and obtaining a thermal history in this model, particularly at high masses.<sup>6</sup>

Another question is whether the model of Eq. (2.3) is a valid effective field theory (EFT) for describing DM annihilations in the early universe. The relevant scale for these annihilations is  $2m_\chi$ . Without appreciable separation between  $\Lambda$  and  $m_\chi$  we expect higher dimension operators to be important. To illustrate regions where the EFT is precarious, we take

$$g_A \lesssim 0.095 \left( \frac{c_{\max}}{4\pi} \right) \left( \frac{\text{TeV}}{m_\chi} \right)^2 \left( \frac{2m_\chi}{\Lambda_{\min}} \right)^2, \quad (2.8)$$

where for illustrative purposes we have set  $\Lambda_{\min} = 2m_\chi$ . In the figure, we have shown two regions: a light green one where we have set  $c_{\max} = 1$ , and a dark green one where we have allowed  $c_{\max} = 4\pi$ . At large  $m_\chi$ , describing the physics with an EFT becomes more difficult.

There are two main take-aways from this section. First, current direct detection bounds constrain thermal  $Z$ -mediated dark matter  $\lesssim 200$  GeV aside from a tiny window where annihilations are resonant. For a recent discussion of possibilities of probing this region, see [179]. Increased sensitivity by next-generation experiments will probe higher masses near the limit of validity for the EFT. Second, this coupling of the DM to the  $Z$  boson maximally breaks custodial  $SU(2)$ . As such, there is tension with constraints on the  $\rho$  parameter for the entirety of the thermal relic space with  $m_\chi > m_Z/2$ , except perhaps at very large DM mass where the validity of the EFT is questionable.

---

<sup>6</sup>Writing down  $|H^\dagger D_\mu H|^2$  directly, with  $\Lambda$  suppression comparable to that in Eq. (2.3), would be even worse. The estimate of Eq. (2.7) corresponds to the idea there is an approximate custodial  $SU(2)$  symmetry, broken only via the DM- $Z$  coupling.

## 2.2 Embedding $Z$ -Mediated Dark Matter in the Singlet-Doublet Model

A simple embedding of this  $Z$ -mediated DM model that moves beyond an EFT, drastically lessens the tension with the  $\rho$  parameter, and is consistent with approximate gauge coupling unification is the singlet-doublet model [161]. Early analysis of this model appeared in [180–182], with focus on the direct detection phenomenology in [183]. It is closely related to the DM story in split supersymmetry [184, 185] or the well-tempered neutralino [186]. More recent studies of the DM phenomenology appear in [159, 187, 188]. Related collider studies appear in [189, 190].

The singlet-doublet model is obtained by adding to the Standard Model a vectorlike pair of electroweak doublets  $D$  and  $D^c$  with hypercharge  $Y = \pm\frac{1}{2}$  and an electroweak singlet  $N$  with  $Y = 0$ . The relevant interactions in the Lagrangian are

$$\mathcal{L} \supset -yDHN - y^c D^c \tilde{H}N - M_D D D^c - \frac{M_N}{2} N^2 + \text{h.c.} \quad (2.9)$$

The Yukawa couplings generate mixing between  $N$  and the electromagnetically-neutral components of  $D, D^c$ , giving rise to three Majorana fermions, the lightest of which is a DM candidate. Because the DM is descended in part from an  $SU(2)$  doublet, it will couple not only to the  $Z$ , but also to the  $W$  boson. This generates a correction to the  $W$  self-energy, which contributes to mitigating the constraints from  $\rho$ , but also generically affects the early universe cosmology.

The Majorana nature of the DM ensures that it does not exhibit vectorial couplings to the  $Z$  boson, avoiding contributions to SI scattering that would be far in excess of current limits. Thus, the coupling to the  $Z$  is of the same form as the right hand side of Eq. (2.3),

but with the coupling determined by

$$g_A = \frac{1}{2} \frac{\Delta^2 v^2 (y^{c2} - y^2)}{\Delta^2 + v^2 ((y^2 + y^{c2})(M_D^2 + m_\chi^2) + 4yy^c M_D m_\chi)}, \quad (2.10)$$

where  $\Delta^2 = M_D^2 - m_\chi^2$  and  $m_\chi$  is the DM mass determined by the mixing of  $N$  and the neutral states in  $D$  and  $D^c$ .

However, bounds on SI scattering are sufficiently strong that they also constrain DM that interacts via the Higgs boson. As such, it is of particular interest to consider this model in the so-called ‘‘Higgs blind spot’’ [183, 187, 191], wherein the coupling to the Higgs boson vanishes

$$y_{\text{BS}}^c = -y \frac{M_N}{M_D} \left( 1 \pm \sqrt{1 - \left( \frac{M_N}{M_D} \right)^2} \right)^{-1}. \quad (2.11)$$

In this blind spot, the DM will retain a diagonal coupling to the  $Z$  boson (as in the previous section) but will also exhibit off-diagonal couplings to the  $Z$  as well as to the  $W$  boson. So, while the DM phenomenology in certain regions of the singlet-doublet parameter space will correspond to that of the  $Z$ -mediated case, these additional couplings can play a significant role elsewhere.

In Fig. 2.2, we have fixed two of the four free parameters of the model as follows:  $y^c$  is fixed to the Higgs blind spot value, so that  $\sigma_{\text{SI}}$  vanishes at tree level, and  $M_D$  is fixed to agree with the observed (thermal) relic density calculated using `Micromegas`.<sup>7</sup> We specialize to the regime  $-|y| < y^c < 0$ , which corresponds to taking the plus solution in Eq. (2.11)—the sign choice is simply a manifestation of the fact that the physics is left invariant by the exchange  $y \leftrightarrow y^c$ . For instance, if we were to write  $\cos \theta = \frac{M_N}{M_D}$ , Eq. (2.11) could be rewritten as

$$\frac{y_{\text{BS}}^c}{y} = -\sqrt{\frac{1 \mp \sin \theta}{1 \pm \sin \theta}}. \quad (2.12)$$

Choosing the opposite sign would reproduce the plot with  $y$  and  $y^c$  exchanged.

---

<sup>7</sup>There is one physical phase among the parameters  $\{y, y^c, M_D, M_N\}$ , which for simplicity we set to zero. Effects of a non-zero phase are discussed in [180, 181].

Several quantities of interest are then plotted as a function of the remaining two free parameters,  $y$  and the DM mass  $m_\chi$  ( $= M_N$  in the blind spot, see, *e.g.*, [183, 188]). The shaded red region is excluded by PandaX bounds on  $\sigma_{\text{SD}}$ . Also shown are projected bounds from Xenon1T [6] and LZ, which if no detection is made would exclude the regions to the right of each line. Blue contours represent  $M_D$ , or equivalently the mass of the charged state. We do not display bounds from IceCube because at present LUX and PandaX provide the strongest constraints throughout the parameter space shown, and as noted in the previous section direct detection experiments will scale in sensitivity much faster than IceCube.

The gray shaded region to the right of the plot indicates where the DM- $Z$  coupling is  $\gtrsim 0.95$  of the value prescribed in the pure  $Z$ -mediated case. Here, the DM cosmology is well-described by the simple  $Z$ -mediated model, but the additional states nearby in mass that fill out complete electroweak representations provide targets for collider searches and render contributions to the  $T$  parameter small, as we now detail.

LHC searches for charginos and neutralinos decaying via electroweak bosons could probe this model. However, current limits [192–195] are mild and do not appear in this region of parameter space. Future limits likely will—see, *e.g.*, [196].

Far into the gray region (at large  $y$  beyond what is plotted),  $\Delta T$  is similar in size to the expectation from the  $Z$ -mediated model above. However, there are additional comparably-sized contributions from the doublet, as the splitting within the doublet is related to the DM- $Z$  coupling. Partial cancellation between these contributions leads to a value somewhat smaller than the naïve expectation. For instance, in the limit  $M_D \gg m_\chi, yv$ ,

$$\delta\rho = \frac{y^4 v^2}{48\pi^2 M_D^2} \left\{ 1 + \frac{17m_\chi^2}{4M_D^2} - \frac{6m_\chi^2}{M_D^2} \log\left(\frac{M_D}{m_\chi}\right) \right\}. \quad (2.13)$$

For larger  $m_\chi/M_D$  (smaller  $y$ ), higher-order terms are relevant and result in further suppression. As such,  $|\Delta T| \lesssim 2 \times 10^{-3}$  throughout Fig. 2.2.

To the left of the plot, away from where the  $Z$ -mediated description is sufficient,  $M_D$  is not

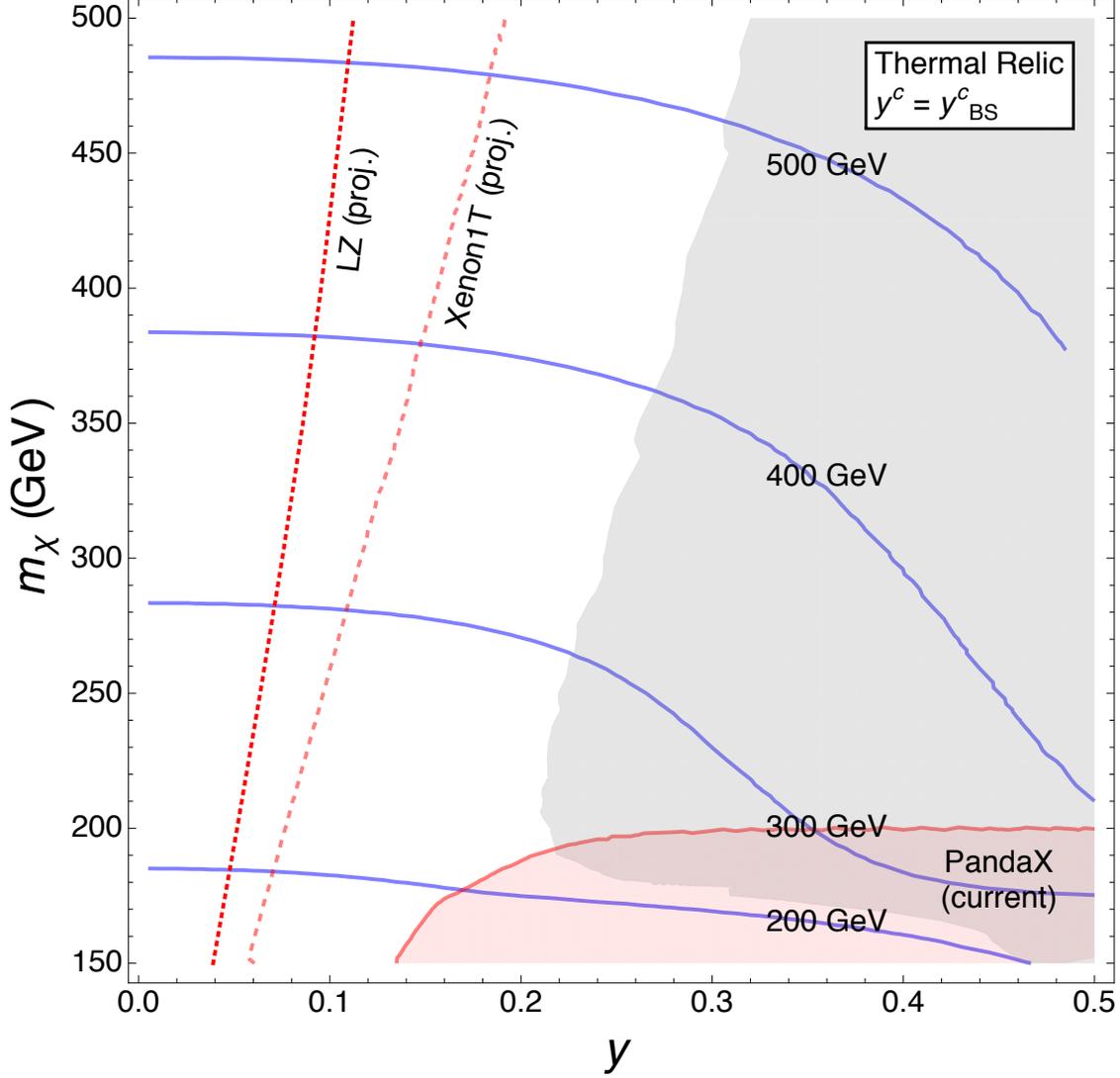


Figure 2.2: Constraints on the singlet-doublet model at the Higgs blind spot. The blind spot condition fixes  $y^c$  via (2.11), while  $M_D$  (contours in blue) is set by demanding the thermal relic abundance matches the observed DM abundance. Shown in red are current bounds from PandaX [2] (solid, shaded) and prospective bounds from Xenon1T [6] (thick dashed) and LZ [3] (thin dashed) on  $\sigma_{SD}$ . In the shaded gray region, the DM- $Z$  is at least 95% of that in the  $Z$ -mediated model of the previous section.

too much larger than  $m_\chi$ , such that coannihilation and  $t$ -channel annihilation to  $WW$  become increasingly important in determining the relic density. Interestingly, future direct searches will be able to probe a sizable portion of this regime. Though future direct detection will have trouble constraining the entire region, and the model will continue to evade these constraints for sufficiently small couplings, collider searches may provide a promising alternative probe.

For instance, the small mass splittings make this region of parameter space susceptible to searches such as [195] based on soft leptons.

So, while the pure  $Z$ -mediated model is a good proxy for the singlet-doublet model in the blind spot at large  $y$  (where direct detection constraints may be directly translated), the reach of future experiments extends well beyond this regime to smaller  $y$ . Even if the  $Z$ -mediated model of the preceding section is excluded up to a given  $m_\chi$ , this model presents a minimal variation in which that DM mass remains viable, and yet meaningful constraints can still be achieved. As such, the singlet-doublet model in the Higgs blind spot represents a worthy target for future WIMP searches, and it would be valuable for experiments to quote constraints in terms of this parameter space.

Finally, we comment on the effect of tuning away from the exact blind spot. This will result in a non-zero DM- $h$  coupling that, as mentioned previously, can result in strong constraints from SI scattering—in fact, this is the case even if the DM- $h$  coupling is sufficiently small to have a negligible impact on the DM thermal history. Parameterizing the deviation from the blind spot as

$$\delta_{y^c} = \frac{y^c}{y_{\text{BS}}^c} - 1, \quad (2.14)$$

we show in Fig. 2.3 the parameter space for  $\delta_{y^c} = -0.3$ . Note that, for  $y^c \neq y_{\text{BS}}^c$ ,  $m_\chi \neq M_N$ , so here we plot with respect to  $\{y, M_N\}$ —however, for the values of  $y$  and  $\delta_{y^c}$  considered,  $m_\chi \simeq M_N$ .

To the left, where  $t$ -channel and coannihilation play a significant role in determining the relic density, changing  $y^c$  simply changes the DM- $Z$  and DM- $h$  couplings, altering the exact values of  $\sigma_{\text{SD,SI}}$  (and hence the future experimental reach) relative to Fig. 2.2 in this region of parameter space. To the right however, where the relic density is predominantly determined by  $s$ -channel  $Z$ -exchange,  $M_D$  changes to compensate the change in  $y^c$  and maintain approximately the same DM- $Z$  coupling as in Fig. 2.2. As a result, the current SD exclusions (and the region well-described by the pure  $Z$ -mediated model) do not change significantly. While it is not plotted, we note that the parameter space is not symmetric

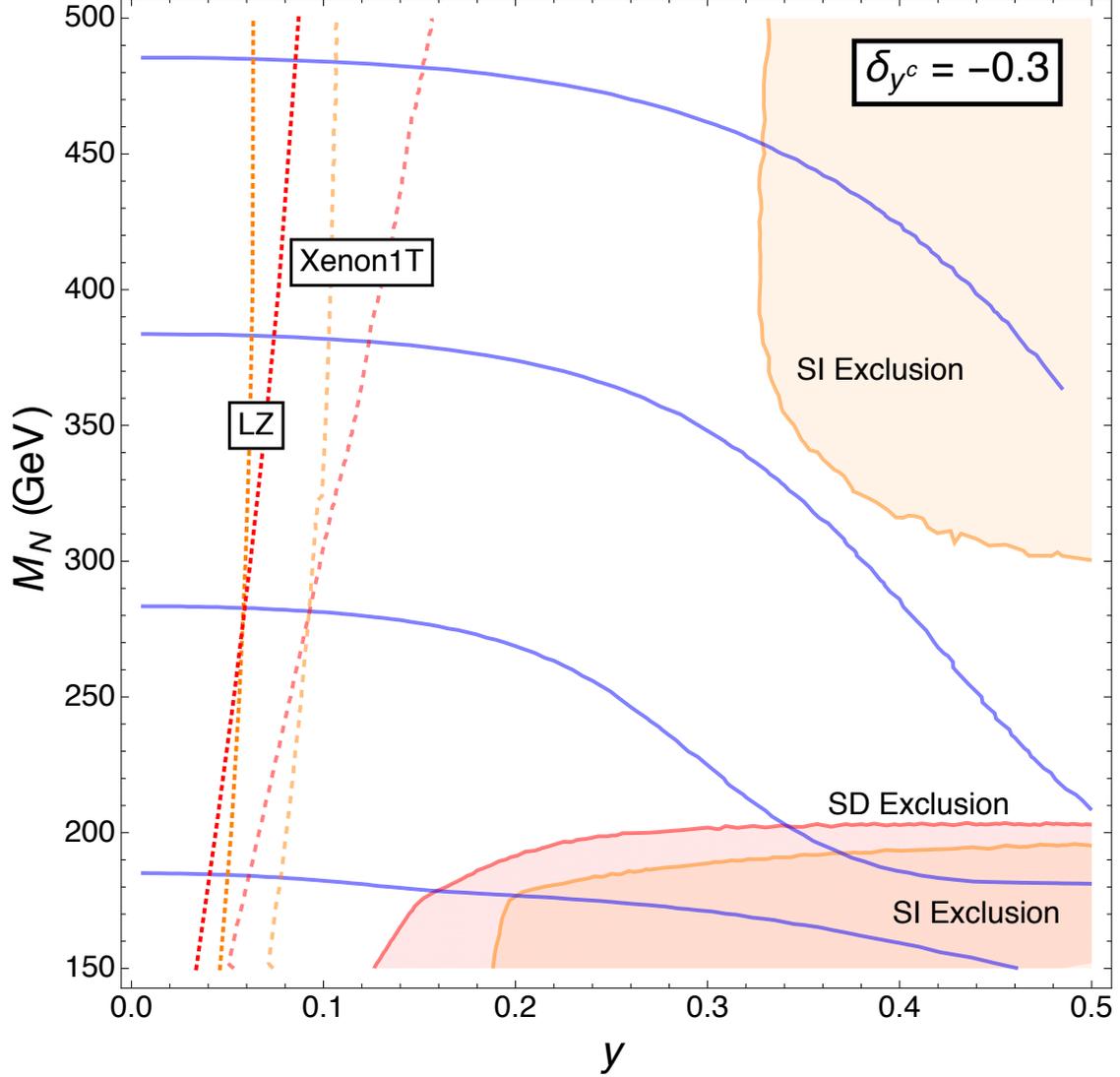


Figure 2.3: Similar to Fig. 2.2, but with  $y^c$  deviating from the Higgs blind spot by  $\delta_{y^c} = -0.3$ , see Eq. (2.14). Mirroring the previous figure, contours of  $M_D$  are shown in blue. Current and projected limits on  $\sigma_{SD}$  ( $\sigma_{SI}$ ) are shown in red (orange).

about  $y_{BS}^c$ .

The two disjoint regions at larger  $y \gtrsim 0.35$  are related to the top quark threshold. For  $m_\chi \gtrsim m_t$ , the observed relic abundance is achieved for a smaller DM- $Z$  coupling, which is accompanied by a similar suppression of the DM- $h$  coupling. This allows the model to evade present SI limits near threshold ( $m_\chi \gtrsim 190$  GeV). However, while the relic density constraint largely fixes the DM- $Z$  coupling at larger  $M_N$ , the DM- $h$  coupling exhibits different parametric dependence and modestly increases with  $M_N$ . Eventually this increase results in a

second excluded region for  $m_\chi \gtrsim 300$  GeV. For  $\delta_{y^c} = -0.3$ ,  $\sigma_{\text{SI}}$  for  $200 \text{ GeV} \lesssim M_N \lesssim 300 \text{ GeV}$  is just below current LUX limits while, for larger  $\delta_{y^c}$ , this gap does not appear.

For the value of  $\delta_{y^c}$  shown, constraints from SI and SD scattering are complementary today, excluding slightly different regions of parameter space, and are comparable in the future. This point was chosen specifically to show where the future constraints may be roughly similar.<sup>8</sup> For larger  $\delta_{y^c}$ , SI constraints rapidly dominate, *e.g.*, already excluding much of the parameter space for  $|\delta_{y^c}| = 0.5$  while still not significantly altering the thermal history. For  $|\delta_{y^c}| \lesssim 0.1$ , SD constraints dominate throughout the parameter space. But, in all cases, an order of magnitude improvement in limits would require the model to lie squarely in the small coupling/coannihilation regime, though the exact regions of parameter space probed will depend on the proximity of  $y^c$  to the blind spot value.

Incidentally, there are two notable regions of singlet-doublet parameter space that provide a thermal DM candidate via coannihilation but with suppressed direct detection cross sections. First, in the Higgs blind spot discussed here, there is a “double blind spot” where both of the DM- $h$  and DM- $Z$  couplings vanish. This occurs at  $M_N \simeq M_D \simeq 880$  GeV (with a slight  $y \simeq y^c$  dependence). In this case, there is a single Majorana “Higgsino” degenerate with the chargino. The second interesting case is that of the nearly pure doublet “Higgsino” when  $M_N \gg M_D$ . The neutral components of the doublets are split into a pseudo-Dirac state, suppressing direct detection cross sections, and the observed relic abundance is obtained for  $M_D \simeq 1.1$  TeV.

---

<sup>8</sup>Because the SI and SD constraints will be comparable in the future for the  $\delta_{y^c}$  in Fig. 2.3, the exclusions should be combined to yield a somewhat stronger bound on  $y$ . We have chosen not to do so in order to demonstrate the relative strength of each, and anyway for most values of  $\delta_{y^c}$  only one of the SI or SD constraints will dominate.

## 2.3 Vectorlike Leptons as the Tip of the Dark Matter Iceberg

The previous sections have taken the approach that forbids the vector  $Z$ -mediated operator relevant for SI scattering  $\mathcal{O}_Z = (\bar{q}\gamma_\mu q)(\bar{X}\gamma^\mu X)$  by making the dark matter Majorana. However, if  $X$  comprises only a tiny fraction of the dark matter, it will not be forbidden by present experimental data. Indeed, in such a case, its enormous direct detection scattering cross section can lead to an interesting signal (see [162] for related work on detecting a subdominant component of the DM in the context of the MSSM). We will see dilution by the necessary amount is possible in simple cosmologies. We briefly discuss ways in which such a particle—which could be the first discovery at direct detection experiments—might be disentangled from the dominant dark matter using information from both colliders and direct detection.

As a concrete example, consider the addition of a vectorlike pair of doublets  $X, \bar{X}$ . This is similar to the singlet-doublet model of Section 2.2, except without the singlet. If  $X, \bar{X}$  have an unbroken  $X \rightarrow -X$  symmetry, the Dirac fermion will comprise a component of the dark matter. Because it is Dirac, it has full-strength direct detection cross section per nucleon, e.g. [37, 158]:

$$\sigma \approx \frac{G_F^2}{2\pi} \mu_{XN}^2 \frac{1}{A^2} [(1 - 4 \sin^2 \theta_W)Z - (A - Z)]^2, \quad (2.15)$$

where  $\mu_{XN}$  is the dark matter-nucleon reduced mass.

A potential motivation for novel vectorlike doublets arises via string theory. There Ref. [197] has shown that there exist additional constraints on the chiral spectrum of  $SU(2)$  gauge theories which ensure anomaly cancellation in nucleated D-brane theories. These constraints go beyond standard anomaly cancellation in the  $SU(2)$  theories and can require the existence of electroweak exotics; see [198, 199] for particle physics implications. In weakly

coupled orientifold compactifications, doublet quantum numbers for the exotics are a likely possibility. The exotic states arise from open strings, which selects out  $SU(2)$  singlets, doublets, and triplets as the only possibilities. If one further requires that one end of the open string ends on a D-brane corresponding to a novel symmetry (perhaps related to the stability of the  $X$ ), then doublet quantum numbers are uniquely selected. A string scale mass term for the doublets is forbidden by symmetry, rendering their presence at the TeV scale even more plausible. If the low energy theory is the Standard Model plus these exotics, gauge couplings approximately unify at  $10^{14}$  GeV [161, 180].

In this section we will work under the assumption that a new Dirac  $SU(2)$  doublet  $X$  exists with weak-scale mass. We address implications for direct detection experiments and sketch how the dilution necessary to bring it into compliance with experimental bounds might be accomplished in a simple non-thermal cosmology. Finally, we discuss prospects for probing such a doublet at the LHC.

Note that experimental results in this section are current at the time of publication of [153], although newer results like the ones used in the previous sections are now available that tighten the constraints presented here.

### 2.3.1 Direct Detection and Cosmology

We have used `Micromegas` [200] to calculate the spin-independent cross section of  $X$  and verified that it is consistent with Eq. (2.15). We also used it to calculate the  $X$  relic abundance assuming a standard thermal freeze-out. It is well approximated by  $\Omega_X h^2 \simeq 0.1 \left(\frac{\mu}{1 \text{ TeV}}\right)^2$  [201]. An  $X$  produced with a standard thermal history is well excluded by current direct detection bounds. To evade current bounds from LUX [7],  $X$  with  $\mu_X = 100$  GeV (1 TeV) must have a tiny relic density  $\Omega_X/\Omega_{\text{cdm}} \lesssim 5 \times 10^{-7}$  ( $4 \times 10^{-6}$ ), where  $\Omega_{\text{cdm}} h^2 = 0.1199 \pm 0.0027$  [1].

One possibility is to simply declare a smaller relic abundance by fiat. Indeed, we could imagine that there is thermal freeze-out with a subsequent dilution by e.g. late time inflation. Interestingly, however, the maximum dilution is limited if the baryon number is generated

before this dilution. Even if the baryon asymmetry proceeds by an extraordinarily efficient mechanism like Affleck-Dine (for a review see [202]), where the baryon to photon ratio could be as large as  $\mathcal{O}(1)$ , consistency with the current ratio imposes a maximum dilution factor of  $10^9$  [203]. Then, the dark matter densities would range from  $\Omega_{\text{dil}}h^2 \approx 10^{-12} - 10^{-10}$  for  $\mu_X = 100 - 1000$  GeV. But given the large direct detection cross sections, it should be possible to probe relic abundances of  $\Omega_{\text{dil}}h^2 \approx 5 \times 10^{-11} - 10^{-11}$  without running afoul of the neutrino background [204]. A 1 ton Xe experiment might be sensitive to relic densities perhaps a hundred times these; therefore, it is possible to almost completely probe this scenario of arbitrary dilution. We explore a perhaps better motivated possibility below, where we discuss a more concrete cosmology. In that case, the relic abundance is expected to be less diluted, and therefore the likelihood of direct detection is even greater.

This model gives a characteristic material dependence at direct detection experiments. The ratio of measured cross section per nucleon at experiments composed of Xenon, Germanium, and Argon would be  $1 : 0.89 : 0.86$ . Observing the deviation of these ratios from unity will be challenging but would be powerful evidence for this scenario. Also, if the mass is close to 100 GeV, it is possible to make a determination of the  $X$  mass via an examination of the recoil spectrum, e.g., [205]. This mass could then be correlated with collider discoveries, see below.

### **Non-Thermal Production via Modulus Decay**

The late decay of a scalar field  $\phi$  can modify the dark matter relic abundance. This occurs if the energy density of the universe becomes  $\phi$ -dominated until the time of  $\phi$  decay, which can then both produce dark matter and provide substantial entropy generation as it reheats the universe to a temperature  $T_{\text{RH}}$ . Such cosmologies are well-motivated in string compactifications, which typically contain many light scalar fields in the form of stabilized moduli. Another possible motivation is supersymmetric axion models—the saxion could play the role of  $\phi$  and the axion and lightest neutralino or axino could make up (some or all of) the

remaining dark matter [206].

$T_{\text{RH}}$  is model dependent, but it is bounded by phenomenological requirements. First, to ensure that the successful predictions of big bang nucleosynthesis (BBN) are not spoiled,  $T_{\text{RH}} \gtrsim T_{\text{BBN}} \simeq 5 \text{ MeV}$  [207, 208]. Second, to accommodate an alternative production mechanism, it must be below the thermal freeze-out temperature,  $T_{\text{RH}} \lesssim T_{\text{fo}} \simeq \mu_X/20$ . We will see that consistency with direct detection bounds will place further limits on  $T_{\text{RH}}$ .

$T_{\text{RH}}$  is determined by the decay rate  $\Gamma_\phi = c_\phi^2 m_\phi^3/M_P^2$  (perhaps arising from an operator like  $\phi G\tilde{G}$ ). Under the assumption that  $\phi$  decay and the subsequent thermalization are instantaneous:

$$T_{\text{RH}} = \left[ \left( \frac{8}{90} \pi^3 g_\star \right)^{-1/2} M_P \Gamma_\phi \right]^{1/2} \simeq 10 c_\phi \left( \frac{m_\phi}{100 \text{ TeV}} \right)^{3/2} \text{ MeV}, \quad (2.16)$$

where  $c_\phi$  is a presumably  $\mathcal{O}(1)$  constant computable in specific models.

The  $X$  relic abundance  $\Omega_X h^2$  depends on  $T_{\text{RH}}$  and  $b$ , the number of dark matter particles produced per  $\phi$  decay. The Boltzmann equation in (1.43) is modified to include the contribution from  $\phi$  in the following way [207, 208]:

$$\frac{dn_X}{dt} + 3Hn_X = -\langle \sigma |v| \rangle \left( n_X^2 - \left( n_X^{\text{EQ}} \right)^2 \right) + \frac{b}{m_\phi} \Gamma_\phi \rho_\phi, \quad (2.17)$$

$$\frac{d\rho_\phi}{dt} + 3H\rho_\phi = -\Gamma_\phi \rho_\phi, \quad (2.18)$$

$$\frac{ds}{dt} + 3Hs = \frac{\Gamma_\phi \rho_\phi}{T}. \quad (2.19)$$

The first equation now includes a term for decays of  $\phi$  to  $X$ . Likewise, the results in Section 1.2 assumed the right hand side of the last equation was zero so that comoving entropy  $S = sa^3$  was conserved. Note that only the combination  $b/m_\phi$  appears in these equations; accordingly, we will employ the dimensionless parameter  $\eta \equiv b(100 \text{ TeV}/m_\phi)$ .

If the  $\phi$  branching ratio to  $X$  is very small or zero, the  $X$  relic density is set by thermal

production and freeze-out followed by its dilution via the entropy produced in  $\phi$  decays. Since the  $X$  interaction cross section is large enough to reach chemical equilibrium prior to freeze-out, the (diluted) thermal relic density is parametrized by [207–209],

$$\Omega_X \simeq \frac{T_{\text{RH}}^3 T_{\text{fo}}}{(T_{\text{fo}}^{\text{new}})^4} \Omega_{\text{std}} \simeq \left( \frac{T_{\text{RH}}}{T_{\text{fo}}} \right)^4 \Omega_{\text{std}}, \quad (\eta \text{ tiny}), \quad (2.20)$$

where  $\Omega_{\text{std}}$  is the  $X$  relic abundance assuming a standard thermal history (i.e.,  $T_{\text{RH}} > T_{\text{fo}}$ ).

On the other hand,  $\eta$  may be large enough for non-thermal production to dominate over thermal production. Since direct detection bounds require that the relic abundance be much less than the standard thermal abundance,  $\eta$  must nonetheless be small. Thus, non-thermal production will not be compensated by annihilations. In this regime [207, 208],

$$\frac{\Omega_X}{\Omega_{\text{cdm}}} \simeq 2 \times 10^3 \eta \left( \frac{\mu_X}{100 \text{ GeV}} \right) \left( \frac{T_{\text{RH}}}{\text{MeV}} \right), \quad (\eta \text{ small}). \quad (2.21)$$

Using the sum of expressions (2.20) and (2.21) (a good approximation to the numerical solution in this small- $\eta$  regime when  $T_{\text{RH}}$  is not too close to  $T_{\text{fo}}$ ) and assuming  $T_{\text{fo}} \simeq \mu_X/25$  (in good agreement with `Micromegas` and with numerical solutions in [207, 208]), the relic density for  $\mu_X = 100 \text{ GeV}$  is plotted as a function of  $T_{\text{RH}}$  for various values of  $\eta$  in Figure 2.4, see also Ref. [207, 208]. Also shown are bounds from LUX [7] and BBN as well as prospective bounds from a ton-scale Xe experiment [8].

Figure 2.4 indicates that bounds from LUX require  $\eta \lesssim 10^{-10}$ . This is approximately true for any value of  $\mu_X > 100 \text{ GeV}$  because the LUX bound on  $\sigma_{\text{SI}} \propto \mu_X$ ,  $\Omega_X \propto \eta \mu_X$  from (2.21), and the cross section (2.15) is approximately constant with respect to  $\mu_X$ .

Such small values of  $\eta$  require that the Yukawa coupling which determines  $\Gamma_{\phi \rightarrow X\bar{X}}$  is very small. Is it reasonable to expect such suppression? If  $\phi$  is uncharged under the Peccei-Quinn-like symmetry protecting  $\mu_X$ , then the bare operator  $\phi X\bar{X}$  is forbidden. However, the effect which generates  $\mu_X$  will typically also give rise to an effective Yukawa coupling which is  $\frac{\mu_X}{M_P}$  suppressed. For example, if  $\mu_X$  is generated via singlet expectation value via a

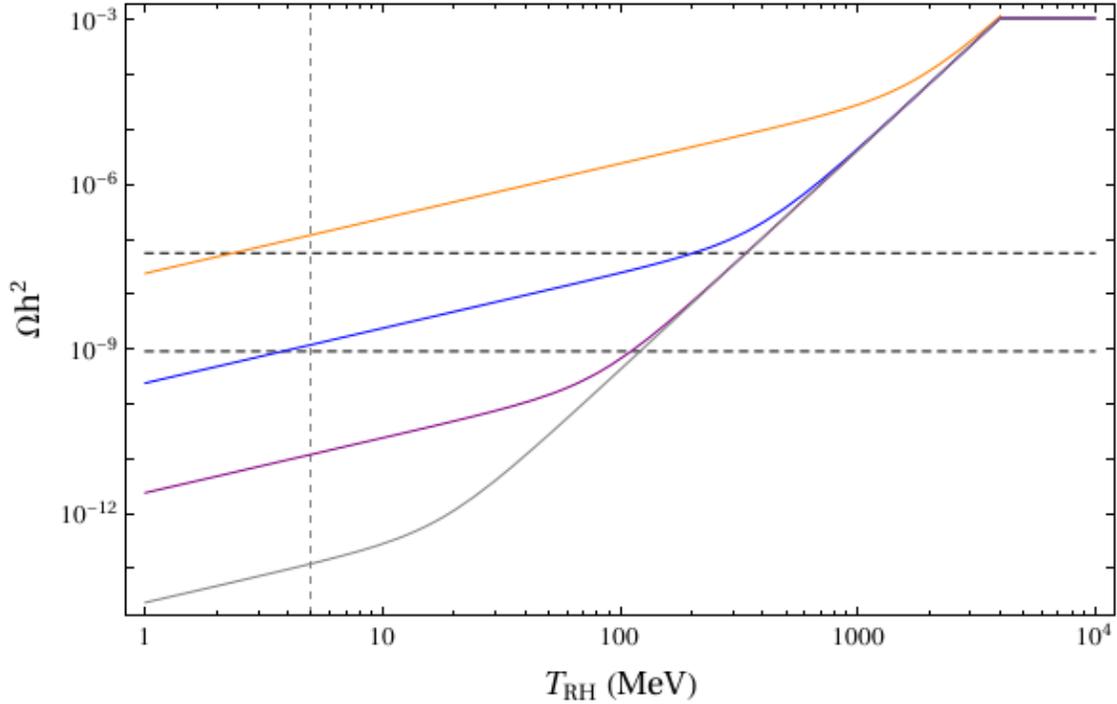


Figure 2.4: Relic density as a function of  $T_{\text{RH}}$  for various values of  $\eta$  for  $\mu_X = 100$  GeV. The solid lines correspond to, from top to bottom,  $\eta = 10^{-10}$  (orange),  $10^{-12}$  (blue),  $10^{-14}$  (purple), and  $10^{-16}$  (gray). The horizontal dashed lines correspond to the maximum allowed relic density to evade LUX [7] (top) and prospective XelT [8] (bottom) bounds. The vertical dashed line represents the cutoff of allowable  $T_{\text{RH}}$  due to BBN.

coupling  $sX\bar{X}$ , then the invariant coupling  $\frac{1}{M_P}\phi sX\bar{X}$  gives an effective Yukawa  $\frac{\langle s \rangle}{M_P}\phi X\bar{X}$  with  $\langle s \rangle \simeq \mu_X$ .<sup>9</sup> We accordingly parameterize the effective Yukawa coupling as  $c_X \frac{\mu_X}{M_P}$ , giving,

$$\Gamma_{\phi XX} = \frac{m_\phi}{4\pi} \left( c_X \frac{\mu_X}{M_P} \right)^2 \Rightarrow b = \frac{1}{2\pi} \left( \frac{c_X \mu_X}{c_\phi m_\phi} \right)^2. \quad (2.22)$$

Here,  $\Gamma_{\phi XX}$  is the width to both  $X^0\bar{X}^0$  and  $X^+X^-$ .

The present constraints on  $c_X$  are shown in Figure 2.5. Using Eqs. (2.21) and (2.22):

$$\frac{\Omega_X}{\Omega_{\text{cdm}}} \simeq \frac{c_X^2}{10\pi} \left( \frac{\mu_X}{100 \text{ GeV}} \right)^3 \left( \frac{\text{MeV}}{T_{\text{RH}}} \right). \quad (2.23)$$

<sup>9</sup>Similar statements can be made about operators arising from non-renormalizable Kahler potential terms e.g.  $\phi^\dagger X\bar{X}$  after using the equation of motion for  $X$ .

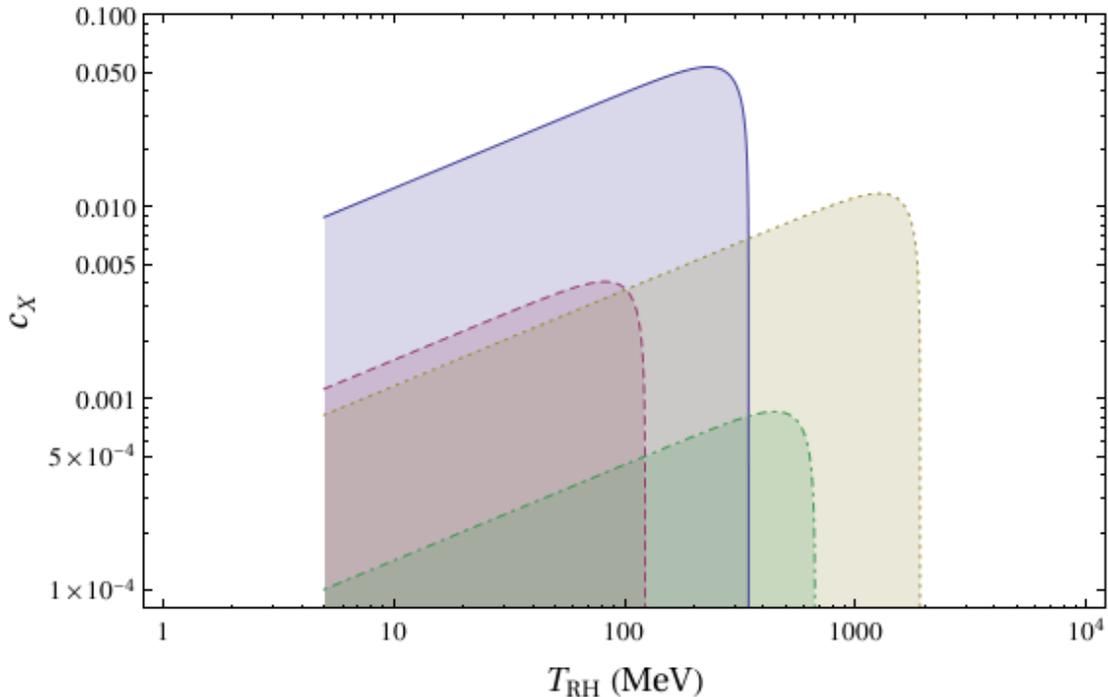


Figure 2.5: Allowed regions (shaded) of  $c_X$  for  $\mu_X = 100$  GeV from LUX [7] (blue solid) and prospective Xe1T [8] (red dashed) bounds and for  $\mu_X = 1$  TeV from LUX (yellow dotted) and prospective Xe1T (green dot-dashed) bounds.

This equation determines the upper boundary of the shaded regions. The left boundary is set by BBN:  $T_{\text{RH}} \gtrsim T_{\text{BBN}} \simeq 5$  MeV. The right boundary occurs when Eq. (2.20) is sufficient to violate (present or future) direct detection bounds.

Absent additional model building, we expect a number of  $\mathcal{O}(1)$  contributions to  $c_X$ , so the stringent direct detection bounds already necessitate tuning at the few percent level. The most favored  $c_X$  region is the one where the relic lies just outside current bounds in Figure 2.5.

The above assumes a single late-decaying modulus. However, string compactifications often contain  $\mathcal{O}(100)$  moduli. For the case of many moduli, the details of the evolution of the dark matter density depend on the initial abundances of the moduli. The single-modulus case is a good approximation to the many-moduli case under the assumption that one modulus  $\phi_i$  dominates the energy density of the universe before it decays, and that no other modulus comes to dominate the energy density any time after  $\phi_i$  decays. Otherwise,

several moduli may contribute to the non-thermal production of dark matter. Additionally, decays of the later moduli might dilute contributions from earlier moduli.

### 2.3.2 Collider Phenomenology

The strongest current collider bound on the  $SU(2)$  doublet comes from LEP2, which bounds  $\mu_X \gtrsim 95$  GeV [210–213]. We present the prospects for LHC14 in the monojet and disappearing track channels.<sup>10</sup>

#### Monojet + $\cancel{E}_T$

The mass splitting between the  $X^\pm$  and  $X^0$  particles for mass  $\mu_X = 100$  (200) GeV is  $\delta m \simeq 256$  (295) MeV. This is a finite, calculable effect due to electroweak symmetry breaking [215]. With these splittings, the  $X^\pm$  decay promptly into  $X^0 +$  invisibly soft  $\pi^\pm$ ,  $e^\pm\nu$ , or  $\mu^\pm\nu$ . Charged and neutral doublet particles will appear as missing energy in the detector. The most recent results for this channel are from ATLAS [216] and CMS [217].

To estimate signal and background, we use MadGraph5 [218], pass to Pythia [219] for MLM matching [220, 221], showering, and hadronization, and use PGS [222] for detector simulation, using an anti- $k_t$  jet clustering algorithm with  $R = .5$ . Simulated parton-level events include one or two jets. We simulate the dominant backgrounds  $j(j) + (Z \rightarrow \nu\nu, W \rightarrow l\nu, \text{ or } W \rightarrow \tau\nu)$ , with  $l = e, \mu$ ; the signal has  $j(j) + (X^0\bar{X}^0, X^+X^0, X^-\bar{X}^0, \text{ or } X^+X^-)$ . Following [216], we apply the following cuts:

$$(i) \quad p_T(j_1) > p_T^{\text{cut}} \text{ and } |\eta(j_1)| < 2$$

$$(ii) \quad \cancel{E}_T > p_T^{\text{cut}}$$

$$(iii) \quad \text{No more than 2 jets with } p_T > 30 \text{ GeV and } |\eta| < 4.5$$

$$(iv) \quad \Delta\phi(j_2, \cancel{E}_T) > 0.5$$

---

<sup>10</sup>Ref. [214] studies prospects for the vector boson fusion (VBF) channel but concludes that it is not sensitive to a pure Higgsino.

	Signal ( $\mu_X$ in GeV)		Backgrounds		
	$\mu_X = 100$	$\mu_X = 150$	$B_{j\nu\nu}$	$B_{jl\nu}$	$B_{j\tau\nu}$
Cross section (fb)	5.78	3.28	136	28.0	29.7
$\chi$ ( $\beta_{\text{tot}} = 0.03$ )	1.0	0.6			
$\chi$ ( $\beta_{\text{tot}} = 0.01$ )	3.0	1.7			

Table 2.1: Cross sections of backgrounds and signal (following cuts (i) – (vi) in the text) and signal significances for the monojet +  $\cancel{E}_T$  channel with  $\sqrt{s} = 14$  TeV,  $\mathcal{L} = 3000 \text{ fb}^{-1}$ , and  $p_T^{\text{cut}} = 500$  GeV.

(v) Lepton vetoes:  $p_T(e) > 20$  GeV and  $|\eta(e)| < 2.47$ ,  $p_T(\mu) > 7$  GeV and  $|\eta(\mu)| < 2.5$ , or  $p_T(\tau) > 20$  GeV and  $|\eta(\tau)| < 2.3$

(vi) Veto on b-jets

The significance can be found using

$$\chi^2 = \frac{S^2}{S + B + \sigma_B^2}. \quad (2.24)$$

Its square root gives the significance. We parameterize the background uncertainty as  $\sigma_B = \beta_{\text{tot}} B$ , remaining agnostic about where the uncertainties originate.<sup>11</sup>

Taking  $\mathcal{L} = 3000 \text{ fb}^{-1}$ , we present significances in Table 2.1 assuming that  $\beta_{\text{tot}}$  can be made to be either 0.03 or 0.01 at  $p_T^{\text{cut}} = 500$  GeV. For comparison, ATLAS and CMS have background uncertainties of  $\beta_{\text{tot}} \simeq 0.04$  and 0.03, respectively, for cuts in present monojet analyses which have a comparable number of background events as our projected  $p_T^{\text{cut}}$  and luminosity. We note that *if* such a small uncertainty could be maintained at larger  $p_T^{\text{cut}}$ , the significance could be modestly increased.

Based on our analysis, we conclude the discovery sensitivity of this channel to a  $SU(2)$  doublet appears weak. Our signal and background cross sections are in rough agreement with [224–227], although our estimated backgrounds tend to be a little smaller and our

<sup>11</sup>Another approach is to estimate the use of data-driven uncertainties to reduce the background uncertainty as in [223]. We found this approach tended to underestimate the uncertainties when applied to current experimental data [216, 217], and does not substantively change the conclusions.

signals a little larger. Thus, we reach the same general conclusions as [224–226] that  $5\sigma$  detection is unlikely in this channel, while a small mass window may be excluded at  $2\sigma$ .

## Disappearing Track

Because of the mass splitting discussed in the previous subsection, the path length of the  $X^\pm$  in its own rest frame for mass  $\mu_X = 100$  (200) GeV is a modest  $c\tau = 1.93$  (1.19) cm [215]. While these path lengths are difficult to detect, it is possible that some of the particles in the tail of the lifetime distribution might be observed if the production rate is sufficiently high. Thus, low masses may be accessible to future disappearing track studies that search for  $X^\pm$  before they decay. The most recent results from the ATLAS experiment can be found in [228].

Following the cuts in [228], to obtain an estimate for the expected signal, we use MadGraph to simulate  $pp \rightarrow j + (X^+X^0, X^-\bar{X}^0, \text{ or } X^+X^-)$  at parton level, stipulating that  $p_T(j) > 90$  GeV and  $|\eta(j)| < 5$ . Then, in each event with  $p_T(j_1) > p_T^{\text{cut}}$  (to be varied), we select the  $X^\pm$  with  $0.1 < |\eta^{\text{track}}| < 1.9$ ,  $p_T^{\text{track}} > 500$  GeV, and  $p_T^{\text{track}} < 1000$  GeV at  $\sqrt{s} = 8$  TeV (to match what is done in ATLAS) and 1500 GeV at  $\sqrt{s} = 14$  TeV. Next, using the known lifetime, we calculate the probability that each passing  $X^\pm$  would achieve a transverse length of at least 30 cm before decay, corresponding to the beginning of the first SCT layer in the ATLAS detector. As alluded to above,  $X^\pm$  that reach the SCT are either highly boosted and/or are in the tail of the lifetime distribution. We assume the efficiency for detection after these cuts is 100%.

Comparing to current limits at  $\sqrt{s} = 8$  TeV, for  $\mu_X = 100$  GeV and  $p_T^{\text{cut}} = 200$  GeV, we estimate  $\sigma_{\text{vis}} = 0.27$  fb. This is just below the ATLAS 95% exclusion of  $\sigma_{\text{vis}} < 0.44$  fb (smaller  $p_T^{\text{track}}$  cuts in [228] set weaker bounds).

To make projections for LHC14, we must estimate the background. A reliable estimate is difficult, as the dominant background (see Figure 5 of [228]) is from mismeasured tracks.

We parameterize the background at  $\sqrt{s} = 14$  TeV and luminosity  $\mathcal{L}$  as

$$B = B_{8\text{ TeV}} \left( \frac{\mathcal{L}}{\mathcal{L}_{8\text{ TeV}}} \right) \left( \frac{\sigma_{14\text{ TeV}}}{\sigma_{8\text{ TeV}}} \right) \left( \frac{\epsilon_{p_T(j_1) > p_T^{\text{cut}}}}{\epsilon_{p_T(j_1) > 90\text{ GeV}}} \right) P_{\text{mis}}, \quad (2.25)$$

where  $\mathcal{L}_{8\text{ TeV}} = 20.3\text{ fb}^{-1}$  is the luminosity in [228],  $B_{8\text{ TeV}}$  is the estimated background in [228],  $\sigma_{14\text{ TeV}}/\sigma_{8\text{ TeV}}$  accounts for the increased cross section of the background as collision energy increases (leaving all cuts constant),  $\epsilon_{p_T(j_1) > p_T^{\text{cut}}}/\epsilon_{p_T(j_1) > 90\text{ GeV}}$  accounts for a cut intended to reduce background, and  $P_{\text{mis}}$  parameterizes the potential that the probability for mismeasured tracks may be greater with increased energy and pile-up.

We obtain  $B_{8\text{ TeV}}$  by integrating the background in Figure 5 of Ref. [228] from the  $p_T^{\text{track}}$  cut up to 1500 GeV. We approximate  $\sigma_{14\text{ TeV}}/\sigma_{8\text{ TeV}} \approx 3$  based on **MadGraph** simulations of  $pp \rightarrow j\nu\nu$ , (the dominant monojet +  $\cancel{E}_T$  background). We estimate  $\epsilon_{p_T(j_1) > p_T^{\text{cut}}}/\epsilon_{p_T(j_1) > 90\text{ GeV}}$  by applying cuts to our simulation of  $pp \rightarrow j\nu\nu$ . Finally, we assume either  $P_{\text{mis}} = 1$  or 10 and choose  $p_T(j_1)$  cuts to optimize the significance for each case. We underscore that many assumptions have been made to approximate  $B$ . Quoted backgrounds and significances are estimates.

We show estimated  $p_T^{\text{track}}$  distributions of the background and signal at various masses for  $\sqrt{s} = 14$  TeV applying the cut  $p_T(j_1) > 300$  GeV in Figure 2.6. For simplicity, we chose a single range  $500\text{ GeV} < p_T^{\text{track}} < 1500\text{ GeV}$  for all  $\mu_X$ . We found optimizing this range does not affect the significance much.

The estimated backgrounds and signals for various doublet masses at  $\sqrt{s} = 14$  TeV and  $\mathcal{L} = 3000\text{ fb}^{-1}$  with  $500\text{ GeV} < p_T^{\text{track}} < 1500\text{ GeV}$  for different  $p_T(j_1)$  cuts are shown in Table 2.2.

Significances are estimated using Eq. (2.24), again parameterizing  $\sigma_B = \beta_{\text{tot}}B$ , and are shown in Figure 2.7 for luminosities  $\mathcal{L} = 300$  and  $3000\text{ fb}^{-1}$ , taking various  $P_{\text{mis}}$  and optimizing the  $p_T(j_1)$  cut. The most recent ATLAS study has cuts with expected backgrounds of 18 – 48.5 events and uncertainties of about 25%. Naively, because a large  $p_T^{\text{cut}}$  reduces

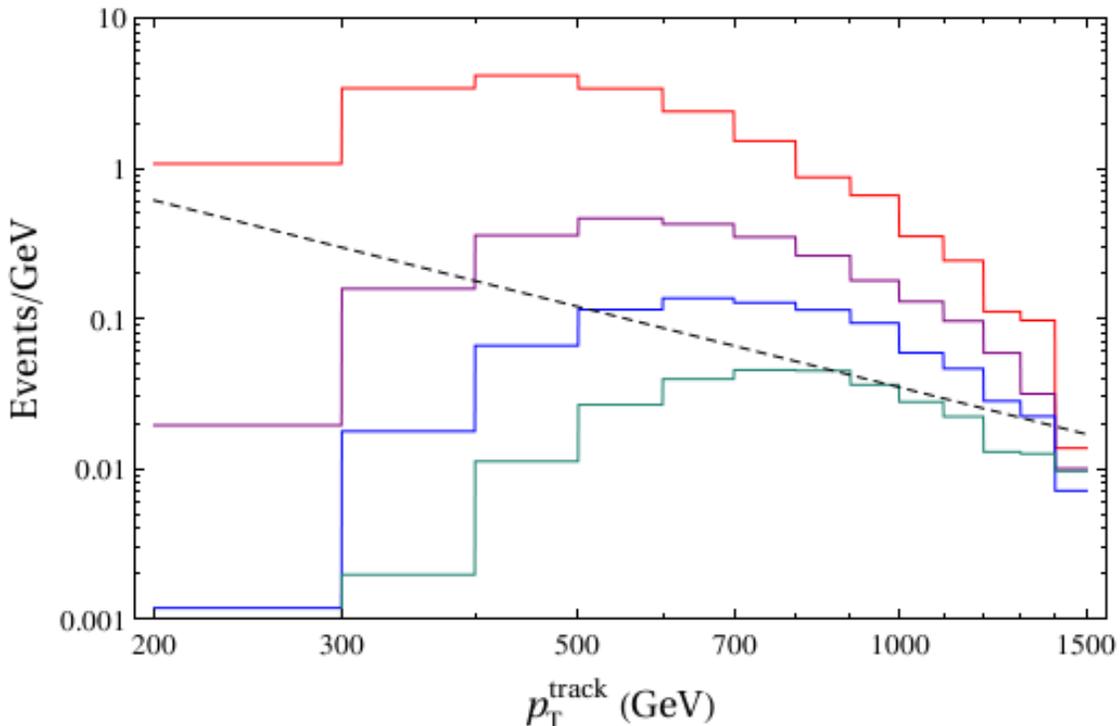


Figure 2.6: The  $p_T^{\text{track}}$  distribution of the background (black, dashed) and signal (solid) with  $p_T(j_1) > 300$  GeV and  $P_{\text{mis}} = 1$  at  $\sqrt{s} = 14$  TeV and  $\mathcal{L} = 3000$  fb $^{-1}$ . Signal spectra correspond to, from top to bottom,  $\mu_X = 100$  GeV (red), 130 GeV (purple), 150 GeV (blue), and 170 GeV (green).

the background much more than the signal, a very hard  $p_T^{\text{cut}}$  may give the best significance. However, because the backgrounds are estimated from data, if the background rate is much smaller than that in the present data, the fractional uncertainty may increase. Thus we limit our chosen cuts to where  $B$  is roughly in the same range as the current ATLAS backgrounds, where we assume that the uncertainty can be approximated by  $\beta_{\text{tot}} = 0.25$ <sup>12</sup>. Nevertheless, an even harder cut might ultimately be effective.

Thus, the LHC14 has the potential to probe the low-mass region of the parameter space for an  $SU(2)$  doublet. Optimistically, if  $P_{\text{mis}} = 1$ , we estimate a  $5\sigma$  discovery reach of about  $\mu_X = 150$  GeV and a  $2\sigma$  exclusion reach of about  $\mu_X = 170$  GeV; however, a signal can still be found with larger  $P_{\text{mis}}$ . Further, some parameter space will likely be accessible at

<sup>12</sup>We also entertained the possibility that data-driven methods could decrease  $\beta_{\text{tot}}$  as low as, e.g., 0.05 with softer cuts and larger  $B$ , but find harder cuts still produce better significance.

$p_T(j_1)$ cut	$B/P_{\text{mis}}$	$S$ ( $\mu_X$ in (GeV))			
		100	130	150	170
200 GeV	227	1190	255	97	36
300 GeV	44.1	963	200	75	28
500 GeV	3.96	646	137	49	18

Table 2.2: Number of the disappearing track background and signal events for  $\sqrt{s} = 14$  TeV,  $\mathcal{L} = 3000 \text{ fb}^{-1}$ , and  $500 \text{ GeV} < p_T^{\text{track}} < 1500 \text{ GeV}$ .

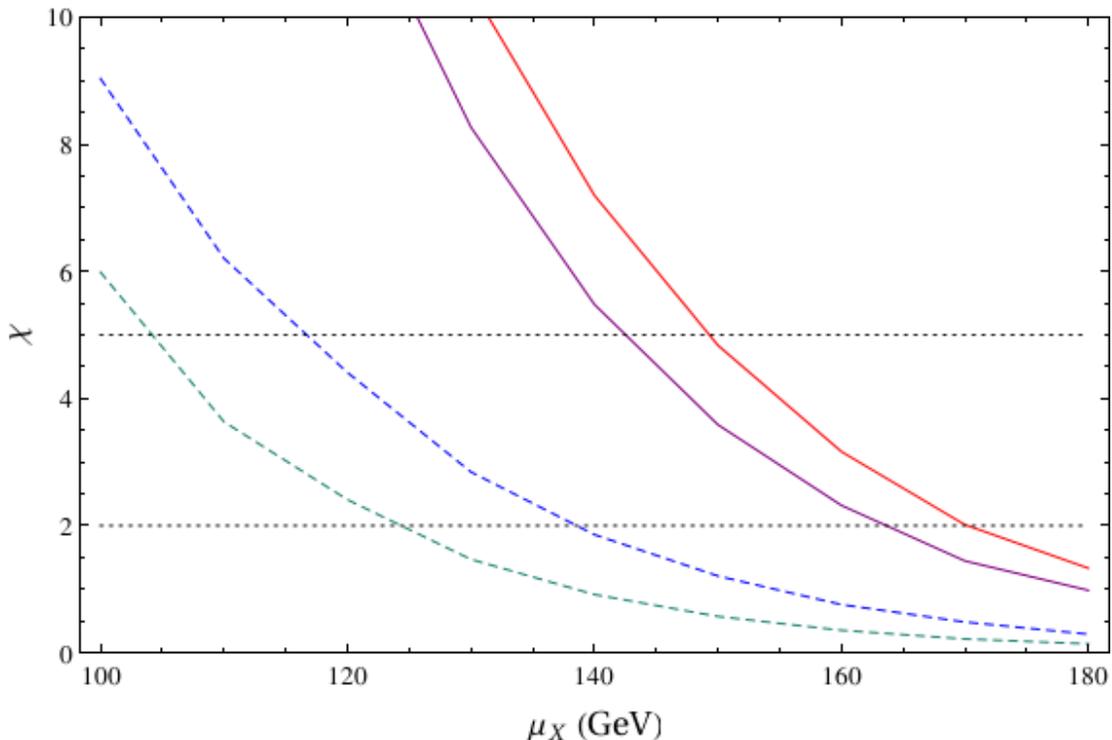


Figure 2.7: The significance  $\chi$  as a function of  $\mu_X$  with  $\sqrt{s} = 14$  TeV,  $500 \text{ GeV} < p_T^{\text{track}} < 1500 \text{ GeV}$ , and  $\sigma_B = \beta_{\text{tot}} B$ ,  $\beta_{\text{tot}} = .25$ . The dashed curves are for  $\mathcal{L} = 300 \text{ fb}^{-1}$  assuming  $P_{\text{mis}} = 10$  and  $p_T(j_1) > 300 \text{ GeV}$  (green, lower) or  $P_{\text{mis}} = 1$  and  $p_T(j_1) > 200 \text{ GeV}$  (blue, upper). The solid curves are for  $\mathcal{L} = 3000 \text{ fb}^{-1}$  assuming  $P_{\text{mis}} = 10$  and  $p_T(j_1) > 500 \text{ GeV}$  (purple, lower) or  $P_{\text{mis}} = 1$  and  $p_T(j_1) > 300 \text{ GeV}$  (red, upper). All  $p_T(j_1)$  cuts have been chosen to optimize the significance as described in the text. Dotted lines indicate  $2\sigma$  exclusion and  $5\sigma$  discovery thresholds.

lower luminosities. However, the exact reach will depend on how the backgrounds and their uncertainties scale with energy, instantaneous luminosity, and  $p_T^{\text{cut}}$ .

If a signal is detected, the track length and  $p_T^{\text{track}}$  distributions could provide clues about the type of particle that is detected. For example, because of the shortness of the  $X^\pm$

lifetime, this model will have tracks with higher  $p_T$  and shorter path lengths, while other models with longer lifetimes that have so far avoided detection will have smaller production cross sections that compensate for the greater probability for long path lengths.

## 2.4 Conclusions

A simplified DM EFT in which the dark matter communicates with the SM through a  $Z$  boson represents a valuable target for WIMP searches. If this particle is to make up all of the observed DM abundance, it must be Majorana in order to forbid SI interactions through a vector coupling to the  $Z$ . However, recent improvements in direct detection limits have begun to force the simplest gauge-invariant version of such a Majorana DM into a region of parameter space exhibiting large contributions to precision electroweak parameters and in which one might question the validity of the EFT. As such, the time is ripe to consider how models of WIMPs beyond the simplest examples fare in the face of current and future direct searches.

To illustrate this, we have discussed the singlet-doublet model, which exhibits similar phenomenology to the  $Z$ -mediated model in certain regions of parameter space. However, the presence of additional states nearby in mass that fill out complete electroweak representations prevents overly large contributions to precision electroweak parameters. Moreover, the extra DM couplings, notably to the  $W$  boson, allowed by these additional states lead to different DM phenomenology. In particular, contributions to the DM annihilation cross section from  $t$ -channel DM partner exchange or coannihilation can allow the correct thermal relic density to be achieved with a small DM- $Z$  coupling, opening new regions of parameter space that represent exciting targets for future experiments. Moreover, the additional partners of the DM could be discovered at the high-luminosity LHC.

Future direct detection will probe well beyond where the DM cosmology is described by the simplified  $Z$ -mediated model, and null results would allow only the case in which the

Yukawa couplings are relatively small and the thermal relic density is achieved through coannihilation. This would require a somewhat striking coincidence of parameters (which could perhaps arise from renormalization group fixed ratios as in [229]), with not only the mass of the charged state lying close to the DM mass but also the Yukawa couplings conspiring such that DM- $h$  coupling approximately vanished to evade the very stringent SI scattering constraints.

Alternatively, or perhaps in addition to the above models, we have explored the possibility that new vectorlike doublets may be present at the TeV scale. If stable, these particles must make up a tiny fraction of the dark matter. Nevertheless, they may be phenomenologically relevant. They could actually be the first signal observed at direct detection experiments, perhaps presenting a background to the true dark matter.

Comparing the two collider search channels presented for the vectorlike doublets, the disappearing track channel has the potential to probe a significantly larger mass range than the monojet channel. Interestingly, it is this low mass window where the non-thermal cosmology realizes this scenario most easily, i.e. with the largest values of  $c_X$ , see Fig. 2.5. Observing larger-mass doublets at hadron colliders is challenging—unless, perhaps, they are part of a larger dark sector that boosts either the production and/or the visibility of the events containing the  $X$  particles. If a missing energy signal is found in the monojet channel, it will be difficult to determine what type of particle is responsible for it, and indeed whether it corresponds to a significant fraction of the dark matter. A much smaller set of models will (simultaneously) produce a disappearing track signature. Moreover, with enough statistics, further inferences can in principle be made from the lifetime distribution in the detector.

These added clues, perhaps along with material dependence at direct detection experiments, would be enough to indicate that a putative direct detection signal actually came from the “tip of the dark matter iceberg”. A future lepton collider could also probe this scenario contingent on kinematic accessibility.

# Chapter 3

## High-Scale Axions without Isocurvature from Inflationary Dynamics

*This chapter was completed in collaboration with John Kearney and Aaron Pierce [155].*

As discussed in Section 1.3, axions may simultaneously solve the strong CP problem and be produced copiously enough to explain the observed dark matter abundance. Axion production is quite different than the thermal production of WIMPs, and instead proceeds via coherent oscillations and, in some cases, topological defects. In particular, axion cosmology depends on whether the PQ symmetry is spontaneously broken before inflation ends. The Hubble parameter during inflation  $H_I$  can be related to the tensor perturbation amplitude in the cosmic microwave background (CMB) [230]. Current bounds from BICEP2/Keck/Planck constrain the tensor-to-scalar ratio  $r < 0.07$  [9, 10], while planned near-future detectors will probe the region  $r \gtrsim 2 \times 10^{-3}$  [34]. If such a primordial gravity wave detection is made by these experiments, it will indicate

$$H_I = \left( \frac{\pi^2}{2} M_P^2 \Delta_R^2 r \right)^{1/2} \gtrsim 10^{13} \text{ GeV}, \quad (3.1)$$

where the measured scalar perturbation amplitude  $\Delta_R^2 = 2.142 \times 10^{-9}$  [20] and the reduced Planck mass  $M_P = 2.435 \times 10^{18}$  GeV.

If the PQ symmetry breaks after inflation ( $f_a < \frac{H_I}{2\pi}$ ), topological defects will form. These topological defects will lead to overclosure due to either domain wall stability if the domain wall number  $N_{\text{DW}} > 1$ , or axion overproduction from cosmic strings and domain walls if  $f_a \gtrsim 10^{11}$  GeV [84–89]. Larger PQ-breaking scales and domain wall numbers other than one are allowed only if the PQ symmetry is broken before the end of inflation ( $f_a > \frac{H_I}{2\pi}$ ) so that topological defects are inflated away.<sup>1</sup>

In the case where  $f_a > \frac{H_I}{2\pi}$ , the (massless) axion will obtain inflationary fluctuations  $\sim \frac{H_I}{2\pi}$  similar to the inflaton [80–83]. These fluctuations in the axion field lead to density perturbations when the axion field begins coherent oscillations. At that point, energy in the quark-gluon plasma is converted into axion rest mass energy, resulting in a relationship between the axion and radiation density perturbations arising from this process:

$$\delta\rho_a = -\delta\rho_{\text{rad}}. \quad (3.2)$$

This is different than the adiabatic perturbations seeded during inflation, which satisfy

$$\frac{\delta(n/s)}{n/s} = 0 \implies \frac{1}{3} \frac{\delta\rho_{\text{matter}}}{\rho_{\text{matter}}} = \frac{1}{4} \frac{\delta\rho_{\text{rad}}}{\rho_{\text{rad}}} = \frac{\delta T}{T}. \quad (3.3)$$

Note that the perturbations created by the axion field fluctuations in (3.2) do not initially change the total density perturbations  $\delta\rho_{\text{tot}} = \delta\rho_{\text{matter}} + \delta\rho_{\text{rad}}$ , whereas the adiabatic perturbations in (3.3) affect  $\delta\rho_{\text{tot}}$ . This difference in initial conditions is reflected in the CMB, where the axion fluctuations appear as isocurvature fluctuations that are out of phase with the adiabatic fluctuations. Thus, the isocurvature fluctuations resulting from the axion and are bounded by observations of the CMB [20].

---

<sup>1</sup>Alternatively, topological defects may be destabilized via a very delicately chosen tilt to the potential [231, 232].

Remaining on the case that the PQ symmetry is broken before the end of inflation, once inflation ends the axial component of the PQ field remains at the value taken during inflation until the temperature-dependent QCD instanton axion mass  $m_{\text{QCD}}^2$  becomes comparable to the Hubble expansion rate (typically near the QCD phase transition). At this point, the field will begin oscillating around the minimum of the low-energy potential. These coherent oscillations correspond to an energy density stored in the axion field, meaning that the axion can behave as cold DM. The QCD axion DM fraction generated in this manner is given by [26, 233]

$$R_a = \frac{\Omega_a}{\Omega_{\text{DM}}} \simeq \begin{cases} 5.6 \times 10^7 \langle \theta^2 \rangle (f_a/M_P)^{7/6}, & f_a < \hat{f}_a, \\ 1.6 \times 10^8 \langle \theta^2 \rangle (f_a/M_P)^{3/2}, & f_a > \hat{f}_a, \end{cases} \quad (3.4)$$

where  $\langle \theta^2 \rangle = \theta_i^2 + \delta\theta^2$  is the axion misalignment angle including contributions from the initial misalignment  $\theta_i$  and the primordial fluctuations  $\delta\theta$  (neglecting anharmonic factors [75]). The scale  $\hat{f}_a \simeq .991 \times 10^{17} \text{ GeV} \simeq \frac{M_P}{2.5}$  corresponds to the transition between oscillations beginning before or after the QCD phase transition (see [234, 235] for corrections to this approximation).

Using this expression for the axion DM fraction and assuming a detection of primordial tensor modes—indicating a large  $H_I$  and thus large isocurvature fluctuations  $\delta\theta$ —current constraints on isocurvature would rule out the simplest axion models for  $f_a > \frac{H_I}{2\pi}$ , see, *e.g.*, [74, 233]. For high-scale axion models to be viable in the presence of such a detection, the axion fluctuations must be suppressed.

It is important to understand the robustness of isocurvature constraints for two reasons. First, string theories generically predict a PQ-breaking scale around the grand unified theory (GUT) scale,  $f_a \sim 10^{16} \text{ GeV}$  [90, 91]. Second, several experiments have recently been proposed to look for large- $f_a$  axion DM [11, 117–120]. If primordial gravity waves are observed, it will be important to know whether such theories are permissible and thus which regions of parameter space these experiments should target. More optimistically, if signals are seen both in tensor modes and in a hunt for large- $f_a$  axions, we should know what types of new

physics would be required to reconcile the two signals.

We examine two classes of solutions that have been discussed to circumvent isocurvature constraints and resurrect models with large  $f_a$ . Both involve an inflationary shift of the PQ sector away from its zero temperature minimum and/or potential. In the first, the PQ scale is large during inflation  $f_I \gg f_a$  [83, 236–239]. In the second, the axion has a large mass during inflation related to an explicit PQ symmetry breaking [237, 240].<sup>2</sup> We explore the measures necessary to suppress isocurvature and the constraints that must be taken into account as these measures are implemented.

In Section 3.1, we discuss the viability of exclusively suppressing isocurvature via an inflationary PQ-breaking scale that is larger than the present scale. While this has been previously considered as a viable mechanism to reduce isocurvature, we show that it often faces insurmountable constraints due to the large hierarchy required between the two scales. In Section 3.2, we explore the possibility that isocurvature can be suppressed with an enhanced explicit breaking of the PQ symmetry, demonstrating the model-building and experimental constraints that ultimately limit the range that  $f_a$  may take in such models. Numerous previous works have observed that explicit PQ-breaking modifies the axion potential, and so in general could be a suitable strategy for suppressing isocurvature. Here, however, we focus on developing concrete models, which allows us to study the various interrelated field dynamics and constraints, and we are thus able to more fully address the strengths and shortcomings of this approach.

As we are interested in modifying the behavior of the axion during inflation, our primary focus here will be on how the solutions implemented affect the dynamics of the axion and inflaton fields. However, the solutions under consideration in the above sections can also affect the behavior of the radial component of the PQ-breaking field. This leads to model-dependent, but potentially severe, constraints related to ensuring a consistent cosmological history. We review these issues in Section 3.3. Many of these constraints could be evaded

---

<sup>2</sup>Other mechanisms to suppress isocurvature have also been proposed [241–250], though notably many of these do not work for GUT-scale axion DM with high-scale inflation.

if the PQ field relaxed adiabatically to its minimum today without oscillation. However, in Section 3.4, we discuss how models that are designed to permit such adiabatic relaxation will generally only partially mitigate constraints. Moreover, we highlight some previously unappreciated obstacles to constructing such models, which make it unclear whether adiabatic relaxation can in fact be implemented. We conclude in Section 3.5.

### 3.1 Suppressing Isocurvature via Wave Function Renormalization

The complex PQ field  $S$  can be written in terms of its radial and axial components,  $\sigma$  and  $a$ , and the PQ breaking scale,  $f_{\text{PQ}}$ ,

$$S = \frac{1}{\sqrt{2}} (\sigma + f) \exp \left[ \frac{ia}{f} \right], \quad (3.5)$$

where  $f = \frac{f_{\text{PQ}}}{N_{\text{DW}}}$ , with  $f = f_a$  and  $f = f_I$  denoting the effective breaking scale today and during inflation, respectively. During inflation, fluctuations  $\sim \frac{H_I}{2\pi}$  are induced in the canonically normalized axion field, which correspond to fluctuations in the field today given by [83, 236],

$$\delta a = \frac{H_I f_a}{2\pi f_I}. \quad (3.6)$$

Thus,  $f_I > f_a$  will reduce axion fluctuations. However, due to the stringent bounds on isocurvature,  $\alpha_{\text{iso}} < 0.0019$  at the 95% confidence level [20],  $f_I \gg f_a$  is necessary to achieve the required suppression. The isocurvature parameter is

$$\frac{\alpha_{\text{iso}}}{1 - \alpha_{\text{iso}}} \simeq \frac{1}{\Delta_R^2} \left( 4R_a^2 \frac{(H_I/(2\pi))^2}{f_I^2 \langle \theta^2 \rangle} \right). \quad (3.7)$$

As such, employing Eq. (3.4),

$$\frac{f_I}{f_a} \gtrsim 1.4 \times 10^4 \sqrt{R_a \left( \frac{r}{0.002} \right) \left( \frac{0.0019}{\alpha_{\text{iso,max}}} \right) \left( \frac{10^{-1} M_P}{f_a} \right)^{5/12}} \quad (3.8)$$

(valid for  $f_a < \hat{f}_a$ , while a similar expression can be derived for the other case). Thus, for a theory with detectable gravitational waves  $r \gtrsim 2 \times 10^{-3}$ , a large separation is required between  $f_I$  and  $f_a$  to evade bounds on isocurvature.

This displacement must be generated in a way consistent with inflation. For instance,  $\langle S \rangle$  could be displaced via a model that leverages inflationary dynamics to give an explicit inflationary minimum for  $|S|$  larger than its present one. However, as seen from Eq. (3.8),  $f_I$  must be super-Planckian for  $f_a \gtrsim 4 \times 10^{-7} M_P$ . Such a large displacement is likely to seriously disrupt the inflaton potential via the couplings between the inflationary and PQ sectors, making such a mechanism unfavorable; further details are provided in footnote 13. As such, the likely source of the inflationary radial field displacement is Hubble friction, *i.e.*, the situation in which the radial field is displaced from the minimum today at the onset of inflation and remains light (compared to  $H_I$ ) during inflation.

In fact, regardless of the source of the radial field displacement, such super-Planckian field values could still disrupt inflation. Depending on the form of the potential for  $S$ , at large field values it may dominate the Universe’s energy density, superseding the supposed inflationary sector. This effect is exacerbated as the required  $f_I$  increases. Moreover, such large values for  $f$  have resisted embedding in string theory [251], perhaps related to the weak-gravity conjecture [252–256].

This concern could be avoided if the PQ sector were responsible for inflation, as suggested in [239]. We will return to this example at the end of this section, but, as we shall see, the other constraints outlined below generally require additional late-time dilution of axions for such “PQ sector inflation” to be a viable solution.

Following inflation, once  $H$  decreases below the radial field’s mass,  $\sigma$  will start to oscil-

late with a large initial amplitude corresponding approximately to its displacement during inflation.<sup>3</sup> As we discuss in Section 3.3, the amplitude and decay of these oscillations are constrained by cosmological observations—the energy density must be dissipated efficiently and appropriately (Section 3.3.3) and fluctuations induced in a light, Hubble-trapped  $\sigma$  may also be bounded (Section 3.3.1). Both of these constraints can be evaded via appropriate model building that is largely independent of the axion. Alternately, if  $\sigma$  is the inflaton, these constraints are superseded by requiring generation of the observed curvature perturbations and appropriate reheating.

Our main focus here, however, is that the axion field itself can experience nonperturbative effects from large  $\sigma$  oscillations, along the lines of the parametric resonance effects that can be induced by inflaton oscillations during preheating [257,258]. Constraints from these effects turn out to be most relevant for smaller  $f_a$ , and thus complementary to the concerns about the disruption of the inflation potential. Specifically, radial oscillations can excite fluctuations in the axial field via parametric resonance. If the fluctuations in the axial field grow to order  $f_a$  so that  $\theta = a/f_a$  takes all possible values throughout the universe with roughly equal probability, topological defects form and symmetry can be considered restored [259].<sup>4</sup> Since topological defects overclose the universe for the models with larger  $f_a$  (or  $N_{\text{DW}}$ ) that are of interest here, PQ symmetry restoration must be avoided.<sup>5</sup> Moreover, if the symmetry is restored, this clearly defeats the purpose of suppressing the axion fluctuations in the first place.

Axion fluctuations are more strongly enhanced for larger initial  $\sigma$  oscillation amplitude, which results in an increased duration of oscillations and oscillation speed. Based on lattice

---

<sup>3</sup>Adiabatic relaxation as described in Section 3.4, which requires an inflationary minimum set by inflationary dynamics, is not possible in this model which relies purely on Hubble friction.

<sup>4</sup>The fluctuations can also be viewed as inducing a large, positive effective mass squared for the radial direction such that the potential no longer exhibits a symmetry-breaking form [259].

<sup>5</sup>In addition to the possibility of symmetry restoration, large local fluctuations could contribute to the axion relic density (though not isocurvature, as the relevant modes are very small wavelength) or even result in the field locally relaxing to different minima after inflation, producing domain walls. As fluctuations grow exponentially, though, these latter concerns are only likely to be relevant very close to the region where there is risk of fully restoring the symmetry.

simulations [260], topological defects form when the initial amplitude

$$|S|_i \gtrsim 10^4 \left( \frac{f_a}{\sqrt{2}} \right) \quad (3.9)$$

in a quartic potential and matter-dominated background.<sup>6</sup> The bound is even stronger in the case where the PQ field dominates the energy density [259, 261], in which case topological defects form when

$$|S|_i \gtrsim 2 \times 10^2 \left( \frac{f_a}{\sqrt{2}} \right). \quad (3.10)$$

A similar bound should apply in the case of radiation domination as a PQ field oscillating in a quartic potential behaves as radiation. Thus, avoiding symmetry restoration requires that the initial amplitude of oscillations (or, at least, the initial amplitude when the potential starts behaving quartically [263]) is somewhat small.<sup>7</sup>

Comparing Eqs. (3.8) and (3.9), for  $f_a \lesssim 10^{-1} M_P$ , one cannot simultaneously suppress isocurvature using wave function renormalization alone and avoid symmetry restoration. Note  $f_I$  in Eq. (3.8) is the field value when the modes relevant to the CMB exit the horizon, as opposed to when oscillations commence. However, we expect  $|S|_i$  to differ from  $\frac{f_I}{\sqrt{2}}$  only by  $\mathcal{O}(10\%)$  due to the flatness of the PQ potential required to avoid dominating the inflationary energy density. Thus, suppressing isocurvature via this mechanism is challenging and likely only viable for a limited range of  $f_a$  and potentials.

The constraint from nonthermal symmetry restoration could be relaxed somewhat if the PQ field began oscillating in a potential dominated by  $|S|^{2M}$  with  $M \geq 3$  [263]. However, constraints still apply once quartic or quadratic terms come to dominate. Alternatively, the required hierarchy in  $f_I$  and  $f_a$  could be reduced in the case of a late-decaying scalar that

---

<sup>6</sup>Refs. [259–261] use the linear basis,  $S = \frac{1}{\sqrt{2}}(X + iY)$ , to explore this effect, but it can also be understood in the nonlinear basis of Eq. (3.5) via derivative interactions between  $\sigma$  and  $a$ , see [262].

<sup>7</sup>Another more model-dependent nonthermal effect that can restore the PQ symmetry is described in Section 3.3.2.

dilutes the axion relic abundance [74, 92, 264]. In this case,

$$R_a \simeq 1.7 \times 10^7 \left( \frac{T_{\text{RH}}}{6 \text{ MeV}} \right) \langle \theta^2 \rangle \left( \frac{f_a}{M_P} \right)^2, \quad (3.11)$$

where  $T_{\text{RH}}$  is the reheat temperature at which the scalar decays, with the requirements  $T_{\text{RH}} \gtrsim 6 \text{ MeV}$  to not spoil Big Bang nucleosynthesis (BBN) [265] and  $T_{\text{RH}} \lesssim \Lambda_{\text{QCD}}$  for the dilution to occur. Then, the isocurvature bound requires

$$\frac{f_I}{f_a} \gtrsim 3.0 \times 10^3 \sqrt{R_a \left( \frac{r}{0.002} \right) \left( \frac{0.0019}{\alpha_{\text{iso,max}}} \right) \left( \frac{T_{\text{RH}}}{6 \text{ MeV}} \right)}, \quad (3.12)$$

which could still be permitted by Eq. (3.9) depending on  $r$  and  $\alpha_{\text{iso}}$ . However, future nonobservation of isocurvature would ultimately restrict even this case.

Finally, even if axion fluctuations are made negligible, the initial displacement of the axion field is also constrained to avoid the overproduction of DM, particularly for larger values of  $f_a$ . Eq. (3.4) implies

$$\theta_i \lesssim \begin{cases} 1.3 \times 10^{-4} (f_a/M_P)^{-7/12}, & f_a < \hat{f}_a, \\ 7.9 \times 10^{-5} (f_a/M_P)^{-3/4}, & f_a > \hat{f}_a. \end{cases} \quad (3.13)$$

In other words, the inflationary value for the PQ phase must not differ significantly from the value today. Here, the initial misalignment at the onset of inflation corresponds to that in the Hubble patch that gave rise to our Universe. Such a small misalignment angle could be justified by an anthropic argument—the axion takes on different initial misalignments in different inflationary patches, and we reside in one such patch that gives rise to a not-too-large axion abundance [93, 94]. Alternatively, dilution of the axion abundance by a late-decaying particle would permit larger initial misalignment angles for a given  $f_a$ .

Overall, it seems challenging to implement a solution in which axion isocurvature is suppressed by wave function renormalization alone.

### 3.1.1 PQ Sector Inflation

Some of these challenges, namely those arising from the interplay between the PQ and inflationary sectors, could be surmounted if these sectors were identified. However, this approach is complicated by nonperturbative effects that result in the PQ symmetry being restored in much of the same parameter space where the isocurvature fluctuations are sufficiently suppressed. As we shall see, this leads to the additional requirement of late-time dilution of the axion abundance to yield a viable inflationary model.

If the PQ field is in a “wine-bottle” potential,

$$V = \lambda \left( |S|^2 - \frac{f_a^2}{2} \right)^2, \quad (3.14)$$

it will have a  $|S|^4$ -type inflationary potential [266], which is inconsistent with BICEP2/Keck/Planck data on  $r$  and the scalar tilt  $n_s$  [9, 10, 20]. The introduction of a nonminimal coupling of  $S$  to gravity [267]

$$V \supset \xi |S|^2 R, \quad (3.15)$$

where  $R$  is the Ricci scalar and  $\xi$  is a constant, allows for an inflationary model potentially consistent with existing data. The measured value of  $\Delta_R^2$  specifies the relationship between  $\xi$  and  $\lambda$ . With the potential so fixed, the slow-roll parameters and inflaton field value throughout inflation (and thus  $f_I$ ) can be calculated. One finds the requirement  $\xi \gtrsim (\text{few}) \times 10^{-3}$  to satisfy bounds on  $r$  and  $n_s$ . The details are worked out in Ref. [239] (see also [268, 269]), which identifies a window  $10^{12} \text{ GeV} \lesssim f_a \lesssim 10^{15} \text{ GeV}$  where isocurvature constraints can be avoided.

However, this model suffers from the previously described parametric resonance constraints, wherein oscillations in  $\sigma$  can restore the symmetry. Because the PQ field dominates the energy density, a stronger bound more analogous to Eq. (3.10) applies. However, due to the nonminimal coupling, we cannot directly apply the bound of Eq. (3.10). First, the

potential is modified: in the Einstein frame,

$$V_E = \frac{\lambda(|S|^2 - f_a^2/2)^2}{(1 + 2\xi |S|^2/M_P^2)^2}, \quad (3.16)$$

which for large values of  $|S|$  and  $\xi > 0$  is flatter than the wine bottle. Second,  $\sigma$  has a rolling mass parametrized by,

$$Z(\sigma) = \frac{(1 + \xi\sigma^2)^2}{1 + \xi\sigma^2 + 6\xi^2\sigma^2}, \quad (3.17)$$

where a heavier rolling mass,  $Z(\sigma) < 1$ , corresponds to slower motion.

Rather than attempting a full reanalysis similar to that of [259, 261] including these contributions, we present two bounds that we believe bracket the true one. They both come from the rule of thumb of Eq. (3.10), but use different choices of  $|S|_i$ . In the first,  $|S|_i \simeq |S|_{\text{end}}$ , where  $|S|_{\text{end}}$  is the value for  $|S|$  at the end of inflation, defined to be when the slow-roll parameter  $\epsilon \simeq 1$ . This requirement is likely a bit too restrictive, especially for larger values of  $\xi$ , since the radial field will begin rolling more slowly due both to its rolling mass and the shallower potential compared to the case of  $\xi = 0$ . The second possibility is to take  $|S|_i$  to be the point after inflation has ended where both  $Z(\sigma) \simeq 1$  and  $\frac{d \log V_E}{d \log \sigma} \simeq 4$ , *i.e.*, where the rolling mass is close to its standard value of 1 and the potential is similar to Eq. (3.14). This second bound is likely not quite constraining enough, as the radial field will already have begun rolling from a larger field value and will have attained some “velocity” from doing so.

These symmetry restoration bounds are plotted in Fig. 3.1, with the solid blue curve and shaded region corresponding to constraints with  $|S|_i$  determined by conditions on  $Z(\sigma)$  and  $V_E$ ,<sup>8</sup> and the dashed blue curve corresponding to taking  $|S|_i = |S|_{\text{end}}$ . Also shown are isocurvature bounds if the axion makes up all of the DM abundance (*i.e.*,  $R_a = 1$ ) assuming either Eq. (3.4) (solid red region)<sup>9</sup> or Eq. (3.11) with maximal possible dilution

<sup>8</sup>In our study, we require  $Z \geq 0.9$  and  $\frac{d \log V_E}{d \log \sigma} \geq 3.9$ . The solid blue curve is mildly sensitive to this choice, but the overlap of the regions excluded by the solid blue and red curves is not.

<sup>9</sup>Our bounds on isocurvature are stronger by roughly a factor of  $8\pi$  compared to Ref. [239] because their Eq. (15) uses the Planck mass while their other expressions use the reduced Planck mass. We also use the

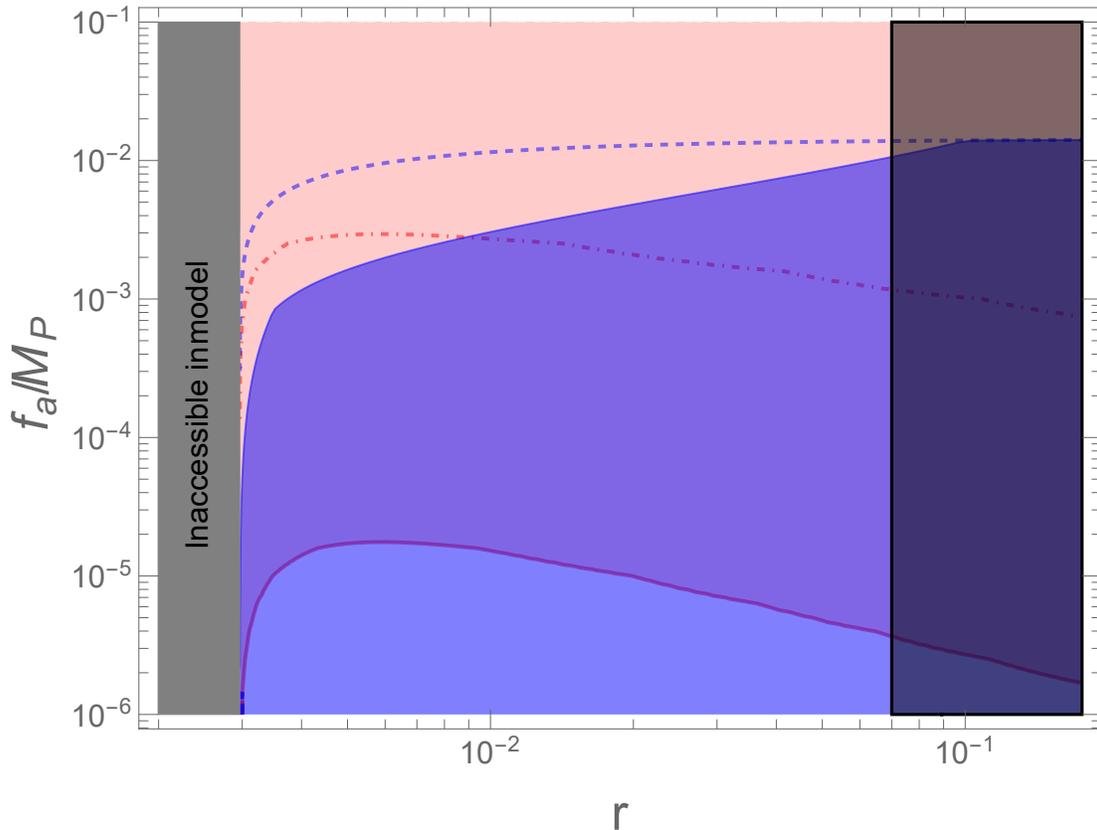


Figure 3.1: Constraints on the radial part of the PQ field acting as the inflaton. The red curves denote isocurvature constraints after taking into account  $f_I > f_a$  assuming  $R_a = 1$  with an abundance determined by either Eq. (3.4) (solid, shaded upwards) or Eq. (3.11) with  $T_{\text{RH}} = 6$  MeV (dot-dashed). The two blue curves show two prescriptions for determining where the PQ symmetry may be restored by parametric resonance (see text for details); the solid shaded bound should be viewed as definitively excluded. The black vertical region on the right shows the current BICEP2/Keck/Planck bounds on the tensor to scalar ratio [9,10], and the gray vertical region on the left is inaccessible in this model.

(corresponding to  $T_{\text{RH}} = 6$  MeV) (dot-dashed red line). For the purposes of calculating the isocurvature constraints we take  $f_I$  to be the field value 60 e-folds before the end of inflation, approximately corresponding to the scale when modes relevant to the CMB exit the horizon. Also shown are current BICEP2/Keck/Planck bounds on  $r$  [9,10]. Note that future experiments will be able to probe all values of  $r$  for this model [34].

Comparing even the conservative estimate of the symmetry restoration bounds to the updated Planck bounds on  $\alpha_{\text{iso}}$  [20].

isocurvature bounds assuming the axions are produced in the manner of Eq. (3.4) (*i.e.*, without a late-decaying scalar), it is clear that some additional cosmological model building is necessary to make this a viable inflationary model. With late-time dilution, some parameter space may not yet be excluded depending on where the true symmetry restoration bounds fall. Nevertheless, this parameter space is narrow and only potentially allows PQ-breaking scales  $f_a \lesssim 2 \times 10^{-3} M_P$ . Additionally, a detection of  $r \gtrsim 10^{-2}$  would definitively exclude this model. Of course, larger values of  $f_a$  and  $r$  could be allowed if the requirement that  $R_a = 1$  is relaxed, though this would require an even smaller initial misalignment angle without any particular motivation for such a tuning.

This model illustrates an important tension. Sufficient suppression of isocurvature requires a very large hierarchy between  $f_I$  and  $f_a$ . For just such a hierarchy, nonperturbative dynamics have the potential to restore the PQ symmetry. Consequently, additional cosmological mechanisms are necessary to permit models that suppress isocurvature via wave function renormalization, and the allowed values for  $f_a$  are still restricted. Moreover, if isocurvature perturbations continue unobserved, even these remaining values for  $f_a$  may eventually be excluded.

Having demonstrated the difficulties inherent in models that seek to suppress isocurvature via axion wave function renormalization alone, we now turn to models in which field displacement enhances explicit PQ breaking in the early Universe.

## 3.2 Axion Mass from Enhanced Explicit $U(1)_{\text{PQ}}$ Breaking During Inflation

Scalar field fluctuations are very efficiently suppressed for fields with large masses  $m \gtrsim H_I$  (see, *e.g.*, [270]). With this motivation, in this section we will consider models in which  $U(1)_{\text{PQ}}$  is only an approximate symmetry that is explicitly broken to a discrete subgroup  $\mathbb{Z}_N$ . Then, higher dimension operators allowed by the discrete symmetry but forbidden by

the continuous  $U(1)_{\text{PQ}}$  generate contributions to the axion mass in addition to those from QCD. If these operators were enhanced during inflation, the large inflationary mass could sufficiently suppress axion isocurvature while still yielding a model consistent with strong CP constraints.

Supposing the axial component of the field is heavy during inflation, it will evolve to the minimum favored by the operator(s) responsible for generating its large mass. This minimum need not coincide with those of the QCD potential and in general would not be expected to as the two contributions to the axion potential arise from different sources. However, it will determine the initial misalignment angle  $\theta_i$ . So, for axions with large  $f_a$ , barring a mechanism for the axion to adiabatically evolve to the vicinity of the minimum today, cosmological considerations do require a rather dissatisfying coincidence between operators to avoid overproduction of axion DM, see Eq. (3.13).<sup>10</sup> This incredible coincidence could be mitigated if, for instance, a late-decaying particle diluted the axion abundance, allowing larger initial misalignment.

Moreover, if non-QCD contributions to the axion potential are incompletely turned off, measurements of CP-violating observables today tightly constrain the size of such contributions [271]. For the axion to still provide a solution to the strong CP problem, the minimum today must correspond to the QCD minimum  $\theta_0$  to a very high degree [64],

$$\left\langle \frac{a}{f_a} - \theta_0 \right\rangle \leq \bar{\theta} \simeq 10^{-11}, \quad (3.18)$$

where  $\bar{\theta}$  represents the current constraints on the effective  $\theta$  angle. Let  $m_{\text{eff},0}^2$  represent additional contributions to  $m_a^2$  today and  $\theta_N$  represent the minimum favored by the additional contributions (such that  $\theta_i \simeq |\theta_N - \theta_0|$  is the initial displacement). Then, assuming  $\theta_i$  is

---

<sup>10</sup>The anthropic argument that we simply reside in an inflationary patch where the initial misalignment is small is no longer valid. Any attempted anthropic argument must instead consider a multiverse selection amongst appropriate potential parameters.

tuned to be as small as required by cosmology but not more so,

$$m_{\text{eff},0}^2 \lesssim \frac{\bar{\theta} m_{\text{QCD}}^2}{|\theta_N - \theta_0|}. \quad (3.19)$$

Note, owing to the necessarily small denominator, this constraint is somewhat weaker than the usual constraint for  $\mathcal{O}(1)$  displacement between the operators—constraints would be more stringent for larger displacements as allowed by, *e.g.*, dilution. Here we have approximated the QCD potential as quadratic in the vicinity of the minimum,

$$V \simeq \frac{1}{2} m_{\text{QCD}}^2 (a - f_a \theta_0)^2, \quad (3.20)$$

and  $m_{\text{QCD}} \simeq 6 \mu\text{eV} (10^{12} \text{ GeV}/f_a)$  [178].

This severe constraint makes it impossible to arrange for a large inflationary mass via explicit PQ symmetry breaking under the assumption that  $f_I = f_a$ . However, displacement of the PQ field during inflation from its minimum today can enhance the non-QCD contributions to  $m_a^2$ . In fact, these models can achieve an effective inflationary axion mass  $m_{\text{eff,inf}}^2 \gtrsim H_I^2$  even for  $f_I \lesssim M_P$ , so the required field displacement can be substantially less than the (super-)Planckian values of  $\langle S \rangle$  needed to suppress isocurvature by changing the normalization of the axion field alone (Section 3.1).

However, the symmetry breaking responsible for giving the axion a large inflationary mass will also give the radial field a mass of the same order,  $m_\sigma \sim H_I$ . Thus, in this case,  $f_I > f_a$  cannot simply arise from Hubble trapping, but instead requires a modification of the PQ potential during inflation. As such, the PQ field must couple to inflationary dynamics. Importantly, such couplings to the inflationary sector can disrupt the flatness of the fragile inflaton potential. To be concrete, the contributions of couplings between the PQ and inflationary sectors, denoted  $\Delta V$ , will result in additional contributions to the slow-roll

parameters,

$$\epsilon \equiv \frac{M_P^2}{2} \left( \frac{V'}{V} \right)^2, \quad \eta \equiv M_P^2 \frac{V''}{V}, \quad (3.21)$$

where primes denote derivatives with respect to  $I$ . There are no hard and fast constraints on such contributions, as the part of the potential that depends only on the inflaton  $V_I$  could always be tuned to give values consistent with CMB observations. However, large contributions from  $\Delta V$  would require particularly severe (and potentially dynamical, if  $f_I$  changes over the course of inflation) tunings. As such,  $V_I$  would have to be very carefully arranged to appropriately balance the contribution from  $\Delta V$  throughout inflation and maintain an appropriate inflationary trajectory.

We study the restrictions arising from these contributions to the inflationary potential as well as from strong CP constraints in the context of specific models below. We will consider three models. The first will consist of only a single PQ field  $S$  charged under  $\mathbb{Z}_N$ , and we will see that the combination of strong CP constraints and avoiding excessive disruption of the inflaton potential limits the values of  $f_a$  that can be reasonably allowed. Consequently, we will consider other models in which additional fields charged under  $U(1)_{\text{PQ}}$  (or the discrete symmetry)—either the inflaton or additional PQ fields—acquire large values during inflation, potentially relaxing constraints on models with larger  $f_a$ .

### 3.2.1 A Simple Model with a Single PQ Field

Perhaps the simplest example of an operator that explicitly breaks  $U(1)_{\text{PQ}}$  to a discrete symmetry (in this case,  $\mathbb{Z}_N$ ) is  $S^N$ . We consider a basic model containing this operator as well as  $U(1)_{\text{PQ}}$  invariant terms,

$$V \supset \lambda \left( |S|^2 - \frac{f_a^2}{2} \right)^2 - \frac{\delta}{2} I^2 |S|^2 + \left( \frac{k S^N}{M_P^{N-4}} + \text{h.c.} \right). \quad (3.22)$$

The inclusion of the  $\delta$  coupling is motivated by the fact that it is not forbidden by any symmetries. In addition, for  $\delta > 0$ , this coupling can be responsible for generating the necessary radial displacement  $f_I > f_a$ .

The operator in the last term of this equation will generate an additional contribution to the axion mass

$$m_{\text{eff}}^2 = \frac{|k| N^2 \langle S \rangle^{N-2}}{M_P^{N-4}}. \quad (3.23)$$

If  $\theta_N$  represents the minimum favored by the  $S^N$  operator, strong CP constraints require the contribution to the axion mass today,

$$\frac{\frac{m_{\text{eff},0}^2}{N} \sin(N |\theta_N - \theta_0|)}{m_{\text{QCD}}^2 + m_{\text{eff},0}^2 \cos(N |\theta_N - \theta_0|)} \lesssim \bar{\theta}, \quad (3.24)$$

where as above we have approximated the QCD potential as quadratic in the vicinity of the minimum, see Eq. (3.20). For small  $m_{\text{eff},0}^2$  (*i.e.*, supposing  $\sin(N |\theta_N - \theta_0|)$  not incredibly small and so neglecting the subdominant term in the denominator), we write the constraint as

$$m_{\text{eff},0}^2 \lesssim \frac{N \bar{\theta} m_{\text{QCD}}^2}{\sin(N |\theta_N - \theta_0|)}. \quad (3.25)$$

Taking  $\theta_i = |\theta_N - \theta_0|$  small as required by cosmology, see Eq. (3.13), and expanding this equation reproduces Eq. (3.19). Unless  $\sin(N |\theta_N - \theta_0|)$  is extremely close to zero—in particular, much closer even than required by cosmology—this places a stringent lower bound on  $N$  for a given  $(|k|, f_a)$  that generally requires  $N$  to be large,  $N \gtrsim \mathcal{O}(10)$ . In other words, as one might expect,  $U(1)_{\text{PQ}}$  needs to be a good symmetry to a high degree in order to solve the strong CP problem.

Clearly, this precludes the  $S^N$  operator from giving a large mass to the axion during inflation if  $f_I = f_a$ , even for  $|\theta_N - \theta_0| \ll 1$ . When  $I$  takes on large values, though, the  $I^2 |S|^2$  coupling can drive  $\langle S \rangle = \frac{f_I}{\sqrt{2}} > \frac{f_a}{\sqrt{2}}$ . Then, a large effective mass for the axial direction due to the  $S^N$  term may stabilize the phase of the PQ field at  $\theta_N$  and suppress fluctuations. After

inflation, the effect of the inflaton-PQ field cross-coupling will disappear, and the field will evolve to a minimum with  $\langle S \rangle = \frac{f_a}{\sqrt{2}}$  and  $\arg(S) \simeq \theta_N$  after a period of rolling and oscillation, where it will remain until  $m_{\text{QCD}}^2$  turns on.<sup>11</sup>

However, this coupling also influences the inflaton potential. The inflationary trajectory is constrained by limits on the slow-roll parameters given in Eq. (3.21),

$$r \approx 16\epsilon < 0.07, \quad (3.26)$$

$$n_s - 1 \approx -6\epsilon + 2\eta \in (-0.0413, -0.0253), \quad (3.27)$$

where we have given the 95% confidence bounds from [9, 20]. A reasonable constraint to avoid an overly tuned inflationary model is to require that the additional contributions to  $\epsilon, \eta$  do not exceed the maximal values consistent with present CMB observations.<sup>12</sup> For instance, requiring that the contribution be  $\lesssim 1$  of the maximal value corresponds to

$$\left| \frac{\Delta V'}{V} \right| \leq 0.094, \quad \left| \frac{\Delta V''}{V} \right| \leq 0.021. \quad (3.28)$$

We stress though that, while it represents a severe, *ad hoc*, dynamical tuning and so should be taken seriously, this constraint is aesthetic rather than experimental.<sup>13</sup>

In Fig. 3.2, we present an example of the constraints on such a model. Blue regions are excluded by strong CP constraints with cosmological bounds on the axion abundance taken into account, *i.e.*,  $|\theta_N - \theta_0|$  set by Eq. (3.13). The constraints are indiscernibly more stringent if one assumes  $|\theta_N - \theta_0| \simeq \mathcal{O}(1)$  as might be expected in the case of late-time dilution. The red region denotes where the inflaton potential is significantly modified, in

<sup>11</sup>While the potential barriers in the axial direction are small at small  $\langle S \rangle$ , the larger barrier at large  $\langle S \rangle$  will produce a “funnel” directing the phase towards  $\theta_N$  or, if oscillations are sufficiently large to take  $S$  through  $\langle S \rangle = 0$ , towards  $\theta_N + \pi$ .

<sup>12</sup>An analogous constraint on  $\xi$  or  $|\Delta V'''/V|$  from the running spectral index  $\alpha_s$  is subdominant. In addition, we have confirmed that  $\Delta V \ll V_I$  in these models.

<sup>13</sup>This constraint generally precludes the  $\langle S \rangle > M_P$  required in Section 3.1 from arising in this manner. For instance, supposing  $\Delta V = -\frac{\delta}{2} I^2 |S|^2$  we expect  $\delta I^2 \gtrsim H_I^2$  such that  $m_\sigma^2 \gtrsim H_I^2$  and  $|S|$  evolves to the large vev. This implies  $|\Delta V'/V| \sim f_I^2/(IM_P) \gtrsim 1$  and  $|\Delta V''/V| \sim f_I^2/I^2 \gtrsim 1$ , exceeding the stated bounds.

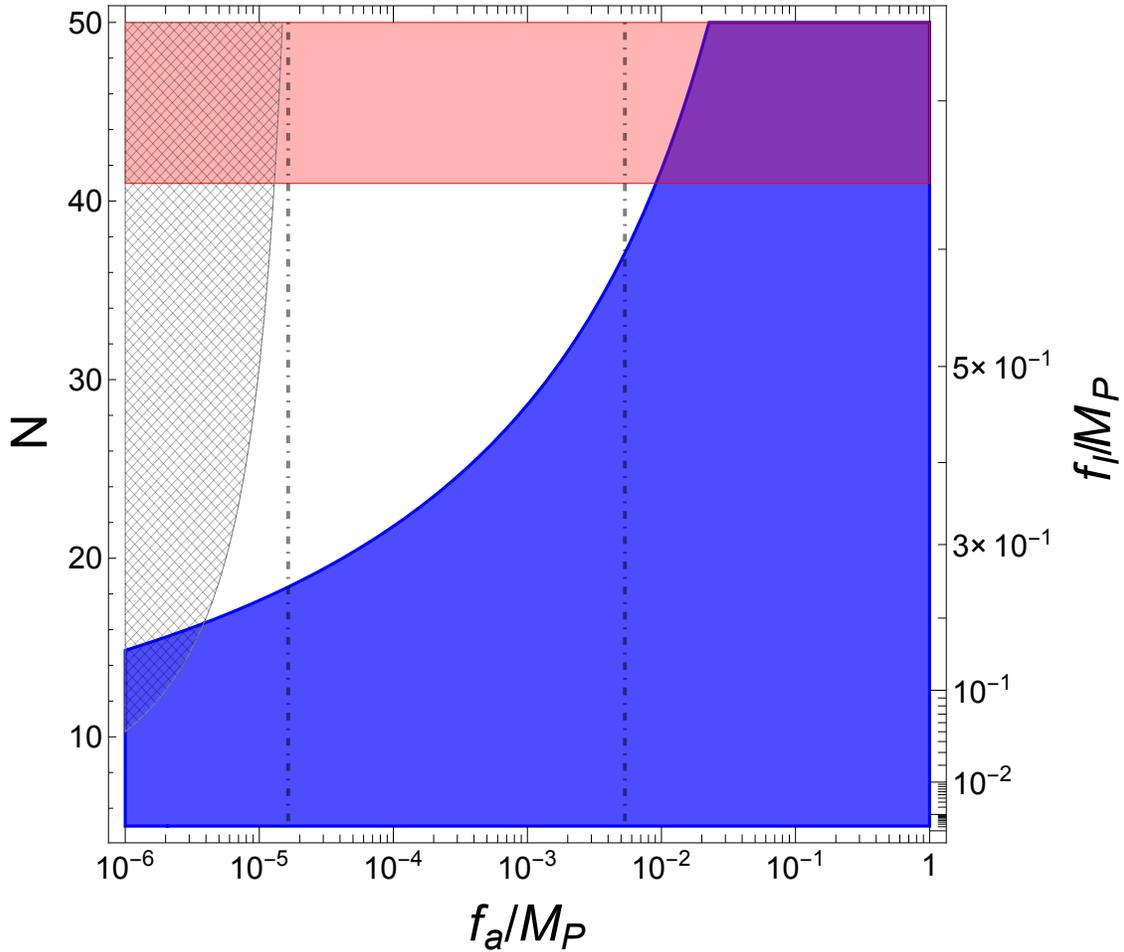


Figure 3.2: Bounds on the  $S^N$  model in the parameter space of  $f_a$  and  $N$  for  $|k| = 1$ . The right axis shows the required value for  $f_I$  to give the axion a large enough inflationary mass to suppress isocurvature. Blue corresponds to the strong CP constraint. Red indicates where the contributions to the slow-roll parameters from the  $S$  potential are greater than present bounds. The hatched region indicates where  $\langle S \rangle_e > 10^4 \frac{f_a}{\sqrt{2}}$  so PQ symmetry restoration via parametric resonance may occur (see text for caveats). Vertical lines indicate the CASPER Phase 2 (right) and ideal (left) reach (prospective bounds extend to the right) [11].

violation of Eq. (3.28). If these constraints are weakened by a factor of 3, this exclusion region moves up to  $N \gtrsim 130$ , potentially allowing  $f_a \simeq 3 \times 10^{-1}$ . We take  $H_I \simeq 10^{-5} M_P$ , consistent with  $10^{-3} \lesssim r \lesssim 10^{-2}$  that will be probed in near future experiments.<sup>14</sup> The Lyth bound [35] implies that, for such values of  $r$ , the excursion of the inflaton field over the course

<sup>14</sup>Note that  $f_I > \frac{H_I}{2\pi}$  throughout the parameter space, such that the PQ symmetry is broken during inflation even for  $f_a < \frac{H_I}{2\pi}$ .

of inflation is  $\frac{\Delta I}{M_P} \gtrsim 1$ . So, we set  $I_* = 5M_P$  at CMB mode horizon crossing (comparable to the value for a Starobinsky-like model at the edge of observability [272]).

The constraints shown are conservative in the sense that, for each value of  $N$ , we choose  $\lambda, \delta$  to be the smallest possible values such that the radial and axial directions both have masses  $\gtrsim H_I$ —larger couplings would result in larger contributions to the slow-roll parameters.<sup>15</sup> The axion mass condition is equivalent to requiring

$$f_I \simeq \left( \frac{H_I^2 M_P^{N-4}}{|k| N^2} \right)^{N-2}. \quad (3.29)$$

The CASPEr experiment [11]—with its Phase 2 reach of  $f_a \gtrsim 1.3 \times 10^{16}$  GeV and ideal reach of  $f_a \gtrsim 4 \times 10^{13}$  GeV—may eventually be able to probe much of the allowed region as indicated by the vertical lines.

The cross-hatched region denotes where the field value at the end of inflation  $\langle S \rangle_e > 10^4 \frac{f_a}{\sqrt{2}}$ , supposing that the value of the inflaton field at the end of inflation  $I_e$  is a factor of 5 smaller than at CMB horizon crossing (such that  $\frac{\langle S \rangle_e}{f_I/\sqrt{2}} \approx \frac{1}{5}$ ). According to the analysis of [260], for such values radial PQ field oscillations may excite large oscillations in the axial field, potentially leading to nonthermal symmetry restoration (see Eq. (3.9) and surrounding discussion). In this model, the situation is complicated by the presence of additional operators and the lack of initial fluctuations. For instance, the curvature of the potential in the axial direction due to the  $S^N$  operator may suppress fluctuations. Alternatively, self-couplings of the axial field resulting from this operator may mitigate the growth of fluctuations [273]. Thus, while this region is not necessarily excluded, a further analysis of field dynamics post-inflation would be required to ensure a consistent cosmology.

This analysis suggests that it is difficult to reach larger values of  $f_a \gtrsim 10^{-2} M_P$ . While the conclusion that a significant modification of the PQ potential can disrupt inflation is robust, the exact limits do depend on the inflationary model. For larger values of  $I_*$ , the same  $f_I$

---

<sup>15</sup>The required values of  $\delta$  are sufficiently small that parametric excitation of  $S$  due to inflaton oscillations is not a concern.

can be achieved for smaller values of  $\delta$  and  $\lambda$ , resulting in smaller contributions to slow-roll parameters. In addition, as long as  $I_e \simeq M_P$ ,  $\langle S \rangle_e$  would be reduced in the case of large  $I_*$ , shrinking the nonthermal symmetry restoration region. As such, perhaps paradoxically, high-scale inflation may reconcile more readily with a solution of this type—the steeper potential could be less susceptible to disruption. It should be noted, however, that larger  $I_*$  would also likely correspond to larger values of  $H_I \simeq 10^{-4} M_P$  ( $r \simeq 10^{-2}$ ), implying imminent tensor mode observation. Meanwhile, in a specific inflationary model, violation of Eq. (3.28) may result in a worse than  $\mathcal{O}(1)$  tuning. Recall, these equations were derived using current experimental constraints. For a Starobinsky-like model with  $r \simeq 2 \times 10^{-3}$ , the red region in Fig. 3.2 would more approximately correspond to tuning  $\gtrsim$  few, owing to the smaller denominators in Eq. (3.28). Requiring tuning  $\leq 1$  in this model would exclude  $N \gtrsim 31$ ,  $f_a \gtrsim 2 \times 10^{-3} M_P$ .

In sum, strong CP constraints require large values of  $N$  to ensure that the contribution to the axion mass due to the explicit breaking is small today. As such,  $f_I$  must be somewhat larger than  $f_a$  to ensure the axion is sufficiently heavy during inflation. In other words, large modifications to the PQ potential are required. These in turn “backreact” on the inflaton potential, at worst threatening to destabilize the fragile inflaton potential and at best constituting a very severe (field-dependent) tuning.

One reason that it is difficult to achieve a sufficiently large mass whilst satisfying strong CP constraints is that the same field is responsible for both the enhancement of the explicit breaking and solving the strong CP problem—the latter requires the potential of this field be largely  $U(1)_{\text{PQ}}$  invariant today, such that drastic modifications of its potential are required during inflation, readily disrupting inflationary dynamics. Consequently, in the next subsections, we discuss the possibility of realizing a large axion mass during inflation by coupling  $S$  to additional fields that acquire large vevs during inflation but are not subject to the same strong CP constraints. We will see that such approaches do indeed extend the reach in  $f_a$  relative to this simple model.

### 3.2.2 Inflaton-Sourced Axion Mass

At first glance, coupling to the inflaton would appear an ideal method for boosting the axion mass during inflation—the large (super-Planckian) value of  $I_*$  could give a significant contribution to  $m_{\text{eff,inf}}^2$  even via Planck-suppressed operators. Meanwhile, for small  $\langle I \rangle$  today, this contribution would be suppressed, alleviating strong CP constraints. However, as we discuss, the difficulty in this approach lies in generating the desired operators without generating either additional unwanted operators or symmetries (the latter of which, for instance, may prevent fields from acquiring a necessary mass).

One straightforward approach is to consider a term that explicitly breaks  $U(1)_{\text{PQ}}$  such as

$$V \supset \frac{kIS^N}{M_P^{N-3}} + \text{h.c.} \quad (3.30)$$

Such an operator was considered in, *e.g.*, [237].<sup>16</sup> If the rest of the potential maintains the symmetry  $I \rightarrow -I, S^N \rightarrow -S^N$ , this explicitly breaks  $U(1)_{\text{PQ}} \rightarrow \mathbb{Z}_{2N}$ . Then, for  $\langle I \rangle = 0$ , the leading operator expected to contribute to the axion mass would involve  $S^{2N}$ , permitting smaller values of  $N$  for a given  $(|k|, f_a)$  while still maintaining a solution to the strong CP problem.

However, for  $\langle S \rangle = \frac{f_a}{\sqrt{2}} \neq 0$ , Eq. (3.30) generates a tadpole for  $I$ , leading to  $\langle I \rangle \neq 0$  unless the potential is specifically tuned to stabilize  $\langle I \rangle \approx 0$ . If the inflaton potential is approximately quadratic near the origin,

$$V_I \simeq \frac{1}{2} m_I^2 I^2, \quad (3.31)$$

then,

$$\langle I \rangle \simeq \frac{2|k|(f_a/\sqrt{2})^N}{m_I^2 M_P^{N-3}} \cos \left( N \left\langle \frac{a}{f_a} - \theta_N \right\rangle \right). \quad (3.32)$$

This in turn drives  $\left\langle \frac{a}{f_a} - \theta_0 \right\rangle \neq 0$ —in fact, due to additional  $\left(\frac{M_P}{m_I}\right)^2$  enhancement, the effect

---

<sup>16</sup>Refs. [274, 275] consider the potential for similar operators to generate a cross-correlation spectrum in the case where isocurvature is unsuppressed.

of this term is expected to dominate over that of an  $S^{2N}$  operator supposing both operators are generated with similarly sized coefficients (unless  $m_I \sim M_P$ ). Consequently, strong CP considerations will still imply a somewhat stringent bound on  $N$  that is not quite a factor of 2 weaker than that for the model considered in Section 3.2.1.

Regardless, the combination of lower  $N$  and large  $I_*$  does allow a sufficiently large axion mass to be achieved with a smaller increase of  $\langle S \rangle$  during inflation, and hence less risk of destabilizing the inflaton potential. This opens up some parameter space that was not available in the model of Section 3.2.1. In Fig. 3.3, we show the analog of Fig. 3.2 with the  $S^N$  operator of Eq. (3.22) replaced by the  $IS^N$  operator of Eq. (3.30) assuming  $m_I \approx 10^{-5}M_P$  (representative of large-field models such as chaotic or Starobinsky-like inflation). Again, caveats about inflation model dependence such as those in the previous subsection apply.

This analysis assumes that Eq. (3.30) is the leading contribution to the axion mass. While operators with more powers of  $S$  will be suppressed as  $\langle S \rangle < M_P$ , this is not the case for operators including higher powers of  $I$ . For instance, the  $I \rightarrow -I, S^N \rightarrow -S^N$  symmetry discussed above would allow operators of the form  $I^{2M+1}S^N$ . Such operators may well be subdominant, though, as assumedly there exists some symmetry (for instance, a shift symmetry) responsible for maintaining the flatness of the inflaton potential. Hence, all operators containing  $I$  would also come with a spurion representing the breaking of this symmetry—if the appropriate combination to consider were, for instance,  $(kI)^{2M+1}$ , (where  $k$  is now the aforementioned spurion) then for  $|k| \ll \frac{I_*}{M_P}$  such higher-dimension operators would also be suppressed and the above analysis would hold. Reducing  $|k|$  reduces both the inflationary mass (such that larger  $f_I$  would be necessary to yield  $m_a \gtrsim H_I$ ) and the inflaton vev today, Eq. (3.32). As such, the red and blue regions in Fig. 3.3 would both move down, but by different amounts, resulting in a slightly reduced reach in  $f_a$ .

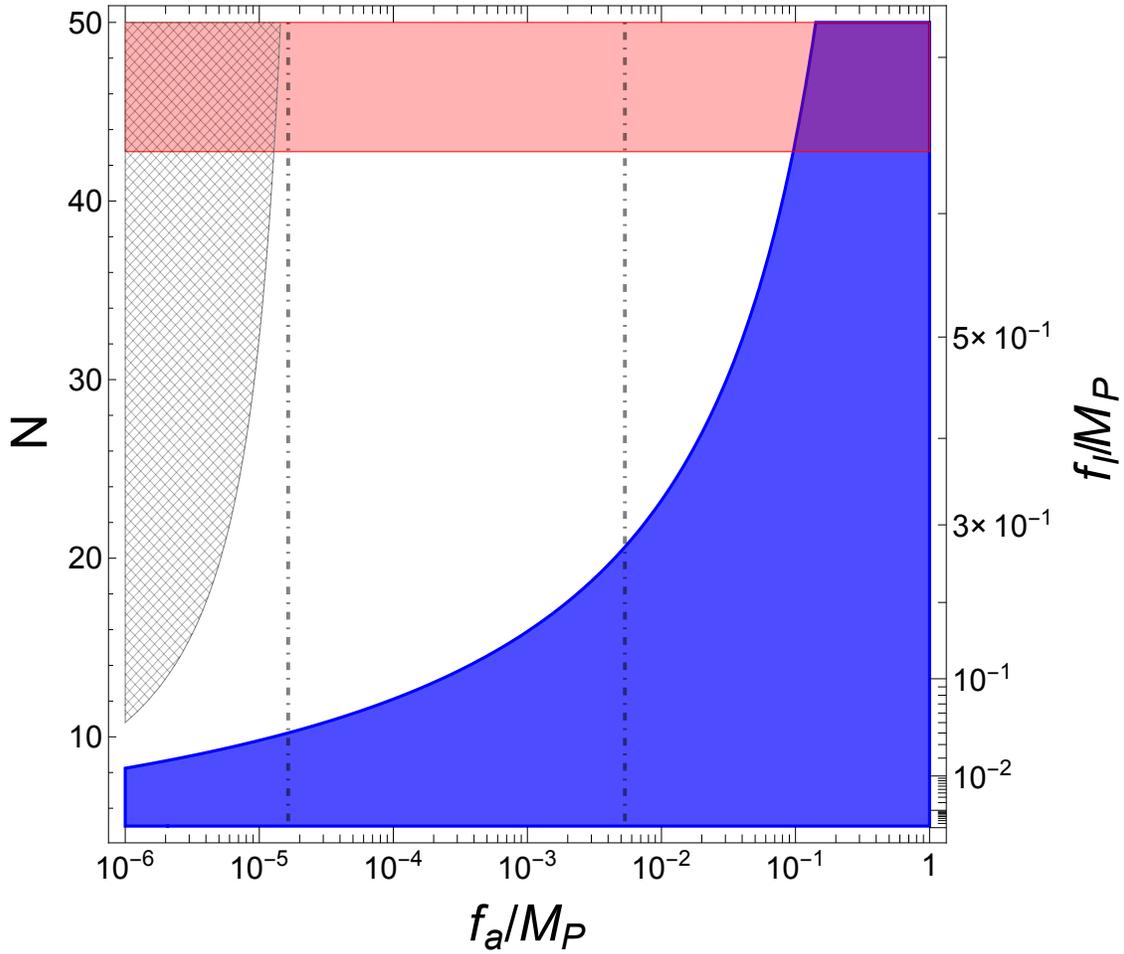


Figure 3.3: Bounds on  $IS^N$  model with  $m_I = 10^{-5}M_P$  and  $|k| = 1$ . All curves are the same as in Fig. 3.2.

### Challenges for Realizing $I^M S^N$ with $M > 1$

A tadpole for  $I$  would not arise if the leading explicit-breaking operator contained a higher power of  $I$ ,  $I^M S^N$ . But, in such a case, it is hard to imagine what (beyond coincidental cancellation) could effectively suppress lower dimension operators.

If  $I$  is real, then for odd  $M$  this operator maintains the  $I \rightarrow -I, S^N \rightarrow -S^N$  symmetry that permits the operator  $IS^N$ . For the reasons discussed above, we may well expect  $IS^N$  to dominate. Even less promising is the scenario in which  $M$  is even, as then the explicit breaking term violates  $S^N \rightarrow -S^N$  and so nothing should forbid the operator  $S^N$ .

If  $I$  were a complex field, one could imagine a global  $U(1)$  with charges  $q_S = \frac{1}{N}$ ,  $q_I = -\frac{1}{M}$

that, for appropriate choices of  $N$  and  $M$ , would give rise to  $I^M S^N$  without additional lower dimension operators. However, this global  $U(1)$  must be explicitly broken to avoid the presence of a massless field during inflation—in fact, it is indistinguishable from  $U(1)_{\text{PQ}}$ . In this setup, it becomes difficult to achieve a single, light inflaton (with mass  $\ll H$ ) without additional light fields. If this  $U(1)$  is a good symmetry during inflation (*i.e.*, broken sufficiently negligibly to suppress the generation of other worrisome operators), it should also relate the masses of the real and imaginary components of  $I$  to a good degree. Thus, if some component of  $I$  is light, the orthogonal component would be as well, potentially leading to isocurvature. Alternatively, if the  $U(1)$  is sufficiently badly broken that one component of  $I$  is quite massive while the orthogonal component is light (the inflaton), then it would also not be good enough to forbid other operators.

In supersymmetry (SUSY), holomorphy and nonrenormalization can suppress or forbid operators that one may have expected to be present from a more naïve symmetry analysis. Thus, SUSY appears promising for generating the desired  $I^M S^N$  operator without dangerous operators involving fewer powers of the inflaton field. However, once SUSY and  $U(1)_{\text{PQ}}$  are broken, dangerous operators contributing to the axion mass will be generated. So, while SUSY can ensure that these operators exhibit additional suppression relative to nonsupersymmetric models, for instance by the scale of SUSY breaking, this suppression is generally insufficient to circumvent the very stringent strong CP constraints.

Moreover, there is the well-known challenge that it can be difficult to realize large-field models of inflation within supergravity (SUGRA) (for a review, see, *e.g.*, [276]). The SUGRA scalar potential is given as a function of  $K$  and the superpotential  $W$  by

$$V_F = e^{K/M_P^2} \left( D_{\Phi_i} W K_{ij}^{-1} D_{\Phi_j} W^\dagger - 3 \frac{|W|^2}{M_P^2} \right), \quad (3.33)$$

with  $F_{\Phi_i} \equiv D_{\Phi_i} W = W_i + \frac{W K_i}{M_P^2}$  where  $W_i \equiv \frac{\partial W}{\partial \Phi_i}$  (and similarly for  $K$ ). The exponential factor  $\exp(K/M_P^2)$  induces large curvature at super-Planckian field values, a challenge for ensuring

slow roll. Thus, a viable model must generally invoke a symmetry (such as a shift symmetry) to prevent the inflaton from appearing in  $K$  [277]. In addition, supersymmetry implies that the field containing the inflaton is necessarily complex, so any continuous symmetries relating the various components of the field must be badly broken to ensure only one component remains light. This breaking combined with the presence of operators  $I^M S^N$  contributes to the breaking of  $U(1)_{\text{PQ}}$ .

To illustrate these points, consider a SUSY model with the following Kähler and superpotentials

$$K = \sum_{\Phi=X,Y,S,\bar{S}} \Phi^\dagger \Phi + \frac{1}{2}(I + I^\dagger)^2, \quad (3.34)$$

$$W = \lambda Y \left( S\bar{S} - \frac{f_a^2}{2} \right) + m_I X I + k X I \frac{S^N}{M_P^{N-1}}. \quad (3.35)$$

The Kähler potential respects a shift symmetry  $I \rightarrow I + ic$  for  $c \in \mathbb{R}$ , which is softly broken by  $m_I$  and  $k$ . This precludes the imaginary component of  $I$  from appearing in  $\exp(K/M_P^2)$ , giving rise to a viable model of chaotic inflation in which the imaginary component serves as the inflaton.<sup>17</sup> The fields  $S$  and  $\bar{S}$  carry opposite PQ charge and the PQ symmetry is spontaneously broken today by  $\langle S\bar{S} \rangle = \frac{f_a^2}{2}$  due to the first term in  $W$ .

Supposing  $X$  and  $Y$  are stabilized at the origin,  $X = Y = 0$ , the SUGRA scalar potential is

$$V = e^{K/M_P^2} \left\{ |\lambda|^2 \left| S\bar{S} - \frac{f_a^2}{2} \right|^2 + \left| m_I + k \frac{S^N}{M_P^{N-1}} \right|^2 |I|^2 \right\}, \quad (3.36)$$

in which the leading inflaton-PQ coupling comes from  $|I|^2 S^N$ , without an  $S^N$ -type term at this level. While no symmetry could prevent the generation of a term in the Kähler potential such as

$$\Delta K = \frac{k m_I^\dagger S^N}{M_P^{N-1}} + \text{h.c.}, \quad (3.37)$$

---

<sup>17</sup>Models of chaotic inflation are currently disfavored by bounds on  $r$  from Planck, so such a model would have to be modified to be consistent with experimental results. This could be done by, for instance, adding a nonminimal curvature coupling for  $I$  similar to Eq. (3.15). However, as our focus is the difficulty of embedding explicit PQ breaking in SUSY, we refrain from constructing a complete inflationary model here.

the contribution to the axion mass would vanish for  $I = 0$  and  $\langle S\bar{S} \rangle = \frac{f_a^2}{2}$  today. It would appear that the constrained form of the potential permitted in SUSY allows the desired inflationary explicit-breaking operators to be generated without correspondingly large explicit breaking today.

However, this approach is disrupted by the inclusion of SUSY breaking. For instance, if SUSY breaking is realized via a spurion superfield  $\xi = M_P + \theta^2 F_\xi$ , where  $|F_\xi|^2 \sim m_{3/2}^2 M_P^2$ , then operators such as

$$\Delta K = |\xi^\dagger \xi| \left( \frac{km_I^\dagger S^N}{M_P^{N+1}} + \text{h.c.} \right), \quad (3.38)$$

will generate terms in the scalar potential

$$V \supset \frac{km_I^\dagger m_{3/2}^2 S^N}{M_P^{N-1}} + \text{h.c.} \quad (3.39)$$

Even with the additional  $m_{3/2}$  suppression,  $N$  is constrained to be relatively large by strong CP considerations [278], reducing the efficacy of the  $|I|^2 S^N$  operator in giving mass to the axion during inflation. For instance we find that, for low-scale SUSY breaking  $m_{3/2} \sim \text{TeV}$  and  $m_I \simeq 10^{-5} M_P$ , the values of  $N$  required preclude  $m_{\text{eff,inf}}^2 > H_I^2$  for inflationary parameters such as those considered above if  $\langle S \rangle \sim f_a$  during inflation.

Additional parameter space would likely open if  $\langle S \rangle$  were boosted during inflation, but such a model would require a mechanism for generating the displacement and gains would be limited relative to the model of Section 3.2.1. Indeed, even for the model above, it would be necessary to explain why the PQ fields exhibited  $\langle S \rangle \sim \langle \bar{S} \rangle \sim f_a$ , as opposed to, *e.g.*, a minimum in which  $\langle S \rangle$  was such that the effective inflaton mass vanished.

So, while the large value of the inflaton makes it seemingly an ideal candidate for enhancing explicit PQ breaking during inflation, it is in fact difficult to generate the desired operators without residual dangerous contributions to the axion mass. This motivates considering whether the presence of another field, which still acquires a large vev during inflation but whose nature and symmetry properties are not as constrained as those of the inflaton,

might be able to enhance explicit breaking during inflation.

### 3.2.3 Additional $U(1)_{\text{PQ}}$ Fields

A challenge faced by models with a single PQ field is that the explicit breaking experienced by the axion pNGB during inflation, responsible for generating the large mass, is necessarily related to the explicit breaking today, which is constrained to be small by the strong CP problem. However, if there are multiple fields that contribute to the breaking of  $U(1)_{\text{PQ}}$ , the QCD axion will be a linear combination of the axial components of these fields. As such, strong CP constraints on each field will depend on the amount of the axion contained within that field. If the identity of the axion changes with time, it could “feel” more explicit breaking during inflation than today without running afoul of strong CP constraints.

To get a sense of the issues one must consider in constructing a viable model, we consider a “toy” consisting of two PQ fields,  $S$  and  $\bar{S}$ , with charges  $q_S = 1, q_{\bar{S}} = -K$  under  $U(1)_{\text{PQ}}$ , which is explicitly broken to  $\mathbb{Z}_{KN}$ . As such, the lowest dimension Planck-suppressed local operators are

$$V \supset \frac{k\bar{S}S^K}{M_P^{K-3}} + \frac{k'\bar{S}^N}{M_P^{N-4}} + \frac{k''S^{KN}}{M_P^{KN-4}} + \text{h.c.} \quad (3.40)$$

In the limit where  $U(1)_{\text{PQ}}$  is not explicitly broken ( $k' = k'' = 0$ ), the QCD axion and PQ breaking scale are

$$a_{\text{QCD}} = \frac{1}{v_a} \sum_{i=S,\bar{S}} q_i v_i a_i, \quad v_a^2 = \sum_{i=S,\bar{S}} q_i^2 v_i^2, \quad (3.41)$$

where  $a_S$  denotes the axial component of  $S$  and  $\langle S \rangle \equiv v_S$ , and similar for  $a_{\bar{S}}, v_{\bar{S}}$ . The  $\bar{S}^N$  and  $S^{KN}$  operators both contribute to a mass for the axion, with the dominant contribution coming from the former.  $\bar{S}S^K$  is  $U(1)_{\text{PQ}}$  invariant and so, while it gives mass to the orthogonal axial field  $a_{\perp}$ , it only contributes to  $m_{\text{eff}}^2$  via heavily suppressed mass mixing.

The vevs of  $S$  and  $\bar{S}$  are set by the potential,

$$V \supset \lambda_S \left( |S|^2 - \frac{f_S^2}{2} \right)^2 + m_{\bar{S}}^2 |\bar{S}|^2 + \frac{\lambda_{\bar{S}}}{4} |\bar{S}|^4 - \frac{\delta_{\bar{S}}}{2} I^2 |\bar{S}|^2. \quad (3.42)$$

Today, when  $\langle I \rangle = 0$  is assumed,  $v_{\bar{S},0} \neq 0$  is driven by the tadpole for  $\bar{S}$  induced via the first term of Eq. (3.40) when  $v_{S,0} \simeq \frac{f_S}{\sqrt{2}}$ . For instance, neglecting the terms with powers of  $\bar{S}$  greater than two,<sup>18</sup>

$$\langle \bar{S} \rangle_0 \equiv v_{\bar{S},0} \simeq \frac{kv_{S,0}^K}{m_{\bar{S}}^2 M_P^{K-3}}. \quad (3.43)$$

Thus,  $v_{\bar{S},0}$  can be considerably smaller than  $v_{S,0}$  due to the  $\left(\frac{v_{S,0}}{M_P}\right)^{K-3}$  suppression, supposing  $m_{\bar{S}}^2 \ll v_{S,0}^2$ . In this case, the present day axion resides predominantly in  $S$  and  $v_a \simeq v_{S,0}$ , allowing  $N$  significantly smaller than  $KN$  consistent with strong CP constraints.

Meanwhile, during inflation,  $v_{\bar{S}}$  is enhanced to  $v_{\bar{S},I} > v_{S,I}$  by the coupling to the inflaton in the second line of Eq. (3.42). However, we assume  $v_{S,I} \simeq v_{S,0} \simeq \frac{f_S}{\sqrt{2}}$  by taking  $\lambda_S$  large enough that the cross coupling for  $S$ ,  $I^2 |S|^2$ , is negligible.<sup>19</sup> Because  $v_{\bar{S},I} > v_{S,I}$ ,  $a_{\text{QCD}}$  is dominantly composed of  $a_{\bar{S}}$  during inflation and so receives a large explicit PQ-breaking mass from  $\bar{S}^N$ . The mass of  $a_{\perp} \simeq a_S$  from the  $\bar{S}S^K$  term is similarly enhanced by large  $v_{\bar{S},I}$ . Therefore, all fields are heavy during inflation, eliminating isocurvature constraints.

The modification to the PQ potential required to enhance  $v_{\bar{S}}$  during inflation still risks destabilizing the inflaton potential. In Fig. 3.4, we show an example of the constraints on such a model with  $|k| = |k'| = |k''| = 1$  and  $m_{\bar{S}}^2 = 10^{-11} M_P^2$ . For a given  $K$  and  $N$ , we plot in black contours the maximal allowed value of  $f_a \equiv \sqrt{2}v_a$  such that the contributions to the axion mass today due to explicit breaking do not disrupt the solution to the strong CP problem—*i.e.*, both  $N$  and  $KN$  are sufficiently large that the  $U(1)_{\text{PQ}}$ -breaking operators of Eq. (3.40)

<sup>18</sup> The quartic term can be neglected for  $\lambda_{\bar{S}} \ll \frac{2m_{\bar{S}}^6 M_P^{2K-6}}{|k|^2 f_S^{2K}}$ .

<sup>19</sup> Similarly, we neglect terms such as  $|S|^2 |\bar{S}|^2$ . For constant  $\langle S \rangle$ , this term can be considered a contribution to  $m_{\bar{S}}^2$ .

give small contributions to  $m_{\text{eff}}^2$ . As above, we implement strong CP constraints subject to the cosmological requirement that the minimum to which fields evolve after inflation is not significantly displaced from that favored by QCD (*i.e.*, analogous to Eq. (3.19)).

To simplify the analysis, we take  $v_{\bar{S},0}$  to be given by Eq. (3.43), which is of order  $10^{-5}$  to  $10^{-3}M_P$  in the allowed region. As discussed below Eq. (3.43),  $v_{\bar{S},0}$  is not suppressed relative to  $v_{S,0}$  for small  $K$ , and thus  $a_{\text{QCD},0} \simeq a_{\bar{S}}$ —*i.e.*, the composition of the axion is largely the same during inflation and today. This results in dynamics essentially equivalent to those of the model of Section 3.2.1. As such, in this (horizontal-hatched) region there is no particular advantage to multiple PQ fields, though the model is viable.

Regions corresponding to (overly) large shifts to the slow-roll parameters in excess of 1 (10) times the current experimental constraints on the potential (Eq. (3.28)) are shown in light (dark) red.<sup>20</sup> Within the dashed curve, the tadpole approximation of Eq. (3.43) breaks down. Since this region is well excluded by fine-tuning considerations, we do not attempt to improve upon this approximation.

In the allowed region,  $\frac{M_P}{5} \lesssim f_{\bar{S},I} = \sqrt{2}v_{\bar{S},I} \lesssim \frac{M_P}{2}$  is the minimum value required to give a sufficiently large mass  $\gtrsim H_I$  to both  $a_{\text{QCD}}$  and  $a_{\perp}$  (outside the allowed region it differs by less than an order of magnitude except in the darker red region). The cross-hatched region denotes where the initial amplitude of the resulting  $\bar{S}$  fluctuations is very large,  $v_{\bar{S},e} > 10^4 v_{\bar{S},0}$  (as before, subscript  $e$  denotes the end of inflation). Such large oscillations potentially produce large fluctuations in the phases of  $S, \bar{S}$  that could result in additional contributions to the axion abundance, domain walls, or nonthermal symmetry restoration. But this is not necessarily the case, and further investigation would be required to determine whether or not this region yields a consistent cosmology—see discussion in Section 3.2.1.

In Fig. 3.4, we have tuned  $m_{\bar{S}}^2$  to be just less than  $H_I^2$ . That this allows for near maximal reach in  $f_a$  can be understood as follows. The non-QCD contribution to the axion mass goes

---

<sup>20</sup>As above,  $\delta_{\bar{S}}, \lambda_{\bar{S}}$  are conservatively fixed by the requirements that the potential is minimized at  $\langle \bar{S} \rangle = v_{\bar{S},I}$  and that the mass of the radial direction is sufficiently large  $\simeq H_I$  that the field does indeed rapidly evolve to this minimum during inflation.

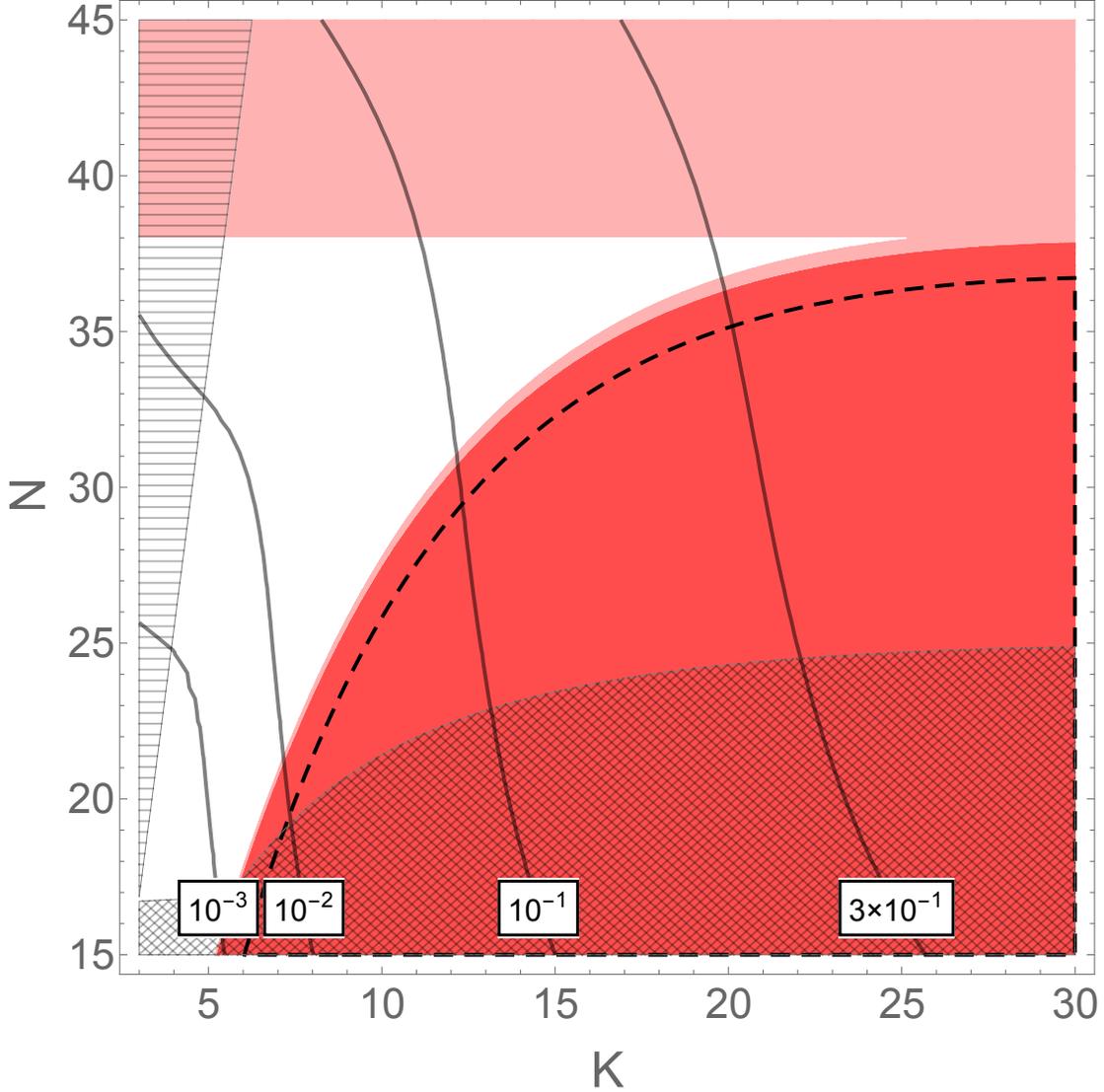


Figure 3.4: Bounds on multiple PQ field model for  $|k| = |k'| = |k''| = 1$ ,  $m_{\bar{S}}^2 = 10^{-11}M_P^2$ . Contours of the maximal  $\frac{f_a}{M_P}$  allowed by strong CP constraints are solid black. Regions where the contributions to the slow-roll parameters are greater than 1 (10) times current bounds are shaded light (dark) red. The region where  $v_{\bar{S},e} > 10^4 v_{\bar{S},0}$  so that symmetry restoration via parametric resonance may be a concern is crosshatched. Within the dashed black contour denotes where our assumption that the  $\lambda_{\bar{S}}$  term is negligible today breaks down. In the horizontal-hatched region,  $a_{\text{QCD}} \simeq a_{\bar{S}}$  always.

as  $m_{\text{eff}}^2 \propto \frac{v_{\bar{S},0}^N}{v_{\bar{S},0}^2} \propto \frac{f_S^{KN-2}}{m_{\bar{S}}^{2N}}$ . So, for smaller  $m_{\bar{S}}^2$  and fixed  $f_a$ ,  $K$  and/or  $N$  must be increased slightly to ensure that the contribution to  $m_{\text{eff}}^2$  is sufficiently small today. However, small increases in  $K$  exponentially increase the necessary  $v_{\bar{S},I}$  required to give a large enough mass to  $a_{\perp}$  during inflation, since its mass is proportional to only a single power of  $\bar{S}$ .

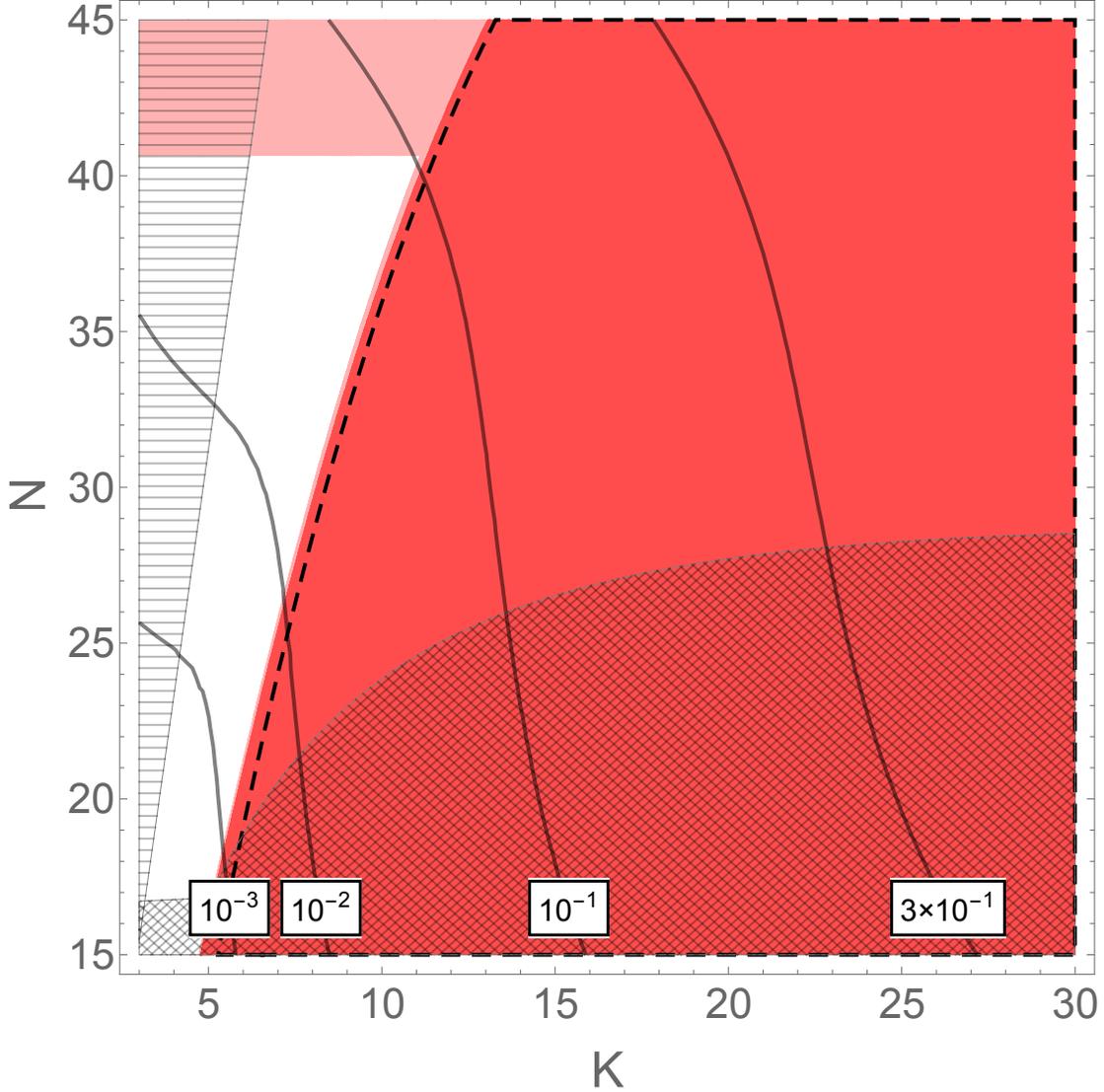


Figure 3.5: Bounds on multiple PQ field model for  $m_{\bar{S}}^2 = 10^{-12}M_P^2$ . All curves are the same as in Fig. 3.4.

Thus, the bottom right red region moves significantly to the left while the black contours move slightly to the right—when taken together, this can drastically reduce the values of  $f_a$  achievable. Along similar lines, slightly larger  $f_a$  may be accommodated by tuning  $m_{\bar{S}}^2$  marginally closer to  $H_I^2$ . But gains are limited because, as  $m_{\bar{S}}^2$  approaches (or exceeds)  $H_I^2$ , larger  $\delta_{\bar{S}}$  is also eventually required to overcome the positive mass-squared parameter, such that slow-roll constraints rapidly exclude the whole parameter space. Thus, Fig. 3.4 represents close to the maximal (rather than typical) reach for a model of this type. For

comparison, in Fig. 3.5 we show the analog of Fig. 3.4 for  $m_{\bar{S}}^2 = 10^{-12}M_P^2$ . As  $m_{\bar{S}}^2$  decreases, the region consistent with slow-roll constraints moves to overlap with the horizontal-hatched region where  $a_{\text{QCD}} \simeq a_{\bar{S}}$  always, with (horizontal) slow-roll constraints excluding  $N \gtrsim 40$  in agreement with Section 3.2.1.

Further parameter space could also be reached if  $v_{S,I} > v_{S,0}$ . Such a setup requires a mechanism for boosting  $\langle S \rangle$  as well. Constraints on  $N$  would vanish entirely for  $v_{\bar{S},0} = 0$ , which could perhaps be achieved for different  $q_{S,\bar{S}}$ , but for vanishing vev it is difficult to achieve efficient  $\bar{S}$  decay. For instance, if  $\bar{S}$  couples to PQ quarks via  $\bar{S}Q\bar{Q}$ , couplings such as  $S^M Q\bar{Q}$  are forbidden, precluding the quarks from acquiring mass.

As for the models explored in the previous subsection, this setup with multiple PQ fields and changing vevs can open parameter space at larger  $f_a$ , potentially even allowing models with  $f_a \sim M_P$ . However, for these largest values, remaining consistent with isocurvature constraints while avoiding excessive disruption of inflaton dynamics is still difficult and requires some conspiracy between  $K, N, m_{\bar{S}}^2$  and inflationary parameters. Similar model building may even allow  $f_a \gtrsim M_P$ . But, in this regime, the analysis presented here may well be insufficient as higher-order operators, no longer necessarily suppressed, could play a role. Moreover,  $f_a \gtrsim 10^{-1}M_P$  is excluded by bounds from black hole superradiance [97].

### 3.3 Additional Constraints on Radial Field Displacement

There are various other constraints that we have not so far considered when the radial field is significantly displaced during inflation from its present minimum. Mostly these are model dependent and can be avoided under appropriate circumstances that we detail.

### 3.3.1 Perturbations in the Radial Field

A light PQ field that was displaced due to Hubble friction would obtain primordial fluctuations  $\delta\sigma = \frac{H_I}{2\pi}$ , which are orthogonal to the curvature perturbations seeded by the inflaton. The decay products of  $\sigma$  will inherit these fluctuations, potentially providing an additional source of isocurvature.

This may be erased if local thermal equilibrium is achieved [279]. Alternatively, if the radial mode comes to dominate the energy density and its decay products reheat the universe, it will effectively act as a “curvaton” [280]—the observed perturbations result from  $\sigma$  fluctuations as opposed to those of a separate inflaton, so are not observed as isocurvature modes. If  $\sigma$  decays when its energy density is subdominant, it can still seed the observed curvature perturbations in a curvaton-like fashion if inflationary curvature perturbations are subdominant or absent. Although, in this case, the perturbations arising from  $\sigma$  must be larger to compensate for its subdominance, which frequently results in sizable non-Gaussianities, see *e.g.* [281]. In the event that  $\sigma$  does indeed decay while subdominant and is not a curvaton, exact constraints depend on the epoch—for a detailed analysis in the context of moduli, see [282].

### 3.3.2 Scalar Trapping

If the radial field is responsible for giving the PQ quarks (and squarks, if present) mass, symmetry restoration may occur as a result of scalar trapping as described in [283]. This trapping proceeds as follows. When  $|S|$  becomes small during its oscillations, the PQ (s)quarks become light enough to be produced via thermal effects or parametric resonance. These (s)quarks then backreact on  $|S|$ , leading to it becoming trapped at the origin and thus effectively restoring the PQ symmetry. For this to occur, the initial oscillation amplitude must be large enough that  $|S|$  passes near the origin, which only occurs for  $|S|_i \gtrsim 10 \frac{f_a}{\sqrt{2}}$ . Scalar trapping may be avoided if there are no PQ squarks, there are additional sources of mass for the PQ (s)quarks (*e.g.*, in multifield PQ models), or the radial field is sufficiently heavy that it does

not spend a significant time in the critical regime during each oscillation.

### 3.3.3 Radial Field Energy Density

As the  $\sigma$  oscillations decay, the corresponding energy density is transferred into its decay products, leading to constraints that depend in detail upon when the decay occurs and to what final state(s) [284]. If it decays to axions, they act like dark radiation, which is constrained by measurements of the CMB and the success of BBN. Alternatively, the radial field can decay predominantly to colored states (including gluons or PQ quarks), which would avoid dark radiation constraints. The decays of the radial field must not disrupt BBN, so it must either decay before the start of BBN or remain a (substantially) subdominant component of the energy density. Both conditions favor earlier decay of the radial field—before BBN or before its proportional energy density has increased significantly—which generally corresponds to heavier PQ fields. Such a heavy field is possible in a generic model, but could pose a problem in supersymmetric models where there is necessarily a light saxion field with mass of order the SUSY-breaking scale due to the complexification of the  $U(1)_{\text{PQ}}$  symmetry [285].

Decays before BBN can still be constrained by the observed DM relic density if  $\sigma$  decays to DM after thermal freeze-out. Again, this constraint may be more difficult to avoid in supersymmetric modes with R-parity, *e.g.*, for a saxion decaying to superpartners whose decays yield stable lightest supersymmetric particles. The extent to which this is a concern depends on whether supergravity effects yield  $\text{Br}(\sigma \rightarrow \text{gauge bosons}) \simeq \text{Br}(\sigma \rightarrow \text{gauginos})$ , as suggested in [286], in which case a sizable branching ratio to gluons (desired to avoid dark radiation constraints) would be accompanied by a sizable ratio to gluinos. However, other analyses suggest that decays to gauginos suffer from additional chiral suppression (*i.e.*, by the mass of the gaugino) [287], in which case decays to superpartners would be subdominant.

### 3.4 Can the Radial Field Evolve Adiabatically?

Constraints arising from  $\sigma$  oscillations (both thermal and nonthermal) as discussed in the previous sections would be evaded if  $\langle S \rangle$  evolved gradually to  $f = f_a$ . This can only occur if the PQ potential does not change violently at the end of inflation, which does not occur for generic couplings between the PQ and inflationary sectors. Typically, the end of inflation corresponds to a radical change in dynamics in the inflationary sector—in slow-roll models, the inflaton leaves the slow-roll regime and begins to coherently oscillate around the minimum of its potential—and as such a similarly drastic change to the PQ potential is to be expected. In all likelihood, the mechanism responsible for maintaining  $\langle S \rangle = \frac{f_I}{\sqrt{2}}$  rapidly disappears and the radial component of the PQ field begins oscillating around the zero temperature minimum  $\langle S \rangle = \frac{f_a}{\sqrt{2}}$ .

However, if the size of the operator coupling the PQ and inflationary sectors decreased gradually,  $\sigma$  and  $\langle S \rangle$  could conceivably evolve adiabatically to a lower value. The obvious candidate for such a solution is for  $S$  to couple to the full energy density of the inflaton  $\rho_I$ , which is nonoscillatory but rather decreases gradually up until inflaton decay. An analogous approach has been considered previously to alleviate constraints on moduli energy density through assuming moduli couple directly to  $H^2 \propto \rho_{\text{total}}$  (which is equal to the inflaton energy density prior to reheating) [288]. Couplings proportional to  $H$  have also been invoked to reduce the amplitude of saxion oscillations [289] and in the context of models to suppress axion isocurvature via a sufficiently high  $f_I$  [238, 290].

Unfortunately, there are two obstacles to invoking such a solution. First, even supposing a coupling such as  $cH^2 |S|^2$  does dominate, coherent oscillations are generally diminished but not completely avoided. In a SUSY model, for instance, the minimum today is determined by the soft masses, so is likely displaced  $\gtrsim f_a$  from the minimum preferred by  $\mathcal{O}(H^2)$  masses (see, *e.g.*, [284]). Moreover, in a general model, adiabatic tracking reduces the initial amplitude of oscillations, but oscillations still commence eventually, with the extent of the reduction depending on the magnitude of  $c$  and the other terms in the  $S$  potential [263, 288, 291].

Suppressing the initial amplitude by even an order of magnitude requires  $c \gtrsim \mathcal{O}(10)$  [288, 291]—*e.g.*, for  $c = (4\pi)^2$  and a quartic PQ potential, the initial oscillation amplitude is only reduced by a factor of  $\sim 10$  [263]. As alluded to in previous sections, too large  $c$  risks interfering with inflationary dynamics, particularly for models requiring (super-)Planckian  $f_I$ .

Second, implementing such a coupling to  $\rho_I$  appears difficult from a model-building standpoint, as it requires tuning between the couplings of the PQ field and the inflaton kinetic and potential terms. We elaborate on this issue below, but stress again that even if such a model can be successfully constructed it only realistically reduces the oscillation amplitude by a factor of  $\mathcal{O}(10)$ . So, at the end of the day, the constraints of the previous sections are likely still significant.

### 3.4.1 Implementing Adiabatic Relaxation of the PQ Field

As mentioned, at the most basic level, coupling to a nonoscillatory or smoothly varying quantity requires a tuning between the couplings of the PQ field to the inflaton kinetic and potential terms. One might hope to invoke a symmetry to enforce the required coincidence; a frequently considered candidate is supersymmetry. For instance, a SUSY inflationary model containing a Kähler potential coupling

$$K \supset -c \frac{|I|^2 |S|^2}{M_P^2}, \quad (3.44)$$

where  $I$  is an inflaton field whose potential is dominated by its  $F$ -term,  $V_I \simeq |F_I|^2$ , exhibits a coupling between  $|S|$  and the energy density of  $I$  as  $\langle \int d^4\theta |I|^2 \rangle = \rho_I$  [292, 293],

$$\mathcal{L} \supset -\frac{c}{M_P^2} (|\partial_\mu I|^2 + V_I) |S|^2 = -3cH^2 |S|^2, \quad (3.45)$$

where the last equality holds when  $\rho_I$  is the dominant component of the energy density of the universe. Supposing other operators were subdominant, the evolution of  $\langle S \rangle$  would be

determined by the gradually red-shifting  $\rho_I$ . Moreover, for  $c \gg 1$ , the highly curved nature of the potential in the vicinity of the vev would cause  $S$  to evolve rapidly to the minimum, such that the field approximately tracked the adiabatically evolving vev without large oscillations.

However, it is not obviously sufficient to consider only these terms—for instance, as observed in [293], a number of comparably sized operators will generically arise from various sources. In particular, SUGRA corrections will generate additional couplings that disrupt the relationship between the coupling of  $S$  to inflaton kinetic and potential terms. Notably, expanding the exponential in Eq. (3.33) will lead to additional couplings of  $S$  to  $I$  in the scalar potential without corresponding couplings to the kinetic terms for  $I$ .

One might worry that these terms would spoil the success of this solution in suppressing large coherent oscillations. Indeed, when  $I$  starts to oscillate, the additional terms, which necessarily play a role in determining  $\langle S \rangle$ , would rapidly change or disappear. As such, the vev would also rapidly change and so  $S$  too would be expected to start oscillating. But, for the large values of  $c \gg 1$  required for this adiabatic tracking mechanism to work, the coupling proportional to  $H^2$  may dominate, in which case the change to  $\langle S \rangle$  could be small. Furthermore, the coherent oscillations would occur around the new nearby minimum as opposed to the minimum today and would be rapidly damped due to the large effective mass  $\sim \sqrt{c}H$ . So, small perturbations introduced by couplings to the oscillating inflaton field  $I$  as opposed to  $H^2$  do not necessarily disrupt the tracking. However, the post-inflationary dynamics of both the inflaton and PQ fields are potentially complicated in this scenario, and depend on their coupled equations of motion. The exact behavior will depend on model-specific details and an analysis of the viability of this solution in well-motivated examples is an interesting question, albeit beyond the scope of this work.

A more serious concern is that the absence of tuning permitted by SUSY in this case—*i.e.*, that a single term can generate the desired coupling to  $\rho_I \propto H^2$ —relies on the inflaton potential being dominated by its  $F$ -term,  $F_I$ . As is well known, SUSY inflationary models of this type exhibit a severe  $\eta$  problem, namely that  $\eta \simeq \mathcal{O}(1)$  if various contributions are not

tuned against one another (for a review see, *e.g.*, [28, 294])—notably, sizable contributions to  $\eta$  arise from the exponential in Eq. (3.33). The  $\eta$  problem is potentially exacerbated in this case, especially for  $c \gg 1$  and  $|S| \sim M_P$ . So, while tuning might not be required to get the desired coupling between the inflaton and PQ field, it may still be necessary to yield a viable model of inflation.

On the other hand, models do exist in which the  $\eta$  problem is solved without tuning by an additional symmetry for the inflaton, such as the model considered in Section 3.2.2. Then, though, the inflaton potential is not dominated by  $F_I$  and so additional tuning would be required to achieve a dominant coupling to  $H^2$ . For instance, in the model discussed previously with

$$K \supset |X|^2 + \frac{1}{2} (I + I^\dagger)^2, \quad W = m_I X I, \quad (3.46)$$

the shift symmetry  $I \rightarrow I + ic$  prevents the imaginary scalar component of  $I$ ,  $\phi_I$ , from appearing nonderivatively from the Kähler potential. As such,  $\phi_I$  can take on large field values without an associated  $\eta$  problem, allowing it to act as the inflaton while the heavy real component of  $I$  and the  $X$  scalar are stabilized at the origin. In this model, though, inflation is driven by the  $F$ -term for  $X$  rather than that for  $I$ ,

$$V_F \simeq |F_X|^2 = \frac{m_I^2}{2} \phi_I^2, \quad (3.47)$$

where we have taken  $X = \text{Re}(I) = 0$ . So, equal coupling to the kinetic and potential terms for  $\phi_I$  requires multiple additional terms in the Kähler potential

$$K \supset |S|^2 - \frac{c}{2M_P^2} (I + I^\dagger)^2 |S|^2 - \frac{d}{M_P^2} |X|^2 |S|^2 \quad (3.48)$$

and a tuning  $c \simeq d - 1 \gg 1$ . Moreover, higher order terms such as  $\frac{|S|^4 |F_X|^2}{M_P^4}$  are still generated, which may affect the dynamics of the fields after inflation.

Furthermore, a similar “ $\eta$ ” problem exists for the PQ field, making it difficult to stabilize

$\langle S \rangle$  at a large field value. Specifically, the exponential barrier tends to drive  $\langle S \rangle \rightarrow 0$  or  $\langle S \rangle \rightarrow \infty$  during inflation. Unlike for a generic modulus, a shift symmetry solution cannot be invoked for  $S$  to prevent it from appearing explicitly in  $K$  as such a symmetry is incompatible with  $U(1)_{\text{PQ}}$ . The tunings required to avoid the inflaton  $\eta$  problem are not the same as those required to generate the desired potential for  $S$ , such that the solutions to these problems are not necessarily related.

Both problems are more severe for large-field models as the field values  $I_* \gtrsim M_P$  during inflation mean higher-order operators are effectively unsuppressed. As these are exactly the models that yield observable scalar-to-tensor ratios and for which axion isocurvature is a major concern, it is unclear that an adiabatic solution can be readily implemented to suppress isocurvature without large coherent oscillations. At the very least, both a precise tuning of various terms and large couplings appear necessary to simultaneously achieve inflation, the desired inflationary minimum, and adiabatic tracking behavior, even with supersymmetry.

### 3.5 Conclusions

In models where the PQ symmetry breaks before the end of inflation, a high scale of inflation naïvely induces too large isocurvature perturbations in the CMB. However, if the PQ sector is modified from its zero temperature form during inflation, this need not be the case.

One possibility is that, during inflation, the PQ field is still evolving towards its minimum from a large initial value. This occurs, for example, in a model wherein the radial component of the PQ field acts as the inflaton. In this scenario, however, post-inflationary radial oscillations can potentially induce the restoration of the PQ symmetry, with disastrous cosmological consequences. Indeed, the “PQ sector inflation” approach only seems viable if there is a late-time release of entropy. Even then, this is only the case for a narrow window  $f_a \sim 10^{-3} M_P$ , and more detailed studies would be required to ensure symmetry restoration would not occur. Moreover, as constraints on isocurvature tighten, this class of solution will

become increasingly untenable.

Another approach is to induce modifications of the PQ potential by explicitly breaking the PQ symmetry and coupling the PQ sector to the inflaton. However, this has its own set of challenges. The inflaton potential is necessarily delicate—it must satisfy the slow-roll conditions. Thus, maintaining a sufficiently flat potential in the presence of large couplings between the PQ sector and the inflation sector can necessitate aesthetically unpleasing fine-tunings. And the axion potential must be dominated by QCD—additional contributions run the risk of spoiling the elegant solution to the strong CP problem. With judicious choice of potential, it is indeed possible to avoid these concerns, and GUT-scale or even Planck-scale  $f_a$  is allowed. These models typically rely on a discrete symmetry, with a delicate explicit breaking which is amplified during the inflationary period. Furthermore, this solution does not allow a small axion abundance via an anthropically chosen initial misalignment angle  $\theta_i$ , so either a coincidence in the phase of the operator, or a late-time dilution (perhaps due to a modulus) is required.

Given the fragility of both the inflaton and the axion, it might be productive to introduce yet another field that amplifies the PQ breaking. This could insulate the inflaton against the fine-tuning effects discussed above. One possibility might be a modulus that obtains a large vev during inflation. However, potential reintroduction of the strong-CP problem by, *e.g.*, generation of a tadpole, is still a concern. The viability of this solution is an interesting direction for further work.

Even if isocurvature is suppressed by one of the above mechanisms or similar, precautions must be taken to ensure a consistent cosmology. This is particularly true if the transition to the field configuration today is violent, involving large coherent field oscillations or significant energy density stored in late-decaying fields. While these challenges are well known from the physics of moduli and saxions and can be evaded by appropriate model building, they must be taken into account and may well constitute the dominant constraints, especially as coherent oscillations can perhaps be reduced but generally not eliminated. Additional fields (multiple

PQ fields, separate inflatons, moduli) offer more alternatives for suppressing isocurvature, but their nontrivial post-inflationary dynamics will also yield additional constraints.

If primordial tensor perturbations are indeed observed in the CMB, models with large  $f_a$  will require some additional physics coupled to the PQ sector. If residual isocurvature is also observed, it may be the case that isocurvature is simply suppressed by  $f_I \gg f_a$ . Alternatively, if no isocurvature is visible, it may be that a large mass for the axion is generated during inflation. In either case, future precision probes of the CMB and axion DM stand to tell us much about high-scale axions, the inflationary sector, and, perhaps, an interplay between the two.

# Chapter 4

## Inflationary Theory and Pulsar Timing Investigations of Primordial Black Holes and Gravitational Waves

*This chapter was completed in collaboration with Aaron Pierce and James Wells [156].*

The LIGO detection of gravitational waves (GWs) resulting from the merger of black holes with masses  $\sim 10$  to  $30M_{\odot}$  [134, 135] may be the first detection of primordial black holes (PBHs). The constraints on PBHs discussed in Section 1.4 may indeed be compatible with this interpretation, with most constraints allowing  $f \sim 0.1$  to 1 and disputed CMB constraints limiting  $f \lesssim 10^{-2}$  to  $10^{-4}$  in this mass range. Section 1.4 also discussed how the rate of mergers observed by LIGO may be compatible with the expectations for PBHs making up all [136] or a portion [12, 137] of the DM abundance. In this chapter, we investigate the consequences for theory and experiment by assuming the PBH merger explanation.

If these black holes were produced primordially, they would have resulted from the collapse of large density perturbations. While these would be sourced as scalar perturbations, because of their size they can lead to non-trivial tensor perturbations at second order in cosmological perturbation theory [146–151]. A key observation is PBHs in the mass range

detected by LIGO generate tensor perturbations that may be detected as gravitational waves at pulsar timing arrays (PTAs) [152].

To make a clear distinction from the GWs detected at LIGO, we call the GWs probed at PTAs “secondary gravitational waves” (SGWs). SGWs that are correlated with the LIGO GW signal may or may not be detectable at PTAs depending on the details of the formation mechanism of the PBHs. For example, it has been observed that for scalar perturbations that are highly peaked—behaving essentially like Dirac- $\delta$  functions—PTA probes of SGWs already exclude the formation of PBHs with masses in the range  $10^{-2}M_{\odot} \lesssim M_{\text{PBH}} \lesssim 10M_{\odot}$  [295–297]. However, constraints have not been applied to explicit models for forming PBHs with non-idealized scalar power spectra except recently in [298, 299].

In this chapter, we discuss production mechanisms that may give rise to a somewhat narrow spectrum of  $\sim 10$  to  $30M_{\odot}$  PBHs that make up some or all of the dark matter and are consistent with the LIGO GW signal. We place constraints on these models using PTA sensitivities to SGWs. Models with strongly peaked primordial spectra produce SGWs that can be probed in the near future by PTAs. Meanwhile, models with extended primordial spectra are already excluded by present PTA data. However, some models that explain PBH formation consistent with LIGO will not be probed by present or future PTA experiments. The most important limiting factor for probing these models is PTA observing time. SGW detection by a PTA would not only bolster the case for the merging black holes detected by LIGO as being formed primordially, it would also provide insight into the physics of the very early universe.

The remainder of this chapter is organized as follows. Section 4.1 discusses the details and uncertainties in the calculation of the PBH binary merger rate. While this is important for estimating how much of DM abundance resides in PBHs, it will ultimately not impact PTA probes of PBHs to a large degree. In Section 4.2, we review the calculation of PBH and SGW spectra from a primordial scalar spectrum. Section 4.3 gives a demonstration of the SGW spectrum for an idealized  $\delta$  function primordial spectrum. Section 4.4 describes

explicit modes for PBH formation and places bounds on their resulting SGW spectra. In Section 4.5, we discuss assumptions and uncertainties in our calculations and their effects on the bounds. We conclude in Section 4.6.

## 4.1 PBH Binary Merger Rate

In calculating the merger rate, Ref. [136] estimate binary formation inside DM halos, while Refs. [12, 137] estimate binaries formed primordially, just after PBH pairs can gravitate strongly enough to decouple from Hubble expansion (see also earlier work by [300, 301]). As the predicted rate of the latter works is much higher than the former, the mergers of primordially produced binaries may be expected to dominate. However, there are uncertainties in this the calculation of the primordial binary formation that have not before been estimated. We quantify some of these uncertainties here.

First, let us review how the PBH binaries are formed. In isolation, a primordial PBH pair would merge essentially immediately after they decouple from the Hubble flow, with a time scale set by the gravitational free-fall time. However, other PBHs may perturb the binary system by creating tidal forces that lead the pair to form eccentric orbits. Once in this eccentric orbit, the time scale to merge via the emission of gravitational radiation is significantly prolonged. To estimate this new infall time, the semi-major and -minor axes for these binaries must be computed, a point on which Refs. [137, 301] disagree numerically (although Ref. [301] uses a more realistic Poisson probability distribution compared to the flat distribution in [12, 137]).

The binary formation and merger rate are calculated as follows. With the fraction of DM composed of PBHs  $f = \Omega_{\text{PBH}}/\Omega_{\text{DM}}$ , assuming a monochromatic mass distribution, the average separation at matter-radiation equality is

$$\bar{x} = \left( \frac{M_{\text{PBH}}}{\rho_{\text{PBH}}(z_{\text{eq}})} \right)^{1/3} = \frac{1}{(1 + z_{\text{eq}})f^{1/3}} \left( \frac{M_{\text{PBH}}}{3M_{\text{P}}^2 H_0^2 \Omega_{\text{PBH}}} \right)^{1/3}, \quad (4.1)$$

where  $M_P$  is the reduce Planck mass,  $H_0$  is the Hubble parameter today, and  $z$  is the redshift. A pair of PBHs separated by  $R$  decouple when  $M_{\text{PBH}}R^{-3} > \rho(z)$ . Thus, only PBHs with separation  $x$  at matter-radiation equality satisfying  $x < f^{1/3}\bar{x}$  will merge. The next-closest PBH to the decoupled PBH pair, which was a distance  $y$  at  $z_{\text{eq}}$ , will perturb the system and lead to an orbiting binary pair. Once formed, the binaries will have semi-major and semi-minor axes  $a$  and  $b$  and eccentricity  $e$  at formation time given by

$$a = \frac{\alpha x^4}{f \bar{x}^3}, \quad (4.2)$$

$$b = \beta \left(\frac{x}{y}\right)^3 a, \quad (4.3)$$

$$e = \sqrt{1 - \left(\frac{b}{a}\right)^2}, \quad (4.4)$$

where  $\alpha$  and  $\beta$  are order one numerical factors.

The probability distribution for  $(x, y)$  is

$$dP = \frac{9x^2y^2}{\bar{x}^6} e^{-(y/\bar{x})^3} dx dy, \quad (4.5)$$

normalized so  $\int_0^\infty dx \int_x^\infty dy dP = 1$ . Once binaries form, they will eventually merge with coalescence time

$$t = Qa^4(1 - e^2)^{7/2}, \quad Q \equiv \frac{5}{512} \left(\frac{8\pi M_P^2}{M_{\text{PBH}}}\right)^3. \quad (4.6)$$

To evaluate  $dP/dt$ , use  $(t/T) = (x/\bar{x})^{37}(y/\bar{x})^{-21}$  (derived from the above equations) to eliminate  $x$  in favor of  $y$  and  $t$  in (4.5), then integrate  $y$ . The bounds of integration on  $y$  corresponding to  $y = x \rightarrow \infty$  are  $y/\bar{x} = (T\alpha^4\beta^7/t)^{-1/16} \rightarrow (T\alpha^4\beta^7/t)^{1/21} f^{37/63}$ , where  $T \equiv Q\bar{x}^4/f^4$ .

Numerical studies [137, 301] have numerically estimated  $\alpha$  and  $\beta$ , and these estimates are expected to be most accurate for  $y \lesssim f^{1/3}\bar{x}$ . However,  $y$  ought to be integrated to significantly larger values where numerical simulations may not be as accurate. Thus, we

split the integral as:

$$\begin{aligned}
\frac{dP}{dt} &= \int_{(T\alpha^4\beta^7/t)^{-1/16}}^{\gamma f^{1/3}\bar{x}} dy dP|_{\alpha^4\beta^7=\delta^{11}} + \int_{\gamma f^{1/3}\bar{x}}^{(T\alpha^4\beta^7/t)^{1/21} f^{37/63}\bar{x}} dy dP|_{\alpha^4\beta^7=\epsilon^{11}} \\
&= \frac{3}{37} \left( \frac{t}{T\delta^{11}} \right)^{3/37} \frac{1}{t} \left( \Gamma \left[ \frac{58}{37}, \left( \frac{t}{T\delta^{11}} \right)^{3/16} \right] - \Gamma \left[ \frac{58}{37}, f\gamma^3 \right] \right) \\
&\quad - \frac{3}{37} \left( \frac{t}{T\epsilon^{11}} \right)^{3/37} \frac{1}{t} \left( \Gamma \left[ \frac{58}{37}, \left( \frac{t}{T\epsilon^{11}} \right)^{-1/7} f^{37/21} \right] - \Gamma \left[ \frac{58}{37}, f\gamma^3 \right] \right),
\end{aligned} \tag{4.7}$$

where we set  $\delta^{11} = (.68)^4(2.8)^7$  in accordance with [137] (the values  $\alpha = .4$  and  $\beta = .8$  in the earlier work [301] give similar results), while we vary  $10^{-1} < \gamma, \delta < 10$  to account for uncertainties.

Finally, the merger rate today ( $t = t_0$ ) can be calculated from

$$R = \frac{n_{\text{PBH}}}{2} \frac{dP}{dt}(t = t_0). \tag{4.8}$$

The results are show in Fig. 4.1, along with a comparison to the LIGO preferred merger rate. For comparison, we show in red the results from an earlier version of [12], which uses a flat (rather than Poisson) probability distribution in (4.5), only integrates up to  $y < \bar{x}$ , and sets  $\alpha = \beta = 1$ . The calculational uncertainties admit a wide band of merger rates. This translates into a range of PBH DM fractions  $f$  consistent with the LIGO rate from  $f \sim 10^{-3}$  to  $10^{-1}$ . The theoretical uncertainties that may be included in this estimate include those discussed in [301], such as the treatment of angular dependence, three-body collisions, other fluctuations beyond the three bodies considered, initial conditions, and radiation drag. Additionally, all these calculations assume an idealized monochromatic mass spectrum of PBHs. Finally, there is a further question of whether these binaries, which initially have major axes  $\gtrsim$  the size of our solar system, would survive the process of halo formation. These questions require further detailed numerical study beyond the scope of this work. An interesting observation is that this collection of binary merger assumptions appears inconsistent with

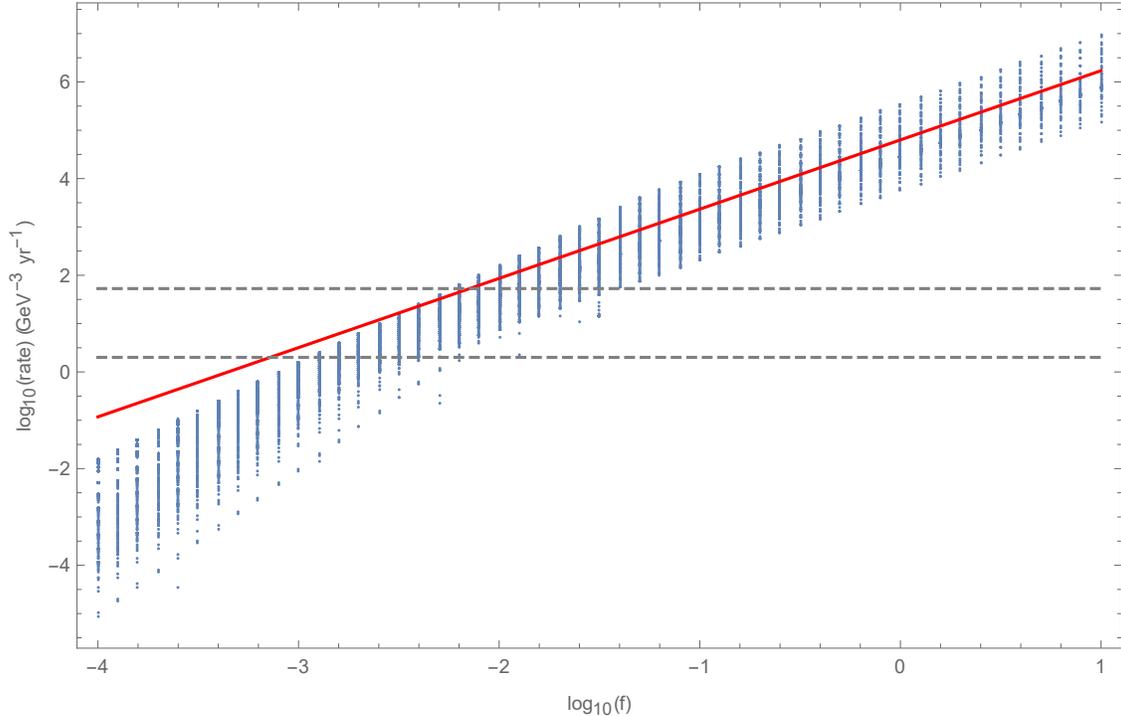


Figure 4.1: Uncertainty in the merger rate of binary PBHs today from varying  $10^{-1} < \gamma, \delta < 10$  as described in the text (blue band). The band outlined by gray dashed lines is the LIGO observed merger rate. For comparison, the red line shows a similar calculation in an older version of [12].

PBH DM fraction  $f = 1$ . For our purposes, we simply note the present theoretical and experimental uncertainties may allow a wide range of PBH abundance. For concreteness, we focus on the case where PBHs make up all of the DM, which might obtain if the above mechanism were ineffective (and in which case the mechanism of [136] might dominate), but we will see that reducing the PBH abundance by even several orders of magnitude will have only a modest impact on PTA probes of SGWs.

## 4.2 PBH and GW Spectra from a Primordial Scalar Spectrum

We consider a primordial curvature perturbation spectrum  $\mathcal{P}_{\mathcal{R}}(k)$  whose form is determined by early universe dynamics. Our approach is to ensure that PBHs are formed in the right

mass range and with the right abundance to explain the LIGO GW signal. The spectrum then has the potential to give rise to a background of stochastic SGWs, whose rate and strength depend on details of the inflationary theory. These details will be described in Section 4.4. Here we review the formalism for determining PBH formation and SGWs from a generic perturbation spectrum.

First, let us consider PBH formation. The curvature perturbations result in density perturbations, which during radiation domination are described by power spectra<sup>1</sup>  $\mathcal{P}_\delta(k) = \frac{4}{9} \left(\frac{k}{aH}\right)^4 \mathcal{P}_\Phi(k) = \left(\frac{4}{9}\right)^2 \left(\frac{k}{aH}\right)^4 \mathcal{P}_\mathcal{R}(k)$  for the matter perturbations  $\delta = \delta\rho/\rho$ , Bardeen potential  $\Phi$ , and curvature  $\mathcal{R}$ . Here,  $a$  is the scale factor and  $H = \frac{1}{a} \frac{da}{dt}$  is the Hubble parameter. When the perturbations are large enough, an overdense region can collapse into a PBH when the overdensity reenters the horizon at scale  $k_f$ , resulting in a PBH approximated to have a horizon mass

$$M_{\text{PBH}} = \frac{4\pi}{3} \rho_r H^{-3} \simeq 10M_\odot \left(\frac{g_*}{100}\right)^{-1/6} \left(\frac{\text{pc}^{-1}}{k_f}\right)^2, \quad (4.9)$$

where the universe is assumed to be radiation dominated with energy density  $\rho_r$ , and  $g_*$  is the number of relativistic degrees of freedom.

Assuming these perturbations are Gaussian, the energy fraction of PBHs with mass in the interval  $(M, M + dM)$  at their formation time  $t_f$  can be calculated using Press-Schechter formalism [303, 304],

$$\beta(M) = \frac{d}{d \log M} \frac{\rho_{\text{PBH}}(t_f)}{\rho_{\text{tot}}(t_f)} = 2 \int_{\delta_c}^{\infty} d\delta \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\delta^2}{2\sigma^2}} = \text{Erfc} \left( \frac{\delta_c}{\sqrt{2}\sigma_{R_M}} \right), \quad (4.10)$$

where [305]

$$\sigma_{R_M}^2 = \int \frac{dk}{k} W(kR_M)^2 \mathcal{P}_\delta(k) = \int \frac{dk}{k} W(kR_M)^2 \left(\frac{4}{9}\right) (kR_M)^4 \mathcal{P}_\Phi(k) \quad (4.11)$$

---

<sup>1</sup>More properly,  $\mathcal{P}_\delta(k) = \frac{16}{3} \left(\frac{k}{aH}\right)^2 j_1^2 \left(\frac{k}{\sqrt{3}aH}\right) \mathcal{P}_\mathcal{R}(k)$  [302], where  $j_1$  is a spherical Bessel function, though the difference mainly appears on sub-horizon scales that are suppressed in (4.11) by the window function  $W(kR)$ .

is the variance for the Gaussian probability distribution for primordial density perturbations on length scale  $R_M$ . Here,  $W(x) = e^{-x^2/2}$  is a Gaussian window function. We take the threshold for PBH formation  $\delta_c = 1/3$  [127] in (4.10) which will result in a conservative estimate of the SGW abundance, though higher values of *e.g.*  $\delta_c \simeq 0.45$  have also been indicated [128–130], and non-sphericity effects [130–133] can make  $\delta_c$  higher still. See the Discussion section for the effect of these different choices. Finally, the relic abundance of PBH today is given by

$$\frac{d}{d \log M} \Omega_{\text{PBH}} h^2 \simeq 2 \times 10^7 \beta(M) \left( \frac{g_{*,i}}{106.75} \right)^{-1/4} \left( \frac{M}{M_\odot} \right)^{-1/2}. \quad (4.12)$$

Next, let us consider SGW production. By definition here, SGWs are GWs that are produced at second order in perturbation theory and could be probed by PTAs. Precisely because the scalar power spectrum must be large to produce PBHs, these secondary tensor modes may be detectable.

The second-order tensor power spectrum can be calculated from the scalar power spectrum using [296, 297, 306–309],<sup>2</sup>

$$\mathcal{P}_h(k, \eta) = \int_0^\infty d\tilde{k} \int_{-1}^1 d\mu \mathcal{P}_\Phi(|\mathbf{k} - \tilde{\mathbf{k}}|) \mathcal{P}_\Phi(\tilde{k}) \mathcal{F}(k, \tilde{k}, \mu, \eta), \quad (4.13)$$

where

$$\begin{aligned} \mathcal{F}(k, \tilde{k}, \mu, \eta) = & \frac{(1 - \mu^2)^2}{a^2(\eta)} \frac{k^3 \tilde{k}^3}{|\mathbf{k} - \tilde{\mathbf{k}}|^3} \int_0^\eta d\eta_1 a(\eta_1) g_k(\eta, \eta_1) f(\mathbf{k}, \tilde{\mathbf{k}}, \eta_1) \\ & \int_0^\eta d\eta_2 a(\eta_2) g_k(\eta, \eta_2) \left[ f(\mathbf{k}, \tilde{\mathbf{k}}, \eta_2) + f(\mathbf{k}, \mathbf{k} - \tilde{\mathbf{k}}, \eta_2) \right], \end{aligned} \quad (4.14)$$

with

$$f(\mathbf{k}, \tilde{\mathbf{k}}, \eta) = 12\Phi(\tilde{k}\eta)\Phi(|\mathbf{k} - \tilde{\mathbf{k}}|\eta) + 8\eta\Phi(\tilde{k}\eta)\Phi'(|\mathbf{k} - \tilde{\mathbf{k}}|\eta) + 4\eta^2\Phi'(\tilde{k}\eta)\Phi'(|\mathbf{k} - \tilde{\mathbf{k}}|\eta). \quad (4.15)$$

---

<sup>2</sup>See [306, 309] for variable changes to make this computationally simpler, where we have resolved some inconsistencies and ambiguities.

Here,  $\mu = \mathbf{k} \cdot \tilde{\mathbf{k}} / (k\tilde{k})$  and  $\eta$  is the conformal time. The Bardeen potential  $\Phi$  during radiation domination (RD) (after dropping the decaying mode) is

$$\Phi(\mathbf{k}, \eta) = \frac{A(\mathbf{k})}{(\sqrt{wk\eta})^2} \left( \frac{\sin(\sqrt{wk\eta})}{\sqrt{wk\eta}} - \cos(\sqrt{wk\eta}) \right), \quad (4.16)$$

with equation of state  $w = 1/3$  during RD. Its power spectrum is defined by

$$\langle \Phi(\mathbf{k})\Phi(\mathbf{k}') \rangle = \frac{2\pi^2}{k^3} \delta^3(\mathbf{k} + \mathbf{k}') \mathcal{P}_\Phi(k). \quad (4.17)$$

The Green's function in (4.14) is

$$g_k(\eta, \eta') = \frac{\sin(k(\eta - \eta'))}{k}. \quad (4.18)$$

Finally, the relic abundance of gravitational waves can be calculated in terms of Eq. (4.13) as [308, 310, 311],<sup>3</sup>

$$\frac{d}{d \ln k} \Omega_{\text{SGW}}(k, \eta) = \frac{1}{12} \left( \frac{k}{a(\eta)H(\eta)} \right)^2 \mathcal{P}_h(k, \eta). \quad (4.19)$$

Since  $\Omega_{\text{SGW}}$  scales as radiation, it is convenient to evaluate this quantity at matter-radiation equality (denoted by subscript “eq”) and then scale to today (denoted by subscript “0”), giving approximately,

$$\frac{d}{d \ln k} \Omega_{\text{SGW}}(k, \eta_0) \simeq \frac{1}{12} \frac{1}{1 + z_{\text{eq}}} (k\eta_{\text{eq}})^2 \mathcal{P}_h(k, \eta_{\text{eq}}), \quad (4.20)$$

where  $z$  is the redshift.

Finally, let us review how experimental searches connect with gravitational wave abundance. Gravitational wave experiments typically quote results in terms of the characteristic

---

<sup>3</sup>Refs. [296, 306, 307, 309] give a range of differing values for this expression with which we do not agree.

strain  $h_c$ , which is related to an abundance of stochastic gravitational waves by [310, 311],

$$\frac{d}{d \ln k} \Omega_{\text{GW}}(k = 2\pi f, \eta_0) = \frac{2\pi}{3H_0^2} f^2 h_c^2(f). \quad (4.21)$$

The  $\Omega_{\text{SGW}}$  abundance computed in Eq. (4.20) can be directly translated to characteristic strain constraints using this formula.

### 4.3 Idealized Delta Function Spectrum

Perhaps the simplest model for PBH production is to assume a sharp narrow spike in  $k$  space in the scalar power spectrum. For a narrow enough spike, this can be approximated by a  $\delta$ -function [296]. While a  $\delta$  function is not physical, it is useful to consider as a mathematical construct.

The SGW spectrum for a  $\delta$ -function scalar spectrum is plotted in Fig. 4.2. For reference, the  $\delta$ -function is chosen to be peaked at a scale  $k_f$  corresponding to a horizon PBH mass of  $30M_\odot$  (see Eq. (4.9)), and its amplitude is chosen so that PBHs that form from this spectrum make up all of the observed DM abundance.

A constructive interference between  $g_k$  and  $f(\mathbf{k}, \tilde{\mathbf{k}}, \eta)$  in (4.14) leads to a resonance at  $k = 2\sqrt{w}k_f$  where the amplitude continues to grow at late times, and there is a zero at  $k = \sqrt{2w}k_f$ . The spectrum extends up to  $k = 2k_f$  where the incoming scalar modes are aligned.

Of course, physical spectra will be extended. Let us briefly discuss the changes to the PBH and SGW spectra as we go to more extended primordial scalar spectra. Regarding the PBH spectrum, note that PBH formation is exponentially sensitive to  $\sigma_{R_M}$  and the integration in (4.11) samples a somewhat narrow window in  $k$  of  $\mathcal{P}_\Phi(k)$ . Thus, PBHs predominately form near where  $\mathcal{P}_\Phi(k)$  peaks. Nevertheless, the integration in (4.11) may lead the PBH mass spectrum to peak at a smaller or larger scale than  $\mathcal{P}_\Phi(k)$  depending on the detailed shape of  $\mathcal{P}_\Phi(k)$ . Critical collapse effects [128–130, 312–316] will lead to further corrections to the peak

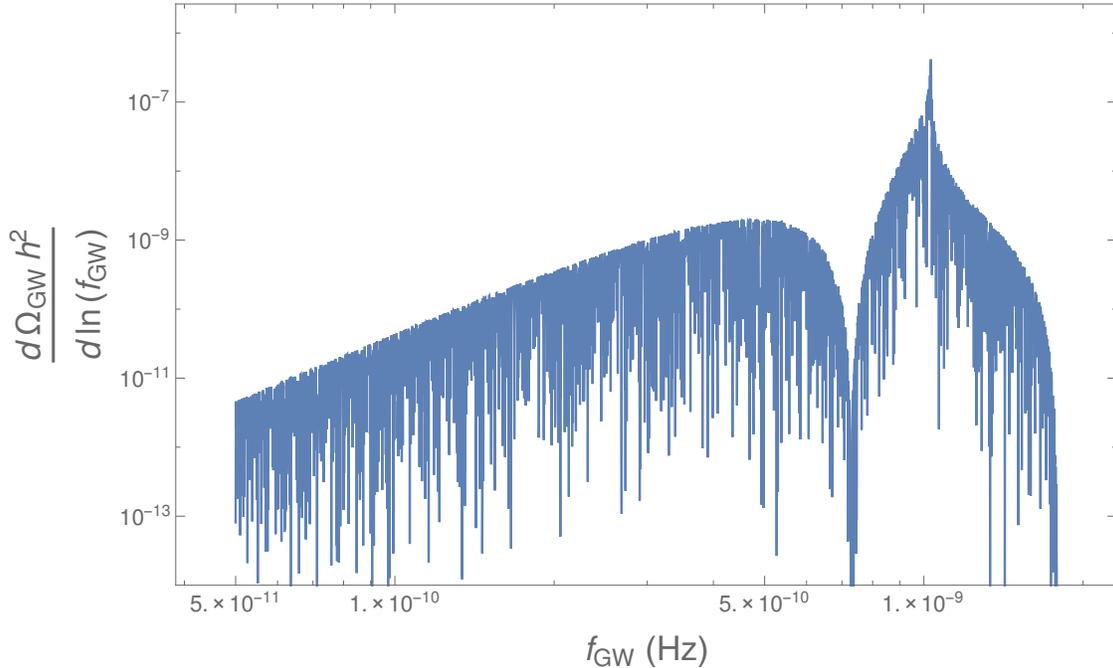


Figure 4.2: Gravitational wave abundance for an idealized delta function scalar spectrum peaked at the scale  $k_f$  corresponding to PBH mass  $30M_{\odot}$  according to Eq. (4.9) and normalized so that  $\Omega_{\text{PBH}} = \Omega_{\text{DM}}$

mass—see comments in the Discussion section. In regards to SGWs, the SGW abundance spectra will be smoothed out by the integral in (4.13) for extended  $\mathcal{P}_{\Phi}(k)$ , so features like the resonance and destructive interference in the  $\delta$ -function spectrum will not be present for extended spectra. Additionally, the SGW abundance is only quadratically sensitive to  $\mathcal{P}_{\Phi}(k)$  and depends on integration over a larger range of  $k$  in (4.13). This leads to an enhanced SGW abundance over a larger range in frequency for extended scalar spectra relative to a narrower spectrum.

## 4.4 Models of Primordial Scalar Spectra

There are many models of inflationary dynamics that can induce primordial power spectra giving rise to PBH formation. Here, we review several classes of models capable of producing PBHs with large enough abundance at masses relevant to LIGO. We do not attempt a full accounting of all models that have been proposed for PBH production. Rather, we

survey several models that predict different primordial spectra, allowing us to draw some general conclusions. Other models of PBH production not considered here can be found in Refs. [298, 317–321]. For each model, we calculate the resulting SGW spectrum. Results and present bounds from the European Pulsar Timing Array (EPTA) [13], Parkes Pulsar Timing Array (PPTA) [15], and NANOGrav [14] experiments and projections for SKA [16] are displayed in conjunction with these spectra in Fig. 4.3. The plots show envelopes of SGW spectra that correspond to a peak in the PBH mass spectrum at masses  $M_{\text{PBH}} = 30$  and  $10M_{\odot}$ . We show results for  $\Omega_{\text{PBH}} = \Omega_{\text{DM}}$ , but the SGW amplitude is only log-dependent on the PBH relic abundance. Thus, results are only slightly changed for other choices. For example, for a spectra peaking at  $M_{\text{PBH}} = 30M_{\odot}$ , reducing the PBH abundance from  $f = 1$  (all the DM) to  $f = 10^{-4}$  reduces the SGW relic abundance by a factor of 0.4, which would have a minor effect on SGW detectability, whereas the LIGO rate may be accommodated in either case depending on the details of the binary formation and merger rate calculation as discussed in the Introduction.

#### 4.4.1 Double Inflation with Parametric Resonance

A close approximation to a highly-peaked  $\delta$ -function scalar spectrum can be realized with a period of parametric resonance after the end of inflation. Oscillations of the inflaton can lead to specific modes being exponentially enhanced [257]. Since parametric resonance occurs after inflation has ended, the resonantly excited modes would not have a large enough length scale to produce PBHs at masses relevant to LIGO. A solution is to have two periods of inflation, with parametric resonance occurring between the two periods. The second inflation stretches the resonantly amplified modes to the relevant length scale. A model of this type is given in Refs. [322–324], which construct a period of hybrid inflation [325, 326] followed by new inflation [327, 328]. Such a model can be engineered to give a peak at any scale—by varying the length of the second inflation—and with any amplitude—depending on the relationship between the efficiency of the resonance and the decay width of the inflaton

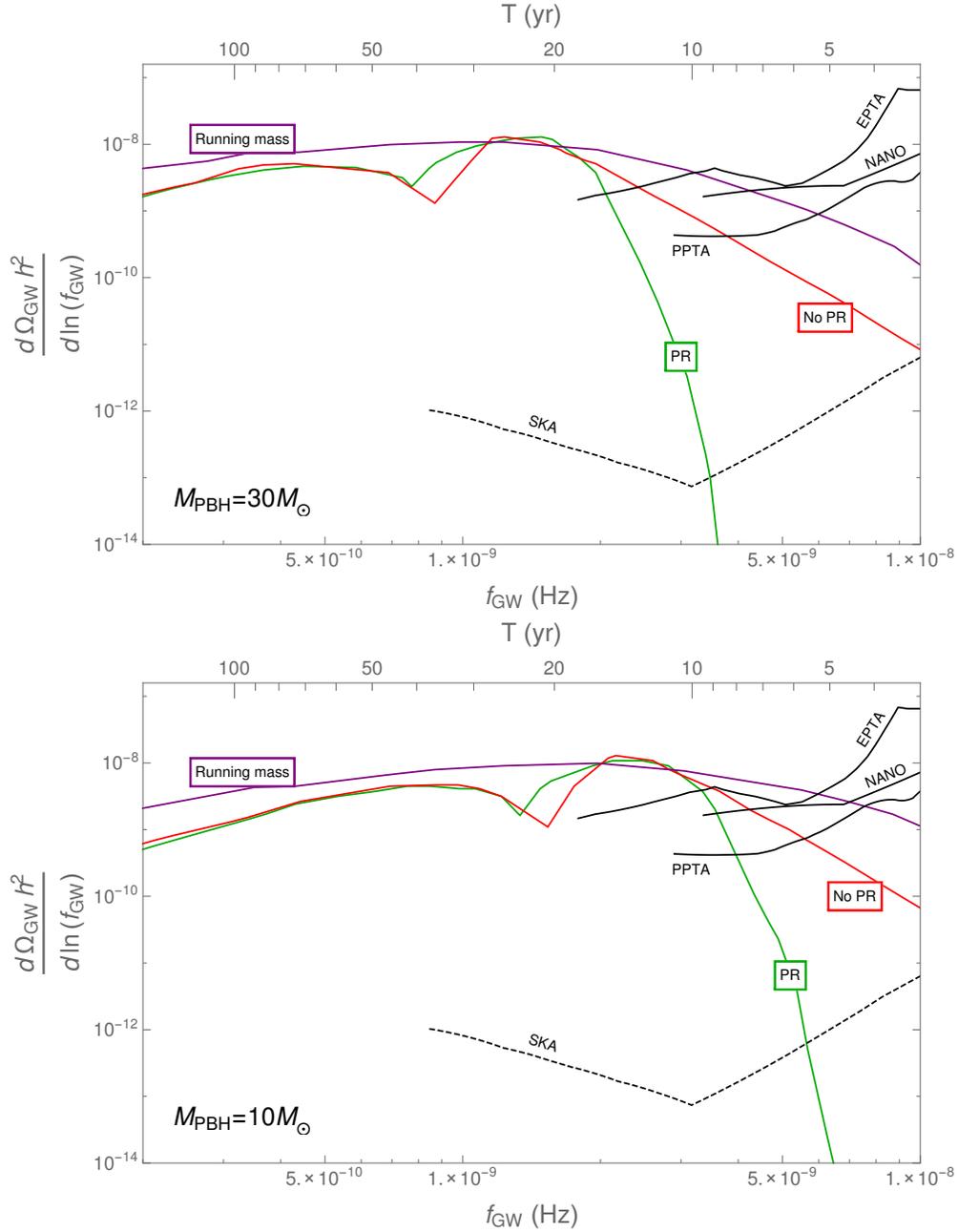


Figure 4.3: Gravitational wave abundance (envelopes) as a function of frequency assuming  $\Omega_{\text{PBH}} = \Omega_{\text{DM}}$ . The PBH abundance spectrum is peaked at 30 (top) or 10 (bottom)  $M_{\odot}$ . We display SGW for a top-hat spectrum with width set by expectations from parametric resonance (green “PR” curve), a red-tilted scalar spectrum with spectral index  $n_s = -1$  supplemented by a cutoff at a minimum frequency (red “No PR” curve), and the spectrum from the running mass model (purple curve). Black solid lines are current spectrum-independent bounds from EPTA (upper) [13], NANOGrav (middle) [14], and PPTA (lower) [15]. The black dashed line is a projection for bounds from SKA [16]. The top axis indicates the approximate observing time  $T$  to be sensitive to a given minimum frequency  $f_{\text{min}} \sim 1/T$ .

causing the resonance.

Because parametric resonance produces a sharply peaked scalar perturbation spectrum, it is the model that most closely mimics the  $\delta$ -function spectrum of the previous subsection. However, even in this case, the  $\delta$ -function approximation is not quite applicable; parametric resonance at a scale  $k$  results in a resonant band of width  $\sim k$  [258]. Such a spectrum can be approximated more closely by a top-hat [296], and the resulting SGW spectrum is shown in Fig. 4.3 by a green curve.<sup>4</sup> This more realistic spectrum does not have as pronounced a peak as the  $\delta$ -function case. Furthermore, for a fixed peak mass of PBH production, the scalar power spectrum peak must be shifted to larger  $k$  due to the integration in (4.11), and therefore the SGW frequencies in this more realistic spectrum are larger than those of the  $\delta$ -function spectrum.

At present, pulsar timing constraints have sensitivity to this model up to PBH mass spectra peaked at  $M_{\text{PBH}} \lesssim 30M_{\odot}$ . However, it should be noted that critical collapse effects will somewhat reduce the mass reach shown here; see details in the Discussion section. Thus, this model cannot be definitively excluded at present as an explanation for the black holes observed by LIGO. Note that although EPTA is not as sensitive to small GW abundance as the other experiments, its longer data collection time allows it to probe smaller frequencies which are critical here to detecting SGW signals. Thus, even absent any gain in sensitivity to  $h_c$ , all that would be needed to probe larger masses is increased data collection time  $T$ . The minimum frequency that can be probed is  $f_{\text{min}} \sim 1/T$ ; therefore, as observing time increases, the maximal mass probed will go as  $M \propto T^2$  (see Eq. (4.9)). We translate frequency to observing time along the top axis of Fig. 4.3.

*Summary:* Double inflation with parametric resonance has enough flexibility to allow a narrow range of PBHs needed to produce the LIGO GW signal. SGW spectra for masses  $M_{\text{PBH}} \lesssim 30M_{\odot}$  are constrained before accounting for critical collapse effects. An increase

---

<sup>4</sup>In the notation of [296], the resonant band has amplitude  $\mathcal{P}_{\Phi} \simeq A^2/(2\Delta)$  on the domain  $|\ln(k/k_p)| < \Delta$ . For parametric resonance peaked at  $k_p$  with width  $k_p$ , this implies  $\Delta \simeq \sinh^{-1}(1/2) \simeq 0.48$ . PBHs are predominantly produced at the smallest masses within this window corresponding to the scale  $k = k_p e^{\Delta}$ .

in PTA data collection time will enable detection of the SGWs for spectra peaked at masses greater than this, potentially detecting or excluding this mechanism as an explanation for LIGO GWs.

#### 4.4.2 Double Inflation without Parametric Resonance

Double inflation can produce a somewhat peaked spectrum even without parametric resonance. The earlier period of inflation gives rise to perturbations observed in the CMB and large scale structure. This early inflation is thus bounded to have small perturbations  $\delta\rho/\rho \simeq 10^{-5}$ . However, a later period of inflation can have much larger perturbations. This can be constructed with a model similar to that of the previous subsection; hybrid inflation followed by new inflation [329,330]. The detectability of this model using PTAs was recently discussed in Ref. [299], with which our conclusions are in agreement. To produce a large peak in the power spectrum, the model is designed so that the scale  $k_*$  at which the second inflation starts roughly coincides with the scale relevant to PBH formation. The power spectrum of the second inflation starts out very large but has a red spectral tilt ( $n_s < 1$ ). Thus, the spectrum is  $\mathcal{P}_{\mathcal{R}} = \mathcal{P}_{\mathcal{R}}(k_*)(k/k_*)^{n_s-1}$  for  $k > k_*$  and  $\mathcal{P}_{\mathcal{R}} \ll \mathcal{P}_{\mathcal{R}}(k_*)$  for  $k < k_*$ . In this model, the slow roll parameters satisfy  $|\eta| \gg |\epsilon|$ , so using the relationship  $n_s - 1 = 2\eta - 6\epsilon$  and the slow roll condition  $|\eta| < 1$  implies  $-1 \lesssim n_s \lesssim 3$  [327].

For this model, given a peak PBH mass  $M_{\text{peak}}$ ,  $k_*$  is smaller than the formation scale  $k_f(M_{\text{peak}})$  given by (4.9). This results from the integration in (4.11), which is maximized when  $R_M^{-1} > k_*$ . Thus, the SGW spectrum will be peaked at smaller frequencies in this model compared to the case of a  $\delta$ -function giving the same peak PBH mass. This effect is more pronounced for larger  $n_s$  corresponding to wider spectra, whereas for the minimum  $n_s = -1$ ,  $k_* \simeq k_f(M_{\text{peak}})$ . On the other hand, the SGW abundance away from the peak scale  $k_*$  is enhanced relative to narrower spectra.

Fig. 4.3 shows the resulting SGW spectra as red curves assuming  $n_s = -1$ , the minimum consistent with slow roll. This choice gives the narrowest possible peak in the perturbation

spectrum. Larger values of  $n_s$  would be more strongly constrained because they represent a more extended perturbation spectrum with larger resulting SGW abundance at high frequencies, which more than makes up for the necessary reduction in  $k_*$  to ensure the peak PBH mass  $M_{\text{peak}}$  remains constant as  $n_s$  increases. Already, the EPTA and PPTA experiments exclude this model of PBH formation up to masses peaked at greater than  $30M_\odot$  due to the contributions to SGWs at high frequency.<sup>5</sup> However, it is possible that critical collapse effects may still narrowly allow this model.

*Summary:* This double inflation model can produce PBHs in a narrow mass range with an abundance consistent with the rate of binary mergers observed by LIGO. However, because of its more extended spectrum, this model produces SGWs across a wider range of frequencies. Thus, present PTA experiments are more sensitive to this model than the previous and exclude this as an explanation for LIGO GWs up to the details of critical collapse effects.

### 4.4.3 Running Mass Model

The running mass model [331–335] supposes just one period of inflation, but with significant running of the spectral index. This approach can achieve a large perturbation amplitude at scales much smaller than those relevant to the CMB (here denoted  $k_0 = 0.05 \text{ Mpc}^{-1}$ ). The amplitude at any point can be parametrized by,

$$\mathcal{P}_{\mathcal{R}}(k) = \mathcal{P}_{\mathcal{R}}(k_0) \left( \frac{k}{k_0} \right)^{n(k)-1}, \quad (4.22)$$

where,

$$n(k) = n_s(k_0) + \frac{1}{2!}\alpha_s(k_0) \ln \left( \frac{k}{k_0} \right) + \frac{1}{3!}\beta_s(k_0) \ln^2 \left( \frac{k}{k_0} \right) + \frac{1}{4!}\gamma_s(k_0) \ln^3 \left( \frac{k}{k_0} \right) + \dots \quad (4.23)$$

---

<sup>5</sup>Data from multiple frequencies can be combined to yield stronger bounds on extended spectra. NANOGrav gives such a bound, but it is no stronger than the bin-by-bin exclusion for the sharply falling spectrum considered here.

Here,  $\alpha_s$ ,  $\beta_s$ , and  $\gamma_s$  are the runnings of  $n_s$ , giving sequentially higher derivatives of  $n_s$  with respect to  $\log k$ . When they are all allowed to vary, the parameters are bounded at  $k = k_0$  as  $\mathcal{P}_{\mathcal{R}} = (2.142 \pm .049) \times 10^{-9}$  [20],  $n_s = 0.9586 \pm 0.0056$ ,  $\alpha_s = 0.009 \pm 0.010$ , and  $\beta_s = 0.025 \pm 0.013$  [10]. To obtain a peaked spectrum, we will generally want one or both of  $\alpha_s$  and  $\beta_s$  to be positive, while  $\gamma_s$  (or a higher order term) must be negative. The coefficients are chosen so that the peaking occurs for scales relevant for PBH formation, well outside of the region probed by the CMB.

For an explicit model of this type, see the above references. We will simply set at  $k = k_0$  the values  $\mathcal{P}_{\mathcal{R}} = 2.142 \times 10^{-9}$ ,  $n_s = 0.96$ , and  $\alpha_s = 0.009$  near their preferred experimental values. We then vary  $\beta_s$  and  $\gamma_s$  to achieve a PBH fractional abundance vs. mass spectrum that peaks at a given mass (here, as before, either  $30$  or  $10M_{\odot}$ ) and gives the desired total relic abundance equal to the DM abundance.<sup>6</sup> Of note, this requires a tuning of several significant digits in the running parameters to obtain the correct peak position and amplitude in the primordial spectrum. Additionally, for the masses of interest,  $\beta_s$  is several standard deviations away from its measured value.

Another shortcoming is that generically  $|\beta_s|$  should be suppressed by a factor of  $(n_s - 1)$  in relation to  $|\alpha_s|$  if the third derivative of the inflaton potential is not much larger than the lower derivatives. This is the case, *e.g.*, for the explicit model in [334]. However, to produce a large enough peak in the curvature spectrum for the PBH masses of interest,  $|\beta_s| \gtrsim |\alpha_s|$  is required. Constructing an explicit model to circumvent this generic suppression poses a challenge.

As in the previous subsection, PBH production occurs predominantly on scales where  $\mathcal{P}_{\Phi}$  is largest, while SGWs are produced at all frequencies. The result is plotted in purple in Fig. 4.3. While in principle stronger bounds could be placed by combining PTA data over many frequencies, even the sensitivity of the bin-by-bin exclusions surpasses the SGW abundance for this model. Thus, present PTA experiments exclude this model for PBH

---

<sup>6</sup>Specifically,  $\beta_s = 0.0903091$  (0.0813319) and  $\gamma_s = -0.0166717$  (-0.0145293) for peak  $M_{\text{PBH}} = 30M_{\odot}$  ( $10M_{\odot}$ ).

production as an explanation for the LIGO events.

*Summary:* Like the double inflation without parametric resonance model of the previous subsection, the running mass model has an extended  $\mathcal{P}_\Phi$  spectrum. It can produce a somewhat narrow range of PBH masses in the LIGO window, though there are theoretical challenges with constructing such a model related to the size of the runnings. The necessary runnings may also be inconsistent with present bounds from the CMB. Even if these issues are ignored, this model leads to a large enough SGW abundance over a wide range of frequencies to exclude this model as an explanation for LIGO GWs.

#### 4.4.4 Axion-Curvaton Model

Unlike the previous models discussed, the axion-curvaton model [336, 337] (see also [338]) supposes that primordial fluctuations on small scales are sourced after inflation from a separate curvaton field [280, 339] that need not induce a second period of inflation. The model consists of a complex field  $\Phi = (\varphi/\sqrt{2})e^{i\sigma/f_\sigma}$ , where  $\sigma$  is the curvaton. Once  $\varphi$  reaches the minimum of its potential at  $\varphi = f_\sigma$  and begins oscillating about it,  $\sigma$  becomes well-defined. After this point, corresponding to comoving scale  $k_*$ , the curvaton can induce a blue scalar spectrum ( $n_s > 1$ ) so that  $\mathcal{P}_\mathcal{R} = \mathcal{P}_\mathcal{R}(k_*)(k/k_*)^{n_s-1}$  for  $k < k_*$  and  $\mathcal{P} \ll \mathcal{P}_\mathcal{R}(k_*)$  for  $k > k_*$ . Here, a blue spectrum with  $n_s \sim 2$  to 4 can be obtained. The power spectrum from the curvaton is constrained on large scales to be less than that observed to be coming from inflation— $\mathcal{P}_\mathcal{R}(k) \lesssim 2 \times 10^{-9}$  for  $k \lesssim \text{Mpc}^{-1}$ .

Because in this model the universe is assumed to always be radiation dominated, the curvaton, which redshifts as matter, decays before it dominates the universe’s energy density. So, the perturbations sourced by the curvaton will grow as  $(\rho_\sigma/\rho_r)^2 \propto a^2$  until the curvaton decays. Thus, PBHs—whose formation depends exponentially on the perturbation amplitude—will preferentially form at the time of the curvaton decay. Since the scale at which the curvaton decays is unrelated to the scale  $k_*$  at which  $\varphi$  reaches its minimum, PBHs are expected to form long after the scale at which their primordial overdensities reenter

the horizon.<sup>7</sup> But during the intervening time, the (radiation) energy within the comoving volume containing the primordial overdensity will redshift. The resulting PBH is thus potentially much smaller in mass than if it had formed immediately after horizon reentry. To produce a black hole of equal mass,  $k_*$  must in turn be far smaller than if the PBH collapsed immediately after horizon reentry. This drives the expected SGW signal to smaller frequencies than can be probed at pulsar timing arrays. Thus, barring some coincidence between the time it takes  $\varphi$  to reach its minimum and the decay time of the curvaton, PTA experiments will not be sensitive to this PBH production model. Indeed, these time scales are expected to be far apart, with the Hubble parameter at which  $\varphi$  oscillations start corresponding to  $H \simeq 2\pi f_\sigma$  and the Hubble parameter when  $\varphi$  decays corresponding to  $H \simeq \Gamma_\sigma \simeq m_\sigma^3/f_\sigma^2 \simeq \Lambda^6/f_\sigma^5 \ll f_\sigma$ , where  $\Lambda \ll f_\sigma$  is the explicit  $U(1)$  breaking scale where  $\sigma$  receives a non-perturbative mass.

*Summary:* This model can explain LIGO GWs through merging PBHs, but it has primordial fluctuations sourced after inflation. This generically gives rise to SGWs at frequencies too low for PTAs to discover.

## 4.5 Discussion

It is worth noting that we have neglected effects of critical collapse [128–130, 312–316], wherein detailed numerical work has shown that the mass of PBHs formed following horizon reentry may differ from the horizon mass depending on the size of the overdensity in  $\delta$ . Most importantly for our discussion, critical collapse effects shift the peak in the mass spectrum to slightly lower masses. This shift in the peak of the mass spectrum can be at most of order a few for the models considered here [316]. To compensate,  $\mathcal{P}_\Phi(k)$  must shift to smaller  $k$  (corresponding to larger mass, see (4.9)) once critical collapse effects are included in order to keep the PBH abundance spectrum peaked at the same mass. This change would require a slightly longer collection time, with the collection time depending on the mass as  $T \propto \sqrt{M}$ .

---

<sup>7</sup>For example, Ref. [337] considers benchmark points of  $M_{\min}/M_{\text{BH}} = 10^{-3}$  and  $10^{-8}$ .

Of the models we have discussed, the one whose detection prospects are most sensitive to this effect is double inflation with parametric resonance. In addition, critical collapse may also affect the PBH abundance by an order one factor, which will not significantly impact the SGW abundance.

We have also neglected effects of non-sphericity [130–133], which tend to raise the threshold  $\delta_c$  on the matter perturbation spectrum for PBHs to form. For an increase from  $\delta_c$  to  $\delta'_c$ , the amplitude of the scalar spectrum  $\mathcal{P}_\Phi$  must increase by a factor of  $\sim (\delta'_c/\delta_c)^2$ , and the corresponding GW abundance increases by a factor  $\sim (\delta'_c/\delta_c)^4$ . This may increase the SGW abundances in Fig. 4.3 by as much as a factor of  $\sim 9$ . This is not enough to change the qualitative picture of which models are probed by PTA experiments, though it may partially compensate for the effects of critical collapse on the needed observation time.

Another effect that can change the necessary threshold on  $\delta_c$  is a soft equation of state during the period relevant to PBH formation [340–343]. If new physics exists such that the equation of state is more matter-like ( $w < 1/3$ ), a smaller  $\delta_c$  is necessary to induce collapse. While for a fixed abundance of PBHs, this can reduce the amplitude of SGWs, it would not change the qualitative picture.

It is worth commenting that the merging of supermassive black holes (SMBH) will also create a stochastic background of gravitational waves. These obey a power-law spectrum with  $\Omega_{\text{GW}} \propto f^{2/3}$  [344–346]. This differs markedly from the spectra considered in this chapter, but disentangling the SGW considered here from the GW from the SMBH will present an additional challenge. This challenge may be acute depending on the amplitude of the SMBH GW.

We now revisit our assumption that PBHs constitute all of the dark matter. Recent work [142] has called into question whether an extended spectrum of PBHs making up all of the DM is allowed by MACHO and faint dwarf cooling constraints. This, perhaps along with bounds from WMAP [144] and Planck [145], may indicate that PBHs are allowed to be at most only a fraction of the DM. Nevertheless, even a small abundance of PBHs may

still be consistent with LIGO observations [12, 137]. But we reiterate that a reduction in the the PBH abundance by even several orders of magnitude will have only a small impact on the SGW spectra considered here because the PBH abundance is exponentially sensitive to the primordial power spectrum amplitude, whereas the SGW abundance is only power-law sensitive to it. As discussed in the introduction, the observed LIGO merger rate may be accommodated by PBH DM fractions much smaller than unity. Since our primary study target is explaining LIGO GWs, not the full abundance of DM, our results hold.

## 4.6 Conclusions

If LIGO GWs are to be explained by merging PBHs, they will need to be produced in the early universe via a peaked primordial curvature spectrum. A by-product of this will be the production of SGWs. We have shown that the detectability of SGWs at current and future experiments will depend sensitively on the physics that gives rise to PBHs. PTAs thus represent a powerful discriminator between PBH production mechanisms with very little dependence on the PBH relic abundance. Models that give rise to extended initial power spectra are already excluded, while highly peaked models like ones that use parametric resonance give SGW signals detectable by present experiments provided more data collecting time. Thus, a future detection of SGWs would give valuable insight into inflation and the formation of PBHs. Meanwhile, the absence of SGWs at present experiments or even SKA does not rule out every production mechanism for PBHs. So, if the rate of binary mergers at LIGO continues to agree with the rates calculated in [12, 136, 137] with increased data collection time, a non-detection of SGWs could indicate a model similar to that of the axion-curvaton as a source of PBHs. To support the hypothesis that the black holes are primordial and not astrophysical in nature, other probes may be necessary [347–349]. In any case, inflationary models that explain LIGO GWs through merging PBHs have important implications for SGWs, which will be probed effectively at PTAs.

# Chapter 5

## Summary and Outlook

The hunt continues to explain the nature of dark matter—one of the few solid experimental indications for BSM physics. This thesis has examined three of the most well-motivated and studied classes of DM models—weakly interacting massive particles, axions, and primordial black holes—in the context of their cosmological history in order to understand what models and areas of parameter space remain viable.

The WIMP miracle has generated plenty of theoretical interest in WIMP DM models, and this has been bolstered by theories to explain naturalness that often contain WIMP candidates, such as supersymmetry. In this thesis we have examined WIMPs that carry electroweak charge. We have shown how increasingly stringent limits are pushing such models into more constrained regions of parameter space. Models with Dirac fermions that predict a spin-independent direct detection cross section through the  $Z$  boson cannot make up all of the DM abundance. Nevertheless, we have shown how such models could still comprise a subdominant component of DM and be detectable in the near future.

Changing the particle nature of the DM from Dirac to Majorana can avoid the large spin independent direct detection cross sections associated with vectorial couplings to the  $Z$ . Indeed, we have shown that current experiments still allow a thermal relic with an axial vector coupling to the  $Z$  boson to make up all of the DM abundance—so long as the coupling to the

Higgs boson is not too large. Though, we have noted that the inclusion of additional states beyond the simplest effective field theory (as in the example of the singlet-doublet model) is important to ensure compatibility with precision electroweak measurements. These additional states may also present an important target for future collider searches. However, if future direct detection experiments like LZ do not observe a DM signal, even these models will become more constrained. In that instance, assuming a standard cosmological history, DM models with electroweak charges will need to contain multiple states with nearly degenerate masses to allow for coannihilations in the early universe that suppress the abundance of the DM without increasing the direct detection cross section. Though, it is not clear why these states should be degenerate to such a degree. Alternatively, a non-standard cosmological history such as a late-decaying scalar can dilute the DM abundance and still allow these relics. On the whole, as limits continue to improve, alternatives to the simplest WIMP models, or even to WIMPs themselves, will become increasingly attractive objects for study.

Axions are also well-motivated as a DM candidate, having first been proposed as the most robust explanation for the strong CP problem as well as being predicted by many string theories. As very light pseudo-Nambu-Goldstone bosons, their presence can have important effects in the early universe. In particular, here we have studied constraints on axion models arising from the non-observation of isocurvature perturbations in the CMB. These perturbations become larger as the scale of inflation increases. Future CMB experiments will probe for tensor modes, and if a detection is made, it will indicate that inflation occurred at a high scale with Hubble parameter  $H_I \gtrsim 10^{13}$  GeV. Then, under the simplest assumptions, if  $f_a > H_I/(2\pi)$ , PQ symmetry breaking occurs before inflation ends and the massless axion obtains isocurvature perturbations in excess of current limits. Thus, an important question will be whether axion models can be made compatible with models of high-scale inflation so that isocurvature bounds are not violated. This is especially important because direct detection experiments are being developed to search for axion DM with larger  $f_a$ . Here we have revisited two models wherein a combination of axion and inflationary dynamics suppress

these isocurvature perturbations.

The first model requires that the PQ breaking scale during inflation  $f_I$  be much larger than today, i.e.  $f_I > f_a$ . This has previously been suggested as a promising solution, but we have shown that in general this will not work. The problem occurs after inflation, when the radial field that sets the PQ breaking scale must change from  $f_I$  to  $f_a$ . Because the required hierarchy between these scales is so large, when the radial field begins oscillating about its minimum it induces parametric resonance effects that lead to the restoration of the PQ symmetry, which would in turn produce topological defects and lead to the overclosure of the universe. A very narrow workaround for this exists involving a late stage of reheating to dilute the isocurvature perturbations. However, this involves some tuning of initial conditions and is anyways close to being ruled out.

The second model arranges for some large explicit breaking of the PQ symmetry that leads to an axion mass. While the mass must be large during inflation ( $\gtrsim H_I$ ) to suppress isocurvature, it must nearly vanish today so that the axion still solves the strong CP problem. We have built specific models that demonstrate some previously unappreciated challenges in arranging for this. In so doing, we have shown it is impossible to arrange for the explicit breaking to completely vanish today, contrary to the conclusions of earlier works, so additional model building is required to enhance the explicit breaking during inflation. We have demonstrated how this can be done and that such a solution does allow isocurvature constraints to be sufficiently reduced for PQ breaking scales up to the Planck scale. However, a major caveat is that Lagrangian parameters must be tuned to give a small enough initial displacement of the axion field from its minimum so as not to overclose the universe, and this cannot be explained by an anthropic argument [93] about human existence being possible only in specific Hubble patches without too much DM. While this setup is certainly possible in principle and would represent a smaller degree of tuning than what is present in the strong CP problem, it does not seem like the most likely model for inflation and DM in our universe.

Finally, PBHs have received renewed interest with the LIGO detections of merging black hole binaries. An important question is whether the black holes detected by LIGO were indeed produced primordially or whether they resulted from an astrophysical process. If they were produced primordially, these black holes would have formed from overdensities resulting from some initial density power spectrum seeded during or soon after inflation. Because these overdensities must be large, there must also be a large peak in the scalar power spectrum, and this can give rise to gravitational waves at second order in cosmological perturbation theory (SGWs). Detecting these SGWs would help to disentangle whether the black holes detected by LIGO are astrophysical or primordial in origin. In addition, SGWs will offer hints about the physics of the very early universe, giving a window to physics at much higher energy scales than what is accessible at current experiments.

In this thesis, we have examined several models of inflation and a curvaton model that can produce PBHs to determine what spectra of SGWs may be expected. While past bounds have focused on a highly-peaked, near  $\delta$ -function spectrum, this is not physical. Most models have a much more extended spectrum, and even the most peaked spectra still have a non-negligible width that affects the GW spectrum. We have shown how this impacts the sensitivity of pulsar timing array experiments to models of PBH formation at masses detected by LIGO. Inflationary models with more extended primordial scalar spectra are excluded, and models with narrower spectra will soon be probed. Still, other models like the axion-curvaton cannot be probed by PTAs.

While there have been no definitive detections for models of DM (beyond its gravitational effects) or BSM, many exciting results are expected in the coming decade that will help to guide theoretical research and possibly even make new discoveries in DM, cosmology, and BSM physics. These include:

- The LHC has only begun to publish early results from data taken at 13 TeV collisions. Future increases in luminosity and energy (to 14 TeV) will probe models of DM and its interactions, as well as explanations for naturalness, such as supersymmetry.

Future higher-energy proton colliders or precision electron colliders will offer further experimental reach.

- The recent detection of GWs at LIGO promises an entirely new type of astronomy, and ground-based interferometers are being joined by pulsar timing arrays and space-based interferometers like LISA to hunt for GWs at other frequencies.
- Future CMB stage 3 and 4 experiments will, among other things, determine the sum of the neutrino masses and possibly their hierarchy, more tightly constrain the number of light species (via  $\Delta N_{\text{eff}}$ ), probe ever smaller tensor mode amplitudes looking for hints of inflation, and study large scale structure via lensing.
- Direct detection experiments looking for WIMPs and for axions are gaining in sensitivity, and either a detection or null results will have implications for models of DM.
- Observatories like the Dark Energy Survey are discovering an ever-increasing array of Milky Way dwarf satellites that can be used to place strong bounds on DM annihilations, of which there are already tantalizing hints from the galactic center and from other galaxies. It will also probe large scale structures and dark energy, giving insight into the history (and future) of the universe that can provide further constraints on cosmological models.
- Astrophysical neutrino experiments like IceCube and Super-Kamiokande will continue to look for neutrinos from the annihilation of DM in the Sun or Earth and will also continue to perform exciting astronomy such as investigating IceCube's PeV neutrino excess.
- Precision short- and long-baseline neutrino oscillation experiments such as T2K, NO $\nu$ A, DUNE, and MicroBooNE will measure neutrino properties with the hopes of measuring their masses, mixings, and CP-violating phase. Their findings could have implications for BSM models that seek to explain the presence of neutrino masses.

Should a detection of new physics be made at any one of these, it will lead to follow-up studies across many other of these experiments with the goal of determining how the new physics may be embedded within a more complete theoretical framework, similar to how the  $Z$ -mediated EFT model of Section 2.1 was embedded in the singlet doublet model of the following section. Thus, a determination of the nature of DM could lead to a broader understanding of BSM physics, or alternatively a detection of BSM physics relevant to, *e.g.*, the problem of naturalness could provide a hint for where to look for DM. On the other hand, if future observations continue to find no deviations from the Standard Model, theoretical research will be driven away from the current most popular paradigms, forcing physicists to consider fresh directions for model building and different probes for new physics.

# Bibliography

- [1] Planck, P. A. R. Ade *et al.*, *Astron. Astrophys.* **571**, A16 (2014), 1303.5076.
- [2] PandaX-II, C. Fu *et al.*, *Phys. Rev. Lett.* **118**, 071301 (2017), 1611.06553.
- [3] LZ, D. S. Akerib *et al.*, (2015), 1509.02910.
- [4] IceCube, M. G. Aartsen *et al.*, *JCAP* **1604**, 022 (2016), 1601.00653.
- [5] SLD Electroweak Group, DELPHI, ALEPH, SLD, SLD Heavy Flavour Group, OPAL, LEP Electroweak Working Group, L3, S. Schael *et al.*, *Phys. Rept.* **427**, 257 (2006), hep-ex/0509008.
- [6] XENON, E. Aprile *et al.*, *JCAP* **1604**, 027 (2016), 1512.07501.
- [7] LUX, D. S. Akerib *et al.*, *Phys. Rev. Lett.* **112**, 091303 (2014), 1310.8214.
- [8] XENON1T, E. Aprile, *Springer Proc. Phys.* **148**, 93 (2013), 1206.6288.
- [9] BICEP2, Keck Array, P. A. R. Ade *et al.*, *Phys. Rev. Lett.* **116**, 031302 (2016), 1510.09217.
- [10] Planck, P. A. R. Ade *et al.*, *Astron. Astrophys.* **594**, A20 (2016), 1502.02114.
- [11] D. Budker, P. W. Graham, M. Ledbetter, S. Rajendran, and A. Sushkov, *Phys. Rev.* **X4**, 021030 (2014), 1306.6089.
- [12] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, *Phys. Rev. Lett.* **117**, 061101 (2016), 1603.08338.
- [13] L. Lentati *et al.*, *Mon. Not. Roy. Astron. Soc.* **453**, 2576 (2015), 1504.03692.
- [14] NANOGrav, Z. Arzoumanian *et al.*, *Astrophys. J.* **821**, 13 (2016), 1508.03024.
- [15] R. M. Shannon *et al.*, *Science* **349**, 1522 (2015), 1509.07320.
- [16] G. Janssen *et al.*, *PoS AASKA14*, 037 (2015), 1501.00127.
- [17] F. Zwicky, *Helv. Phys. Acta* **6**, 110 (1933), [Gen. Rel. Grav.41,207(2009)].
- [18] V. C. Rubin and W. K. Ford, Jr., *Astrophys. J.* **159**, 379 (1970).

- [19] N. Jarosik *et al.*, *Astrophys. J. Suppl.* **192**, 14 (2011), 1001.4744.
- [20] Planck, P. A. R. Ade *et al.*, *Astron. Astrophys.* **594**, A13 (2016), 1502.01589.
- [21] J. R. Primack, Dark matter and structure formation, in *Midrasha Mathematicae in Jerusalem: Winter School in Dynamical Systems Jerusalem, Israel, January 12-17, 1997*, 1997, astro-ph/9707285.
- [22] D. Walsh, R. F. Carswell, and R. J. Weymann, *Nature* **279**, 381 (1979).
- [23] M. Markevitch *et al.*, *Astrophys. J.* **567**, L27 (2002), astro-ph/0110468.
- [24] M. Markevitch *et al.*, *Astrophys. J.* **606**, 819 (2004), astro-ph/0309303.
- [25] J. R. Bond, G. Efstathiou, and J. Silk, *Phys. Rev. Lett.* **45**, 1980 (1980).
- [26] E. W. Kolb and M. S. Turner, *Front. Phys.* **69**, 1 (1990).
- [27] S. Dodelson, *Modern Cosmology* (Academic Press, Amsterdam, 2003).
- [28] D. Baumann, Inflation, in *Physics of the large and the small, TASI 09, proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, USA, 1-26 June 2009*, pp. 523–686, 2011, 0907.5424.
- [29] F. Beutler *et al.*, *Mon. Not. Roy. Astron. Soc.* **416**, 3017 (2011), 1106.3366.
- [30] BOSS, L. Anderson *et al.*, *Mon. Not. Roy. Astron. Soc.* **441**, 24 (2014), 1312.4877.
- [31] E. A. Kazin *et al.*, *Mon. Not. Roy. Astron. Soc.* **441**, 3524 (2014), 1401.0358.
- [32] A. J. Ross *et al.*, *Mon. Not. Roy. Astron. Soc.* **449**, 835 (2015), 1409.3242.
- [33] A. D. Linde, *Phys. Lett.* **B108**, 389 (1982).
- [34] P. Creminelli, D. L. López Nacir, M. Simonović, G. Trevisan, and M. Zaldarriaga, *JCAP* **1511**, 031 (2015), 1502.01983.
- [35] D. H. Lyth, *Phys. Rev. Lett.* **78**, 1861 (1997), hep-ph/9606387.
- [36] G. 't Hooft *et al.*, *NATO Sci. Ser. B* **59**, pp.1 (1980).
- [37] M. W. Goodman and E. Witten, *Phys.Rev.* **D31**, 3059 (1985).
- [38] G. Jungman, M. Kamionkowski, and K. Griest, *Phys. Rept.* **267**, 195 (1996), hep-ph/9506380.
- [39] XENON100 Collaboration, E. Aprile *et al.*, *Phys.Rev.Lett.* **109**, 181301 (2012), 1207.5988.
- [40] LUX, D. S. Akerib *et al.*, *Phys. Rev. Lett.* **116**, 161301 (2016), 1512.03506.
- [41] LUX, D. S. Akerib *et al.*, *Phys. Rev. Lett.* **116**, 161302 (2016), 1602.03489.

- [42] LUX, D. S. Akerib *et al.*, Phys. Rev. Lett. **118**, 021303 (2017), 1608.07648.
- [43] W. H. Press and D. N. Spergel, Astrophys. J. **296**, 679 (1985).
- [44] J. Silk, K. A. Olive, and M. Srednicki, Phys. Rev. Lett. **55**, 257 (1985).
- [45] M. Srednicki, K. A. Olive, and J. Silk, Nucl. Phys. **B279**, 804 (1987).
- [46] Super-Kamiokande, K. Choi *et al.*, Phys. Rev. Lett. **114**, 141301 (2015), 1503.04858.
- [47] D. Hooper and B. L. Dingus, Phys. Rev. **D70**, 113007 (2004), astro-ph/0210617.
- [48] T. Daylan *et al.*, Phys. Dark Univ. **12**, 1 (2016), 1402.6703.
- [49] Fermi-LAT, M. Ackermann *et al.*, Phys. Rev. **D89**, 042001 (2014), 1310.0828.
- [50] Z. Li, X. Huang, Q. Yuan, and Y. Xu, JCAP **1612**, 028 (2016), 1312.7609.
- [51] M. Ackermann *et al.*, Astrophys. J. **836**, 208 (2017).
- [52] L. Bergstrom, J. Edsjo, and P. Ullio, Phys. Rev. Lett. **87**, 251301 (2001), astro-ph/0105048.
- [53] A. Tasitsiomi and A. V. Olinto, Phys. Rev. **D66**, 083006 (2002), astro-ph/0206040.
- [54] AMS, M. Aguilar *et al.*, Phys. Rev. Lett. **110**, 141102 (2013).
- [55] S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978).
- [56] F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978).
- [57] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).
- [58] C. G. Callan, Jr., R. F. Dashen, and D. J. Gross, Phys. Lett. **B63**, 334 (1976).
- [59] R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 172 (1976).
- [60] S. Weinberg, Phys. Rev. **D11**, 3583 (1975).
- [61] G. 't Hooft, Phys. Rev. Lett. **37**, 8 (1976).
- [62] G. 't Hooft, Phys. Rev. **D14**, 3432 (1976), [Erratum: Phys. Rev.D18,2199(1978)].
- [63] R. J. Crewther, P. Di Vecchia, G. Veneziano, and E. Witten, Phys. Lett. **88B**, 123 (1979), [Erratum: Phys. Lett.91B,487(1980)].
- [64] C. A. Baker *et al.*, Phys. Rev. Lett. **97**, 131801 (2006), hep-ex/0602020.
- [65] R. D. Peccei, Lect. Notes Phys. **741**, 3 (2008), hep-ph/0607268, [,3(2006)].
- [66] P. Sikivie, Lect. Notes Phys. **741**, 19 (2008), astro-ph/0610440, [,19(2006)].
- [67] J. E. Kim, Phys. Rev. Lett. **43**, 103 (1979).

- [68] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. **B166**, 493 (1980).
- [69] M. Dine, W. Fischler, and M. Srednicki, Phys. Lett. **B104**, 199 (1981).
- [70] A. R. Zhitnitsky, Sov. J. Nucl. Phys. **31**, 260 (1980), [Yad. Fiz.31,497(1980)].
- [71] L. F. Abbott and P. Sikivie, Phys. Lett. **B120**, 133 (1983).
- [72] J. Preskill, M. B. Wise, and F. Wilczek, Phys. Lett. **B120**, 127 (1983).
- [73] M. Dine and W. Fischler, Phys. Lett. **B120**, 137 (1983).
- [74] P. Fox, A. Pierce, and S. D. Thomas, (2004), hep-th/0409059.
- [75] M. S. Turner, Phys. Rev. **D33**, 889 (1986).
- [76] L. Visinelli and P. Gondolo, Phys. Rev. **D80**, 035024 (2009), 0903.4377.
- [77] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* Cambridge Monographs on Mathematical Physics (Cambridge Univ. Press, Cambridge, UK, 1984).
- [78] L. H. Ford and A. Vilenkin, Phys. Rev. **D25**, 2569 (1982).
- [79] A. A. Starobinsky, Phys. Lett. **B117**, 175 (1982).
- [80] D. H. Lyth and E. D. Stewart, Phys. Lett. **B283**, 189 (1992).
- [81] D. H. Lyth, Phys. Rev. **D45**, 3394 (1992).
- [82] D. H. Lyth, Phys. Lett. **B236**, 408 (1990).
- [83] A. D. Linde, Phys. Lett. **B259**, 38 (1991).
- [84] D. Harari and P. Sikivie, Phys. Lett. **B195**, 361 (1987).
- [85] S. Chang, C. Hagmann, and P. Sikivie, Nucl. Phys. Proc. Suppl. **72**, 99 (1999), hep-ph/9808302.
- [86] T. Hiramatsu, M. Kawasaki, K. Saikawa, and T. Sekiguchi, Phys. Rev. **D85**, 105020 (2012), 1202.5851, [Erratum: Phys. Rev.D86,089902(2012)].
- [87] R. L. Davis, Phys. Lett. **B180**, 225 (1986).
- [88] R. A. Battye and E. P. S. Shellard, Nucl. Phys. **B423**, 260 (1994), astro-ph/9311017.
- [89] R. A. Battye and E. P. S. Shellard, Phys. Rev. Lett. **73**, 2954 (1994), astro-ph/9403018, [Erratum: Phys. Rev. Lett.76,2203(1996)].
- [90] K. Choi and J. E. Kim, Phys. Lett. **B165**, 71 (1985).
- [91] T. Banks and M. Dine, Nucl. Phys. **B505**, 445 (1997), hep-th/9608197.

- [92] M. Kawasaki, T. Moroi, and T. Yanagida, Phys. Lett. **B383**, 313 (1996), hep-ph/9510461.
- [93] A. D. Linde, Phys. Lett. **B201**, 437 (1988).
- [94] F. Wilczek, (2004), hep-ph/0408167.
- [95] A. Ringwald, Phys. Dark Univ. **1**, 116 (2012), 1210.5081.
- [96] G. G. Raffelt, Lect. Notes Phys. **741**, 51 (2008), hep-ph/0611350, [,51(2006)].
- [97] A. Arvanitaki and S. Dubovsky, Phys. Rev. **D83**, 044026 (2011), 1004.3558.
- [98] H.E.S.S., F. Aharonian *et al.*, Astron. Astrophys. **475**, L9 (2007), 0709.4584.
- [99] MAGIC, E. Aliu *et al.*, Science **320**, 1752 (2008), 0807.2822.
- [100] W. Essey and A. Kusenko, Astrophys. J. **751**, L11 (2012), 1111.0815.
- [101] D. Horns and M. Meyer, JCAP **1202**, 033 (2012), 1201.4711.
- [102] D. Horns and M. Meyer, Pair-production opacity at high and very-high gamma-ray energies, in *Proceedings, 9th Patras Workshop on Axions, WIMPs and WISPs (AXION-WIMP 2013): Mainz, Germany, June 24-28, 2013*, 2013, 1309.3846.
- [103] A. De Angelis, M. Roncadelli, and O. Mansutti, Phys. Rev. **D76**, 121301 (2007), 0707.4312.
- [104] M. Simet, D. Hooper, and P. D. Serpico, Phys. Rev. **D77**, 063001 (2008), 0712.2825.
- [105] M. A. Sanchez-Conde, D. Paneque, E. Bloom, F. Prada, and A. Dominguez, Phys. Rev. **D79**, 123511 (2009), 0905.3270.
- [106] D. Horns *et al.*, Phys. Rev. **D86**, 075024 (2012), 1207.0776.
- [107] O. Mena and S. Razzaque, JCAP **1311**, 023 (2013), 1306.5865.
- [108] M. Meyer, D. Horns, and M. Raue, Phys. Rev. **D87**, 035027 (2013), 1302.1208.
- [109] J. P. Conlon and M. C. D. Marsh, Phys. Rev. Lett. **111**, 151301 (2013), 1305.3603.
- [110] S. Angus, J. P. Conlon, M. C. D. Marsh, A. J. Powell, and L. T. Witkowski, JCAP **1409**, 026 (2014), 1312.3947.
- [111] J. Isern, E. Garcia-Berro, S. Torres, and S. Catalan, Astrophys. J. **682**, L109 (2008), 0806.2807.
- [112] J. Isern, E. Garcia-Berro, L. G. Althaus, and A. H. Corsico, Astron. Astrophys. **512**, A86 (2010), 1001.5248.
- [113] A. H. Corsico *et al.*, JCAP **1212**, 010 (2012), 1211.3389.

- [114] N. Viaux *et al.*, Phys. Rev. Lett. **111**, 231301 (2013), 1311.1669.
- [115] ADMX, S. J. Asztalos *et al.*, Phys. Rev. **D64**, 092003 (2001).
- [116] ADMX, L. D. Duffy *et al.*, Phys. Rev. **D74**, 012006 (2006), astro-ph/0603108.
- [117] V. Popov, J. Exp. Theor. Phys. **122**, 236 (2016), 1410.6682, [Zh. Eksp. Teor. Fiz.149,no.2,272(2016)].
- [118] P. Sikivie, Phys. Rev. Lett. **113**, 201301 (2014), 1409.2806.
- [119] P. Sikivie, N. Sullivan, and D. B. Tanner, Phys. Rev. Lett. **112**, 131301 (2014), 1310.8545.
- [120] Y. Kahn, B. R. Safdi, and J. Thaler, Phys. Rev. Lett. **117**, 141801 (2016), 1602.01086.
- [121] E. Armengaud *et al.*, JINST **9**, T05002 (2014), 1401.3233.
- [122] R. Bähre *et al.*, JINST **8**, T09001 (2013), 1302.5647.
- [123] A. Kogut *et al.*, JCAP **1107**, 025 (2011), 1105.2044.
- [124] PRISM, P. André *et al.*, JCAP **1402**, 006 (2014), 1310.1554.
- [125] B. J. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, Phys. Rev. **D81**, 104019 (2010), 0912.5297.
- [126] B. Carr, F. Kuhnel, and M. Sandstad, Phys. Rev. **D94**, 083504 (2016), 1607.06077.
- [127] B. J. Carr, Astrophys. J. **201**, 1 (1975).
- [128] I. Musco, J. C. Miller, and L. Rezzolla, Class. Quant. Grav. **22**, 1405 (2005), gr-qc/0412063.
- [129] I. Musco, J. C. Miller, and A. G. Polnarev, Class. Quant. Grav. **26**, 235001 (2009), 0811.1452.
- [130] I. Musco and J. C. Miller, Class. Quant. Grav. **30**, 145009 (2013), 1201.2379.
- [131] R. K. Sheth, H. J. Mo, and G. Tormen, Mon. Not. Roy. Astron. Soc. **323**, 1 (2001), astro-ph/9907024.
- [132] F. Kuhnel and M. Sandstad, Phys. Rev. **D94**, 063514 (2016), 1602.04815.
- [133] J. R. Bond and S. T. Myers, Astrophys. J. Suppl. **103**, 1 (1996).
- [134] Virgo, LIGO Scientific, B. P. Abbott *et al.*, Phys. Rev. Lett. **116**, 061102 (2016), 1602.03837.
- [135] Virgo, LIGO Scientific, B. P. Abbott *et al.*, Phys. Rev. Lett. **116**, 241103 (2016), 1606.04855.

- [136] S. Bird *et al.*, Phys. Rev. Lett. **116**, 201301 (2016), 1603.00464.
- [137] Yu. N. Eroshenko, (2016), 1604.04932.
- [138] EROS-2, P. Tisserand *et al.*, Astron. Astrophys. **469**, 387 (2007), astro-ph/0607207.
- [139] L. Wyrzykowski *et al.*, Mon. Not. Roy. Astron. Soc. **416**, 2949 (2011), 1106.2925.
- [140] Macho, R. A. Allsman *et al.*, Astrophys. J. **550**, L169 (2001), astro-ph/0011506.
- [141] T. D. Brandt, Astrophys. J. **824**, L31 (2016), 1605.03665.
- [142] A. M. Green, Phys. Rev. **D94**, 063530 (2016), 1609.01143.
- [143] K. Schutz and A. Liu, Phys. Rev. **D95**, 023002 (2017), 1610.04234.
- [144] M. Ricotti, J. P. Ostriker, and K. J. Mack, Astrophys. J. **680**, 829 (2008), 0709.0524.
- [145] L. Chen, Q.-G. Huang, and K. Wang, JCAP **1612**, 044 (2016), 1608.02174.
- [146] K. Tomita, Prog. Theor. Phys. **37**, 831 (1967).
- [147] S. Matarrese, O. Pantano, and D. Saez, Phys. Rev. **D47**, 1311 (1993).
- [148] S. Matarrese, O. Pantano, and D. Saez, Phys. Rev. Lett. **72**, 320 (1994), astro-ph/9310036.
- [149] S. Matarrese, S. Mollerach, and M. Bruni, Phys. Rev. **D58**, 043504 (1998), astro-ph/9707278.
- [150] H. Noh and J.-c. Hwang, Phys. Rev. **D69**, 104011 (2004).
- [151] C. Carbone and S. Matarrese, Phys. Rev. **D71**, 043508 (2005), astro-ph/0407611.
- [152] S. L. Detweiler, Astrophys. J. **234**, 1100 (1979).
- [153] J. Halverson, N. Orlofsky, and A. Pierce, Phys. Rev. **D90**, 015002 (2014), 1403.1592.
- [154] J. Kearney, N. Orlofsky, and A. Pierce, Phys. Rev. **D95**, 035020 (2017), 1611.05048.
- [155] J. Kearney, N. Orlofsky, and A. Pierce, Phys. Rev. **D93**, 095026 (2016), 1601.03049.
- [156] N. Orlofsky, A. Pierce, and J. D. Wells, (2016), 1612.05279.
- [157] PandaX-II, A. Tan *et al.*, Phys. Rev. Lett. **117**, 121303 (2016), 1607.07400.
- [158] R. Essig, Phys. Rev. **D78**, 015004 (2008), 0710.1668.
- [159] A. Basirnia, S. Macaluso, and D. Shih, (2016), 1605.08442.
- [160] A. Beniwal *et al.*, Phys. Rev. **D93**, 115016 (2016), 1512.06458.
- [161] N. Arkani-Hamed, S. Dimopoulos, and S. Kachru, (2005), hep-th/0501082.

- [162] G. Duda, G. Gelmini, and P. Gondolo, *Phys.Lett.* **B529**, 187 (2002), hep-ph/0102200.
- [163] G. Giudice and A. Masiero, *Phys.Lett.* **B206**, 480 (1988).
- [164] H. P. Nilles, M. Srednicki, and D. Wyler, *Phys.Lett.* **B120**, 346 (1983).
- [165] J. Frere, D. Jones, and S. Raby, *Nucl.Phys.* **B222**, 11 (1983).
- [166] J. Derendinger and C. A. Savoy, *Nucl.Phys.* **B237**, 307 (1984).
- [167] R. Blumenhagen, M. Cvetič, and T. Weigand, *Nucl. Phys.* **B771**, 113 (2007), hep-th/0609191.
- [168] L. Ibanez and A. Uranga, *JHEP* **0703**, 052 (2007), hep-th/0609213.
- [169] B. Florea, S. Kachru, J. McGreevy, and N. Saulina, *JHEP* **0705**, 024 (2007), hep-th/0610003.
- [170] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, *Comput. Phys. Commun.* **192**, 322 (2015), 1407.6129.
- [171] A. L. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers, and Y. Xu, *JCAP* **1302**, 004 (2013), 1203.3542.
- [172] N. Anand, A. L. Fitzpatrick, and W. C. Haxton, *Phys. Rev.* **C89**, 065501 (2014), 1308.6288.
- [173] G. Arcadi, Y. Mambrini, and F. Richard, *JCAP* **1503**, 018 (2015), 1411.2985.
- [174] M. Escudero, A. Berlin, D. Hooper, and M.-X. Lin, *JCAP* **1612**, 029 (2016), 1609.09079.
- [175] A. De Simone, G. F. Giudice, and A. Strumia, *JHEP* **06**, 081 (2014), 1402.6287.
- [176] G. Belanger, J. Da Silva, T. Perrillat-Bottonet, and A. Pukhov, *JCAP* **1512**, 036 (2015), 1507.07987.
- [177] Fermi-LAT, MAGIC, M. L. Ahnen *et al.*, *JCAP* **1602**, 039 (2016), 1601.06590.
- [178] Particle Data Group, K. A. Olive *et al.*, *Chin. Phys.* **C38**, 090001 (2014).
- [179] K. Hamaguchi and K. Ishikawa, *Phys. Rev.* **D93**, 055009 (2016), 1510.05378.
- [180] R. Mahbubani and L. Senatore, *Phys. Rev.* **D73**, 043510 (2006), hep-ph/0510064.
- [181] F. D’Eramo, *Phys. Rev.* **D76**, 083522 (2007), 0705.4493.
- [182] R. Enberg, P. J. Fox, L. J. Hall, A. Y. Papaioannou, and M. Papucci, *JHEP* **11**, 014 (2007), 0706.0918.
- [183] T. Cohen, J. Kearney, A. Pierce, and D. Tucker-Smith, *Phys. Rev.* **D85**, 075003 (2012), 1109.2604.

- [184] A. Pierce, Phys. Rev. **D70**, 075006 (2004), hep-ph/0406144.
- [185] G. F. Giudice and A. Romanino, Nucl. Phys. **B699**, 65 (2004), hep-ph/0406088, [Erratum: Nucl. Phys.B706,487(2005)].
- [186] N. Arkani-Hamed, A. Delgado, and G. F. Giudice, Nucl. Phys. **B741**, 108 (2006), hep-ph/0601041.
- [187] C. Cheung and D. Sanford, JCAP **1402**, 011 (2014), 1311.5896.
- [188] S. Banerjee, S. Matsumoto, K. Mukaida, and Y.-L. S. Tsai, JHEP **11**, 070 (2016), 1603.07387.
- [189] L. Calibbi, A. Mariotti, and P. Tziveloglou, JHEP **10**, 116 (2015), 1505.03867.
- [190] A. Freitas, S. Westhoff, and J. Zupan, JHEP **09**, 015 (2015), 1506.04149.
- [191] C. Cheung, L. J. Hall, D. Pinner, and J. T. Ruderman, JHEP **05**, 100 (2013), 1211.4873.
- [192] ATLAS, G. Aad *et al.*, JHEP **05**, 071 (2014), 1403.5294.
- [193] ATLAS, G. Aad *et al.*, Eur. Phys. J. **C75**, 208 (2015), 1501.07110.
- [194] CMS, C. Collaboration, Report No. CMS-PAS-SUS-16-024, 2016 (unpublished).
- [195] CMS, C. Collaboration, Report No. CMS-PAS-SUS-16-025, 2016 (unpublished).
- [196] CERN Report No. ATL-PHYS-PUB-2015-032, 2015 (unpublished).
- [197] J. Halverson, Phys.Rev.Lett. **111**, 261601 (2013), 1310.1091.
- [198] M. Cvetič, J. Halverson, and P. Langacker, JHEP **11**, 058 (2011), 1108.5187.
- [199] M. Cvetič, J. Halverson, and H. Piragua, JHEP **1302**, 005 (2013), 1210.5245.
- [200] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, Comput. Phys. Commun. **185**, 960 (2014), 1305.0237.
- [201] N. Arkani-Hamed, A. Delgado, and G. Giudice, Nucl.Phys. **B741**, 108 (2006), hep-ph/0601041.
- [202] M. Dine and A. Kusenko, Rev.Mod.Phys. **76**, 1 (2003), hep-ph/0303065.
- [203] Particle Data Group, J. Beringer *et al.*, Phys.Rev. **D86**, 010001 (2012).
- [204] P. Cushman *et al.*, Working Group Report: WIMP Dark Matter Direct Detection, in *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2013, 1310.8327.
- [205] A. M. Green, PoS **IDM2008**, 108 (2008), 0809.1904.

- [206] H. Baer, A. Lessa, and W. Sreethawong, JCAP **1201**, 036 (2012), 1110.2491.
- [207] G. B. Gelmini and P. Gondolo, Phys.Rev. **D74**, 023510 (2006), hep-ph/0602230.
- [208] G. Gelmini, P. Gondolo, A. Soldatenko, and C. E. Yaguna, Phys.Rev. **D74**, 083514 (2006), hep-ph/0605016.
- [209] J. McDonald, Phys.Rev. **D43**, 1063 (1991).
- [210] ALEPH Collaboration, A. Heister *et al.*, Phys.Lett. **B533**, 223 (2002), hep-ex/0203020.
- [211] OPAL Collaboration, G. Abbiendi *et al.*, Eur.Phys.J. **C29**, 479 (2003), hep-ex/0210043.
- [212] DELPHI Collaboration, J. Abdallah *et al.*, Eur.Phys.J. **C34**, 145 (2004), hep-ex/0403047.
- [213] L3 Collaboration, M. Acciarri *et al.*, Phys.Lett. **B482**, 31 (2000), hep-ex/0002043.
- [214] A. G. Delannoy *et al.*, (2013), 1308.0355.
- [215] S. D. Thomas and J. D. Wells, Phys.Rev.Lett. **81**, 34 (1998), hep-ph/9804359.
- [216] ATLAS Collaboration, ATLAS Collaboration, ATLAS-CONF-2012-147, ATLAS-COM-CONF-2012-190 (2012).
- [217] CMS Collaboration, CMS-PAS-EXO-12-048 (2013).
- [218] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, JHEP **1106**, 128 (2011), 1106.0522.
- [219] T. Sjostrand, S. Mrenna, and P. Z. Skands, JHEP **0605**, 026 (2006), hep-ph/0603175.
- [220] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, JHEP **0701**, 013 (2007), hep-ph/0611129.
- [221] S. Mrenna and P. Richardson, JHEP **0405**, 040 (2004), hep-ph/0312274.
- [222] J. Conway *et al.*, PGS – Pretty Good Simulation, 2009, <http://physics.ucdavis.edu/~conway/research/software/pgs/pgs4-general.htm>.
- [223] M. Drees, M. Hanussek, and J. S. Kim, Phys.Rev. **D86**, 035024 (2012), 1201.5714.
- [224] H. Baer, A. Mustafayev, and X. Tata, Phys. Rev. **D89**, 055007 (2014), 1401.1162.
- [225] C. Han *et al.*, JHEP **02**, 049 (2014), 1310.4274.
- [226] P. Schwaller and J. Zurita, JHEP **03**, 060 (2014), 1312.7350.
- [227] N. Zhou, D. Berge, L. Wang, D. Whiteson, and T. Tait, (2013), 1307.5327v2 (forthcoming).

- [228] ATLAS Collaboration, G. Aad *et al.*, Phys.Rev. **D88**, 112006 (2013), 1310.3675.
- [229] J. Kearney and A. Pierce, Phys. Rev. **D88**, 095009 (2013), 1309.4447, [Erratum: Phys. Rev.D88,no.11,119902(2013)].
- [230] L. F. Abbott and M. B. Wise, Nucl. Phys. **B244**, 541 (1984).
- [231] M. Kawasaki, K. Saikawa, and T. Sekiguchi, Phys. Rev. **D91**, 065014 (2015), 1412.0789.
- [232] A. Ringwald and K. Saikawa, Phys. Rev. **D93**, 085031 (2016), 1512.06436, [Addendum: Phys. Rev.D94,no.4,049908(2016)].
- [233] L. Visinelli and P. Gondolo, Phys.Rev.Lett. **113**, 011802 (2014), 1403.4594.
- [234] R. Kitano and N. Yamada, JHEP **10**, 136 (2015), 1506.00370.
- [235] G. Grilli di Cortona, E. Hardy, J. Pardo Vega, and G. Villadoro, JHEP **01**, 034 (2016), 1511.02867.
- [236] A. D. Linde and D. H. Lyth, Phys. Lett. **B246**, 353 (1990).
- [237] T. Higaki, K. S. Jeong, and F. Takahashi, Phys. Lett. **B734**, 21 (2014), 1403.4186.
- [238] E. J. Chun, Phys.Lett. **B735**, 164 (2014), 1404.4284.
- [239] M. Fairbairn, R. Hogan, and D. J. Marsh, Phys.Rev. **D91**, 023509 (2015), 1410.1752.
- [240] M. Dine and A. Anisimov, JCAP **0507**, 009 (2005), hep-ph/0405256.
- [241] G. R. Dvali, (1995), hep-ph/9505253.
- [242] S. Folkerts, C. Germani, and J. Redondo, Phys. Lett. **B728**, 532 (2014), 1304.7270.
- [243] K. S. Jeong and F. Takahashi, Phys. Lett. **B727**, 448 (2013), 1304.8131.
- [244] K. Choi, K. S. Jeong, and M.-S. Seo, JHEP **07**, 092 (2014), 1404.3880.
- [245] M. Kawasaki, N. Kitajima, and F. Takahashi, Phys. Lett. **B737**, 178 (2014), 1406.0660.
- [246] K. Choi, E. J. Chun, S. H. Im, and K. S. Jeong, Phys. Lett. **B750**, 26 (2015), 1505.00306.
- [247] K. Nakayama and M. Takimoto, Phys. Lett. **B748**, 108 (2015), 1505.02119.
- [248] F. Takahashi and M. Yamada, JCAP **1510**, 010 (2015), 1507.06387.
- [249] M. Kawasaki, F. Takahashi, and M. Yamada, Phys. Lett. **B753**, 677 (2016), 1511.05030.
- [250] Y. Nomura, S. Rajendran, and F. Sanches, Phys. Rev. Lett. **116**, 141803 (2016), 1511.06347.

- [251] T. Banks, M. Dine, P. J. Fox, and E. Gorbatov, JCAP **0306**, 001 (2003), hep-th/0303252.
- [252] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, JHEP **06**, 060 (2007), hep-th/0601001.
- [253] J. Brown, W. Cottrell, G. Shiu, and P. Soler, JHEP **10**, 023 (2015), 1503.04783.
- [254] J. Brown, W. Cottrell, G. Shiu, and P. Soler, JHEP **04**, 017 (2016), 1504.00659.
- [255] B. Heidenreich, M. Reece, and T. Rudelius, JHEP **12**, 108 (2015), 1506.03447.
- [256] A. de la Fuente, P. Saraswat, and R. Sundrum, Phys. Rev. Lett. **114**, 151303 (2015), 1412.3457.
- [257] L. Kofman, A. D. Linde, and A. A. Starobinsky, Phys. Rev. Lett. **73**, 3195 (1994), hep-th/9405187.
- [258] L. Kofman, A. D. Linde, and A. A. Starobinsky, Phys. Rev. **D56**, 3258 (1997), hep-ph/9704452.
- [259] I. Tkachev, S. Khlebnikov, L. Kofman, and A. D. Linde, Phys. Lett. **B440**, 262 (1998), hep-ph/9805209.
- [260] M. Kawasaki, T. T. Yanagida, and K. Yoshino, JCAP **1311**, 030 (2013), 1305.5338.
- [261] S. Kasuya and M. Kawasaki, Phys. Rev. **D61**, 083510 (2000), hep-ph/9903324.
- [262] A. Mazumdar and S. Qutub, Phys. Rev. **D93**, 043502 (2016), 1508.04136.
- [263] K. Harigaya, M. Ibe, M. Kawasaki, and T. T. Yanagida, JCAP **1511**, 003 (2015), 1507.00119.
- [264] P. J. Steinhardt and M. S. Turner, Phys. Lett. **B129**, 51 (1983).
- [265] M. H. Reno and D. Seckel, Phys. Rev. **D37**, 3441 (1988).
- [266] A. D. Linde, Phys. Lett. **B129**, 177 (1983).
- [267] B. L. Spokoiny, Phys. Lett. **B147**, 39 (1984).
- [268] F. L. Bezrukov and M. Shaposhnikov, Phys. Lett. **B659**, 703 (2008), 0710.3755.
- [269] F. L. Bezrukov, Non-minimal coupling in inflation and inflating with the Higgs boson, in *Proceedings, 15th International Seminar on High Energy Physics (Quarks 2008)*, 2008, 0810.3165.
- [270] A. Riotto, Inflation and the theory of cosmological perturbations, in *Astroparticle physics and cosmology. Proceedings: Summer School, Trieste, Italy, Jun 17-Jul 5 2002*, pp. 317–413, 2002, hep-ph/0210162.

- [271] M. Kamionkowski and J. March-Russell, Phys. Lett. **B282**, 137 (1992), hep-th/9202003.
- [272] A. A. Starobinsky, Phys. Lett. **B91**, 99 (1980).
- [273] T. Prokopec and T. G. Roos, Phys. Rev. **D55**, 3768 (1997), hep-ph/9610400.
- [274] K. Kadota, J.-O. Gong, K. Ichiki, and T. Matsubara, JCAP **1503**, 026 (2015), 1411.3974.
- [275] K. Kadota, T. Kobayashi, and H. Otsuka, JCAP **1601**, 044 (2016), 1509.04523.
- [276] M. Yamaguchi, Class. Quant. Grav. **28**, 103001 (2011), 1101.2488.
- [277] M. Kawasaki, M. Yamaguchi, and T. Yanagida, Phys. Rev. Lett. **85**, 3572 (2000), hep-ph/0004243.
- [278] B. A. Dobrescu, Phys. Rev. **D55**, 5826 (1997), hep-ph/9609221.
- [279] S. Weinberg, Phys. Rev. **D70**, 083522 (2004), astro-ph/0405397.
- [280] D. H. Lyth and D. Wands, Phys. Lett. **B524**, 5 (2002), hep-ph/0110002.
- [281] M. Beltran, Phys. Rev. **D78**, 023530 (2008), 0804.1097.
- [282] L. Iliesiu, D. J. E. Marsh, K. Moodley, and S. Watson, Phys. Rev. **D89**, 103513 (2014), 1312.3636.
- [283] T. Moroi, K. Mukaida, K. Nakayama, and M. Takimoto, JHEP **1306**, 040 (2013), 1304.6597.
- [284] M. Kawasaki, K. Nakayama, and M. Senami, JCAP **0803**, 009 (2008), 0711.3083.
- [285] T. Kugo, I. Ojima, and T. Yanagida, Phys. Lett. **B135**, 402 (1984).
- [286] M. Endo and F. Takahashi, Phys. Rev. **D74**, 063502 (2006), hep-ph/0606075.
- [287] H. Baer, S. Kraml, A. Lessa, and S. Sekmen, JCAP **1104**, 039 (2011), 1012.3769.
- [288] A. D. Linde, Phys. Rev. **D53**, 4129 (1996), hep-th/9601083.
- [289] M. Kawasaki, N. Kitajima, and K. Nakayama, Phys. Rev. **D87**, 023513 (2013), 1112.2818.
- [290] E. J. Chun, K. Dimopoulos, and D. Lyth, Phys.Rev. **D70**, 103510 (2004), hep-ph/0402059.
- [291] K. Nakayama, F. Takahashi, and T. T. Yanagida, Phys. Rev. **D84**, 123523 (2011), 1109.2073.
- [292] M. Dine, L. Randall, and S. D. Thomas, Phys.Rev.Lett. **75**, 398 (1995), hep-ph/9503303.

- [293] M. Dine, L. Randall, and S. D. Thomas, Nucl. Phys. **B458**, 291 (1996), hep-ph/9507453.
- [294] D. H. Lyth and A. Riotto, Phys. Rept. **314**, 1 (1999), hep-ph/9807278.
- [295] R. Saito and J. Yokoyama, Phys. Rev. Lett. **102**, 161101 (2009), 0812.4339, [Erratum: Phys. Rev. Lett.107,069901(2011)].
- [296] R. Saito and J. Yokoyama, . Theor. Phys. **123**, 867 (2010), 0912.5317, [Erratum: Prog. Theor. Phys.126,351(2011)].
- [297] E. Bugaev and P. Klimai, Phys. Rev. **D83**, 083521 (2011), 1012.4697.
- [298] J. Garcia-Bellido, M. Peloso, and C. Unal, JCAP **1612**, 031 (2016), 1610.03763.
- [299] K. Inomata, M. Kawasaki, K. Mukaida, Y. Tada, and T. T. Yanagida, (2016), 1611.06130.
- [300] T. Nakamura, M. Sasaki, T. Tanaka, and K. S. Thorne, Astrophys. J. **487**, L139 (1997), astro-ph/9708060.
- [301] K. Ioka, T. Chiba, T. Tanaka, and T. Nakamura, Phys. Rev. **D58**, 063003 (1998), astro-ph/9807018.
- [302] A. S. Josan, A. M. Green, and K. A. Malik, Phys. Rev. **D79**, 103520 (2009), 0903.3184.
- [303] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).
- [304] M. Kawasaki, A. Kusenko, Y. Tada, and T. T. Yanagida, Phys. Rev. **D94**, 083523 (2016), 1606.07631.
- [305] S. Young, C. T. Byrnes, and M. Sasaki, JCAP **1407**, 045 (2014), 1405.7023.
- [306] K. N. Ananda, C. Clarkson, and D. Wands, Phys. Rev. **D75**, 123518 (2007), gr-qc/0612013.
- [307] D. Baumann, P. J. Steinhardt, K. Takahashi, and K. Ichiki, Phys. Rev. **D76**, 084019 (2007), hep-th/0703290.
- [308] E. Bugaev and P. Klimai, Phys. Rev. **D81**, 023517 (2010), 0908.0664.
- [309] L. Alabidi, K. Kohri, M. Sasaki, and Y. Sendouda, JCAP **1209**, 017 (2012), 1203.4663.
- [310] M. Maggiore, Phys. Rept. **331**, 283 (2000), gr-qc/9909001.
- [311] M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments* Oxford Master Series in Physics (Oxford University Press, 2007).
- [312] C. Gundlach, Living Rev. Rel. **2**, 4 (1999), gr-qc/0001046.
- [313] C. Gundlach, Phys. Rept. **376**, 339 (2003), gr-qc/0210101.

- [314] J. C. Niemeyer and K. Jedamzik, Phys. Rev. Lett. **80**, 5481 (1998), astro-ph/9709072.
- [315] J. C. Niemeyer and K. Jedamzik, Phys. Rev. **D59**, 124013 (1999), astro-ph/9901292.
- [316] F. Kühnel, C. Rampf, and M. Sandstad, Eur. Phys. J. **C76**, 93 (2016), 1512.00488.
- [317] A. Dolgov and J. Silk, Phys. Rev. **D47**, 4244 (1993).
- [318] J. Garcia-Bellido, A. D. Linde, and D. Wands, Phys. Rev. **D54**, 6040 (1996), astro-ph/9605094.
- [319] S. Clesse and J. García-Bellido, Phys. Rev. **D92**, 023524 (2015), 1501.07565.
- [320] S.-L. Cheng, W. Lee, and K.-W. Ng, JHEP **02**, 008 (2017), 1606.00206.
- [321] S. Blinnikov, A. Dolgov, N. K. Porayko, and K. Postnov, JCAP **1611**, 036 (2016), 1611.00541.
- [322] M. Kawasaki, T. Takayama, M. Yamaguchi, and J. Yokoyama, Phys. Rev. **D74**, 043525 (2006), hep-ph/0605271.
- [323] T. Kawaguchi, M. Kawasaki, T. Takayama, M. Yamaguchi, and J. Yokoyama, Mon. Not. Roy. Astron. Soc. **388**, 1426 (2008), 0711.3886.
- [324] P. H. Frampton, M. Kawasaki, F. Takahashi, and T. T. Yanagida, JCAP **1004**, 023 (2010), 1001.2308.
- [325] E. J. Copeland, A. R. Liddle, D. H. Lyth, E. D. Stewart, and D. Wands, Phys. Rev. **D49**, 6410 (1994), astro-ph/9401011.
- [326] A. D. Linde and A. Riotto, Phys. Rev. **D56**, R1841 (1997), hep-ph/9703209.
- [327] K. I. Izawa and T. Yanagida, Phys. Lett. **B393**, 331 (1997), hep-ph/9608359.
- [328] K. I. Izawa, M. Kawasaki, and T. Yanagida, Phys. Lett. **B411**, 249 (1997), hep-ph/9707201.
- [329] M. Kawasaki, N. Sugiyama, and T. Yanagida, Phys. Rev. **D57**, 6050 (1998), hep-ph/9710259.
- [330] M. Kawasaki and T. Yanagida, Phys. Rev. **D59**, 043512 (1999), hep-ph/9807544.
- [331] E. D. Stewart, Phys. Lett. **B391**, 34 (1997), hep-ph/9606241.
- [332] E. D. Stewart, Phys. Rev. **D56**, 2019 (1997), hep-ph/9703232.
- [333] S. M. Leach, I. J. Grivell, and A. R. Liddle, Phys. Rev. **D62**, 043516 (2000), astro-ph/0004296.
- [334] M. Drees and E. Erfani, JCAP **1104**, 005 (2011), 1102.2340.
- [335] M. Drees and E. Erfani, JCAP **1201**, 035 (2012), 1110.6052.

- [336] S. Kasuya and M. Kawasaki, Phys. Rev. **D80**, 023516 (2009), 0904.3800.
- [337] M. Kawasaki, N. Kitajima, and T. T. Yanagida, Phys. Rev. **D87**, 063519 (2013), 1207.2550.
- [338] K. Kohri, C.-M. Lin, and T. Matsuda, Phys. Rev. **D87**, 103527 (2013), 1211.2371.
- [339] D. H. Lyth, C. Ungarelli, and D. Wands, Phys. Rev. **D67**, 023503 (2003), astro-ph/0208055.
- [340] M. Yu. Khlopov and A. G. Polnarev, Phys. Lett. **B97**, 383 (1980).
- [341] K. Jedamzik, Phys. Rev. **D55**, 5871 (1997), astro-ph/9605152.
- [342] K. Jedamzik, Phys. Rept. **307**, 155 (1998), astro-ph/9805147.
- [343] K. Jedamzik and J. C. Niemeyer, Phys. Rev. **D59**, 124014 (1999), astro-ph/9901293.
- [344] E. S. Phinney, (2001), astro-ph/0108028.
- [345] A. H. Jaffe and D. C. Backer, Astrophys. J. **583**, 616 (2003), astro-ph/0210148.
- [346] J. S. B. Wyithe and A. Loeb, Astrophys. J. **590**, 691 (2003), astro-ph/0211556.
- [347] S. Clesse and J. García-Bellido, Phys. Dark Univ. **10**, 002 (2016), 1603.05234.
- [348] S. Clesse and J. García-Bellido, (2016), 1610.08479.
- [349] I. Cholis *et al.*, Phys. Rev. **D94**, 084013 (2016), 1606.07437.