Final Scientific Report

Title: Theoretical Research at the High Energy Frontier: Cosmology and Beyond
PI: Lawrence M. Krauss
Period: 04/01/2015-03/31/2017
Institution: Arizona State University
DOE Award Number: DE-SC0013605
Amount of unexpended funds: 0

Krauss' research has focused on the attempt to explore how phenomena at various extremes of scale can be used to probe fundamental physics, including field theory, general relativity, cosmology, particle phenomenology, condensed matter physics, geophysics, and astrophysics. As well, he has become involved in enhancing the public's understanding of these issues, becoming, through his books and writings, as well as radio, TV, film, and lecture appearances, a leading public advocate of particle physics and astrophysics. Over the past 2 years, his leadership of the Origins Project at ASU included a centennial celebration of General Relativity attended by over 2000 people and viewed online by over 50,000 people, and a lecture and in-depth discussion onstage with Frank Wilczek. He recently published new popular book, entitled "The Greatest Story Ever Told..So Far", released in March 2017, which provides a comprehensive and up to date history of the physics leading to the discovery of the Standard Model, and the Higgs particle.

Krauss's recent research has built on a long history of previous accomplishments, including some of the earliest calculations of remnant dark matter abundances in low energy supersymmetry models and calculations of g-2 in these models, proposals of the first bolometric detectors for WIMP dark matter and neutrinos, calculations of cosmic axion detection in magnetic fields, the first constraints from Big Bang Nucleosynthesis to self consistently include experimental and observational errors, the first detailed exploration of the terrestrial antineutrino flux and the the flux of neutrinos from all historical supernovae, the proposal that dark matter could be captured in the sun and earth and lead to possibly detectable indirect signatures, detailed explorations of the solar neutrino problem and neutrino oscillations, the theoretical proposal that 70% of the energy of the universe resides in something that resembles a cosmological constant, the proposal that discrete gauge symmetries could exist and have detectable consequences--including quantum hair on black holes as well as important implications for GUT model building, the first demonstration that Inflation at the GUT scale could result in a gravitational wave background that could produce a measurable tensor modes need to be taken in to account when analyzing CMB data, comprehensive calculations of experimental and observational uncertainties in globular cluster stars to determine of the age of the universe and its implications for the equation of state of matter and energy, explorations of the possibility of low energy baryogenesis, explorations of gravitational lensing as a probe of dark energy, explorations of new signatures for cosmic phase transitions, new explorations of the quantum nature of black hole collapse, and an exploration of Higgs physics, and of physics beyond the standard model at the LHC.

Final Scientific Report

Title: Theoretical Research at the High Energy Frontier: Cosmology and Beyond
PI: Lawrence M. Krauss
Period: 04/01/2015-03/31/2017
Institution: Arizona State University
DOE Award Number: DE-SC0013605
Amount of unexpended funds: 0

Krauss' research has focused on the attempt to explore how phenomena at various extremes of scale can be used to probe fundamental physics, including field theory, general relativity, cosmology, particle phenomenology, condensed matter physics, geophysics, and astrophysics. As well, he has become involved in enhancing the public's understanding of these issues, becoming, through his books and writings, as well as radio, TV, film, and lecture appearances, a leading public advocate of particle physics and astrophysics. Over the past 2 years, his leadership of the Origins Project at ASU included a centennial celebration of General Relativity attended by over 2000 people and viewed online by over 50,000 people, and a lecture and in-depth discussion onstage with Frank Wilczek. He recently published new popular book, entitled "The Greatest Story Ever Told..So Far", released in March 2017, which provides a comprehensive and up to date history of the physics leading to the discovery of the Standard Model, and the Higgs particle.

Krauss's recent research has built on a long history of previous accomplishments, including some of the earliest calculations of remnant dark matter abundances in low energy supersymmetry models and calculations of g-2 in these models, proposals of the first bolometric detectors for WIMP dark matter and neutrinos, calculations of cosmic axion detection in magnetic fields, the first constraints from Big Bang Nucleosynthesis to self consistently include experimental and observational errors, the first detailed exploration of the terrestrial antineutrino flux and the the flux of neutrinos from all historical supernovae, the proposal that dark matter could be captured in the sun and earth and lead to possibly detectable indirect signatures, detailed explorations of the solar neutrino problem and neutrino oscillations, the theoretical proposal that 70% of the energy of the universe resides in something that resembles a cosmological constant, the proposal that discrete gauge symmetries could exist and have detectable consequences--including quantum hair on black holes as well as important implications for GUT model building, the first demonstration that Inflation at the GUT scale could result in a gravitational wave background that could produce a measurable tensor modes need to be taken in to account when analyzing CMB data, comprehensive calculations of experimental and observational uncertainties in globular cluster stars to determine of the age of the universe and its implications for the equation of state of matter and energy, explorations of the possibility of low energy baryogenesis, explorations of gravitational lensing as a probe of dark energy, explorations of new signatures for cosmic phase transitions, new explorations of the quantum nature of black hole collapse, and an exploration of Higgs physics, and of physics beyond the standard model at the LHC.

During the term of the DOE award, Vachaspati has continued to work on the interface of particle physics and cosmology, with a focus on topological defects, primordial magnetic fields, and the outcomes of cosmic phase transitions.

Publications (04/01/2015-03/31/2017):

1. J.L. Cook, E. Dimastrogiovanni, D.A. Easson and L.M. Krauss, "Reheating Predictions in Single Field Inflation," JCAP 1504, 047 (2015).
2. J. Aalbers (NIKHEF, Amsterdam & Amsterdam U.) *et al.*). DARWIN: towards the ultimate dark matter detector DARWIN Collaboration Jun 22, 2016. 37 pp. Published in JCAP 1611 (2016) 017
3. Lawrence M. Krauss (Arizona State U., Tempe & Australian Astron. Observ.), Andrew J. Long (Chicago U., KICP). Sound Speed and Viscosity of Semi-Relativistic Relic Neutrinos Apr 4, 2016. 20 pp. Published in JCAP 1607 (2016) no.07, 002
4. J.B. Dent, L.M. Krauss, J.L. Newstead and S. Sabharwal, "A General Analysis of Direct Dark Matter Detection: From Microphysics to Observational Signatures,", Phys. Rev. D 92, 063515 (2015).
5. J.L. Cook and L.M. Krauss, "Large Slow Roll Parameters in Single Field Inflation,", JCAP 1603, no. 03, 028 (2016).
6. SNO+ Collaboration and L.M. Krauss, "Current Status and Future Prospects of the SNO+ Experiment," vol. 2016, 6194250 (2016).
7. E. Dimastrogiovanni, L.M. Krauss and J. Chluba, "Constraints on Gravitino Decay and the Scale of Inflation using CMB Spectral Distortions,", Phys. Rev. D 94, 023518 (2016).
8. A.J. Long and T. Vachaspati, "Implications of a Primordial Magnetic Field for Magnetic Monopoles, Axions and Dirac Neutrinos," Phys. Rev. D 91, 103522 (2015).
9. A.J. Long and T. Vachaspati, "Morphology of blazar-induced gamma ray halos due to a helical intergalactic magnetic field," JCAP 1509, no. 09, 065 (2015).
10. M. Brush, L. Pogosian and T. Vachaspati, "Magnetic monopole- domain wall collisions," Phys. Rev. D 92, 045008 (2015).
11. T.W.B. Kibble and T. Vachaspati, "Monopoles on Strings," J. Phys. G: Nucl. Part. Phys. 42, 094002 (2015).
12. T. Vachaspati, L. Pogosian and D. Steer, "Cosmic Strings,", Scholarpedia 10 (2): 31682 (2015).
13. T. Vachaspati, "Monopole-antimonopole Scattering,", Phys. Rev. D 93, 045008 (2016).
14. M.G. Alford, S.K. Mallavarapu, T. Vachaspati and A. Windisch, "Stability of superfluid vortices in dense quark matter," Phys. Rev. C 93, no. 4, 045801 (2016).
15. T. Vachaspati, "Fundamental Implications of Intergalactic Magnetic Field Observations," Phys. Rev. D 95, 063505 (2017).
16. R.A. Batista, A. Saveliev, G. Sigl and T. Vachaspati, "Probing Intergalactic Magnetic Fields with Simulations of Electromagnetic Cascades," Phys. Rev. D 94, 083005 (2016).
17. T. Vachaspati, "Creation of Magnetic Monopoles in Classical Scattering," Phys. Rev. Lett. 117, 181601 (2016).
18. M.G. Alford, S.K. Mallavarapu, T. Vachaspati and A. Windisch, "Vortex structure in superfluid color-flavor locked quark matter," Conference proceedings 'QCD@Work 2016'.
19. T. Vachaspati, "Gravitational Waves, Gamma Ray Bursts, and Black Stars," IJMPD Vol. 25, No. 12, 1644025 (2016)).
20. F. Duplessis and T. Vachaspati, "Probing Stochastic Inter-Galactic Magnetic Fields Using Blazar-Induced Gamma Ray Halo Morphology," JCAP 05, 005 (2017).

Prepared for submission to JCAP

Reheating predictions in single field inflation

Jessica L. Cook,[a] **Emanuela Dimastrogiovanni,**[a] **Damien A. Easson,**[a] **Lawrence M. Krauss**[a,b]

[a]Department of Physics and School of Earth and Space Exploration
Arizona State University, Tempe, AZ 85827-1404

[b]Research School of Astronomy and Astrophysics, Mt. Stromlo Observatory,
Australian National University, Canberra, Australia 2611

E-mail: jlcook14@asu.edu, emad@asu.edu, easson@asu.edu, krauss@asu.edu

Abstract.

Reheating is a transition era after the end of inflation, during which the inflaton is converted into the particles that populate the Universe at later times. No direct cosmological observables are normally traceable to this period of reheating. Indirect bounds can however be derived. One possibility is to consider cosmological evolution for observable CMB scales from the time of Hubble crossing to the present time. Depending upon the model, the duration and final temperature after reheating, as well as its equation of state, may be directly linked to inflationary observables. For single-field inflationary models, if we approximate reheating by a constant equation of state, one can derive relations between the reheating duration (or final temperature), its equation of state parameter, and the scalar power spectrum amplitude and spectral index. While this is a simple approximation, by restricting the equation of state to lie within a broad physically allowed range, one can in turn bracket an allowed range of n_s and r for these models. The added constraints can help break degeneracies between inflation models that otherwise overlap in their predictions for n_s and r.

Contents

1 Introduction

The inflationary paradigm [1–10] offers, in its numerous constructions (see e.g. [11]), a testable [12, 13] description for the physics of the very early Universe. Inflation addresses several open problems in cosmology, chief among them the question of the origin of cosmological structures. In its simplest realization, the Universe is dominated by the potential energy of a light scalar field, the inflaton, that drives the expansion. In this picture, quantum fluctuations of the scalar field during inflation are precisely the primary source of cosmological perturbations [14–19]. The statistical properties of the Cosmic Microwave Background (CMB) fluctuations and of the Large Scale Structures (LSS) may therefore contain information about the physics of inflation. In addition to scalar density perturbations, inflation generically produces tensor perturbations, resulting in a spectrum of primordial gravitational waves which, via their impacts on the CMB and other astronomical sources, reveal information about inflation [20–23].

The transition from inflation to later stages of the evolution of the Universe (radiation and matter dominance) is referred to as *reheating*. During reheating the inflaton field loses its energy, eventually leading to the production of ordinary matter. Several reheating models have been proposed: the simplest ones, involve the perturbative decay of an oscillating inflaton field at the end of inflation [24–26], while more intricate scenarios include non-perturbative processes such as (broad) parametric resonance decay [27–29], tachyonic instability [30–35], and instant preheating [36][1]. The word *preheating* indicates the initial stage of reheating, especially in the context where decay happens exponentially, generating high occupation numbers in select frequency bands. Immediately after preheating the frequency bands that underwent parametric resonance will have extremely high occupation

[1]See also [37–40] for reviews and, e.g., [41, 42] for more studies on reheating.

numbers while the rest of the space will be basically un-populated, a highly non-thermal state. Over time, scattering events will spread out the distribution, eventually leading to a blackbody spectrum characterized by a final temperature T_{re}, which normally corresponds to the temperature at the beginning of the radiation-dominated era.

For some inflationary scenarios and for given interactions between the inflaton field and other matter fields, numerical studies were performed to derive an effective equation of state (eos). The eos is parametrized by a function $w_{re}(t)$ for the Universe during the various stages of reheating. As inflation ends, the eos parameter is equal to $-1/3$. Assuming a massive inflaton, very quickly the eos climbs to 0, the eos of a massive harmonic oscillator oscillating between potential dominance (eos of -1) and kinetic dominance (eos of 1). During this initial phase of reheating, the frequency of oscillations, characterized by the inflaton mass m, will be larger than the expansion rate. It is therefore correct to approximate the eos of the inflaton as a constant of 0. This is the equation of state of the Universe at the beginning of reheating when the Universe is still dominated by the inflaton field. As the inflaton decays and the decay products compose an increasing percentage of the energy density of the Universe, the eos will increase from 0 to 1/3 at the start of radiation dominance. In [43] it was shown that for a simple chaotic inflation model and for a quartic $g^2\phi^2\chi^2$ interaction (ϕ being the inflation and χ its decay product), the equation of state right after inflation, characterized by $w_{re} = 0$, sharply, within a couple efolds, changes to $w_{re} \sim 0.2 - 0.3$ already during preheating, long before the system reaches thermal equilibrium[2]. The duration of preheating can therefore generally be regarded as "instantaneous" in comparison with the remaining stages of reheating. In cases like the ones described in [43] (see also [27]), w_{re} may therefore be rightfully treated as a constant throughout the entire reheating era.

Aside from its thermalization temperature, T_{re}, and effective equation of state, w_{re}, reheating is also characterized by its duration, which one may quantify in terms of e-foldings $N_{re} \equiv \ln(a_{re}/a_{end})$, occurring between the time inflation ends, t_{end}, and the beginning of the radiation-dominated era, t_{re}.

The reheating era is a difficult one to constrain observationally: except for some non-conventional scenarios (e.g. [45–56]). In the absence of topological defects like monopoles or strings, the fluctuations produced during reheating remain sub-horizon and cannot leave an observable imprint at the level of the CMB or LSS. A lower bound is placed on the reheating temperature by primordial nucleosynthesis (BBN) $T_{BBN} \sim 10^{-2} GeV$ [57][3]; the scale of inflation is merely bounded from above (the CMB B-modes recently measured by BICEP2 [59, 60] do not yet, unfortunately, point to an inflationary signal) and can be as large as $\sim 10^{16} GeV$, leaving for T_{re} an allowed range of many orders of magnitude . Aside for the production of metric fluctuations in the aforementioned scenarios, a variety of signatures (or lack thereof) relative to the production of primordial black holes [61–63], magnetic field [64–66], unwanted relics [67, 68] and also to mechanisms such as baryo-and leptogenesis [69–73] (and more, see [39] for an overview and for a full list of related references), may be traced back to specific preheating/reheating models.

[2] A physical system reaching an effective (macroscopic) state characterized by nearly constant ratio of pressure over energy density while it is, microscopically, still out-of-equilibrium ("pre-thermalization") had been previously investigated in Minkowski spacetime in [44].

[3] Smaller values may be assigned to the lower bound of the reheating temperature in models such as [58].

Another possibility for extracting information about reheating is to consider the expansion history of the Universe between the time the observable CMB scales crossed outside the Hubble radius during inflation and the time they later re-entered, in such a way as to define a relation between inflationary and reheating parameters [74]

$$\ln\left[\frac{k}{a_0 H_0}\right] = -N_k - N_{re} - N_{RD} + \ln\left[\frac{a_{eq}H_{eq}}{a_0 H_0}\right] + \ln\left[\frac{H_k}{H_{eq}}\right]. \tag{1.1}$$

In this equation, k can be chosen as the pivot scale for a specific experiment, N_k is the number of e-foldings between the exit time of the modes at this pivot during inflation and the end of inflation, N_{re} and N_{RD} respectively indicated the e-folds between the end of inflation and the end of reheating and between the end of reheating and the end of the radiation-dominated era. From (1.1) one realizes that from the CMB constraints on the primordial power spectrum (which would correspond to a prediction for N_k), for a given inflationary model one would be able to infer the sum of N_{RD} and N_{re}. To solve for N_{re} and N_{RD} individually one needs more information. For reheating models that can be parametrized by a constant effective pressure to energy ratio w_{re}, one can relate the density at the end of inflation to the density at the end of reheating, and then assuming conservation of entropy after reheating, to the temperature today. This way one obtains another equation with the same two unknowns N_{re} and N_{RD} that can be used to solve for each individually, or to rework the equations to trade the quantity N_{re} for T_{re}, the temperature at the end of reheating. All of this is particularly straightforward for single-field models of inflation that are entirely defined by the form of their potential. In summary, for a given inflationary model and for given equations of state during reheating lying within a reasonable physically plausible range, one may use the CMB data to place constraints on the reheating temperature and its duration. These techniques have been successfully employed in several studies [75–82].

In the same spirit as [75–82], and using similar techniques as in [81] (where the attention was directed specifically to inflation with power-law potentials, $V(\phi) \sim \phi^\alpha$), we consider the constraints imposed by reheating on popular single field inflationary scenarios. We derive predictions for the length of the reheating era, and the temperature at the end of reheating for each model, assuming a constant equation of state during reheating. Accounting for the lower bounds on T_{re} imposed by BBN and considering a physically plausible range of values for w_{re} (likely the average value will fall between 0 and $\frac{1}{3}$) we use the relations between reheating and inflationary parameters and the constraints on the primordial power spectrum amplitude and tilt from Planck [12, 13] to provide new constraints on the parameter space in given inflationary models. This is a useful and relatively new tool for constraining and differentiating between inflation models. Models might overlap in predictions for n_s and r, but not for the same w_{re}. As the constraints on n_s gets tighter, this will translate into an increasingly narrow allowed range for w_{re} for a given inflation model, and so this technique of constraining models with reheating will be increasingly efficient in ruling out some models in favor of others.

This work is organized as follows: in Sec. 2 we detail the derivation of the reheating duration and of the temperature at the end of reheating as a function of the spectral index, for canonical single-field inflationary models and for reheating scenarios that can be described in terms of a constant effective equation of state; in Sec. 3 we review the analysis of [81] for a power law potential and we discuss the constraints from reheating on the inflationary

parameters; in Secs. 4 through 7 we compute the relations between inflationary and reheating parameters in the Starobinsky, Higgs, natural and hilltop inflation models and we discuss the bounds placed on some of these models by reheating; in Sec. 8 we present our conclusions.

2 Calculating N_{re} and T_{re}

A reheating model (or class of models) may be characterized by a thermalization temperature T_{re}, a duration, N_{re} (here defined in terms of the number of e-folds counted from the end of inflation), and an equation of state with an effective pressure-to-energy-density ratio, w_{re}. The latter should have values larger than $-1/3$ for inflation to come to an end, and is assumed to be smaller than 1 in order not to violate causality. A variety of reheating scenarios allow for an equation of state that is nearly constant in time. For the purposes of this work we will thus approximate w_{re} as a constant in all our calculations; in our plots for N_{re} and T_{re}, we assign to w_{re} sample values ranging in the interval $[-1/3, 1]$. We define N_{re} as the time frame from the end of inflation until the equation of state makes a step function transition from the value w_{re} it had during reheating to $w = 1/3$, which we define as the start of radiation dominance. T_{re} is the temperature when this transition occurs. From this definition, N_{re} and T_{re} are not well defined if the equation of state during reheating is also equal to $1/3$ (we will discuss this case more later). Also, we assume a standard expansion history after reheating, with a radiation-dominated (RD) era followed by a matter-dominated (MD) one.

We derive, following [75–82], an expression for the reheating parameters (N_{re}, T_{re} and w_{re}) in terms of a set of physical quantities that are specific to inflation and to the cosmological epochs subsequent to reheating. Considering the evolution of the Universe between the Hubble-exit time during inflation (henceforth indicated by t_k) for observable scales and the time of observation of the same scales (t_0), one can write matching conditions for the total energy density as well as for the scale factor, $a(t)$, during the intermediate eras. Fig. (1) summarizes the evolution of the comoving horizon distance throughout this length of time, marked by the transitions between consecutive epochs at t_{end}, the end of inflation, t_{re}, the end of reheating/beginning of RD era, and t_{eq}, the beginning of the MD era.

In the figure we equate the size the comoving horizon far back into inflation, corresponding to modes $l = 2$, to the size of the horizon today. In order to solve the horizon problem, the span of comoving scales that leave the horizon from $l = 2$ to the end of inflation must equal the span of comoving scales that reenter the horizon after inflation till today. Note the factor by which the comoving horizon shrinks between scales $l = 2$ and the end of inflation (the length of the first line in the figures) is not known. The slope of that line is set by the fact that the equation of state is ≈ -1 during inflation. Depending on the model, that line could be longer or shorter. While there is a minimum length in order to solve the horizon problem while having Inflation occur before BBN, there is no upper bound. The value of w_{re} will set the slope of the second line, the rate by which modes reenter the horizon during reheating. In the figure we display the two extreme cases of $w_{re} = 1$ and $w_{re} = -1/3$. One can see from comparing the two plots, the smaller w_{re} is during reheating, the less efficiently modes re-enter the horizon, and the more efolds will be necessary in the post-inflation period.

We consider single-field inflationary models with background field equations, $\ddot{\phi} + 3H\dot{\phi} + V^{'} = 0$ and $3H^2 M_P^2 \simeq V(\phi)$. We also assume that both ϵ and η remain smaller than 1 throughout the inflationary regime.

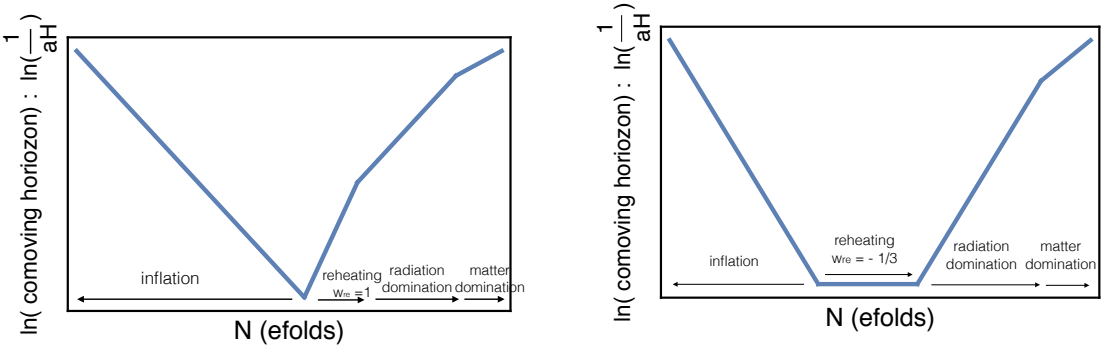

Figure 1: Each figure shows the evolution of the comoving horizon distance over time. Each figure shows the extreme cases for w_{re}: the first figure for $w_{re} = 1$ and the second for $w_{re} = -\frac{1}{3}$.

If one assumes a constant equation of state, the change in the scale factor during reheating is easily related to the change in the energy density. Using $\rho \propto a^{-3(1+w)}$, the reheating epoch is described by

$$\frac{\rho_{end}}{\rho_{re}} = \left(\frac{a_{end}}{a_{re}}\right)^{-3(1+w_{re})}, \tag{2.1}$$

where the subscript *end* refers to the end of inflation (the start of reheating), and *re* refers to the end of reheating. Writing this in terms of e-foldings

$$N_{re} = \frac{1}{3(1+w_{re})} \ln\left(\frac{\rho_{end}}{\rho_{re}}\right) = \frac{1}{3(1+w_{re})} \ln\left(\frac{3}{2}\frac{V_{end}}{\rho_{re}}\right), \tag{2.2}$$

where the last step of (2.2) is obtained by replacing $\rho_{end} = (3/2)V_{end}$, derived by setting $w = -1/3$ at the end of inflation.
The temperature is related to the density by

$$\rho_{re} = \frac{\pi^2}{30} g_{re} T_{re}^4, \tag{2.3}$$

where g_{re} is the number of relativistic species at the end of reheating. Combining Eqs. (2.2) and (2.3) one finds

$$N_{re} = \frac{1}{3(1+w)} \ln\left(\frac{30 \cdot \frac{3}{2} V_{end}}{\pi^2 g_{re} T_{re}^4}\right). \tag{2.4}$$

Making the standard assumption that entropy is conserved between the end of reheating and today, one can relate the reheating temperature to the temperature today by taking into account the changing number of helicity states in the radiation gas as a function of temperature,

$$T_{re} = T_0 \left(\frac{a_0}{a_{re}}\right) \left(\frac{43}{11 g_{re}}\right)^{\frac{1}{3}} = T_0 \left(\frac{a_0}{a_{eq}}\right) e^{N_{RD}} \left(\frac{43}{11 g_{re}}\right)^{\frac{1}{3}}, \tag{2.5}$$

where N_{RD} is the length in e-folds of radiation dominance, $e^{-N_{RD}} \equiv a_{re}/a_{eq}$. The ratio a_0/a_{eq} can be rewritten as

$$\frac{a_0}{a_{eq}} = \frac{a_0 H_k}{k} e^{-N_k} e^{-N_{re}} e^{-N_{RD}} \,, \tag{2.6}$$

where one uses the relation $k = a_k H_k$ for the time at which the pivot scale k [4] crosses outside the Hubble radius and N_k is defined as the number of e-foldings between the latter and the time inflation ends. Inserting (2.6) into (2.5) one finds

$$T_{re} = \left(\frac{43}{11 g_{re}}\right)^{\frac{1}{3}} \left(\frac{a_0 T_0}{k}\right) H_k e^{-N_k} e^{-N_{re}}. \tag{2.7}$$

Notice that larger values of N_{re} corresponds to smaller T_{re} and vice versa. In other words, as expected, the quicker and more efficiently reheating takes place, the larger the temperature. Plugging (2.7) into Eq. (2.4)

$$N_{re} = \frac{4}{3(1+w_{re})} \left[\frac{1}{4} \ln\left(\frac{3^2 \cdot 5}{\pi^2 g_{re}}\right) + \ln\left(\frac{V_{end}^{\frac{1}{4}}}{H_k}\right) + \frac{1}{3} \ln\left(\frac{11 g_{re}}{43}\right) + \ln\left(\frac{k}{a_0 T_0}\right) + N_k + N_{re}\right]. \tag{2.8}$$

One can first solve for N_{re} assuming $w_{re} \not\equiv \frac{1}{3}$

$$N_{re} = \frac{4}{(1-3w_{re})} \left[-\frac{1}{4} \ln\left(\frac{3^2 \cdot 5}{\pi^2 g_{re}}\right) - \frac{1}{3} \ln\left(\frac{11 g_{re}}{43}\right) - \ln\left(\frac{k}{a_0 T_0}\right) - \ln\left(\frac{V_{end}^{\frac{1}{4}}}{H_k}\right) - N_k\right]. \tag{2.9}$$

Notice that the values of the last two terms in Eq. (2.9) depend on the specific inflationary model. Assuming $g_{re} \approx 100$ and using Planck's pivot of $0.05 Mpc^{-1}$ [5], one obtains a simplified expression for N_{re}, before specifying a particular inflationary model:

$$N_{re} = \frac{4}{(1-3w_{re})} \left[61.6 - \ln\left(\frac{V_{end}^{\frac{1}{4}}}{H_k}\right) - N_k\right]. \tag{2.10}$$

One can then use Eq. (2.7) to obtain

$$T_{re} = \left[\left(\frac{43}{11 g_{re}}\right)^{\frac{1}{3}} \frac{a_0 T_0}{k} H_k e^{-N_k} \left[\frac{3^2 \cdot 5 V_{end}}{\pi^2 g_{re}}\right]^{-\frac{1}{3(1+w_{re})}}\right]^{\frac{3(1+w_{re})}{3w_{re}-1}}. \tag{2.11}$$

[4] Note in the following when we repeatedly refer to the pivot scale, we will use throughout Planck's pivot scale of $0.05 Mpc^{-1}$.

[5] The convention in the Planck analysis defines the pivot scale such that the comoving momentum k becomes horizon sized when $k a_0 = aH$, where we have been using $k = aH$, so using our conventions $\frac{k}{a_0} = 0.05 Mpc^{-1}$.

2.1 Special case $w_{re} = \frac{1}{3}$

The final result for N_{re} in Eq. (2.10) only applies for $w_{re} \neq 1/3$. Going back to Eq. (2.8), notice that if $w_{re} = \frac{1}{3}$, N_{re} cancels from both sides of the equation, and one is left with

$$0 = \frac{1}{4}\ln\left(\frac{30}{\pi^2 g_{re}}\right) + \frac{1}{4}\ln\left(\frac{3}{2}\right) + \ln\left(\frac{V_{end}^{\frac{1}{4}}}{H_k}\right) + \frac{1}{3}\ln\left(\frac{11 g_{re}}{43}\right) + \ln\left(\frac{k}{a_0 T_0}\right) + N_k\,. \tag{2.12}$$

Assuming $g_{re} = 100$, and Planck's pivot scale, this simplifies to:

$$61.6 = \ln\left(\frac{V_{end}^{\frac{1}{4}}}{H_k}\right) + N_k\,. \tag{2.13}$$

For $w = 1/3$, it is not possible to derive a prediction for N_{re} or T_{re} but instead, for a particular inflation model, one finds a prediction for n_s. Note the ambiguity in N_{re} and T_{re} is due to the fact that we are defining the start of radiation dominance as the moment w_{re} reaches $1/3$. If w_{re} is already equal to $1/3$ during reheating, then there is ambiguity in when to differentiate between the two regimes.

2.2 Model dependent part

In order to solve for N_{re} in Eq. (2.10) (or to solve for n_s in Eq. (2.13) if $w_{re} = 1/3$) for a particular model, one needs to compute N_k, H_k, and V_{end}. N_k can be calculated starting from the definition of e-foldings:

$$\Delta N = \int H dt\,. \tag{2.14}$$

Recasting the r.h.s. of (2.14) as an integral over ϕ and using the background equation of motion for the inflaton, $3H\dot{\phi} + V' \simeq 0$, and the Friedmann equation, $H^2 \simeq V/(3M_P^2)$, one finds

$$N_k \simeq \frac{1}{M_P^2}\int_{\phi_{end}}^{\phi_k} \frac{V}{V'}\, d\phi\,. \tag{2.15}$$

Next, H_k can be written as a function of n_s. Using the definition of the tensor-to-scalar ratio $r = P_h/P_\zeta$ (where $P_h = (2H^2)/(\pi^2 M_P^2)$ and $P_\zeta = A_s$ at the pivot scale)

$$r_k = \frac{2H_k^2}{\pi^2 M_P^2 A_s}. \tag{2.16}$$

Then using $r = 16\epsilon$ this gives

$$H_k \simeq \pi M_P \sqrt{8 A_s \epsilon_k}. \tag{2.17}$$

Once the form of $V(\phi)$ is specified for a given model, one can express V_{end} as a function of model parameters calculated at the pivot scale. The explicit form of V_{end} along with (2.15) and (2.17) can be plugged into Eqs. (2.10) and (2.11) to derive N_{re} and T_{re} as a function of inflationary model parameters (or into Eq. (2.13) in the case $w_{re} = 1/3$).

3 Polynomial potentials

Consider a polynomial type potential

$$V = \frac{1}{2} m^{4-\alpha} \phi^{\alpha}. \tag{3.1}$$

This was considered in the context of reheating in [11, 75, 77, 81, 82]. We quickly review this specific application. At the end of this section, we discuss with some quantitative examples how closely the constraints from inflation compare to the ones from reheating.
The first step is to calculate the model dependent parameters in Eq. (2.10), i.e. N_k, H_k, and V_{end}. The number of e-folds between the time the pivot scale exited the Hubble radius and the end of inflation can be derived using Eq. (2.15)

$$N_k = \frac{1}{2\alpha M_P^2} \left(\phi_k^2 - \phi_{end}^2 \right). \tag{3.2}$$

The potential in these polynomial models is generally steep enough so that $\phi_k \gg \phi_{end}$ and it is appropriate to approximate

$$N_k \approx \frac{1}{2\alpha M_P^2} \phi_k^2 \,. \tag{3.3}$$

We now require N_k as a function of n_s. From the expression of the spectral index as a function of the slow-roll parameters, $n_s = 1 - 6\epsilon + 2\eta$ (where $\epsilon = (M_P^2/2)(V'/V)^2$ and $\eta = M_P^2 V''/V$), and using (3.3) to rewrite ϵ and η as functions of N_k, one finds

$$N_k = \frac{\alpha + 2}{2(1 - n_s)}. \tag{3.4}$$

From Eq. (2.17) and using the previous equation, H_k is given by

$$H_k = \pi M_P \sqrt{\frac{4\pi A_s}{\alpha + 2}(1 - n_s)}. \tag{3.5}$$

Lastly one computes V_{end} in terms of n_s and A_s,

$$V_{end} = 3 M_P^2 H_k^2 \frac{\phi_{end}^{\alpha}}{\phi_k^{\alpha}} = 6\pi^2 M_P^4 A_s (1 - n_s) \left(\frac{\alpha(1 - n_s)}{2(\alpha + 2)} \right), \tag{3.6}$$

where the value of the inflaton field at the end of inflation was computed by solving for ϕ_{end} from the condition $\epsilon = 1$.

Thus N_k, H_k, and V_{end} are all expressed as functions only of α, n_s and A_s and one may plot N_{re} (and T_{re}) as a function of n_s for some fixed values of w_{re} and α. We use $n_s = 0.9682 \pm 0.0062$ and Planck's central value $A_s = 2.196 \times 10^{-9}$ (small variations in A_s have negligible effects on reheating predictions).

We plot in Fig. 2 N_{re} and T_{re} predictions for $\alpha = 2/3$, 1, 2 and 4. The case $\alpha = 2/3$ is favored by axion-monodromy models, and $\alpha = 1$ and $\alpha = 2$ give promising predictions when compared with the Planck data. The case $\alpha = 4$ is difficult to reconcile with $w_{re} \leq 1$ even

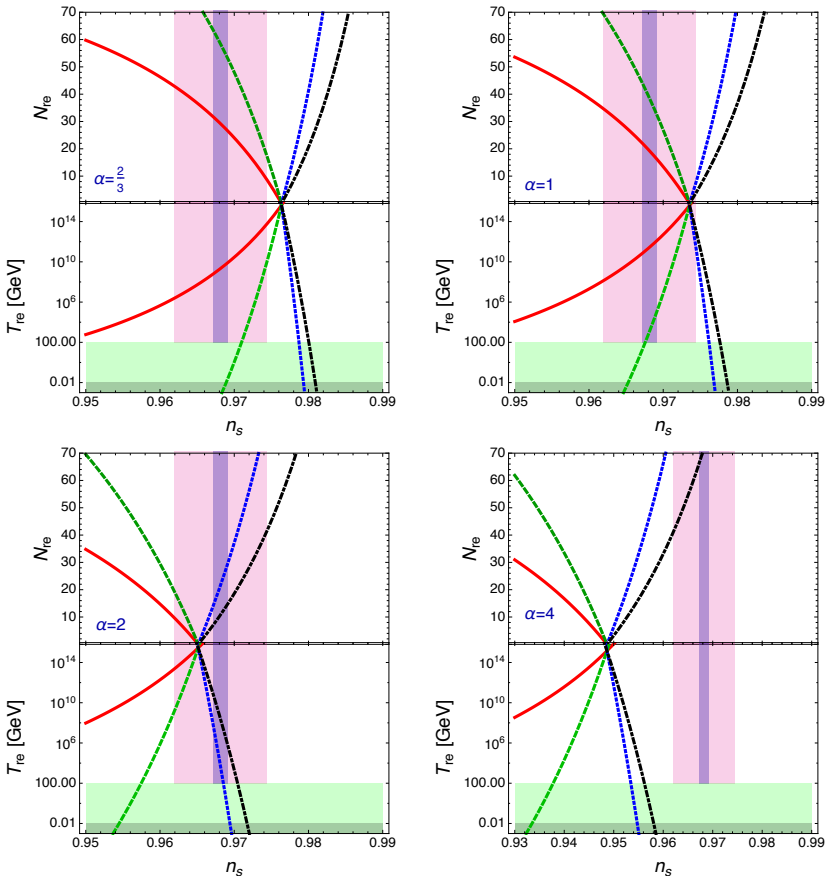

Figure 2: Plots of N_{re} and T_{re}, the length of reheating and the temperature at the end of reheating respectively, for polynomial potentials with exponent α. The solid red line corresponds to $w_{re} = -1/3$, the dashed green line to $w_{re} = 0$, the dotted blue line to $w_{re} = 2/3$, and the dot-dashed black line to $w_{re} = 1$. The pink shaded region corresponds to the 1σ bounds on n_s from Planck. The purple shaded region corresponds to the 1σ bounds of a further CMB experiment with sensitivity $\pm 10^{-3}$ [83, 84], using the same central n_s value as Planck. Temperatures below the dark green shaded region are ruled out by BBN. The light green shaded region is below the electroweak scale, assumed 100 GeV for reference. This region is not disallowed but would be interesting in the context of baryogenesis.

considering the 2σ bounds on n_s[6].

Instantaneous reheating is defined as the limit $N_{re} \to 0$, visualized in the figure as the point where all the lines converge. Such instantaneous reheating leads to the maximum temperature at the end of reheating, and the equation of state parameter is irrelevant.

(Thus, while not shown, a $w_{re} = \frac{1}{3}$ solution would correspond to a vertical line passing through the instantaneous reheat point.)

From Fig. 2, $\alpha = 2/3$ can be consistent with Planck bounds, but assuming an equation of state $w_{re} \geq 0$, the model would tend to predict smaller reheating temperatures if one considers Planck's 1σ bound on n_s; using Planck's 2σ bounds, any reheating temperature up

[6] An exception where ϕ^4 may still be viable is in the context of warm inflation [85, 86].

to the maximum instantaneous case is still allowed.

For $\alpha = 1$ and $\alpha = 2$ all the lines in Fig. 2 are shifted towards the central value of n_s when compared to the $\alpha = 2/3$ case, thus allowing for a wider range of reheating temperatures as well as values of the equation of state parameter.

Consider now the case $w_{re} = 1/3$. Solving Eq. (2.13) for the polynomial potential, one obtains

$$61.6 = \frac{1}{4}\ln\left(\frac{3\alpha}{4\pi^2 A_s(\alpha+2)}\right) + \frac{\alpha+2}{2(1-n_s)}. \tag{3.7}$$

Using Planck's central value for A_s, Eq. (3.7) gives specific predictions for n_s

$$\begin{cases} n_s = 0.977 & \text{for} \quad \alpha = \frac{2}{3}\,, \\ n_s = 0.974 & \text{for} \quad \alpha = 1\,, \\ n_s = 0.965 & \text{for} \quad \alpha = 2\,. \end{cases} \tag{3.8}$$

Notice that larger values of α require smaller values of n_s. With the 2σ bounds on n_s from Planck, $0.956 < n_s < 0.981$, $w_{re} = \frac{1}{3}$ would be consistent with all three values of α.

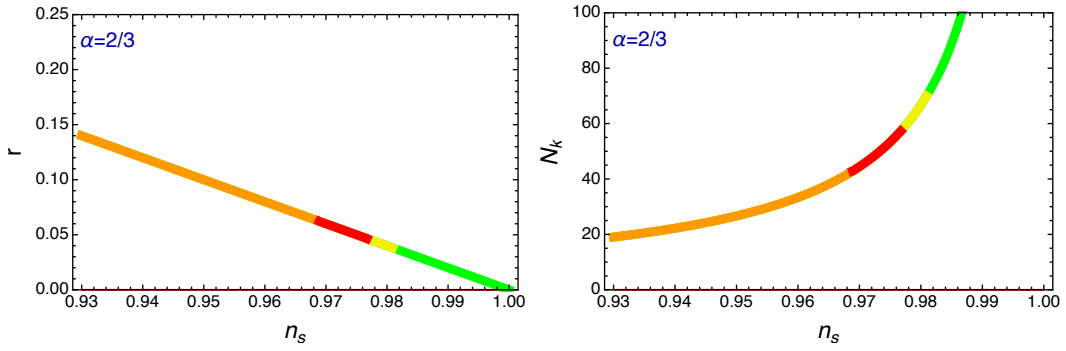

Figure 3: Parameter space for $\phi^{2/3}$ inflation. The figures show r and N_k predictions that give the correct A_s for the plotted n_s at the pivot scale. The green portion of the line comprises the region of parameter space corresponding to reheating models with $w_{re} > 1$, the yellow part corresponds to $w_{re} > 1/3$, red to $w_{re} < 1/3$ and orange to $w_{re} < 0$. Note the most likely w_{re}, between 0 and 1/3, falls in the red region.

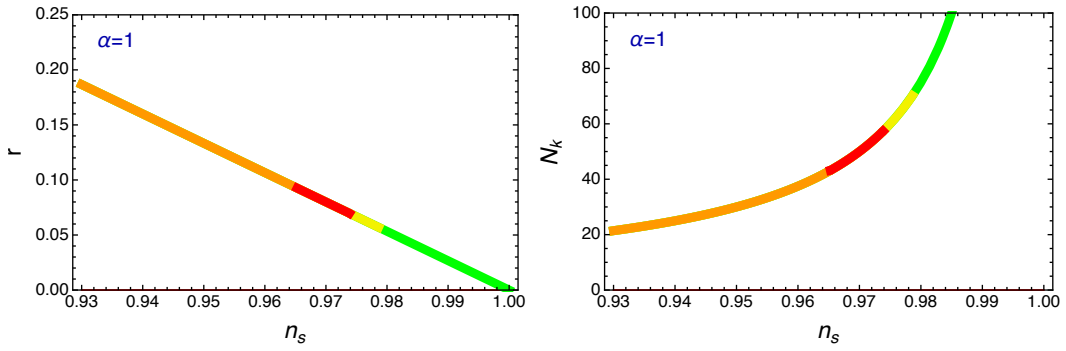

Figure 4: Parameter space for ϕ inflation. Shading is as for Fig. (3).

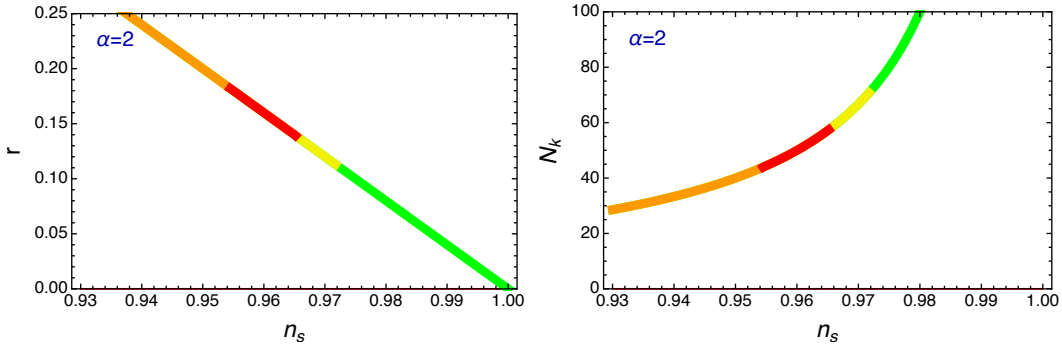

Figure 5: Parameter space for ϕ^2 inflation. Shading is as for Fig. (3).

Figs. 3-5 shows the parameter space in the r and N_k vs. n_s plane, corresponding to the different reheating scenarios. We allow for any $N_k > 19$. We note again that there is no maximum allowed N_k. A minimum on N_k is determined by the temperature at the end of reheating in order to solve the horizon and flatness problems. One finds that $N > 24.9$ if reheating after inflation is to be above the BBN scale and $N > 34.8$ for reheating above the electroweak scale in order for scales on the order of the horizon today (i.e $l = 2$) to have left the horizon during inflation. A simple estimate of the ratio of expansion scales between $l = 2$, and Planck's pivot scale, at $l \approx 685.8$, if the expansion rate during Inflation were constant, is $\Delta N \approx \ln(l_2/l_1) \approx 6.5$. However, in the large field modes we are considering, the variation in H is not negligible and the exact $\Delta N = \ln(\frac{k_2}{k_1}\frac{H_1}{H_2})$ is closer to $\Delta N \approx 5.9$. This means that for reheating greater than the BBN scale, one finds $N_k \geq 19$ (or $N_k \geq 29$ for reheating above the electroweak symmetry breaking scale).

The green part of the line in Fig. 3 corresponds to the region of parameter space that requires reheating models with w_{re} larger than one, the yellow part corresponds to $w_{re} > 1/3$, red to $w_{re} < 1/3$ and orange to $w_{re} < 0$. We stress that a value of w_{re} between 0 and 1/3 is most likely and these solutions fall in the red band in Fig. 3.
One can see that requiring $0 \leq w_{re} \leq 1/3$ corresponds to respectively setting an upper and a lower bound on the tensor-to-scalar ratio

$$\begin{cases} 0.05 \leq r \leq 0.06 & \text{for} \quad \alpha = \frac{2}{3}\,, \\ 0.07 \leq r \leq 0.09 & \text{for} \quad \alpha = 1\,, \\ 0.14 \leq r \leq 0.18 & \text{for} \quad \alpha = 2\,. \end{cases} \tag{3.9}$$

Since it now appears that the majority of BICEP2's signal is comprised of dust [87], it is difficult to find a viable reheating scenario for ϕ^2 inflation; if we loosen our restriction to just requiring $w_{re} < 1$ then one obtains a bound $r \geq 0.11$, which is just inside the 2σ limit [87].

The assumption $0 \leq w_{re} \leq 1/3$ results in tighter constraints on r than Planck's 2σ bound on n_s alone. For ϕ^2, the n_s 2σ bound yields $0.08 \leq r \leq 0.18$. Restricting w_{re} also provides stronger constraints on N_k: for ϕ^2, the n_s 2σ bound yields $45 \leq N_k \leq 103$, whereas $0 \leq w_{re} \leq 1/3$ yields $44 \leq N_k \leq 57$.

4 Starobinsky model

The action for the Starobinsky model [1] has the form

$$S = \int d^4x\sqrt{-g}\left[\frac{M_P^2}{2}(R+\alpha R^2) + \mathcal{L}_{matter}\right], \tag{4.1}$$

where R is the Ricci scalar. Performing a conformal transformation [88, 89]

$$\tilde{g}_{\mu\nu} = \omega^2 g_{\mu\nu}, \tag{4.2}$$

where $\omega^2 = 1 + 2\alpha R$, the action (4.1) is rewritten as the canonical Einstein-Hilbert action plus other terms which form a modified $\mathcal{L}_{matter}$

$$S = \int d^4x\sqrt{-\tilde{g}}\left[\frac{M_P^2}{2}\left[\tilde{R} - \frac{\alpha\phi^2}{(1+2\alpha\phi)^2} - \frac{6\alpha^2}{(1+2\alpha\phi)^2}(\tilde{\partial}\phi)^2\right] + \mathcal{L}_{matter}\right], \tag{4.3}$$

where what we now call ϕ is equal to R, the original, untransformed Ricci scalar. Notice that $\tilde{\partial}^\alpha$ carries factors of the metric, therefore $\not\equiv \partial^\alpha$. Next one defines $\bar{\phi}$, a canonically normalized version of ϕ

$$\bar{\phi} = \sqrt{\frac{3}{2}} M_P \ln(1+2\alpha\phi). \tag{4.4}$$

Rewriting the action in terms of $\bar{\phi}$ one finds

$$S = \int d^4x\sqrt{-\tilde{g}}\left[\frac{M_P^2}{2}\left[\tilde{R} - \frac{1}{4\alpha}\left(1 - e^{-\sqrt{\frac{2}{3}}\frac{\bar{\phi}}{M_P}}\right)^2\right] - \frac{1}{2}(\tilde{\partial}\bar{\phi})^2 + e^{-2\sqrt{\frac{2}{3}}\frac{\bar{\phi}}{M_P}}\mathcal{L}_{matter}\right]. \tag{4.5}$$

If one assumes that the other fields in $\mathcal{L}_{matter}$ are subdominant during inflation and can be ignored, then one can verify that this Einstein frame action behaves as normal gravity plus a canonical scalar field with the potential

$$V = \frac{M_P^2}{8\alpha}\left(1 - e^{-\sqrt{\frac{2}{3}}\frac{\bar{\phi}}{M_P}}\right)^2. \tag{4.6}$$

Dropping the bar on ϕ from now on, but continuing to work with the canonical version of the field, one can easily compute the number of e-foldings between the horizon exit of the pivot scale and the end of inflation

$$N_k = \frac{1}{M_P^2}\int_{\phi_{end}}^{\phi_k}\frac{V}{V'}\,d\phi = \frac{1}{2M_P^2}\sqrt{\frac{3}{2}}\left[M_P\sqrt{\frac{3}{2}}e^{\sqrt{\frac{2}{3}}\frac{\phi}{M_P}} - \phi\right]\Big|_{\phi_{end}}^{\phi_k}. \tag{4.7}$$

With the approximations $\phi_k \gg \phi_{end}$, and $M_P e^{\sqrt{\frac{2}{3}}\frac{\phi_k}{M_P}} \gg \phi_k$, the previous expression simplifies to

$$N_k = \frac{3}{4}e^{\sqrt{\frac{2}{3}}\frac{\phi_k}{M_P}} \tag{4.8}$$

which can be inverted for ϕ_k

$$\phi_k = \sqrt{\frac{3}{2}} M_P \ln\left(\frac{4}{3}N_k\right) \tag{4.9}$$

The next step is to compute ϵ_k and η_k in order to derive N_k as a function of n_s using $n_s = 1 - 6\epsilon + 2\eta$. The slow-roll parameters have the following form

$$\epsilon_k \simeq \frac{3}{4N_k^2}, \qquad \eta_k = -\frac{1}{N_k}, \tag{4.10}$$

where Eq. (4.9) was used along with the approximation $N_k \gg 1$. From Eq. (4.10) then one finds

$$N_k = \frac{2}{1 - n_s}. \tag{4.11}$$

Using the expressions above, one derives H_k as a function of n_s and A_s

$$H_k = \pi M_P \sqrt{\frac{3}{2} A_s}(1 - n_s), \tag{4.12}$$

$$V_{end} = \frac{9}{2}\pi^2 M_P^4 A_s (1 - ns)^2 \frac{\left(\frac{1}{\frac{\sqrt{3}}{2}+1}\right)^2}{\left(1 - \frac{3}{8}(1 - n_s)\right)^2} \tag{4.13}$$

Eqs. (4.11)-(4.13) are all that is needed to derive the results for the duration and for the temperature of reheating.

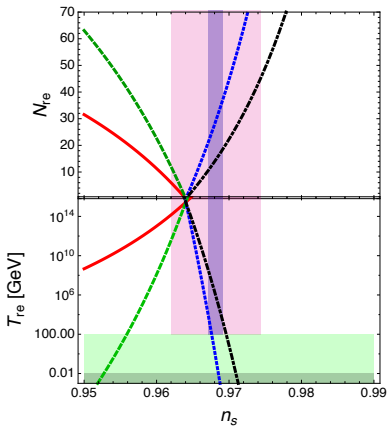

Figure 6: Plots of N_{re} and T_{re}, the length of reheating and the temperature at the end of reheating respectively, for Starobinsky and Higgs inflation. All curves and shaded regions are as for Fig. 2

Fig. 6 shows good compatibility with Planck's 1σ bounds on n_s for all the possible w_{re} values. Also, if one does not put any restrictions on the value of w_{re} then any temperature between the BBN bound and the instantaneous reheating value is allowed within the 1σ bound.

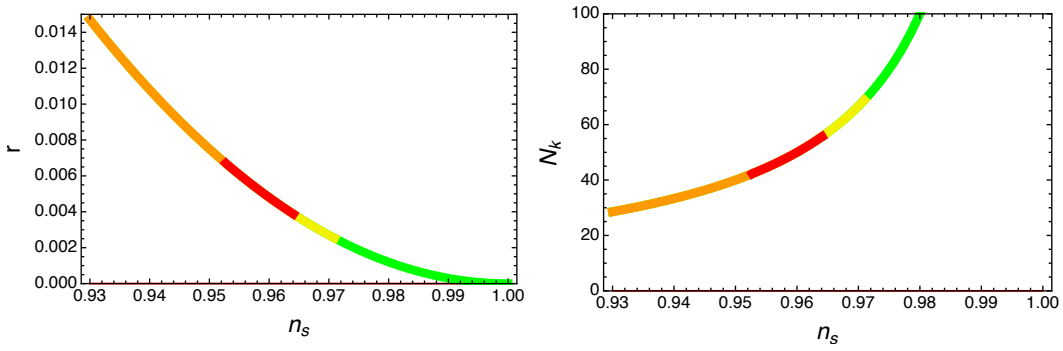

Figure 7: Parameter space for Starobinsky inflation. Shading is as for Fig. 3.

We plot in Fig. 7 the parameter space in the r and N_k vs. n_s plane for Starobinsky inflation, for different ranges of w_{re}[7]. For $0 < w_{re} < 1/3$, the corresponding range for the spectral index is $0.953 < n_s < 0.964$. This also corresponds to the range $0.004 \leq r \leq 0.007$ and $42 \leq N_k \leq 56$.

5 Higgs Inflation

The idea behind Higgs inflation [90] is to allow the Standard Model Higgs field to be the inflaton by adding a non-minimal coupling to gravity. The Jordan frame action is

$$S = \int d^4x\sqrt{-g}\left[\frac{M_P^2}{2}R\left(1+2\xi\frac{H^\dagger H}{M_P^2}\right)+\mathcal{L}_{matter}\right], \tag{5.1}$$

where H is the Higgs doublet. We may again perform a conformal transformation to write the action in the form of Einstein gravity plus a modified $\mathcal{L}_{matter}$. The transformation is given by $\tilde{g}_{\mu\nu} = \omega^2 g_{\mu\nu}$ with $\omega^2 = 1 + 2\xi\frac{H^\dagger H}{M_P^2}$. Rewriting the action in terms of the transformed metric, we find

$$S = \int d^4x\sqrt{-\tilde{g}}\left[\frac{M_P^2}{2}\tilde{R} - \frac{3\xi^2}{\omega^4 M_P^2}\left(\tilde{\partial}H^\dagger H\right)^2 + \frac{1}{\omega^4}\mathcal{L}_{matter}\right]. \tag{5.2}$$

Next, one extracts the kinetic and potential terms for the Higgs field contained within $\mathcal{L}_{matter}$. One can use $V_h = \frac{\lambda}{4}(H^\dagger H - \frac{\nu^2}{2})^2$, dropping the ν part (we are interested in inflation scales much larger than electroweak scale). Ignoring all the Higgs interactions with other fields, and only considering its self coupling (which we assume is the dominant term in the Higgs potential at inflation scales)

$$S = \int d^4x\sqrt{-\tilde{g}}\left[\frac{M_P^2}{2}\tilde{R} - \frac{3\xi^2}{\omega^4 M_P^2}\left(\tilde{\partial}H^\dagger H\right)^2 - \frac{1}{\omega^4}\left(\partial H^\dagger\right)^2 - \frac{\lambda}{4\omega^4}\left(H^\dagger H\right)^2 + \frac{1}{\omega^4}\mathcal{L}_{matter}\right], \tag{5.3}$$

where now $\mathcal{L}_{matter}$ comprises all the matter fields except the Higgs. Note: one needs to convert $\partial^\alpha \to \omega^2\tilde{\partial}^\alpha$. The Higgs is no longer canonical because of the effect of the non-minimal coupling. To canonically normalize all four of the Higgs degrees of freedom, one

[7] See also [11, 82] for n_s vs r plots in the Starobinsky model for $w_{re} = 0$.

must work in unitary gauge, where three of the four degrees of freedom are equal to zero, $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ h \end{pmatrix}$

$$S = \int d^4x \sqrt{-\tilde{g}} \left[\frac{M_P^2}{2} \tilde{R} - \frac{3\xi^2 h^2}{\omega^4 M_P^2} \left(\tilde{\partial} h \right)^2 - \frac{1}{2\omega^2} \left(\tilde{\partial} h \right)^2 - \frac{\lambda}{4\omega^4} h^4 + \frac{1}{\omega^4} \mathcal{L}_{matter} \right] . \tag{5.4}$$

The canonically normalized version of h is $\bar{h}$, defined as

$$\frac{\partial \bar{h}}{\partial h} = \frac{1}{\left(1 + \frac{\xi}{M_P^2} h^2\right)} \sqrt{1 + \frac{\xi}{M_P^2} h^2 (6\xi + 1)}. \tag{5.5}$$

Before integrating the previous equation, it is useful to introduce a few approximations. First one uses $6\xi \gg 1$. To get a successful inflation model, one should require $\xi \approx 10^4$. Next one uses the condition $(6\xi)/(M_P^2)h^2 \gg 1$. $h \approx M_P$ when inflation ends, and therefore $h > M_P$ for the duration of inflation. This allows one to rewrite Eq. (5.5) as

$$\bar{h} = \frac{\sqrt{6}\xi}{M_P} \int dh \, \frac{h}{1 + \frac{\xi h^2}{M_P^2}} , \tag{5.6}$$

which integrates to

$$\bar{h} = \sqrt{\frac{3}{2}} M_P \ln \left(1 + \frac{\xi h^2}{M_P^2} \right) . \tag{5.7}$$

Rewriting the action in terms of $\bar{h}$, one finds

$$S = \int d^4x \sqrt{-\tilde{g}} \left[\frac{M_P^2}{2} \tilde{R} - \frac{1}{2} \left(\tilde{\partial} \bar{h} \right)^2 - \frac{\lambda M_P^4}{4\xi^2} \left(1 - e^{-\sqrt{\frac{2}{3}} \frac{\bar{h}}{M_P}} \right)^2 + e^{-2\sqrt{\frac{2}{3}} \frac{\bar{h}}{M_P}} \mathcal{L}_{matter} \right] , \tag{5.8}$$

The potential term for the canonical field takes the same form as the Starobinsky potential with the identification $(1)/(8\alpha) = (\lambda M_P^2)(4\xi^2)$. Since we have a canonical field evolving in the same potential as the Starobinsky case, the Higgs inflation model gives the same predictions for N_{re} and T_{re} (see also [11, 82]). We note Starobinsky and Higgs inflation have different low scale behavior [91–95] and so while the allowed parameter space as a function of w_{re} is the same, the w_{re} that is most likely for Starobinsky vs. Higgs inflation is likely to differ. Tighter constraints could be obtained by considering gravitational, Planck suppressed couplings in the Starobinsky case, and standard model couplings in the Higgs case [91–93]. Of course new physics may modify the running of the couplings or add new couplings at these high scales (see for example [94]); in this respect, our approach of characterizing an allowed parameter space by assuming a range of w_{re} between 0 and 1/3 can usefully help bracket different allowed scenarios.

6 Natural Inflation

The potential for natural inflation is [96]

$$V(\phi) = \Lambda^4 \left[1 + \cos \left(\frac{\phi}{f} \right) \right] . \tag{6.1}$$

The number of e-folds N_k between the time the pivot scale modes crossed outside the horizon and the end of inflation is given by

$$N_k = \left(\frac{f}{M_P}\right)^2 \ln\left[\frac{\sin^2\left(\chi_{end}/2\right)}{\sin^2\left(\chi_{in}/2\right)}\right] \tag{6.2}$$

where $\chi \equiv \phi/f$. The slow-roll parameters have the following form

$$\epsilon = \frac{1}{2}\left(\frac{M_P}{f}\right)^2 \left[\frac{1-\cos(\chi)}{1+\cos(\chi)}\right], \qquad \eta = -\left(\frac{M_P}{f}\right)^2 \frac{\cos\chi}{1+\cos\chi}. \tag{6.3}$$

The field value at the end of inflation can be determined by setting $\epsilon = 1$; this leads to the following equation for χ_{end}

$$\frac{1}{2}\frac{M_P^2}{f^2}\frac{\sin^2\left(\chi_{end}\right)}{\left[1+\cos\left(\chi_{end}\right)\right]^2} = 1. \tag{6.4}$$

The solution is

$$\cos(\chi_{end}) = \frac{-1+b}{1+b}, \qquad b \equiv \left(\frac{M_P}{\sqrt{2}f}\right)^2, \tag{6.5}$$

The number of e-folds in Eq. (6.2) can be written as

$$N_k = \left(\frac{f}{M_P}\right)^2 \ln\left[\frac{1-\cos(\chi_{end})}{1-\cos(\chi_{in})}\right] = \left(\frac{f}{M_P}\right)^2 \ln\left[\frac{2}{(1+b)}\frac{1}{(1-\cos(\chi_{in}))}\right]. \tag{6.6}$$

The value of the field at the pivot scale during inflation is then given by

$$\cos(\chi_{in}) = 1 - z, \qquad z \equiv \frac{2}{(1+b)}\exp\left[-N_k\left(\frac{M_P}{f}\right)^2\right]. \tag{6.7}$$

Using (6.3) and (6.7), one finds

$$n_s - 1 \equiv -6\epsilon + 2\eta = -\left(\frac{M_P}{f}\right)^2\left(\frac{2+z}{2-z}\right), \tag{6.8}$$

which leads to

$$N_k = -\left(\frac{f}{M_P}\right)^2 \ln\left[\left(1+\frac{M_P^2}{2f^2}\right)\left(\frac{(1-n_s)-\frac{M_P^2}{f^2}}{(1-n_s)+\frac{M_P^2}{f^2}}\right)\right]. \tag{6.9}$$

Notice that the previous expression is positive and real only if the argument of the logarithm is defined between zero and one

$$0 < \left(1+\frac{M_P^2}{2f^2}\right)\left(\frac{(1-n_s)-\frac{M_P^2}{f^2}}{(1-n_s)+\frac{M_P^2}{f^2}}\right) < 1. \tag{6.10}$$

The conditions (6.10) are equivalent to requiring that

$$\left(\frac{f}{M_P}\right)^2 > \frac{1}{(1-n_s)} \qquad \text{and} \qquad 3 + n_s + \left(\frac{M_P}{f}\right)^2 > 0. \tag{6.11}$$

The second condition in (6.11) is always true. The first condition implies a minimum f for each n_s, and the bound on f increases with increasing n_s. Using the central value for the Planck constraints on the spectral index, then (6.11) gives $f > 5.6\, M_P$.

The tensor-to-scalar ratio can be expressed in terms of n_s

$$r = 16\,\epsilon = 4\left[(1 - n_s) - \frac{M_P^2}{f^2}\right]. \tag{6.12}$$

Using $V_{in} = \Lambda^4\,[1 + \cos(\chi_{in})] \simeq 3H^2 M_P^2$, one finds

$$V_{end} = 3H^2 M_P^2 \left[\frac{1 + (1 - n_s)\left(\frac{f}{M_P}\right)^2}{2 + 4\left(\frac{f}{M_P}\right)^2}\right]. \tag{6.13}$$

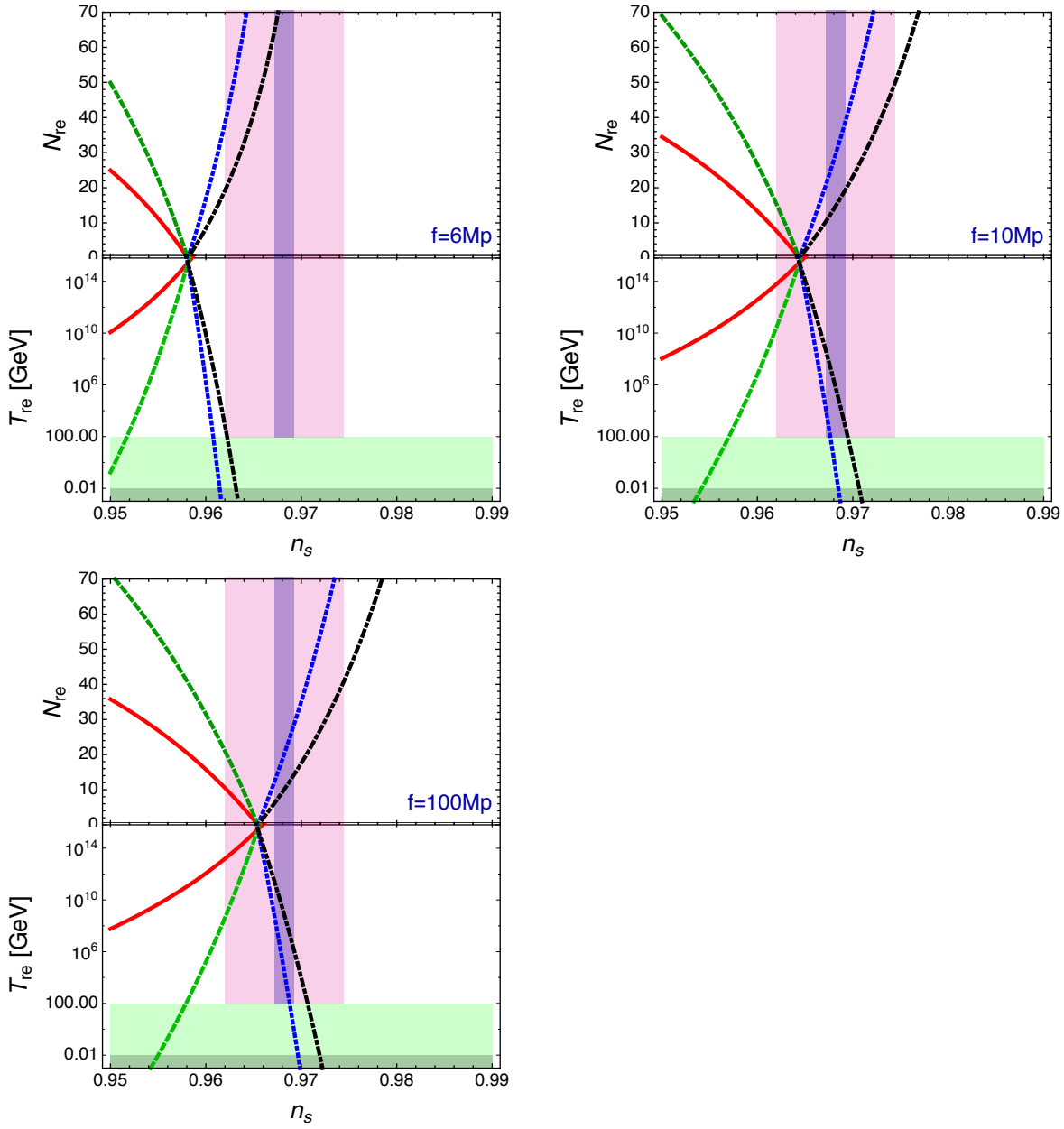

Figure 8: We show N_{re} and T_{re}, the length of reheating and the temperature at the end of reheating respectively, for natural inflation, for 3 values of the coupling f. Again, shading is as in Fig. 2

Fig. 8 shows N_{re} and T_{re} solutions for various reheating parameters w_{re} and for various couplings f in natural inflation. Unlike polynomial inflation, or Starobinsky/ Higgs inflation, natural inflation has an extra free parameter, and so one no longer gets a precise prediction for the temperature and length of reheating once a reheating model, w_{re}, and n_s are specified. But one can get reasonable bounds on the coupling f such that a viable reheating model exists.

One can obtain separate, and stronger constraints on f based on the requirement for

viable reheating. These constraints likewise are functions of n_s, and their effects are displayed in Figure 9[8].

There is no upper limit on f. For $f \gtrsim 14 M_P$ the various w_{re} lines reach an asymptotic form. As a result even for very large f, there is a valid solution for each w_{re} value consistent with Planck's 1σ bounds. The asymptotic solution for large f for $0 \leq w_{re} \leq 1/3$ corresponds to a solution for the spectral index in the range $0.956 \leq n_s \leq 0.965$.

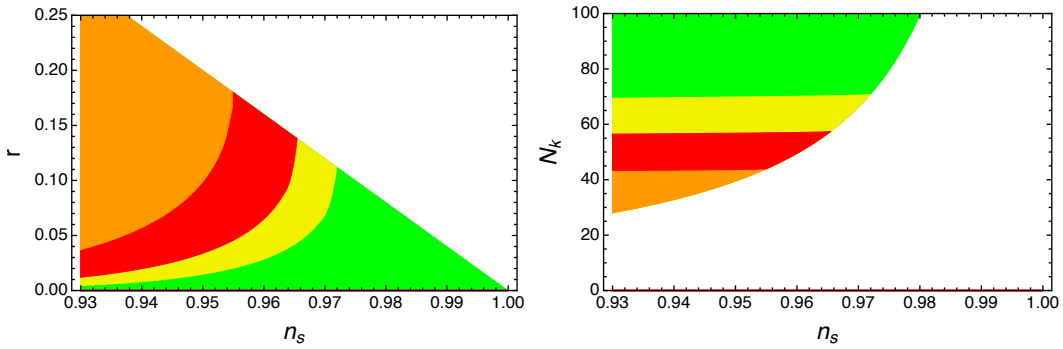

Figure 9: Representation of the r and N_k vs. n_s plane for natural inflation. The colored region is the entire allowed parameter space that can produce the measured A_s value at the pivot scale, and the n_s value at the pivot scale as plotted. Shading again follows Fig. 3

Note the minimum on f increases with increasing n_s, such that for larger n_s, a larger f is needed to find a solution consistent with the reheating model being considered. The top limit in Fig. 9 A (the bottom limit in Fig. 9 B) is approached asymptotically for large f. The bottom part of the parameter space in Fig. 9 A (top part in Fig. 9 B) corresponds to small f. Everywhere in the figure $N_k > 19$, such that inflation lasts long enough to allow for BBN.

Using Planck's 2σ bounds on n_s, requiring $w_{re} \leq 1$ gives $r \geq 0.02$, and requiring $w_{re} \leq 1/3$ gives $r \geq 0.05$. For $w_{re} \leq 1/3$, values of n_s smaller than Planck's central value would be favored for any f. The weakest constraint is for large values f, for which $n_s \leq 0.965$.

7 Hilltop inflation

The potential is given by [4, 97]

$$V(\phi) = M^4 \left[1 - \left(\frac{\phi}{\mu}\right)^p\right]. \tag{7.1}$$

We begin by considering $p > 2$. The exact expression for the number of e-foldings between the time the pivot scale crossed outside the horizon and the end of inflation is

$$N_k = \frac{\mu^2}{2pM_P^2}\left[\chi_{in}^2 - \chi_{end}^2 + \frac{2}{p-2}\chi_{in}^{2-p} - \frac{2}{p-2}\chi_{end}^{2-p}\right]. \tag{7.2}$$

[8] See also [11, 82] for n_s vs r plots in the natural inflation model for $w_{re} = 0$.

where one defines $\chi \equiv \phi/\mu$.
The slow-roll parameters are

$$\epsilon = \frac{p^2}{2}\frac{M_P^2}{\mu^2}\frac{\chi^{2(p-1)}}{(1-\chi^p)^2}\,, \qquad \eta \equiv -p(p-1)\frac{M_P^2}{\mu^2}\frac{\chi^{p-2}}{(1-\chi^p)}. \tag{7.3}$$

Setting $\epsilon = 1$ at the end of inflation one derives the equation for χ_{end}

$$\frac{p^2}{2}\frac{M_P^2}{\mu^2}\frac{\chi_{end}^{2(p-1)}}{\left(1-\chi_{end}^p\right)^2} = 1. \tag{7.4}$$

Let us consider the case where $\mu > M_P$ and define $q \equiv M_P/\mu$. For small values of q, one can search for a solution for χ_{end} in the form of a Taylor expansion around $q = 0$

$$\chi_{end} = a_0 + a_1\, q + \frac{1}{2}\, a_2\, q^2 + \mathcal{O}(q^3). \tag{7.5}$$

One can show that, up to order q^2, a solution to Eq. (7.4) is (see e.g. [11])

$$\chi_{end} = 1 - \frac{1}{\sqrt{2}}\, q + \frac{(p-1)}{4}\, q^2. \tag{7.6}$$

Similarly, one can look for a solution for the initial value of the scalar field using (7.6) and (7.2), to find

$$\chi_{in} = 1 - \sqrt{\frac{1+4\,N_k}{2}}\, q + \mathcal{O}(q^2). \tag{7.7}$$

Using (7.6) in the expression for the slow-roll parameters, (7.3), the spectral index as a function of N_k is

$$n_s - 1 \simeq -\frac{6}{1+4N_k}, \qquad \longrightarrow \qquad N_k \simeq \frac{1}{4}\left(\frac{6}{1-n_s} - 1\right). \tag{7.8}$$

The tensor-to-scalar ratio is

$$r \simeq \frac{8}{3}(1-n_s). \tag{7.9}$$

Notice that (7.8) and(7.9) only apply for small values of q, more precisely for

$$q < \frac{\sqrt{2}}{(p-1)\sqrt{1+4\,N_k}}. \tag{7.10}$$

For $p \in (3, 8)$ and for $N_k \in (30, 100)$, the previous condition is satisfied if $q \leq 0.01$.
Within the same range of validity, the potential at the end of inflation is given by

$$V_{end} \simeq \sqrt{\frac{3}{2}} H^2 M_P^2 \sqrt{1-n_s}. \tag{7.11}$$

For $p \leq 2$ one derives the same results as Eqs. (7.8)-(7.9) and (7.11). For $p = 2$, however, a new expression for N_k is required. In this case we find

$$N_k = \frac{\mu^2}{2M_P^2}\left[\frac{\chi_{in}^2}{2} - \frac{\chi_{end}^2}{2} - \ln\chi_{in} + \ln\chi_{end}\right]. \tag{7.12}$$

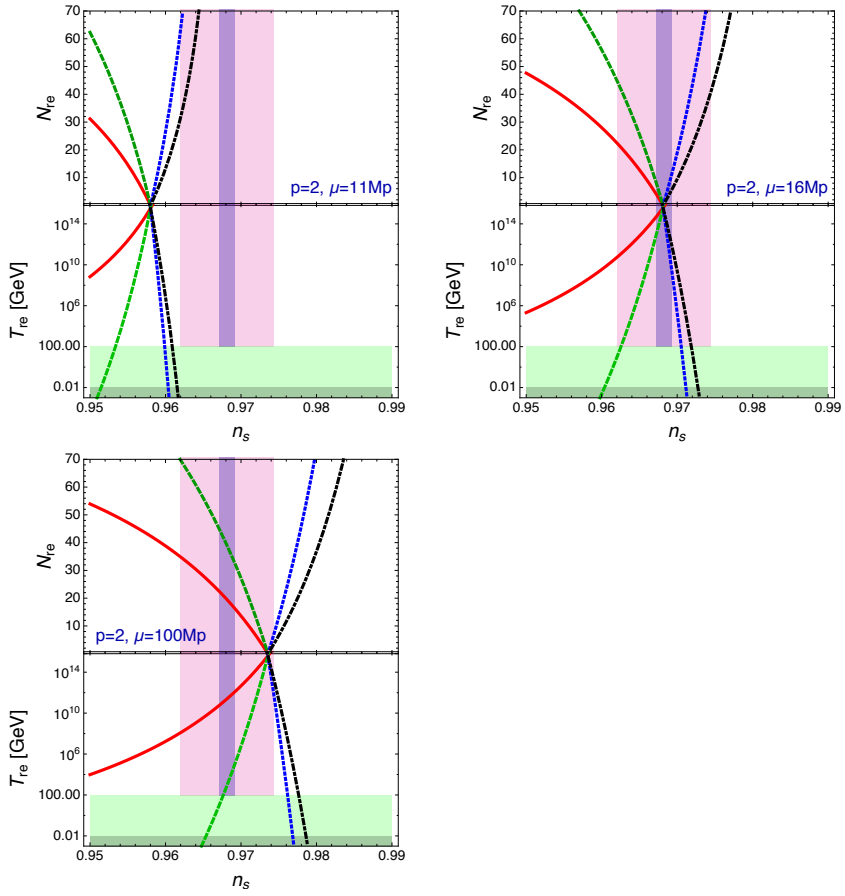

Figure 10: Plots of N_{re} and T_{re}, for hilltop inflation with $p = 2$ and for three different values of μ. Shading is as in Fig. 2

The plots for hilltop inflation are derived using a numerical procedure and therefore they convey more information than one would obtain with the above analytic results since they cover a range in which the latter would not apply (i.e. for smaller values of μ)[9].

Note that for $p = 1$ the potential is just a straight line, and so should give the same predictions as the $V \propto \phi$ inflation model considered above. In Figs. 10-12 we therefore plot N_{re} and T_{re} for various reheating scenarios parametrized by w_{re}, and various values of μ for $p = 2$ and larger.

Just as with natural inflation, hilltop inflation has two free parameters, in this case M and μ. This extra freedom means for each different p value, there are μ values that are readily consistent with Planck data and μ values that are not. One can give bounds on μ for each p model such that reasonable reheating solutions exist, and these results are shown in Figure 13 and 14.

As with the bound on f for natural inflation, there is a minimum on μ required for $p = 2$ to get any solution at all, even before reheating is considered, and that bound is a function

[9]Notice that hilltop inflation was previously studied in the context of reheating in [11, 77, 82].

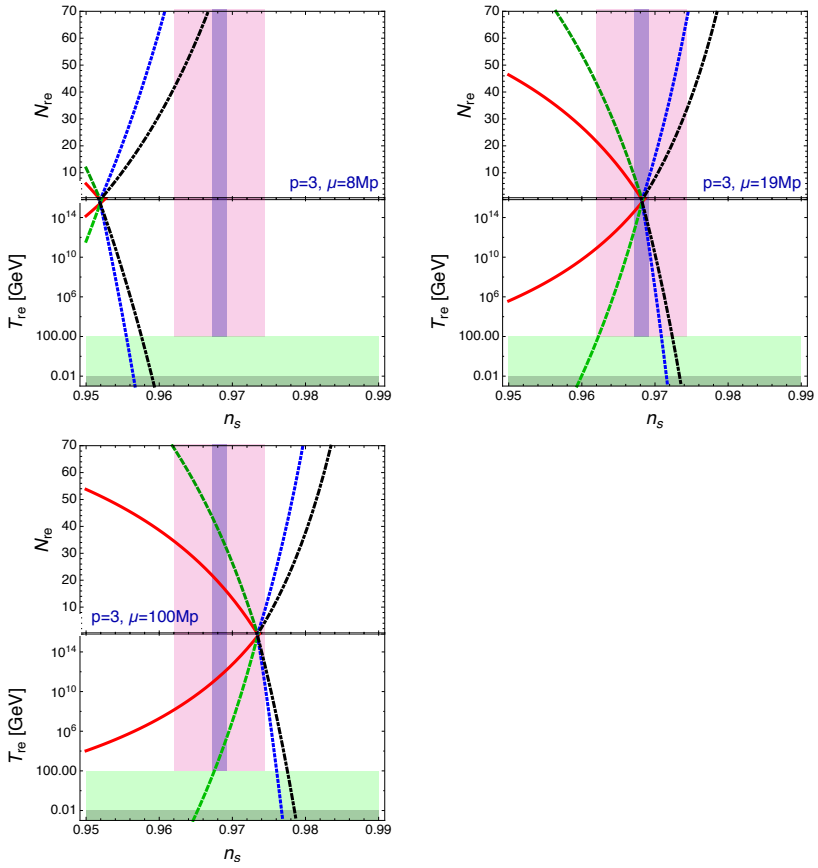

Figure 11: Plots of N_{re} and T_{re}, for hilltop inflation with $p = 3$ and for three different values of μ. Shading is as in Fig. 2

of n_s. This μ_{min} corresponds to $r \to 0$, so in this case there is no minimum on r before reheating is considered. For $p = 3$, 4 there is no such minimum on μ to get a solution in the regime $\mu \geq M_P$. There appears to be no observational constraint from Planck for very large values of μ.

The upper bounds in the plots in Figure 13 (the lower bounds in Figure 14) correspond to larger μ, and the lower bounds in Figure 13 (or the upper bounds in Figure 14) correspond to smaller μ. Using the 2σ bounds on n_s and requiring $w_{re} \leq 1/3$ gives the following lower bounds on the tensor-to-scalar ratio: $r \geq 0.02$ (for $p = 2$) and $r \geq 0.007$ ($p = 3$) and $r \geq 0.003$ ($p = 4$). Using the central value of n_s and requiring $w_{re} \leq 1/3$ gives the bounds: $r \geq 0.03$ (for $p = 2$ and $p = 3$) and $r \geq 0.02$ ($p = 4$). The region in parameter space that is associated with these more likely values of w_{re} then allows for fairly small r values.

8 Discusion and Conclusions

Inflation includes a wide variety of models that give similar predictions for the fairly small number of available inflationary observables. The physics of reheating can provide an ad-

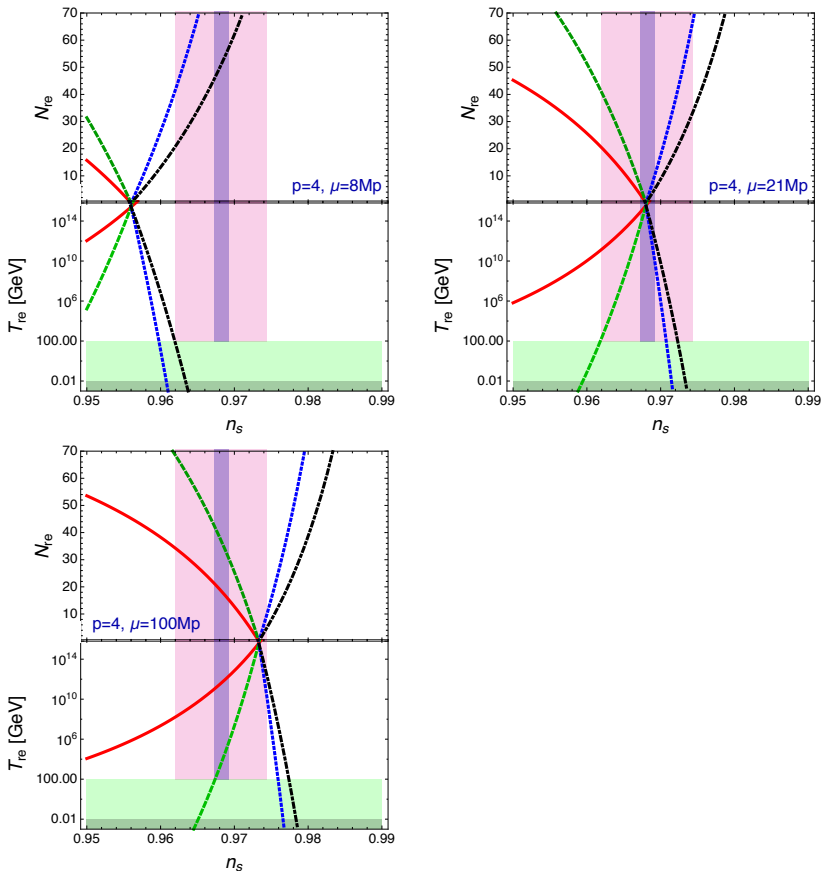

Figure 12: Plots of N_{re} and T_{re}, for hilltop inflation with $p = 4$ and for three different values of μ. Shading is as in Fig. 2

ditional opportunity to break this degeneracy. While CMB fluctuations themselves do not supply direct probes of the physics during the reheating era, the details of reheating affect the predictions for inflation (and vice versa) because they determine the nature of the cosmic thermal history after inflation (see e.g., [75–82]). Although we do not know exactly what occurred during reheating, we can make reasonable assumptions such as that the average equation of state during reheating was very likely between 0 and 1/3. This leads to independent constraints on observables like n_s and r, that can then be tested against CMB data.

One can parametrize our ignorance about reheating in terms of an equation of state, w_{re}, a length N_{re} (measured in terms of number of e-folds elapsed from the end of inflation), and a final temperature, T_{re}. For any given inflationary model, one can write relations between the specific model parameters, the amplitude of the scalar power spectrum A_s, the spectral index n_s, and the reheating parameters (w_{re}, N_{re} and T_{re}). These relations are derived by accounting for the total expansion history between the time the observable CMB modes crossed outside the Hubble radius during inflation and the time of observation, and employing a continuity equation for the energy density during the different cosmological epochs. We also assume that w_{re} is constant. For single-field models the derivation is particularly straightforward. The main results are summarized in Eq. (2.11) for T_{re} and w_{re} (and in Eq. (2.10) for N_{re} and w_{re}) as a function of inflationary parameters and observables.

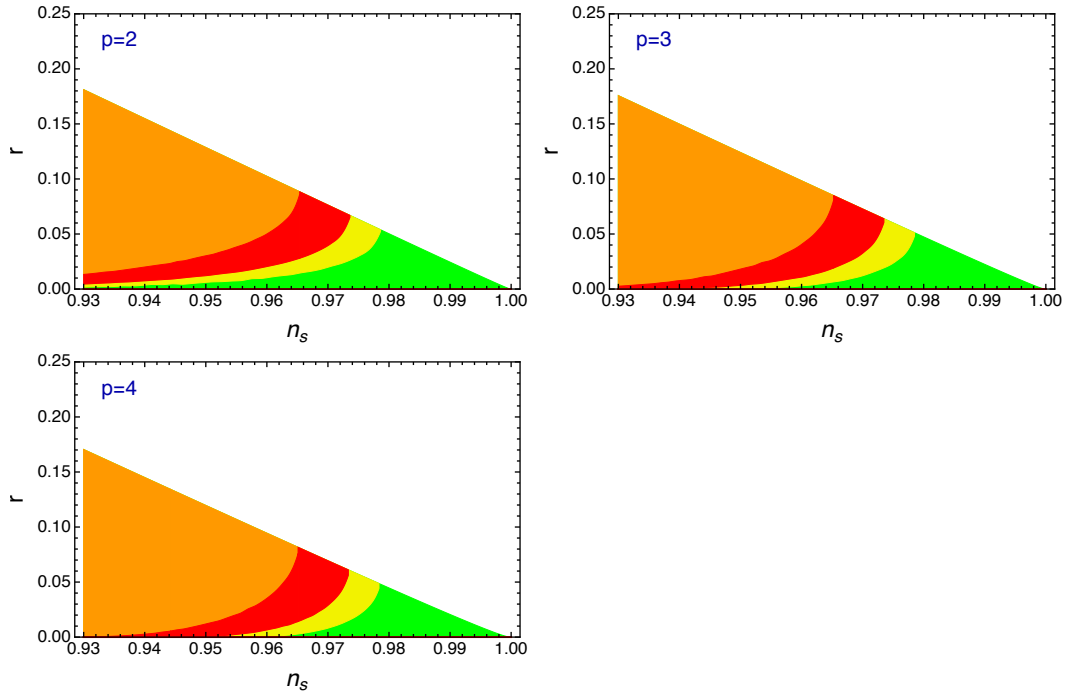

Figure 13: Plot of the parameter space in the r vs. n_s plane for the three hilltop models with $p = 2$, 3, and 4. Shading is as in Fig. 3

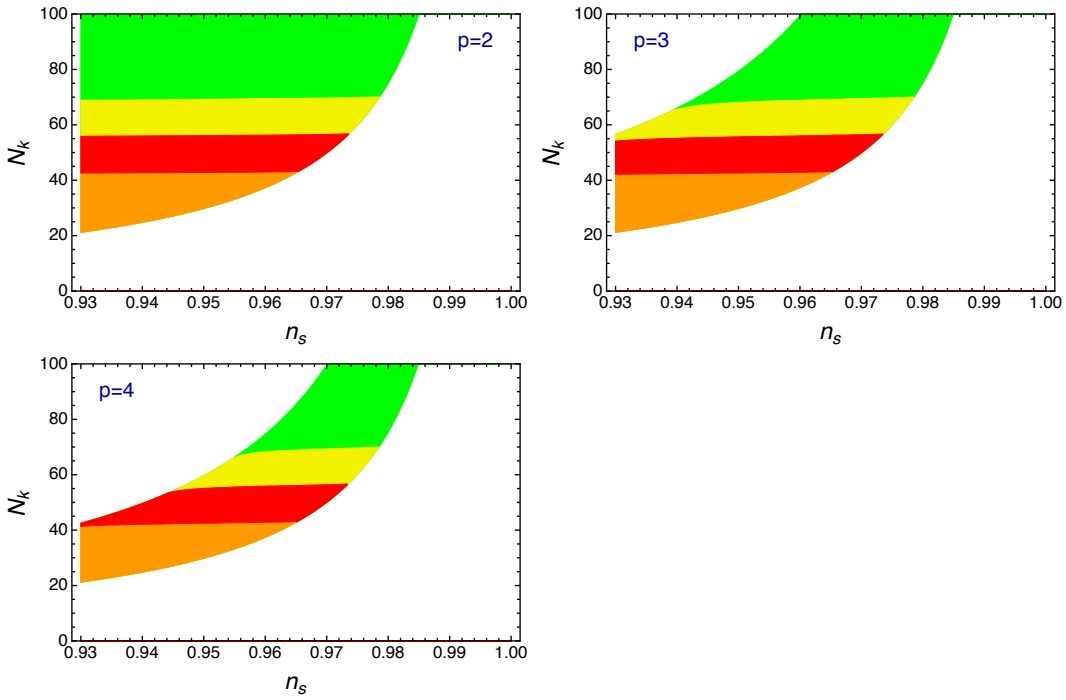

Figure 14: Plot of the parameter space in the N_k vs. n_s plane for the three hilltop models with $p = 2$, 3, and 4. Shading is as in Fig. 3

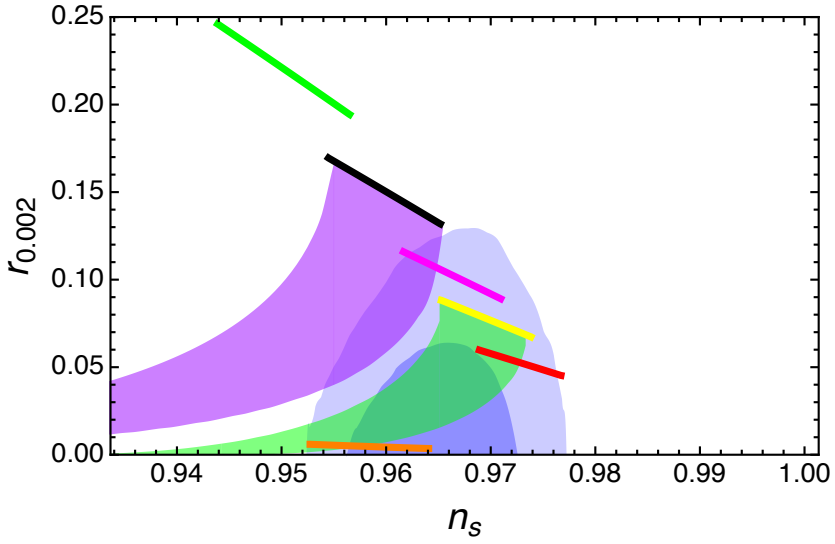

Figure 15: We recreate a version of the Planck Figure 12 in their recent Constraints on Inflation paper, using their 1 and 2 σ TT, TE, EE + lowP constraints on n_s and r [13], but plotting the parameter space for models such that there exists a reheating solution for $0 \leq w_{re} \leq \frac{1}{3}$, as opposed to Planck's choice of parameter space for which there is a solution with N_k between 50 and 60. Following the conventions in Planck's version of the plot, the green line is ϕ^3, the black is ϕ^2, the pink is $\phi^{4/3}$, the yellow ϕ, the red $\phi^{2/3}$, the orange Starobinsky/ Higgs model, the puple region is natural inflation, and the green region is the quartic hilltop model.

We consider a broad range for the equation of state parameter, $-1/3 \leq w_{re} \leq 1$, and the corresponding limits on CMB observables for different inflationary models. We notice that a ϕ^2 potential would favor relatively large values of r: a reheating model with $w_{re} \leq 1$ implies $r \geq 0.11$; to allow for a reheating model with $w_{re} \leq 1/3$ which is very probable, requires $r \geq 0.14$. Since it appears BICEP2's signal is dust instead of primordial gravitational waves, it is difficult to reconcile ϕ^2 inflation with the data. We also consider Starobinsky/Higgs inflation, natural inflation and the hilltop models. For Starobinsky and Higgs inflation, requiring $w_{re} \leq 1/3$ corresponds to $r \geq 0.004$. Because natural and hilltop inflation models have two free parameters, there are ranges of parameter space that can fit well the data for any value of the reheating parameter w_{re}. For natural inflation, we find that Planck's 2σ bound on n_s favors a tensor-to-scalar ratio $r \geq 0.05$ for $w_{re} \leq 1/3$ (Fig. 9). For the same range of w_{re}, the hilltop model, on the other hand, allows for smaller r values, specifically $r \geq 0.02$ for $p = 2$, $r \geq 0.007$ for $p = 3$, or $r \geq 0.003$ for $p = 4$ (Fig. 13).

We show this parameter space in Fig. (15), where we recreate a version of the Planck Figure 12 in their recent Constraints on Inflation paper, using their 1 and 2 σ TT, TE, EE + lowP constraints on n_s and r [13]. To get their model parameter space they impose N_k between 50 and 60. Instead here, we don't specify N_k but plot the parameter space for which there exists a reheating solution with $0 \leq w_{re} \leq \frac{1}{3}$. Constraining models in this way, using w_{re}, is a nice model dependent but straightforward and well motivated way of representing the parameter space.

To conclude we find that considering broad, well-motivated physical constraints on the reheating equation of state indeed allows one to narrow the viable parameter space for inflation models, offering an improvement over merely specifying whether or not an inflation model can reproduce the correct predictions at the pivot scale. These methods will become increasingly effective with future more precise CMB data.

Note added:
Our analysis was initially performed considering the Planck 2013 results for the scalar power spectrum parameters. Just after completion, but before submission, Planck released their 2015 data, so we have updated our analysis using the new observational bounds on n_s and A_s. All presented results are now based on the Planck 2015 data. While completing the first version of this work, it was brought to our attention that a similar approach was carried out by [98]. Some of the results on natural inflation were reproduced in [98] and, where there is overlap, we find agreement if we consider the Planck 2013 bounds on the scalar power spectrum parameters. Furthermore, when this work was near completion, two papers concerning Higgs inflation and reheating were released [99, 100]. We find agreement with the results of [99] if we consider their pivot scale ($k_p = 0.002 Mpc^{-1}$ as opposed to $0.05 Mpc^{-1}$).

Acknowledgments

It is a pleasure to thank M. Liguori and J. Wang for useful correspondence. This research was supported in part by a grant from the DOE.

References

[1] A. A. Starobinsky, *A New Type of Isotropic Cosmological Models Without Singularity*, *Phys.Lett.* **B91** (1980) 99–102, [doi:10.1016/0370-2693(80)90670-X].

[2] A. H. Guth, *The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems*, *Phys.Rev.* **D23** (1981) 347–356, [doi:10.1103/PhysRevD.23.347].

[3] K. Sato, *First Order Phase Transition of a Vacuum and Expansion of the Universe*, *Mon.Not.Roy.Astron.Soc.* **195** (1981) 467–479.

[4] A. D. Linde, *A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems*, *Phys.Lett.* **B108** (1982) 389–393, [doi:10.1016/0370-2693(82)91219-9].

[5] A. Albrecht and P. J. Steinhardt, *Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking*, *Phys.Rev.Lett.* **48** (1982) 1220–1223, [doi:10.1103/PhysRevLett.48.1220].

[6] A. D. Linde, *Chaotic Inflation*, *Phys.Lett.* **B129** (1983) 177–181, [doi:10.1016/0370-2693(83)90837-7].

[7] D. H. Lyth and A. Riotto, *Particle physics models of inflation and the cosmological density perturbation*, *Phys.Rept.* **314** (1999) 1–146, [arXiv:hep-ph/9807278], [doi:10.1016/S0370-1573(98)00128-8].

[8] A. Riotto, *Inflation and the theory of cosmological perturbations*, arXiv:hep-ph/0210162.

[9] W. H. Kinney, *Cosmology, inflation, and the physics of nothing*, *NATO Sci.Ser.II* **123** (2003) 189–243, [arXiv:astro-ph/0301448], [doi:10.1007/978-94-010-0076-5-5].

[10] D. Baumann, *TASI Lectures on Inflation*, arXiv:0907.5424.

[11] J. Martin, C. Ringeval, and V. Vennin, *Encyclopdia Inflationaris*, *Phys.Dark Univ.* (2014) [arXiv:1303.3787], [doi:10.1016/j.dark.2014.01.003].

[12] **Planck Collaboration** Collaboration, *Planck 2015 results. XIII. Cosmological parameters*, arXiv:1502.01589.

[13] **Planck Collaboration** Collaboration, P. Ade *et al.*, *Planck 2015. XX. Constraints on inflation*, arXiv:1502.02114.

[14] V. F. Mukhanov and G. V. Chibisov, *Quantum Fluctuation and Nonsingular Universe. (In Russian)*, *JETP Lett.* **33** (1981) 532–535.

[15] A. A. Starobinsky, *Dynamics of Phase Transition in the New Inflationary Universe Scenario and Generation of Perturbations*, *Phys.Lett.* **B117** (1982) 175–178, [doi:10.1016/0370-2693(82)90541-X].

[16] A. H. Guth and S. Pi, *Fluctuations in the New Inflationary Universe*, *Phys.Rev.Lett.* **49** (1982) 1110–1113, [doi:10.1103/PhysRevLett.49.1110].

[17] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, *Spontaneous Creation of Almost Scale - Free Density Perturbations in an Inflationary Universe*, *Phys.Rev.* **D28** (1983) 679, [doi:10.1103/PhysRevD.28.679].

[18] L. Abbott and M. B. Wise, *Constraints on Generalized Inflationary Cosmologies*, *Nucl.Phys.* **B244** (1984) 541–548, [doi:10.1016/0550-3213(84)90329-8].

[19] L. Abbott and M. B. Wise, *Large Scale Anisotropy of the Microwave Background and the Amplitude of Energy Density Fluctuations in the Early Universe*, *Astrophys.J.* **282** (1984) L47–L50, [doi:10.1086/184302].

[20] A. A. Starobinsky, *Spectrum of relict gravitational radiation and the early state of the universe*, *JETP Lett.* **30** (1979) 682–685.

[21] V. Rubakov, M. Sazhin, and A. Veryaskin, *Graviton Creation in the Inflationary Universe and the Grand Unification Scale*, *Phys.Lett.* **B115** (1982) 189–192, [doi:10.1016/0370-2693(82)90641-4].

[22] L. M. Krauss and M. J. White, *Grand unification, gravitational waves, and the cosmic microwave background anisotropy*, *Phys.Rev.Lett.* **69** (1992) 869–872, [arXiv:hep-ph/9205212], [doi:10.1103/PhysRevLett.69.869].

[23] L. M. Krauss and F. Wilczek, *Using Cosmology to Establish the Quantization of Gravity*, *Phys.Rev.* **D89** (2014), no. 4 047501, [arXiv:1309.5343], [doi:10.1103/PhysRevD.89.047501].

[24] L. Abbott, E. Farhi, and M. B. Wise, *Particle Production in the New Inflationary Cosmology*, *Phys.Lett.* **B117** (1982) 29, [doi:10.1016/0370-2693(82)90867-X].

[25] A. Dolgov and A. D. Linde, *Baryon Asymmetry in Inflationary Universe*, *Phys.Lett.* **B116** (1982) 329, [doi:10.1016/0370-2693(82)90292-1].

[26] A. Albrecht, P. J. Steinhardt, M. S. Turner, and F. Wilczek, *Reheating an Inflationary Universe*, *Phys.Rev.Lett.* **48** (1982) 1437, [doi:10.1103/PhysRevLett.48.1437].

[27] L. Kofman, A. D. Linde, and A. A. Starobinsky, *Reheating after inflation*, *Phys.Rev.Lett.* **73** (1994) 3195–3198, [arXiv:hep-th/9405187], [doi:10.1103/PhysRevLett.73.3195].

[28] J. H. Traschen and R. H. Brandenberger, *Particle Production During Out-of-equilibrium Phase Transitions*, *Phys.Rev.* **D42** (1990) 2491–2504, [doi:10.1103/PhysRevD.42.2491].

[29] L. Kofman, A. D. Linde, and A. A. Starobinsky, *Towards the theory of reheating after inflation*, *Phys.Rev.* **D56** (1997) 3258–3295, [arXiv:hep-ph/9704452], [doi:10.1103/PhysRevD.56.3258].

[30] B. R. Greene, T. Prokopec, and T. G. Roos, *Inflaton decay and heavy particle production with negative coupling*, *Phys.Rev.* **D56** (1997) 6484–6507, [arXiv:hep-ph/9705357], [doi:10.1103/PhysRevD.56.6484].

[31] N. Shuhmaher and R. Brandenberger, *Non-perturbative instabilities as a solution of the cosmological moduli problem*, *Phys.Rev.* **D73** (2006) 043519, [arXiv:hep-th/0507103], [doi:10.1103/PhysRevD.73.043519].

[32] J. F. Dufaux, G. N. Felder, L. Kofman, M. Peloso, and D. Podolsky, *Preheating with trilinear interactions: Tachyonic resonance*, *JCAP* **0607** (2006) 006, [arXiv:hep-ph/0602144], [doi:10.1088/1475-7516/2006/07/006].

[33] A. A. Abolhasani, H. Firouzjahi, and M. Sheikh-Jabbari, *Tachyonic Resonance Preheating in Expanding Universe*, *Phys.Rev.* **D81** (2010) 043524, [arXiv:0912.1021], [doi:10.1103/PhysRevD.81.043524].

[34] G. N. Felder, J. Garcia-Bellido, P. B. Greene, L. Kofman, A. D. Linde, *et al.*, *Dynamics of symmetry breaking and tachyonic preheating*, *Phys.Rev.Lett.* **87** (2001) 011601, [arXiv:hep-ph/0012142], [doi:10.1103/PhysRevLett.87.011601].

[35] G. N. Felder, L. Kofman, and A. D. Linde, *Tachyonic instability and dynamics of spontaneous symmetry breaking*, *Phys.Rev.* **D64** (2001) 123517, [arXiv:hep-th/0106179], [doi:10.1103/PhysRevD.64.123517].

[36] G. N. Felder, L. Kofman, and A. D. Linde, *Instant preheating*, *Phys.Rev.* **D59** (1999) 123523, [arXiv:hep-ph/9812289], [doi:10.1103/PhysRevD.59.123523].

[37] D. Boyanovsky, H. de Vega, and R. Holman, *Erice lectures on inflationary reheating*, arXiv:hep-ph/9701304.

[38] B. A. Bassett, S. Tsujikawa, and D. Wands, *Inflation dynamics and reheating*, *Rev.Mod.Phys.* **78** (2006) 537–589, [arXiv:astro-ph/0507632], [doi:10.1103/RevModPhys.78.537].

[39] R. Allahverdi, R. Brandenberger, F.-Y. Cyr-Racine, and A. Mazumdar, *Reheating in Inflationary Cosmology: Theory and Applications*, *Ann.Rev.Nucl.Part.Sci.* **60** (2010) 27–51, [arXiv:1001.2600], [doi:10.1146/annurev.nucl.012809.104511].

[40] M. A. Amin, M. P. Hertzberg, D. I. Kaiser, and J. Karouby, *Nonperturbative Dynamics Of Reheating After Inflation: A Review*, *Int.J.Mod.Phys.* **D24** (2015) 1530003, [arXiv:1410.3808], [doi:10.1142/S0218271815300037].

[41] M. Drewes and J. U. Kang, *The Kinematics of Cosmic Reheating*, *Nucl.Phys.* **B875** (2013) 315–350, [arXiv:1305.0267], [doi:10.1016/j.nuclphysb.2013.07.009, 10.1016/j.nuclphysb.2014.09.008].

[42] M. Drewes, *On finite density effects on cosmic reheating and moduli decay and implications for Dark Matter production*, *JCAP* **1411** (2014), no. 11 020, [arXiv:1406.6243], [doi:10.1088/1475-7516/2014/11/020].

[43] D. I. Podolsky, G. N. Felder, L. Kofman, and M. Peloso, *Equation of state and beginning of thermalization after preheating*, *Phys.Rev.* **D73** (2006) 023501, [arXiv:hep-ph/0507096], [doi:10.1103/PhysRevD.73.023501].

[44] J. Berges, S. Borsanyi, and C. Wetterich, *Prethermalization*, *Phys.Rev.Lett.* **93** (2004) 142002, [arXiv:hep-ph/0403234], [doi:10.1103/PhysRevLett.93.142002].

[45] A. Taruya and Y. Nambu, *Cosmological perturbation with two scalar fields in reheating after inflation*, *Phys.Lett.* **B428** (1998) 37–43, [arXiv:gr-qc/9709035], [doi:10.1016/S0370-2693(98)00378-5].

[46] B. A. Bassett and F. Viniegra, *Massless metric preheating*, *Phys.Rev.* **D62** (2000) 043507, [arXiv:hep-ph/9909353], [doi:10.1103/PhysRevD.62.043507].

[47] F. Finelli and R. H. Brandenberger, *Parametric amplification of metric fluctuations during reheating in two field models*, *Phys.Rev.* **D62** (2000) 083502, [arXiv:hep-ph/0003172], [doi:10.1103/PhysRevD.62.083502].

[48] S. Tsujikawa and B. A. Bassett, *When can preheating affect the CMB?*, *Phys.Lett.* **B536** (2002) 9–17, [arXiv:astro-ph/0204031], [doi:10.1016/S0370-2693(02)01813-0].

[49] A. Chambers and A. Rajantie, *Lattice calculation of non-Gaussianity from preheating*, *Phys.Rev.Lett.* **100** (2008) 041302, [arXiv:0710.4133], [doi:10.1103/PhysRevLett.100.041302].

[50] J. R. Bond, A. V. Frolov, Z. Huang, and L. Kofman, *Non-Gaussian Spikes from Chaotic Billiards in Inflation Preheating*, *Phys.Rev.Lett.* **103** (2009) 071301, [arXiv:0903.3407], [doi:10.1103/PhysRevLett.103.071301].

[51] B. A. Bassett, D. I. Kaiser, and R. Maartens, *General relativistic preheating after inflation*, *Phys.Lett.* **B455** (1999) 84–89, [arXiv:hep-ph/9808404], [doi:10.1016/S0370-2693(99)00478-5].

[52] B. A. Bassett, F. Tamburini, D. I. Kaiser, and R. Maartens, *Metric preheating and limitations of linearized gravity. 2.*, *Nucl.Phys.* **B561** (1999) 188–240, [arXiv:hep-ph/9901319], [doi:10.1016/S0550-3213(99)00495-2].

[53] B. A. Bassett, C. Gordon, R. Maartens, and D. I. Kaiser, *Restoring the sting to metric preheating*, *Phys.Rev.* **D61** (2000) 061302, [arXiv:hep-ph/9909482], [doi:10.1103/PhysRevD.61.061302].

[54] L. Bethke, D. G. Figueroa, and A. Rajantie, *Anisotropies in the Gravitational Wave Background from Preheating*, *Phys.Rev.Lett.* **111** (2013), no. 1 011301, [arXiv:1304.2657], [doi:10.1103/PhysRevLett.111.011301].

[55] R. Easther, R. Galvez, O. Ozsoy, and S. Watson, *Supersymmetry, Nonthermal Dark Matter and Precision Cosmology*, *Phys.Rev.* **D89** (2014), no. 2 023522, [arXiv:1307.2453], [doi:10.1103/PhysRevD.89.023522].

[56] H. B. Moghaddam, R. H. Brandenberger, Y.-F. Cai, and E. G. M. Ferreira, *Parametric Resonance of Entropy Perturbations in Massless Preheating*, arXiv:1409.1784.

[57] G. Steigman, *Primordial Nucleosynthesis in the Precision Cosmology Era*, *Ann.Rev.Nucl.Part.Sci.* **57** (2007) 463–491, [arXiv:0712.1100], [doi:10.1146/annurev.nucl.56.080805.140437].

[58] M. Kawasaki, K. Kohri, and N. Sugiyama, *Cosmological constraints on late time entropy production*, *Phys.Rev.Lett.* **82** (1999) 4168, [arXiv:astro-ph/9811437], [doi:10.1103/PhysRevLett.82.4168].

[59] **BICEP2 Collaboration** Collaboration, P. Ade *et al.*, *Detection of B-Mode Polarization at Degree Angular Scales by BICEP2*, *Phys.Rev.Lett.* **112** (2014) 241101, [arXiv:1403.3985], [doi:10.1103/PhysRevLett.112.241101].

[60] **BICEP2 Collaboration** Collaboration, P. Ade *et al.*, *BICEP2 II: Experiment and Three-Year Data Set*, *Astrophys.J.* **792** (2014) 62, [arXiv:1403.4302], [doi:10.1088/0004-637X/792/1/62].

[61] J. Garcia-Bellido, A. D. Linde, and D. Wands, *Density perturbations and black hole formation in hybrid inflation*, *Phys.Rev.* **D54** (1996) 6040–6058, [arXiv:astro-ph/9605094], [doi:10.1103/PhysRevD.54.6040].

[62] B. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, *New cosmological constraints on primordial black holes*, *Phys.Rev.* **D81** (2010) 104019, [arXiv:0912.5297], [doi:10.1103/PhysRevD.81.104019].

[63] E. Torres-Lomas, J. C. Hidalgo, K. A. Malik, and L. A. Urea-Lpez, *Formation of subhorizon black holes from preheating*, *Phys.Rev.* **D89** (2014) 083008, [arXiv:1401.6960], [doi:10.1103/PhysRevD.89.083008].

[64] E. A. Calzetta and A. Kandus, *Selfconsistent estimates of magnetic fields from reheating*, *Phys.Rev.* **D65** (2002) 063004, [arXiv:astro-ph/0110341], [doi:10.1103/PhysRevD.65.063004].

[65] A. Diaz-Gil, J. Garcia-Bellido, M. Garcia Perez, and A. Gonzalez-Arroyo, *Magnetic field production during preheating at the electroweak scale*, *Phys.Rev.Lett.* **100** (2008) 241301, [arXiv:0712.4263], [doi:10.1103/PhysRevLett.100.241301].

[66] A. Diaz-Gil, J. Garcia-Bellido, M. G. Perez, and A. Gonzalez-Arroyo, *Primordial magnetic fields from preheating at the electroweak scale*, *JHEP* **0807** (2008) 043, [arXiv:0805.4159], [doi:10.1088/1126-6708/2008/07/043].

[67] G. Giudice, I. Tkachev, and A. Riotto, *Nonthermal production of dangerous relics in the early universe*, *JHEP* **9908** (1999) 009, [arXiv:hep-ph/9907510], [doi:10.1088/1126-6708/1999/08/009].

[68] G. Giudice, A. Riotto, and I. Tkachev, *The Cosmological moduli problem and preheating*, *JHEP* **0106** (2001) 020, [arXiv:hep-ph/0103248], [doi:10.1088/1126-6708/2001/06/020].

[69] G. Giudice, M. Peloso, A. Riotto, and I. Tkachev, *Production of massive fermions at preheating and leptogenesis*, *JHEP* **9908** (1999) 014, [arXiv:hep-ph/9905242], [doi:10.1088/1126-6708/1999/08/014].

[70] L. M. Krauss and M. Trodden, *Baryogenesis below the electroweak scale*, *Phys.Rev.Lett.* **83** (1999) 1502–1505, [arXiv:hep-ph/9902420], [doi:10.1103/PhysRevLett.83.1502].

[71] J. Garcia-Bellido, D. Y. Grigoriev, A. Kusenko, and M. E. Shaposhnikov, *Nonequilibrium electroweak baryogenesis from preheating after inflation*, *Phys.Rev.* **D60** (1999) 123504, [arXiv:hep-ph/9902449], [doi:10.1103/PhysRevD.60.123504].

[72] S. Davidson, M. Losada, and A. Riotto, *A New perspective on baryogenesis*, *Phys.Rev.Lett.* **84** (2000) 4284–4287, [arXiv:hep-ph/0001301], [doi:10.1103/PhysRevLett.84.4284].

[73] E. J. Copeland, D. Lyth, A. Rajantie, and M. Trodden, *Hybrid inflation and baryogenesis at the TeV scale*, *Phys.Rev.* **D64** (2001) 043506, [arXiv:hep-ph/0103231], [doi:10.1103/PhysRevD.64.043506].

[74] A. R. Liddle and S. M. Leach, *How long before the end of inflation were observable perturbations produced?*, *Phys.Rev.* **D68** (2003) 103503, [arXiv:astro-ph/0305263], [doi:10.1103/PhysRevD.68.103503].

[75] J. Martin and C. Ringeval, *Inflation after WMAP3: Confronting the Slow-Roll and Exact Power Spectra to CMB Data*, *JCAP* **0608** (2006) 009, [arXiv:astro-ph/0605367], [doi:10.1088/1475-7516/2006/08/009].

[76] L. Lorenz, J. Martin, and C. Ringeval, *Brane inflation and the WMAP data: A Bayesian analysis*, *JCAP* **0804** (2008) 001, [arXiv:0709.3758], [doi:10.1088/1475-7516/2008/04/001].

[77] J. Martin and C. Ringeval, *First CMB Constraints on the Inflationary Reheating Temperature*, *Phys.Rev.* **D82** (2010) 023511, [arXiv:1004.5525], [doi:10.1103/PhysRevD.82.023511].

[78] P. Adshead, R. Easther, J. Pritchard, and A. Loeb, *Inflation and the Scale Dependent Spectral Index: Prospects and Strategies*, *JCAP* **1102** (2011) 021, [arXiv:1007.3748], [doi:10.1088/1475-7516/2011/02/021].

[79] J. Mielczarek, *Reheating temperature from the CMB*, *Phys.Rev.* **D83** (2011) 023502, [arXiv:1009.2359], [doi:10.1103/PhysRevD.83.023502].

[80] R. Easther and H. V. Peiris, *Bayesian Analysis of Inflation II: Model Selection and Constraints on Reheating*, *Phys.Rev.* **D85** (2012) 103533, [arXiv:1112.0326], [doi:10.1103/PhysRevD.85.103533].

[81] L. Dai, M. Kamionkowski, and J. Wang, *Reheating constraints to inflationary models*, *Phys.Rev.Lett.* **113** (2014) 041302, [arXiv:1404.6704], [doi:10.1103/PhysRevLett.113.041302].

[82] J. Martin, C. Ringeval, and V. Vennin, *Observing the Inflationary Reheating*, arXiv:1410.7958.

[83] **Euclid Theory Working Group** Collaboration, L. Amendola *et al.*, *Cosmology and fundamental physics with the Euclid satellite*, *Living Rev.Rel.* **16** (2013) 6, [arXiv:1206.1225].

[84] **PRISM Collaboration** Collaboration, P. Andre *et al.*, *PRISM (Polarized Radiation Imaging and Spectroscopy Mission): A White Paper on the Ultimate Polarimetric Spectro-Imaging of the Microwave and Far-Infrared Sky*, arXiv:1306.2259.

[85] S. Bartrum, M. Bastero-Gil, A. Berera, R. Cerezo, R. O. Ramos, *et al.*, *The importance of being warm (during inflation)*, *Phys.Lett.* **B732** (2014) 116–121, [arXiv:1307.5868], [doi:10.1016/j.physletb.2014.03.029].

[86] M. Bastero-Gil, A. Berera, R. O. Ramos, and J. G. Rosa, *Observational implications of mattergenesis during inflation*, *JCAP* **1410** (2014), no. 10 053, [arXiv:1404.4976], [doi:10.1088/1475-7516/2014/10/053].

[87] **BICEP2 Collaboration, Planck Collaboration** Collaboration, P. Ade *et al.*, *A Joint Analysis of BICEP2/Keck Array and Planck Data*, *Phys.Rev.Lett.* (2015) [arXiv:1502.00612].

[88] C. Wetterich, *Cosmologies With Variable Newton's 'Constant'*, *Nucl.Phys.* **B302** (1988) 645, [doi:10.1016/0550-3213(88)90192-7].

[89] S. Kalara, N. Kaloper, and K. A. Olive, *Theories of Inflation and Conformal Transformations*, *Nucl.Phys.* **B341** (1990) 252–272, [doi:10.1016/0550-3213(90)90270-N].

[90] F. L. Bezrukov and M. Shaposhnikov, *The Standard Model Higgs boson as the inflaton*, *Phys.Lett.* **B659** (2008) 703–706, [arXiv:0710.3755], [doi:10.1016/j.physletb.2007.11.072].

[91] F. Bezrukov, D. Gorbunov, and M. Shaposhnikov, *On initial conditions for the Hot Big Bang*, *JCAP* **0906** (2009) 029, [arXiv:0812.3622], [doi:10.1088/1475-7516/2009/06/029].

[92] J. Garcia-Bellido, D. G. Figueroa, and J. Rubio, *Preheating in the Standard Model with the Higgs-Inflaton coupled to gravity*, *Phys.Rev.* **D79** (2009) 063531, [arXiv:0812.4624], [doi:10.1103/PhysRevD.79.063531].

[93] F. Bezrukov and D. Gorbunov, *Distinguishing between R^2-inflation and Higgs-inflation*, *Phys.Lett.* **B713** (2012) 365–368, [arXiv:1111.4397], [doi:10.1016/j.physletb.2012.06.040].

[94] D. Gorbunov and A. Tokareva, *R^2-inflation with conformal SM Higgs field*, *JCAP* **1312** (2013) 021, [arXiv:1212.4466], [doi:10.1088/1475-7516/2013/12/021].

[95] A. Kehagias, A. M. Dizgah, and A. Riotto, *Remarks on the Starobinsky model of inflation and its descendants*, *Phys.Rev.* **D89** (2014), no. 4 043527, [arXiv:1312.1155], [doi:10.1103/PhysRevD.89.043527].

[96] K. Freese, J. A. Frieman, and A. V. Olinto, *Natural inflation with pseudo - Nambu-Goldstone bosons*, *Phys.Rev.Lett.* **65** (1990) 3233–3236, [doi:10.1103/PhysRevLett.65.3233].

[97] A. D. Linde, *PRIMORDIAL INFLATION WITHOUT PRIMORDIAL MONOPOLES*, *Phys.Lett.* **B132** (1983) 317–320, [doi:10.1016/0370-2693(83)90316-7].

[98] J. B. Munoz and M. Kamionkowski, *The Equation-of-State Parameter for Reheating*, arXiv:1412.0656.

[99] R.-G. Cai, Z.-K. Guo, and S.-J. Wang, *Reheating Phase Diagram for Higgs Inflation*, arXiv:1501.07743.

[100] J.-O. Gong, S. Pi, and G. Leung, *Probing reheating with primordial spectrum*, arXiv:1501.03604.

Prepared for submission to JCAP

DARWIN: towards the ultimate dark matter detector

DARWIN collaboration

J. Aalbers,[a] F. Agostini,[b,n] M. Alfonsi,[c] F.D. Amaro,[r] C. Amsler,[d] E. Aprile,[e] L. Arazi,[f] F. Arneodo,[g] P. Barrow,[h] L. Baudis,[h] M.L. Benabderrahmane,[g] T. Berger, [i] B. Beskers,[c] A. Breskin,[f] P.A. Breur,[a] A. Brown,[a] E. Brown, [i] S. Bruenner,[j] G. Bruno,[n] R. Budnik,[f] L. Bütikofer,[d] J. Calvén,[k] J.M.R. Cardoso,[r] D. Cichon,[j] D. Coderre,[d] A.P. Colijn,[a] J. Conrad,[k] J.P. Cussonneau,[l] M.P. Decowski,[a] S. Diglio,[l] G. Drexlin,[m] E. Duchovni,[f] E. Erdal,[f] G. Eurin,[j] A. Ferella,[k] A. Fieguth,[w] W. Fulgione,[n] A. Gallo Rosso,[n] P. Di Gangi,[b] A. Di Giovanni,[g] M. Galloway,[h] M. Garbini,[b] C. Geis,[c] F. Glueck,[m] L. Grandi,[o] Z. Greene,[e] C. Grignon,[c] C. Hasterok,[j] V. Hannen,[w] E. Hogenbirk,[a] J. Howlett,[e] D. Hilk,[m] C. Hils,[c] A. James,[h] B. Kaminsky,[d] S. Kazama,[h] B. Kilminster,[h] A. Kish,[h] L.M. Krauss,[p] H. Landsman,[f] R.F. Lang,[q] Q. Lin,[e] F.L. Linde,[a] S. Lindemann,[j] M. Lindner,[j] J.A.M. Lopes,[r] T. Marrodán Undagoitia,[j] J. Masbou,[l] F.V. Massoli,[b] D. Mayani,[h] M. Messina,[e] K. Micheneau,[l] A. Molinario,[n] K.D. Morå,[k] E. Morteau,[l] M. Murra,[w] J. Naganoma,[t] J.L. Newstead,[p] K. Ni,[s] U. Oberlack,[c] P. Pakarha,[h] B. Pelssers,[k] P. de Perio,[e] R. Persiani,[l] F. Piastra,[h] M.C. Piro, [i] G. Plante,[e] L. Rauch,[j] S. Reichard,[q] A. Rizzo,[e] N. Rupp,[j] J.M.F. Dos Santos,[r] G. Sartorelli,[b] M. Scheibelhut,[c] S. Schindler,[c] M. Schumann,[d] J. Schreiner,[j] L. Scotto Lavina,[l] M. Selvi,[b] P. Shagin,[t] M.C. Silva,[r] H. Simgen,[j] P. Sissol,[c] M. von Sivers,[d] D. Thers,[l] J. Thurn,[x] A. Tiseni,[a] R. Trotta,[u] C.D. Tunnell,[a] K. Valerius,[m] M.A. Vargas,[w] H. Wang,[v] Y. Wei,[h] C. Weinheimer,[w] T. Wester,[x] J. Wulf,[h] Y. Zhang,[e] T. Zhu,[e] K. Zuber[x]

[a]Nikhef and the University of Amsterdam, Netherlands

[b]Department of Physics and Astrophysics, University of Bologna and INFN-Bologna, Italy

[c]Institut für Physik & Exzellenzcluster PRISMA, Johannes Gutenberg-Universität Mainz, Germany
[d]Albert Einstein Center for Fundamental Physics, Universität Bern, Switzerland
[e]Physics Department, Columbia University, New York, NY, USA
[f]Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot, Israel
[g]New York University Abu Dhabi, United Arab Emirates
[h]Physik-Institut, Universität Zürich, Switzerland
[i]Department of Physics, Applied Physics and Astronomy, Rensselaer Polytechnic Institute, Troy, NY, USA
[j]Max-Planck-Institut für Kernphysik, Heidelberg, Germany
[k]Department of Physics, Stockholm University, Sweden
[l]Subatech, Ecole des Mines de Nantes, CNRS/In2p3, Université de Nantes, France
[m]Institut für Experimentelle Kernphysik, Karlsruhe Institute of Technology (KIT), Germany
[n]INFN-Laboratori Nazionali del Gran Sasso and Gran Sasso Science Institute, L'Aquila, Italy
[o]Kavli Institute, Enrico Fermi Institute and Dept. of Physics, University of Chicago, IL, USA
[p]Physics Department, Arizona State University, Tempe, AZ, USA
[q]Department of Physics and Astronomy, Purdue University, West Lafayette, IN, USA
[r]Department of Physics, University of Coimbra, Portugal
[s]Department of Physics, University of California, San Diego, CA, USA
[t]Department of Physics and Astronomy, Rice University, Houston, TX, USA
[u]Astrophysics Group & Data Science Institute, Imperial College London, UK
[v]Physics & Astronomy Department, University of California, Los Angeles, CA, USA
[w]Institut für Kernphysik, Westfälische Wilhelms-Universität Münster, Germany
[x]Institute for Nuclear and Particle Physics, TU Dresden, Germany
E-mail: lior.arazi@weizmann.ac.il, laura.baudis@uzh.ch, amos.breskin@weizmann.ac.il, decowski@nikhef.nl, marc.schumann@lhep.unibe.ch

Abstract. DARk matter WImp search with liquid xenoN (DARWIN[1]) will be an experiment for the direct detection of dark matter using a multi-ton liquid xenon time projection chamber at its core. Its primary goal will be to explore the experimentally accessible parameter space for Weakly Interacting Massive Particles (WIMPs) in a wide mass-range, until neutrino interactions with the target become an irreducible background. The prompt scintillation light and the charge signals induced by particle interactions in the xenon will be observed by VUV sensitive, ultra-low background photosensors. Besides its excellent sensitivity to WIMPs above a mass of $5\,\mathrm{GeV}/c^2$, such a detector with its large mass, low-energy threshold and ultra-low background level will also be sensitive to other rare interactions. It will search for solar axions, galactic axion-like particles and the neutrinoless double-beta decay of ^{136}Xe, as well as measure the low-energy solar neutrino flux with $<1\%$ precision, observe coherent neutrino-nucleus interactions, and detect galactic supernovae. We present the concept of the DARWIN detector and discuss its physics reach, the main sources of backgrounds and the ongoing detector design and R&D efforts.

Keywords: Direct dark matter detection, WIMPs, Neutrinos, Large detector systems for particle and astroparticle physics, Time projection chambers, Liquid xenon detector, Axions, Neutrinoless double beta decay, Supernovae

[1]www.darwin-observatory.org

Contents

1 Introduction

Astronomical and cosmological observations reveal that the vast majority of the matter and energy content of our universe is invisible – or dark – and interacts neither strongly nor electromagnetically with ordinary matter. Results from the Planck satellite [1] show that about 68% of the overall budget is dark energy, leading to the observed accelerated expansion of the cosmos. Another 27% is composed of dark matter, a yet-undetected form of matter whose presence is needed to explain the observed large-scale structures and galaxies. While dark matter interacts gravitationally with baryonic matter, any additional interactions, if existing, must be very weak with extremely small cross sections [2]. Because the standard model of particle physics does not accommodate dark matter, the observationally-driven need

for its existence is one of the strongest indications for physics beyond the standard model. The direct detection and subsequent characterisation of dark matter particles is, therefore, one of the major experimental challenges of modern particle and astroparticle physics [3, 4].

Many theories beyond the standard model predict viable candidates; one particular class, receiving the attention of most current and planned experiments, is that of Weakly Interacting Massive Particles (WIMPs) [3, 5]. Worldwide, more than a dozen experiments are prepared to observe low-energy nuclear recoils induced by galactic WIMPs in ultra-sensitive, low-background detectors [6–9]. Since the predicted WIMP masses and scattering cross sections are model-dependent and essentially unknown, these searches must cover a vast parameter space [10, 11]. Most promising are detectors based on liquefied noble gas targets such as liquid xenon (LXe) or liquid argon (LAr). This technology is by now well-established and can be scaled up to ton-scale, homogeneous target masses [12–14], taking data over several years.

Two detector concepts are in use. The first uses a single-phase noble-liquid WIMP target, surrounded by photosensors to record the emitted scintillation light. Examples are the XMASS detector, operating a 850 kg total LXe target [15], as well as DEAP-3600 [16] and miniCLEAN [17], large LAr detectors currently under commissioning. The LAr instruments employ the powerful rejection of electronic recoil background based on pulse shape discrimination (PSD) [18]. With a 3600 kg LAr target, the larger detector DEAP-3600 aims at a sensitivity of $\sim 1 \times 10^{-46}\,\mathrm{cm}^2$ for spin-independent WIMP-nucleon interactions at a WIMP mass of $100\,\mathrm{GeV/c}^2$.

The second concept is based on dual-phase noble gas time projection chambers (TPCs), where the prompt scintillation light (S1) and the delayed proportional scintillation light signal from the charge (S2) are measured. Both signals are employed for a precise reconstruction of the event vertex and, thus, to suppress backgrounds by rejection of multiple-scatter interactions, as WIMPs are expected to interact only once. The charge-to-light ratio, S2/S1, is exploited to separate the expected signal, namely nuclear recoils (NR), from the dominant electronic recoil (ER) background.

TPCs filled with LXe were pioneered by the ZEPLIN [19, 20] and XENON10 [21, 22] collaborations. The XENON100 experiment [23, 24], a TPC with a 62 kg active target, has reached its sensitivity goal and excluded spin-independent WIMP-nucleon cross sections above $2 \times 10^{-45}\,\mathrm{cm}^2$ at a WIMP mass of $55\,\mathrm{GeV}/c^2$ [25]. These constraints were superseded by the results from the LUX collaboration [26, 27], which operates a 250 kg TPC and excludes spin-independent WIMP-nucleon scattering cross sections above $4 \times 10^{-46}\,\mathrm{cm}^2$ at $33\,\mathrm{GeV/c}^2$. The second phase of PandaX has published first result from its run with a 500 kg active LXe target [28]. Liquid argon dual-phase TPCs were pioneered by WArP [29] and ArDM [30]. In addition to the S2/S1 discrimination, they also exploit the considerably more powerful PSD rejection of ER background [18]. DarkSide-50, using 50 kg of active LAr mass, has presented first results from a low radioactivity run [31]. The experiment reduces its target radioactivity by using underground argon in which the radioactive ^{39}Ar is depleted by a factor of $1.4{\times}10^3$ with respect to atmospheric argon.

Probing lower cross sections at WIMP masses above a few GeV/c^2 requires larger detectors. XENON1T, the current phase in the XENON collaboration programme, aims to reach spin-independent cross sections of $1.6 \times 10^{-47}\,\mathrm{cm}^2$ after 2 years of continuous operation of its 2 t LXe target [32]. The next phase, XENONnT, to be designed and constructed during XENON1T operation, will increase the sensitivity by another order of magnitude, assuming 20 t×y exposure [32]. A similar sensitivity is sought by LUX-ZEPLIN (LZ), the next phase

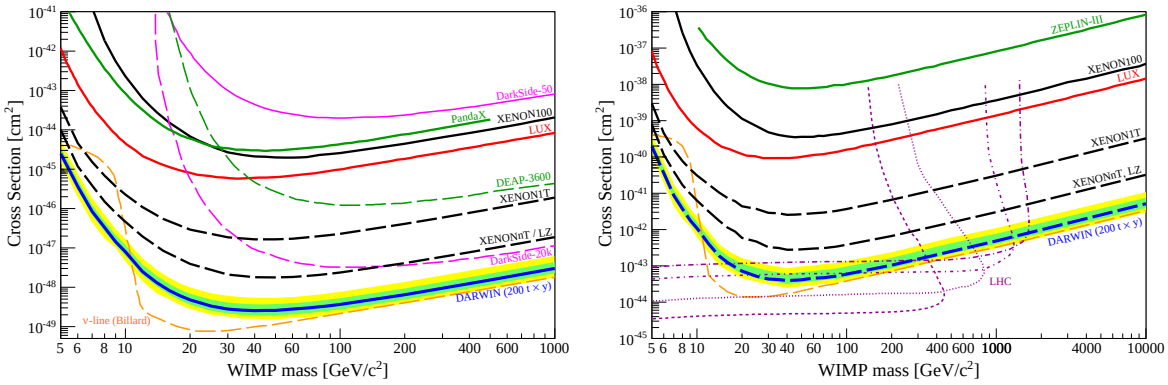

Figure 1. **(left)** The attainable bound on spin-independent WIMP nucleon cross-section as a function of the WIMP mass of present and future nobel liquid detectors. Shown are upper limits from PandaX-II [28], DarkSide-50 [31], XENON100 [25], and LUX [27] as well as the sensitivity projections for DEAP3600 [16], XENON1T [32], XENONnT [32], LZ [33], DarkSide-20k [34] and DARWIN [35], for which we also show the 1-σ (yellow band) and 2-σ (green band) regions. DARWIN is designed to probe the entire parameter region for WIMP masses above $\sim$5 GeV/c^2, until the neutrino background (ν-line, dashed orange [36]) will start to dominate the recoil spectrum. **(right)** Upper limits on the spin-dependent WIMP-*neutron* cross section of ZEPLIN-III [19], XENON100 [37] and LUX [38] as well as projections for XENON1T, XENONnT, LZ and DARWIN. DARWIN and the high-luminosity LHC will cover a common region of the parameter space. The 14 TeV LHC limits for the coupling constants $g_\chi = g_q = 0.25$, 0.5, 1.0, 1.45 (bottom to top) are taken from [39]. The LHC reach for spin-independent couplings is above 10^{-41} cm^2. Argon has no stable isotopes with non-zero nuclear spins and is thus not sensitive to spin-dependent couplings. Figures updated from [35], using the xenon response model to nuclear recoils from [27] for the DARWIN sensitivity.

in the LUX programme, which plans to operate a 7 t LXe detector with an additional scintillator veto to suppress the neutron background [33]. The DarkSide collaboration proposes a 20 t LAr dual-phase detector, with the goal to reach 9×10^{-48}cm^2 at 1 TeV/c^2, based on extrapolations of the demonstrated PSD efficiency of the smaller detector [34, 40].

The DARk matter WImp search with liquid xenoN (DARWIN) observatory, which is the subject of this article, aims at a $\sim$10-fold increase in sensitivity compared to these projects. Figure 1 (left) summarises the status and expected sensitivities to spin-independent WIMP-nucleon interactions as a function of the WIMP mass for noble liquid detectors including DARWIN. Figure 1 (right) shows the situation for the spin-dependent case, assuming WIMP coupling to neutrons only. DARWIN and the high-luminosity LHC will cover common parameter space [39] in this channel.

This article is structured as follows: After a brief introduction to the DARWIN project in Section 2, we discuss its reach for several astroparticle and particle physics science channels in Section 3. DARWIN's main background sources are introduced in Section 4, followed by a detailed discussion on design considerations and the status of the ongoing R&D towards the ultimate WIMP dark matter detector in Section 5.

2 The DARWIN project

DARWIN will be an experiment using a multi-ton liquid xenon TPC, with the primary goal to explore the experimentally accessible parameter space for WIMPs. DARWIN's 50 t total

(40 t active) LXe target will probe particles with masses above 5 GeV/c^2, and WIMP-nucleon cross sections down to the few $\times 10^{-49}$ cm^2 region for masses of $\sim$50 GeV/c^2 [35]. Should dark matter particles be discovered by existing or near-future experiments, DARWIN will measure WIMP-induced nuclear recoil spectra with high statistics and constrain the mass and the scattering cross section of the dark matter particle [41, 42]. Other physics goals are the real-time detection of solar *pp*-neutrinos with high statistics, detection of coherent neutrino-nucleus interactions, searches for solar axions and galactic axion-like particles (ALPs) and the search for the neutrinoless double beta decay ($0\nu\beta\beta$). The latter would establish whether the neutrino is its own anti-particle, and can be detected via the double beta emitter ^{136}Xe, which has a natural abundance of 8.9% in xenon. The facility will also be able to observe neutrinos of all flavours from supernova explosions [43], providing complementary information to large-scale water-Cherenkov or LAr detectors. DARWIN is included in the European Roadmap for Astroparticle Physics (APPEC) and additional national roadmaps.

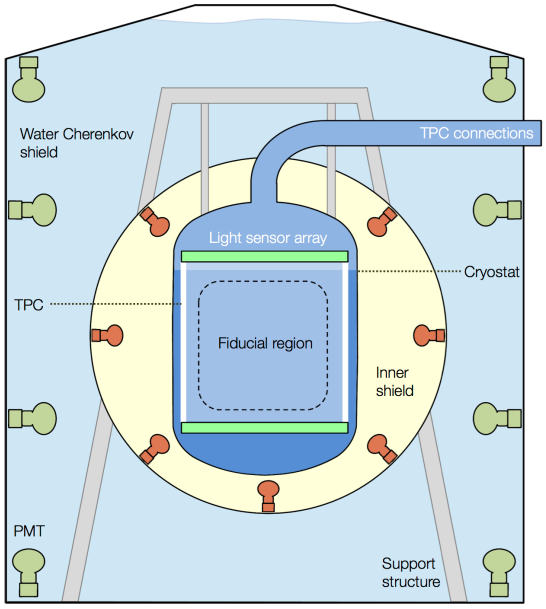

Figure 2. Sketch of the DARWIN detector inside a tank, operated as a water-Cherenkov muon veto. The need for an additional liquid-scintillator neutron veto inside the water shield, as shown in the figure ('inner shield'), is subject to further studies. The dual-phase time projection chamber is enclosed within a double-walled cryostat and contains 40 t of liquid xenon (50 t total in the cryostat). In the baseline scenario, the prompt and delayed VUV scintillation signals, induced by particle interactions in the sensitive volume, are recorded by two arrays of photosensors installed above and below the liquid xenon target.

The experiment will operate a large volume of liquid xenon in a low-background cryostat, surrounded by concentric shielding structures, as shown schematically in Figure 2. Future studies will reveal whether a liquid scintillator detector inside the water Cherenkov shield is required for this massive detector. The core of the experiment is a dual-phase TPC containing the active xenon mass. The high density of liquid xenon, $\sim$3 kg/l, results in a short radiation length and allows for a compact detector geometry with efficient self-shielding. The fiducial target mass is not fixed *a priori*: it will be defined during the analysis process, based on the relevant backgrounds and on the studied physics case. A drift field of $\mathcal{O}(0.5)$ kV/cm across the liquid target is required to drift the electrons from the interaction vertex. This will be achieved by biasing the cathode at the bottom of the TPC with voltages on the order of -100 kV or above. Large field shaping rings made from oxygen-free high conductivity (OFHC) copper, optimised for such high voltages, will ensure the field homogeneity. The main materials to be used for the TPC construction are OFHC copper as a conductor and polytetrafluoroethylene (PTFE) as an insulator, with the latter also acting as an efficient reflector for vacuum ultra-violet (VUV) scintillation light [44]. The TPC will be housed in a double-walled cryostat made out of stainless steel, titanium or copper, and all the materials

are to be selected for ultra-low intrinsic radioactivity. The structure will be suspended from a support frame which will allow for the levelling of the TPC with ∼100 μm precision once the outer shields and the detector are filled with liquids.

In the baseline scenario, the prompt and proportional scintillation signals will be recorded by two arrays of photosensors installed above and below the target. The photosensors could be future versions (3" or 4" in diameter) of the photomultiplier tubes (PMTs) employed in XENON1T (Hamamatsu R11410-21). These sensors feature a very low intrinsic radioactivity, high quantum efficiency (QE) at 178 nm, high gain and low dark count rate at low temperatures [45–47]. However, albeit a proven and reliable technology, PMTs are bulky, expensive and generate a significant fraction of the radioactive background in a dark matter detector, especially in terms of radiogenic nuclear recoils [35]. Thus, several alternative light readout schemes are under consideration. In addition, new challenges may arise in scaling up the 'traditional' dual-phase scheme to the multi-ton regime. To meet these potential challenges the DARWIN R&D programme further incorporates feasibility studies of other, non-traditional, light and charge readout concepts and novel TPC configurations, as outlined in Section 5.

3 Science channels

This section outlines the science capabilities of the DARWIN facility. Due to its low energy threshold, ultra-low ER and NR-induced background and large target volume, DARWIN will not only be sensitive to WIMP dark matter, but also to a wide variety of other rare-event searches including solar *pp*-neutrinos, supernova neutrinos, coherent neutrino scattering, axions and axion-like-particles, and neutrinoless double beta decay [48]. Section 3.1 summarises the WIMP dark matter sensitivity reach, which we have investigated under various assumptions [35, 42]. The additional physics channels are described in Section 3.2.

3.1 WIMP dark matter

The primary purpose of DARWIN is to investigate dark matter interactions and a large part of our activity is focused on optimising the sensitivity for WIMP dark matter. We have shown in a recent study [35] that one can exploit the full discovery potential of this technique with a 40 t LXe TPC (50 t total, and 30 t in the fiducial target), considering all known backgrounds listed in Section 4. These include backgrounds from detector construction materials (γ-radiation, neutrons), β-decays of ^{85}Kr (0.1 ppt of $^{\mathrm{nat}}$Kr) and the progeny of ^{222}Rn (0.1 μBq/kg) in the liquid target, two-neutrino double beta-decays ($2\nu\beta\beta$) of ^{136}Xe, electronic recoil interactions from low energy solar neutrinos (*pp*, ^{7}Be), as well as higher energy neutrino interactions with xenon nuclei in coherent neutrino-nucleus scattering (CNNS). Under these assumptions and with an exposure of 200 t×y, we find that a spin-independent WIMP sensitivity of 2.5×10^{-49} cm^2 can be reached at a WIMP mass $m_\chi = 40\,\mathrm{GeV}/c^2$, as shown in Figure 1 (left) on page 3. Increasing the exposure to 500 t×y improves this sensitivity to $\sim 1.5 \times 10^{-49}$ cm^2, under identical assumptions.

Natural xenon includes two isotopes with non-zero total nuclear angular momentum, ^{129}Xe and ^{131}Xe, at a combined abundance of ∼50%. For spin-dependent WIMP-neutron couplings and WIMP masses up to $\sim 1\,\mathrm{TeV}/c^2$, the searches that can be conducted by DARWIN will be complementary to those of the future high-luminosity LHC, at 14 TeV center-of-mass energy [39], as shown in Figure 1 (right). If the WIMP-nucleus interaction is indeed spin-dependent, DARWIN can also probe inelastic scattering, where the ^{129}Xe and ^{131}Xe nuclei are excited into low-lying states, with subsequent prompt deexcitation [49].

The projected sensitivity critically depends on the ability to discriminate NR signals from ER background, as the background from low-energetic solar neutrinos cannot be mitigated by other methods. Our study assumes an ER rejection level of 99.98% at 30% nuclear recoil acceptance, which is a factor 5 above the one of XENON100 [25] and has already been achieved by ZEPLIN-III [50]. Crucial requirements for reaching this rejection level include a uniform and high light yield for S1 and an S2 signal detection with uniform electron extraction and gas amplification. The statistical fluctuations in the S1 signal close to threshold significantly affect the width of the electronic and nuclear recoil distributions. Uniformity in S1 and S2 signal detection minimises any instrument-related fluctuations affecting the width of the S2/S1 distributions and hence the ER rejection power. While an increased light yield will also reduce the energy threshold, the dominating CNNS background will render thresholds below 5 keV nuclear recoil energy (5 $\mathrm{keV_{nr}}$) less relevant for the WIMP search at spin-independent cross sections below $\sim 10^{-45}\mathrm{cm}^2$. A further consideration is that the steeply falling CNNS spectrum requires the best possible energy resolution also at threshold. An energy scale derived from the charge signal or from a combination of light and charge is therefore necessary to optimise the sensitivity, as discussed in Section 5.7.

We have studied the reconstruction of WIMP properties, namely mass and scattering cross section, from the measured recoil spectra. In a numerical model, we have incorporated realistic detector parameters, backgrounds and astrophysical uncertainties [42]. Our primary study was directed towards spin-independent WIMP-nucleon interactions; however, given DARWIN's excellent sensitivity to spin-dependent interactions, especially for ^{129}Xe [51], it can be extended to axial vector couplings as well. Figure 3 (left) shows the reconstructed parameters for three hypothetical particle masses and a fixed cross section of $2 \times 10^{-47}\,\mathrm{cm}^2$, assuming an exposures of 200 t×y [42]. The corresponding number of events are 154, 224 and 60, for WIMP masses of 20 GeV/c^2, 100 GeV/c^2 and 500 GeV/c^2, respectively. Using the same exposure, Figure 3 (right) shows the reconstructed mass and cross section

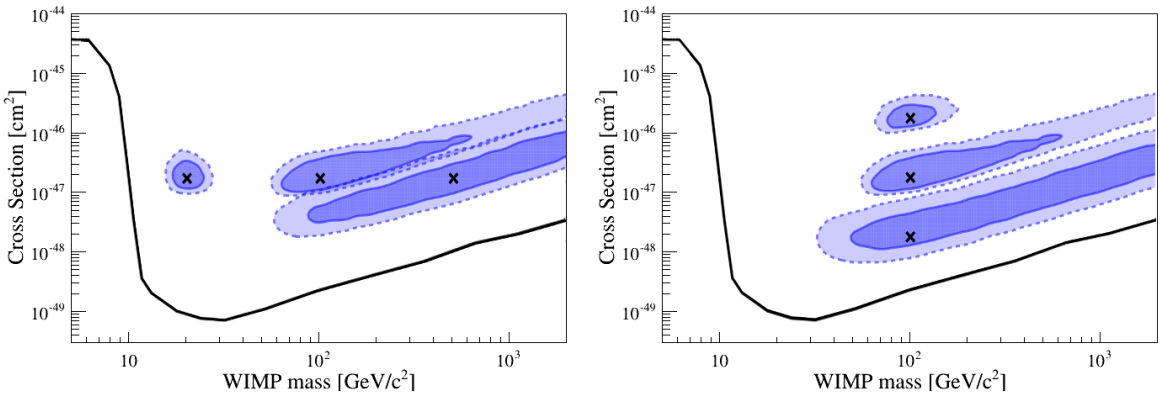

Figure 3. The 1 σ and 2 σ credible regions of the marginal posterior probabilities for simulations of WIMP signals assuming various masses and spin-independent (scalar) cross sections with DARWIN's LXe target. The width and length of these contours demonstrate how well the WIMP parameters can be reconstructed in DARWIN after a 200 t×y exposure. The '×' indicate the simulated benchmark models. **(left)** Reconstruction for three different WIMP masses of 20 GeV/c^2, 100 GeV/c^2 and 500 GeV/c^2 and a cross section of $2 \times 10^{-47}\,\mathrm{cm}^2$, close to the sensitivity limit of XENON1T. **(right)** Reconstruction for cross sections of $2 \times 10^{-46}\,\mathrm{cm}^2$, $2 \times 10^{-47}\,\mathrm{cm}^2$ and $2 \times 10^{-48}\,\mathrm{cm}^2$ for a WIMP mass of 100 GeV/c^2. The black curve indicates where the WIMP sensitivity will start to be limited by neutrino-nucleus coherent scattering. Figure adapted from [42].

values for a 100 GeV/c^2 WIMP and several cross sections. The study applies a conservative nuclear recoil energy threshold of 6.6 $\mathrm{keV_{nr}}$, a Helm nuclear form factor [52] and a Maxwell-Boltzmann distribution for the WIMP speed. The uncertainties on the dark matter halo parameters ρ_0=(0.3±0.1) GeV cm^{-3}, v_0=(220±20) km s^{-1} and v_{esc}=(544±40) km s^{-1} have been marginalised over, and lead to extended regions in the mass-cross section parameter space [53, 54]. The parameter reconstructions were performed on a representative data set, where the number of observed events was equal to the expected number of events for the given WIMP and detector parameters. A real experiment would be subject to realisation noise, which would induce a shift in the reconstructed regions from the underlying ground truth, as quantified in [55].

The tightest constraints are obtained for WIMP masses up to a few hundred GeV/c^2. We also find that for masses ≥500 GeV/c^2 only lower limits on the WIMP mass can be derived due to the fact that the shape of the nuclear recoil spectra depends on the WIMP-nucleus reduced mass. Figure 3 shows that, even with a large exposure such as 200 t×y, a substantial uncertainty on the reconstruction of the WIMP properties remains, depending on the mass and cross section. The extraction of dark matter properties is complicated by the astrophysical uncertainties, in particular from the underlying phase space distribution in our Galactic halo and the local normalisation – both of which can induce systematic errors if not properly accounted for. This systematic bias can be converted into a more manageable statistical error by introducing a parametric astrophysical model and marginalising over its parameters [53].

3.2 Other rare event searches

In this section we describe several other searches for rare events which can be pursued by multi-ton liquid xenon experiments. Due to the expected low background of electronic recoils, DARWIN will be sensitive not only to WIMPs, but also to some additional, hypothetical particles which are expected to have non-vanishing couplings to electrons. It will also be able to detect solar neutrinos, which constitute part of the background for the WIMP search channel, and neutrinos from supernova explosions in the galaxy or the Magellanic Clouds.

3.2.1 Axions and axion-like particles

Galactic axions and axion-like particles (ALPs) are well-motivated dark matter candidates [4, 56]. Even if axions do not represent the majority of the dark matter in our Universe, they could still exist and be abundantly produced in the Sun. By exploiting the axio-electric effect [57, 58], DARWIN can search for galactic and solar axions. In this process, axions couple to the electrons of the xenon atoms in the target and lead to atomic ionisation. They can thus be detected in the electronic recoil channel, down to energies of a few keV. The process is analogous to that of the photoelectric effect, and the expected signature would be a mono-energetic peak at the axion mass, spread only by the energy resolution of the detector.

XENON100 was the first to report results on solar axions and galactic ALPs using a dual-phase xenon detector [59]. It excluded axion-electron couplings $g_{Ae}^{\mathrm{ALP}} \gtrsim 2 \times 10^{-12}$ (90% CL) for galactic ALPs in a mass range of $1 < m_A < 40\,\mathrm{keV}/c^2$. In the case of solar axions, it excluded axion-electron couplings $g_{Ae}^{\mathrm{solar}} > 7.7 \times 10^{-12}$ (90% CL) in a mass range of $10^{-5} < m_A < 1\,\mathrm{keV}/c^2$, as shown in Figure 4.

The sensitivity of large liquid xenon detectors to axion signals was first studied in [65]. Extrapolating from these results, and assuming the total ER background estimated in [35], a similar energy threshold, a 30% superior energy resolution to XENON100, and an exposure

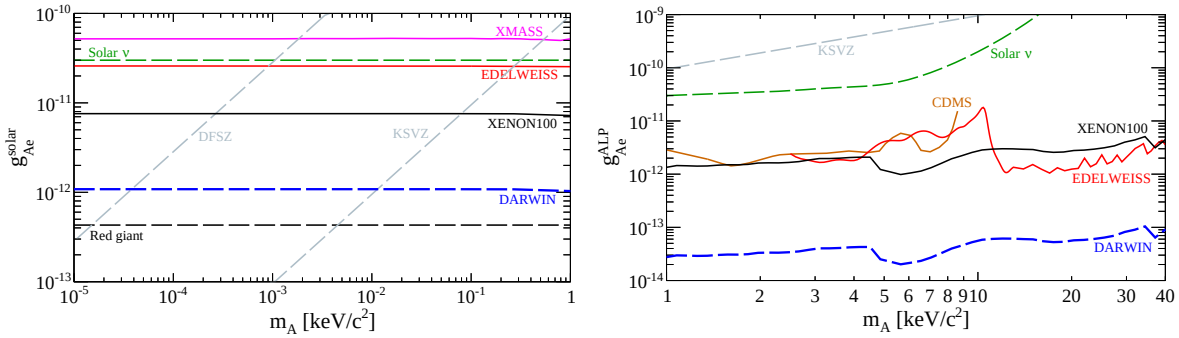

Figure 4. Sensitivity of DARWIN to solar axions **(left)** and axion-like-particles (ALPs) **(right)** which could constitute the entire galactic dark matter. While the increase in sensitivity compared to the XENON100 result [59] is only moderate for solar axions, due to the very weak dependence on the exposure ($x^{-1/8}$), DARWIN could improve the sensitivity to galactic ALPs by almost two order of magnitude in a 200 t × y exposure. Direct upper limits from the dark matter experiments XMASS [60], EDELWEISS [61] and CDMS [62], indirect limits from solar neutrinos and red giants, as well as two generic axion models, DFSZ [63] and KSVZ [64], are also shown.

of 200 t×y, we find that DARWIN could improve the sensitivity of XENON100 for galactic ALPs by almost two orders of magnitude, as shown in Figure 4, right. For solar axions, the sensitivity improvement will be more modest, equaling about one order of magnitude, see Figure 4, left. This is due to the rather weak dependence of the coupling on the exposure (target mass $M \times$ time T), with $g_{Ae}^{\rm ALP} \propto (M\,T)^{-1/4}$ for galactic ALPs and $g_{Ae}^{\rm solar} \propto (M\,T)^{-1/8}$ for solar axions. The dominating background for these searches will come from irreducible solar neutrino interactions and from the $2\nu\beta\beta$ of ^{136}Xe, see Sections 3.2.2 and 3.2.3, respectively.

3.2.2 Solar neutrinos

The most restrictive background for DARWIN's dark matter physics program will come from solar neutrino interactions, see Section 4.3. On the other hand, the DARWIN detector can also study neutrinos, and this capability opens up another relevant physics channel, as detailed in [48]. A precise measurement of the *pp*-neutrino flux would test the main energy production mechanism in the Sun, since the *pp*- and ^{7}Be-neutrinos together account for more than 98% of the total neutrino flux predicted by the Standard Solar Model. A total ^{7}Be-neutrino flux of $(4.84 \pm 0.24) \times 10^9\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ has been measured by the Borexino experiment [66], assuming MSW-LMA solar oscillations, a flux which was confirmed by KamLAND [67]. However, the most robust prediction of the Standard Solar Model is for the *pp*-neutrino flux, which is heavily constrained by the solar luminosity in photons. A high-precision real-time comparison between the solar photon luminosity and the luminosity inferred by the direct measurement of the solar *pp*-neutrino flux would therefore severely limit any other energy production mechanism, besides nuclear fusion, in the Sun. Borexino has recently reported the first direct measurement of the *pp*-flux, $(6.6 \pm 0.7) \times 10^{10}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, with ∼10% precision [68].

The detection of low-energy solar neutrinos is through elastic neutrino-electron scattering $\nu + e^- \rightarrow \nu + e^-$. We estimated the potential of the DARWIN detector to measure their spectrum in real time [48]. Figure 5 (left) shows the recoil spectrum from *pp* and ^{7}Be neutrinos. (Figure 7 (left) on page 13 focuses on the low energy region.) The total expected number of events above an energy threshold of 2 keV$_{\rm ee}$ (electronic recoil equivalent) and below an upper limit of 30 keV$_{\rm ee}$, imposed by the rising $2\nu\beta\beta$ spectrum of ^{136}Xe, is $n_{pp} = 7.2$ events/day

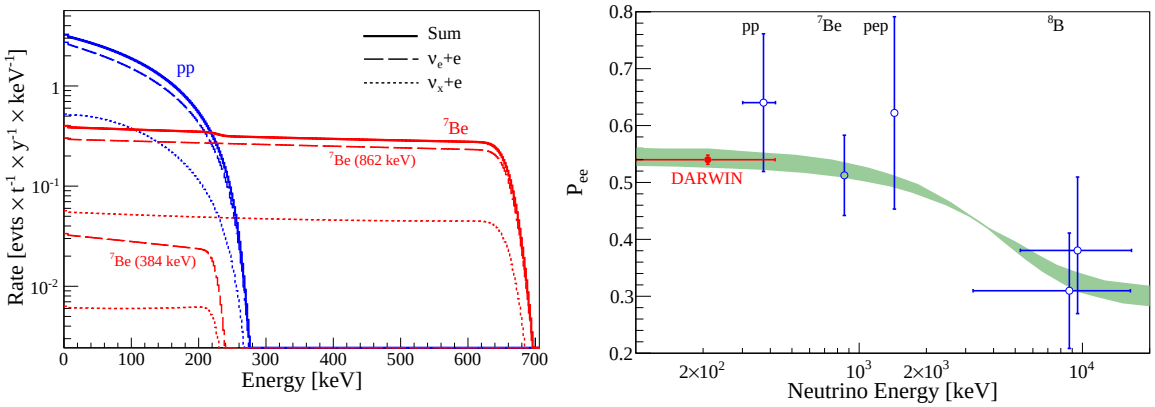

Figure 5. **(left)** Differential electron recoil spectra for *pp*- (blue) and ^{7}Be neutrinos (red) in liquid xenon. The sum contribution (solid line) is split into the contributions from ν_e (dashed) and the other flavours (dotted). Figure adapted from [48]. **(right)** Survival probability of solar, electron-neutrinos. The expected sensitivity of DARWIN (red) is shown together with existing measurements from Borexino and the MSW neutrino oscillation prediction ($\pm1\sigma$, green) for the large mixing angle scenario [68]. The precise measurement of the *pp*-flux with sub-percent precision with DARWIN will allow for testing neutrino and solar models.

and $n_{^7\mathrm{Be}} = 0.9\,\mathrm{events/day}$. These numbers assume a fiducial target mass of 30 tons of natural xenon and take into account the most recent values for the neutrino mixing angles [69]. More than $2{\times}10^3$ *pp*-neutrino events will be observed per year, allowing for a measurement of the flux with 2% statistical precision. A precision below 1% would be reached after 5 years of data taking. DARWIN would therefore address one of the remaining experimental challenges in the field of solar neutrinos, namely the comparison of the Sun's neutrino and electromagnetic luminosities with a precision of <1% [70]. The high statistics measurement of the *pp*-neutrino flux would open the possibility to test the solar model and neutrino properties, see Figure 5 (right). For example, non-standard neutrino interactions [71, 72] can modify the survival probability of electron neutrinos in the transition region around 1 MeV but also at *pp*-neutrino energies.

3.2.3 Neutrinoless double-beta decay

The question about whether neutrinos are Majorana fermions (i.e., their own antiparticles) is of intense scientific interest [73]. The most practical investigation of the Majorana nature of neutrinos, and of lepton number violation, is through the search for neutrinoless double-beta decay ($0\nu\beta\beta$). ^{136}Xe is an interesting $0\nu\beta\beta$-decay candidate and has an abundance of 8.9% in natural xenon. Its $Q_{\beta\beta}$-value is at 2.458 MeV, well above the energy-range expected from a WIMP recoil signal.

Two experiments, EXO-200 [74] and KamLAND-Zen [75, 76], have already reported very competitive lower limits on the half-life using a few hundred kilograms of ^{136}Xe. Even without isotopic enrichment, DARWIN's target contains more than 3.5 t of ^{136}Xe and can be used to perform a search for its $0\nu\beta\beta$-decay in an ultra-low background environment. The main challenge for this measurement will be to optimise the detector's sensors and readout electronics to perform at both the $\mathcal{O}(10)$ keV energy-scale and at the $\mathcal{O}(1)$ MeV scale relevant for the expected $0\nu\beta\beta$-decay peak. Once a resolution $\sigma/E \sim$ 1-2% at ~2.5 MeV is achieved and the background is reduced by a strong fiducialisation or the selection of ultra-low radioactiv-

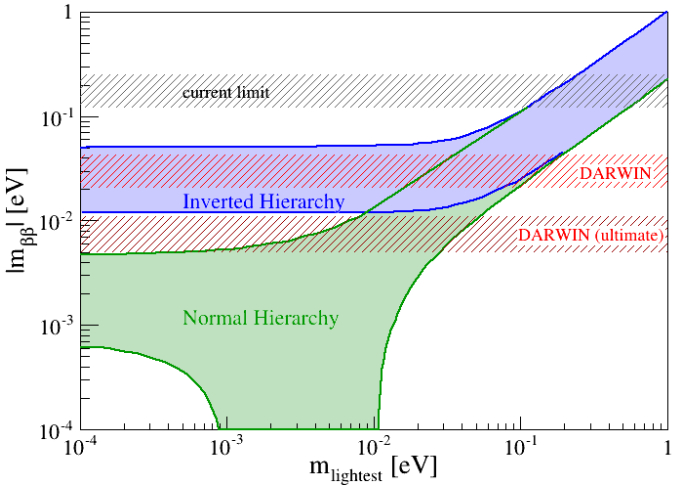

Figure 6. Expected sensitivity for the effective Majorana neutrino mass. The sensitivity band widths reflect the uncertainties in the nuclear matrix element of the ^{136}Xe $0\nu\beta\beta$-decay. The 'DARWIN' sensitivity assumes a 30 t×y exposure of natural xenon and a background dominated by γ-rays from detector materials. The 'ultimate' case with 140 t×y exposure assumes this background being absent, thus only ^{222}Rn, $2\nu\beta\beta$ and ^{8}B solar neutrinos contribute. For details, see [48]. Also shown are the expected regions for the two neutrino mass hierarchy scenarios.

ity detector materials, DARWIN's $0\nu\beta\beta$ sensitivity will become comparable to other future ton-scale experiments. In Figure 6, we show its reach for the effective Majorana neutrino mass $|m_{\beta\beta}|$ versus the mass of the lightest neutrino, as calculated in [48]. The corresponding sensitivities to the half-life of the decay are $T^{0\nu}_{1/2} > 5.6{\times}10^{26}$ y and $T^{0\nu}_{1/2} > 8.5{\times}10^{27}$ y (at 90% C.L.) for assumed natural xenon exposures of 30 t×y and 140 t×y, respectively. The latter value assumes that the material backgrounds can be completely mitigated and that the only background sources are 0.1 μBq/kg of ^{222}Rn, $2\nu\beta\beta$-decays and interactions of solar ^{8}B neutrinos. The width of the bands reflect the uncertainties in the nuclear matrix element calculations of the ^{136}Xe $0\nu\beta\beta$-decay. We expect smaller exposures compared to the WIMP search, caused by a more stringent fiducialisation to reach the required background level at $Q_{\beta\beta}$. Other rare nuclear processes, such as the double-beta decays of ^{134}Xe, ^{126}Xe and ^{124}Xe, might be observable as well [77].

3.2.4 Coherent neutrino-nucleus scattering

The rate of low-energy signals in all multi-ton WIMP detectors will eventually be dominated by interactions of cosmic neutrinos via coherent neutrino-nucleus scattering (CNNS) [78]. DARWIN will be able to detect and study this yet-unobserved standard model process, which produces a nuclear recoil signal like the WIMP. (For the implications on the WIMP search, see Section 4.3.) The largest CNNS rate comes from the relatively high-energy ^{8}B solar neutrinos which produce nuclear recoils $\leq$3 keV$_{\rm nr}$. Neutrinos from the solar hep-process induce a similar maximal recoil energy but their flux is much lower. Events from neutrinos created in the upper atmosphere and from the diffuse supernova neutrino background will extend to slightly higher energies of $\mathcal{O}(10)$ keV$_{\rm nr}$, however at significantly lower rates.

Because LXe detectors operate at low energy thresholds, even the low-energy CNNS signal from ^{8}B neutrinos is readily accessible. LUX has demonstrated an energy threshold of 1.1 keV$_{\rm nr}$ in a re-analysis of their data [27], and XENON10 reached a threshold of 1.4 keV$_{\rm nr}$ using an energy scale based on the S2 signal only [79]. Such thresholds would lead to an observed rate of ~90 events t^{-1}y^{-1} from ^{8}B neutrinos. The rate from atmospheric neutrinos will be much lower, around $3{\times}10^{-3}$ events t^{-1}y^{-1} [78]. Being a standard model process, the cross section of coherent neutrino-nucleus scattering can be calculated to a high precision [80, 81]. A deviation of the experimentally measured value from the expectation is thus a clear

sign of new physics, e.g., from oscillations into sterile neutrino states or from electromagnetic interactions of the neutrino [82].

3.2.5 Galactic supernova neutrinos

Neutrinos and anti-neutrinos of all flavours are emitted by core-collapse supernovae in a burst lasting a few tens of seconds [83]. The measurement of their energy spectrum and time structure can provide information about supernovae properties, but also about the intrinsic properties of neutrinos [84].

One avenue to detect such neutrinos is via their coherent scattering off xenon nuclei in DARWIN. In contrast to large water Cherenkov or liquid scintillator detectors, where the most significant interaction is from $\overline{\nu}_e$ via inverse β-decay on free protons, $\overline{\nu}_e + p \rightarrow e^+ + n$, DARWIN would be sensitive to all six neutrino species ν_x via neutral current interactions. The measured recoil spectrum would provide direct information about the neutrino energy spectrum, which is important to study neutrino oscillations etc. [84]. The expected nuclear recoil event rate in a LXe detector depends on the distance of the supernova explosion, the progenitor mass, the neutrino emission spectrum and on the differential neutrino-nucleus scattering cross section. It also depends on the detector properties, such as its fiducial mass, energy threshold, and detection efficiency for nuclear recoils.

The XMASS collaboration has calculated the expected event rates in their detector [85] and finds results consistent with those of [43]. However, XMASS is a single-phase detector and thus greatly limited in its sensitivity to this channel due to the higher energy threshold compared to an S2-based analysis. For DARWIN, we expect between 10 and 20 events t^{-1} from a supernova at a distance of 10 kpc, depending on the supernova neutrino emission model and progenitor mass [86]. Due to the extremely transient nature of a neutrino burst, lasting only a few seconds, the background in the low-energy S2 range can be expected to be negligible and will allow us to use a larger fiducial target [79, 87]. Thus, with $\mathcal{O}(100)$ neutrino events from a Galactic supernova, DARWIN will be a supernova neutrino detector that can contribute to the Supernova Early Warning System SNEWS [88]. By looking at the time evolution of the event rate from a nearby supernova, DARWIN could possibly distinguish between different supernova models [86] and provide complementary neutrino flavour-independent information to supernova physics.

4 Expected backgrounds

As a detector searching for rare events, DARWIN requires a very good understanding of all possible background sources and an extremely low absolute background level. The inner detector will hence be surrounded by a large, instrumented water shield, which passively reduces the environmental radioactivity as well as muon-induced neutrons, and acts as an active Cherenkov muon veto. Massive shields made from lead and copper are neither practical nor cost-effective at this scale. Typical ambient γ-ray and radiogenic neutron fluxes of $0.3\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and $9 \times 10^{-7}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ [89] are reduced by a factor of 10^6 after 3 m and 1 m of water shield, respectively [90], and can therefore be considered negligible. The γ-induced background from the radioactivity of detector materials is also irrelevant for WIMP searches at the DARWIN mass-scale [35].

Here we discuss the most important background sources for DARWIN, namely the backgrounds related to cosmogenic and radiogenic neutrons, background sources intrinsic to the target material, and finally neutrino-induced backgrounds.

4.1 Neutron backgrounds

A major background source for the WIMP search, where the expected signature is nuclear recoils from elastic WIMP-nucleus collisions, are neutron-induced nuclear recoils. The neutrons mainly originate from (α,n) reactions and spontaneous fission of heavy isotopes in detector materials, and can also be generated by interactions of cosmic ray muons that penetrate deep underground. Single-scatter nuclear recoils are in principle indistinguishable from a WIMP signal, hence the absolute neutron fluxes from all sources must be minimised *a priori*, e.g., by selecting materials with low intrinsic radioactive contamination [91]. Remaining neutron backgrounds will be reduced by self-shielding and by efficient multiple-scatter rejection, thanks to the large size and the excellent position resolution of the detector. DARWIN will reduce the muon-induced nuclear recoil background by its underground location and by using an active Cherenkov muon veto. We estimate that at the depth of the LNGS laboratory (3600 meters water equivalent) an active water shield of ∼14 m diameter will reduce the rate of cosmogenic neutrons to negligible levels. Most critical are interactions from (α, n) and fission neutrons from ^{238}U and ^{232}Th decays in detector components close to the LXe, such as PTFE reflectors or photosensors.

The expected neutron background was studied for a 40 t target mass, where the detector is mostly made of copper, PTFE, and photosensors [35]. The neutron energy spectra and yields in these materials were calculated taking into account their composition and reference levels of ^{238}U and ^{232}Th. Because secular equilibrium in the primordial decay chains is usually lost in processed materials, the ^{238}U and ^{232}Th activities are often determined via mass spectrometry or neutron activation analysis, while the activities of the late part of these chains, ^{226}Ra, ^{228}Ac, ^{228}Th, are determined via gamma spectrometry using ultra-low background high-purity germanium (HPGe) detectors [91–95]. With the detector material and radioactivity assumptions from [35], we obtain an expected single-scatter nuclear recoil rate of about 3.8×10^{-5} events t^{-1} y^{-1} keV$_{\rm nr}^{-1}$, in the central detector region of 30 t. This background level is sufficiently low for an exposure of 200 t×y at ∼30% nuclear recoil acceptance, but it requires us to identify materials with reduced radioactivity compared to currently measured levels. This is particularly true of the light reflector PTFE. This material contributes to the background significantly through (α, n) reactions due to the presence of ^{19}F. The neutron background could be further reduced by stronger fiducialisation, as well.

4.2 Xenon-intrinsic backgrounds

In dark matter detectors based on liquefied noble gases, radioactivity intrinsic to the WIMP target, such as ^{39}Ar, ^{85}Kr and ^{222}Rn provide sources of ER backgrounds. The activation of the xenon gas itself by exposure to cosmic rays becomes irrelevant after underground storage for a few months [96]. As ^{39}Ar is absent in LXe, the main challenges for DARWIN are ^{85}Kr and ^{222}Rn. ^{85}Kr is an anthropogenic radioactive isotope present in noble liquids extracted from air. Currently achieved $^{\rm nat}$Kr-levels after purification using krypton distillation or gas chromatography are (1.0±0.2) ppt by XENON100 [97], with a gas chromatography and mass spectrometry detection limit of 0.008 ppt, (3.5±1.0) ppt by LUX [98] and <2.7 ppt by XMASS [15]. The new $^{\rm nat}$Kr-removal apparatus of XENON1T has recently delivered a sample with a concentration of <0.03 ppt in a test run [99, 100]. This is a factor of 3 below the 0.1 ppt assumed for the WIMP sensitivity study presented in Section 3.1. Even at such low $^{\rm nat}$Kr concentrations, the background from $2\nu\beta\beta$ decays of ^{136}Xe in a natural xenon target only contributes at a much lower level [35], with a spectrum that decreases towards the threshold (see Figure 7, left).

^{222}Rn is part of the ^{238}U natural decay chain and constantly emitted by detector surfaces, which therefore have to be selected for low Rn-emanation [101, 102]. The ^{222}Rn concentrations in LXe achieved so far are 65 μBq/kg in XENON100 [25], 32 μBq/kg in LUX [98], 22 μBq/kg in PandaX-I [103], 9.8 μBq/kg in XMASS [15], and (3.65 ± 0.37) μBq/kg in EXO [104]. With the exception of EXO, none of these experiments was particularly optimised for low radon emanation. The target concentration for XENON1T is $\sim$10 μBq/kg, and the smaller surface-to-volume ratio will further help with the reduction in larger detectors. Nonetheless, we anticipate that achieving a low radon level will be the largest background reduction challenge. Concentrations of $\sim$0.1 μBq/kg must be achieved to probe WIMP-nucleon cross sections down to a few 10^{-49}cm^2, assuming an S2/S1-based rejection of ERs at the 2×10^{-4} level at 30% NR acceptance [35]. Such rejection levels have already been achieved [19]. The emanation of ^{220}Rn, a part of the ^{232}Th chain, could lead to similar backgrounds. However, it has a considerably shorter half-life than ^{222}Rn and is observed to be less abundant in existing LXe detectors, hence its impact is considered sub-dominant.

At DARWIN's current background goal of 0.1 ppt of $^{\text{nat}}$Kr in Xe, 0.1 μBq/kg of ^{220}Rn and a target of natural ^{136}Xe abundance, a total Xe-intrinsic background rate of $\sim$17 events t^{-1} y^{-1} is expected in a 2-10 keV$_{\text{ee}}$ WIMP search energy interval [35]. An ER rejection efficiency of 2×10^{-4} reduces this rate to 3.5×10^{-3} events t^{-1} y^{-1}.

4.3 Neutrino backgrounds

Neutrino signals in DARWIN will provide important science opportunities (see Sections 3.2.2, 3.2.4 and 3.2.5). However, they will also constitute the ultimate background source for many

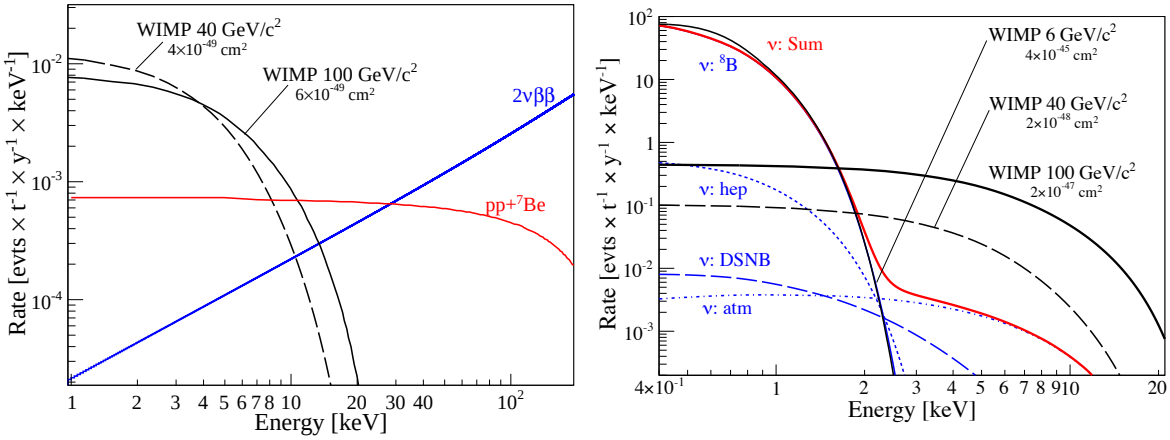

Figure 7. **(left)** Summed differential energy spectrum for *pp* and ^{7}Be neutrinos (red) undergoing neutrino-electron scattering in a LXe detector. Also shown are the electron recoil spectrum from the double-beta decay of ^{136}Xe (blue), as well as the expected nuclear recoil spectrum from WIMPs for a spin-independent WIMP-nucleon cross section of 6×10^{-49} cm^2 (solid black) and 4×10^{-49} cm^2 (dashed black) and WIMP masses of 100 GeV/c^2 and 40 GeV/c^2, respectively. A 99.98% discrimination of ERs at 30% NR acceptance is assumed and the recoil energies are derived using the S1 signal only (see [48]). **(right)** The differential nuclear recoil spectrum from coherent scattering of neutrinos (red) from the Sun, the diffuse supernova background (DSNB), and the atmosphere (atm), compared to the one from WIMPs for various masses and cross sections (black). The coherent scattering rate will provide an irreducible background for low-mass WIMPs, limiting the cross section sensitivity to $\sim 4 \times 10^{-45}$ cm^2 for WIMPs of 6 GeV/c^2 mass, while WIMP masses above $\sim$10 GeV/c^2 will be significantly less affected. No finite energy resolution but a 50% NR acceptance is taken into account. Figure from [48].

searches, especially for dark matter, as the neutrino flux can neither be shielded nor avoided by the design of the experiment. Solar *pp*-neutrinos (and ^{7}Be-neutrinos at the ∼10% level) will contribute to the electronic recoil background via neutrino-electron scattering at the level of ∼26 events $\mathrm{t}^{-1}\mathrm{y}^{-1}$ in the low-energy, dark matter signal region of the detector [48]. Solar neutrinos will thus become a relevant background source at WIMP-nucleon cross sections below $10^{-48}\mathrm{cm}^2$ and electronic recoil rejection levels around 99.98% are required to reach the envisaged ultimate WIMP sensitivity [35], reducing the solar neutrino background to 5.2×10^{-3} events $\mathrm{t}^{-1}\,\mathrm{y}^{-1}$ in a 2-10 $\mathrm{keV_{ee}}$ WIMP search region. Figure 7 (left) compares the combined background from *pp* and ^{7}Be solar neutrinos with two WIMP spectra.

Neutrino-induced nuclear recoils from coherent neutrino-nucleus scatters cannot be distinguished from a WIMP-induced signal, if no additional (statistical) discriminants such as track directionality are taken into account [105]. The ^{8}B solar neutrinos yield up to 10^3 events $\mathrm{t}^{-1}\mathrm{y}^{-1}$ for heavy targets such as xenon [78], however, all the events are at very low recoil energies below 4 keV, see Figure 7 (right). About 90 events $\mathrm{t}^{-1}\mathrm{y}^{-1}$ are expected above 1 $\mathrm{keV_{nr}}$. Nuclear recoils from atmospheric neutrinos and the diffuse supernovae neutrino background will yield event rates which are orders of magnitude lower but at slightly higher recoil energies. These will dominate the measured spectra at WIMP-nucleon cross sections around $10^{-49}\,\mathrm{cm}^2$ for WIMP masses above ∼10 GeV/c^2 [35, 36, 48, 78, 106, 107]. The total observed rates will strongly depend on the detector's energy threshold and energy resolution.

5 Design considerations and associated research and development

DARWIN will incorporate techniques which were successfully probed in the current generation of liquid xenon detectors and which will be tested in upcoming dual-phase dark matter experiments. At the same time, new design features will be evaluated and possibly implemented. Some of these were demonstrated with the XENON1T Demonstrator [108]. It established the ability to drift electrons over distances of 1 m in LXe, high-speed purification up to 100 standard liters per minute (slpm) [109], and high-voltage capabilities beyond −100 kV. Valuable input is also obtained from projects pursued outside the DARWIN consortium which employ TPCs filled with LXe to address various physics questions [26, 74]. In addition to these full-scale TPCs, several smaller instruments, with an R&D program focused at specific questions, are in place at various DARWIN institutions. The detailed design of DARWIN is not yet finalised, and further R&D towards such an ultimate WIMP detector is needed. In the following sections, we discuss potential designs based on state-of-the art concepts and technologies, and introduce some non-standard concepts as well.

5.1 Cryostat and time projection chamber

To reach the desired WIMP sensitivity within a reasonable timescale, DARWIN requires a total (target) LXe mass of ∼50 (40) tons, and hence a cylindrical detector with linear dimensions >2.5 m [35]. The WIMP search target after fiducialisation would be around 30 tons. The vacuum-insulated cryostats must be constructed from materials with a very low specific radioactivity level, where the cleanest available metal is copper. However, it imposes tight mechanical constraints, and, given the excellent self-shielding capabilities of LXe, it might be beneficial to use alternatives such as titanium or stainless steel instead, as in the XENON, LUX/LZ and PandaX projects.

All dual-phase LXe TPCs in current dark matter projects utilise the same concepts and mostly differ from each other in the details (high-voltage generation, aspect ratio, PMT

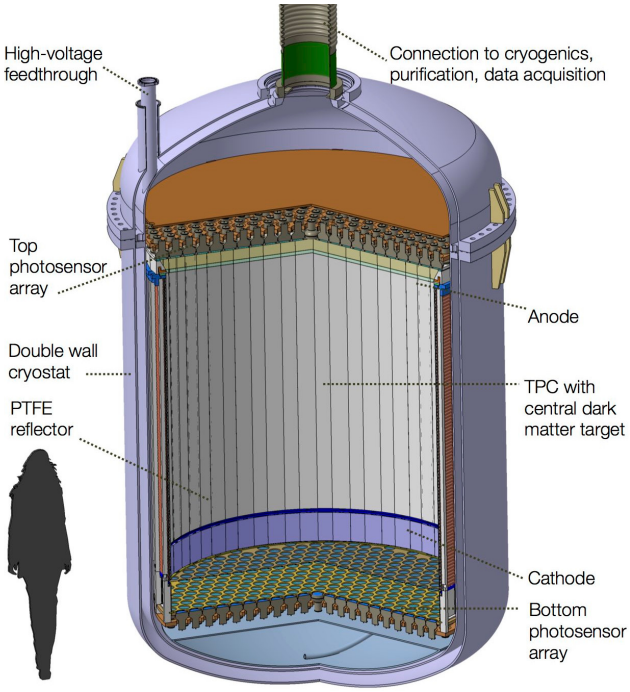

Figure 8. A possible realisation of a ∼50 t (40 t) total (target) LXe mass DARWIN detector, inside a double-walled stainless steel cryostat. The TPC is surrounded by highly reflective PTFE walls, closed by the cathode and anode electrodes on bottom and top, respectively. The sketch shows a TPC with two photosensor arrays made of circular PMTs with 3" diameter. The final sensor type, however, is not yet defined and all details regarding the cryostat and TPC are subject to R&D.

granularity, liquid level control, etc.). DARWIN, in its baseline configuration, will feature this well-established dual-phase TPC design scheme with light detected by photosensor arrays above and below the LXe target, see Figure 8. The light collection efficiency is constant for a fixed height-diameter ratio. With an optimal design of the reflecting inner TPC surfaces, it is only affected by the LXe absorption length. The working hypothesis of DARWIN's baseline design is that the absorption length can be kept much larger than the TPC diameter by continuous purification of the xenon, see Section 5.3. Under this assumption, with state-of-the-art PMTs, it is expected that the currently achieved thresholds of $\sim$1 keV$_{\rm nr}$ [27] can also be established with DARWIN. To cope with the possibility of smaller values for the absorption length – or, alternatively, to further increase the light collection efficiency – a potential scheme with the TPC surrounded by photosensors in $\sim 4\pi$, similar to a single-phase detector, is being evaluated as part of the DARWIN R&D program. This option is outlined in Section 5.4, which also discusses alternative photosensor technologies. A novel scheme relying on the concept of liquid hole multipliers (LHMs), with a potentially significant light yield improvement, is discussed in Section 5.4, as well.

Insulating materials are essential to construct the TPC, as components biased with very high voltages above −100 kV (cathode, field shaping electrodes) must be supported and insulated from grounded components. The primary choice is PTFE providing excellent insulation, good UV reflectivity [44], reasonable mechanical strength, and low radioactivity. A possible cylindrical DARWIN TPC of 260 cm diameter and height, enclosing a target mass of 40 t of xenon, is illustrated in Figure 8.

The type and dimension of the light sensors installed on the two arrays, above and below the target are still an active part of the DARWIN study. Under the assumption that the charge signal is detected via proportional scintillation in the gas phase, the 40 t LXe TPC would require ∼1800 sensors of 3" diameter (∼1000 of 4") assuming the use of identical, circular photosensors on both arrays. If available, larger low-radioactivity photosensors on

the bottom array could reduce the number of channels, as discussed in Section 5.4. A finite detector granularity on the top array is required for the xy-vertex identification.

5.2 High voltage system

To create an electron drift field of ∼0.5 kV/cm across the TPC, the cathode must be biased by a high, negative voltage. The field homogeneity is ensured by a set of circular field shaping rings, interconnected with high-ohmic resistors, realising a voltage divider which is gradually approaching ground potential. For a TPC of 2.6 m length, a cathode potential of 130 kV is required to establish the design field. We note that the TPC can be successfully operated at lower fields, as well (LUX: 0.18 kV/cm [98]); this reduces the field quenching, resulting in larger S1 signals. On the other hand, the lower electron drift velocity increases the pile-up rate during calibration runs and there are indications that the S2/S1 discrimination power deteriorates for reduced drift fields [50].

While the field shaping rings, which are made from massive copper with a smooth surface, do not impose a problem for the high voltage, the cathode electrode and the HV-feedthrough must withstand the high operation potentials. In order to optimise the optical transparency, the cathode will be made of single wires of ∼100 μm diameter, spot-welded to a sturdy, low-background metal frame. The cathode currently operated in XENON1T can be seen as a first prototype. A low-radioactivity feedthrough with demonstrated stable operation in LXe up to ∼130 kV has been built for XENON1T. As noted above, it would allow for reaching the design drift field of ∼0.5 kV/cm, similar to the one in XENON100 [25]. The extraction field across the liquid-gas interface requires a moderate, positive anode bias voltage for 100% electron extraction efficiency [110], which is typically between +5 and +10 kV, depending on the distance between the gate electrode, at ground potential, and the anode. The main challenge is to keep the anode, with a diameter of ∼2.6 m, parallel to the liquid surface, and to maintain a constant gap between gate and anode. This is required for achieving a homogenous S2 response across the detector's surface. The LHM-based TPC scheme discussed in Section 5.4 may provide a non-traditional solution.

The electrostatic configuration of DARWIN's TPC field cage, which is composed of very large (grid electrodes diameter) and small (wire diameter) elements, is being designed with *KEMField* [111]. This simulation tool, developed within the KATRIN collaboration, has been adapted for dual-phase TPCs. Optimised for the simulation of electric fields in large-scale geometries with small-scale structures, it takes advantage of the Boundary Element Method (BEM) to compute electric fields and potentials with the highest precision. For DARWIN-type geometries, BEM performs superior to Finite Element Methods (FEM), that require large computer memory for the meshing of the 3D-space.

5.3 Cryogenic and purification systems

The cryogenic system will consist of several sub-systems, which are required to initially liquefy the xenon target, maintain its constant low temperature during operation ($\Delta T/T < 0.05\%$), store it during down-time, and cleanse it of electronegative impurities which affect the light and charge yields, as well as from radioactive backgrounds. On-line measurements of the xenon purity will be performed by various diagnostic systems.

Cooling system The cooling system will have to maintain a stable cryogenic environment for many years. It can be accomplished by means of cryocoolers, such as pulse tube refrigerators (PTRs) [112] used in XENON, XMASS and PandaX, or by liquid nitrogen cooling [113],

as chosen for LUX/LZ. The design of the vacuum-insulated cryostat will be optimised in order to avoid heat leaks, and it will be equipped with super-insulation radiation shields. Given that the surface-to-volume ratio improves with the detector size, the cooling power required to keep the target cold will only be a few hundred Watts. Assuming similar cryogenics as in XENON1T, several redundant PTR cold-heads will be installed far away from the target volume, outside of the water shield. Xenon gas evaporated from the liquid target will be liquefied there, collected and returned into the main cryostat.

Storage The storage of the noble liquid inventory of a multi-ton liquid xenon detector requires a dedicated solution. In addition, the purity must be maintained, thus continuous storage in a closed system is essential. The system must also be able to store and purify the noble gas before the detector is available, in order to considerably shorten the time required for detector filling as well as for initial gas purification. It should allow for quick recuperation of the noble gas in liquid form in case of emergency or in case of maintenance operations, as transfer to the gas phase and storage in bottles would take several weeks. The solution developed for XENON1T is a new storage system, *ReStoX*, which satisfies all these requirements [114]. *ReStoX* consists of a vacuum-insulated stainless steel sphere of 2.2 m diameter, capable of holding 7.6 tons of xenon in liquid phase. Xenon is kept in liquid form by means of a liquid nitrogen-based cooling system and can be constantly purified during storage. As the system is designed to withstand pressures of up to 72 bar, the xenon can also be stored in gaseous phase at room temperature in case of longer shut-downs or emergency situations (longer power loss, etc.). Thus, a series of 7 interconnected *ReStoX* units is sufficient to store the full xenon inventory.

Purification from electronegative impurities To achieve a low energy threshold, good signal-background discrimination, and small signal corrections, the light and charge yields ought to be maximised. This requires a leak-tight detector and gas systems (metal seals) and constant purification of the noble-gas target from electronegative impurities such as N_2, O_2, H_2O, etc., which mainly stem from outgassing of material surfaces. During operation, the purification system constantly extracts gas from the detector and returns it after purification. While the MEG experiment [115] purifies its LXe target very efficiently in liquid form, this approach is not feasible for dark matter searches, due to the radioactive background induced by the recirculation pump for the liquid and due to radioactive contamination of the molecular sieves used for the cleaning. However, purification in the gas phase by means of hot zirconium getters at flow rates reaching up to ∼100 slpm (corresponding to 880 kg/day) has been achieved in the XENON1T R&D programme, using effective heat exchangers [116]. Different recirculation pumps suitable for noble gases have been investigated, specifically addressing their performance in terms of gas purity. Since the entire xenon inventory will be in contact with the pump during the purification process, it must be leak tight and not emanate ^{222}Rn. Standard diaphragm pumps, which were used for gas transport in previous experiments, do not fulfil the leak-tightness requirement. Thus, novel ultra-clean magnetically driven piston pumps with hermetically sealed pumping volumes based on the design of [117] are being developed and optimised. The performance of the purification system will be monitored online with dedicated purity monitors [118] and commercial systems, based on light absorption.

Purification from radioactive contaminants Another aspect of purity is the cleanliness of the target with respect to radioactive noble gas isotopes, see also Section 4.2, which can easily mix with LXe. Sub-ppt purification of xenon from krypton (necessary due to the

radioactive isotope ^{85}Kr) was already achieved [97, 119] by an ultra-clean cryogenic distillation column developed within the XENON project [100, 119, 120] together with a method to use the short-lived metastable ^{83m}Kr as a tracer to optimise the process [99, 100, 119, 121]. Another problematic isotope is ^{222}Rn, a daughter of the ^{238}U chain which is present in the underground environment and also continuously emanated from the surfaces of the inner detector and cryostat itself. While metal seals mitigate the external ^{222}Rn completely, all materials which are in contact with the target (liquid or gas) have to be specially selected for low Rn-emanation [101, 102]. Removing their topmost surface layer by electropolishing or etching can help to reduce this contribution. ^{220}Rn, a daughter in the ^{232}Th chain, could in principle also lead to backgrounds. However, due to its shorter half-life and lower abundance, it is less problematic. In addition, all efforts to reduce ^{222}Rn are expected to also work for ^{220}Rn.

Due to the minute amounts of radioactive impurities, separate online monitoring of the contamination level is not possible. Instead, the science data must constantly be checked for characteristic signatures of the contaminants. This is possible while the search regions for new physics remain blind. ^{222}Rn (^{220}Rn) can be tagged efficiently using high-energy α-events or the delayed $\beta - \alpha$ coincidence from the ^{214}Bi-^{214}Po (^{212}Bi-^{212}Po) decay, and ^{85}Kr can be identified using delayed $\beta - \gamma$ coincidence in $^{85}\text{Kr} \rightarrow {}^{85m}\text{Rb} \rightarrow {}^{85}\text{Rb}$. Due to the low branching ratio of 0.454%, the latter signature has a low efficiency and is thus not practical for low concentrations. For this reason, several off-line diagnostic methods were developed such as rare gas mass spectrometry (RGMS) [97, 100] and atom trap trace analysis (ATTA) [122]. These have sensitivities below the 1 ppt level and detect ^{nat}Kr rather than the radioactive isotope ^{85}Kr itself, which is present at the 2×10^{-11} level in ^{nat}Kr. A less precise (40 ppt) but on-site and approximately on-line method using a quadrupole mass spectrometer following a cold-trap has been established as well [123, 124].

5.4 Signal readout

The sensitivity of liquid xenon detectors to dark matter is closely related to the detector's light collection and detection efficiency, as the expected WIMP scattering spectra show an exponential rise towards low nuclear recoil energies. As also explained in Section 5.7, the relevant quantity to be optimised is the light yield (LY), which depends on the photon detection efficiency (PDE) of the photosensors and on the detector's light collection efficiency; the latter is determined by the photocathode and light-reflector coverage, the transparency of the TPC electrodes, the reflectivity of the TPC walls, and the VUV photon absorption length. The detector configurations (single- or dual-phase), with their different geometrical photon detector coverage (4π for single-phase detectors, top and bottom in present dual-phase detectors), have a large impact on the light yield as well. This is because a large fraction of the emitted light can be absorbed if multiple reflections of VUV photons are required to reach the sensors.

A common reference for the light yield is the detector response to the full absorption of 122 keV gamma rays at zero drift field: XENON100 has a light yield of 4.3 PE/keV [24], LUX, which has superior PMTs and a higher PTFE reflectivity, reaches 8.8 PE/keV [98], and the single-phase XMASS detector, where 62% of the spherical surface is covered by PMTs with a QE of 28%, has reached 14.7 PE/keV [15]. In the following, we review the relevant aspects for the light readout as studied within the DARWIN project, in order to achieve a design light yield around 8 PE/keV.

5.4.1 Photomultipliers

At the operating cryogenic temperatures, standard bialkali photocathodes used in PMTs can have extremely low saturation currents due to the increase in resistivity of the photosensitive material. This problem was solved with Hamamatsu's Bialkali LT (Low Temperature) photocathode, which operates down to liquid nitrogen temperatures with high saturation current and high QE. This photocathode is available with 3" diameter metal bulb photomultipliers. The Hamamatsu R11410 PMT with 12 dynode stages, optimised for use in liquid xenon with a mean QE of 35% at 178 nm, reaching up to ∼40% in some cases, and a collection efficiency (CE) of >90%, has been investigated in setups relevant for the next-generation dark-matter search experiments using LXe [45, 46]. The results show a stable gain of $\sim 5 \times 10^6$ at ∼1500 V and an excellent peak-to-valley ratio around 3 or higher. A subset has been operated continuously in LXe for several months and repeated cooling cycles from room temperature to LXe temperatures were performed successfully, without damage to the PMTs and without changes in their response. The PMTs underwent several successful high-voltage tests in xenon gas and in strong electric fields [45]. After several iterations, and collaborative efforts of Hamamatsu with XENON, the radioactivity of these 3" tubes was reduced to levels of <0.4 mBq/unit from ^{238}U, 0.02 mBq/unit from ^{226}Ra, and 0.01 mBq/unit from ^{228}Th [47].

We note that the high QE of bialkali PMTs comprises a significant contribution (∼20%) of events where a single VUV photon releases two photoelectrons from the photocathode [125]. Thus, for a PMT with QE=35% and CE=90%, the PDE for photons impinging on the photocathode is ∼25% (i.e., ∼80% of the product QE×CE). With ∼55% of the instrumented area covered by a photocathode (fill factor), the overall PDE for a densely-packed PMT array is thus ∼14% for state-of-the-art PMTs. However, this number is increased if the area between the PMTs is covered by efficient VUV-light reflectors.

5.4.2 Novel photosensors

PMTs, even if they are still the only photosensors used in current noble liquid dark matter detectors, have several important shortcomings: the residual radioactivity levels (although less relevant for very large detectors), cost, bulkiness, and stability at cryogenic conditions. Thus several alternative technologies are under consideration for DARWIN. The final choice of photosensors will be made during the design phase of the project, based on the technological maturity, performance characteristics (PDE, dark count rate, stability), radiopurity and cost of the different alternatives.

SiPM The silicon photomultiplier (SiPM) technology is rapidly developing and may become viable for readout of large detectors, offering very low radioactivity levels, compact geometry and low operation voltages. SiPMs may allow us to increase the photosensitive area coverage of the TPC and could in principle be suitable for 4π coverage. Arrays of SiPMs sensitive to visible light and suitable for operation in LAr are becoming commercially available on surface-mount boards with >75% fill factors over a $50 \times 50\,\mathrm{mm}^2$ area and are considered for the DarkSide-20k experiment [34]. Several vendors are also developing VUV-sensitive SiPMs suitable for operation in LXe, and PDE values >10% at 175-178 nm have been reported [126]. However, photon detection in large-volume dark matter LXe detectors by means of SiPM arrays still requires significant optimisation. In particular, their present best dark count rate ($\gtrsim 1\,\mathrm{Hz/mm}^2$ at LXe temperatures [126]), should be reduced by ∼2 orders of magnitude to keep accidental coincidence low enough for the desired detection thresholds. Improvements are

further required in the array PDE and correlated noise. The total correlated noise probability, namely the after-pulse and cross-talk probability, requires further studies as well.

SiGHT Another solution could come from the development of novel vacuum photosensors with reduced radioactivity and a simpler internal structure than a multi-stage PMT. A new photosensor concept, the Silicon Geiger Hybrid Tube (SiGHT), is under development [127]. SiGHT, a descendant of the QUPID detector [128], consists of a 3" diameter cylindrical tube with a hemispherical photocathode biased at $-3\,\mathrm{kV}$. A SiPM is placed on a pillar within the cylinder. The entire structure is made from ultra-clean synthetic fused silica to ensure very low levels of radioactivity. Electrons released from the photocathode are focused and accelerated onto the SiPM. A photoelectron hits a single pixel of the SiPM, yielding an output signal analogous to a single-photon hit on the SiPM. The SiPM pixels allow for excellent charge resolution, which translates to the device's ability to count integer numbers of photoelectrons. Good linearity can be achieved by using SiPMs with a high number of pixels. The SiGHT photosensor was designed to operate at low temperatures, expecting dark count rates of the same order as PMTs at LXe temperatures (tens of Hz per tube). Although the diameter of the photosensor currently being developed is 3", future developments may include larger 4" or 5" versions. The SiGHT photosensor could thus be well suited for ton- and multi-ton scale direct dark matter detection experiments.

Gaseous Photomultipliers Cryogenic Gaseous Photomultipliers (GPMs) [129] could become an economic alternative to PMTs for DARWIN, offering superior spatial resolution, compact geometry and similar overall PDE. GPMs combine a high-QE CsI-photocathode and cascaded gas-avalanche multipliers, e.g, Gas Electron Multipliers (GEMs) [130] or Thick Gas Electron Multipliers (THGEMs) [131]) coupled to an anode segmented into small pixels. Modular units, with a typical size of $\gtrsim 20 \times 20\,\mathrm{cm}^2$, equipped with UV-windows and embedded readout electronics, can be shaped to provide filling factors of $\sim$90% (compared to present $\sim$55% with circular PMTs), with comparable low radioactivity. For a nominal QE of $\sim$25% at 175 nm [132] and optimised choice of hole geometry and counting gas (Ne/CH_4 or Ar/CH_4 mixtures), an overall PDE of $\sim$15% over the entire instrumented area can be expected, similar to PMT arrays. While the superior granularity is not an *a priori* critical requirement, it may prove to be useful for precise event topology reconstruction.

The electron multipliers would be either cascaded THGEMs or hybrid structures, e.g., CsI-coated THGEM followed by thin-mesh multipliers [133, 134] with high gains ($>10^5$), allowing for high single-photon detection efficiency at LXe temperatures. Such GPMs, with reflective CsI photocathode, are suitable for the top photosensor array. GPMs could be also deployed at the TPC walls to provide $\sim 4\pi$ coverage with a considerable improvement in the S1 photon detection due to the reduction in multiple reflections (the degree of improvement will depend on the ratio between the absorption length and TPC diameter). The wall GPM should include an additional semi-transparent photocathode, on the inner window surface, to prevent loss of photons by total internal reflection. Stable operation of a 4" triple-THGEM GPM with a reflective CsI photocathode, coupled to a dual-phase LXe TPC, was recently demonstrated. It displayed a broad dynamic range, namely the capability to detect both single photons and massive α-induced S2 signals under the same conditions [135]. The feasibility of the 4π concept and the GPM PDE optimisation are subject to ongoing studies.

5.4.3 Liquid Hole-Multipliers: charge and light readout in a single-phase TPC

The Liquid Hole-Multiplier (LHM) is a new, "non-traditional", detection concept [136]. It originated from the concern that it may be quite difficult to maintain the liquid-gas interface, anode and gate completely parallel to each other across the diameter of a large dual-phase TPC; this may result in degraded S2 resolution and, as a possible consequence, reduced S2/S1-based background discrimination capability. This problem could be solved, in principle, if S2 light were to be generated in a high-field region within the liquid itself rather than in the vapour phase [137]. Some groups studied the possibility of generating S2 light around thin wires immersed in LXe [137–139], following works from the 1970s [140]. Parallel efforts focus on immersed THGEM and GEM electrodes. First experiments with a THGEM immersed in LXe demonstrated large S2 signals for alpha-particle induced ionisation electrons [141]. Subsequent studies [142, 143] proved that the light was in fact generated at the bottom part of the THGEM hole, in a xenon gas bubble trapped below the electrode, see Figure 9 (left). It was shown that the process can be controlled and maintained stable over many days. An S2 resolution (σ/E) of ∼7.5% was demonstrated for ∼6000 ionisation electrons, significantly better than in XENON100 (∼10-12%) [24]. Similar results were reached with immersed GEM electrodes [144]. With an appropriate field configuration the process can yield up to a few hundred photons per electron. Coating the electrode with CsI permits the detection of S1 photons in addition to ionisation electrons for the same field configuration, with a drift field of ∼0.5–1 kV/cm in the liquid [145]. Immersed GEMs are preferable over THGEMs for photon detection, as they allow for higher electric fields at the CsI photocathode surface and thus better photoelectron extraction. Based on a previous study, which indicated that the QE of CsI immersed in LXe is ∼30% at 175 nm for a sufficiently large field on the photocathode surface [146], the PDE across the CsI-coated GEM electrode is expected to be >15%.

Looking forward, GEM-based LHM modules with individual heating elements to generate the trapped bubbles, may be tiled to form a large instrumented surface sensitive to both

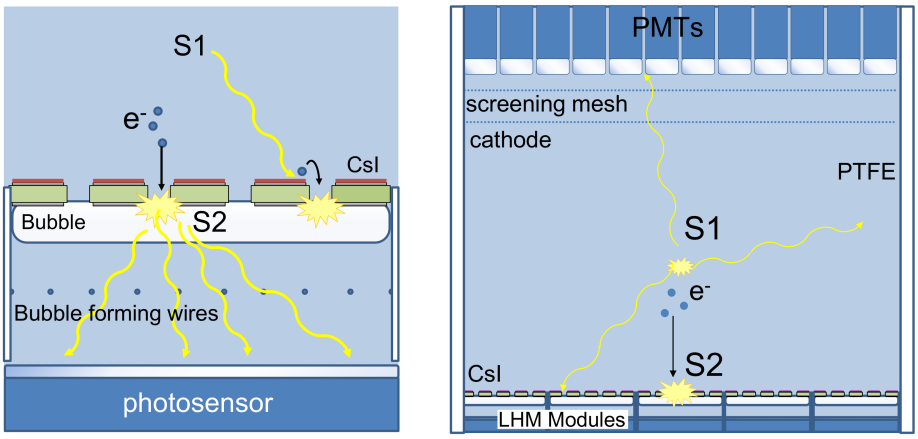

Figure 9. The liquid hole multiplier concept: **(left)** A single LHM element, coated with CsI to allow for the detection of S1 photons in addition to ionisation electrons. Controlled bubble formation can be achieved by ohmic heating of resistive wires below the electrode. **(right)** Schematic design of a liquid-only single-phase TPC with bubble-assisted LHMs at the bottom. S2 signals are created by electrons drifting down towards the LHM. Accurate S2-based position reconstruction is permitted by an array of photosensors below the LHMs. Figure adapted from [142].

ionisation electrons and S1 photons. This configuration, with an inverted drift field, is shown schematically in Figure 9 (right). It has two potential merits: first, it allows for a high S2 resolution over the entire area, by local control of the liquid-gas interface inside the module; second, it can considerably boost the TPC's light yield by the removal of three grids (gate, anode and screening electrode) as well as the liquid-gas interface itself. Even if the absorption length is not much larger than the TPC diameter, the loss of S1 photons is prevented by the effective elimination of multiple reflections: photons will typically reflect only once off the TPC PTFE wall before being detected by either the PMT array or the LHM array. Light readout of such LHM modules can be performed by PMTs, SiPMs or GPMs (the latter with reflective CsI photocathodes); the high dark count rate of SiPMs will be of no concern in this case as each primary S1 photon will produce a 'flash' of light comprising dozens of secondary photons inside the LHM module. While the basic LHM configuration comprises a bubble trapped below the electrode (and thus requires using LHMs horizontally at the bottom of the TPC), ongoing studies look into the possibility of trapping a gas layer above the electrode – opening the possibility of using LHMs at the top of the TPC.

5.5 Calibration

The signal response of liquid xenon detectors to particle energy deposits in the active volume are determined through regular calibration campaigns. The first generation of experiments has mainly relied on the use of external γ-sources (such as ^{137}Cs, ^{60}Co, ^{228}Th) and broad-band (^{241}AmBe, ^{252}Cf) or mono-energetic (generators) neutron sources for detector calibration. The considerably larger ton-scale detectors and in particular the multi-ton instrument DARWIN render calibration more challenging, as the effective background reduction by self-shielding also suppresses the detection of the majority of particles from external sources. It is however the innermost, central detector region which is used for new physics searches and thus must be calibrated precisely.

Several groups have started to investigate the use of sources dissolved in the liquid target, so-called internal sources. Only isotopes with a very short half-life can be employed for regular calibrations, followed by science runs. Some of these sources are neutron-activated xenon, providing rather high-energy γ-lines at 164 keV (^{131m}Xe, $T_{1/2}$ = 11.8 d) and 236 keV (^{129m}Xe, $T_{1/2}$ = 8.9 d) [147]. However, the half-lives of the activated Xe isotopes are too long, and their energies too high, to be useful for ton-scale detectors.

The best-studied example of a short-lived internal source is ^{83m}Kr [149, 150], a daughter of ^{83}Rb. It has been demonstrated that no long-lived Rb-isotopes were emitted into a ^{83m}Kr-sample for calibration [151]. ^{83m}Kr has a half-life of 1.83 h, and decays via 32.1 keV and 9.4 keV conversion electrons, where the intermediate state has a lifetime of 154 ns. The second low-energy process takes place close to the first one and has an increased light signal due

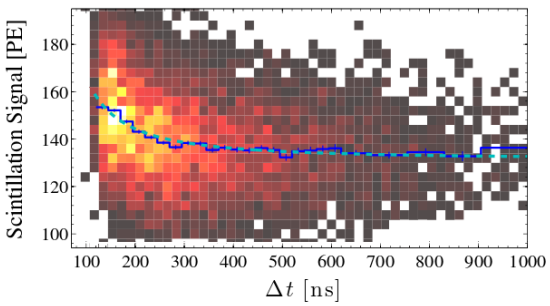

Figure 10. Scintillation light from the 9.4 keV line of ^{83m}Kr, following the initial 32.1 keV γ-ray. The half-life of the intermediate state is 154 ns. The light signal depends on the time difference Δt between the two events, due to the presence of ionisation from the first interaction. Therefore only 9.4 keV events with large Δt can be used for calibrations. Figure adapted from [148].

to an additional supply of ions and electrons that failed to recombine from the preceding transition. This limits the use of the 9.4 keV line to the tail of the time distribution where the effect becomes irrelevant [148], see also Figure 10. ^{83m}Kr has been successfully used in DarkSide-50 [152], XENON100 and LUX [98].

The LUX collaboration has also successfully employed tritiated methane (CH_3T) as an internal calibration source [153]. With $T_{1/2} = 12.3$ y, the half-life of tritium is too long to leave it to decay in the detector, however, the collaboration has demonstrated that it can be efficiently removed by a hot zirconium getter [154], designed to remove gases (N_2, O_2, H_2O, etc.) and other electronegative impurities from the liquid noble gas target (see Section 5.3). The XENON collaboration has used tritiated methane in XENON100, and is currently performing R&D towards the use of ^{220}Rn, emanated from a ^{228}Th source [155]. The use of the same source was also studied by XMASS [156]. It exploits the short half-life of ^{220}Rn and the subsequent β-decays of its daughters ^{212}Pb and ^{208}Tl, which are in radioactive equilibrium with the rest of the chain.

The light and charge response of dual-phase TPCs for dark matter searches must be calibrated for various reasons:

Energy calibration The use of the relative scintillation efficiency $\mathcal{L}_{\rm eff}(E_{\rm nr})$ and the charge yield $Q_y(E_{\rm nr})$, which describe the size of the signal from a nuclear recoil of energy $E_{\rm nr}$ with respect to a fixed-energy γ-source, requires a calibration of the detector with mono-energetic γ-lines. The traditionally used response to 122 keV γ's from a ^{57}Co source was recently replaced by the intrinsic standard 32.1 keV line from ^{83m}Kr. This single line allows for the light/charge calibration of a DARWIN detector using the relative efficiencies measured in dedicated experimental setups, under tightly controlled conditions, as elaborated in Section 5.6.

For the characterisation of the electronic recoil background up to several MeV, an energy scale set by the full absorption peaks of various γ-sources is desirable. Intrinsic sources are mandatory to populate the inner detector regions with a sufficient number of events. One possibility is using the various decay lines seen in ^{220}Rn and its daughters [155]. Another possibility is the known shape of β-spectra from intrinsic contaminations and sources, e.g., the $2\nu\beta\beta$-spectrum from ^{136}Xe.

Nuclear recoil band (signal-like events) The detector response to single-scatter nuclear recoils down to lowest energies ($\sim$1 keV$_{\rm nr}$) must be known precisely. Neutron-induced recoil spectra generated by broad-spectrum sources such as ^{241}AmBe and ^{252}Cf, or by mono-energetic MeV-neutrons from generators, are representative for WIMP-generated spectra of $m_\chi \gtrsim 50\,\mathrm{GeV}/c^2$. The correct light and charge distribution for lower-mass WIMPs is to be determined by feeding the results from these calibrations into Monte Carlo codes, which take into account the signal resolutions [157, 158].

A direct calibration of the signal distribution with external neutron sources, preferentially mono-energetic neutrons from a generator in order to exploit the kinematics of multiple-scatter processes, is still feasible for DARWIN, thanks to the rather long mean-free path of neutrons in LXe. A direct calibration of the low-energy nuclear recoil response with an ^{88}YBe-source, as recently proposed by [159] and currently studied in XENON100, will be only possible in much smaller experimental setups.

Electronic recoil band (background-like events) To establish a background model for the WIMP dark matter search, the distribution of the ER signal in the TPC must be well understood. The same holds for all searches using the ER signal itself, such as solar

neutrinos and axions, see Section 3.2. This background is mostly due to target-intrinsic β-contaminations. Line sources, such as ^{83m}Kr, cannot be used for calibration as a continuous spectrum is required. We thus expect to use intrinsic sources such as tritiated-methane and ^{220}Rn. These provide a continuous spectrum of low energy recoils through ground-state to ground-state beta decays either directly (TCH_3) or through their daughters (^{220}Rn). The comparison of signal and background calibration samples allows for the determination of the background discrimination level. The charge-to-light ratio exploits the different (position-corrected and energy-dependent) mean values of the S2/S1-ratio for signal and background, thanks to the different energy loss mechanisms of the recoils [160].

5.6 Light and charge yield of electronic and nuclear recoils

The response of LXe to particle interactions with energy depositions around the detection threshold is of high relevance, as the differential nuclear recoil spectrum induced by WIMP-nucleus elastic scattering is exponentially decreasing. In particular, WIMPs with masses below 10 GeV/c^2 could potentially leave signatures only in the lowest energy bins of a liquid xenon detector. Such measurements have been performed within the DARWIN consortium in the last years [157, 161, 162] and new measurements are currently ongoing.

The response of large LXe detectors to low-energy nuclear recoils can be measured *in situ* using monoenergetic neutrons from deuterium-deuterium (D-D) fusion generators, as shown by LUX [27] and planned for the XENON1T experiment. Such a measurement, in combination with an empirical response model that is fitted to the data, allows for the determination of the charge and light yields down to 1 keV$_{\mathrm{nr}}$ energy, or even below [27].

To date, estimates of the dark matter sensitivity of LXe TPCs assume that electric fields have a small effect on the light yield from nuclear recoils. This assumption is supported by initial measurements [163], indirect analyses [157] as well as semi-empirical models (NEST) [164, 165]. Direct measurements in small LXe detectors are nonetheless ongoing.

For standard WIMP searches, electronic recoils are expected to stem from background events, however, for axion searches, leptophilic dark matter models or low-energy neutrino measurements, ERs turn into the expected signal (see also Sections 3.2.1 and 3.2.2). The scintillation yields of ERs in liquid xenon were measured with small cryogenic cells using Compton-scattered photons from collimated, high-activity ^{137}Cs sources in coincidence with HPGe and NaI detectors placed under various scattering angles, corresponding to energies down to 1.5 keV [148, 166]. Results from data at zero electric field show a decrease of the yield of recoiling electrons below 20 keV, to a level of $\sim$ 40% of its value at higher energies at around 1.5 keV. Measurements of the light quenching in an electric field have also been performed [148]. The results have been used to set an energy scale for axion searches with XENON100 [59], to test leptophilic dark matter models [167], and for the first search for an annual modulation signal with a LXe TPC [168].

5.7 Detector resolution

The light and charge signals are both employed in the data analysis process. While the energy scale can be derived either using one of these [25] or from their linear combination [98], background discrimination via the S2/S1 ratio always requires the precise knowledge of both quantities. The smaller and hence more-difficult-to-detect S1 signal sets the energy threshold of the detector. The energy resolution is derived from the individual signals or their combination, and is related to the number of detected physical quanta. For the S1 channel, this is the number of photons recorded by the photosensors, giving rise to signals measured in PE.

Because of the finite quantum efficiency (∼35% for state-of-the-art PMTs), photoelectron collection efficiency (∼90%) and light collection efficiency (LCE), which describes the fraction of primary photons reaching a photosensitive area, the detected number of PE is considerably smaller than the number of initially generated photons. The LCE depends on the target purity, the reflectivity of the inner TPC surfaces, the transparency of the TPC electrodes, the photocathode coverage of the detector, as well as on the TPC height-to-diameter ratio. All these parameters can be optimised for a given detector. The light yield of the LUX detector, 8.8 PE/keV$_{ee}$ for a 122 keV line (electron recoil equivalent) at zero-field [98], is about 2× higher than the one of XENON100 (4.3 PE/keV$_{ee}$ [24]), mainly due to an improved PTFE reflectivity, optimised TPC electrodes transparency and a higher collection efficiency of the PMTs. Typically, PTFE is used as an efficient reflector for the scintillation light of LXe at 178 nm, with a reflectivity above 90% [44].

Due to electron-ion recombination effects, the light yield (LY) is always highest for zero-field and decreases (in an energy-dependent way) with increasing drift field, see for example [150]. The numbers for XENON100 and LUX quoted above reduce to 2.3 PE/keV$_{ee}$ and ∼4.6 PE/keV$_{ee}$ at 122 keV$_{ee}$ and $|\vec{E}| \sim 0.5$ kV/cm, respectively. Figure 11 (left) illustrates this situation for a nuclear recoil signal of 5 keV$_{nr}$: a higher light yield will improve the S1-resolution at a given nuclear recoil energy E_{nr}. However, current dark matter TPCs are already highly optimised, hence not much improvement beyond values of $4-5$ PE/keV$_{ee}$ at a 0.5 kV/cm drift field (about $8-10$ PE/keV$_{ee}$ at zero field) can be realistically expected for DARWIN, leading to an anticipated S1-resolution of around 40% at the detector threshold.

Due to the larger number of quanta involved, the resolution of the proportional S2 signal

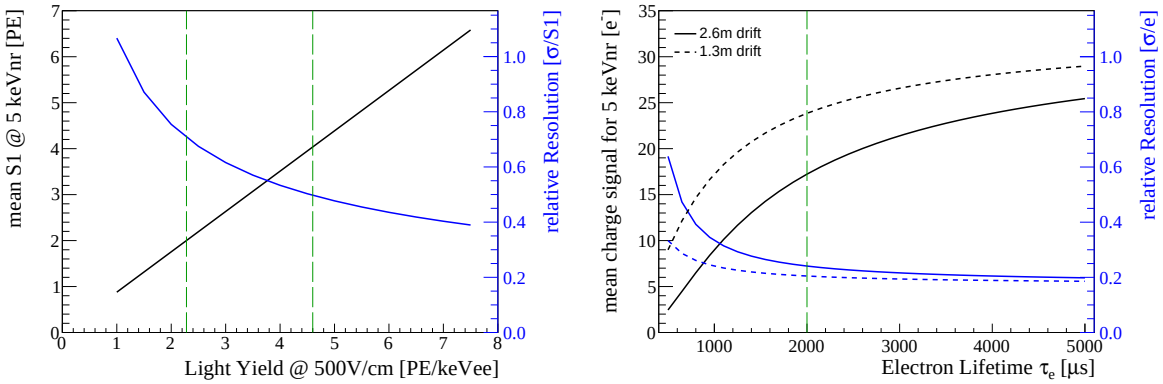

Figure 11. **(left)** Mean size of the S1 light signal in a dual-phase LXe TPC, expected from a nuclear recoil of 5 keV$_{nr}$ (left axis), plotted against the detector's light yield for a 122 keV γ-line at a drift field of ∼500 V/cm. The relative scintillation efficiency $\mathcal{L}_{eff}$ of LXe for this study is taken from [23]. Also shown is the relative resolution (σ/S1) at 5 keV$_{nr}$, assuming that it is dominated by a Poisson process (right axis). The LXe TPCs XENON100 [24] and LUX [169] have achieved light yields of ∼2.3 PE/keV$_{ee}$ and ∼4.6 PE/keV$_{ee}$ at this drift field, respectively, as indicated by the green lines. **(right)** Mean charge signal in electrons, before gas amplification, for a 5 keV$_{nr}$ recoil signal as a function of the electron lifetime τ_e. A drift field of ∼500 V/cm is assumed. The charge yield Q_y for LXe is taken from [157]. The right axis shows the relative Gaussian resolution for 1.3 m and 2.6 m electron drift, corresponding to the central and the maximal value in a DARWIN detector. Lifetimes of 2 ms have already been achieved in LXe detectors (green line). In both cases, it is unrealistic that DARWIN can improve significantly with respect to the numbers already achieved, leading to S1 resolutions of about 40% for 5 keV$_{nr}$ recoil signals and to charge resolutions of about 20%.

(from the ionisation electrons) is superior to the S1 resolution. The W-value, describing the energy required to create an electron-ion pair, is 15.6 eV in xenon [170]. The number has to be corrected for recombination effects in the non-zero electric field, leading to a somewhat larger effective W-value. However, the total number of electrons liberated by low-energy interactions is still rather high. As the typical electron extraction fields of >9 kV/cm lead to 100% extraction efficiency [110], the main loss mechanism is the effective electron lifetime, τ_e, due to capture of electrons by electronegative impurities in the target. Electron lifetimes of 2 ms have been demonstrated in large LXe detectors [108] and its impact on the resolution is illustrated in Figure 11 (right): for $\tau_e \geq 2$ ms and 100% extraction efficiency, the charge resolution is better than 24% for 5 $\mathrm{keV_{nr}}$ recoils throughout the TPC. For $\tau_e = 5$ ms, the resolution improves to 20%. The number of photoelectrons detected by the PMT arrays is typically about 20 PE/e^- [110], and depends on the gas pressure, the distance between liquid-gas interface and anode as well as the extraction field. Due to the much larger number of quanta, the fluctuations in the number of PE are subdominant compared to the fluctuations in the number of electrons.

The linear combination of light and charge signals, exploiting an anti-correlation between the two in ERs [171], allows for energy resolutions similar to NaI crystals ($\sigma/E = 1.53\%$ has been reached at 2480 keV [74]). Such a resolution would be relevant for many of the non-WIMP searches which focus on ERs, such as searches for neutrinoless double beta-decays or for axions and ALPs. For low-energy nuclear recoils, no strong anti-correlation is expected in LXe because the recombination-fluctuations are sub-dominant compared to the uncorrelated S1 and S2 fluctuations [160, 172]. The resolution of a combined light and charge signal close to threshold is therefore simply given by the sum of the physical quanta. It is dominated by the charge signal and therefore only leads to a minor improvement of the resolution compared to an S2-only scale [35]. As charge and light signals are both smaller for NRs compared to ERs, due to quenching effects, the energy resolution of an ER signal will always be superior to the one of a NR of the same energy, regardless of how the energy scale is reconstructed.

5.8 Data acquisition and trigger schemes

This section addresses issues regarding the electronics and data acquisition (DAQ). While the total number of channels in DARWIN is still moderate compared to accelerator-based physics experiments, the small, keV-sized signal in the dark matter channel requires an extremely low energy threshold and excellent noise conditions. The digitized waveform of every channel is recorded in order to use the maximum amount of information for further data analysis. The DAQ system must also be able to handle the different data taking rates in science mode and during calibration runs.

The DAQ system of a large dual-phase TPC has to address several challenges. The number of channels, ≥1000 photosensors, depending on the sensor size, will be several times higher than in present (e.g., XENON100: 242 [24], LUX: 122 [26], PandaX-II: 110 [28], XENON1T: 248 [32]) or upcoming (e.g., LZ: 488 [33]) detectors. The maximum time difference between the S1 and S2 signals will be of the order of one millisecond. The detailed waveforms, at $\mathcal{O}(2-10)$ ns resolution, must be digitised as they contain information for the S2/S1 background discrimination, noise rejection, etc. However, there is little information stored in the comparatively long time between the S1 and S2 peaks and already present-day detectors do not digitize this part of the waveforms to reduce the amount of data [24, 26].

The causal connection between S1 and S2 signals limits the maximum achievable detector rate due to the possibility of event pile-up. While conservatively high trigger rates of ∼10 Hz

in dark matter mode would only lead to ∼1.5% of events suffering from pile-up in DARWIN, large LXe detectors with their excellent self-shielding capabilities will generally require very large amounts of calibration data to reach a sufficient number of calibration events within the fiducial volume, hence a high calibration rate, see Section 5.5. In a "classical" event-based readout using a common trigger, the events will overlap and a constant acquisition window will lead to information loss. Finally, the trigger threshold must be as low as possible to ensure a low energy threshold.

The increased number of channels in DARWIN can be handled through parallelisation, the level of which can be increased according to the actual requirements of the experiment. All readout channels will operate independently from one another, they are not triggered globally but run in "self-triggering"-mode. The data from all channels will be correlated and reconstructed in real-time on commodity computing hardware, which only keeps the information associated with an event for storage. Such a flexible software trigger can also be further parallelised in order to increase the computation speed. It will allow for detector calibration in pile-up mode, as overlapping events can be accepted while the correlation analysis of the S1 and S2 peaks is postponed. This analysis could be based on the reconstructed xy-positions of the S1 and the S2 signals, as well as on the size of the S2/S1 ratio.

Since the data is fully reconstructed before making a storage decision, more sophisticated filtering algorithms can be used to enhance DARWIN's physics capabilities. Examples are: (i) xy-position reconstruction to only store events reaching the central volume during calibration runs, (ii) a relatively small, random subset of events can be stored to study high-energy and background events, and (iii) specialised triggers can be used to study specific background topologies (e.g., delayed coincidence triggers to select ^{214}Bi-Po events to study ^{222}Rn backgrounds).

A DAQ system designed according to this concept was developed for the XENON1T detector. Because of its scalability by increasing the level of parallelisation, it can be regarded as the first step towards a DAQ system for DARWIN.

6 Summary and Outlook

DARWIN will be the ultimate liquid xenon dark matter detector with a sensitivity for spin-independent WIMP-nucleon cross sections down to $\sim 10^{-49}\,\mathrm{cm}^2$ capable to detect or exclude WIMPs with masses above $\sim 5\,\mathrm{GeV}/c^2$. With its large target mass, low energy threshold, and ultra-low background level, DARWIN will also provide a unique opportunity for other rare event searches such as axions and other weakly interacting light particles. It will address open questions in neutrino physics, e.g., by measuring the low-energy solar neutrino spectrum with better than 1% precision or by searching for the neutrinoless double beta decay of ^{136}Xe. At its lowest energies, the DARWIN detector will provide the possibility to observe coherent neutrino-nucleus interactions from solar ^{8}B neutrinos, to precisely test the standard solar model flux-prediction, and to detect neutrinos from galactic supernovae.

DARWIN will employ a time projection chamber filled with 40 t of liquid xenon; the full instrument will require about 50 t of liquid xenon. A vigorous R&D and design effort is ongoing within the international DARWIN consortium. It comprises technical aspects such as the design and prototyping of the time projection chamber, Monte Carlo studies of the expected radiogenic and cosmogenic backgrounds, investigation of new light and charge readout schemes and of novel sensors to operate in liquid xenon, selection of low-background construction materials by means of high-purity germanium spectroscopy and other techniques,

radon emanation measurement and removal, new data acquisition and trigger schemes, data analysis and also addresses the scientific reach of the facility. Charge and light yield measurements of nuclear and electronic recoils at lowest energies, necessary to define accurate energy scales in a LXe dark matter detector, are ongoing at several institutions. In parallel to the baseline dual-phase detector geometry, single-phase TPC concepts are being evaluated and prototyped. The R&D and design phase will end by 2019, after which the construction of the various sub-systems will start. Following detector installation and commissioning in the underground laboratory, a first science run could start by 2023. To fully exploit its WIMP sensitivity, the facility would be operated for at least 7 years.

In summary, DARWIN has a unique discovery potential in the areas of astroparticle and low-energy neutrino physics.

Acknowledgments

This work has been supported by the ASPERA first common call (EU), the University of Zurich (CH), the Albert Einstein Center at the University of Bern (CH), the Swiss National Foundation (SNF), the FP7 Marie Curie-ITN action *Invisibles* (EU), the Max-Planck Society (DE), the PRISMA Cluster of Excellence in Mainz (DE), the Helmholtz Alliance for Astroparticle Physics (DE), the Israel Science Foundation (IL), the MINERVA Foundation (DE), the Stichting Fundamenteel Onderzoek der Materie (NL), the Fundação para a Ciência e Tecologia (PT), the Imperial College Trust and the Science and the Technology Facilities Council (UK), the Istituto Nazionale di Fisica Nucleare (IT), and the National Science Foundation (USA).

References

[1] Planck collaboration, P. A. R. Ade et al., *Planck 2013 results. XVI. Cosmological parameters*, *Astron. Astrophys.* **571** (2014) A16, [1303.5076].

[2] M. W. Goodman and E. Witten, *Detectability of Certain Dark Matter Candidates*, *Phys. Rev.* **D31** (1985) 3059.

[3] G. Bertone, D. Hooper and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279–390, [hep-ph/0404175].

[4] L. Baudis, *Dark matter detection*, *J. Phys.* **G43** (2016) 044001.

[5] G. Jungman, M. Kamionkowski and K. Griest, *Supersymmetric dark matter*, *Phys. Rept.* **267** (1996) 195, [hep-ph/9506380].

[6] L. Baudis, *Direct dark matter detection: the next decade*, *Phys. Dark Univ.* **1** (2012) 94, [1211.7222].

[7] M. Schumann, *Dark Matter 2014*, *EPJ Web Conf.* **96** (2015) 01027, [1501.01200].

[8] L. Baudis, *Dark matter searches*, *Annalen Phys.* **528** (2016) 74, [1509.00869].

[9] T. Marrodán Undagoitia and L. Rauch, *Dark matter direct-detection experiments*, *J. Phys.* **G43** (2016) 013001, [1509.08767].

[10] S. Arrenberg et al., *Working Group Report: Dark Matter Complementarity*, 1310.8621.

[11] C. Strege et al., *Profile likelihood maps of a 15-dimensional MSSM*, *JHEP* **09** (2014) 081, [1405.0622].

[12] V. Chepel and H. Araujo, *Liquid noble gas detectors for low energy particle physics*, *JINST* **8** (2013) R04001, [1207.2292].

[13] M. Schumann, *Dual-Phase Liquid Xenon Detectors for Dark Matter Searches*, *JINST* **9** (2014) C08004, [1405.7600].

[14] L. Baudis, *WIMP Dark Matter Direct-Detection Searches in Noble Gases*, *Phys. Dark Univ.* **4** (2014) 50–59, [1408.4371].

[15] XMASS collaboration, K. Abe et al., *XMASS detector*, *Nucl. Instrum. Meth.* **A716** (2013) 78, [1301.2815].

[16] DEAP-3600 collaboration, M. Boulay, *DEAP-3600 Dark Matter Search at SNOLAB*, *J. Phys. Conf. Ser.* **375** (2012) 012027, [1203.0604].

[17] MINICLEAN collaboration, K. Rielage et al., *Update on the MiniCLEAN Dark Matter Experiment*, *Phys. Procedia* **61** (2015) 144–152, [1403.4842].

[18] M. Boulay and A. Hime, *Technique for direct detection of weakly interacting massive particles using scintillation time discrimination in liquid argon*, *Astropart. Phys* **25** (2006) 179.

[19] ZEPLIN-III collaboration, D. Y. Akimov et al., *WIMP-nucleon cross-section results from the second science run of ZEPLIN-III*, *Phys. Lett.* **B709** (2012) 14, [1110.4769].

[20] ZEPLIN-III collaboration, D. Y. Akimov, *The ZEPLIN-III dark matter detector*, *Nucl. Instrum. Meth.* **A623** (2010) 451.

[21] XENON10 collaboration, J. Angle et al., *First Results from the XENON10 Dark Matter Experiment at the Gran Sasso National Laboratory*, *Phys. Rev. Lett.* **100** (2008) 021303, [0706.0039].

[22] XENON10 collaboration, E. Aprile et al., *Design and Performance of the XENON10 Dark Matter Experiment*, *Astropart. Phys.* **34** (2011) 679, [1001.2834].

[23] XENON100 collaboration, E. Aprile et al., *Dark Matter Results from 100 Live Days of XENON100 Data*, *Phys. Rev. Lett.* **107** (2011) 131302, [1104.2549].

[24] XENON100 collaboration, E. Aprile et al., *The XENON100 Dark Matter Experiment*, *Astropart. Phys.* **35** (2012) 573, [1107.2155].

[25] XENON100 collaboration, E. Aprile et al., *Dark Matter Results from 225 Live Days of XENON100 Data*, *Phys. Rev. Lett.* **109** (2012) 181301, [1207.5988].

[26] LUX collaboration, D. Akerib et al., *The Large Underground Xenon (LUX) Experiment*, *Nucl. Instrum. Meth.* **A704** (2013) 111, [1211.3788].

[27] LUX collaboration, D. S. Akerib et al., *Improved Limits on Scattering of Weakly Interacting Massive Particles from Reanalysis of 2013 LUX Data*, *Phys. Rev. Lett.* **116** (2016) 161301, [1512.03506].

[28] PANDA-X collaboration, A. Tan et al., *Dark Matter Search Results from the Commissioning Run of PandaX-II*, 1602.06563.

[29] WArP collaboration, P. Benetti et al., *First results from a dark matter search with liquid argon at 87K in the Gran Sasso underground laboratory*, *Astropart. Phys.* **28** (2008) 495–507, [astro-ph/0701286].

[30] ArDM collaboration, C. Amsler et al., *First results on light readout from the 1-ton ArDM liquid argon detector for dark matter searches*, *JINST* **5** (2010) P11003, [1009.3641].

[31] DarkSide collaboration, P. Agnes et al., *Results from the first use of low radioactivity argon in a dark matter search*, *Phys. Rev.* **D93** (2016) 081101, [1510.00702].

[32] XENON collaboration, E. Aprile et al., *Physics reach of the XENON1T dark matter experiment*, *JCAP* **1604** (2016) 027, [1512.07501].

[33] LZ collaboration, D. S. Akerib et al., *LUX-ZEPLIN (LZ) Conceptual Design Report*, 1509.02910.

[34] DarkSide collaboration, D. Franco et al., *The Search for Dark Matter with low-radioactivity Argon at LNGS, Letter of Intent, https://indico.in2p3.fr/event/12120/contribution/1/material/paper/1.pdf* (2015) .

[35] M. Schumann et al., *Dark matter sensitivity of multi-ton liquid xenon detectors*, *JCAP* **1510** (2015) 016, [1506.08309].

[36] J. Billard, L. Strigari and E. Figueroa-Feliciano, *Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments*, *Phys. Rev.* **D89** (2014) 023524, [1307.5458].

[37] XENON100 collaboration, E. Aprile et al., *Limits on spin-dependent WIMP-nucleon cross sections from 225 live days of XENON100 data*, *Phys. Rev. Lett.* **111** (2013) 021301, [1301.6620].

[38] LUX collaboration, D. S. Akerib et al., *Results on the Spin-Dependent Scattering of Weakly Interacting Massive Particles on Nucleons from the Run 3 Data of the LUX Experiment*, *Phys. Rev. Lett.* **116** (2016) 161302, [1602.03489].

[39] S. A. Malik et al., *Interplay and Characterization of Dark Matter Searches at Colliders and in Direct Detection Experiments*, *Phys. Dark Univ.* **9-10** (2015) 51, [1409.4075].

[40] DarkSide collaboration, T. Alexander et al., *DarkSide search for dark matter*, *JINST* **8** (2013) C11021.

[41] M. Pato et al., *Complementarity of Dark Matter Direct Detection Targets*, *Phys. Rev.* **D83** (2011) 083505, [1012.3458].

[42] J. L. Newstead et al., *The Scientific Reach of Multi-Ton Scale Dark Matter Direct Detection Experiments*, *Phys. Rev.* **D88** (2013) 076011, [1306.3244].

[43] S. Chakraborty, P. Bhattacharjee and K. Kar, *Observing supernova neutrino light curve in future dark matter detectors*, *Phys. Rev.* **D89** (2014) 013011, [1309.4492].

[44] M. Yamashita, T. Doke et al., *Scintillation response of liquid Xe surrounded by PTFE reflector for gamma rays*, *Nucl. Instrum. Meth.* **A535** (2004) 692.

[45] L. Baudis et al., *Performance of the Hamamatsu R11410 Photomultiplier Tube in cryogenic Xenon Environments*, *JINST* **8** (2013) P04026, [1303.0226].

[46] K. Lung et al., *Characterization of the Hamamatsu R11410-10 3-Inch Photomultiplier Tube for Liquid Xenon Dark Matter Direct Detection Experiments*, *Nucl. Instrum. Meth.* **A696** (2012) 32–39, [1202.2628].

[47] XENON collaboration, E. Aprile et al., *Lowering the radioactivity of the photomultiplier tubes for the XENON1T dark matter experiment*, *Eur. Phys. J.* **C75** (2015) 546, [1503.07698].

[48] L. Baudis et al., *Neutrino physics with multi-ton scale liquid xenon detectors*, *JCAP* **01** (2014) 044, [1309.7024].

[49] L. Baudis et al., *Signatures of Dark Matter Scattering Inelastically Off Nuclei*, *Phys. Rev.* **D88** (2013) 115014, [1309.0825].

[50] ZEPLIN-III collaboration, V. Lebedenko et al., *Result from the First Science Run of the ZEPLIN-III Dark Matter Search Experiment*, *Phys. Rev.* **D80** (2009) 052010, [0812.1150].

[51] J. Menendez, D. Gazit and A. Schwenk, *Spin-dependent WIMP scattering off nuclei*, *Phys. Rev.* **D86** (2012) 103511, [1208.1094].

[52] L. Vietze et al., *Nuclear structure aspects of spin-independent WIMP scattering off xenon*, *Phys. Rev.* **D91** (2015) 043520, [1412.6091].

[53] L. E. Strigari and R. Trotta, *Reconstructing WIMP Properties in Direct Detection Experiments Including Galactic Dark Matter Distribution Uncertainties*, *JCAP* **0911** (2009) 019, [0906.5361].

[54] M. Pato et al., *Taming astrophysical bias in direct dark matter searches*, *JCAP* **1302** (2013) 041, [1211.7063].

[55] C. Strege et al., *Fundamental statistical limitations of future dark matter direct detection experiments*, *Phys. Rev.* **D86** (2012) 023507, [1201.3631].

[56] A. Ringwald, *Exploring the Role of Axions and Other WISPs in the Dark Universe*, *Phys. Dark Univ.* **1** (2012) 116, [1210.5081].

[57] M. Pospelov, A. Ritz and M. B. Voloshin, *Bosonic super-WIMPs as keV-scale dark matter*, *Phys. Rev.* **D78** (2008) 115012, [0807.3279].

[58] A. Derevianko et al., *Axio-electric effect*, *Phys. Rev.* **D82** (2010) 065006, [1007.1833].

[59] XENON100 collaboration, E. Aprile et al., *First Axion Results from the XENON100 Experiment*, *Phys. Rev.* **D90** (2014) 062009, [1404.1455].

[60] K. Abe et al., *Search for solar axions in XMASS, a large liquid-xenon detector*, *Phys. Lett. B* **724** (2013) 46–50, [1212.6153].

[61] EDELWEISS collaboration, E. Armengaud et al., *Axion searches with the EDELWEISS-II experiment*, *JCAP* **1311** (2013) 067, [1307.1488].

[62] CDMS collaboration, Z. Ahmed et al., *Search for Axions with the CDMS Experiment*, *Phys. Rev. Lett.* **103** (2009) 141802, [0902.4693].

[63] M. Dine, W. Fischler and M. Srednicki, *A Simple Solution to the Strong CP Problem with a Harmless Axion*, *Phys. Lett.* **B104** (1981) 199.

[64] M. A. Shifman, A. Vainshtein and V. I. Zakharov, *Can Confinement Ensure Natural CP Invariance of Strong Interactions?*, *Nucl. Phys.* **B166** (1980) 493.

[65] K. Arisaka et al., *Expected Sensitivity to Galactic/Solar Axions and Bosonic Super-WIMPs based on the Axio-electric Effect in Liquid Xenon Dark Matter Detectors*, *Astropart. Phys.* **44** (2013) 59, [1209.3810].

[66] Borexino collaboration, G. Bellini et al., *Precision measurement of the ^{7}Be solar neutrino interaction rate in Borexino*, *Phys. Rev. Lett.* **107** (2011) 141302, [1104.1816].

[67] KamLAND collaboration, A. Gando et al., *^{7}Be Solar Neutrino Measurement with KamLAND*, *Phys. Rev.* **C92** (2015) 055808, [1405.6190].

[68] Borexino collaboration, G. Bellini et al., *Neutrinos from the primary proton-proton fusion process in the Sun*, *Nature* **512** (2014) 383–386.

[69] Particle Data Group collaboration, K. A. Olive et al., *Review of Particle Physics*, *Chin. Phys.* **C38** (2014) 090001 and 2015 update.

[70] W. C. Haxton, R. G. Hamish Robertson and A. M. Serenelli, *Solar Neutrinos: Status and Prospects*, *Ann. Rev. Astron. Astrophys.* **51** (2013) 21, [1208.5723].

[71] A. Friedland, C. Lunardini and C. Pena-Garay, *Solar neutrinos as probes of neutrino matter interactions*, *Phys. Lett.* **B594** (2004) 347, [hep-ph/0402266].

[72] M. Maltoni and A. Yu. Smirnov, *Solar neutrinos and neutrino physics*, *Eur. Phys. J.* **A52** (2016) 87, [1507.05287].

[73] S. M. Bilenky and C. Giunti, *Neutrinoless double-beta decay: A brief review*, *Mod. Phys. Lett.* **A27** (2012) 1230015, [1203.5250].

[74] EXO-200 collaboration, J. Albert et al., *Search for Majorana neutrinos with the first two years of EXO-200 data*, *Nature* **510** (2014) 229, [1402.6956].

[75] KamLAND-Zen collaboration, A. Gando et al., *Limit on Neutrinoless $\beta\beta$ Decay of Xe-136 from the First Phase of KamLAND-Zen and Comparison with the Positive Claim in Ge-76*, *Phys. Rev. Lett.* **110** (2013) 062502, [1211.3863].

[76] KamLAND-Zen collaboration, Y. Gando, *Latest results from KamLAND-Zen second phase*, *Nucl. Part. Phys. Proc.* **273-275** (2016) 1842–1846.

[77] N. Barros, J. Thurn and K. Zuber, *Double beta decay searches of ^{134}Xe, ^{126}Xe and ^{124}Xe with large scale Xe detectors*, *J. Phys.* **G41** (2014) 115105, [1409.8308].

[78] L. E. Strigari, *Neutrino Coherent Scattering Rates at Direct Dark Matter Detectors*, *New J. Phys.* **11** (2009) 105011, [0903.3630].

[79] XENON10 collaboration, J. Angle et al., *A search for light dark matter in XENON10 data*, *Phys. Rev. Lett.* **107** (2011) 051301, [1104.3088].

[80] A. Drukier and L. Stodolsky, *Principles and Applications of a Neutral Current Detector for Neutrino Physics and Astronomy*, *Phys. Rev.* **D30** (1984) 2295.

[81] B. Cabrera, L. M. Krauss and F. Wilczek, *Bolometric Detection of Neutrinos*, *Phys. Rev. Lett.* **55** (1985) 25.

[82] C. Giunti and A. Studenikin, *Neutrino electromagnetic interactions: a window to new physics*, *Rev. Mod. Phys.* **87** (2015) 531, [1403.6344].

[83] A. Mirizzi et al., *Supernova Neutrinos: Production, Oscillations and Detection*, *Riv. Nuovo Cim.* **39** (2016) 1, [1508.00785].

[84] C. J. Horowitz, K. J. Coakley and D. N. McKinsey, *Supernova observation via neutrino - nucleus elastic scattering in the CLEAN detector*, *Phys. Rev.* **D68** (2003) 023005, [astro-ph/0302071].

[85] XMASS collaboration, K. Abe et al., *Detectability of galactic supernova neutrinos coherently scattered on xenon nuclei in XMASS*, 1604.01218.

[86] R. Lang et al., *Supernova Neutrino Physics with Xenon Dark Matter Detectors: A Timely Perspective*, *in preparation* (2016) .

[87] XENON100 collaboration, E. Aprile et al., *A low-mass dark matter search using ionization signals in XENON100*, 1605.06262.

[88] P. Antonioli et al., *SNEWS: The Supernova Early Warning System*, *New J. Phys.* **6** (2004) 114, [astro-ph/0406214].

[89] M. Haffke et al., *Background Measurements in the Gran Sasso Underground Laboratory*, *Nucl. Instrum. Meth.* **A643** (2011) 36, [1101.5298].

[90] XENON1T collaboration, E. Aprile et al., *Conceptual design and simulation of a water Cherenkov muon veto for the XENON1T experiment*, *JINST* **9** (2014) 11006, [1406.2374].

[91] G. Heusser, *Low-radioactivity background techniques*, *Ann. Rev. Nucl. Part. Sci.* **45** (1995) 543.

[92] L. Baudis et al., *Gator: a low-background counting facility at the Gran Sasso Underground Laboratory*, *JINST* **6** (2011) P08010, [1103.2125].

[93] G. Heusser, M. Laubenstein and H. Neder, *Low-Level Germanium Gamma-ray Spectrometry at the Bq/kg Level and Future Developments towards higher Sensitivity*, *In Povinec, P.P. and Sanchez-Cabeza, J.A. (eds.): Radionuclides in the Environment* (2006) 495.

[94] G. Heusser et al., *GIOVE - A new detector setup for high sensitivity germanium spectroscopy at shallow depth*, *Eur. Phys. J.* **C75** (2015) 531, [1507.03319].

[95] M. von Sivers et al., *The GeMSE Facility for Low-Background γ-Ray Spectrometry*, 1606.03983.

[96] L. Baudis et al., *Cosmogenic activation of xenon and copper*, *Eur. Phys. J.* **C75** (2015) 485, [1507.03792].

[97] S. Lindemann and H. Simgen, *Krypton assay in xenon at the ppq level using a gas chromatographic system and mass spectrometer*, *Eur. Phys. J.* **C74** (2014) 2746, [1308.4806].

[98] LUX collaboration, D. Akerib et al., *First results from the LUX dark matter experiment at the Sanford Underground Research Facility*, *Phys. Rev. Lett.* **112** (2014) 091303, [1310.8214].

[99] S. Rosendahl, *Gas purification of the XENON dark matter search*, *PhD thesis, University of Münster* (2015) .

[100] D. Stolzenburg, *On the krypton background of the XENON100 and XENON1T dark matter search experiments*, *MSc thesis, University of Heidelberg* (2014) .

[101] W. Rau and G. Heusser, *^{222}Rn emanation measurements at extremely low activities*, *Appl. Rad. Isot.* **53** (2000) 371.

[102] J. Kiko, *Detector for ^{222}Rn measurements in air at the 1 mBq/m^3 level*, *Nucl. Instrum. Meth.* **A460** (2001) 272.

[103] PandaX collaboration, M. Xiao et al., *First dark matter search results from the PandaX-I experiment*, *Sci. China Phys. Mech. Astron.* **57** (2014) 2024–2030, [1408.5114].

[104] J. B. Albert et al., *Investigation of radioactivity-induced backgrounds in EXO-200*, *Phys. Rev.* **C92** (2015) 015503, [1503.06241].

[105] P. Grothaus, M. Fairbairn and J. Monroe, *Directional Dark Matter Detection Beyond the Neutrino Bound*, *Phys. Rev.* **D90** (2014) 055018, [1406.5047].

[106] A. Gutlein et al., *Solar and atmospheric neutrinos: Background sources for the direct dark matter search*, *Astropart. Phys.* **34** (2010) 90, [1003.5530].

[107] A. Anderson et al., *Coherent Neutrino Scattering in Dark Matter Detectors*, *Phys. Rev.* **D84** (2011) 013008, [1103.4894].

[108] M. Messina, *Columbia University R&D program for large mass Dark Matter detector with LXe TPC*, *Talk at TIPP 2014* (2014) .

[109] E. Aprile et al., *Performance of a cryogenic system prototype for the XENON1T Detector*, *JINST* **7** (2012) P10001, [1208.2001].

[110] XENON100 collaboration, E. Aprile et al., *Observation and applications of single-electron charge signals in the XENON100 experiment*, *J. Phys.* **G41** (2014) 035201, [1311.1088].

[111] F. Glueck et al., *Electromagnetic design of the large-volume air coil system of the KATRIN experiment*, *New J. Phys.* **15** (2013) 083025, [1304.6569].

[112] T. Haruyama et al., *LN_2-free operation of the MEG liquid xenon calorimeter by using a high-power pulse tube cryocooler*, *AIP Conf. Proc.* **823** (2006) 1695.

[113] LUX collaboration, A. W. Bradley et al., *LUX Cryogenics and Circulation*, *Phys. Procedia* **37** (2012) 1122–1130.

[114] E. Aprile et al., *Facility and method for supplying liquid xenon*, *patent FR2986061 (A1)* (2013) .

[115] MEG collaboration, J. Adam et al., *The MEG detector for $\mu^+ \to e^+\gamma$ decay search*, *Eur. Phys. J.* **C73** (2013) 2365, [1303.2348].

[116] K. Giboni et al., *Xenon Recirculation-Purification with a Heat Exchanger*, *JINST* **6** (2011) P03002, [1103.0986].

[117] F. LePort et al., *A Magnetically-driven piston pump for ultra-clean applications*, *Rev. Sci. Instrum.* **82** (2011) 105114, [1104.5041].

[118] A. Ferella, *Measuring electron lifetime and V(0) in liquid xenon*, *Nucl. Instrum. Meth.* **A572** (2007) 488.

[119] S. Rosendahl et al., *Determination of the separation efficiencies of a single-stage cryogenic distillation setup to remove krypton out of xenon by using a ^{83m}Kr tracer method*, *Rev. Sci. Instrum.* **86** (2015) 115104.

[120] S. Rosendahl et al., *A cryogenic distillation column for the XENON1T experiment*, *J. Phys. Conf. Ser.* **564** (2014) 012006.

[121] S. Rosendahl et al., *A novel ^{83m}Kr tracer method for characterizing xenon gas and cryogenic distillation systems*, *JINST* **9** (2014) P10010, [1407.3981].

[122] E. Aprile et al., *An atom trap trace analysis system for measuring krypton contamination in xenon dark matter detectors*, *Rev. Sci. Instrum.* **84** (2013) 093105, [1305.6510].

[123] A. Dobi et al., *Detection of krypton in xenon for dark matter applications*, *Nucl. Instrum. Meth.* **A665** (2011) 1–6, [1103.2714].

[124] E. Brown et al., *In situ measurements of Krypton in Xenon gas with a quadrupole mass spectrometer following a cold-trap at a temporarily reduced pumping speed*, *JINST* **8** (2013) P02011, [1212.5136].

[125] C. H. Faham et al., *Measurements of wavelength-dependent double photoelectron emission from single photons in VUV-sensitive photomultiplier tubes*, *JINST* **10** (2015) P09010, [1506.08748].

[126] I. Ostrovskiy et al., *Characterization of Silicon Photomultipliers for nEXO*, *IEEE Trans. Nucl. Sci.* (2015) 1825, [1502.07837].

[127] B. Rossi, *Developments in light readout for noble liquid experiments*, *Talk at TIPP 2014* (2014) .

[128] A. Teymourian et al., *Characterization of the QUartz Photon Intensifying Detector (QUPID) for Noble Liquid Detectors*, *Nucl. Instrum. Meth.* **A654** (2011) 184, [1103.3689].

[129] R. Chechik and A. Breskin, *Advances in Gaseous Photomultipliers*, *Nucl. Instrum. Meth.* **A595** (2008) 116, [0807.2086].

[130] F. Sauli, *GEM: A new concept for electron amplification in gas detectors*, *Nucl. Instrum. Meth.* **A386** (1997) 531.

[131] A. Breskin et al., *A concise review on THGEM detectors*, *Nucl. Instrum. Meth.* **A598** (2009) 107, [0807.2026].

[132] A. Breskin, *CsI UV photocathodes: History and mystery*, *Nucl. Instrum. Meth.* **A371** (1996) 116.

[133] S. Duval et al., *Hybrid Multi Micropattern Gaseous Photomultiplier for detection of liquid-xenon scintillation*, *Nucl. Instrum. Meth.* **A695** (2012) 163, [1110.6053].

[134] S. Duval et al., *On the operation of a Micropattern Gaseous UV-Photomultiplier in Liquid-Xenon*, *JINST* **6** (2011) P04007, [1101.3747].

[135] L. Arazi et al., *First results of a large-area cryogenic gaseous photomultiplier coupled to a dual-phase liquid xenon TPC*, *JINST* **10** (2015) P10020, [1508.00410].

[136] A. Breskin, *Liquid Hole-Multipliers: A potential concept for large single-phase noble-liquid TPCs of rare events*, *J. Phys. Conf. Ser.* **460** (2013) 012020, [1303.4365].

[137] K. Giboni et al., *On Dark Matter detector concepts with large-area cryogenic Gaseous Photo Multipliers*, *JINST* **9** (2014) C02021.

[138] E. Aprile et al., *Measurements of proportional scintillation and electron multiplication in liquid xenon using thin wires*, *JINST* **9** (2014) 11012, [1408.6206].

[139] T. Ye, K. L. Giboni and X. Ji, *Initial evaluation of proportional scintillation in liquid Xenon for direct dark matter detection*, *JINST* **9** (2014) P12007.

[140] T. Doke, *Recent development of liquid xenon detectors*, *Nucl. Instrum. Meth.* **196** (1982) 87.

[141] L. Arazi et al., *First observation of liquid-xenon proportional electroluminescence in THGEM holes*, *JINST* **8** (2013) C12004, [1310.4074].

[142] L. Arazi et al., *Liquid Hole Multipliers: bubble-assisted electroluminescence in liquid xenon*, *JINST* **10** (2015) P08015, [1505.02316].

[143] E. Erdal et al., *Direct observation of bubble-assisted electroluminescence in liquid xenon*, *JINST* **10** (2015) P11002, [1509.02354].

[144] L. Arazi et al., *Simultaneous detection of single photons and ionisation electrons using liquid hole multipliers, in preparation*, *JINST* **11** (2016) P012345, [1602.12345].

[145] E. Erdal et al., *First demonstration of VUV photon detection in liquid xenon with THGEM and GEM-based Liquid Hole Multipliers*, *Nucl. Instrum. Meth. A* (2016) , [1603.07669].

[146] E. Aprile et al., *Performance of CsI photocathodes in liquid Xe, Kr, and Ar*, *Nucl. Instrum. Meth.* **A338** (1994) 328–335.

[147] K. Ni et al., *Preparation of Neutron-activated Xenon for Liquid Xenon Detector Calibration*, *Nucl. Instrum. Meth.* **A582** (2007) 569, [0708.1976].

[148] L. Baudis et al., *Response of liquid xenon to Compton electrons down to 1.5 keV*, *Phys. Rev.* **D87** (2013) 115015, [1303.6891].

[149] L. Kastens et al., *Calibration of a Liquid Xenon Detector with Kr-83m*, *Phys. Rev.* **C80** (2009) 045809, [0905.1766].

[150] A. Manalaysay et al., *Spatially uniform calibration of a liquid xenon detector at low energies using ^{83m}Kr*, *Rev. Sci. Instrum.* **81** (2010) 073303, [0908.0616].

[151] V. Hannen et al., *Limits on the release of Rb isotopes from a zeolite based ^{83m}Kr calibration source for the XENON project*, *JINST* **6** (2011) P10013, [1109.4270].

[152] DarkSide collaboration, P. Agnes et al., *First Results from the DarkSide-50 Dark Matter Experiment at Laboratori Nazionali del Gran Sasso*, *Phys. Lett.* **B743** (2015) 456, [1410.0653].

[153] LUX collaboration, D. S. Akerib et al., *Tritium calibration of the LUX dark matter experiment*, *Phys. Rev.* **D93** (2016) 072009, [1512.03133].

[154] A. Dobi et al., *Study of a zirconium getter for purification of xenon gas*, *Nucl. Instrum. Meth.* **A620** (2010) 594, [1002.2791].

[155] R. F. Lang et al., *A ^{220}Rn source for the calibration of low-background experiments*, *JINST* **11** (2016) P04004, [1602.01138].

[156] M. Kobayashi et al., *Using ^{220}Rn to Calibrate Liquid Noble Gas Detectors*, in *14th International Conference on Topics in Astroparticle and Underground Physics (TAUP 2015) Torino, Italy, September 7-11, 2015*, 2016. 1603.02904.

[157] XENON100 collaboration, E. Aprile et al., *Response of the XENON100 Dark Matter Detector to Nuclear Recoils*, *Phys. Rev.* **D88** (2013) 012006, [1304.1427].

[158] P. Sorensen, *Importance of upgraded energy reconstruction for direct dark matter searches with liquid xenon detectors*, *Phys. Rev.* **D86** (2012) 101301, [1208.5046].

[159] J. Collar, *Applications of an ^{88}Y/Be photo-neutron calibration source to Dark Matter and Neutrino Experiments*, *Phys. Rev. Lett.* **110** (2013) 211101, [1303.2686].

[160] E. Dahl, *The Physics of Background Discrimination in Liquid Xenon, and First Results from XENON10 in the Hunt for WIMP Dark Matter*, *PhD thesis, Princeton University* (2009) .

[161] E. Aprile et al., *New Measurement of the Relative Scintillation Efficiency of Xenon Nuclear Recoils Below 10 keV*, *Phys. Rev.* **C79** (2009) 045807, [0810.0274].

[162] G. Plante et al., *New Measurement of the Scintillation Efficiency of Low-Energy Nuclear Recoils in Liquid Xenon*, *Phys. Rev.* **C84** (2011) 045805, [1104.2587].

[163] A. Manzur et al., *Scintillation efficiency and ionization yield of liquid xenon for mono-energetic nuclear recoils down to 4 keV*, *Phys. Rev.* **C81** (2010) 025808, [0909.1063].

[164] M. Szydagis et al., *NEST: A Comprehensive Model for Scintillation Yield in Liquid Xenon*, *JINST* **6** (2011) P10002, [1106.1613].

[165] M. Szydagis et al., *Enhancement of NEST Capabilities for Simulating Low-Energy Recoils in Liquid Xenon*, *JINST* **8** (2013) C10003, [1307.6601].

[166] E. Aprile et al., *Measurement of the Scintillation Yield of Low-Energy Electrons in Liquid Xenon*, *Phys. Rev.* **D86** (2012) 112004, [1209.3658].

[167] XENON100 collaboration, E. Aprile et al., *Exclusion of Leptophilic Dark Matter Models using XENON100 Electronic Recoil Data*, *Science* **349** (2015) 851–854, [1507.07747].

[168] XENON100 collaboration, E. Aprile et al., *Search for Event Rate Modulation in XENON100 Electronic Recoil Data*, *Phys. Rev. Lett.* **115** (2015) 091302, [1507.07748].

[169] LUX collaboration, D. Akerib et al., *Technical Results from the Surface Run of the LUX Dark Matter Experiment*, *Astropart. Phys.* **45** (2013) 34, [1210.4569].

[170] T. Takahashi et al., *Average energy expended per ion pair in liquid xenon*, *Phys. Rev. A* **12** (1975) 1771.

[171] E. Aprile et al., *Simultaneous Measurement of Ionization and Scintillation from Nuclear Recoils in Liquid Xenon as Target for a Dark Matter Experiment*, *Phys. Rev. Lett.* **97** (2006) 081302, [astro-ph/0601552].

[172] P. Sorensen, *A coherent understanding of low-energy nuclear recoils in liquid xenon*, *JCAP* **1009** (2010) 033, [1007.3549].

Prepared for submission to JCAP

Sound Speed and Viscosity of Semi-Relativistic Relic Neutrinos

Lawrence Krauss[a,b] **and Andrew J. Long**[c]

[a]Physics Department and School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287, USA.

[b]Research School of Astronomy and Astrophysics, Mt. Stromlo Observatory, Australia National University, Canberra, Australia 2611.

[c]Kavli Institute for Cosmological Physics, The University of Chicago, Chicago, Illinois 60637, USA.

E-mail: krauss@asu.edu, andrewjlong@kicp.uchicago.edu

Abstract. Generalized fluid equations, using sound speed $c^2_{\rm eff}$ and viscosity $c^2_{\rm vis}$ as effective parameters, provide a convenient phenomenological formalism for testing the relic neutrino "null hypothesis," *i.e.* that that neutrinos are relativistic and free-streaming prior to recombination. In this work, we relax the relativistic assumption and ask "to what extent can the generalized fluid equations accommodate finite neutrino mass?" We consider both the mass of active neutrinos, which are largely still relativistic at recombination $m^2/T^2 \sim 0.2$, and the effect of a semi-relativistic sterile component. While there is no one-to-one mapping between mass/mixing parameters and $c^2_{\rm eff}$ and $c^2_{\rm vis}$, we demonstrate that the existence of a neutrino mass could induce a bias to measurements of $c^2_{\rm eff}$ and $c^2_{\rm vis}$ at the level of $0.01m^2/T^2 \sim 10^{-3}$.

Keywords: relic neutrinos, sterile neutrino, sound speed, viscosity, cosmic microwave background

Contents

1 Introduction

Precision measurements of the cosmic microwave background (CMB) and large scale structure (LSS) are providing a wealth of information about the early universe and its constituents. This information is particularly valuable in the neutrino sector where a number of fundamental questions have yet to be answered: What is the absolute neutrino mass scale? Are some neutrinos sterile? Do neutrinos self-interact through a long range force? The next-generation of CMB and LSS experiments will bring dramatic improvements in sensitivity and the promise of new insight into the physics of neutrinos [1].

To address the questions listed above in a model-independent way, it is customary to use phenomenological parameters. These parameters are introduced "by hand" into the equations of motion (Einstein or Boltzmann equations). They are not defined by any underlying fundamental parameters, such as Lagrangian couplings or masses.

The most familiar phenomenological parameters are the effective number of neutrino species $N_{\rm eff}$ and the total neutrino mass $\sum m_\nu$. Since the relic neutrinos are decoupled at the time of recombination and structure formation, their effect on the CMB and LSS are only gravitational. Thus, the phenomenological parameters encode how much the neutrinos contribute to the energy densities (see [1, 2] for notation)

$$\rho_{\rm rad} = \rho_\gamma + N_{\rm eff}\frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\rho_\gamma \qquad \text{and} \qquad \Omega_\nu h^2 = \frac{\sum m_\nu}{93.1\ \text{eV}}\ . \tag{1.1}$$

Since $N_{\rm eff}$ and $\sum m_\nu$ are not defined from fundamental parameters, there does not necessarily exist a one-to-one mapping from any specific microphysical model onto the parameters $(N_{\rm eff}, \sum m_\nu)$. Rather, the phenomenological parameters are most useful

as a test of the "null hypothesis." A combination of the concordance cosmology and Standard Model of particle physics predicts $N_{\rm eff} = 3.046$ and $\sum m_\nu = m_1 + m_2 + m_3 > 0.05$ eV where m_i are the three neutrino mass eigenvalues. Measurements compiled by the Planck collaboration [3] (*Planck*, TT + lensing + ext),

$$N_{\rm eff} \simeq 3.15 \pm 0.40 \qquad \text{and} \qquad \sum m_\nu < 0.234 \text{ eV} \quad \text{at } 95\% \text{ CL} \,, \tag{1.2}$$

are consistent with the null hypothesis.

Two additional phenomenological parameters affect the evolution of neutrino density inhomogeneities. These are the effective sound speed $c^2_{\rm eff}$ and viscosity $c^2_{\rm vis}$ [4, 5]. The effective sound speed sets the sound horizon, which in turn controls the growth of neutrino density perturbations, and the viscosity parameter leads to an anisotropic stress and the damping of neutrino density perturbations. (See Refs. [6] for a discussion of these effects on the CMB.) Once again, the phenomenological parameters provide a model-independent formalism to test the null hypothesis: if the relic neutrinos are relativistic and free-streaming then one expects $c^2_{\rm eff} = 1/3$ and $c^2_{\rm vis} = 1/3$. The Planck collaboration furnishes the measurements [3] (*Planck*, TT, TE, EE + lowP + BAO)

$$c^2_{\rm eff} \simeq 0.3242 \pm 0.0059 \qquad \text{and} \qquad c^2_{\rm vis} \simeq 0.331 \pm 0.037 \,, \tag{1.3}$$

which are consistent with the null hypothesis.

As measurements of the four phenomenological parameters improve with the next generation of CMB and LSS experiments, we must be mindful of any deviation from the null hypothesis, as this would indicate the presence of new physics. In order to probe the nature of the new physics, we must understand how a specific microphysical model maps onto the phenomenological parameters. For instance, many studies have investigated how eV-scale sterile neutrinos (motivated in part by the short baseline and reactor anomalies [7, 8]) manifest themselves in the CMB and LSS (for one such recent paper see Ref. [9]). This provides a mapping from the sterile mass and abundance to $N_{\rm eff}$ and $\sum m_\nu$. We seek to extend that correspondence to the perturbation parameters $c^2_{\rm eff}$ and $c^2_{\rm vis}$.

In Sec. 2 we study the formalism (generalized fluid equations) in which the phenomenological parameters $c^2_{\rm eff}$ and $c^2_{\rm vis}$ arise. While this formalism is convenient for testing the null hypothesis, we will see that it cannot generally accommodate realistic deviations from the null hypothesis. Specifically, if the neutrinos are assumed to be free-streaming but allowed to be semi-relativistic (such is the case for sterile neutrinos) then the fluid equations describing their evolution cannot be mapped onto the generalized fluid equations. In Sec. 3 we estimate the dependence of $c^2_{\rm eff}$ and $c^2_{\rm vis}$ on neutrino mass, and we calculate the predicted deviations from the null hypothesis, $1/3 - c^2_{\rm eff}$ and $1/3 - c^2_{\rm vis}$, for a model of sterile neutrinos that saturates the Planck limits in Eq. (1.2). We summarize our results and discuss directions for future work in Sec. 4. The main paper is accompanied by Appendix A, where we derive the fluid equations for a free-streaming species from the Boltzmann hierarchy. Appendix B contains formulas relevant to a semi-relativistic Fermi-Dirac phase space distribution.

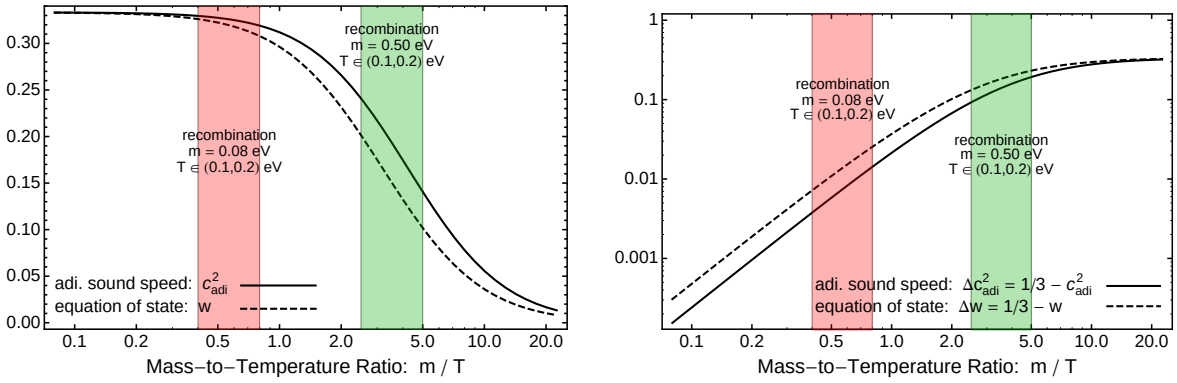

Figure 1. The equation of state w (solid) and squared adiabatic sound speed $c^2_{\rm adi}$ (dashed) for a Fermi-Dirac distribution of relic neutrinos with mass m and temperature T. The shaded regions indicates the epoch of recombination, $T_{\rm rec} \approx 0.1-0.2$ eV, for $m = 0.08$ eV and 0.5 eV.

2 Fluid Equations for Semi-Relativistic, Free-Streaming Neutrinos

We are interested in the background of relic neutrinos at temperatures $T \lesssim 3$ MeV when the weak interactions have frozen out and the neutrinos are decoupled from the plasma. Standard Model neutrinos experience no additional interactions, and they are free-streaming. Later we will extend the model to include eV-scale sterile neutrinos, which are also assumed to be free-streaming.

To leading order, the medium is homogeneous with energy density $\bar{\rho}(\tau)$ and pressure $\bar{P}(\tau)$. The corresponding equation of state and adiabatic sound speed are $w = \bar{P}/\bar{\rho}$ and $c^2_{\rm adi} = \dot{\bar{P}}/\dot{\bar{\rho}}$, where the dot indicates differentiation with respect to conformal time τ. After decoupling the neutrino background maintains its Fermi-Dirac distribution with temperature T. Using the notation established in Appendix A we calculate w and $c^2_{\rm adi}$ in Appendix B and present the result in Fig. 1. For simplicity we assume that the neutrino spectrum is in the degenerate regime, and the common neutrino mass is $m \approx (1/3)\sum m_\nu$. Initially the neutrino temperature is high, $m \ll T$, and $w, c^2_{\rm adi} \approx 1/3$ for the relativistic neutrinos. As the temperature is lowered, the deviations $\Delta w = 1/3 - w$ and $\Delta c^2_{\rm adi} = 1/3 - c^2_{\rm adi}$ start to grow as the neutrinos become semi-relativistic. For a Fermi-Dirac distribution we find

$$\Delta c^2_{\rm adi} \approx \frac{\Delta w}{2} \approx \frac{5}{21\pi^2}\frac{m^2}{T^2} \tag{2.1}$$

for small m/T. The anomalously small prefactor, $5/21\pi^2 \simeq 0.02$, invalidates the naive dimensional analysis prediction $\Delta c^2_{\rm adi} \sim m^2/T^2$.

During recombination, the photon temperature is $T_\gamma \approx 0.2 - 0.3$ eV and the

ℓ	ϵ	q	q^2/ϵ	q^3/ϵ^2	q^4/ϵ^3	$\cdots$
0	δ		Π		$\tilde{\Pi}$	$\cdots$
1		θ		$\tilde{\theta}$		$\cdots$
2			σ		$\tilde{\sigma}$	$\cdots$
3				χ		$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Table 1. The perturbation variables discussed in the text can be organized according the multipole moment ℓ of the phase space distribution function from which they were calculated, and the factors of energy ϵ and momentum q that were included in the momentum integral. Numerical factors of 1/3, etc., are not shown; see Eqs. (A.12) and (A.13) for detailed expressions.

neutrino temperature is smaller by a factor of $(4/11)^{1/3}$, which corresponds to $T_{\rm rec} \approx 0.1 - 0.2$ eV. Taking a fiducial neutrino mass of $m = 0.08$ eV, which saturates the Planck bound in Eq. (1.2), the deviations fall into the range $0.4\% \lesssim \Delta c^2_{\rm adi} \lesssim 2\%$ at the time of recombination. For a heavier eV-scale sterile neutrino, the deviation is $(10 - 20)\%$ (assuming the phase space distribution function is also Fermi-Dirac). The observation that $\Delta w/w \ll 1$ and $\Delta c^2_{\rm adi}/c^2_{\rm adi} \ll 1$ has two implications for our analysis. It indicates that we can study deviations from the relativistic relic neutrino background by perturbing in the small quantities $\Delta w/w$ and $\Delta c^2_{\rm adi}/c^2_{\rm adi}$. Additionally it suggests that the effects of finite neutrino mass will be at most $\sim 0.02 m^2/T^2$ in magnitude.

Let us now consider perturbations to the homogenous Fermi-Dirac distribution. The details of this calculation appear in Appendix A. Since the inhomogenous phase space distribution function depends on both momentum $\boldsymbol{q}$ and position (or wavevector $\boldsymbol{k}$ in Fourier space), it is convenient to organize the perturbations into multipole moments with index ℓ. Each moment of the phase space distribution function can be integrated over momentum $q = |\boldsymbol{q}|$. It is possible to include additional factors of the momentum-to-energy ratio q/ϵ in the integrand. For the lowest order multipole moments ($\ell = 0, 1, 2$) one obtains the energy density contrast $\delta(\boldsymbol{k}, \tau)$, energy flux $\theta(\boldsymbol{k}, \tau)$, and anisotropic stress $\sigma(\boldsymbol{k}, \tau)$. Other combinations of ℓ and q/ϵ lead to a doubly-infinite tower of perturbation variables, shown in Table 1; see Eqs. (A.12) and (A.13) for detailed expressions. Specifically, $\Pi(\boldsymbol{k}, \tau)$ is the pressure perturbation. In the non-relativistic limit, $T/m \ll 1$, the perturbation variables constructed from additional factors of q/ϵ are suppressed by powers of T/m. In the relativistic limit, $T/m \gg 1$, additional factors of the momentum-to-energy ratio simplify $q/\epsilon \approx 1$ leading to

$$\tilde{\Pi} \approx \Pi \approx \delta/3 \,, \quad \tilde{\theta} \approx \theta \,, \quad \text{and} \quad \tilde{\sigma} \approx \sigma \,. \tag{2.2}$$

The evolution of perturbations in a system of freely streaming particles is described by the collisionless Boltzmann equation [10]. Upon performing the multipole expansion described above, the Boltzmann equation yields a hierarchy of coupled first order

differential equations describing the evolution of each moment. See Appendix A for details of this calculation. Focusing on the first few multipole moments, we perform the momentum integrals to obtain

$$\dot{\delta} = -(1+w)\Big(\theta + \frac{1}{2}\dot{h}\Big) + 3\frac{\dot{a}}{a}\Big(w\delta - \Pi\Big) \tag{2.3a}$$

$$\dot{\theta} = -3\frac{\dot{a}}{a}\Big(\frac{1}{3} - c_{\rm adi}^2\Big)\theta + k^2\frac{\Pi}{1+w} - k^2\sigma \tag{2.3b}$$

$$\begin{aligned}\dot{\sigma} = {}& \frac{4}{15}\theta + (3c_{\rm adi}^2)\frac{2}{15}\big(\dot{h} + 6\dot{\eta}\big) - \frac{3}{5}k\chi \\ & - 3\frac{\dot{a}}{a}\Big(\frac{1}{3} - c_{\rm adi}^2\Big)\sigma + \frac{\dot{a}}{a}\big(\tilde{\sigma} - \sigma\big) + \frac{4}{15}\big(\tilde{\theta} - \theta\big)\end{aligned} \tag{2.3c}$$

$$\dot{\Pi} = -3\frac{\dot{a}}{a}\Big(\frac{1}{3} - w\Big)\Pi + \frac{\dot{a}}{a}\big(\tilde{\Pi} - \Pi\big) - \frac{1}{3}(1+w)\tilde{\theta} - \frac{1}{6}\dot{h}\big(5w - \tilde{w}\big)\ , \tag{2.3d}$$

which we call the collisionless fluid equations[1]. We are working in the synchronous gauge where the metric perturbations are denoted as $h(\boldsymbol{k},\tau)$ and $\eta(\boldsymbol{k},\tau)$, and their evolution is given by Einstein's equations. The equations for $\dot{\delta}$ and $\dot{\theta}$ are the familiar continuity and Euler equations[2]. Note that $k = |\boldsymbol{k}|$ and $a(\tau)$ is the FRW scale factor. The parameter $\tilde{w}$ is the pseudo-equation of state, defined in the appendix. The equation for $\dot{\sigma}$ depends on the next moment ($\ell = 3$) in the multipole expansion $\chi(\boldsymbol{k},\tau)$. This is the familiar result for the Boltzmann hierarchy: the evolution of lower-order multipole moments depends on the higher-order moments. In addition, the equations for $\dot{\sigma}$ and $\dot{\Pi}$ also depend on the tilde'd variables $\tilde{\theta}$, $\tilde{\sigma}$, and $\tilde{\Pi}$. Consequently the equations shown in Eq. (2.3) do not form a closed system. However, we are only interested in comparing the form of these equations with the generalized fluid equations below, and for that purpose we do not require the rest of the hierarchy.

In the ultra-relativistic regime, $m \ll T$, we can approximate $w \approx c_{\rm adi}^2 \approx 1/3$. Additionally, the perturbation variables reduce as in Eq. (2.2). Then Eq. (2.3) becomes

$$\dot{\delta} = -\frac{4}{3}\Big(\theta + \frac{1}{2}\dot{h}\Big) \tag{2.4a}$$

$$\dot{\theta} = \frac{1}{4}k^2\delta - k^2\sigma \tag{2.4b}$$

$$\dot{\sigma} = \frac{4}{15}\theta + \frac{2}{15}\big(\dot{h} + 6\dot{\eta}\big) - \frac{3}{5}k\chi\ , \tag{2.4c}$$

and $\Pi = \delta/3$, which are the fluid equations for free-streaming, relativistic particles.

A phenomenological generalization of the fluid equations was proposed in Ref. [4, 5]. By introducing the sound speed and the viscosity parameters, $c_{\rm eff}^2$ and $c_{\rm vis}^2$, one can write

[1] This name is something of an oxymoron. In a perfect fluid, collisions occur frequently and tend to isotropize the perturbations. This enforces a vanishing of the anisotropic stress σ and higher multipole moments. One should view Eq. (2.3) as the analog of the fluid equations for a free-streaming species.

[2] These can also be derived from the conservation of stress-energy [10].

[11][3]

$$\dot{\delta} = -\frac{4}{3}\Big(\theta + \frac{1}{2}\dot{h}\Big) + 3\frac{\dot{a}}{a}\Big(\frac{1}{3} - c_{\rm eff}^2\Big)\delta + 12\Big(\frac{\dot{a}}{a}\Big)^2\Big(\frac{1}{3} - c_{\rm eff}^2\Big)\frac{\theta}{k^2} \tag{2.5a}$$

$$\dot{\theta} = -3\frac{\dot{a}}{a}\Big(\frac{1}{3} - c_{\rm eff}^2\Big)\theta + \frac{1}{4}(3c_{\rm eff}^2)k^2\delta - k^2\sigma \tag{2.5b}$$

$$\dot{\sigma} = (3c_{\rm vis}^2)\frac{4}{15}\theta + (3c_{\rm vis}^2)\frac{2}{15}\big(\dot{h} + 6\dot{\eta}\big) - \frac{3}{5}k\chi \ , \tag{2.5c}$$

which we call the generalized fluid equations (GFE). The rest of the Boltzmann hierarchy, *e.g.* the equation for $\dot{\chi}$, is unmodified. Relativistic and free-streaming neutrinos obey the fluid equations in Eq. (2.4), which corresponds to the limit $c_{\rm eff}^2 = c_{\rm vis}^2 = 1/3$ in the GFE. Therefore, measuring a deviation from 1/3 would refute the "null hypothesis," *i.e.* that the relic neutrinos are relativistic and free-streaming. A number of studies have investigated the effects of $c_{\rm eff}^2$ and $c_{\rm vis}^2$ on the cosmic microwave background [6, 12–22], and recently the Planck collaboration reported the measurements in Eq. (1.3) using a combination of CMB and BAO data. These measurements illustrate the utility of the generalized fluid equations for testing – and thus far confirming – the null hypothesis of relativistic and free-streaming neutrinos.

However, it is not clear the extent to which the GFE is able to capture specific models when we relax the assumptions of relativistic free-streaming particles. For instance, it is often said that $(c_{\rm eff}^2\,,\, c_{\rm vis}^2) = (1/3, 0)$ corresponds to a relativistic perfect fluid, and therefore this limit has been used to model the effect of neutrino self-interactions [23–28] (see also [29]). However, while $c_{\rm vis}^2 = 0$ allows for solutions in which the anisotropic stress and higher moments vanish as in a perfect fluid, it also allows for solutions where they are nonzero and static, which is not the case for a perfect fluid. These criticisms were recently raised by Refs. [21, 28, 30].

In this work, we consider the effect of finite neutrino mass either arising from the active neutrinos themselves or a heavier sterile neutrino component. This problem has been investigated recently in Ref. [6] by numerically solving the Boltzmann hierarchy, and it was found that there is no clear degeneracy between neutrino mass and the sound speed parameters. Our goal is to develop an analytic understanding of this result while also deriving a parametric relationship between the parameters of the GFE and the neutrino mass.

If we relax the assumption of relativistic neutrinos but maintain the assumption of free-streaming neutrinos, then the density perturbations satisfy the collisionless fluid equations of Eq. (2.3). Clearly it is not possible to put the GFE of Eq. (2.5) into the form of Eq. (2.3) even with a judicious choice of the parameters $(c_{\rm eff}^2\,,\, c_{\rm vis}^2)$; the equations have different structures. However, the neutrinos are still semi-relativistic at the time of recombination, see Fig. 1, and this observation motivates us to expand

[3] In comparing with Eqs. (2–4) of Ref. [11], note that $q_\nu(\boldsymbol{k},\tau) = 4\theta(\boldsymbol{k},\tau)/(3k)$ and $\pi_\nu(\boldsymbol{k},\tau) = 2\sigma(\boldsymbol{k},\tau)$ and $F_{\nu,3}(\boldsymbol{k},\tau) = 2\chi(\boldsymbol{k},\tau)$ in the massless limit.

around the relativistic limit. Using the results of Appendix B the equation of state, pseudo-equation of state, and sound speed are written as

$$w = \frac{1}{3} - 2\Delta c_{\rm adi}^2 \ , \quad \tilde{w} = \frac{1}{3} - 4\Delta c_{\rm adi}^2 \ , \quad \text{and} \quad c_{\rm adi}^2 = \frac{1}{3} - \Delta c_{\rm adi}^2 \tag{2.6}$$

where all of the perturbations are proportional to m^2/T^2 and we have used Eq. (2.1). We can similarly expand the perturbation variables around Eq. (2.2) as[4]

$$\begin{aligned} &\Pi = \delta/3 - \Delta\Pi \ , \quad \tilde{\Pi} = \delta/3 - 3\Delta\Pi \ , \\ &\tilde{\theta} = \theta - \Delta\tilde{\theta} \ , \quad \text{and} \quad \tilde{\sigma} = \sigma - \Delta\tilde{\sigma} \end{aligned} \tag{2.7}$$

where the deviations are $O(m^2/T^2)$. Making these replacements the collisionless fluid equations Eq. (2.3) become

$$\dot{\delta} = -\frac{4}{3}\Big(\theta + \frac{1}{2}\dot{h}\Big) + 2\Delta c_{\rm adi}^2\Big(\theta + \frac{1}{2}\dot{h}\Big) - 3\frac{\dot{a}}{a}\Big(2\Delta c_{\rm adi}^2\delta - \Delta\Pi\Big) \tag{2.8a}$$

$$\dot{\theta} = -3\frac{\dot{a}}{a}\Delta c_{\rm adi}^2\theta + \frac{1}{4}k^2\delta - k^2\sigma + \frac{3}{16}k^2\big(2\Delta c_{\rm adi}^2\delta - 4\Delta\Pi\big) \tag{2.8b}$$

$$\dot{\sigma} = \frac{4}{15}\theta + \big(1 - 3\Delta c_{\rm adi}^2\big)\frac{2}{15}\big(\dot{h} + 6\dot{\eta}\big) - \frac{3}{5}k\chi - 3\frac{\dot{a}}{a}\Delta c_{\rm adi}^2\sigma - \frac{\dot{a}}{a}\Delta\tilde{\sigma} - \frac{4}{15}\Delta\tilde{\theta} \ . \tag{2.8c}$$

Here we keep only terms up to linear order in the deviations. In summary, a system of free-streaming particle obeys the collisionless fluid equations of Eq. (2.3), and if the particles are semi-relativistic these equations can be approximated as in Eq. (2.8).

Now we seek to compare Eq. (2.8) with the generalized fluid equations of Eq. (2.5). To facilitate the comparison we difference the two sets of equations to obtain

$$\dot{\delta} : 3\frac{\dot{a}}{a}\Big(\frac{1}{3} - c_{\rm eff}^2 + 2\Delta c_{\rm adi}^2\Big)\delta + 12\Big(\frac{\dot{a}}{a}\Big)^2\Big(\frac{1}{3} - c_{\rm eff}^2\Big)\frac{\theta}{k^2} - 2\Delta c_{\rm adi}^2\Big(\theta + \frac{1}{2}\dot{h}\Big) - 3\frac{\dot{a}}{a}\Delta\Pi \tag{2.9a}$$

$$\dot{\theta} : \ -3\frac{\dot{a}}{a}\Big(\frac{1}{3} - c_{\rm eff}^2 - \Delta c_{\rm adi}^2\Big)\theta + \frac{3}{4}\Big(c_{\rm eff}^2 - \frac{1}{3} - \frac{\Delta c_{\rm adi}^2}{2}\Big)k^2\delta + \frac{3}{4}k^2\Delta\Pi \tag{2.9b}$$

$$\dot{\sigma} : \Big(c_{\rm vis}^2 - \frac{1}{3}\Big)\frac{4}{5}\theta + \Big(c_{\rm vis}^2 - \frac{1}{3} + \Delta c_{\rm adi}^2\Big)\frac{2}{5}\big(\dot{h} + 6\dot{\eta}\big) + 3\frac{\dot{a}}{a}\Delta c_{\rm adi}^2\sigma + \frac{\dot{a}}{a}\Delta\tilde{\sigma} + \frac{4}{15}\Delta\tilde{\theta} \ . \tag{2.9c}$$

Evidently, there is no choice of $c_{\rm eff}^2$ and $c_{\rm vis}^2$ that brings the two expressions into the same form, *i.e.* causes the three lines of Eq. (2.9) to vanish. The generalized fluid equations thus fail to capture even this minor deviation from the null hypothesis.

3 Estimate Deviations from Null Hypothesis

While we have shown that there is no choice of $c_{\rm eff}^2$ and $c_{\rm vis}^2$ for which the generalized fluid equations reduce to the collisionless fluid equations, nevertheless, it is reasonable

[4]Explicit calculation using Eqs. (A.12) and (A.13) reveals that $\delta/3 - \tilde{\Pi} \approx 3(\delta/3 - \Pi)$ to leading order in m^2/T^2.

to ask the following question. Suppose that the neutrinos have a small mass and are semi-relativistic at the time of recombination. This affects the evolution of their density perturbations according to Eq. (2.8) and ultimately impacts the CMB temperature anisotropies. However, suppose one (naively) analyzes the observed CMB data using the generalized fluid equations, Eq. (2.5), which do not capture the physics of the semi-relativistic neutrinos. How will the best fit parameters $c^2_{\rm eff}$ and $c^2_{\rm vis}$ depend on the neutrino mass?

Inspecting Eq. (2.9), we ask what choice of the phenomenological sound speed and viscosity parameters would give the best agreement between the GFE and collisionless fluid equations. Taking $c^2_{\rm vis} = 1/3 - \Delta c^2_{\rm adi}$ causes the gravitational source term to vanish from the equation for $\dot{\sigma}$, and taking $c^2_{\rm eff} = 1/3 - \Delta c^2_{\rm adi}$ causes a number of other terms to exactly or partially cancel. This observation suggests that as the neutrinos start becoming semi-relativistic, the sound speed and viscosity will deviate from the null hypothesis, $(c^2_{\rm eff}, c^2_{\rm vis}) = (1/3, 1/3)$, according to[5]

$$c^2_{\rm eff} \approx c^2_{\rm adi} \qquad \text{and} \qquad c^2_{\rm vis} \approx c^2_{\rm adi} \ . \tag{3.1}$$

While the identification of $c^2_{\rm eff}$ and $c^2_{\rm vis}$ with the adiabatic sound speed is not rigorous, we propose here that it quantitatively reflects the correct parametric behavior and order of magnitude of the effect.

It is interesting to note that *both* $c^2_{\rm eff}$ and $c^2_{\rm vis}$ begin to deviate from 1/3 as the neutrinos become semi-relativistic. This is somewhat surprising, because if we relax only the free streaming assumption, it is possible to describe a relativistic perfect fluid with $(c^2_{\rm eff}, c^2_{\rm vis}) = (1/3, 0)$ in which only $c^2_{\rm vis}$ deviates from its value in the null hypothesis.

One additional comment is in order. Whereas $c^2_{\rm adi}$ is temperature dependent, see Fig. 1, the phenomenological parameters $c^2_{\rm eff}$ and $c^2_{\rm vis}$ are assumed to be static. Thus we should interpret Eq. (3.1) to mean that $c^2_{\rm eff}$ and $c^2_{\rm vis}$ are derived from a weighted time average of $c^2_{\rm adi}$ between the epoch of neutrino decoupling and recombination. Our analytic approximation does not determine which function will appear in the time averaging. However, since $c^2_{\rm adi}$ decreases monotonically from 1/3, any arbitrarily weighted time average must satisfy

$$\Delta c^2_{\rm eff} \ , \ \Delta c^2_{\rm vis} \ \leq \ \Delta c^2_{\rm adi}(T_{\rm rec}) \ , \tag{3.2}$$

where the deviation in the adiabatic sound speed is evaluated at the time of recombination when the neutrino temperature was $T_{\rm rec} \simeq 0.2$ eV. The largest effect of finite neutrino mass on the phenomenological parameters occurs if the inequality is saturated. We will make this assumption for determining our upper limits in numerical estimates below.

[5] A similar identification was employed in the mixed dark matter scenario of Ref. [4].

Using the analytic expression for $\Delta c^2_{\rm adi}$ from Eq. (2.1) we estimate

$$\Delta c^2_{\rm eff} \approx \Delta c^2_{\rm vis} \lesssim 0.01 \frac{m^2}{T^2_{\rm rec}} \ . \tag{3.3}$$

The sum of the relic neutrino masses is constrained as in Eq. (1.2) using Planck data. If the limit is saturated, the neutrinos are in the degenerate regime, and we can take $m \sim 0.08$ eV as a reference point. For this mass, the anticipated deviation in the sound speed and viscosity parameters at the time of recombination are

$$\Delta c^2_{\rm eff} \approx \Delta c^2_{\rm vis} \lesssim 0.002 \left(\frac{m}{0.08 \text{ eV}}\right)^2 \left(\frac{T_{\rm rec}}{0.2 \text{ eV}}\right)^{-2} \ . \tag{3.4}$$

Comparing with Eq. (1.3), we see that the expected deviation is smaller than Planck's sensitivity to $c^2_{\rm eff}$ and $c^2_{\rm vis}$. If the sensitivity to $c^2_{\rm eff}$ improves by an order of magnitude, the estimate of Eq. (3.4) suggests that the effect of finite neutrino mass could become relevant. In that case, a more detailed numerical analysis would be necessary to determine actual constraints..

Next we consider the possibility that the relic neutrino background contains a subdominant component of eV-scale sterile neutrinos. The fact that the neutrinos are sterile, *i.e.* not weakly interacting, will not actually be relevant for this discussion. Rather, it only matters that they are semi-relativistic and free-streaming at the time of recombination. Once again we ask the question: suppose that the CMB sky generated in this model is studied (naively) using the generalized fluid equations, which do not explicitly account for the sterile neutrino component. How will the best fit phenomenological parameters, $c^2_{\rm eff}$ and $c^2_{\rm vis}$, depend on the sterile mass and abundance?

To study this model, one writes down two sets of collisionless fluid equations with each taking the form of Eq. (2.3) but labeled by subscripts "a" for active and "s" for sterile. This significantly complicates the analysis, but we now proceed to argue that one can reduce the system to a single dynamical degree of freedom in the limit where both active and sterile neutrinos are relativistic. Since the neutrinos are free-streaming, they only influence the densities of other species (*e.g.*, photons) through their gravitational effect on the metric perturbations. Einstein's equations, which govern the evolution of the metric perturbations, only depend on the diagonal linear combinations, *e.g.* $\bar{\rho}_{\rm a} + \bar{\rho}_{\rm s}$ and $\delta\rho_{\rm a} + \delta\rho_{\rm s}$ (see Ref. [10] for complete expressions). Thus, as far as Einstein's equations are concerned, we do not need to know the separate evolution of the active and sterile neutrino perturbation variables, but only their sums are relevant:

$$\bar{\rho}_\nu = \bar{\rho}_{\rm a} + \bar{\rho}_{\rm s} \quad , \quad \bar{P}_\nu = \bar{P}_{\rm a} + \bar{P}_{\rm s} \quad , \quad \delta\rho_\nu = \delta\rho_{\rm a} + \delta\rho_{\rm s} \tag{3.5}$$

and so on for the other perturbation variables, θ_ν, σ_ν, etc. In this way, we can model the combined active and sterile neutrino background as a two-component fluid. The

corresponding adiabatic sound speed is given by

$$c^2_{\mathrm{adi},\nu} = \frac{\dot{\bar{P}}_\nu}{\dot{\bar{\rho}}_\nu} = \frac{c^2_{\mathrm{adi,a}}(1+w_\mathrm{a})\bar{\rho}_\mathrm{a} + c^2_{\mathrm{adi,s}}(1+w_\mathrm{s})\bar{\rho}_\mathrm{s}}{(1+w_\mathrm{a})\bar{\rho}_\mathrm{a} + (1+w_\mathrm{s})\bar{\rho}_\mathrm{s}} \tag{3.6}$$

where we have used Eq. (A.8). In the subsequent analysis we will assume, as above, that the effective sound speed and viscosity will follow the adiabatic sound speed as the neutrinos become semi-relativistic.

Note that while the sound speed formula bears a similarity to the baryon-photon fluid, the physics is very different. Before recombination the baryons and photons are tightly coupled due to frequent Thompson scattering [10]. Consequently, the baryon perturbation variables tend to track the photon perturbation variables, *e.g.* $\theta_\gamma \approx \theta_\mathrm{b}$ and $\sigma_\gamma \approx \sigma_\mathrm{b} \approx 0$, and the single coupled fluid evolves as if it had an adiabatic sound speed given by the analog of Eq. (3.6). In the case of free-streaming neutrinos, on the other hand, the active and sterile perturbations are not directly coupled. However, in the relativistic regime, $m_\mathrm{a}, m_\mathrm{s} \ll T$, the two sets of Boltzmann equations describing the evolution of the active and sterile neutrinos are reduced to the same form, *i.e.* $w_\mathrm{a} \approx w_\mathrm{s} \approx 1/3$ and $c^2_{\mathrm{adi,a}} \approx c^2_{\mathrm{adi,s}} \approx 1/3$. If the isocurvature modes vanish initially, *e.g.* $\theta_\mathrm{a} \approx \theta_\mathrm{s}$, then they remain vanishing as long at the both species are ultra-relativistic. Consequently the active and sterile neutrino perturbations evolve in the same way, even through they are not directly coupled, and they can be modeled as a single fluid[6]. Once the sterile neutrinos become non-relativistic, the isocurvature modes will grow, and the two species will start evolving differently. Until that time, in the semi-relativistic regime, the sound speed given in Eq. (3.6) is appropriate.

Further, we assume that the sterile neutrinos have a phase space distribution function of the Fermi-Dirac form with the same temperature as the active neutrinos but a different overall normalization:

$$f_{0,\mathrm{a}}(q) = \frac{g}{(2\pi)^3}\frac{1}{e^{q/aT}+1} \qquad \text{and} \qquad f_{0,\mathrm{s}}(q) = \alpha\, f_{0,\mathrm{a}}(q)\ . \tag{3.7}$$

The proportionality constant α controls the relative number densities, $\bar{n}_\mathrm{s} = \alpha\,\bar{n}_\mathrm{a}$. In the relativistic limit this proportionality implies $\bar{\rho}_\mathrm{s}/\bar{\rho}_\mathrm{a} \approx \alpha$, and Eq. (1.1) gives

$$\Delta N_\mathrm{eff} \approx \frac{8}{7}\left(\frac{11}{4}\right)^{4/3}\frac{\bar{\rho}_\mathrm{s}}{\rho_\gamma} \approx 3\alpha\ . \tag{3.8}$$

In the non-relativistic limit the proportionality implies

$$\Delta\sum m_\nu \approx m_\mathrm{s}\frac{\bar{n}_\mathrm{s}}{\bar{n}_\mathrm{a}} \approx \alpha\, m_\mathrm{s}\ . \tag{3.9}$$

The effective number of neutrinos is measured with an error of $\delta N_\mathrm{eff} \approx 0.4$, see Eq. (1.2), which implies $\bar{\rho}_\mathrm{s}/\bar{\rho}_\mathrm{a} = \alpha \lesssim \delta N_\mathrm{eff}/3 \approx 0.1$. Similarly, imposing the bound on $\sum m_\nu$ implies $m_\mathrm{s} \lesssim (0.2\ \mathrm{eV})/\alpha \approx 2\ \mathrm{eV}$ for $\alpha \approx 0.1$.

[6] One makes a similar reduction when modeling the Standard Model relic neutrino background as a single fluid, even though it is composed of three non-interacting components, corresponding to the three neutrino mass eigenstates.

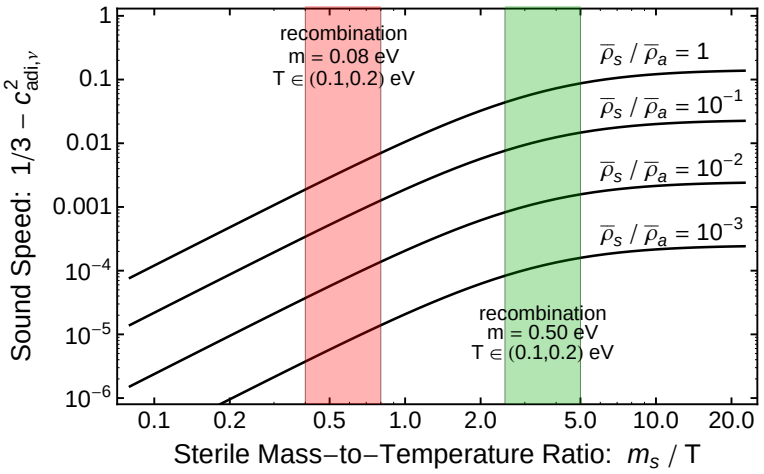

Figure 2. The relic neutrino adiabatic sound speed from Eq. (3.6). Varying the sterile neutrino mass, $m_{\rm s}$, affects the equation of state and sound speed, $w_{\rm s}$ and $c^2_{\rm adi,s}$, which appear in Eq. (3.6). The three lines correspond to different values of the sterile-to-active energy ratios, $\bar{\rho}_{\rm s}/\bar{\rho}_{\rm a} = 1, 10^{-1}, 10^{-2}$, and 10^{-3} from top to bottom, which is a proxy for $\Delta N_{\rm eff}/3$.

We evaluate the relic neutrino adiabatic sound speed using Eq. (3.6). The active neutrinos are still relativistic at recombination, and we can set $w_{\rm a} \approx c^2_{\rm adi,a} \approx 1/3$. The sterile equation of state and sound speed, $w_{\rm s}$ and $c^2_{\rm adi,s}$, are calculated from Eq. (3.7); they vary with the sterile neutrino mass $m_{\rm s}$ as shown in Fig. 1. Figures 2 and 3 show how the sound speed deviation $\Delta c^2_{{\rm adi},\nu} = 1/3 - c^2_{{\rm adi},\nu}$ varies with the sterile neutrino mass $m_{\rm s}$ and relative abundance $\bar{\rho}_{\rm s}/\bar{\rho}_{\rm a}$. If the sterile neutrino is sufficiently light, then it is still relativistic at recombination, and its effect on the sound speed is small. Similarly, if the relative sterile abundance is small, $\alpha = \bar{\rho}_{\rm s}/\bar{\rho}_{\rm a} \ll 1$, then it also has a suppressed impact on the sound speed. As the sterile neutrino mass is increased, the sound speed begins to deviate further from the null hypothesis value, $c^2_{{\rm adi},\nu} = 1/3$. However, if the mass is so large that the sterile neutrino is non-relativistic at recombination, then the approximations used in our calculation are no longer valid, which is why we cut off Figures 2 at $m_s/T \approx 20$. Fig. 3 also indicates the parameter space that is excluded by bounds on $\sum m_\nu$ and $\Delta N_{\rm eff}$ as the red and blue shaded regions, respectively. Focusing on $m_{\rm s} \approx T_\nu \simeq 0.2$ eV we see that the sound speed can deviate from the null hypothesis by as much as $O(10^{-3})$ before running into the bound on $N_{\rm eff}$.

4 Conclusion and Discussion

Generalized fluid equations (GFEs) provide a phenomenological formalism for testing the relic neutrino "null hypothesis," *i.e.* that the neutrinos are both relativistic and free-streaming in the epoch prior to recombination. This formalism has two key advan-

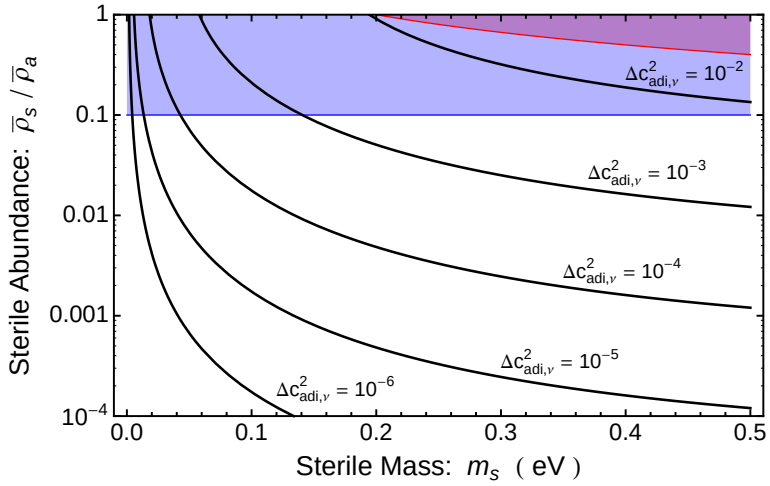

Figure 3. The relic neutrino adiabatic sound speed from Eq. (3.6). The figure is calculated in the same way as Fig. 2, but we hold $T = 0.2$ eV fixed. Curves of constant $\Delta c^2_{\mathrm{adi},\nu} = 1/3 - c^2_{\mathrm{adi},\nu}$ are shown. The red shaded region is excluded by the bound on $\sum m_\nu$, and the blue shaded region is excluded by the bound on ΔN_{eff}.

tages: it requires only minimal modifications to the fluid equations, which are easily implemented in a numerical Boltzmann solver; and these modifications are captured by just two parameters, the effective sound speed c^2_{eff} and the viscosity c^2_{vis}, which quantify deviations from the null hypothesis. However, as we have demonstrated in Sec. 2, where we consider the effects of finite neutrino masses, specific microphysical models that deviate from the null hypothesis cannot always be accommodated into the GFE formalism.

One can nevertheless investigate how the presence of a finite neutrino mass would affect the best fit values of c^2_{eff} and c^2_{vis} if the CMB sky were analyzed using a GFE analysis. In Sec. 3 we propose that one can use the adiabatic sound speed at recombination, $c^2_{\mathrm{adi}}(T_{\mathrm{rec}})$, to gauge the magnitude of deviations in c^2_{eff} and c^2_{vis} from 1/3, the null hypothesis prediction. Taking this as our measure, we estimate that a value of the neutrino mass saturating the Planck limit, $m \simeq 0.08$ eV, could induce a deviation in the effective sound speed and viscosity by as much as 0.2%. If the relic neutrino background contains a sterile component, similar estimates suggest that deviations could be as large as $\Delta c^2_{\mathrm{eff}} \approx \Delta c^2_{\mathrm{vis}} \approx 10^{-3}$ if Plank's limits on ΔN_{eff} and $\sum m_\nu$ are saturated. Since Planck's error bars on the phenomenological parameters are relatively large, $\delta c^2_{\mathrm{eff}} \simeq 0.0059$ and $\delta c^2_{\mathrm{vis}} \simeq 0.037$ respectively, the effect of finite neutrino mass is currently imperceptible. However, if the next generation of CMB telescopes achieve an order of magnitude improvement in sensitivity to the GFE parameters, then our estimates suggest that the effect of finite neutrino mass cannot be neglected, and analytical and phenomenological approximations will need to be supplemented by detailed numerical estimates.

In this regard we conclude by noting that our numerical estimates of Sec. 3 rely on the plausible argument that the best fit values of $c^2_{\rm eff}$ and $c^2_{\rm vis}$ will begin to deviate from 1/3 in the same way as $c^2_{\rm adi}$. If experimental sensitivities improve sufficiently one could test this ansatz in detail using a numerical Boltzmann solver and the following algorithm[7]: for a particular neutrino mass solve the full Boltzmann hierarchy, Eq. (A.4), to generate realizations of the CMB sky; then for a particular $(c^2_{\rm eff}, c^2_{\rm vis})$ solve the generalized fluid equations Eq. (2.5); using MCMC techniques find the values of $(c^2_{\rm eff}, c^2_{\rm vis})$ that best fit the sky generated from the Boltzmann hierarchy.

Acknowledgments

We are grateful for discussions with Sean Bryan, Daniel Grin, Wayne Hu, and Marius Millea, and we are grateful to Cecilia Lunardini for collaboration in the initial stages of this project. This work was supported at Arizona State University by the Department of Energy under Grant No. DE-SC0008016 and the National Science Foundation under Grant No. PHY-1205745. This work was supported at the University of Chicago by the Kavli Institute for Cosmological Physics through grant NSF PHY-1125897 and an endowment from the Kavli Foundation and its founder Fred Kavli.

A Derivation of Fluid Equations

In this appendix we derive the fluid equations from the Boltzmann hierarchy. Our notation mostly follows Ma & Bertschinger [10]: $a(\tau)$ is the scale factor, $\mathrm{d}\tau = \mathrm{d}t/a(\tau)$ is the conformal time, $\mathrm{d}\boldsymbol{x} = \mathrm{d}\boldsymbol{r}/a(\tau)$ is the comoving coordinate, $\boldsymbol{k}$ is the corresponding wave vector, $\boldsymbol{q} = a(\tau)\boldsymbol{p}$ is the comoving momentum, and $\epsilon(\tau) = \sqrt{q^2 + a(\tau)^2 m^2}$ is $a(\tau)$ times the proper energy measured by a comoving observer.

The phase space distribution function is written as

$$f(\boldsymbol{k}, \boldsymbol{q}, \tau) = f_0(q, \tau)\big(1 + \Psi(\boldsymbol{k}, \boldsymbol{q}, \tau)\big) \ . \tag{A.1}$$

For freely streaming particles, f satisfies the collisionless Boltzmann equation. For the homogenous term[8] this is simply $\partial f_0(q,\tau)/\partial\tau = 0$, and the perturbations satisfy

$$\frac{\partial\Psi}{\partial\tau} + i\frac{qk}{\epsilon}(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}})\Psi + \frac{d\ln f_0}{d\ln q}\left(\dot\eta - \frac{\dot h + 6\dot\eta}{2}(\hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}})^2\right) = 0 \ , \tag{A.2}$$

[7] The algorithm we outline here is similar in spirit to the approach taken by Refs. [31, 32] to infer the effect of neutrino mass on $N_{\rm eff}$.

[8] The analysis in this appendix does not assume any specific form for $f_0(q,\tau)$.

which has been written here in synchronous gauge (with $\eta(\boldsymbol{k},\tau)$ and $h(\boldsymbol{k},\tau)$ the metric perturbations). The perturbation is decomposed onto the Legendre polynomials as

$$\Psi(\boldsymbol{k},\boldsymbol{q},\tau) = \sum_{l=0}^{\infty}(-i)^l(2l+1)\,\Psi_l(\boldsymbol{k},q,\tau)\,P_l(\mu) \tag{A.3}$$

with $\mu = \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}$. The Boltzmann equation resolves to the set of coupled, first-order differential equations

$$\dot{\Psi}_0 = -\frac{qk}{\epsilon}\Psi_1 + \frac{1}{6}\frac{d\ln f_0}{d\ln q}\dot{h} \tag{A.4a}$$

$$\dot{\Psi}_1 = \frac{qk}{3\epsilon}\Big(\Psi_0 - 2\Psi_2\Big) \tag{A.4b}$$

$$\dot{\Psi}_2 = \frac{qk}{5\epsilon}\Big(2\Psi_1 - 3\Psi_3\Big) - \frac{1}{15}\frac{d\ln f_0}{d\ln q}\big(\dot{h}+6\dot{\eta}\big) \tag{A.4c}$$

$$\dot{\Psi}_l = \frac{1}{2l+1}\frac{qk}{\epsilon}\Big(l\Psi_{l-1} - (l+1)\Psi_{l+1}\Big) \qquad \text{for } l \geq 3\ , \tag{A.4d}$$

which are collectively known as the Boltzmann hierarchy.

If one is not interested in the momentum dependence of the perturbations, it would seem that the problem is simplified by integrating Eq. (A.4) over q. In the case of massless particles ($\epsilon = q$) one can identify a new dynamical variable $F_l(\boldsymbol{k},\tau) \propto \int_0^\infty q^2\mathrm{d}q\,q f_0(q,\tau)\Psi_l(\boldsymbol{k},q,\tau)$ for each original Ψ_l, and in fact, the problem is simplified. However in the massive case ($\epsilon \neq q$) the number of dynamical variables increases. For instance, Eq. (A.4a) gives the evolution of $A_0 = \int \epsilon\Psi_0$ in terms of $B_1 = \int q\Psi_1$ (written schematically), but Eq. (A.4b) gives the evolution of B_1 in terms of $C_0 = \int(q^2/\epsilon)\Psi_0$. This second moment of Ψ_0 requires its own evolution equation, and thus it is typically easier to solve Eq. (A.4) directly. Nevertheless, the first few equations obtained by integrating Eq. (A.4) correspond to the familiar fluid equations, and we now proceed to derive them.

First we define the spatially averaged energy density, pressure, and pseudo-pressure:

$$\bar{\rho}(\tau) = 4\pi a(\tau)^{-4}\int_0^\infty q^2\mathrm{d}q\,f_0(q,\tau)\,\epsilon(q,\tau) \tag{A.5}$$

$$\bar{P}(\tau) = 4\pi a(\tau)^{-4}\int_0^\infty q^2\mathrm{d}q\,f_0(q,\tau)\,\frac{q^2}{3\epsilon(q,\tau)} \tag{A.6}$$

$$\tilde{P}(\tau) = 4\pi a(\tau)^{-4}\int_0^\infty q^2\mathrm{d}q\,f_0(q,\tau)\,\frac{q^4}{3\epsilon(q,\tau)^3}\ . \tag{A.7}$$

For freely streaming particles ($\partial f_0/\partial\tau = 0$), the energy density satisfies the homogenous continuity equation

$$\dot{\bar{\rho}}(\tau) = -3\frac{\dot{a}}{a}\big(\bar{\rho}+\bar{P}\big)\ , \tag{A.8}$$

and the pressure satisfies

$$\dot{\bar{P}}(\tau) = -\frac{\dot{a}}{a}\big(5\bar{P} - \tilde{P}\big) \ . \tag{A.9}$$

We define the equation of state, pseudo-equation of state, and adiabatic sound speed,

$$w(\tau) = \frac{\bar{P}(\tau)}{\bar{\rho}(\tau)} \ , \quad \tilde{w}(\tau) = \frac{\tilde{P}(\tau)}{\bar{\rho}(\tau)} \ , \quad \text{and} \quad c^2_{\text{adi}}(\tau) = \frac{\dot{\bar{P}}(\tau)}{\dot{\bar{\rho}}(\tau)} \ . \tag{A.10}$$

They obey the useful relations

$$\frac{\dot{w}}{1+w} = 3\frac{\dot{a}}{a}\big(w - c^2_{\text{adi}}\big) \quad \text{and} \quad c^2_{\text{adi}} = \frac{5w - \tilde{w}}{3(1+w)} \ . \tag{A.11}$$

Next we define a few of the lower order moments of Ψ_l as[9]

$$\delta\rho(\boldsymbol{k},\tau) = 4\pi a(\tau)^{-4} \int_0^\infty q^2 \mathrm{d}q \, f_0(q,\tau) \Psi_0(\boldsymbol{k},q,\tau) \, \epsilon(q,\tau) \tag{A.12a}$$

$$\delta P(\boldsymbol{k},\tau) = 4\pi a(\tau)^{-4} \int_0^\infty q^2 \mathrm{d}q \, f_0(q,\tau) \Psi_0(\boldsymbol{k},q,\tau) \, \frac{q^2}{3\epsilon(q,\tau)} \tag{A.12b}$$

$$\delta \tilde{P}(\boldsymbol{k},\tau) = 4\pi a(\tau)^{-4} \int_0^\infty q^2 \mathrm{d}q \, f_0(q,\tau) \Psi_0(\boldsymbol{k},q,\tau) \, \frac{q^4}{3\epsilon(q,\tau)^3} \tag{A.12c}$$

$$\delta Q(\boldsymbol{k},\tau) = 4\pi a(\tau)^{-4} \int_0^\infty q^2 \mathrm{d}q \, f_0(q,\tau) \Psi_1(\boldsymbol{k},q,\tau) \, qk \tag{A.12d}$$

$$\delta \tilde{Q}(\boldsymbol{k},\tau) = 4\pi a(\tau)^{-4} \int_0^\infty q^2 \mathrm{d}q \, f_0(q,\tau) \Psi_1(\boldsymbol{k},q,\tau) \, \frac{q^3 k}{\epsilon(q,\tau)^2} \tag{A.12e}$$

$$\delta S(\boldsymbol{k},\tau) = 4\pi a(\tau)^{-4} \int_0^\infty q^2 \mathrm{d}q \, f_0(q,\tau) \Psi_2(\boldsymbol{k},q,\tau) \, \frac{2q^2}{3\epsilon(q,\tau)} \tag{A.12f}$$

$$\delta \tilde{S}(\boldsymbol{k},\tau) = 4\pi a(\tau)^{-4} \int_0^\infty q^2 \mathrm{d}q \, f_0(q,\tau) \Psi_2(\boldsymbol{k},q,\tau) \, \frac{2q^4}{3\epsilon(q,\tau)^3} \tag{A.12g}$$

$$\delta C(\boldsymbol{k},\tau) = 4\pi a(\tau)^{-4} \int_0^\infty q^2 \mathrm{d}q \, f_0(q,\tau) \Psi_3(\boldsymbol{k},q,\tau) \, \frac{2q^3}{3\epsilon(q,\tau)^2} \ . \tag{A.12h}$$

These correspond to perturbations in the energy density $\delta\rho$, pressure δP, pseudo-pressure $\delta\tilde{P}$, energy flux δQ, anisotropic stress δS, etc. We can also write

$$\delta\rho(\boldsymbol{k},\tau) = \bar{\rho}(\tau) \, \delta(\boldsymbol{k},\tau) \tag{A.13a}$$

$$\delta P(\boldsymbol{k},\tau) = \bar{\rho}(\tau) \, \Pi(\boldsymbol{k},\tau) \tag{A.13b}$$

$$\delta \tilde{P}(\boldsymbol{k},\tau) = \bar{\rho}(\tau) \, \tilde{\Pi}(\boldsymbol{k},\tau) \tag{A.13c}$$

$$\delta Q(\boldsymbol{k},\tau) = \big(1 + w(\tau)\big) \, \bar{\rho}(\tau) \, \theta(\boldsymbol{k},\tau) \tag{A.13d}$$

$$\delta \tilde{Q}(\boldsymbol{k},\tau) = \big(1 + w(\tau)\big) \, \bar{\rho}(\tau) \, \tilde{\theta}(\boldsymbol{k},\tau) \tag{A.13e}$$

$$\delta S(\boldsymbol{k},\tau) = \big(1 + w(\tau)\big) \, \bar{\rho}(\tau) \, \sigma(\boldsymbol{k},\tau) \tag{A.13f}$$

$$\delta \tilde{S}(\boldsymbol{k},\tau) = \big(1 + w(\tau)\big) \, \bar{\rho}(\tau) \, \tilde{\sigma}(\boldsymbol{k},\tau) \tag{A.13g}$$

$$\delta C(\boldsymbol{k},\tau) = \big(1 + w(\tau)\big) \, \bar{\rho}(\tau) \, \chi(\boldsymbol{k},\tau) \ , \tag{A.13h}$$

[9] Here our notation diverges from that of Ma & Bertschinger [10].

which defines the dimensionless perturbation variables δ, Π, etc. For massless particles ($\epsilon = q$) we have $w = \tilde{w} = c_{\rm adi}^2 = 1/3$, and the higher order moments simplify to the lower order ones, *e.g.* $\delta\tilde{Q} = \delta Q$, $\tilde{\theta} = \theta$, $\tilde{P} = P$, and so on.

Finally we are prepared to derive the fluid equations from the Boltzmann hierarchy. Taking the appropriately-weighted momentum integral of Eq. (A.4a) leads to the inhomogenous continuity equation, which can be written in three equivalent forms:

$$\dot{\delta\rho} = -\delta Q - \frac{1}{2}\dot{h}\left(\bar{\rho} + \bar{P}\right) - 3\frac{\dot{a}}{a}\left(\delta\rho + \delta P\right) \tag{A.14a}$$

$$\dot{\delta} = -(1+w)\left(\theta + \frac{1}{2}\dot{h}\right) + 3\frac{\dot{a}}{a}\left(w\delta - \Pi\right) \tag{A.14b}$$

$$\left(\frac{\delta}{1+w}\right)^{\cdot} = -\left(\theta + \frac{1}{2}\dot{h}\right) + 3\frac{\dot{a}}{a}\left(c_{\rm adi}^2\frac{\delta}{1+w} - \frac{\Pi}{1+w}\right) \ . \tag{A.14c}$$

The relation $\Pi = (\delta P/\delta\rho)\delta$ puts the second equation into a more familiar form. Using a different weighting in the momentum integral yields,

$$\dot{\delta P} = -5\frac{\dot{a}}{a}\delta P + \frac{\dot{a}}{a}\delta\tilde{P} - \frac{1}{3}\delta\tilde{Q} - \frac{1}{6}\dot{h}\left(5\bar{P} - \tilde{P}\right) \tag{A.15a}$$

$$\dot{\Pi} = -3\frac{\dot{a}}{a}\left(\frac{1}{3} - w\right)\Pi + \frac{\dot{a}}{a}\left(\tilde{\Pi} - \Pi\right) - \frac{1}{3}(1+w)\tilde{\theta} - \frac{1}{6}\dot{h}\left(5w - \tilde{w}\right) \tag{A.15b}$$

$$\left(\frac{\Pi}{1+w}\right)^{\cdot} = -3\frac{\dot{a}}{a}\left(\frac{1}{3} - c_{\rm adi}^2\right)\frac{\Pi}{1+w} + \frac{\dot{a}}{a}\frac{\tilde{\Pi} - \Pi}{1+w} - \frac{1}{3}\tilde{\theta} - \frac{1}{2}c_{\rm adi}^2\dot{h} \ , \tag{A.15c}$$

which gives the evolution of the momentum perturbation. Integrating Eq. (A.4b) leads to the Euler equation,

$$\dot{\delta Q} = -4\frac{\dot{a}}{a}\delta Q + k^2\delta P - k^2\delta S \tag{A.16a}$$

$$\dot{\theta} = -3\frac{\dot{a}}{a}\left(\frac{1}{3} - c_{\rm adi}^2\right)\theta + k^2\frac{\Pi}{1+w} - k^2\sigma \ , \tag{A.16b}$$

and integrating Eq. (A.4c) gives the shear equation,

$$\dot{\delta S} = -5\frac{\dot{a}}{a}\delta S + \frac{\dot{a}}{a}\delta\tilde{S} + \frac{4}{15}\delta\tilde{Q} + \frac{2}{15}(\dot{h} + 6\dot{\eta})\left(5\bar{P} - \tilde{P}\right) + \frac{3}{5}k\delta C \tag{A.17a}$$

$$\dot{\sigma} = -3\frac{\dot{a}}{a}\left(\frac{1}{3} - c_{\rm adi}^2\right)\sigma + \frac{\dot{a}}{a}\left(\tilde{\sigma} - \sigma\right) + \frac{4}{15}\tilde{\theta} + \frac{2}{5}c_{\rm adi}^2(\dot{h} + 6\dot{\eta}) - \frac{3}{5}k\chi \ . \tag{A.17b}$$

Equations (A.14)–(A.17) do not form a closed system, since the evolution of $\tilde{\Pi}$, $\tilde{\theta}$, $\tilde{\sigma}$, and χ are undetermined.

B Fermi-Dirac Distribution

For the relic neutrinos, which decoupled while they were relativistic, $f_0(q,\tau)$ maintains the Fermi-Dirac distribution

$$f_0 = \frac{g}{(2\pi)^3}\frac{1}{e^{q/aT}+1} \tag{B.1}$$

where $aT = a_0T_0$ is independent of τ and $g = 6$ counts the two spin and three flavor degrees of freedom. The energy density, pressure, and pseudo-pressure are calculated from Eqs. (A.5), (A.6), and (A.7) with $\epsilon = \sqrt{q^2 + a(\tau)^2m^2}$. In the limit $m^2/T^2 \ll 1$ the integrals can be evaluated analytically, and we find

$$\bar{\rho}(\tau) \approx \frac{7}{240}g\pi^2T^4 + \frac{g}{48}m^2T^2 \tag{B.2}$$

$$\bar{P}(\tau) \approx \frac{7}{720}g\pi^2T^4 - \frac{g}{144}m^2T^2 \tag{B.3}$$

$$\tilde{P}(\tau) \approx \frac{7}{720}g\pi^2T^4 - \frac{g}{48}m^2T^2 \tag{B.4}$$

up to terms that are $O(m^4)$. The equation of state, pseudo-equation of state, and adiabatic sound speed are calculated using Eq. (A.10). The exact expressions, determined numerically, are shown in Fig. 1. In the limit $m^2/T^2 \ll 1$ we can approximate

$$w \approx \frac{1}{3} - \frac{10}{21\pi^2}\frac{m^2}{T^2} \tag{B.5}$$

$$\tilde{w} \approx \frac{1}{3} - \frac{20}{21\pi^2}\frac{m^2}{T^2} \tag{B.6}$$

$$c_{\rm adi}^2 \approx \frac{1}{3} - \frac{5}{21\pi^2}\frac{m^2}{T^2}\ , \tag{B.7}$$

up to terms of order $O(m^4/T^4)$. These expressions give Eq. (2.1).

References

[1] **Topical Conveners: K.N. Abazajian, J.E. Carlstrom, A.T. Lee** Collaboration, K. N. Abazajian et al., *Neutrino Physics from the Cosmic Microwave Background and Large Scale Structure*, Astropart. Phys. **63** (2015) 66–80, [arXiv:1309.5383].

[2] G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti, and P. D. Serpico, *Relic neutrino decoupling including flavor oscillations*, Nucl. Phys. **B729** (2005) 221–234, [hep-ph/0506164].

[3] **Planck** Collaboration, P. A. R. Ade et al., *Planck 2015 results. XIII. Cosmological parameters*, arXiv:1502.0158.

[4] W. Hu, *Structure formation with generalized dark matter*, Astrophys. J. **506** (1998) 485–494, [astro-ph/9801234].

[5] W. Hu, D. J. Eisenstein, M. Tegmark, and M. J. White, *Observationally determining the properties of dark matter*, Phys. Rev. **D59** (1999) 023512, [astro-ph/9806362].

[6] B. Audren et al., *Robustness of cosmic neutrino background detection in the cosmic microwave background*, JCAP **1503** (2015) 036, [arXiv:1412.5948].

[7] K. N. Abazajian et al., *Light Sterile Neutrinos: A White Paper*, arXiv:1204.5379.

[8] C. Giunti, *Light Sterile Neutrinos*, 2015. arXiv:1512.0475.

[9] M. Costanzi, B. Sartoris, M. Viel, and S. Borgani, *Neutrino constraints: what large-scale structure and CMB data are telling us?*, JCAP **1410** (2014), no. 10 081, [arXiv:1407.8338].

[10] C.-P. Ma and E. Bertschinger, *Cosmological perturbation theory in the synchronous and conformal Newtonian gauges*, Astrophys. J. **455** (1995) 7–25, [astro-ph/9506072].

[11] M. Archidiacono, E. Calabrese, and A. Melchiorri, *The Case for Dark Radiation*, Phys. Rev. **D84** (2011) 123008, [arXiv:1109.2767].

[12] R. Trotta and A. Melchiorri, *Indication for primordial anisotropies in the neutrino background from WMAP and SDSS*, Phys. Rev. Lett. **95** (2005) 011305, [astro-ph/0412066].

[13] F. De Bernardis, L. Pagano, P. Serra, A. Melchiorri, and A. Cooray, *Anisotropies in the Cosmic Neutrino Background after WMAP 5-year Data*, JCAP **0806** (2008) 013, [arXiv:0804.1925].

[14] J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) IV: efficient implementation of non-cold relics*, JCAP **1109** (2011) 032, [arXiv:1104.2935].

[15] T. L. Smith, S. Das, and O. Zahn, *Constraints on neutrino and dark radiation interactions using cosmological observations*, Phys. Rev. **D85** (2012) 023001, [arXiv:1105.3246].

[16] R. Diamanti, E. Giusarma, O. Mena, M. Archidiacono, and A. Melchiorri, *Dark Radiation and interacting scenarios*, Phys. Rev. **D87** (2013), no. 6 063509, [arXiv:1212.6007].

[17] M. Archidiacono, E. Giusarma, A. Melchiorri, and O. Mena, *Dark Radiation in extended cosmological scenarios*, Phys. Rev. **D86** (2012) 043509, [arXiv:1206.0109].

[18] M. Gerbino, E. Di Valentino, and N. Said, *Neutrino Anisotropies after Planck*, Phys. Rev. **D88** (2013), no. 6 063538, [arXiv:1304.7400].

[19] M. Archidiacono, E. Giusarma, S. Hannestad, and O. Mena, *Cosmic dark radiation and neutrinos*, Adv. High Energy Phys. **2013** (2013) 191047, [arXiv:1307.0637].

[20] M. Archidiacono, E. Giusarma, A. Melchiorri, and O. Mena, *Neutrino and dark radiation properties in light of recent CMB observations*, Phys. Rev. **D87** (2013), no. 10 103519, [arXiv:1303.0143].

[21] E. Sellentin and R. Durrer, *Detecting the cosmological neutrino background in the CMB*, Phys. Rev. **D92** (2015), no. 6 063012, [arXiv:1412.6427].

[22] E. Di Valentino, S. Gariazzo, M. Gerbino, E. Giusarma, and O. Mena, *Dark Radiation and Inflationary Freedom after Planck 2015*, arXiv:1601.0755.

[23] J. F. Beacom, N. F. Bell, and S. Dodelson, *Neutrinoless universe*, Phys. Rev. Lett. **93** (2004) 121302, [astro-ph/0404585].

[24] S. Hannestad, *Structure formation with strongly interacting neutrinos - Implications for the cosmological neutrino mass bound*, JCAP **0502** (2005) 011, [astro-ph/0411475].

[25] N. F. Bell, E. Pierpaoli, and K. Sigurdson, *Cosmological signatures of interacting neutrinos*, Phys. Rev. **D73** (2006) 063523, [astro-ph/0511410].

[26] R. F. Sawyer, *Bulk viscosity of a gas of neutrinos and coupled scalar particles, in the era of recombination*, Phys. Rev. **D74** (2006) 043527, [astro-ph/0601525].

[27] A. Basboll, O. E. Bjaelde, S. Hannestad, and G. G. Raffelt, *Are cosmological neutrinos free-streaming?*, Phys. Rev. **D79** (2009) 043512, [arXiv:0806.1735].

[28] I. M. Oldengott, C. Rampf, and Y. Y. Y. Wong, *Boltzmann hierarchy for interacting neutrinos I: formalism*, JCAP **1504** (2015), no. 04 016, [arXiv:1409.1577].

[29] M. Archidiacono and S. Hannestad, *Updated constraints on non-standard neutrino interactions from Planck*, JCAP **1407** (2014) 046, [arXiv:1311.3873].

[30] M. Shoji and E. Komatsu, *Massive Neutrinos in Cosmology: Analytic Solutions and Fluid Approximation*, Phys. Rev. **D81** (2010) 123516, [arXiv:1003.0942]. [Erratum: Phys. Rev.D82,089901(2010)].

[31] T. D. Jacques, L. M. Krauss, and C. Lunardini, *Additional Light Sterile Neutrinos and Cosmology*, Phys.Rev. **D87** (2013), no. 8 083515, [arXiv:1301.3119].

[32] E. Grohs, G. M. Fuller, C. T. Kishimoto, and M. W. Paris, *Effects of neutrino rest mass on N_eff and ionization equilibrium freeze-out*, arXiv:1412.6875.

A General Analysis of Direct Dark Matter Detection: From Microphysics to Observational Signatures

James B. Dent[a], Lawrence M. Krauss[b,c], Jayden L. Newstead[b], and Subir Sabharwal[b]

[a] *Department of Physics, University of Louisiana at Lafayette, Lafayette, LA 70504, USA,*

[b] *Department of Physics and School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA, and*

[c] *Research School of Astronomy and Astrophysics, Mt. Stromlo Observatory, Australian National University, Canberra 2614, Australia*

(Dated: May 12 2015)

Abstract

Beginning with a set of simplified models for spin-0, spin-$\frac{1}{2}$, and spin-1 dark matter candidates using completely general Lorentz invariant and renormalizable Lagrangians, we derive the full set of non-relativistic operators and nuclear matrix elements relevant for direct detection of dark matter, and use these to calculate rates and recoil spectra for scattering on various target nuclei. This allows us to explore what high energy physics constraints might be obtainable from direct detection experiments, what degeneracies exist, which operators are ubiquitous and which are unlikely or subdominant. We find that there are operators which are common to all spins as well operators which are unique to spin-$\frac{1}{2}$ and spin-1 and elucidate two new operators which have not been previously considered. In addition we demonstrate how recoil energy spectra can distinguish fundamental microphysics if multiple target nuclei are used. Our work provides a complete roadmap for taking generic fundamental dark matter theories and calculating rates in direct detection experiments. This provides a useful guide for experimentalists designing experiments and theorists developing new dark matter models.

I. INTRODUCTION

The existence of non-baryonic dark matter has been inferred from measurements including galactic rotation curves [1], large scale structure surveys [2–4], X-ray observations [5], gravitational lensing [6, 7], and cosmic microwave background anisotropy measurements [8], spanning cosmological eras from the present day to the remote past. This widespread and robust data has led to cold dark matter models with a cosmological constant, labeled ΛCDM becoming entrenched as the standard cosmological model.

Nevertheless, this impressive array of observations has only been sensitive to the *gravitational* influence of dark matter and constrained its relic abundance, leaving its particle nature as one of the most important open questions in physics. The search for dark matter includes indirect astrophysical searches ([9–13]), collider production efforts (for some examples of dark matter searches at the LHC, see [14–18]) which will examine new territory soon with LHC run 2 which will commence this year, and attempts to observe dark matter interactions with Standard Model (SM) particles via dark matter-nucleus scattering processes in direct detection experiments, to which we now turn.

The search for dark matter via direct detection goes back at least three decades [19, 20] and has been particularly vigorous over the last decade or so with experiments such as LUX [21], Xenon100 [22], CDMS II (Ge) [23], CDMS I (Si) [24], DAMA/LIBRA [25], COGENT [26], and CRESST [27] pushing ever deeper into weakly interacting dark matter mass and scattering cross-section parameter space, but has thus far failed to yield a convincing signal. In the near future detectors such as Super CDMS [28] (which has recently released its first results on low mass dark matter searches [29, 30]), XENON1T [31], and DARWIN [32] are expected to push the limits of direct detection orders of magnitude below the current levels.

In order to connect observations to microphysical models one needs a general framework within which to interpret the observations of direct detection experiments. For quite some time the prevailing method of analyzing dark matter-nucleus interactions has been to assume that dark matter is a weakly interacting massive particle (WIMP), and then to categorize the interactions as elastic and isospin conserving and either spin-independent or spin-dependent [33, 34]. For some well studied models of dark matter, such as the weakly interacting Majorana neutralino found in supersymmetry models, this assumption is reasonable.

With an absence of observed dark matter signals, there has of late been a surge in

interest in exploring more general types of interactions between dark matter and nuclei. Generalizations include inelastic and momentum dependent interactions, which may arise due to additional structure in the dark sector including excited dark matter states, or dark gauge bosons giving rise to electric and magnetic form factors [35–42].

The formalism of choice for many of these investigations is relativistic effective field theory, which provides a model independent framework to analyse dark matter-SM interactions [43–45]. It has been shown that these effective theories break down when applied to high-momentum transfer experiments, such as the LHC [46]. Therefore analyses moved beyond this framework and have moved to what are labeled as 'simplified models' instead [47–49]. Simplified models are field theories which extend the SM by a single dark matter particle and a single mediator particle which allows the WIMP to communicate with quarks and/or leptons. The newly added dark matter content is assumed to be a singlet under the SM gauge groups (we will consider some cases where the particles mediating the interaction have SM charge). In this context it is then possible to calculate collider amplitudes valid at the high energies of interest in such experiments. Given this simple dark sector, one can write down an exhaustive list of every combination of WIMP and mediator spins, and all possible tree level interactions. These simplified models have now gained popularity for analyzing indirect detection signals [50, 51], allowing connections to be made with the growing body of literature which make use of them.

Another step towards placing dark matter-nucleus interactions on a general footing has been accomplished recently by utilizing a non-relativistic effective field theory (EFT) approach [52–55]. Since the interactions in direct detection scenarios are assumed to take place due to an incoming dark matter particle with a typical velocity $\mathcal{O}(100\text{km/s})$, the recoil momenta in such an interaction will be $\mathcal{O}(\lesssim 100\text{keV})$. The particle masses involved, including the nucleons of roughly GeV scale, the dark matter particles, which typically range from the GeV region to several orders of magnitude above, and mediators that can also be quite heavy compared to the typical interaction momenta, produce a situation where an EFT treatment is quite natural.

In order to circumvent as much model dependence as possible, one can construct general interactions which obey Galilean invariance, T-symmetry, and Hermiticity. These operators

will take the standard effective four-particle interaction form, reminiscent of Fermi's original model of weak interactions. The non-relativistic interactions can be shown to be functions of only four parameters including the nucleon spin S_N, the dark matter spin S_χ, the momentum transfer, $\vec{q}$, and a kinematic variable $\vec{v}^{\perp}$ which is a function of the relative incoming ($\vec{v}_{\chi,in} - \vec{v}_{N,in}$) and outgoing velocities $\vec{v}_{\chi,out} - \vec{v}_{N,out}$

$$\vec{v}^{\perp} = \frac{1}{2}\left(\vec{v}_{\chi,in} - \vec{v}_{N,in} + \vec{v}_{\chi,out} - \vec{v}_{N,out}\right) = \vec{v}_{\chi,in} - \vec{v}_{N,in} + \frac{\vec{q}}{2\mu_N} \tag{1}$$

which obeys $\vec{v}^{\perp} \cdot \vec{q} = 0$. It was demonstrated in [53] that there exist fifteen such non-relativistic interactions which arise from twenty possible bi-linear combinations of dark matter and nucleons.

The formalism developed in [53] is unique in being the only analysis to comprehensively develop the nuclear physics of direct detection experiments. From this general framework it is now apparent that there are interactions beyond the standard spin independent/dependent type. The origins of these 'new' interactions are not necessarily exotic and it has been shown, in the context of relativistic EFT, how many of them can be generated [56].

What has been lacking to date however, is a completely general and comprehensive treatment that connects high energy microphysics with low-energy effective nuclear matrix elements in a model independent way. It is possible, for example, that the various interactions listed in [53] can give rise to degeneracies where different fundamental dark matter Lagrangians, describing dark matter and interaction mediators of various spins, can produce the same interaction types. This will obviously pose problems for attempts to discern the properties of dark matter when interpreting the results of experimental data. Furthermore, dark matter may not be spin-$\frac{1}{2}$, which creates a need for extending the parametric framework from the four descriptors listed above. In particular, as we shall show, this allows the existence of new non-relativistic operators to appear in the low energy effective theory.

Motivated by the above we present here a general analysis covering a broad spectrum of particle and interaction types, starting from the microphysics, which will enable one to link experiment with fundamental theory while incorporating the new nuclear responses described in [53].

In this work we build upon the NR-EFT description by examining simplified models which incorporate the most general renormalizable Lagrangians for scalar, spinor, and vector dark matter interacting with nucleons via scalar, spinor, and vector mediators, consistent with

Lorentz invariance and hermiticity while imposing stability of the dark matter candidates. We integrate out the heavy mediator and obtain effective relativistic interaction Lagrangians. Next, we take the non-relativistic limit of these Lagrangians, and identify them with the NR operators from [53], which are reproduced below, in Table 1. Using these, we identify which electroweak nuclear responses are excited by a given fundamental interaction model and determine the relative importance of various models within the context of direct detection experiments consisting of xenon and germanium targets by exploring the relative magnitude of coefficients of these operators, and also their energy dependence.

The paper is organized as follows; in section II the EFT formalism of [53] is summarized, in section III we build the generalized relativistic Lagrangians and in section IV we outline the signatures and distinguishability of these models in the context of direct detection experiments, providing a framework for both experimentalists and theorists to base their future analyses.

II. EFFECTIVE FIELD THEORY OF DIRECT DETECTION

Conventionally, coherent WIMP-nucleus scattering has been considered to come from two types of interactions; spin-independent (SI) and spin-dependent (SD). SI interactions couple to the charge/mass of the nucleus while SD couples to the spin. The nuclear cross section is generally written in terms of the nucleon cross section at zero momentum transfer, σ_0, and a form factor, $F(q)$, to take into account the loss of coherence over the finite size of the nucleus,

$$\frac{d\sigma}{dE_r} = \frac{M}{2\pi\mu_{\chi M}v^2}\left(\sigma_0^{SI}F_{SI}^2(q) + \sigma_0^{SD}F_{SD}^2(q)\right). \tag{2}$$

where M is the mass of the target nucleus and $\mu_{\chi M}$ is the WIMP-nucleus reduced mass. This picture has recently been shown to be incomplete, as it is also possible for the WIMP to couple to the nucleus through additional nuclear responses [53]. Working in the language of a non-relativistic (NR) effective field theory Fitzpatrick et al. identified 15 operators to characterize the ways in which a WIMP can couple to the various nuclear responses. These operators are constructed from combinations of non-relativistic vectors which respect Galilean invariance, T symmetry and which are Hermitian. We list them in table I. The

Hermitian vectors are:

$$i\frac{\vec{q}}{m_N},\ \vec{v}^{\perp}=\vec{v}+\frac{\vec{q}}{2\mu_N},\ \vec{S}_\chi,\ \vec{S}_N, \tag{3}$$

where $\vec{q}=\vec{p'}-\vec{p}=\vec{k}-\vec{k'}$ is the momentum transfer, $\vec{v}$ is the velocity of WIMP with respect to the nucleus of the detector, μ_N is the reduced mass of the system and $\vec{S}_\chi$ and $\vec{S}_N$ are the WIMP and nuclear spins respectively. Throughout the paper, we denote by $\vec{p}$ and $\vec{p'}$ the incoming and outgoing WIMP momenta and by $\vec{k}$ and $\vec{k'}$ the incoming and outgoing nuclear momenta respectively. Energy-momentum conservation implies the orthogonality condition $\vec{q}\cdot\vec{v}^{\perp}=0$. Here we will briefly outline the procedure employed in [53] in going from the NR operators to the final differential WIMP-nucleus cross section.

TABLE I. List of NR effective operators described in [53]

$\mathcal{O}_1$	$1_\chi 1_N$
$\mathcal{O}_2$	$(\vec{v}^{\perp})^2$
$\mathcal{O}_3$	$i\vec{S}_N\cdot(\frac{\vec{q}}{m_N}\times\vec{v}^{\perp})$
$\mathcal{O}_4$	$\vec{S}_\chi\cdot\vec{S}_N$
$\mathcal{O}_5$	$i\vec{S}_\chi\cdot(\frac{\vec{q}}{m_N}\times\vec{v}^{\perp})$
$\mathcal{O}_6$	$(\frac{\vec{q}}{m_N}\cdot\vec{S}_N)(\frac{\vec{q}}{m_N}\cdot\vec{S}_\chi)$
$\mathcal{O}_7$	$\vec{S}_N\cdot\vec{v}^{\perp}$
$\mathcal{O}_8$	$\vec{S}_\chi\cdot\vec{v}^{\perp}$
$\mathcal{O}_9$	$i\vec{S}_\chi\cdot(\vec{S}_N\times\frac{\vec{q}}{m_N})$
$\mathcal{O}_{10}$	$i\frac{\vec{q}}{m_N}\cdot\vec{S}_N$
$\mathcal{O}_{11}$	$i\frac{\vec{q}}{m_N}\cdot\vec{S}_\chi$
$\mathcal{O}_{12}$	$\vec{S}_\chi\cdot(\vec{S}_N\times\vec{v}^{\perp})$
$\mathcal{O}_{13}$	$i(\vec{S}_\chi\cdot\vec{v}^{\perp})(\frac{\vec{q}}{m_N}\cdot\vec{S}_N)$
$\mathcal{O}_{14}$	$i(\vec{S}_N\cdot\vec{v}^{\perp})(\frac{\vec{q}}{m_N}\cdot\vec{S}_\chi)$
$\mathcal{O}_{15}$	$-(\vec{S}_\chi\cdot\frac{\vec{q}}{m_N})\left((\vec{S}_N\times\vec{v}^{\perp})\cdot\frac{\vec{q}}{m_N}\right)$

In general one can write down the non-relativistic interaction Lagrangian as

$$\mathcal{L}_{NR}=\sum_{\alpha=n,p}\sum_{i=1}^{15}c_i^\alpha\mathcal{O}_i^\alpha, \tag{4}$$

where the coefficients c_i^α are given by the microphysics of the interaction and in general one could allow for isospin violation by having different couplings to neutron and proton inside the nucleus. This can be rewritten in 2-component isospin space as

$$\mathcal{L}_{NR} = \sum_{\tau=0,1} \sum_{i=1}^{15} c_i^\tau \mathcal{O}_i t^\tau \tag{5}$$

where t^0 and t^1 are the identity matrix and the Pauli matrix σ^3 respectively. The nucleus is composed of nucleons, and these can individually interact with the WIMP. This is incorporated by considering the operator $\mathcal{O}(j)$ as an interaction between a single nucleon, j, and the WIMP, and then summing over the nucleons.

$$\sum_{\tau=0,1} \sum_{i=1}^{15} c_i^\tau \mathcal{O}_i t^\tau \to \sum_{\tau=0,1} \sum_{i=1}^{15} c_i^\tau \sum_{j=1}^{A} \mathcal{O}_i(j) t^\tau(j) \tag{6}$$

where A is the atomic mass number given by the total number of neutrons and protons. One can do the same reduction with $\vec{v}^\perp$,

$$\begin{aligned} \vec{v}^\perp &\to \{\vec{v}_\chi - \vec{v}_N(i), i = 1, ..., A\} \\ &\equiv \vec{v}_T^\perp - \{\dot{\vec{v}}_N(i), i = 1, ..., A-1\} \end{aligned} \tag{7}$$

where $\vec{v}_\chi$ and $\vec{v}_N(i)$ are the symmetrized combination of incoming and outgoing velocities for the WIMP and nucleons respectively. $\vec{v}_T^\perp$ (here T stands for target, i.e., the nuclear center-of-mass) is defined as

$$\vec{v}_T^\perp = \vec{v}_\chi - \frac{1}{2A} \sum_{i=1}^{A} [\vec{v}_{N,in}(i) + \vec{v}_{N,out}(i)] \tag{8}$$

This allows for a decomposition of the nucleon velocities into internal velocities $\dot{\vec{v}}_N(i)$ that act only on intrinsic nuclear coordinates and 'in' and 'out' velocities that evolve as a WIMP scatters off the detector. As an example, the dot product between $\vec{v}_N^\perp$ and $\vec{S}_N$ can be rewritten as

$$\vec{v}^\perp \cdot \vec{S}_N \to \sum_{i=1}^{A} \frac{1}{2} [\vec{v}_{\chi,in} + \vec{v}_{\chi,out} - \vec{v}_{N,in}(i) - \vec{v}_{N,out}(i)] \cdot \vec{S}_N(i) \tag{9}$$

$$= \vec{v}_T^\perp \cdot \sum_{i=1}^{A} \vec{S}_N(i) - \left\{ \sum_{i=1}^{A} \frac{1}{2} [\vec{v}_{N,in}(i) + \vec{v}_{N,out}(i)] \cdot \vec{S}_N(i) \right\}_{int} \tag{10}$$

The second term in the curly brackets is internal to the nucleus and acts as an operator on the 'in' and 'out' nucleon states. $\vec{v}_{N,in}$ can be replaced by $\vec{p}_{N,in}/M$ acting on the incoming state, which can in turn be replaced by $i\overleftarrow{\nabla}/M$, and similarly $\vec{p}_{N,out}/M$ by $-i\overrightarrow{\nabla}/M$ on the

outgoing nuclear state. Finally, since the nucleus is non-zero in size and individual nucleons locally interact with the WIMP, nuclear operators built from $\mathcal{O}_i$ are accompanied by an additional spatial operator $e^{-i\vec{q}\cdot\vec{x}(i)}$ where $x(i)$ is the location of the i^{th} nucleon inside the nucleus.

Starting from Eqn. 6 and using the substitution rules for $\vec{v}^{\perp}$ and including a factor of $e^{-i\vec{q}\cdot\vec{x}_i}$, the interaction Lagrangian can be written as a sum of five distinct terms (nuclear electroweak operators) that only act on internal nucleon states. Their coefficients, on the other hand, act on WIMP 'in' and 'out' states. The WIMP-nucleus interaction can then be written as

$$\sum_{\tau=0,1}\left\{l_0^\tau S + l_0^{A\tau} T + \vec{l}_5^\tau\cdot\vec{P} + \vec{l}_M^\tau\cdot Q + \vec{l}_E^\tau\cdot\vec{R}\right\}t^\tau(i) \tag{11}$$

where

$$\begin{aligned}
S &= \sum_{i=1}^{A} e^{-i\vec{q}\cdot\vec{x}_i}\\
T &= \sum_{i=1}^{A}\frac{1}{2M}\left[-\frac{1}{i}\overleftarrow{\nabla}_i\cdot\vec{\sigma}(i)e^{-i\vec{q}\cdot\vec{x}_i} + e^{-i\vec{q}\cdot\vec{x}_i}\vec{\sigma}(i)\cdot\frac{1}{i}\overrightarrow{\nabla}_i\right]\\
\vec{P} &= \sum_{i=1}^{A}\vec{\sigma}(i)e^{-i\vec{q}\cdot\vec{x}_i}\\
\vec{Q} &= \sum_{i=1}^{A}\frac{1}{2M}\left[-\frac{1}{i}\overleftarrow{\nabla}_i e^{-i\vec{q}\cdot\vec{x}_i} + e^{-i\vec{q}\cdot\vec{x}_i}\frac{1}{i}\overrightarrow{\nabla}_i\right]\\
\vec{R} &= \sum_{i=1}^{A}\frac{1}{2M}\left[\overleftarrow{\nabla}_i\times\vec{\sigma}(i)e^{-i\vec{q}\cdot\vec{x}_i} + e^{-i\vec{q}\cdot\vec{x}_i}\vec{\sigma}(i)\times\overrightarrow{\nabla}_i\right]
\end{aligned} \tag{12}$$

and

$$\begin{aligned}
l_0^\tau &= c_1^\tau + ic_5^\tau\vec{S}_\chi\cdot\left(\frac{\vec{q}}{m_N}\times\vec{v}_T^{\perp}\right) + c_8^\tau(\vec{S}_\chi\cdot\vec{v}_T^{\perp}) + ic_{11}^\tau\frac{\vec{q}\cdot\vec{S}_\chi}{m_N}\\
l_0^{A\tau} &= -\frac{1}{2}\left[c_7^\tau + ic_{14}^\tau\left(\vec{S}_\chi\cdot\frac{\vec{q}}{m_N}\right)\right]\\
\vec{l}_5 &= \frac{1}{2}\left[c_3^\tau i\frac{\left(\vec{q}\times\vec{v}_T^{\perp}\right)}{m_N} + c_4^\tau\vec{S}_\chi + c_6^\tau\frac{(\vec{q}\cdot\vec{S}_\chi)\vec{q}}{m_N^2} + c_7^\tau\vec{v}_T^{\perp} + ic_9^\tau\frac{(\vec{q}\times\vec{S}_\chi)}{m_N} + ic_{10}^\tau\frac{\vec{q}}{m_N}\right.\\
&\qquad \left. c_{12}^\tau(\vec{v}_T^{\perp}\times\vec{S}_\chi) + ic_{13}^\tau\frac{(S_\chi\cdot\vec{v}_T^{\perp})\vec{q}}{m_N} + ic_{14}^\tau\left(\vec{S}_\chi\cdot\frac{\vec{q}}{m_N}\right)\vec{v}_T^{\perp} + c_{15}^\tau\frac{(\vec{q}\cdot\vec{S}_\chi)(\vec{q}\times\vec{v}_T^{\perp})}{m_N^2}\right]\\
\vec{l}_M &= c_5^\tau\left(i\frac{\vec{q}}{m_N}\times\vec{S}_\chi\right) - \vec{S}_\chi c_8^\tau\\
\vec{l}_E &= \frac{1}{2}\left[c_3^\tau\frac{\vec{q}}{m_N} + ic_{12}^\tau\vec{S}_\chi - c_{13}^\tau\frac{(\vec{q}\times\vec{S}_\chi)}{m_N} - ic_{15}^\tau\frac{(\vec{q}\cdot\vec{S}_\chi)\vec{q}}{m_N^2}\right]
\end{aligned} \tag{13}$$

The WIMP-nucleus amplitude, $\mathcal{M}$, can then be succinctly written as

$$\mathcal{M} = \sum_{\tau=0,1} \langle j_\chi, M_\chi; j_N, M_N | \left\{ l_0^\tau S + l_0^{A\tau} T + \vec{l}_5^\tau \cdot \vec{P} + \vec{l}_M^\tau \cdot Q + \vec{l}_E^\tau \cdot \vec{R} \right\} t^\tau(i) | j_\chi, M_\chi; j_N, M_N \rangle. \tag{14}$$

By using spherical decomposition, the internal nuclear operators S, T, P, Q and R can be further rewritten in terms of standard nuclear electroweak responses as follows:

$$\begin{aligned}
\mathcal{M} = \sum_{\tau=0,1} \langle j_\chi, M_{\chi f}; j_N, M_{Nf} | \Bigg(& \sum_{J=0} \sqrt{4\pi(2J+1)}(-i)^J \left[l_0^\tau M_{J0;\tau} - i l_0^{A\tau} \frac{q}{m_N} \tilde{\Omega}_{J0;\tau}(q) \right] \\
& + \sum_{J=1} \sqrt{2\pi(2J+1)}(-i)^J \sum_{\lambda \pm 1} (-1)^\lambda \Big\{ l_{5\lambda}^\tau [\lambda \Sigma_{J-\lambda;\tau}(q) + i \Sigma'_{J-\lambda;\tau}(q)] \\
& - i \frac{q}{m_N} l_{M\lambda}^\tau [\lambda \Delta_{J-\lambda;\tau}(q)] - i \frac{q}{m_N} l_{E\lambda}^\tau [\lambda \tilde{\Phi}_{J-\lambda;\tau}(q) + i \tilde{\Phi}'_{J-\lambda;\tau}(q)] \Big\} \\
& + \sum_{J=0}^{\infty} \sqrt{4\pi(2J+1)}(-i)^J \left[i l_{50}^\tau \Sigma''_{J0;\tau}(q) + \frac{q}{m_N} l_{M0}^\tau \tilde{\Delta}''_{J0;\tau}(q) + \frac{q}{m_N} l_{E0}^\tau \tilde{\Phi}''_{J0;\tau}(q) \right] \Bigg) | j_\chi, M_{\chi i}; j_N, M_{Ni} \rangle
\end{aligned} \tag{15}$$

Where there is an implicit sum over the nucleons,

$$\mathcal{O}_{JM;\tau}(q) \equiv \sum_{i=1}^{A} \mathcal{O}_{JM}(q\vec{x}_i) t^\tau(i), \tag{16}$$

and the various electroweak responses are defined as

$$\begin{aligned}
M_{JM}(q\vec{x}) &\equiv j_J(qx) Y_{JM}(\Omega_x) \\
\vec{M}_{JL}^M &\equiv j_J(qx) \vec{Y}_{JLM}(\Omega_x) \\
\Delta_{JM} &\equiv \vec{M}_{JJ}^M(qx_i) \cdot \frac{1}{q} \vec{\nabla}_i \\
\Sigma'_{JM} &\equiv -i \left\{ \frac{1}{q} \vec{\nabla}_i \times \vec{M}_{JJ}^M(q\vec{x}_i) \right\} \cdot \vec{\sigma}(i) \\
\Sigma''_{JM} &\equiv \left\{ \frac{1}{q} \vec{\nabla}_i M_{JM}(q\vec{x}_i) \right\} \cdot \vec{\sigma}(i) \\
\tilde{\Phi}'_{JM} &\equiv \left[\frac{1}{q} \vec{\nabla}_i \times \vec{M}_{JJ}^M(q\vec{x}_i) \right] \cdot \left[\vec{\sigma}(i) \times \frac{1}{q} \vec{\nabla}_i \right] + \frac{1}{2} \vec{M}_{JJ}^M(q\vec{x}_i) \cdot \vec{\sigma}(i) \\
\Phi''_{JM} &\equiv i \left[\frac{1}{q} \vec{\nabla}_i M_{JM}(q\vec{x}_i) \right] \cdot \left[\vec{\sigma}(i) \times \frac{1}{q} \vec{\nabla}_i \right] \\
\Sigma_{JM} &\equiv \vec{M}_{JJ}^M(q\vec{x}_i) \cdot \vec{\sigma}(i) \\
\tilde{\Omega}_{JM} &\equiv \Omega_{JM}(q\vec{x}_i) + \frac{1}{2} \Sigma''_{JM}(q\vec{x}_i) \\
\tilde{\Phi}_{JM} &\equiv \Phi_{JM}(qx_i) - \frac{1}{2} \Sigma'_{JM}(qx_i) \\
\tilde{\Delta}''_{JM} &\equiv \Delta''_{JM}(qx_i) - \frac{1}{2} M_{JM}(qx_i)
\end{aligned} \tag{17}$$

where Y_{JM} and $\vec{Y}_{JLM}$ are spherical harmonics and vector spherical harmonics respectively. We are only considering elastic transitions, and assuming parity and CP as symmetries of the nuclear ground state. This eliminates some of the responses, and only $M, \Phi'', \Sigma', \Delta, \Sigma'', \tilde{\Phi}'$ survive. To calculate cross-sections, one needs to square the amplitude, average over initial spins and sum over final spins. The matrix element squared for the nuclear portion of the amplitude has been made available by Fitzpatrick et al. [53], and codes have been supplied to calculate the full amplitude and rate [54].

As we shall describe, in the following analysis we discovered that two additional NR operators are required to fully describe the scattering of spin-1 WIMPs off nuclei,

$$\begin{aligned} \mathcal{O}_{17} &\equiv i\frac{\vec{q}}{m_N} \cdot \mathcal{S} \cdot \vec{v}_\perp, \\ \mathcal{O}_{18} &\equiv i\frac{\vec{q}}{m_N} \cdot \mathcal{S} \cdot \vec{S}_N, \end{aligned} \tag{18}$$

where $\mathcal{S}$ is the symmetric combination of polarization vectors. Appendix A contains the details required to include these new operators in the above formalism.

III. SIMPLIFIED MODELS FOR DIRECT DETECTION

From a model building perspective, one would like to know how relevant the novel nuclear responses are in interpreting direct detection data. Previous work [56] demonstrated that using only the SI/SD form factors (even with additional momentum dependence taken into account) can lead one to infer wildly incorrect values of the WIMP mass and cross sections.

Here we go further by starting with simplified models at the Lagrangian level, where 'simplified model' means a single WIMP with a single mediator coupling it to the quark sector. This is useful for two reasons; it allows us to better explore which NR operators arise from a broad set of UV complete theories, and also make connection with the growing body of literature which use simplified models for indirect detection and collider searches.

When it comes to interpreting signals, knowing comprehensively how different interactions with different nuclei arise from different UV complete models will allow us to identify degeneracies between competing models. Further, it can also help optimize target selection for maximum discrimination of the UV model parameter space.

In building these simplified models we remain agnostic about the WIMP's spin, and consider dark matter spins of 0, $\frac{1}{2}$ and 1. We do however only consider renormalizable interactions between quarks and WIMPs. To ensure a stable WIMP, we assume that the WIMP is either charged under some internal gauge group or a discrete symmetry group (for example Z_2). However, we assume that this gauge charge is not shared by quarks. We will couple the WIMP to the quarks via a heavy mediator in two distinct ways: charged and uncharged mediators, each with all possible spins consistent with angular momentum conservation. The mediator mass is chosen to be the heaviest scale in the problem (and certainly much greater than the momentum exchange which characterizes the scattering process) so that we can integrate it out (see appendix B for details). This leads to relativistic effective WIMP-nucleon interactions, whose NR limit can then be examined. In the uncharged mediator case we will consider mediators that are neutral under all SM and WIMP gauge charges, while in the charged case, the mediator must have both WIMP and SM gauge charges. Given the above as a guide, our Lagrangian construction is then constrained only by gauge invariance, Lorentz invariance, renormalizability and hermiticity. In certain cases which follow, the requirement of hermiticity demands coupling constants be complex. Unless explicitly noted, the coupling constants are dimensionless and can be assumed to be real.

A. Uncharged-mediator Lagrangians

1. Scalar Dark Matter

We begin with a spin-0 scalar WIMP, S, which has some internal charge to ensure stability, and $S^\dagger$ is its Hermitian conjugate. To have renormalizable interactions, the neutral mediator can only be a scalar or a vector. We denote the scalar mediator by ϕ and the vector mediator by G^μ with field strength tensor $\mathcal{G}_{\mu\nu}$.

The most general renormailzable Lagrangian for scalar mediation consistent with the

above assumptions is given by

$$\begin{aligned}\mathcal{L}_{S\phi q} = &\partial_\mu S^\dagger \partial^\mu S - m_S^2 S^\dagger S - \frac{\lambda_S}{2}(S^\dagger S)^2 \\ &+\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m_\phi^2\phi^2 - \frac{m_\phi\mu_1}{3}\phi^3 - \frac{\mu_2}{4}\phi^4 \\ &+i\bar{q}\not{D}q - m_q\bar{q}q \\ &-g_1 m_S S^\dagger S\phi - \frac{g_2}{2}S^\dagger S\phi^2 - h_1\bar{q}q\phi - ih_2\bar{q}\gamma^5 q\phi, \end{aligned} \tag{19}$$

where we have suppressed all the SM quark interactions. Similarly, the Lagrangian for vector mediation (up to gauge fixing terms) is

$$\begin{aligned}\mathcal{L}_{SGq} = &\partial_\mu S^\dagger \partial^\mu S - m_S^2 S^\dagger S - \frac{\lambda_S}{2}(S^\dagger S)^2 \\ &-\frac{1}{4}\mathcal{G}_{\mu\nu}\mathcal{G}^{\mu\nu} + \frac{1}{2}m_G^2 G_\mu G^\mu - \frac{\lambda_G}{4}(G_\mu G^\mu)^2 \\ &+i\bar{q}\not{D}q - m_q\bar{q}q \\ &-\frac{g_3}{2}S^\dagger S G_\mu G^\mu - ig_4(S^\dagger\partial_\mu S - \partial_\mu S^\dagger S)G^\mu \\ &-h_3(\bar{q}\gamma_\mu q)G^\mu - h_4(\bar{q}\gamma_\mu\gamma^5 q)G^\mu. \end{aligned} \tag{20}$$

2. Spin-$\frac{1}{2}$ Dark Matter

If the WIMP has spin-$\frac{1}{2}$ (denoted by χ below), then, as in the scalar WIMP case, mediation will only occur via scalar or vector mediators. The most general renormalizable interactions for the scalar (ϕ) and vector mediator (G_μ) cases respectively are given below,

$$\begin{aligned}\mathcal{L}_{\chi\phi q} = &i\bar{\chi}\not{D}\chi - m_\chi\bar{\chi}\chi \\ &+\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m_\phi^2\phi^2 - \frac{m_\phi\mu_1}{3}\phi^3 - \frac{\mu_2}{4}\phi^4 \\ &+i\bar{q}\not{D}q - m_q\bar{q}q \\ &-\lambda_1\phi\bar{\chi}\chi - i\lambda_2\phi\bar{\chi}\gamma^5\chi - h_1\phi\bar{q}q - ih_2\phi\bar{q}\gamma^5 q, \end{aligned} \tag{21}$$

$$\begin{aligned}\mathcal{L}_{\chi Gq} = & i\bar{\chi}\not{D}\chi - m_\chi\bar{\chi}\chi \\ & -\frac{1}{4}\mathcal{G}_{\mu\nu}\mathcal{G}^{\mu\nu} + \frac{1}{2}m_G^2 G_\mu G^\mu \\ & +i\bar{q}\not{D}q - m_q\bar{q}q \\ & -\lambda_3\bar{\chi}\gamma^\mu\chi G_\mu - \lambda_4\bar{\chi}\gamma^\mu\gamma^5\chi G_\mu \\ & -h_3\bar{q}\gamma_\mu q G^\mu - h_4\bar{q}\gamma_\mu\gamma^5 q G^\mu. \end{aligned} \tag{22}$$

3. Spin-1 Dark Matter

If the WIMP is a massive spin-1 particle, uncharged mediation to the quark sector can occur via a heavy scalar or a vector particle. For the case of vector mediation, there are many possible interactions because the Lorentz indices on the vectors afford a more diverse set of terms. The general interaction Lagrangian for the scalar mediation case is

$$\begin{aligned}\mathcal{L}_{X\phi q} = & -\frac{1}{2}\mathcal{X}^\dagger_{\mu\nu}\mathcal{X}^{\mu\nu} + m_X^2 X^\dagger_\mu X^\mu - \frac{\lambda_X}{2}(X^\dagger_\mu X^\mu)^2 \\ & +\frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{2}m_\phi^2\phi^2 - \frac{m_\phi\mu_1}{3}\phi^3 - \frac{\mu_2}{4}\phi^4 \\ & +i\bar{q}\not{D}q - m_q\bar{q}q \\ & -b_1 m_X\phi X^\dagger_\mu X^\mu - \frac{b_2}{2}\phi^2 X^\dagger_\mu X^\mu - h_1\phi\bar{q}q - ih_2\phi\bar{q}\gamma^5 q. \end{aligned} \tag{23}$$

For the case of vector mediation, there are many possible interactions because the Lorentz indices on the vectors afford a more diverse set of terms. The Lagrangian is given by

$$\begin{aligned}\mathcal{L}_{XGq} = & -\frac{1}{2}\mathcal{X}^\dagger_{\mu\nu}\mathcal{X}^{\mu\nu} + m_X^2 X^\dagger_\mu X^\mu - \frac{\lambda_X}{2}(X^\dagger_\mu X^\mu)^2 \\ & -\frac{1}{4}\mathcal{G}_{\mu\nu}\mathcal{G}^{\mu\nu} + \frac{1}{2}m_G^2 G_\mu^2 - \frac{\lambda_G}{4}(G_\mu G^\mu)^2 \\ & +i\bar{q}\not{D}q - m_q\bar{q}q \\ & -\frac{b_3}{2}G_\mu^2(X^\dagger_\nu X^\nu) - \frac{b_4}{2}(G^\mu G^\nu)(X^\dagger_\mu X_\nu) - \Big[ib_5 X^\dagger_\nu\partial_\mu X^\nu G^\mu \\ & +b_6 X^\dagger_\mu\partial^\mu X_\nu G^\nu + b_7\epsilon_{\mu\nu\rho\sigma}(X^{\dagger\mu}\partial^\nu X^\rho)G^\sigma + h.c.\Big] \\ & -h_3 G_\mu\bar{q}\gamma^\mu q - h_4 G_\mu\bar{q}\gamma^\mu\gamma^5 q \end{aligned} \tag{24}$$

where, for the Lagrangian to be Hermitian, b_6 and b_7 are complex (this implies a new source of CP violation, which will not be considered further here).

A. Charged-mediator Lagrangians

Here we consider the simplest case of mediators that are charged under both the DM internal symmetry group and SM gauge groups. This is motivated by the absence of spin-$\frac{1}{2}$ mediators (s-channel processes) in the previous section. Such a mediator, if neutral, is forbidden by simultaneous requirements of gauge invariance and renormalizability. Dark Matter models with mediators endowed with charges from both DM and SM side have been considered in the literature before [57, 58]. The case of a spin-$\frac{1}{2}$ mediator carrying $SU(3)_c$ is also motivated by studies of heavy quark models. This allows unique interactions as we show below. In particular they necessitate a direct interaction between quarks and WIMPs at the level of the Lagrangian.

1. Scalar Dark Matter

Scalar WIMPs with a charged scalar or vector mediator do not lead to any Lorentz invariant interactions. This is easy to see since both the scalars (or scalar and vector) and the quark are required in the (gauge invariant) interaction, but there is no way to contract the spinor indices consistently if the mediating particle is a scalar or vector. Therefore, the only possibility is that of a spin-1/2 mediator, Q, which acts like a heavy quark. The general renormalizable action is given by

$$\begin{aligned}\mathcal{L}_{SQq} = &\partial_\mu S^\dagger \partial^\mu S - m_S^2 S^\dagger S - \lambda_S (S^\dagger S)^2 \\ &+ i\bar{Q}\not{D}Q - m_Q \bar{Q}Q \\ &+ i\bar{q}\not{D}q - m_q \bar{q}q \\ &- (y_1 S\bar{Q}q + y_2 S\bar{Q}\gamma^5 q + h.c.),\end{aligned} \tag{25}$$

where y_1 and y_2 are again complex.

2. Spin-$\frac{1}{2}$ Dark Matter

For a spin-1/2 WIMP, both a charged scalar and charged vector mediator exchange can

lead to novel interactions. The charged scalar is denoted by Φ and the charged vector by V_μ

$$\begin{aligned}\mathcal{L}_{\chi\Phi q} = & i\bar{\chi}\not{D}\chi - m_\chi\bar{\chi}\chi \\ & +(\partial_\mu\Phi^\dagger)(\partial^\mu\Phi) - m_\Phi^2\Phi^\dagger\Phi - \frac{\lambda_\Phi}{2}(\Phi^\dagger\Phi)^2 \\ & +i\bar{q}\not{D}q - m_q\bar{q}q \\ & -(l_1\Phi^\dagger\bar{\chi}q + l_2\Phi^\dagger\bar{\chi}\gamma^5 q + h.c.), \end{aligned} \tag{26}$$

$$\begin{aligned}\mathcal{L}_{\chi V q} = & i\bar{\chi}\not{D}\chi - m_\chi\bar{\chi}\chi \\ & -\frac{1}{2}\mathcal{V}^\dagger_{\mu\nu}\mathcal{V}^{\mu\nu} + m_V^2 V_\mu^\dagger V^\mu \\ & +i\bar{q}\not{D}q - m_q\bar{q}q \\ & -(d_1\bar{\chi}\gamma^\mu q V_\mu^\dagger + d_2\bar{\chi}\gamma^\mu\gamma^5 q V_\mu^\dagger + h.c.), \end{aligned} \tag{27}$$

where l_1, l_2, d_1 and d_2 are complex.

3. Vector DM

Here again we only have the case of a spin-$\frac{1}{2}$ mediated interaction between vector DM and quarks (again scalar and vector charged mediators aren't possible because they don't lead to Lorentz invariant and renormalizable interactions). The general Lagrangian is given by

$$\begin{aligned}\mathcal{L}_{XQq} = & -\frac{1}{2}\mathcal{X}^\dagger_{\mu\nu}\mathcal{X}^{\mu\nu} + m_X^2 X_\mu^\dagger X^\mu - \frac{\lambda_X}{2}(X_\mu^\dagger X^\mu)^2 \\ & +i\bar{Q}\not{D}Q - m_Q\bar{Q}Q \\ & +i\bar{q}\not{D}q - m_q\bar{q}q \\ & -(y_3 X_\mu\bar{Q}\gamma^\mu q + y_4 X_\mu\bar{Q}\gamma^\mu\gamma^5 q + h.c.), \end{aligned} \tag{28}$$

where y_3 and y_4 are complex.

IV. NON-RELATIVISTIC REDUCTION OF SIMPLIFIED MODELS

After integrating out the heavy mediator we replace quark operators with nucleon operators (see appendix C), take the non-relativistic limit (see appendix B), and match onto the operators given in table I. The results of this calculation are presented in terms of the c_i coefficients from [54], described in section II, facilitating a straightforward computation of

amplitudes and rates. The c_i's are given for each of the WIMP spins in tables II, III and IV. With this general framework in place we can now easily find the leading order NR operators for each distinct WIMP-nucleus interaction. One can imagine a series of minimal scenarios in which a combination of two Lagrangian couplings that give rise to a direct detection signal is non-zero with all others set to zero, and then proceeding in this manner for the entire set. Each of these scenarios is listed with its leading operators in table V and with all operators generated in table VI. Note that in the case of a complex coupling constant we consider purely real and purely imaginary values as separate cases since they produce a distinct set of operators.

TABLE II. Non-zero c_i coefficients for a spin-0 WIMP

	Uncharged Mediator	Charged Mediator
c_1	$\frac{h_1^N g_1}{m_\phi^2}$	$\frac{y_1^\dagger y_1 - y_2^\dagger y_2}{m_Q m_S} f_T^N$
c_{10}	$\frac{-ih_2^N g_1}{m_\phi^2} + \frac{2ig_4 h_4^N}{m_G^2}\frac{m_N}{m_S}$	$i\frac{y_2^\dagger y_1 - y_1^\dagger y_2}{m_Q m_S}\tilde{\Delta}^N$

TABLE III. c_i coefficients for a spin-$\frac{1}{2}$ WIMP

	Uncharged Mediator	Charged Mediator
c_1	$\frac{h_1^N \lambda_1}{m_\phi^2} - \frac{h_3^N \lambda_3}{m_G^2}$	$\left(\frac{l_2^\dagger l_2 - l_1^\dagger l_1}{4m_\Phi^2} + \frac{d_2^\dagger d_2 - d_1^\dagger d_1}{4m_V^2}\right) f_T^N + \left(-\frac{l_2^\dagger l_2 + l_1^\dagger l_1}{4m_\Phi^2} + \frac{d_2^\dagger d_2 + d_1^\dagger d_1}{8m_V^2}\right)\mathcal{N}^N$
c_4	$\frac{4h_4^N \lambda_4}{m_G^2}$	$\frac{l_2^\dagger l_2 - l_1^\dagger l_1}{m_\Phi^2}\delta^N - \left(\frac{l_1^\dagger l_1 + l_2^\dagger l_2}{m_\Phi^2} + \frac{d_2^\dagger d_2 - d_1^\dagger d_1}{2m_V^2}\right)\Delta^N$
c_6	$\frac{h_2^N \lambda_2 m_N}{m_\phi^2 m_\chi}$	$(\frac{l_1^\dagger l_1 - l_2^\dagger l_2}{4m_\Phi^2} + \frac{d_2^\dagger d_2 - d_1^\dagger d_1}{4m_V^2})\frac{m_N}{m_\chi}\tilde{\Delta}^N$
c_7	$\frac{2h_4^N \lambda_3}{m_G^2}$	$(\frac{l_1^\dagger l_2 - l_2^\dagger l_1}{2m_\Phi^2} + \frac{d_1^\dagger d_2 + d_2^\dagger d_1}{4m_V^2})\Delta^N$
c_8	$-\frac{2h_3^N \lambda_4}{m_G^2}$	$(\frac{l_1^\dagger l_2 - l_2^\dagger l_1}{2m_\Phi^2} - \frac{d_1^\dagger d_2 + d_2^\dagger d_1}{4m_V^2})\mathcal{N}^N$
c_9	$-\frac{2h_4^N \lambda_3 m_N}{m_\chi m_G^2} - \frac{2h_3^N \lambda_4}{m_G^2}$	$(\frac{l_1^\dagger l_2 - l_2^\dagger l_1}{2m_\Phi^2} - \frac{d_1^\dagger d_2 + d_2^\dagger d_1}{4m_V^2})\mathcal{N}^N - (\frac{l_1^\dagger l_2 - l_2^\dagger l_1}{2m_\Phi^2} - \frac{d_1^\dagger d_2 + d_2^\dagger d_1}{4m_V^2})\frac{m_N}{m_\chi}\Delta^N$
c_{10}	$\frac{h_2^N \lambda_1}{m_\phi^2}$	$i(\frac{l_1^\dagger l_2 - l_2^\dagger l_1}{4m_\Phi^2} + \frac{d_2^\dagger d_1 - d_1^\dagger d_2}{4m_V^2})\tilde{\Delta}^N - i\frac{l_1^\dagger l_2 - l_2^\dagger l_1}{m_\Phi^2}\delta^N$
c_{11}	$-\frac{h_1^N \lambda_2 m_N}{m_\phi^2 m_\chi}$	$i(\frac{l_2^\dagger l_1 - l_1^\dagger l_2}{4m_\Phi^2} + \frac{d_2^\dagger d_1 - d_1^\dagger d_2}{4m_V^2})\frac{m_N}{m_\chi} f_T^N + i\frac{l_1^\dagger l_2 - l_2^\dagger l_1}{m_\Phi^2}\frac{m_N}{m_\chi}\delta^N$
c_{12}	0	$\frac{l_2^\dagger l_1 - l_1^\dagger l_2}{m_\Phi^2}\delta^N$

As described earlier, we find that it is important to consider operators beyond those incorporated into the standard spin-independent and spin-dependent formalism, i.e. simple models exist in which one would infer an incorrect rate in current experiments by not including these effects. Also importantly, not all of the NR operators are actually generated at

TABLE IV. c_i coefficients for a spin-1 WIMP

	Uncharged Mediator	Charged Mediator
c_1	$\frac{b_1 h_1^N}{m_\phi^2}$	$\frac{y_3^\dagger y_3 - y_4^\dagger y_4}{m_Q m_X} f_T^N$
c_4	$\frac{4\mathrm{Im}(b_7)h_4^N}{m_G^2} + i\frac{q^2}{m_X^2}\frac{\mathrm{Re}(b_7)h_4^N}{m_G^2} - \frac{q^2}{m_X m_N}\frac{\mathrm{Re}(b_6)h_3^N}{m_G^2}$	$2\frac{y_3^\dagger y_3 - y_4^\dagger y_4}{m_Q m_X}\delta^N$
c_5	$\frac{\mathrm{Re}(b_6)h_3^N}{m_G^2}\frac{m_N}{m_X}$	0
c_6	$\frac{\mathrm{Re}(b_6)h_3^N}{m_G^2}\frac{m_N}{m_X} - i\frac{\mathrm{Re}(b_7)h_4^N}{m_G^2}\frac{m_N^2}{m_X^2}$	0
c_8	$\frac{2\mathrm{Im}(b_7)h_3^N}{m_G^2}$	0
c_9	$-\frac{2\mathrm{Re}(b_6)h_4^N}{m_G^2}\frac{m_N}{m_X} + \frac{2\mathrm{Im}(b_7)h_3^N}{m_G^2}$	0
c_{10}	$\frac{b_1 h_2^N}{m_\phi^2} - \frac{3b_5 h_4^N}{m_G^2}\frac{m_N}{m_X}$	$i\frac{y_4^\dagger y_3 - y_3^\dagger y_4}{m_Q m_X}\tilde{\Delta}^N$
c_{11}	$\frac{\mathrm{Re}(b_7)h_3^N}{m_G^2}\frac{m_N}{m_X}$	$i\frac{y_4^\dagger y_3 - y_3^\dagger y_4}{m_Q m_X}\delta^N$
c_{12}	0	$2i\frac{y_3^\dagger y_4 - y_4^\dagger y_3}{m_Q m_X}\delta^N$
c_{14}	$-\frac{2\mathrm{Re}(b_7)h_4^N}{m_G^2}\frac{m_N}{m_X}$	0
c_{17}	$-\frac{4\mathrm{Im}(b_6)h_3^N}{m_G^2}\frac{m_N}{m_X}$	0
c_{18}	$\frac{4\mathrm{Im}(b_6)h_4^N}{m_G^2}\frac{m_N}{m_X}$	$-2i\frac{y_4^\dagger y_3 - y_3^\dagger y_4}{m_Q m_X}\delta^N$

leading order; for example, the operators $\mathcal{O}_2$, $\mathcal{O}_3$, $\mathcal{O}_{13}$ and $\mathcal{O}_{15}$ are missing at leading order. Note that we only consider renormalizable Lagrangians, higher order non-renormalizable operators, which are presumably further suppressed. We have also not considered the case of kinetic mixing, which could be used to generate anapole interactions [56], because the effective interaction doesn't arise from one mediator exchange.

While spin independent interactions are a generic feature of direct couplings to quarks in our charged mediator cases, it is sometimes possbile to suppress them. In the scalar (and vector) WIMP with charged mediator cases, it is possible to suppress the spin independent interaction by ensuring that $|y_1| = |y_2|(|y_3| = |y_4|)$ while keeping their relative phases non-zero (or π). While these non-minimal scenarios require some fine tuning we include it for completeness and label them y_1, y_2 and y_3, y_4.

Aside from scalar WIMPs, each particular spin produces some non-relativistic operators that are unique to that spin. Also, importantly, the operators $\mathcal{O}_1$ and $\mathcal{O}_{10}$ are generic to all spins. In five cases relativistic operators generate unique non-relativistic operators at

leading order. Therefore distinguishing WIMP scenarios in these cases reduces to experimentally discerning between these operators (see also [59]). Given the likely low statistics of any detection in upcoming direct detection experiments, sub-leading operators are not likely to contribute enough to provide any further discriminating power.

V. OBSERVABLES

The principle observable in direct detection experiments is the differential event rate. Since the incoming WIMPs originate in the galactic halo, one must average over the WIMP velocity distribution, $f(v)$, which we assume for the purposes of this paper to be Maxwell-Boltzmann,

$$\frac{dR}{dE_R} = N_T \frac{\rho_\chi M}{2\pi m_\chi} \int_{v_{min}} \frac{f(v)}{v} P_{tot} dv \tag{29}$$

where we use the value $\rho_\chi = 0.3\text{GeV/cm}^3$ for the local dark matter density, N_T is the number of nuclei in the target and P_{tot} can be calculated from the amplitude $\mathcal{M}$ in Eq. 14

$$P_{tot} = \frac{1}{2j_\chi + 1} \frac{1}{2j_N + 1} \sum_{spins} |\mathcal{M}|^2. \tag{30}$$

Thoughout this work we use the mathematica package supplied in [54] to calculate rates. To determine the leading order operator which arises from a given relativistic scenario we first plot the rate for each of the NR operators in xenon-131. To simply compare the operators we set the c_i coefficients to be the same and normalized the overall rate to that of $\mathcal{O}_1$, see Fig. 1. Since operators are either zero, first or second order in momentum transfer q or velocity $\vec{v}^\perp$, the relative strengths of the operators span 16 orders of magnitude. This is an important point to keep in mind when finding the leading operator, as sometimes a term which appears to be higher order in q can dominate the non-relativistic reduction. For example in the $b_7^{\text{Re}} h_4$ scenario, one finds that $q^2\mathcal{O}_4$ dominates over the $\mathcal{O}_6$ and $\mathcal{O}_{14}$ which contain powers of q within the operators.

Since the Lagrangians we have considered are not tied to specific complete and consistent particle physics models, the mediator masses are not fixed in advance and thus specific event rates are not predicted in advance. Clearly one requires a rate that is low enough

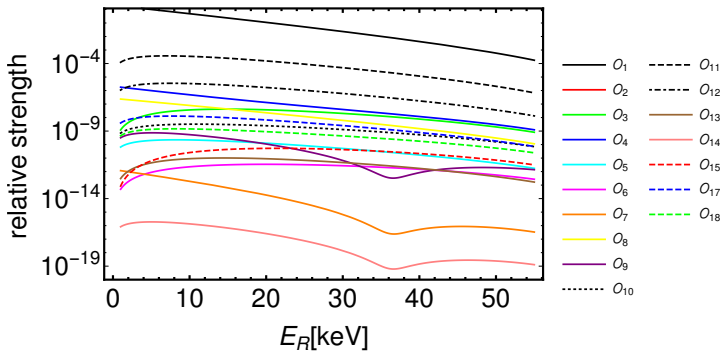

FIG. 1. The relative strength of event rates for a 50GeV spin-$\frac{1}{2}$ WIMP in xenon for each of the non-relativistic operators in table I, where the coefficients of each operator are set to be equal

to evade the current experimental constraints. For example, a 50 GeV WIMP producing 10 events per tonne per year is sufficiently low to evade the bounds from LUX [21]. For demonstration purposes we set the couplings to 0.1 (or $0.1i$ for imaginary) in the various Lagrangians and find a mediator mass that will produce 10 events/t/y in the signal region for xenon $(5-45\text{keV})$. The calculated masses are given in table V. It is perhaps telling that the mediator masses span 6 orders of magnitude, from just a few GeV up to a PeV. While it is unlikely that a full model of thermal relic dark matter could be built around all of these Lagrangians, it is nevertheless a useful metric to estimate the relative strength of the different nuclear responses to each of the operators.

In Figs. 2, 3, 4 and 5 we have plotted rates for two common targets. For simplicity and again for demonstration purposes, we only plot the rates for a single isotope of both germanium and xenon. The choice of isotopes, ^{73}Ge and ^{131}Xe, was made to ensure sensitivity to spin-dependent responses. As can be seen in the figures, many operators produce rates with similar recoil energy dependence in the same target, but different nuclei can have very different responses to the various operators [53]. Thus a complementary choice of nuclear targets can provide important discriminating information.

To illustrate this discriminating power we plot the ratio of the rates in xenon and germanium in Fig. 5 and 6. We choose to only present ratios for the uncharged mediator cases of spinor and vector WIMPs since the other cases produce trival results (all operators being spin independent). To estimate the effect astrophysical uncertainties will have on discriminating between operators, we plot the rate for a range of astrophysical parame-

ters from $v_0 = 200\text{m/s}$, and $v_{esc} = 500\text{m/s}$ (lower) to $v_0 = 240\text{m/s}$ and $v_{esc} = 600\text{m/s}$ (upper). The uncertainty in the dark matter density does not appear since we are considering the ratio of rates. Given the vastly different energy dependence of the ratio of rates of each scenario the astrophysical errors do not completely inhibit their identification. Furthermore, operators $\mathcal{O}_9$ and $\mathcal{O}_{14}$, produced in scenarios $h_4 b_7^{\text{Re}}$ and $h_4 b_6^{\text{Re}}$ respectively, remain indistingushable when considering the ratio of rates. While it appears that in principle almost every operator is discernible, in practice isotopically impure targets and low statistics will further complicate the situation and provide limits on practical discrimination.

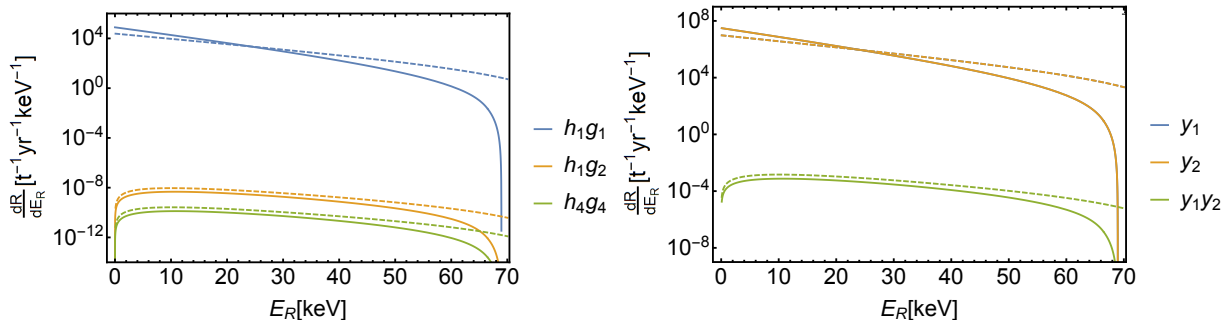

FIG. 2. Rates for a 50GeV spin-0 WIMP in xenon (solid) and germanium (dashed) with uncharged (left) and charged mediators (right), assuming mediator mass of 1TeV and $\mathcal{O}(1)$ coupling constants.

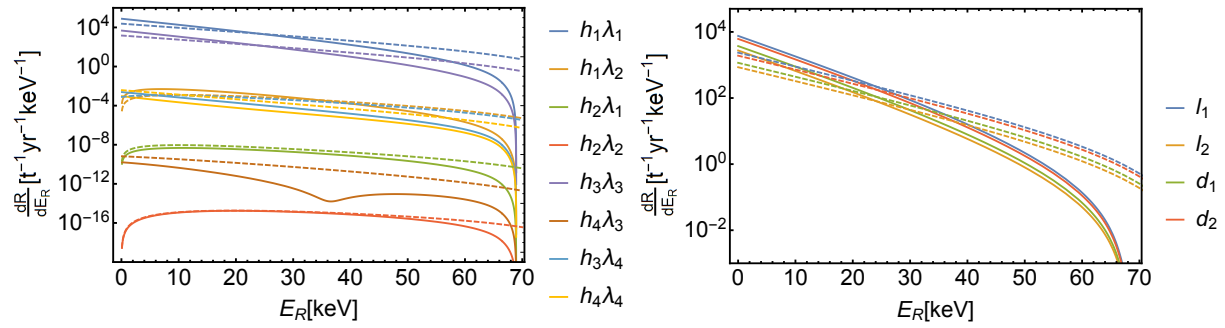

FIG. 3. Rates for a 50GeV spin-$\frac{1}{2}$ WIMP in xenon (solid) and germanium (dashed) with uncharged (left) and charged mediators (right), assuming mediator mass of 1TeV and $\mathcal{O}(1)$ coupling constants.

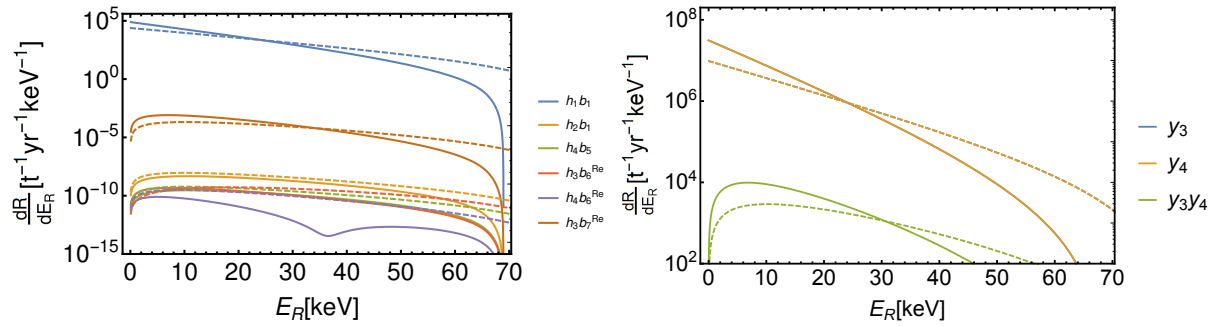

FIG. 4. Rates for a 50GeV spin-1 WIMP in xenon (solid) and germanium (dashed) with uncharged (left) and charged mediators (right), assuming mediator mass of 1TeV and $\mathcal{O}(1)$ coupling constants.

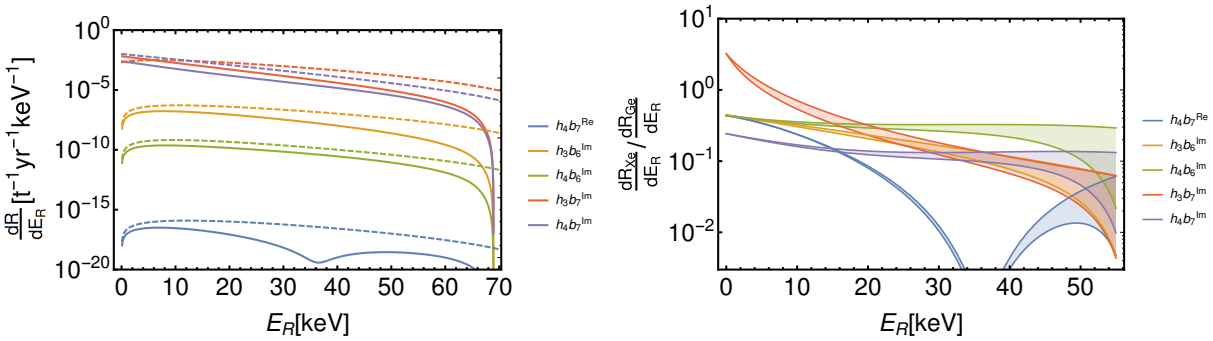

FIG. 5. Rates (left) for a 50GeV spin-1 WIMP in xenon (solid) and germanium (dashed) with uncharged mediators and imaginary couplings, assuming mediator mass of 1TeV and $\mathcal{O}(1)$ coupling constants. Also shown is the ratio of rates in xenon and germanium (right).

VI. CONCLUSION

The analysis we have given here builds on previous analyses to provide, in generality, a roadmap to use event rates in direct dark matter detectors to constrain fundamental dark matter models. We have outlined the steps needed to go from fundamental Lagrangians, first to relativistic operators, then to non-relativistic operators, and finally to produce nuclear matrix elements. In the process several significant facts have been elaborated.

- Not all possible non-relativistic operators contributing to nuclear matrix elements in direct detection will arise from simple UV complete dark matter models.

- Aside from scalar WIMPs each particular spin produces some non-relativistic operators that are unique to that spin.

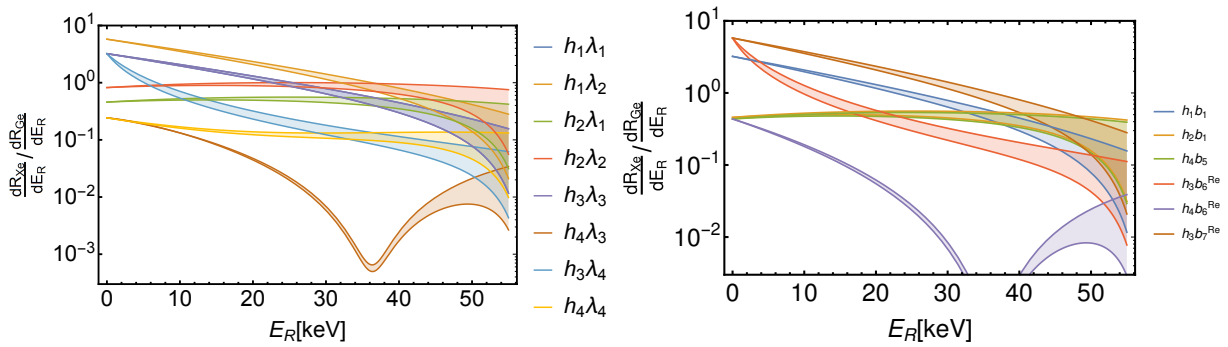

FIG. 6. Ratio of rates in xenon and germanium, illustrating the discriminating power of having multiple nuclear targets. For a 50GeV spin-$\frac{1}{2}$ WIMP with uncharged mediator (left) and a 50GeV spin-1 WIMP with uncharged mediator (right), the shaded regions show the upper and lower bounds due to the astrophysical parameters

- Two non-relativistic operators, $\mathcal{O}_1$ and $\mathcal{O}_{10}$, are ubiquitous and arise for all WIMP spins we have explored.

- In 5 scenarios, relativistic operators generate unique non-relativistic operators at leading order.

- Two new non-relativistic operators not previously considered within the context of the full array of allowed nuclear responses arise at low energies if spin-1 WIMP dark matter is allowed for.

- While the different operators that can contribute to event rates in detectors using specific elements or isotopes cannot be distinguished on the basis of their impact on the differential event rates in these detectors, they can produce radically different energy dependence for scattering off different nuclear targets. Thus, a complementary use of different target materials will be necessary to reliably distinguish between different particle physics model possibilities for WIMP dark matter.

While current detectors have only yielded upper limits, with new generations of larger detectors with greater energy resolution and lower thresholds coming online, the search for WIMP dark matter has never been so vibrant and promising. The tools we have provided here should help experimenters to probe the most useful parameter space, to interpret any non-zero signals in terms of constraints on fundamental models, and should allow theorists

who build fundamental models to frame predictions in an accurate and simple way so that they might be directly compared with experiment.

ACKNOWLEDGMENTS

We would like to thank Richard Lebed for useful discussions. J.B.D. thanks Dr. and Mrs. Sammie W. Cosper at the University of Louisiana at Lafayette, and the Louisiana Board of Regents for support. L.M.K, J.L.N and S.S acknowledge support from the DOE for this work under grant No. DE-SC0008016. We thank the Australian National University for hospitality while part of this work was carried out.

TABLE V. Leading order operators which can arise from the relativistic Lagrangians considered in this work, the column '$\mathcal{L}$ terms' gives the non-zero couplings for that scenario. Each row represents a possible leading order direct detection signal. A '†' indicates that the mediator is charged. The 'Eqv. M_m' column gives the mediator mass required for each scenario to produce ∼10 events $t^{-1}yr^{-1}keV^{-1}$ in xenon, with couplings set to 0.1.

WIMP spin	Mediator spin	$\mathcal{L}$ terms	leading NR operator	Eqv. M_m
0	0	h_1, g_1	$\mathcal{O}_1$	13 TeV
0	0	h_2, g_1	$\mathcal{O}_{10}$	14 GeV
0	1	h_4, g_4	$\mathcal{O}_{10}$	8 GeV
0	$\frac{1}{2}^\dagger$	y_1	$\mathcal{O}_1$	3.2 PeV
0	$\frac{1}{2}^\dagger$	y_2	$\mathcal{O}_1$	3.2 PeV
0	$\frac{1}{2}^\dagger$	y_1, y_2	$\mathcal{O}_{10}$	41 GeV
$\frac{1}{2}$	0	h_1, λ_1	$\mathcal{O}_1$	12.7 TeV
$\frac{1}{2}$	0	h_2, λ_1	$\mathcal{O}_{10}$	293 GeV
$\frac{1}{2}$	0	h_1, λ_2	$\mathcal{O}_{11}$	14 GeV
$\frac{1}{2}$	0	h_2, λ_2	$\mathcal{O}_6$	1.9 GeV
$\frac{1}{2}$	1	h_3, λ_3	$\mathcal{O}_1$	6.3 TeV
$\frac{1}{2}$	1	h_4, λ_3	$\mathcal{O}_9$	6.4 GeV
$\frac{1}{2}$	1	h_3, λ_4	$\mathcal{O}_8$	180 GeV
$\frac{1}{2}$	1	h_4, λ_4	$\mathcal{O}_4$	135 GeV
$\frac{1}{2}$	$0^\dagger$	l_1	$\mathcal{O}_1$	7.1 TeV
$\frac{1}{2}$	$0^\dagger$	l_2	$\mathcal{O}_1$	5.5 TeV
$\frac{1}{2}$	$1^\dagger$	d_1	$\mathcal{O}_1$	5.9 TeV
$\frac{1}{2}$	$1^\dagger$	d_2	$\mathcal{O}_1$	6.7 TeV
1	0	h_1, b_1	$\mathcal{O}_1$	13 TeV
1	0	h_2, b_1	$\mathcal{O}_{10}$	10 GeV
1	1	h_4, b_5	$\mathcal{O}_{10}$	5.1 GeV
1	1	$h_3, b_6^{\mathrm{Re}}(b_6^{\mathrm{Im}})$	$\mathcal{O}_5(\mathcal{O}_{17})$	5.5 GeV(23 GeV)
1	1	$h_4, b_6^{\mathrm{Re}}(b_6^{\mathrm{Im}})$	$\mathcal{O}_9(\mathcal{O}_{18})$	3 GeV(4.6 GeV)
1	1	$h_3, b_7^{\mathrm{Re}}(b_7^{\mathrm{Im}})$	$\mathcal{O}_{11}(\mathcal{O}_8)$	186 GeV(228 GeV)
1	1	$h_4, b_7^{\mathrm{Re}}(b_7^{\mathrm{Im}})$	$\mathcal{O}_4(\mathcal{O}_4)$	78 MeV (172 GeV)
1	$\frac{1}{2}^\dagger$	y_3	$\mathcal{O}_1$	3.2 PeV
1	$\frac{1}{2}^\dagger$	y_4	$\mathcal{O}_1$	3.2 PeV
1	$\frac{1}{2}^\dagger$	y_3, y_4	$\mathcal{O}_{11}$	120 TeV

TABLE VI. List of scenarios with leading operators colored by which are distinguishable via the ratio $\frac{dR_{Xe}}{dE}/\frac{dR_{Ge}}{dE}$.

		$\mathcal{O}_1$	$\mathcal{O}_2$	$\mathcal{O}_3$	$\mathcal{O}_4$	$q^2\mathcal{O}_4$	$\mathcal{O}_5$	$\mathcal{O}_6$	$\mathcal{O}_7$	$\mathcal{O}_8$	$\mathcal{O}_9$	$\mathcal{O}_{10}$	$\mathcal{O}_{11}$	$\mathcal{O}_{12}$	$\mathcal{O}_{13}$	$\mathcal{O}_{14}$	$\mathcal{O}_{15}$	$\mathcal{O}_{17}$	$\mathcal{O}_{18}$
Spin-0 WIMP	(h_1, g_1)	✓																	
	(h_2, g_1)											✓							
	(h_4, g_4)											✓							
	(y_1)	✓										✓							
	(y_2)	✓										✓							
	(y_1, y_2)											✓							
Spin-$\frac{1}{2}$ WIMP	(h_1, λ_1)	✓																	
	(h_2, λ_1)											✓							
	(h_1, λ_2)												✓						
	(h_2, λ_2)							✓											
	(h_3, λ_3)	✓																	
	(h_4, λ_3)								✓		✓								
	(h_3, λ_4)									✓	✓								
	(h_4, λ_4)				✓														[a]
	(l_1)	✓			✓			✓											
	(l_2)	✓			✓			✓											
	(d_1)	✓			✓			✓											
	(d_2)	✓			✓			✓											
Spin-1 WIMP	(h_1, b_1)	✓																	
	(h_2, b_1)											✓							
	(h_4, b_5)											✓							
	(h_3, b_6)					✓	✓	✓										✓*	
	(h_4, b_6)										✓								✓*
	(h_3, b_7)									✓*	✓*		✓						
	(h_4, b_7)				✓*	✓		✓								✓			
	(y_3)	✓			✓							✓	✓	✓					✓
	(y_4)	✓			✓							✓	✓	✓					✓
	(y_3, y_4)											✓	✓	✓					✓

[a] * indicates the purely imaginary scenario for that coupling

Appendix A: Vector Dark Matter

If the WIMP has spin 1, we find two extra operators that haven't been considered previously. Specifically, the operators depend on the symmetric combination of polarization vectors, $S_{ij} = \frac{1}{2}\left(\epsilon_i^\dagger \epsilon_j + \epsilon_j^\dagger \epsilon_i\right)$. This necessitates a modification to the WIMP response functions by first modifying the ℓ coefficients given in Eq. 13. Based on our non-relativistic reduction for vector dark matter, the Lagrangian for vector dark matter and the nucleus, interacting via an uncharged scalar or vector mediator can be written in general as:

$$\mathcal{L}_{vector} = c_1\mathcal{O}_1 + c_4\mathcal{O}_4 + c_5\mathcal{O}_5 + c_8\mathcal{O}_8 + c_9\mathcal{O}_9 + c_{10}\mathcal{O}_{10} + c_{11}\mathcal{O}_{11} + c_{14}\mathcal{O}_{14} + c_{17}\mathcal{O}_{17} + c_{18}\mathcal{O}_{18} \tag{A1}$$

where we've defined $\mathcal{O}_{17} \equiv \frac{i\vec{q}}{m_N} \cdot \mathcal{S} \cdot \vec{v}_\perp$ and $\mathcal{O}_{18} \equiv \frac{i\vec{q}}{m_N} \cdot \mathcal{S} \cdot \vec{S}_N$ and the c_i's are given in table IV. To decompose these new operators we replace $\vec{v}^\perp$ with the target velocity and the internucleon velocities and sum over nucleons. $\mathcal{O}_{17}$ can then be put into the form

$$\mathcal{O}_{17} \to \frac{i\vec{q}}{m_N}.\mathcal{S}.\left[\vec{v}_T^\perp e^{-i\vec{q}.\vec{x}_i} - \sum_{i=1}^{A}\frac{1}{2M}\left(-\frac{1}{i}\overleftarrow{\nabla}_i e^{-i\vec{q}.\vec{x}_i} + e^{-i\vec{q}.\vec{x}_i}\frac{1}{i}\overrightarrow{\nabla}_i\right)_{int}\right]. \tag{A2}$$

$\mathcal{O}_{18}$ can be expanded as

$$\mathcal{O}_{18} \to \frac{1}{2}\frac{i\vec{q}}{m_N} \cdot \mathcal{S} \cdot \vec{\sigma} \tag{A3}$$

Together, all the terms of $\mathcal{L}_{vector}$ give rise to the following ℓ factors from Eq. 13,

$$\begin{aligned}
\ell_0^\tau &= c_1^\tau + i\left(\frac{\vec{q}}{m_N} \times \vec{v}_T^\perp\right) \cdot \vec{S}_\chi c_5^\tau + (\vec{v}_T^\perp \cdot \vec{S}_\chi)c_8^\tau + i\left(\frac{\vec{q}}{m_N} \cdot \vec{S}_\chi\right)c_{11}^\tau + i\left(\frac{\vec{q}}{m_N} \cdot \mathcal{S} \cdot \vec{v}_\perp^T\right)c_{17}^\tau \\
l_0^{A\tau} &= -i\left(\frac{\vec{q}}{2m_N} \cdot \vec{S}_\chi\right)c_{14}^\tau \\
\vec{l}_E^\tau &= 0 \\
\vec{l}_M^\tau &= i\left(\frac{\vec{q}}{m_N} \times \vec{S}_\chi\right)c_5^\tau - \vec{S}_\chi c_8^\tau - i\left(\frac{\vec{q}}{m_N} \cdot \mathcal{S}\right)c_{17}^\tau \\
\vec{l}_5^\tau &= \frac{1}{2}\vec{S}_\chi c_4^\tau + i\left(\frac{\vec{q}}{m_N} \times \vec{S}_\chi\right)c_9^\tau + \frac{1}{2}\left(i\frac{\vec{q}}{m_N}\right)c_{10}^\tau + \frac{1}{2}\vec{v}_T^\perp\left(\frac{\vec{q}}{2m_N} \cdot \vec{S}_\chi\right)c_{14}^\tau + \frac{1}{2}\left(i\frac{\vec{q}}{m_N} \cdot \mathcal{S}\right)c_{18}^\tau
\end{aligned} \tag{A4}$$

Based on the ℓ's above, the coefficients of the various nuclear responses are found by squaring the amplitude and then summing over spins. To simplify calculations, we choose a convenient basis for polarization vectors, $\epsilon_i^s = \delta_i^s$. Recall that the spin can then be written as the anti-symmetric combination $iS_k = \epsilon_{ijk}\epsilon_i^\dagger\epsilon_j$. The WIMP responses unique to the vector case are

then given by:

$$
\begin{aligned}
R_M^{\tau\tau'} &= c_1^\tau c_1^{\tau'} + \frac{2}{3}\left(\frac{\vec{q}^2}{m_N^2} v_T^{\perp 2} c_5^\tau c_5^{\tau'} + v_T^{\perp 2} c_8^\tau c_8^{\tau'} + \frac{q^2}{m_N^2} c_{11}^\tau c_{11}^{\tau'} + \frac{q^2 v_T^{\perp 2}}{4m_N^2} c_{17}^\tau c_{17}^{\tau'}\right) \\
R_{\Phi''}^{\tau\tau'} &= 0 \\
R_{\Phi'' M}^{\tau\tau'} &= 0 \\
R_{\tilde{\Phi}'}^{\tau\tau'} &= 0 \\
R_{\Sigma''}^{\tau\tau'} &= \frac{1}{6} c_4^\tau c_4^{\tau'} + \frac{q^2}{4m_N^2} c_{10}^\tau c_{10}^{\tau'} + \frac{q^2}{12m_N^2} c_{18}^\tau c_{18}^{\tau'} \\
R_{\Sigma'}^{\tau\tau'} &= \frac{1}{6} c_4^\tau c_4^{\tau'} + \frac{q^2}{6m_N^2} c_9^\tau c_9^{\tau'} + \frac{q^2 v_T^{\perp 2}}{2m_N^2} c_{14}^\tau c_{14}^{\tau'} + \frac{q^2}{24m_N^2} c_{18}^\tau c_{18}^{\tau'} \\
R_{\Delta}^{\tau\tau'} &= \frac{2}{3}\left(\frac{\vec{q}^2}{m_N^2} c_5^\tau c_5^{\tau'} + c_8^\tau c_8^{\tau'}\right) + \frac{q^2}{6m_N^2} c_{17}^\tau c_{17}^{\tau'} \\
R_{\Delta\Sigma'}^{\tau\tau'} &= \frac{2}{3}\left(c_5^\tau c_4^{\tau'} - c_8^\tau c_9^{\tau'}\right).
\end{aligned}
\tag{A5}
$$

Appendix B: Non-relativistic Reduction

We find effective relativistic interaction Lagrangians by integrating out heavy mediators. We only keep the leading order interactions (suppressed by m or m^2). To the right of each operator is their non-relativistic reduction expressed in terms of the operators in table I with the coefficient derived from the Lagrangian parameters along with the relevant nucleon form factor. As multiple operators can have the same non-relativistic limit, it is important to include the nucleon form factor at the relativistic level. If this is not performed, erroneous cancellations can occur.

For free spinors we use the Bjorken and Drell normalization and γ matrix conventions. In the non-relativistic limit we make the following replacements:

$$
\begin{aligned}
S &\to \frac{1_S}{\sqrt{m_S}} \\
X_\mu &\to \frac{\epsilon_\mu^s}{\sqrt{m_X}} \\
\chi &\to \sqrt{\frac{E+m_\chi}{2m_\chi}} \begin{pmatrix} \xi \\ \frac{\vec{\sigma}\cdot\vec{p}}{E+m_\chi}\xi \end{pmatrix}
\end{aligned}
\tag{B1}
$$

where $s = 1, 2, 3$ are the different polarization states of the vector. $\xi = (1\ 0)^T$ is the left

handed Weyl spinor. The following Fierz transformation and gamma matrix identites were useful in the charged mediator cases, (a sign difference was found in the final identity when compared with [60]):

$$\begin{aligned}
(\bar{q}\chi)(\bar{\chi}q) =& -\frac{1}{4}\left[\bar{q}q\bar{\chi}\chi + \bar{q}\gamma^\mu q\bar{\chi}\gamma_\mu\chi + \frac{1}{2}\bar{q}\sigma^{\mu\nu}q\bar{\chi}\sigma_{\mu\nu}\chi - \bar{q}\gamma^\mu\gamma^5 q\bar{\chi}\gamma_\mu\gamma^5\chi + \bar{q}\gamma^5 q\bar{\chi}\gamma^5\chi\right] \\
(\bar{q}\gamma^5\chi)(\bar{\chi}\gamma^5 q) =& -\frac{1}{4}\left[\bar{q}q\bar{\chi}\chi + \bar{q}\gamma^5 q\bar{\chi}\gamma^5\chi - \bar{q}\gamma^\mu q\bar{\chi}\gamma_\mu\chi + \bar{q}\gamma^\mu\gamma^5 q\bar{\chi}\gamma_\mu\gamma^5\chi + \frac{1}{2}\bar{q}\sigma^{\mu\nu}q\bar{\chi}\sigma_{\mu\nu}\chi\right] \\
(\bar{q}\chi)(\bar{\chi}\gamma^5 q) =& -\frac{1}{4}\left[\bar{q}q\bar{\chi}\gamma^5\chi + \bar{q}\gamma^5 q\bar{\chi}\chi - \bar{q}\gamma^\mu q\bar{\chi}\gamma_\mu\gamma^5\chi + \bar{q}\gamma^\mu\gamma^5 q\bar{\chi}\gamma_\mu\chi + i\epsilon_{\mu\nu\alpha\beta}\bar{q}\sigma^{\mu\nu}q\bar{\chi}\sigma^{\alpha\beta}\chi\right] \\
(\bar{q}\gamma_\mu\chi)(\bar{\chi}\gamma^\mu q) =& -\left[\bar{q}q\bar{\chi}\chi - \bar{q}\gamma^5 q\bar{\chi}\gamma^5\chi - \frac{1}{2}\bar{q}\gamma^\mu q\bar{\chi}\gamma_\mu\chi - \frac{1}{2}\bar{q}\gamma^\mu\gamma^5 q\bar{\chi}\gamma_\mu\gamma^5\chi\right] \\
(\bar{q}\gamma_\mu\gamma^5\chi)(\bar{\chi}\gamma^\mu\gamma^5 q) =& -\left[-\bar{q}q\bar{\chi}\chi + \bar{q}\gamma^5 q\bar{\chi}\gamma^5\chi - \frac{1}{2}\bar{q}\gamma^\mu q\bar{\chi}\gamma_\mu\chi - \frac{1}{2}\bar{q}\gamma^\mu\gamma^5 q\bar{\chi}\gamma_\mu\gamma^5\chi\right] \\
(\bar{q}\gamma_\mu\chi)(\bar{\chi}\gamma^\mu\gamma^5 q) =& -\left[\bar{q}q\bar{\chi}\gamma^5\chi - \bar{q}\gamma^5 q\bar{\chi}\chi + \frac{1}{2}\bar{q}\gamma^\mu q\bar{\chi}\gamma_\mu\gamma^5\chi + \frac{1}{2}\bar{q}\gamma^\mu\gamma^5 q\bar{\chi}\gamma_\mu\chi\right]
\end{aligned} \tag{B2}$$

$$\sigma^{\mu\nu}\gamma^5 = \frac{i}{2}\epsilon^{\mu\nu\rho\sigma}\sigma_{\rho\sigma} \tag{B3}$$

All of the following operators are collected in terms of the coefficients of the NR operators, c_i, in tables II,III and IV.

TABLE VII. Non-relativistic reduction of operators for a spin-0 WIMP

Scalar Mediator	
$(S^\dagger S)(\bar{q}q)$	$\longrightarrow \left(\frac{h_1^N g_1}{m_\phi^2}\right)\mathcal{O}_1$
$(S^\dagger S)(\bar{q}\gamma^5 q)$	$\longrightarrow \left(\frac{h_2^N g_1}{m_\phi^2}\right)\mathcal{O}_{10}$
Vector Mediator	
$i(S^\dagger\partial_\mu S - \partial_\mu S^\dagger S)(\bar{q}\gamma^\mu q)$	$\longrightarrow 0$
$i(S^\dagger\partial_\mu S - \partial_\mu S^\dagger S)(\bar{q}\gamma^\mu\gamma^5 q)$	$\longrightarrow \left(\frac{2ig_4h_4^N}{m_G^2}\frac{m_N}{m_S}\right)\mathcal{O}_{10}$
Charged Spinor Mediator	
$(S^\dagger S)(\bar{q}q)$	$\longrightarrow \frac{y_1^\dagger y_1 - y_2^\dagger y_2}{m_Q m_S} f_T^N \mathcal{O}_1$
$(S^\dagger S)(\bar{q}\gamma^5 q)$	$\longrightarrow i\frac{y_2^\dagger y_1 - y_1^\dagger y_2}{m_Q m_S}\tilde{\Delta}^N\mathcal{O}_{10}$

TABLE VIII. Operators for a spin-$\frac{1}{2}$ WIMP via a neutral mediator

Scalar Mediator	
$\bar{\chi}\chi\bar{q}q$	$\longrightarrow \left(\frac{h_1^N \lambda_1}{m_\phi^2}\right) \mathcal{O}_1$
$\bar{\chi}\chi\bar{q}\gamma^5 q$	$\longrightarrow \left(\frac{h_2^N \lambda_1}{m_\phi^2}\right) \mathcal{O}_{10}$
$\bar{\chi}\gamma^5\chi\bar{q}q$	$\longrightarrow \left(-\frac{h_1^N \lambda_2 m_N}{m_\phi^2 m_\chi}\right) \mathcal{O}_{11}$
$\bar{\chi}\gamma^5\chi\bar{q}\gamma^5 q$	$\longrightarrow \left(\frac{h_2^N \lambda_2 m_N}{m_\phi^2 m_\chi}\right) \mathcal{O}_6$
Vector Mediator	
$\bar{\chi}\gamma^\mu\chi\bar{q}\gamma_\mu q$	$\longrightarrow \left(-\frac{h_3^N \lambda_3}{m_G^2}\right) \mathcal{O}_1$
$\bar{\chi}\gamma^\mu\chi\bar{q}\gamma_\mu\gamma^5 q$	$\longrightarrow \left(-\frac{2h_4^N \lambda_3}{m_G^2}\right) \left(-\mathcal{O}_7 + \frac{m_N}{m_\chi}\mathcal{O}_9\right)$
$\bar{\chi}\gamma^\mu\gamma^5\chi\bar{q}\gamma_\mu q$	$\longrightarrow \left(-\frac{2h_3^N \lambda_4}{m_G^2}\right) (\mathcal{O}_8 + \mathcal{O}_9)$
$\bar{\chi}\gamma^\mu\gamma^5\chi\bar{q}\gamma_\mu\gamma^5 q$	$\longrightarrow \left(\frac{4h_4^N \lambda_4}{m_G^2}\right) \mathcal{O}_4$

TABLE IX. Non-relativistic reduction of operators for a spin-$\frac{1}{2}$ WIMP via a charged mediator (after using Fierz identities)

Charged Scalar Mediator	
$\bar{\chi}\chi\bar{q}q$	$\longrightarrow \frac{l_2^\dagger l_2 - l_1^\dagger l_1}{4m_\Phi^2} f_{Tq}^N \mathcal{O}_1$
$\bar{\chi}\chi\bar{q}\gamma^5 q$	$\longrightarrow i\frac{l_1^\dagger l_2 - l_2^\dagger l_1}{4m_\Phi^2} \Delta\tilde{q}^N \mathcal{O}_{10}$
$\bar{\chi}\gamma^5\chi\bar{q}q$	$\longrightarrow i\frac{l_2^\dagger l_1 - l_1^\dagger l_2}{4m_\Phi^2} \frac{m_N}{m_\chi} f_{Tq}^N \mathcal{O}_{11}$
$\bar{\chi}\gamma^5\chi\bar{q}\gamma^5 q$	$\longrightarrow \frac{l_1^\dagger l_1 - l_2^\dagger l_2}{4m_\Phi^2} \frac{m_N}{m_\chi} \Delta\tilde{q}^N \mathcal{O}_6$
$\bar{\chi}\gamma^\mu\chi\bar{q}\gamma_\mu q$	$\longrightarrow -\frac{l_1^\dagger l_1 + l_2^\dagger l_2}{4m_\Phi^2} \mathcal{N}_q^N \mathcal{O}_1$
$\bar{\chi}\gamma^\mu\gamma^5\chi\bar{q}\gamma_\mu q$	$\longrightarrow \frac{l_1^\dagger l_2 + l_2^\dagger l_1}{2m_\Phi^2} \mathcal{N}_q^N (\mathcal{O}_8 + \mathcal{O}_9)$
$\bar{\chi}\gamma^\mu\chi\bar{q}\gamma_\mu\gamma^5 q$	$\longrightarrow \frac{l_1^\dagger l_2 + l_2^\dagger l_1}{2m_\Phi^2} \Delta_q^N (\mathcal{O}_7 - \frac{m_N}{m_\chi}\mathcal{O}_9)$
$\bar{\chi}\gamma^\mu\gamma^5\chi\bar{q}\gamma_\mu\gamma^5 q$	$\longrightarrow -\frac{l_1^\dagger l_1 + l_2^\dagger l_2}{m_\Phi^2} \Delta_q^N \mathcal{O}_4$
$\bar{\chi}\sigma^{\mu\nu}\chi\bar{q}\sigma_{\mu\nu}q$	$\longrightarrow \frac{l_2^\dagger l_2 - l_1^\dagger l_1}{m_\Phi^2} \delta_q^N \mathcal{O}_4$
$\epsilon_{\mu\nu\alpha\beta}\bar{\chi}\sigma^{\mu\nu}\chi\bar{q}\sigma^{\alpha\beta}q$	$\longrightarrow \frac{l_2^\dagger l_1 - l_1^\dagger l_2}{m_\Phi^2} \delta_q^N (i\mathcal{O}_{10} - i\frac{m_N}{m_\chi}\mathcal{O}_{11} + 4\mathcal{O}_{12})$
Charged Vector Mediator	
$\bar{\chi}\chi\bar{q}q$	$\longrightarrow \frac{d_2^\dagger d_2 - d_1^\dagger d_1}{4m_V^2} f_{Tq}^N \mathcal{O}_1$
$\bar{\chi}\chi\bar{q}\gamma^5 q$	$\longrightarrow i\frac{d_2^\dagger d_1 - d_1^\dagger d_2}{4m_V^2} \Delta\tilde{q}^N \mathcal{O}_{10}$
$\bar{\chi}\gamma^5\chi\bar{q}q$	$\longrightarrow i\frac{d_2^\dagger d_1 - d_1^\dagger d_2}{4m_V^2} \frac{m_N}{m_\chi} f_{Tq}^N \mathcal{O}_{11}$
$\bar{\chi}\gamma^5\chi\bar{q}\gamma^5 q$	$\longrightarrow \frac{d_2^\dagger d_2 - d_1^\dagger d_1}{4m_V^2} \frac{m_N}{m_\chi} \Delta\tilde{q}^N \mathcal{O}_6$
$\bar{\chi}\gamma^\mu\chi\bar{q}\gamma_\mu q$	$\longrightarrow \frac{d_2^\dagger d_2 + d_1^\dagger d_1}{8m_V^2} \mathcal{N}_q^N \mathcal{O}_1$
$\bar{\chi}\gamma^\mu\gamma^5\chi\bar{q}\gamma_\mu q$	$\longrightarrow -\frac{d_2^\dagger d_1 + d_1^\dagger d_2}{4m_V^2} \mathcal{N}_q^N (\mathcal{O}_8 + \mathcal{O}_9)$
$\bar{\chi}\gamma^\mu\chi\bar{q}\gamma_\mu\gamma^5 q$	$\longrightarrow \frac{d_2^\dagger d_1 + d_1^\dagger d_2}{4m_V^2} \Delta_q^N (\mathcal{O}_7 - \frac{m_N}{m_\chi}\mathcal{O}_9)$
$\bar{\chi}\gamma^\mu\gamma^5\chi\bar{q}\gamma_\mu\gamma^5 q$	$\longrightarrow -\frac{d_2^\dagger d_2 + d_1^\dagger d_1}{2m_V^2} \Delta_q^N \mathcal{O}_4$

TABLE X. Non-relativistic reduction of operators for a spin-1 WIMP

Scalar Mediator	
$X_\mu^\dagger X^\mu \bar{q} q$	$\longrightarrow \left(\frac{b_1 h_1^N}{m_\phi^2}\right) \mathcal{O}_1$
$X_\mu^\dagger X^\mu \bar{q}\gamma^5 q$	$\longrightarrow \left(\frac{b_1 h_2^N}{m_\phi^2}\right) \mathcal{O}_{10}$
Vector Mediator	
$(X_\nu^\dagger \partial_\mu X^\nu - \partial_\mu X_\nu^\dagger X^\nu)(\bar{q}\gamma^\mu q)$	$\longrightarrow 0$
$(X_\nu^\dagger \partial_\mu X^\nu - \partial_\mu X_\nu^\dagger X^\nu)(\bar{q}\gamma^\mu\gamma^5 q)$	$\longrightarrow \left(\frac{-3b_5 h_4^N}{m_G^2}\frac{m_N}{m_X}\right) \mathcal{O}_{10}$
$\partial_\nu(X^{\nu\dagger} X_\mu + X_\mu^\dagger X^\nu)(\bar{q}\gamma^\mu q)$	$\longrightarrow \left(\frac{\mathrm{Re}(b_6) h_3^N}{m_G^2}\frac{m_N}{m_X}\right) (\mathcal{O}_5 + \mathcal{O}_6 - \frac{q^2}{m_N^2}\mathcal{O}_4)$
$\partial_\nu(X^{\nu\dagger} X_\mu + X_\mu^\dagger X^\nu)(\bar{q}\gamma^\mu\gamma^5 q)$	$\longrightarrow \left(-\frac{2\mathrm{Re}(b_6) h_4^N}{m_G^2}\frac{m_N}{m_X}\right) \mathcal{O}_9$
$\partial_\nu(X^{\nu\dagger} X_\mu - X_\mu^\dagger X^\nu)(\bar{q}\gamma^\mu q)$	$\longrightarrow \left(-\frac{4\mathrm{Im}(b_6) h_3^N}{m_G^2}\frac{m_N}{m_X}\right) \mathcal{O}_{17}$
$\partial_\nu(X^{\nu\dagger} X_\mu - X_\mu^\dagger X^\nu)(\bar{q}\gamma^\mu\gamma^5 q)$	$\longrightarrow \left(\frac{4\mathrm{Im}(b_6) h_4^N}{m_G^2}\frac{m_N}{m_X}\right) \mathcal{O}_{18}$
$\epsilon_{\mu\nu\rho\sigma}\left(X^{\nu\dagger}\partial^\rho X^\sigma + X^\nu \partial^\rho X^{\sigma\dagger}\right)(\bar{q}\gamma^\mu q)$	$\longrightarrow \left(\frac{\mathrm{Re}(b_7) h_3^N}{m_G^2}\frac{m_N}{m_X}\right) \mathcal{O}_{11}$
$\epsilon_{\mu\nu\rho\sigma}\left(X^{\nu\dagger}\partial^\rho X^\sigma + X^\nu \partial^\rho X^{\sigma\dagger}\right)(\bar{q}\gamma^\mu\gamma^5 q)$	$\longrightarrow \left(\frac{\mathrm{Re}(b_7) h_4^N}{m_G^2}\frac{m_N}{m_X}\right) (i\frac{q^2}{m_X m_N}\mathcal{O}_4 - i\frac{m_N}{m_X}\mathcal{O}_6 - 2\mathcal{O}_{14})$
$\epsilon_{\mu\nu\rho\sigma}\left(X^{\nu\dagger}\partial^\rho X^\sigma - X^\nu \partial^\rho X^{\sigma\dagger}\right)(\bar{q}\gamma^\mu q)$	$\longrightarrow \left(\frac{2\mathrm{Im}(b_7) h_3^N}{m_G^2}\right) (\mathcal{O}_8 + \mathcal{O}_9)$
$\epsilon_{\mu\nu\rho\sigma}\left(X^{\nu\dagger}\partial^\rho X^\sigma - X^\nu \partial^\rho X^{\sigma\dagger}\right)(\bar{q}\gamma^\mu\gamma^5 q)$	$\longrightarrow \left(\frac{4\mathrm{Im}(b_7) h_4^N}{m_G^2}\right) \mathcal{O}_4$
Charged Spinor Mediator	
$(X_\mu^\dagger X_\nu)(\bar{q}\gamma^\mu\gamma^\nu q)$	$\longrightarrow \left(\frac{y_3^\dagger y_3 - y_4^\dagger y_4}{m_Q m_X}\right) \left(f_{Tq}^N \mathcal{O}_1 + 2\delta_q^N \mathcal{O}_4\right)$
$(X_\mu^\dagger X_\nu)(\bar{q}\gamma^\mu\gamma^\nu\gamma^5 q)$	$\longrightarrow \left(\frac{y_4^\dagger y_3 - y_3^\dagger y_4}{m_Q m_X}\right) (i\Delta_{\tilde{q}}^N \mathcal{O}_{10} + i\delta_q^N \mathcal{O}_{11} - 2i\delta_q^N \mathcal{O}_{12} - 2i\delta_q^N \mathcal{O}_{18})$

Appendix C: Quarks to Nucleons

To go from the fundamental interactions of WIMPs with quarks to scattering from point-like nucleons, one must evaluate the quark (parton) bilinears in the nucleons. For a full discussion see the appendix of [60] and [61]. We write the nucleon couplings in terms of the quark couplings times a form factor (in the limit of zero momentum transfer): The scalar bilinear for light quarks can be evaluated from

$$\langle N | \, m_q \bar{q} q \, | N \rangle = m_N f_{Tq}^N \tag{C1}$$

$$
\begin{aligned}
\langle N_o|\, m_q\bar{q}q\,|N_i\rangle &\longrightarrow f_{Tq}^N \bar{N}N \\
\langle N_o|\, \bar{q}\gamma^5 q\,|N_i\rangle &\longrightarrow \Delta\tilde{q}^N \bar{N}\gamma^5 N \\
\langle N_o|\, \bar{q}\gamma^\mu q\,|N_i\rangle &\longrightarrow \mathcal{N}_q^N \bar{N}\gamma^\mu N \\
\langle N_o|\, \bar{q}\gamma^\mu\gamma^5 q\,|N_i\rangle &\longrightarrow \Delta_q^N \bar{N}\gamma^\mu\gamma^5 N \\
\langle N_o|\, \bar{q}\sigma^{\mu\nu} q\,|N_i\rangle &\longrightarrow \delta_q^N \bar{N}\sigma^{\mu\nu} N
\end{aligned}
$$

while for the heavy quarks

$$
\langle N|\, m_q\bar{q}q\,|N\rangle = \frac{2}{27} m_N F_{TG}^N = \frac{2}{27} m_N \left(1 - \sum_{q=u,d,s} f_{Tq}^N\right). \tag{C2}
$$

Summing over all the quarks one finds

$$
h_1^N = \sum_{q=u,d,s} h_1^q \frac{m_N}{m_q} f_{Tq}^N + \frac{2}{27} f_{TG}^N \sum_{q=c,b,t} h_1^q \frac{m_N}{m_q} \tag{C3}
$$

The psuedo-scalar bilinear was recently revisited in [61]:

$$
h_2^N = \sum_{q=u,d,s} h_2^q \Delta\tilde{q}^N - \Delta\tilde{G}^N \sum_{q=c,b,t} \frac{h_2^q}{m_q} \tag{C4}
$$

The vector bilinear essentially gives the number operator:

$$
h_3^N = \begin{cases} 2h_3^u + h_3^d & N = p \\ h_3^u + 2h_3^d & N = u \end{cases} \tag{C5}
$$

The psuedo-vector bilinear counts the contributions of spin to the nucleon (note that sometimes this coupling has a G_F factored out to make it dimensionless)

$$
h_4^N = \sum_{q=u,d,s} h_4^q \Delta_q^N \tag{C6}
$$

Throughout this paper the following values are used (it should be noted that there are

large uncertainties in these values) [60, 61]:

$$\begin{aligned}
f^n_{Tu} &= \ 0.014 & f^p_{Tu} &= 0.02 \\
f^n_{Td} &= \ 0.036 & f^p_{Td} &= 0.026 \\
f^n_{Ts} &= \ 0.118 & f^p_{Ts} &= 0.118 \\
\Delta^n_u &= -0.427 & \Delta^p_u &= 0.842 \\
\Delta^n_d &= \ 0.842 & \Delta^p_d &= -0.427 \\
\Delta^n_s &= -0.085 & \Delta^p_s &= -0.085 \\
\Delta\tilde{u}^n &= -108.03 & \Delta\tilde{u}^p &= 110.55 \\
\Delta\tilde{d}^n &= \ 108.60 & \Delta\tilde{d}^p &= -107.17 \\
\Delta\tilde{s}^n &= -0.57 & \Delta\tilde{s}^p &= -3.37 \\
\Delta\tilde{G}^n &= 35.7\text{MeV} & \Delta\tilde{G}^p &= 395.2\text{MeV}
\end{aligned} \tag{C7}$$

Assuming a universal coupling of the mediators to all quarks, the nucleon level couplings can then be written as,

$$\begin{aligned}
h^N_1 &= f^N_T h_1 \\
h^N_2 &= \tilde{\Delta}^N h_2 \\
h^N_3 &= \mathcal{N}^N h_3 \\
h^N_4 &= \Delta^N h_4
\end{aligned} \tag{C8}$$

where we have defined,

$$\begin{aligned}
f^n_T &= 11.93 & f^p_T &= 12.31 \\
\tilde{\Delta}^n &= -0.07 & \tilde{\Delta}^p &= -0.28 \\
\mathcal{N}^n &= \ 3 & \mathcal{N}^p &= 3 \\
\Delta^n &= \ 0.33 & \Delta^p &= 0.33 \\
\delta^n &= 0.564 & \delta^p &= 0.564
\end{aligned}$$

. (C9)

This introduces a small amount of isospin violation, and it is known that relaxing the assumption of universal couplings to quarks can lead to interesting isospin violating effects [61, 62].

---

[1] Y. Sofue and V. Rubin, Ann.Rev.Astron.Astrophys. **39** (2001) 137-174, astro-ph/0010594.

[2] 2dFGRS Collaboration (S. Cole *et al.*), Mon.Not.Roy.Astron.Soc. **362** (2005) 505-534, astro-ph/0501174.

[3] F. Beutler *et al.*, Mon.Not.Roy.Astron.Soc. **416** (2011) 3017-3032, arXiv:1106.3366 [astro-ph.CO].

[4] L. Anderson *et al.*, Mon.Not.Roy.Astron.Soc. **427** (2013) 4, 3435-3467, arXiv:1203.6594 [astro-ph.CO].

[5] A. Vikhlinin *et al.*, Astrophys.J. **692** (2009) 1060-1074, arXiv:0812.2720 [astro-ph].

[6] L. Fu *et al.*, Astron.Astrophys. **479** (2008) 9-25, arXiv:0712.0884 [astro-ph].

[7] R. Massey *et al.*, Astrophys.J. Suppl. **172** (2007) 239-253, astro-ph/0701480.

[8] Planck Collaboration (P.A.R. Ade *et al.*), "Planck 2013 results. XVI. Cosmological parameters", Astron.Astrophys. (2014), arXiv:1303.5076 [astro-ph.CO].

[9] PAMELA Collaboration (O. Adriani *et al.*), Nature **458** (2009) 607-609, arXiv:0810.4495 [astro-ph].

[10] M. Su, T.R. Slatyer, and D.P. Finkbeiner, Astrophys.J. **724** (2010) 1044-1082, arXiv:1005.5480 [astro-ph.HE].

[11] D. Hooper and L. Goodenough, Phys.Lett.B **697** (2011) 412-428, arXiv:1010.2752 [hep-ph].

[12] C. Weniger, JCAP **1208** (2012) 007, arXiv:1204.2797 [hep-ph].

[13] E. Bulbul, M. Markevitch, A. Foster, R.K. Smith, M. Loewenstein, and S.W. Randall, Astrophys.J. **789** (2014) 13, arXiv:1402.2301 [astro-ph.CO].

[14] ATLAS Collaboration (G. Aad *et al.*), Phys.Rev. D**90** (2014) 1, 012004, arXiv:1404.0051.

[15] ATLAS Collaboration (G. Aad *et al.*), Eur.Phys.J. C**75** (2015) 2, 92, arXiv:1410.4031.

[16] ATLAS Collaboration (G. Aad *et al.*), arXiv:1502.01518.

[17] ATLAS Collaboration (G. Aad *et al.*), Phys.Rev.Lett. **112** (2014) 4, 041802, arXiv:1309.4017.

[18] CMS Collaboration (S. Lowette for the collaboration), arXiv:1410.3762.

[19] M.W. Goodman and E. Witten, Phys.Rev.D **31** (1985) 3059.

[20] B. Cabrera, L. M. Krauss and F. Wilczek, Phys. Rev. Lett. **55**, 25 (1985).

[21] LUX Collaboration (D.S. Akerib *et al.*), Phys.Rev.Lett. **112** (2014) 091303, arXiv:1310.8214 [astro-ph.CO].

[22] XENON100 Collaboration (E. Aprile *et al.*), Phys.Rev.Lett. **109** (2012) 181301, arXiv:1207.5988 [astro-ph.CO].

[23] CDMS-II Collaboration (Z. Ahmed *et al.*), Phys.Rev.Lett. **106** (2011) 131302, arXiv:1011.2482 [astro-ph.CO].

[24] CDMS Collaboration (R. Agnese *et al.*), Phys.Rev.Lett. **111** (2013) 251301, arXiv:1304.4279 [hep-ex].

[25] DAMA Collaboration (R. Barbieri *et al.*), Eur.Phys.J. C**56** (2008) 333-355, arXiv:0804.2741 [astro-ph].

[26] CoGeNT Collaboration (C.E. Aalseth *et al.*), Phys.Rev.D **88** (2013) 1, 012002, arXiv:1208.5737 [astro-ph.CO].

[27] G. Angloher *et al.*, Eur.Phys.J. C**72** (2012) 1971, arXiv:1109.0702 [astro-ph.CO].

[28] J. Sander *et al.*, AIP Conf.Pocs. **1534** (2012) 129-135.

[29] SuperCDMS Collaboration (R. Agnese *et al.*), Phys.Rev.Lett. **112** (2014) 4, 041302, arXiv:1309.3259 [physics.ins-det].

[30] SuperCDMS Collaboration (R. Agnese *et al.*), Phys.Rev.Lett. **112** (2014) 241302, arXiv:1402.7137 [hep-ex].

[31] XENON1T Collaboration (E. Aprile *et al.*), Springer Proc.Phys. **148** (2013) 93-96, arXiv:1206.6288 [astro-ph.IM].

[32] DARWIN Consortium Collaboration (L. Baudis for the collaboration), J.Phys.Conf.Ser. **375** (2012) 012028, arXiv:1201.2402 [astro-ph.IM].

[33] M. Pato, L. Baudis, G. Bertone, R. Ruiz de Austri, L. E. Strigari and R. Trotta, Phys. Rev. D **83**, 083505 (2011) [arXiv:1012.3458 [astro-ph.CO]].

[34] J. L. Newstead, T. D. Jacques, L. M. Krauss, J. B. Dent and F. Ferrer, Phys. Rev. D **88**, no. 7, 076011 (2013) [arXiv:1306.3244 [astro-ph.CO]].

[35] D. Tucker-Smith and N. Weiner, Phys.Rev.D **64** (2001) 0435002, hep-ph/0101138.

[36] S. Chang. G.D. Kribs, D. Tucker-Smith, and N. Weiner, Phys.Rev.D **79** (2009) 043513, arXiv:0807.2250 [hep-ph].

[37] B. Feldstein, A.L. Fitzpatrick, and E. Katz, JCAP **1001** (2010) 020, arXiv:0908.2991 [hep-ph].

[38] S. Change, A. Pierce, and N. Weiner, JCAP **1001** (2010) 006, arXiv:0908.3192 [hep-ph].

[39] V. Barger, W.-Y. Keung, and D. Marfatia, Phys.Lett.B **696** (2011) 74-78, arXiv:1007.4345 [hep-ph].

[40] T. Banks, J.-F. Fortin, and S. Thomas, RUNHETC-2010-19, SCIPP-10-14, UCSD-PTH-10-06, arXiv:1007.5515 [hep-ph].

[41] R. Foot, Phys.Rev.D **82** (2010) 095001, arXiv:1008.0685 [hep-ph].

[42] C.M. Ho and R.J. Scherrer, Phys.Lett.B **722** (2013) 341-346, arXiv:1211.0503 [hep-ph].

[43] C. Balázs, T. Li and J. L. Newstead, JHEP **1408**, 061 (2014) [arXiv:1403.5829 [hep-ph]].

[44] Y. Bai, P. J. Fox and R. Harnik, JHEP **1012**, 048 (2010) [arXiv:1005.3797 [hep-ph]].

[45] J. Goodman, M. Ibe, A. Rajaraman, W. Shepherd, T. M. P. Tait and H. B. Yu, Phys. Rev. D **82**, 116010 (2010) [arXiv:1008.1783 [hep-ph]].

[46] G. Busoni, A. De Simone, E. Morgante and A. Riotto, Phys. Lett. B **728**, 412 (2014) [arXiv:1307.2253 [hep-ph]].

[47] D. Alves *et al.* [LHC New Physics Working Group Collaboration], J. Phys. G **39**, 105005 (2012) [arXiv:1105.2838 [hep-ph]].

[48] A. Rajaraman, W. Shepherd, T. M. P. Tait and A. M. Wijangco, Phys. Rev. D **84**, 095013 (2011) [arXiv:1108.1196 [hep-ph]].

[49] O. Buchmueller, M. J. Dolan and C. McCabe, JHEP **1401**, 025 (2014) [arXiv:1308.6799 [hep-ph], arXiv:1308.6799].

[50] A. Berlin, D. Hooper and S. D. McDermott, Phys. Rev. D **89**, no. 11, 115022 (2014) [arXiv:1404.0022 [hep-ph]].

[51] C. Balázs and T. Li, Phys. Rev. D **90**, no. 5, 055026 (2014) [arXiv:1407.0174 [hep-ph]].

[52] J. Fan, M. Reece, and L.-T. Wang, JCAP **1011** (2011) 042, arXiv:1008.1591 [hep-ph].

[53] A.L. Fitzpatrick, W. Haxton, E. Katz, N. Lubbers, and Y. Xu, JCAP **1302** (2013) 004, arXiv:1203.3542 [hep-ph].

[54] N. Anand, A.L. Fitzpatrick, and W.C. Haxton, Phys.Rev.C **89** (2014) 065501, arXiv:1308.6288 [hep-ph].

[55] N. Anand, A.L. Fitpatrick, and W.C. Haxton, arXiv:1405.6690 [nucl-th].

[56] M. I. Gresham and K. M. Zurek, Phys. Rev. D **89**, 123521 (2014), arXiv:1401.3739 [hep-ph].

[57] J. Hisano, K. Ishiwata, N. Nagata and M. Yamanaka, Prog. Theor. Phys. **126**, 435 (2011) [arXiv:1012.5455 [hep-ph]].

[58] J. L. Feng and J. Kumar, Phys. Rev. Lett. **101**, 231301 (2008) [arXiv:0803.4196 [hep-ph]].

[59] R. Catena, arXiv:1406.0524 [astro-ph, physics:hep-ph].

[60] P. Agrawal, Z. Chacko, C. Kilic, and R. K. Mishra, arXiv:1003.1912 [hep-ph].

[61] K. R. Dienes, J. Kumar, B. Thomas and D. Yaylali, Phys. Rev. D **90**, no. 1, 015012 (2014) [arXiv:1312.7772 [hep-ph]].

[62] J. L. Feng, J. Kumar and D. Sanford, Phys. Rev. D **88**, no. 1, 015021 (2013) [arXiv:1306.2315 [hep-ph]].

Prepared for submission to JCAP

Large Slow Roll Parameters in Single Field Inflation

Jessica L. Cook,[a] **Lawrence M. Krauss**[a,b]

[a]Department of Physics and School of Earth and Space Exploration
Arizona State University, Tempe, AZ 85827-1404
[b]Research School of Astronomy and Astrophysics, Mt. Stromlo Observatory,
Australian National University, Canberra, Australia 2611

E-mail: jlcook14@asu.edu, krauss@asu.edu

Abstract.

We initially consider two simple situations where inflationary slow roll parameters are large and modes no longer freeze out shortly after exiting the horizon, treating both cases analytically. By modes, we refer to the comoving curvature perturbation R. We then consider applications to transient phases where the slow roll parameters can become large, especially in the context of the common 'fast-roll' inflation frequently used as a mechanism to explain the anomalously low scalar power at low l in the CMB. These transient cases we treat numerically. We find when ϵ, the first slow roll parameter, and only ϵ is large, modes decay outside the horizon, and when δ, the second slow roll parameter, is large, modes grow outside the horizon. When multiple slow roll parameters are large the behavior in general is more complicated, but we nevertheless show in the 'fast-roll' inflation case, modes grow outside the horizon.

Contents

1 Introduction

The simplest inflaton models assume that slow roll is valid during the whole period from when current CMB scales first exited the horizon up to the time just before inflation ends. There are reasons one might suppose this isn't the case. Obviously the slow roll parameters must become large as inflation is ending, but since the modes we do observe were far outside the horizon when inflation ended, we can presumably ignore this complication. We know the slow roll parameters couldn't have been large during the whole of inflation because either: (a) inflation wouldn't have lasted long enough, or (b) the observed n_s would look substantially different from what has been observed ($n_s \ll 1$ if ϵ is close to 1 during much of inflation, or $n_s \approx 1$ in the ultra flat $\delta \approx 3$ case. More on these below.) It would be phenomenologically interesting, however, if the slow roll parameters became large for transient periods during which scales that are currently observable today first left the horizon. For example, something along the lines of [1] where the potential is comprised of a small amplitude, high frequency trig function superimposed on a larger amplitude smaller frequency function would naturally lead to oscillations in the slow roll parameters. One can get oscillations with large amplitudes in the higher order slow roll parameters like δ and ξ^2 while ϵ remains small and inflation last long enough. In [1] this was used as a possible explanation for signs of negative running in SPT data [2], signs that haven't been replicated in Planck data [3].

A second possibility is that the slow roll parameters started out large at the time the largest CMB scales were exiting the horizon, which might be possible if these scales corresponded to the beginning of inflation. There might even be a hint for this found in the CMB in the anomalously low scalar power at low l. There are various reasons one might suppose inflation didn't last much longer than necessary to solve the horizon, flatness problem, etc. For example, one might suppose it's unnatural for a field to end up too far offset from its minimum. If inflation were preceded by a symmetry breaking event, one wouldn't assume the separation between the old and new minima to be arbitrarily large, especially if one assumes the breaking happened below the Planck scale. In another common picture where inflation started from a random quantum fluctuation, small quantum fluctuations

should be exponentially more common than large ones, and many small fluctuations stacking up in the same direction are also disfavored, leading to the inflaton likely not ending up much further offset from its minimum than necessary.

The significance of the low l power anomaly is around $2.5 - 3\sigma$ [4]. The anomaly was first discovered in the COBE data [5], and has since persisted in the WMAP and Planck data [3, 4, 6, 7]. It might just be an effect of cosmic variance, for example see [8]. Moreover, since the CMB comprises so many independent measurements, it would be surprising if none of those measurements ended up deviating in a significant way from the theoretical expectation [9]. Alternatively, it could be a hint of non-slow roll behavior at those scales, possibly providing insight into the beginning of inflation. If large tensors are observed in the future, this will increase the significance of the anomaly. The reason is that tensors as well as scalars contribute to $\langle TT \rangle$ at low l before around $l \approx 100$ after which tensors drop off substantially. Therefore, the fact that $\langle TT \rangle$ at low l is already smaller than expected based on the rest of the power spectrum, observing large tensors and inferring their contributions to $\langle TT \rangle$ would mean the scalar contribution would have to be even more suppressed at those l than previously thought.

For these reasons, it's interesting to see how large slow roll parameters affect the power spectrum. In particular, when slow roll no longer holds, modes no longer necessarily freeze out shortly after exiting the horizon [32–36]. By modes, we refer to the spatial curvature pertrubation R evaluated in comoving gauge, though one could just as well consider ζ, the gravitational potential evaluated in the gauge without density perturbations. This has a direct impact on the power spectrum which can be defined by: $P = \frac{k^3}{2\pi^2} \lim_{\frac{k}{aH} \to 0} |R_k|^2$. R (or ζ) typically asymptote to a constant shortly after a k mode leaves the horizon. This is the case when the perturbations are largely adiabatic, with an irrelevant entropy perturbation. However, there are situations were modes don't immediately freeze out, even for single field inflation. This can be understood as due to an entropy perturbation staying relevant compared to the adiabatic perturbation. We find that the gauge invariant metric perturbations can actually grow (or shrink) exponentially outside the horizon if they exit the horizon when the slow roll parameters are large. Therefore, for these modes, it is important to evaluate them after slow roll is reached and not at horizon crossing, or one will underestimate (or overestimate) the amplitude of the predicted power spectrum for these modes.

The papers [32, 33] gave a particularly useful way of understanding why modes grow in the $\delta \approx 3$ case. Note the equation of motion for the comoving curvature perturbation R:

$$R_{\tau\tau} + 2\frac{z_\tau}{z} R_\tau + k^2 R = 0 \tag{1.1}$$

where τ is conformal time and $z = \frac{a\phi_\tau}{\mathcal{H}}$. The $2\frac{z_\tau}{z} R_\tau$ acts as a friction term, and when it becomes negative, R grows on super horizon scales. $\frac{z_\tau}{z}$ can be written as $1 - \delta + \epsilon$, and note ϵ is always less than 1 if the universe is inflating, so if during inflation, δ is positive and larger than $1 + \epsilon$, then the friction term is negative.

[34–36] consider connections of superhorizon evolution to the low power at low l anomaly. [34] considers superhorizon evolution for DBI inflation, and [35, 36] considers potentials which lead to temporary breaks in inflation leading to dips followed by enhancement of the power spectrum. They show that if such a dip in the power spectrum aligned with the lowest l scales, this helps alleviate the anomaly.

We will need the slow roll parameters written as derivatives of the Hubble parameter since the common definition in terms of derivatives of only the potential assumes slow roll in

the derivation and also doesn't account for different choices of initial conditions. We use the standard definition for ϵ

$$\epsilon = 2M_P^2 \frac{H_\phi^2}{H^2} \tag{1.2}$$

(where subscript $\phi = \frac{d}{d\phi}$, H is the Hubble parameter, and M_P, the reduced Planck mass) and then follow the prescription in [10] for the higher slow roll parameters. These are defined such that the order n parameter is given by

$${}^n\beta = 2M_P^2 \left(\frac{H_\phi^{n-1} (\frac{d}{d\phi})^{n+1} H}{H^n} \right)^{\frac{1}{n}} \tag{1.3}$$

where $n = 1$ gives δ and $n = 2$ gives ξ and so on. While in principle there is an infinite number of slow roll parameters incorporating higher order derivatives of the Hubble parameter, only the first three, ϵ, δ and ξ^2, appear in the general Mukhanov-Sasaki (MS) equation, and are the only ones we will use here. The above prescription gives for δ and ξ^2:

$$\delta = 2M_P^2 \frac{H_{\phi\phi}}{H} \tag{1.4}$$

and

$$\xi^2 = 4M_P^4 \frac{H_\phi H_{\phi\phi\phi}}{H^2} \,. \tag{1.5}$$

It will be convenient to work with these equations with time derivatives instead, where the slow roll parameters can be rewritten:

$$\epsilon = \frac{\phi_N^2}{2M_P^2} \qquad \delta = \frac{\phi_{NN}}{\phi_N} + \frac{\phi_N^2}{2M_P^2} \qquad \xi^2 = \frac{3}{2M_P^2}\phi_{NN}\phi_N + \frac{1}{4M_P^4}\phi_N^4 + \frac{\phi_{NNN}}{\phi_N} - \frac{\phi_{NN}^2}{\phi_N^2} \tag{1.6}$$

in efolding time (N), or in conformal time (τ):

$$\epsilon = \frac{1}{2M_P^2}\frac{\phi_\tau^2}{\mathcal{H}^2} \qquad \delta = 1 - \frac{\phi_{\tau\tau}}{\mathcal{H}\phi_\tau} \qquad \xi^2 = -\frac{\phi_{\tau\tau}}{\mathcal{H}\phi_\tau} + \frac{1}{\mathcal{H}^2}\frac{\phi_{\tau\tau\tau}}{\phi_\tau} - \frac{1}{\mathcal{H}^2}\frac{\phi_{\tau\tau}^2}{\phi_\tau^2} + \frac{\phi_\tau^2}{2M_P^2\mathcal{H}^2} \,, \tag{1.7}$$

where $\mathcal{H} = \frac{a_\tau}{a}$. These can then be applied to the Mukhanov-Sasaki equation. The familiar form of the equation: $u_{\tau\tau} + u(k^2 - \frac{z_{\tau\tau}}{z}) = 0$ with $z = \frac{a\phi_\tau}{\mathcal{H}}$ is general and doesn't assume slow roll. It is at the stage of expanding out z_τ and $z_{\tau\tau}$ that one generally drops higher order slow roll terms, but one can easily leave them in to obtain:

$$u_{\tau\tau} + u\left(k^2 - \frac{\phi_{\tau\tau\tau}}{\phi_\tau} - 2\frac{\phi_{\tau\tau}\phi_\tau}{M_P^2\mathcal{H}} + \frac{\phi_\tau^2}{2M_P^2} - \frac{\phi_\tau^4}{2M_P^4\mathcal{H}^2}\right) = 0 \,. \tag{1.8}$$

The derivative of ϕ terms can then be traded for slow roll parameters:

$$u_{\tau\tau} + u(k^2 - \mathcal{H}^2(2 - 3\delta + 2\epsilon + \xi^2 + \delta^2 - 4\epsilon\delta + 2\epsilon^2)) = 0\,, \tag{1.9}$$

or equivalently in efolding time:

$$u_{NN} + u_N(\epsilon - 1) + u\left(\frac{k^2}{H^2}e^{2N} - 2 + 3\delta - 2\epsilon - \xi^2 - \delta^2 + 4\epsilon\delta - 2\epsilon^2\right) = 0\,. \tag{1.10}$$

In section 2 we consider a simple case where $\epsilon \approx 1$ and stays ≈ 1 for the duration of inflation, and show there is a delayed freeze out effect. In section 3 we explore a simple case of $\delta \approx 3$ for the duration of inflation, and determine how modes actually grow outside the horizon. This case has been considered before, [11–17] but is included here for completeness, and for relevance to transient situations where the slow roll parameters become large. In section 4 we consider numerical studies of modified freeze out with transient periods of large fast roll parameters. In section 5 we consider a simple analytic approximation of one of the transient phases that well approximates the final power spectrum. In section 6 we present our conclusions.

2 ϵ Large

Suppose ϵ is close to but slightly less than one such that the universe still inflates. Note that $\epsilon = \frac{\phi_N^2}{2M_P^2}$, using efolding time, N. Thus for $\epsilon \approx 1$, $\phi_N \approx \sqrt{2}M_P$. One can always start with initial conditions with ϵ this large, but generally this will lead to ϕ_N decreasing exponentially until a slow roll solution is reached. The quintessential example for ϵ large is to inflate on a potential such that the asymptotic solution has $\phi_N \approx \sqrt{2}M_P$.

The equation of motion of ϕ using efolding time is given by:

$$\phi_{NN} + (-3 + \epsilon)\phi_N + \frac{1}{H^2}\partial_\phi V = 0\,. \tag{2.1}$$

For an example where $\epsilon \approx 1$ is maintained for a long period, we need to maintain $\phi_N \approx \sqrt{2}M_P$, and this requires a solution with $\phi_{NN} \approx 0$. Since $\epsilon \approx 1$, the equation of motion becomes:

$$2H^2\phi_N = \partial_\phi V \tag{2.2}$$

Next we use the Friedmann equation in efolding time: $H^2 = \frac{V}{3M_P^2 - \frac{\phi_N^2}{2}}$ and plug this into the equation of motion to give:

$$V = \frac{M_P}{\sqrt{2}}\partial_\phi V\,. \tag{2.3}$$

So a potential that maintains $\epsilon \approx 1$ requires $V \propto \partial_\phi V$. This is obviously satisfied by exponentials, so suppose

$$V = \Lambda^4 e^{\frac{\phi}{f}} \,. \tag{2.4}$$

Plugging this into $V = \frac{M_P}{\sqrt{2}} \partial_\phi V$, we find

$$f = \frac{M_P}{\sqrt{2}} \,. \tag{2.5}$$

Thus if $V = \Lambda^4 e^{\frac{\phi}{f}}$ with $f = \frac{M_P}{\sqrt{2}}$, this allows for inflation with $\epsilon \approx 1$. In order that ϵ is actually a little less than 1, f should be slightly larger that $\frac{M_P}{\sqrt{2}}$.

We consider the other slow roll parameters for this potential (equation 1.6). We find $\frac{\phi_{NN}}{\phi_N} \approx 0$, but $\delta \approx 1$ since $\delta = \frac{\phi_{NN}}{\phi_N} + \epsilon$. Similarly we find $\frac{\phi_{NNN}}{\phi_N} \approx 0$, but $\xi^2 \approx 1$ since $\xi^2 = \frac{\phi_{NNN}}{\phi_N} + 5\epsilon\delta - 3\epsilon^2 - \delta^2$. Note that it is the large ϵ value that makes δ and $\xi^2 \approx 1$. We will use these to solve the full Mukhanov-Sasaki equation without slow roll approximations (equation 1.9). Let $\epsilon = 1 - \alpha$, so α is positive and $\alpha \ll 1$. The MS equation becomes:

$$u_{\tau\tau} + u\left(k^2 - \mathcal{H}^2(1+\alpha)\right) = 0 \tag{2.6}$$

Since $\phi_{NN} \approx 0$, $\epsilon \approx$ constant, then $\alpha \approx$ constant. We plug in for $\mathcal{H}$ using the definition for ϵ in conformal time: $\frac{\mathcal{H}_\tau}{\mathcal{H}^2} = 1 - \epsilon$:

$$\int \frac{1}{\mathcal{H}^2} d\mathcal{H} = \int \alpha d\tau \,. \tag{2.7}$$

After choosing to define τ such that the integration constant is 0, this integrates to:

$$\mathcal{H} = \frac{1}{-\alpha\tau} \,. \tag{2.8}$$

Then the MS equation becomes:

$$u_{\tau\tau} + u\left(k^2 - \frac{1}{\alpha^2\tau^2}(1+\alpha)\right) = 0 \,, \tag{2.9}$$

which can be solved:

$$u = c_1\sqrt{-\tau}H^{(1)}_{\frac{1}{\alpha}+\frac{1}{2}}(-k\tau) + c_2\sqrt{-\tau}H^{(2)}_{\frac{1}{\alpha}+\frac{1}{2}}(-k\tau) \,. \tag{2.10}$$

We recognize this is the standard slow roll result except with $\nu \to \frac{1}{\alpha} + \frac{1}{2}$, and so matching onto Bunch-Davies (BD) initial conditions gives:

$$u = \frac{\sqrt{\pi}}{2}\sqrt{-\tau}H^{(1)}_{\frac{1}{\alpha}+\frac{1}{2}}(-k\tau) \,. \tag{2.11}$$

This can be rewritten using $\tau = \frac{-1}{\alpha a H}$:

$$u = \frac{\sqrt{\pi}}{2}\sqrt{\frac{1}{\alpha a H}} H^{(1)}_{\frac{1}{\alpha}+\frac{1}{2}}\left(\frac{k}{\alpha a H}\right) . \tag{2.12}$$

Normally during slow roll $\nu \approx \frac{3}{2}$, and there are only small deviations from $\frac{3}{2}$ which produce a tilt in the power spectrum. Here we have $\nu \approx \frac{1}{\alpha}$ where $\alpha \ll 1$, which will create an extremely red power spectrum. The closer ϵ gets to one, the more tilted the power spectrum becomes.

Also note that, as always, modes exit the horizon when $k = aH$ but now the long wavelength limit of the mode function only becomes valid when $k \ll \alpha a H$, and since $\alpha \ll 1$, this occurs well after a mode has exited the horizon. This means the amplitude for modes will continue to decay for a long period after they have exited the horizon, but freeze out will eventually occur when $k \ll \alpha a H$. The closer ϵ is to one, the longer it takes for freeze-out to eventually happen.

We next calculate the power spectrum approximated for modes with $k \ll \alpha a H$. Using the gauge invariant perturbation $|R|^2 = \frac{|u|^2}{a^2 \phi_N^2}$ and $P_\zeta = \frac{k^3}{2\pi^2} \lim_{\frac{k}{\alpha a H} \to 0} |R|^2$:

$$P_\zeta = \frac{2^{2\nu}\Gamma^2(\nu)}{8\pi^3 \phi_N^2}(\alpha H)^2 \left(\frac{\alpha a H}{k}\right)^{\frac{2}{\alpha}-2} . \tag{2.13}$$

We find $P_\zeta \propto (\frac{1}{k})^{\text{large number}}$ giving an extremely red power spectrum.

To show that the power spectrum does eventually freeze out, we plug H and ϕ_N into P_ζ to better see the N dependence. Using the equation of motion above:

$$\phi_N = \frac{M_P^2}{f} . \tag{2.14}$$

Let N_0 define the initial value for N with N counting down towards 0 at the end of inflation. Then integrating the above gives:

$$\phi = \frac{M_P^2}{f}(N - N_0) + \phi_0 . \tag{2.15}$$

We use the Friedmann equation to obtain:

$$H = \frac{\Lambda^2}{\sqrt{3M_P^2 - \frac{1}{2}\frac{M_P^4}{f^2}}} e^{\frac{1}{2f}(\phi_0 + \frac{M_P^2}{f}(N-N_0))} . \tag{2.16}$$

We will focus only on the time dependent part:

$$P_\zeta \propto e^{\frac{M_P^2}{f^2}N}\left(e^{N\left(\frac{M_P^2}{2f^2}-1\right)}\right)^{\frac{2}{\alpha}-2} . \tag{2.17}$$

Next, we determine f as a function of α. We use $\epsilon = \frac{\phi_N^2}{2M_P^2}$ and $\phi_N = \frac{M_P^2}{f}$ which gives

$$f^2 = \frac{M_P^2}{2(1-\alpha)}\,. \tag{2.18}$$

This yields the power spectrum:

$$P_\zeta \propto e^{2N(1-\alpha)}\left(e^{N((1-\alpha)-1)}\right)^{\frac{2}{\alpha}-2}\,. \tag{2.19}$$

We find the N dependence cancels out. So, as expected, the modes do freeze out, but not until $k \ll \alpha a H$.

One can then determine how much longer it takes for modes to freeze out after they exit the horizon, or more specifically how many efolds pass between the period when $k = aH$ and when $k = \alpha a H$.

$$e^{-N_1}\frac{\Lambda^2}{\sqrt{3M_P^2 - \frac{1}{2}\frac{M_P^4}{f^2}}}e^{\frac{1}{2f}\left(\phi_0 + \frac{M_P^2}{f}(N_1 - N_0)\right)} = \alpha e^{-N_2}\frac{\Lambda^2}{\sqrt{3M_P^2 - \frac{1}{2}\frac{M_P^4}{f^2}}}e^{\frac{1}{2f}\left(\phi_0 + \frac{M_P^2}{f}(N_2 - N_0)\right)} \tag{2.20}$$

Using $f^2 = \frac{M_P^2}{2(1-\alpha)}$ we find:

$$\Delta N = -\frac{\ln\alpha}{\alpha}\,. \tag{2.21}$$

So for example, for $\epsilon = 0.95$, this corresponds to $\Delta N = 60$. The closer ϵ is to one, the longer the freeze out time. This means when inflation ends, there will be modes which haven't reached the long wavelength limit, and these modes will have more power than they would otherwise have.

Of course this model produces an entirely wrong n_s and can't describe our actual power spectrum. As we shall explore, more realistic models could contain transient periods where ϵ becomes large, which is almost certainly true at least at the end of inflation but possibly elsewhere too, and this model can give useful insights for those cases.

We find this behavior numerically in the arctan potential example below which passes through a period with $\epsilon \approx 1$. There, as we shall see, the behavior is more complicated because as ϵ returns to the slow roll value, some modes actually grow outside the horizon rather than decay, leading to oscillations in the power spectrum. Solving analytically in these scenarios is very difficult because not only are the slow roll parameters large, their derivatives are large, and one can no longer approximate them as constant in the MS equation.

3 ϵ Small but η Large

When the V_ϕ term in the equation of motion during inflation is subdominant, one enters a regime where $\delta \approx 3$ until enough kinetic energy is redshifted that V_ϕ starts to balance the other terms. Then δ will transition to $\delta \ll 1$ and slow roll will be reached. If one also starts

with $\epsilon \approx 1$, and the potential isn't as steep as in the last example, the kinetic energy will drop off, and there will frequently be an in-between regime, where ϵ starts to drop off, but $\delta \approx 3$ until enough kinetic energy is lost that δ too becomes small and slow roll is reached. This case will be seen in the ϕ^2 and R^2 examples in section 4. The flatter the potential, the longer the $\delta \approx 3$ regime. We start from the equation of motion in efolding time:

$$\phi_{NN} + (-3+\epsilon)\phi_N + \frac{1}{H^2}\partial_\phi V = 0\,. \tag{3.1}$$

If ϵ has already become small, or was never large to begin with, but the field has enough kinetic energy such that $\partial_\phi V$ is much smaller than the other terms, we find:

$$\phi_{NN} = 3\phi_N\,. \tag{3.2}$$

Note from equation 1.6 $\delta = \frac{\phi_{NN}}{\phi_N} + \epsilon$. After ϵ has become small, then $\delta = \frac{\phi_{NN}}{\phi_N}$ so that $\delta = 3$.

The flatter the potential, the longer the V_ϕ term will stay irrelevant and $\delta = 3$. So the quintessential example of trying to keep δ large would be a perfectly flat potential, $V_\phi = 0$.

This has been worked out in the past [11–17]. It has been found that freeze out never occurs for the duration of inflation and modes actually grow exponentially outside the horizon. In spite of this, the final power spectrum is perfectly flat, without features.

One can derive this by solving the background/ classical equations, including first the Friedmann equation:

$$H = \frac{\sqrt{V}}{\sqrt{3M_P^2 - \frac{\phi_N^2}{2}}}\,. \tag{3.3}$$

In the regime where ϵ is small, $\frac{\phi_N^2}{2} \ll 3M_P^2$, and with a potential where V is constant:

$$H \approx \sqrt{\frac{V_0}{3M_P^2}}\,, \tag{3.4}$$

the Hubble parameter is also essentially constant. Then we solve the the equation of motion for ϕ, equation 3.2, which gives:

$$\phi = \phi_0 + \frac{\phi_{N\,0}}{3}\left(e^{3(N-N_0)} - 1\right)\,. \tag{3.5}$$

Next we solve the full Mukhanov-Sasaki equation, equation 1.9, where we have both ϵ, $\xi^2 \approx 0$, and $\delta = 3$ is constant:

$$u_{\tau\tau} + u(k^2 - 2\mathcal{H}^2) = 0\,. \tag{3.6}$$

We choose initial conditions for τ such that $\mathcal{H} = -\frac{1}{\tau}$:

$$u_{\tau\tau} + u(k^2 - \frac{2}{\tau^2}) = 0\,. \tag{3.7}$$

This gives the normal slow roll mode function equation which has solution:

$$u = \frac{\sqrt{\pi}}{2}\sqrt{-\tau}H^{(1)}_{\frac{3}{2}}(-k\tau)\,. \tag{3.8}$$

So a potential with $\delta = 3$ actually generates the same mode functions as a power spectrum with $\delta = 0$, assuming ϵ and $\xi^2 \approx 0$. However, the evolution of the metric perturbations outside the horizon is very different. We use the gauge invariant metric perturbation $|R|^2 = \frac{|u|^2}{a^2\phi_N^2}$ and $P_\zeta = \frac{k^3}{2\pi^2}\lim_{\frac{k}{aH}\to 0}|R|^2$. The difference comes from the behavior of the ϕ_N which is approximately constant during slow roll, but is decaying exponentially when $\delta = 3$. This means the behavior of $|R|^2$ for a particular k mode grows exponentially outside the horizon since $|R|^2 \propto \frac{1}{\phi_N^2}$. The power spectrum is:

$$P_\zeta = \frac{2^{2\nu}\Gamma^2(\nu)}{8\pi^3}\frac{H^2}{\phi_N^2}\,, \tag{3.9}$$

which is the form of the slow roll power spectrum except all the k dependence of the power spectrum drops out.

To explore the time dependence, we insert values of H and ϕ_N. H is approximately constant, but $\phi_N = \phi_{N\,0}e^{3(N-N_0)}$.

$$P_\zeta = \frac{2^{2\nu}\Gamma^2(\nu)}{8\pi^3}\frac{H_0^2}{\phi_{N\,0}^2 e^{6(N-N_0)}} \tag{3.10}$$

Thus $P_\zeta \propto e^{-6N}$. Since we are using conventions where N counts down towards the end of inflation, this means the power spectrum is growing exponentially in time. Note that typically inflation doesn't end in this scenario so that one imagines some coupling with another field forces inflation to end at some particular point. Then the power spectrum should freeze out at the transition to reheating (because super-horizon modes are frozen during matter and radiation dominance). The amplitude for the power spectrum should be the amplitude of these modes evaluated at the point inflation ends. Importantly, since there is such a strong time dependence in the power spectrum, changing the time inflation ends in this model by a small amount has a large effect on the amplitude of the power spectrum.

This model, as with the large ϵ one, would give a totally wrong n_s (in this case n_s of exactly one), so this model can't describe our universe. Again, however, there could have been periods, especially early on, or following a sharp transition from a steeper to a flatter potential, when $\delta \approx 3$ for some time. This situation is seen in two of the numerical examples below which do produce a growth of modes outside the horizon until slow roll is reached.

In this case the power spectrum had to come out scale independent because there was essentially no scale to the potential. In more realistic models, scale dependence will appear both in the δ large, and slow roll regimes. The fact that modes grow outside the horizon is independent of this and is a generic feature of large δ evolution.

4 Transient Periods of Fast Roll

It's been pointed out [18] that in situations where not only the slow roll parameters are large, but also their derivatives are large, it becomes extremely difficult to find analytic solutions to the Mukhanov-Sasaki equation. We found an analytic solution for the case of a ϕ^2 potential as inflation is ending, but since the slow roll parameters are only large for about 1/2 an efolding before inflation ends, this case isn't particularly enlightening. Instead we consider numerically cases where the slow roll parameters become large earlier on during inflation, at observable scales, and then become small again.

Situations where one starts the inflaton with extra kinetic energy have been considered in [9, 19–25]. Others have considered possibilities where the shape of the potential changes such that it was steeper initially and then gets flatter to give suppression at low l, but without breaking slow roll throughout the observable part of the CMB spectrum, often considered in the context of tunneling [26–31]. The latter cases where one considers slow roll on a steeper potential and then slow roll on a more shallow potential are much easier to implement since slow roll approximations hold throughout,[1] and because since inflation continues for some time, the Bunch-Davies initial conditions are valid. Both situations tend to give suppressed power at low l.

The question of BD initial conditions becomes an issue if one assumes inflation lasts 'just long enough' such that the largest CMB scales represent the start of inflation, and some non-inflation period came before. We first consider numerically an arctan potential, this way there is a period of slow roll inflation where BD initial conditions are valid, and then a smooth transition to a period where the slow roll parameters become large, followed by another slow roll regime. Next we consider the familiar ϕ^2 potential, but start with initial conditions such that ϕ starts with the maximum kinetic energy such that it's still inflating, 1/2 the potential. In this case, assuming a non-inflating period came before, BD initial conditions are at best a rough approximation which allows us to obtain a power spectrum that can be considered a lower bound. This is because the BD initial conditions imply that the amplitude of the different modes has been reduced down to it's minimum allowed quantum value. Therefore, initial conditions set up by a non-inflating regime will likely give modes a larger initial amplitude. For example, consider [23, 32] where they find that modes which are close to horizon size during a break in inflation are typically associated with enhanced power.

4.1 Arctan Potential

For the arctan example we take:

$$V = \Lambda^4 \left(\arctan\left(-\frac{\phi}{\mu} \right) + C \right) \tag{4.1}$$

where the $+C$ is added to make sure there is positive potential slow roll behavior after the steepening. One can choose to make the potential as steep as one wishes when $\phi = 0$. We choose to keep $\epsilon < 1$ so inflation never actually stops. Figure 1 shows the potential and the evolution of the slow roll parameters around the transition. N stands for number of efolding where we fix $N = 0$ to be the time where ϵ is maximized, and N counts down such that

[1] Except perhaps for a brief period immediately following the transition if the transition isn't smooth.

larger N means earlier time. Figure 2 A shows the ratio of the power spectrum evaluated during the final slow roll regime to the power spectrum evaluated at horizon crossing in order to show the amount that modes either grow or decay outside the horizon during the fast roll regime. In plotting the power spectrum as a function of N, we are plotting different k modes, with N designating the horizon crossing time for each k mode. The figure shows the same information as the more traditional power spectrum plots as a function of k, but with a redefined x-axis. We show the latter in Figure 3 for reference, where the normalization of k is arbitrary in this case, since this isn't meant to represent the actual physical power spectrum. Plotting the power spectrum as a function of N is meant to visually show how long the effects last. Note in the early and late times in the figure, standard slow roll is taking place, and modes decay by about a factor of 2 outside the horizon. The vertical red lines bracket the region $\epsilon \geq 0.25$, the green the region $\delta \geq 0.25$, and the blue the region $\xi^2 \geq 0.25$. We find initially when ϵ is large, the modes decay outside the horizon, but as δ gets large the modes grow. There are also the oscillations/ wiggles one typically finds in the power spectrum following sharp transitions.

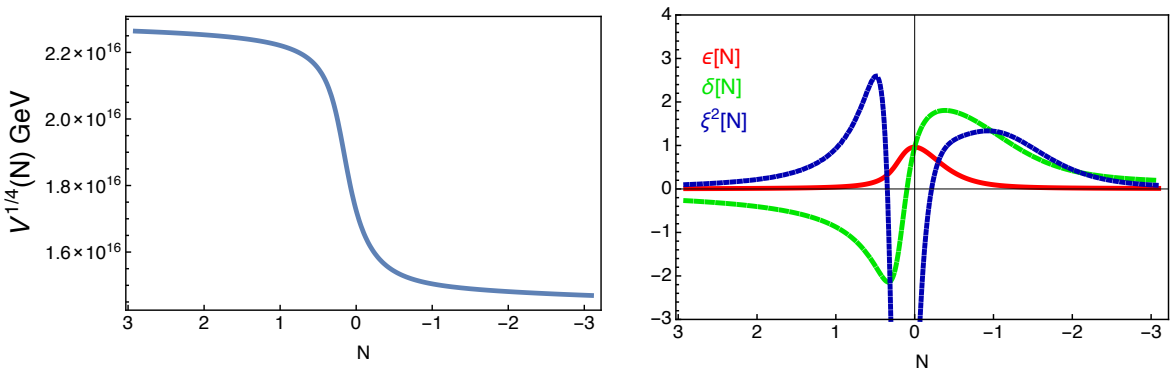

Figure 1: The figure depicts the potential and the three slow roll parameters, which appear in the Mukhanov-Sasaki equation for the arctan potential, plotted in the region where the potential becomes steep. $N = 0$ is taken to be where ϵ is maximized.

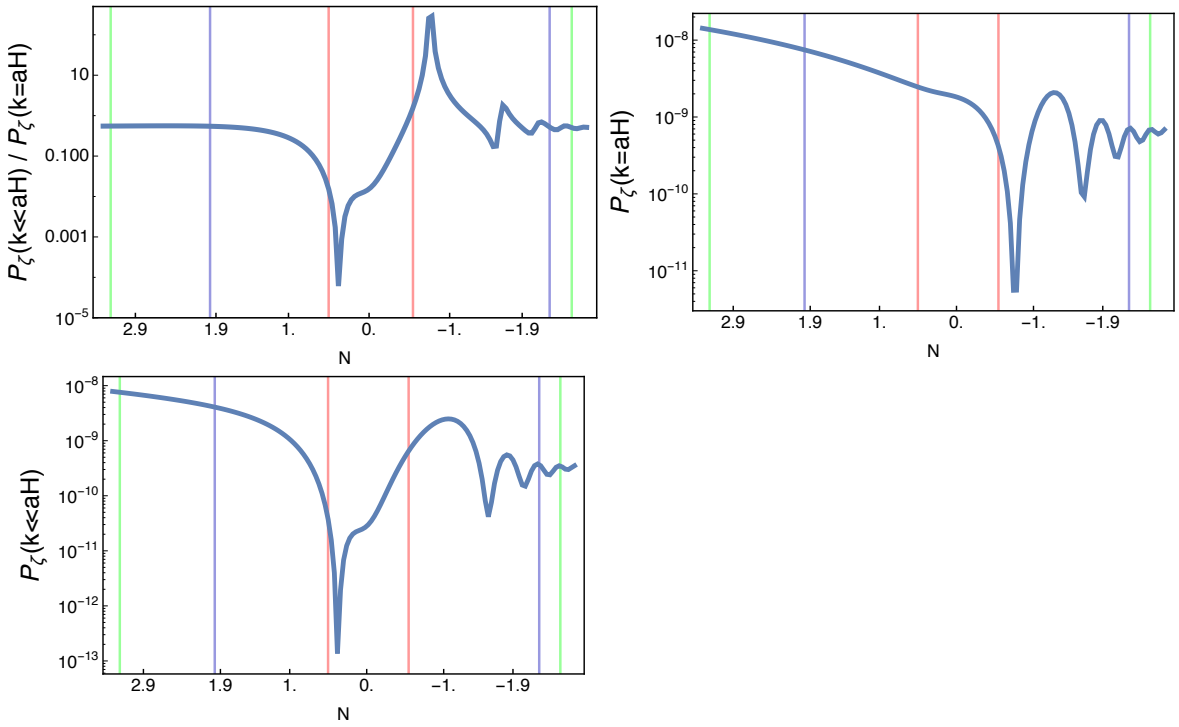

Figure 2: The top left figure depicts the ratio of the final power spectrum amplitude evaluated during the last slow roll regime to the power spectrum amplitude if evaluated at horizon crossing for the arctan potential. This shows how much modes either grow or decay while outside the horizon during the fast roll phase. The top right figure depicts the power spectrum amplitude evaluated at horizon crossing. The bottom left figure depicts the power spectrum amplitude evaluated during the final slow roll regime. $N = 0$ is taken to be where ϵ is maximized. The red vertical lines bracket the region $\epsilon \geq 0.25$, the green the region $\delta \geq 0.25$, and the blue the region $\xi^2 \geq 0.25$.

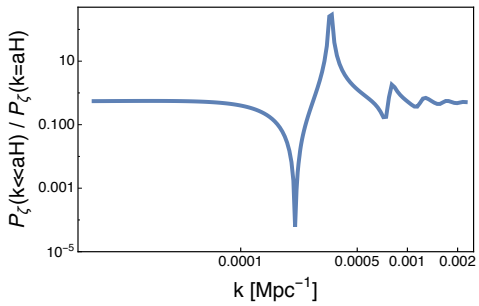

Figure 3: This figure shows the same results as the last figure but plotting in the more traditional way as a function of k.

Just tracing whether modes are growing or shrinking outside the horizon isn't enough to reproduce the full power spectrum. The amplitude of the modes as they reach horizon size is also changing in time, as depicted in Figure 2 B. The modes that start with the smallest amplitude at horizon crossing subsequently grow the most outside the horizon. Incorporating

both of these factors yields the power spectrum evaluated at late times, well into slow roll regime in Figure 2 C.

4.2 ϕ^2 Potential with Fast Roll Initial Conditions

In this example we take a standard ϕ^2 potential but start the field off with extra kinetic energy such that there is a smooth transition from kinetic dominance, with energy loss due to Hubble friction, transitioning to fast roll inflation as ϵ drops to 1, and then eventually slow roll inflation as the kinetic energy continues to diminish and the attractor solution is reached. In this way the background equations are well defined. The difficulty comes in choosing initial conditions for the metric perturbations for the modes that are leaving the horizon shortly after inflation starts. Here as we have described, BD initial conditions are not necessarily valid. Nevertheless, approximating the power spectrum using BD initial conditions should yield an underestimate for the final amplitude. Thus the plots we display should be considered as a minimum for the power spectrum from this fast roll period.

In Figure 4 we show the evolution of the three slow roll parameters appearing in the Mukhanov-Sasaki equation in the fast roll regime, where we define $N = 60$ as the start of inflation when $\epsilon = 1$, and define N so it's counting down towards the end of inflation. In Figure 5 A we show the ratio of the power spectrum evaluated in the late time limit, after slow roll has been established, to the power spectrum evaluated at horizon crossing, to show the evolution of the modes outside the horizon. Again, instead of displaying k on the x-axis, we show N as the time each k mode reached horizon size, to visually show in efolds how long the effects of the fast roll initial conditions last. We show the more traditional power spectrum in Figure 6 for reference. In the slow roll limit, the familiar factor of 2 drop in the power spectrum after a mode leaves the horizon is restored, but initially we find modes grow outside the horizon as for the δ large example above. In Figure 5 B we show the power spectrum at horizon crossing and in Figure 5 C, the power spectrum in the late time limit for reference. We find that although modes initially grow outside the horizon, they start from such a suppressed initial value that the final power spectrum for these initial modes is still smaller than for modes that exit the horizon during slow roll.

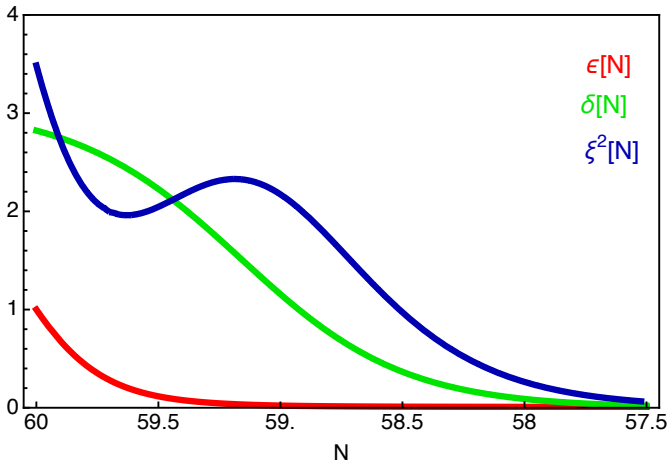

Figure 4: The figure depicts the three slow roll parameters which appear in the Mukhanov-Sasaki equation for the ϕ^2 potential, plotted in the region where the potential is steep. $N = 60$ corresponds to the onset of inflation with N counting down.

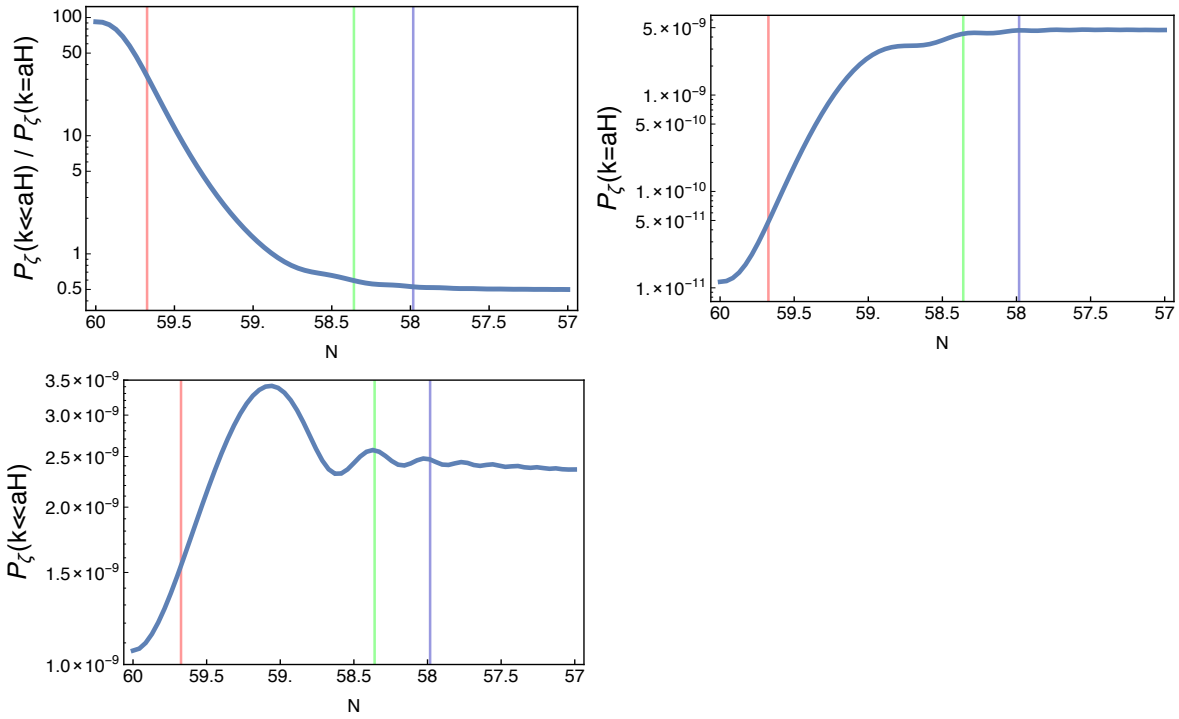

Figure 5: The top left figure depicts the ratio of the final power spectrum amplitude evaluated during the following slow roll regime to the power spectrum amplitude if evaluated at horizon crossing for the ϕ^2 potential. This shows how much modes either grow or decay while outside the horizon during the fast roll phase. The top right figure depicts the power spectrum amplitude evaluated at horizon crossing. The bottom left figure depicts the final power spectrum evaluated during the slow roll regime. The red vertical line marks the region $\epsilon \geq 0.25$, the green the region $\delta \geq 0.25$, and the blue the region $\xi^2 \geq 0.25$.

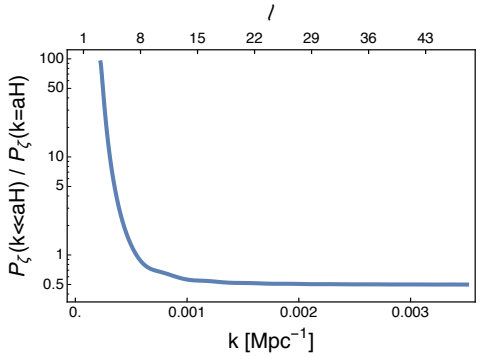

Figure 6: This figure shows the same results as the last figure but plotting in the more traditional way as a function of k. Note the l axis is approximate as P_ζ for each l is really an integral over k. Also we assume the first modes to freeze out with the onset of inflation correspond to the largest observable scales.

4.3 R^2 Potential with Fast Roll Initial Conditions

In this final numeric example we explore the effect of fast roll initial conditions for a flatter potential, the R^2 potential. We again give the field extra kinetic energy initially, and define $N = 60$ as the start of inflation when $\epsilon = 1$. Figure 7 shows the evolution of the slow roll parameters for this case right around the fast roll transition. We find the flatter potential means that the $\delta \approx 3$ region lasts for a longer period of time. As mentioned above, this is because it takes longer for the V_ϕ term in the equation of motion to become relevant. Figure 8 shows the resultant power spectrum, where Figure 8 A shows the ratio of the late time to horizon crossing power spectra, displaying how much modes evolve outside the horizon in this case. Again, we find growth outside the horizon for a period until slow roll is reached.

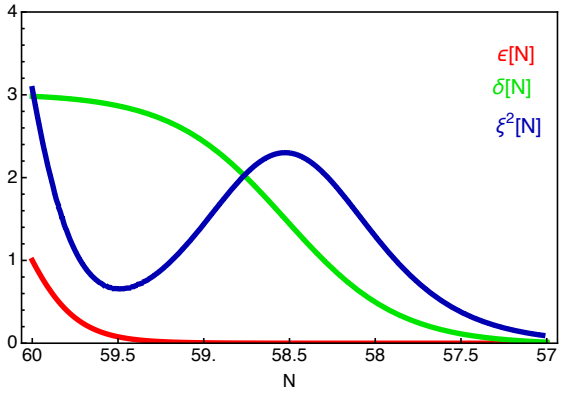

Figure 7: The figure depicts the three slow roll parameters which appear in the Mukhanov-Sasaki equation for the R^2 potential, plotted in the region where the potential is steep. $N = 60$ corresponds to the onset of inflation with N counting down.

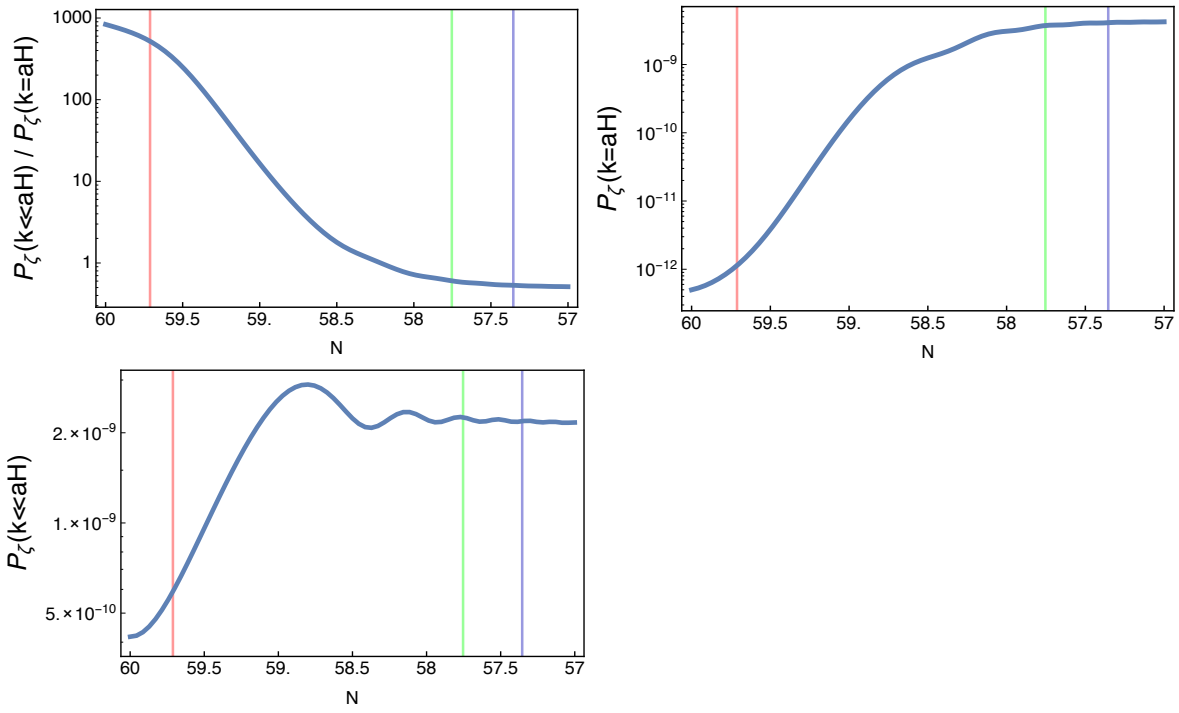

Figure 8: The top left figure depicts the ratio of the final power spectrum amplitude evaluated during the following slow roll regime to the power spectrum amplitude if evaluated at horizon crossing for the R^2 potential. This shows how much modes either grow or decay while outside the horizon during the fast roll phase. The top right figure depicts the power spectrum amplitude evaluated at horizon crossing. The bottom left figure depicts the final power spectrum amplitude evaluated during the slow roll regime. The red vertical line marks the region $\epsilon \geq 0.25$, the green the region $\delta \geq 0.25$, and the blue the region $\xi^2 \geq 0.25$.

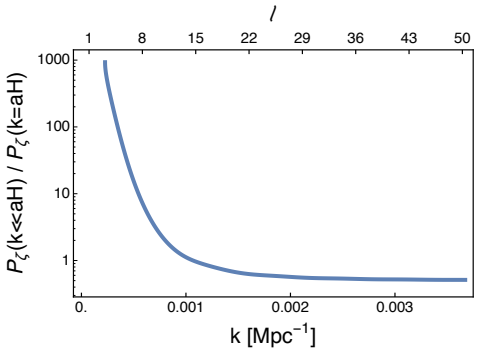

Figure 9: This figure shows the same results as the last figure but plotting in the more traditional way as a function of k. Note the l axis is approximate as P_ζ for each l is really an integral over k. Also we assume the first modes to freeze out with the onset of inflation correspond to the largest observable scales.

5 Analytic Approximation

Others [9, 23] have considered an analytic approximation where one takes a de Sitter solution in one region, a kinetic dominated solution in another range, and then just matches boundary conditions for $\mathcal{H}$, τ, u, and u_τ between the regions where u is the MS mode function. As we show below, this method approximates well the late time suppression of the power spectrum, but doesn't reproduce the behavior of the modes outside the horizon.

First during the kinetic dominated period, one takes the equation of motion and Friedmann equations without the potential terms:

$$\phi_{\tau\tau} + 2\mathcal{H}\phi_\tau = 0 \tag{5.1}$$

and

$$\mathcal{H}^2 = \frac{\phi_\tau^2}{6M_P^2}\,. \tag{5.2}$$

During kinetic dominance, $\epsilon \approx \delta \approx 3$ and $\xi^2 \approx 9$ and all three slow roll parameter's time derivatives are negligible. The general MS equation (eq. 1.9) then reduces to:

$$u_{\tau\tau} + u(k^2 + \frac{1}{4\tau^2}) = 0\,. \tag{5.3}$$

This gives the standard slow roll solution but now with $\nu = 0$. If we wish to match onto BD initial conditions, this gives:

$$u_{KD} = \frac{\sqrt{\pi}}{2}\sqrt{-\tau}H_0^{(1)}(-k\tau)\,. \tag{5.4}$$

Next we take the traditional slow roll equations, making sure $\mathcal{H}$, u and u_τ are continuous across the transition. Figure 10 shows the results where we used slow roll parameters from the ϕ^2 potential.

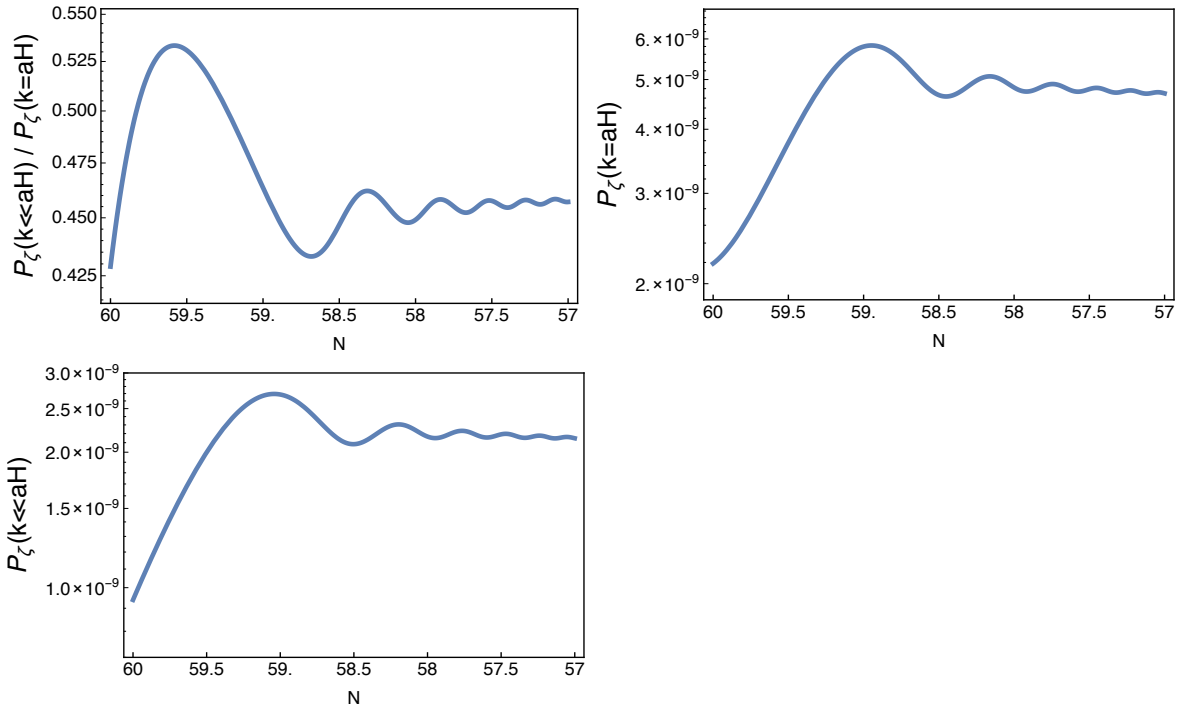

Figure 10: This figure shows power spectra results for a kinetic dominated phase immediately transitioning to slow roll using the ϕ^2 potential. The top left figure shows the ratio of the final amplitude to the horizon crossing amplitude. The top right figure shows the horizon crossing amplitude, and the bottom figure shows the late time amplitude.

This simple analytic approach does well approximate the final power spectrum amplitude when evaluated at late times, almost exactly replicating the power spectrum for the ϕ^2 potential as can be seen by comparing Figure 10 to Figure 5, but the approach doesn't well characterize the evolution of the modes as they are passing outside the horizon.

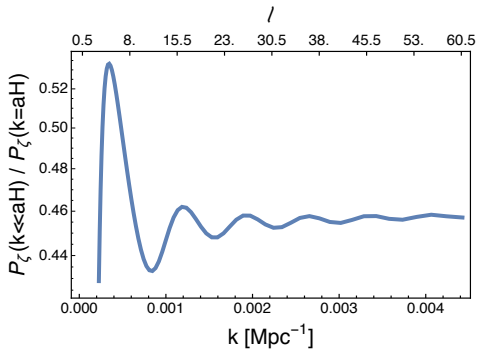

Figure 11: This figure shows the same results as the last figure but plotting in the more trandiational way as a function of k. Note the l axis is approximate as P_ζ for each l is really an integral over k. Also we assume the first modes to freeze out with the onset of inflation correspond to the largest observable scales.

6 Conclusions

We investigate the growth of the comoving metric perturbations outside the horizon in two simple test cases where the slow roll parameters ϵ or δ become large. In the first case where ϵ becomes large, freeze out eventually occurs, but takes longer the closer ϵ is to 1. During the whole period that modes are evolving outside the horizon, they decay exponentially in efolding time. In the second case, when $\delta \approx 3$, freeze out doesn't occur until inflation ends, and the modes instead grow exponentially in efolding time outside the horizon. Neither case can provide the correct n_s and so by itself can't describe what actually occurred during inflation. However, it is possible that during inflation there could be transient periods where the slow roll parameters become temporarily large. In these cases one has to be careful to evaluate modes after slow roll has been reached rather then at horizon crossing, as we have shown.

It is also worth noting that there could even be hints of such transient behavior in the CMB data, given the anomalously low power at low l in the scalar power spectrum. We show that if one uses BD initial conditions and starts inflation with maximum kinetic energy (1/2 the potential), then one obtains a power spectrum that is suppressed at horizon crossing. The modes grow outside the horizon, but the final power spectrum is still suppressed for those initial modes, with the effect lasting for a few efolds. The growth in the modes outside the horizon lasts until δ returns to its slow roll value, which takes longer for flatter potentials.

Such cases, where the initial suppression at low l can be linked to the earliest stages of inflation, are therefore phenomenologically interesting, but there is an open question as to what to use for initial conditions, which is therefore an important issue when trying to compare to actual data. We used BD initial conditions in this case for convenience, but as we have noted, this should only be viewed as a lower bound.

For the one transient case where initial conditions are unambiguous (an arctan potential), where there is an initial and final slow roll regime, and the slow roll parameters get large in-between, we find suppression when ϵ is large, followed by oscillations, typical when there are sudden transitions in power spectra.

Acknowledgments

It is a pleasure to thank Will Kinney for useful correspondence. This research was supported by a grant from the DOE.

References

[1] F. Takahashi, *The Spectral Index and its Running in Axionic Curvaton*, *JCAP* **1306** (2013) 013, [arXiv:1301.2834], [doi:10.1088/1475-7516/2013/06/013].

[2] Z. Hou, C. Reichardt, K. Story, B. Follin, R. Keisler, *et al.*, *Constraints on Cosmology from the Cosmic Microwave Background Power Spectrum of the 2500 deg^2 SPT-SZ Survey*, *Astrophys.J.* **782** (2014) 74, [arXiv:1212.6267], [doi:10.1088/0004-637X/782/2/74].

[3] **Planck** Collaboration, P. Ade *et al.*, *Planck 2015 results. XX. Constraints on inflation*, arXiv:1502.02114.

[4] **Planck** Collaboration, P. Ade *et al.*, *Planck 2013 results. XV. CMB power spectra and likelihood*, *Astron.Astrophys.* **571** (2014) A15, [arXiv:1303.5075], [doi:10.1051/0004-6361/201321573].

[5] C. Bennett, A. Banday, K. Gorski, G. Hinshaw, P. Jackson, *et al.*, *Four year COBE DMR cosmic microwave background observations: Maps and basic results*, *Astrophys.J.* **464** (1996) L1–L4, [arXiv:astro-ph/9601067], [doi:10.1086/310075].

[6] **WMAP** Collaboration, C. Bennett *et al.*, *First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results*, *Astrophys.J.Suppl.* **148** (2003) 1–27, [arXiv:astro-ph/0302207], [doi:10.1086/377253].

[7] **WMAP** Collaboration, D. Spergel *et al.*, *First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters*, *Astrophys.J.Suppl.* **148** (2003) 175–194, [arXiv:astro-ph/0302209], [doi:10.1086/377226].

[8] M. J. White, L. M. Krauss, and J. Silk, *Inflation and the statistics of cosmic microwave background anisotropies: From 1-degree to COBE*, *Astrophys. J.* **418** (1993) 535, [arXiv:astro-ph/9303009], [doi:10.1086/173415].

[9] C. R. Contaldi, M. Peloso, L. Kofman, and A. D. Linde, *Suppressing the lower multipoles in the CMB anisotropies*, *JCAP* **0307** (2003) 002, [arXiv:astro-ph/0303636], [doi:10.1088/1475-7516/2003/07/002].

[10] A. R. Liddle, P. Parsons, and J. D. Barrow, *Formalizing the slow roll approximation in inflation*, *Phys.Rev.* **D50** (1994) 7222–7232, [arXiv:astro-ph/9408015], [doi:10.1103/PhysRevD.50.7222].

[11] N. Tsamis and R. P. Woodard, *Improved estimates of cosmological perturbations*, *Phys.Rev.* **D69** (2004) 084005, [arXiv:astro-ph/0307463], [doi:10.1103/PhysRevD.69.084005].

[12] W. H. Kinney, *Horizon crossing and inflation with large eta*, *Phys.Rev.* **D72** (2005) 023515, [arXiv:gr-qc/0503017], [doi:10.1103/PhysRevD.72.023515].

[13] K. Tzirakis and W. H. Kinney, *Inflation over the hill*, *Phys.Rev.* **D75** (2007) 123510, [arXiv:astro-ph/0701432], [doi:10.1103/PhysRevD.75.123510].

[14] M. H. Namjoo, H. Firouzjahi, and M. Sasaki, *Violation of non-Gaussianity consistency relation in a single field inflationary model*, *Europhys.Lett.* **101** (2013) 39001, [arXiv:1210.3692], [doi:10.1209/0295-5075/101/39001].

[15] J. Martin, H. Motohashi, and T. Suyama, *Ultra Slow-Roll Inflation and the non-Gaussianity Consistency Relation*, *Phys.Rev.* **D87** (2013), no. 2 023514, [arXiv:1211.0083], [doi:10.1103/PhysRevD.87.023514].

[16] H. Motohashi, A. A. Starobinsky, and J. Yokoyama, *Inflation with a constant rate of roll*, arXiv:1411.5021.

[17] S. Mooij, G. A. Palma, and A. E. Romano, *Consistently violating the non-Gaussian consistency relation*, arXiv:1502.03458.

[18] L.-M. Wang, V. F. Mukhanov, and P. J. Steinhardt, *On the problem of predicting inflationary perturbations*, *Phys.Lett.* **B414** (1997) 18–27, [arXiv:astro-ph/9709032], [doi:10.1016/S0370-2693(97)01166-0].

[19] D. Boyanovsky, H. J. de Vega, and N. Sanchez, *CMB quadrupole suppression. 2. The early fast roll stage*, *Phys.Rev.* **D74** (2006) 123007, [arXiv:astro-ph/0607487], [doi:10.1103/PhysRevD.74.123007].

[20] G. Nicholson and C. R. Contaldi, *The large scale CMB cut-off and the tensor-to-scalar ratio*, *JCAP* **0801** (2008) 002, [arXiv:astro-ph/0701783], [doi:10.1088/1475-7516/2008/01/002].

[21] L. Lello and D. Boyanovsky, *Tensor to scalar ratio and large scale power suppression from pre-slow roll initial conditions*, *JCAP* **1405** (2014) 029, [arXiv:1312.4251], [doi:10.1088/1475-7516/2014/05/029].

[22] L. Lello, D. Boyanovsky, and R. Holman, *Pre-slow roll initial conditions: large scale power suppression and infrared aspects during inflation*, *Phys.Rev.* **D89** (2014), no. 6 063533, [arXiv:1307.4066], [doi:10.1103/PhysRevD.89.063533].

[23] M. Cicoli, S. Downes, B. Dutta, F. G. Pedro, and A. Westphal, *Just enough inflation: power spectrum modifications at large scales*, *JCAP* **1412** (2014), no. 12 030, [arXiv:1407.1048], [doi:10.1088/1475-7516/2014/12/030].

[24] W. Handley, S. Brechet, A. Lasenby, and M. Hobson, *Kinetic Initial Conditions for Inflation*, *Phys.Rev.* **D89** (2014), no. 6 063505, [arXiv:1401.2253], [doi:10.1103/PhysRevD.89.063505].

[25] J. White, Y.-l. Zhang, and M. Sasaki, *Scalar suppression on large scales in open inflation*, *Phys. Rev.* **D90** (2014), no. 8 083517, [arXiv:1407.5816], [doi:10.1103/PhysRevD.90.083517].

[26] A. A. Starobinsky, *Spectrum of adiabatic perturbations in the universe when there are singularities in the inflation potential*, *JETP Lett.* **55** (1992) 489–494.

[27] R. Bousso, D. Harlow, and L. Senatore, *Inflation after False Vacuum Decay: Observational Prospects after Planck*, *Phys.Rev.* **D91** (2015), no. 8 083527, [arXiv:1309.4060], [doi:10.1103/PhysRevD.91.083527].

[28] C. R. Contaldi, M. Peloso, and L. Sorbo, *Suppressing the impact of a high tensor-to-scalar ratio on the temperature anisotropies*, *JCAP* **1407** (2014) 014, [arXiv:1403.4596], [doi:10.1088/1475-7516/2014/07/014].

[29] R. Bousso, D. Harlow, and L. Senatore, *Inflation After False Vacuum Decay: New Evidence from BICEP2*, *JCAP* **1412** (2014), no. 12 019, [arXiv:1404.2278], [doi:10.1088/1475-7516/2014/12/019].

[30] D. K. Hazra, A. Shafieloo, G. F. Smoot, and A. A. Starobinsky, *Inflation with Whip-Shaped Suppressed Scalar Power Spectra*, *Phys.Rev.Lett.* **113** (2014), no. 7 071301, [arXiv:1404.0360], [doi:10.1103/PhysRevLett.113.071301].

[31] D. K. Hazra, A. Shafieloo, G. F. Smoot, and A. A. Starobinsky, *Wiggly Whipped Inflation*, *JCAP* **1408** (2014) 048, [arXiv:1405.2012], [doi:10.1088/1475-7516/2014/08/048].

[32] S. M. Leach and A. R. Liddle, *Inflationary perturbations near horizon crossing*, *Phys. Rev.* **D63** (2001) 043508, [arXiv:astro-ph/0010082], [doi:10.1103/PhysRevD.63.043508].

[33] S. M. Leach, M. Sasaki, D. Wands, and A. R. Liddle, *Enhancement of superhorizon scale inflationary curvature perturbations*, *Phys. Rev.* **D64** (2001) 023512, [arXiv:astro-ph/0101406], [doi:10.1103/PhysRevD.64.023512].

[34] R. K. Jain, P. Chingangbam, and L. Sriramkumar, *On the evolution of tachyonic perturbations at super-Hubble scales*, *JCAP* **0710** (2007) 003, [arXiv:astro-ph/0703762], [doi:10.1088/1475-7516/2007/10/003].

[35] R. K. Jain, P. Chingangbam, J.-O. Gong, L. Sriramkumar, and T. Souradeep, *Punctuated inflation and the low CMB multipoles*, *JCAP* **0901** (2009) 009, [arXiv:0809.3915], [doi:10.1088/1475-7516/2009/01/009].

[36] R. K. Jain, P. Chingangbam, L. Sriramkumar, and T. Souradeep, *The tensor-to-scalar ratio in punctuated inflation*, *Phys. Rev.* **D82** (2010) 023509, [arXiv:0904.2518], [doi:10.1103/PhysRevD.82.023509].

Current Status and Future Prospects of the SNO+ Experiment

S. Andringa,[1] E. Arushanova,[2] S. Asahi,[3] M. Askins,[4] D. J. Auty,[5] A. R. Back,[2,6] Z. Barnard,[7] N. Barros,[1,8] E. W. Beier,[8] A. Bialek,[5] S. D. Biller,[9] E. Blucher,[10] R. Bonventre,[8] D. Braid,[7] E. Caden,[7] E. Callaghan,[8] J. Caravaca,[11,12] J. Carvalho,[13] L. Cavalli,[9] D. Chauhan,[1,3,7] M. Chen,[3] O. Chkvorets,[7] K. Clark,[3,6,9,*] B. Cleveland,[7,14] I. T. Coulter,[8,9] D. Cressy,[7] X. Dai,[3] C. Darrach,[7] B. Davis-Purcell,[15,†] R. Deen,[8,9] M. M. Depatie,[7] F. Descamps,[11,12] F. Di Lodovico,[2] N. Duhaime,[7] F. Duncan,[7,14] J. Dunger,[9] E. Falk,[6] N. Fatemighomi,[3] R. Ford,[7,14] P. Gorel,[5,‡] C. Grant,[4] S. Grullon,[8] E. Guillian,[3] A. L. Hallin,[5] D. Hallman,[7] S. Hans,[16] J. Hartnell,[6] P. Harvey,[3] M. Hedayatipour,[5] W. J. Heintzelman,[8] R. L. Helmer,[15] B. Hreljac,[7] J. Hu,[5] T. Iida,[3] C. M. Jackson,[11,12] N. A. Jelley,[9] C. Jillings,[7,14] C. Jones,[9] P. G. Jones,[2,9] K. Kamdin,[11,12] T. Kaptanoglu,[8] J. Kaspar,[17] P. Keener,[8] P. Khaghani,[7] L. Kippenbrock,[17] J. R. Klein,[8] R. Knapik,[8,18] J. N. Kofron,[17,§] L. L. Kormos,[19] S. Korte,[7] C. Kraus,[7] C. B. Krauss,[5] K. Labe,[10] I. Lam,[3] C. Lan,[3] B. J. Land,[11,12] S. Langrock,[2] A. LaTorre,[10] I. Lawson,[7,14] G. M. Lefeuvre,[6,¶] E. J. Leming,[6] J. Lidgard,[9] X. Liu,[3] Y. Liu,[3] V. Lozza,[20] S. Maguire,[16] A. Maio,[1,21] K. Majumdar,[9] S. Manecki,[3] J. Maneira,[1,21] E. Marzec,[8] A. Mastbaum,[8] N. McCauley,[22] A. B. McDonald,[3] J. E. McMillan,[23] P. Mekarski,[5] C. Miller,[3] Y. Mohan,[8] E. Mony,[3] M. J. Mottram,[2,6] V. Novikov,[3] H. M. O'Keeffe,[3,19] E. O'Sullivan,[3] G. D. Orebi Gann,[8,11,12] M. J. Parnell,[19] S. J. M. Peeters,[6] T. Pershing,[4] Z. Petriw,[5] G. Prior,[1] J. C. Prouty,[11,12] S. Quirk,[3] A. Reichold,[9] A. Robertson,[22] J. Rose,[22] R. Rosero,[16] P. M. Rost,[7] J. Rumleskie,[7] M. A. Schumaker,[7] M. H. Schwendener,[7] D. Scislowski,[17] J. Secrest,[24] M. Seddighin,[3] L. Segui,[9] S. Seibert,[8,‖] T. Shantz,[7] T. M. Shokair,[8,**] L. Sibley,[5] J. R. Sinclair,[6] K. Singh,[5] P. Skensved,[3] A. Sörensen,[20] T. Sonley,[3] R. Stainforth,[22] M. Strait,[10] M. I. Stringer,[6] R. Svoboda,[4] J. Tatar,[17] L. Tian,[3] N. Tolich,[17] J. Tseng,[9] H. W. C. Tseung,[17,††] R. Van Berg,[8] E. Vázquez-Jáuregui,[14,25] C. Virtue,[7] B. von Krosigk,[20] J. M. G. Walker,[22] M. Walker,[3] O. Wasalski,[15,‡‡] J. Waterfield,[6] R. F. White,[6] J. R. Wilson,[2] T. J. Winchester,[17] A. Wright,[3] M. Yeh,[16] T. Zhao,[3] and K. Zuber[20]

[1] *Laboratório de Instrumentação e Física Experimental de Partículas (LIP), Av. Elias Garcia, 14, 1°, 1000-149, Lisboa, Portugal*
[2] *Queen Mary, University of London, School of Physics and Astronomy, 327 Mile End Road, London, E1 4NS, UK*
[3] *Queen's University, Department of Physics, Engineering Physics & Astronomy, Kingston, ON K7L 3N6, Canada*
[4] *University of California, 1 Shields Avenue, Davis, CA 95616, USA*
[5] *University of Alberta, Department of Physics, 4-181 CCIS, Edmonton, AB T6G 2E, Canada*
[6] *University of Sussex, Physics & Astronomy, Pevensey II, Falmer, Brighton, BN1 9QH, UK*
[7] *Laurentian University, 935 Ramsey Lake Road, Sudbury, ON P3E 2C6, Canada*
[8] *University of Pennsylvania, Department of Physics & Astronomy, 209 South 33rd Street, Philadelphia, PA 19104-6396, USA*
[9] *University of Oxford, The Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK*
[10] *The Enrico Fermi Institute and Department of Physics, The University of Chicago, Chicago, IL 60637, USA*
[11] *University of California, Department of Physics, Berkeley, CA 94720, USA*
[12] *Lawrence Berkeley National Laboratory, Nuclear Science Division, 1 Cyclotron Road, Berkeley, CA 94720-8153, USA*
[13] *Universidade de Coimbra, Laboratório de Instrumentação e Física Experimental de Partículas and Departamento de Física, 3004-516, Coimbra, Portugal*

[14] *SNOLAB, Creighton Mine #9, 1039 Regional Road 24, Sudbury, ON P3Y 1N2, Canada*
[15] *TRIUMF, 4004 Wesbrook Mall, Vancouver, BC V6T 2A3, Canada*
[16] *Brookhaven National Laboratory, Chemistry Department, Building 555, P.O. Box 5000, Upton, NY 11973-500, USA*
[17] *University of Washington, Center for Experimental Nuclear Physics and Astrophysics, and Department of Physics, Seattle, WA 98195, USA*
[18] *Norwich University, 158 Harmon Drive, Northfield, VT 05663, USA*
[19] *Lancaster University, Physics Department, Lancaster, LA1 4YB, UK*
[20] *Technische Universität Dresden, Institut für Kern- und Teilchenphysik, Zellescher Weg 19, Dresden, 01069, Germany*
[21] *Universidade de Lisboa, Departamento de Física, Faculdade de Ciências, Campo Grande, Edifício C8, 1749-016 Lisboa, Portugal*
[22] *University of Liverpool, Department of Physics, Liverpool, L69 3BX, UK*
[23] *University of Sheffield, Department of Physics and Astronomy, Hicks Building, Hounsfield Road, Sheffield, S3 7RH, UK*
[24] *Armstrong Atlantic State University, Department of Chemistry & Physics, 11935 Abercorn Street, Savannah, GA 31419, USA*
[25] *Universidad Nacional Autónoma de México (UNAM), Instituto de Física, Apartado Postal 20-364, México D.F., 01000, México*

[*] Current address: University of Toronto, Department of Physics, 60 St. George Street, Toronto, M5S 1A7, Canada
[†] Current address: McMaster University, 1280 Main St W, Hamilton, ON L8S 4L8, Canada
[‡] Current address: SNOLAB, Creighton Mine #9, 1039 Regional Road 24, Sudbury, ON P3Y 1N2, Canada
[§] Current address: Porch 2200 1st Ave. South, Seattle, WA 98134, USA
[¶] Current address: Micron Semiconductor Ltd., 1 Royal Buildings, Marlborough Road, Lancing Business Park, Lancing, Sussex, BN15 8SJ, UK
[‖] Current address: Continuum Analytics, 221 W 6th St #1550, Austin, TX 78701, USA
[**] Current address: Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA
[††] Current address: Mayo Clinic, Department of Radiation Oncology, Rochester, MN 55905, USA
[‡‡] Current address: Safe Software, 7445 132 St #2017, Surrey, BC V3W 1J8, Canada

SNO+ is a large liquid scintillator-based experiment located 2 km underground at SNOLAB, Sudbury, Canada. It reuses the Sudbury Neutrino Observatory detector, consisting of a 12 m diameter acrylic vessel which will be filled with about 780 tonnes of ultra-pure liquid scintillator. Designed as a multipurpose neutrino experiment, the primary goal of SNO+ is a search for the neutrinoless double-beta decay ($0\nu\beta\beta$) of ^{130}Te. In Phase I, the detector will be loaded with 0.3% natural tellurium, corresponding to nearly 800 kg of ^{130}Te, with an expected effective Majorana neutrino mass sensitivity in the region of 55-133 meV, just above the inverted mass hierarchy. Recently, the possibility of deploying up to ten times more natural tellurium has been investigated, which would enable SNO+ to achieve sensitivity deep into the parameter space for the inverted neutrino mass hierarchy in the future. Additionally, SNO+ aims to measure reactor antineutrino oscillations, low energy solar neutrinos, and geoneutrinos, to be sensitive to supernova neutrinos, and to search for exotic physics. A first phase with the detector filled with water will begin soon, with the scintillator phase expected to start after a few months of water data taking. The $0\nu\beta\beta$ Phase I is foreseen for 2017.

1. Introduction

SNO+ is a large-scale liquid scintillator experiment located at a depth of 5890 ± 94 meter water equivalent (m.w.e.) in Vale's Creighton mine in Sudbury, Canada. The deep underground location, the high purity of materials used, and the large volume make SNO+ an ideally suited detector to study several aspects of neutrino physics.

The main goal of SNO+ is a search for the neutrinoless double-beta decay ($0\nu\beta\beta$) of ^{130}Te. $0\nu\beta\beta$-decay is a rare nuclear process that will happen if neutrinos are Majorana-type particles, that is, they are their own antiparticles. Understanding the Majorana nature of neutrinos is one of the most active areas of research in modern neutrino physics. The observation of the $0\nu\beta\beta$-decay would demonstrate lepton number violation, a key ingredient in the theory of leptogenesis. The process can be seen as two simultaneous β-decays, in which two neutrons are converted into two protons and two electrons, as the neutrinos from the two weak vertices mutually annihilate. The signature is a peak at the Q-value of the process in the summed energy spectrum of the two electrons. The measured quantity is the half-life of the decay. The effective Majorana neutrino mass, $m_{\beta\beta}$, which is highly dependent on the nuclear matrix elements, is derived from the half-life as described in [1]. A half-life of the order of 10^{25} years corresponds to a neutrino mass range of about 200–400 meV. The large mass and low background of SNO+ allow the investigation of such a rare event.

The large volume and the high radio-purity are also the reason why SNO+ can explore several other physics topics. Observation of geo-neutrinos will help in understanding the mechanisms for heat production in the Earth. Reactor antineutrino measurements constrain the neutrino oscillation parameters. Neutrinos and antineutrinos coming from supernova explosions would help to answer many unresolved questions in neutrino astronomy. Additionally, SNO+ has the potential to search for exotic physics like axion-like particles and invisible nucleon decay.

The depth of SNOLAB also provides the opportunity to measure low energy solar neutrinos, like *pep* and CNO neutrinos. The *pep* neutrinos are monoenergetic, with an energy of 1.44 MeV and a very well predicted flux, with an uncertainty of 1.2%, constrained by the solar luminosity [2]. A precise measurement of the flux can probe the Mikheyev, Smirnov and Wolfenstein (MSW) effect of neutrino mixing as well as alternate models like Non Standard Interactions [3]. Another open question in the solar neutrino field is related to the solar metallicity. The Standard Solar Model was always in excellent agreement with helioseismology until recent analyses suggested a metallicity about 30% lower than the previous model. This raised the question of the homogeneous distribution of elements heavier than helium in the Sun. The measurement of the CNO neutrino flux could be used to solve the problem [4].

This paper is structured as follows. In Sections 2 and 3 the SNO+ experiment is described, including the current status and detector upgrades. The expected background sources are presented in Section 4. In Sections 5 to 9 the broad physics program of SNO+ is described: the neutrinoless double-beta decay search (Section 5), the measurement of low energy solar neutrinos (Section 6), the measurements of geo and reactor antineutrinos (Section 7), the supernova neutrino watch (Section 8), and the exotic physics searches (Section 9). A brief conclusion follows at the end.

2. The SNO+ Experiment

The SNO+ experiment [5] is located in the underground laboratory of SNOLAB, Sudbury, Canada. A flat overburden of 2092 m of rock provides an efficient shield against cosmic muons corresponding to 5890 ± 94 m.w.e. [6]. The resulting muon rate through a 8.3 m radius circular area is 63 muons per day. SNO+ will make use of the SNO detector structure [7, 8] consisting of a spherical acrylic vessel (AV) of 6 m radius and 5.5 cm thickness located within a cavity excavated in the rock. The vessel will be filled with about 780 tonnes of liquid scintillator and will be viewed by ∼9300 PMTs supported by a geodesic stainless steel structure (PSUP) of approximately 8.9 m radius. The volume between the AV and the PSUP, as well as the rest of the cavity, will be filled

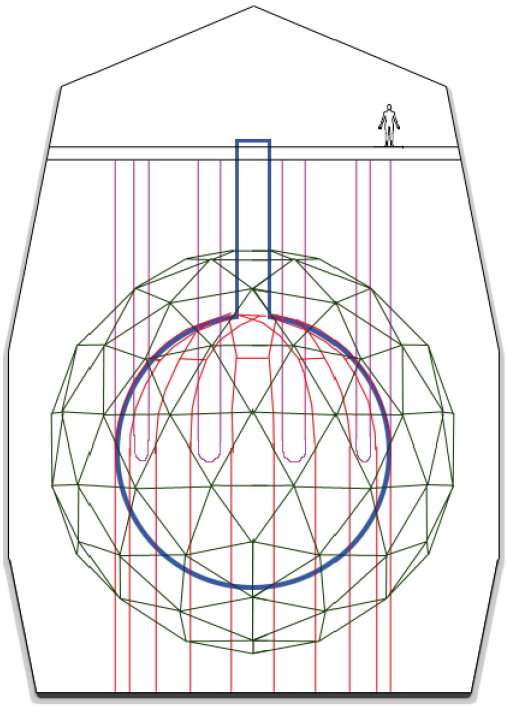

FIGURE 1: The SNO+ detector, figure from [9]. The 12-m diameter acrylic vessel (blue) is viewed by ∼9300 PMTs supported by a ∼18-m diameter geodesic structure (green) and is held by a system of high purity ropes (purple). The AV and the PSUP are within a volume of highly purified water. A rope net (red) will be used to offset the buoyancy of the liquid scintillator contained within the AV.

with about 7000 tonnes of ultra-pure water, which acts as a shield for the radioactivity coming from the rock (cavity walls) and the PMT array. A system of hold-up ropes suspends the acrylic vessel inside the PSUP. Additionally, in order to balance the buoyant force due to the lower density of the liquid scintillator compared to the external water, a new system of hold-down ropes has been installed on the top part of the AV and anchored at the cavity floor. A sketch of the detector is shown in Figure 1.

The major detector upgrades, including the liquid scintillator process systems, are described here.

2.1. Liquid Scintillator. The SNO+ liquid scintillator (LS) is composed of an aromatic hydrocarbon, linear alkylbenzene (LAB), as a solvent, and a concentration of 2 g/L 2,5-diphenyloxazole (PPO) as a fluor. LAB was selected as the liquid scintillator for SNO+ because of (1) its long time stability, (2) compatibility with the acrylic, (3) high purity levels directly from the manufacturer, (4) long attenuation and scattering length, (5) high light yield, and (6) linear response in energy. Additionally, it has a high flash point and is environmentally safe. LAB will be produced very close to the detector location (at the Cepsa plant in Becancour, Quebec, less than 900 km away), allowing short transport times which are important to reduce the possibility of cosmogenic activation.

2.2. Te-Loading. One of the main advantages of using LAB as liquid scintillator is the possibility of dissolving heavy metals with long term stability and good optical properties. For the $0\nu\beta\beta$-decay phase of the experiment, SNO+ will load tellurium into the liquid scintillator. An innovative technique has been developed to load tellurium at concentration levels of several percent into LAB maintaining good optical properties and reasonably high light emission levels [10]. Telluric acid, $Te(OH)_6$, is first dissolved in water and then, adding a surfactant, loaded into the scintillator. To better match the PMT quantum efficiency a secondary wavelength shifter will also be added to the mixture. Currently, we are investigating two different secondary wavelength shifters: perylene and bis-MSB. The former shifts the emission peak's range from 350–380 nm to ∼450–480 nm with a predicted light yield in SNO+ of about 300 Nhits (detected photoelectron hits) per MeV of energy. The latter shifts the emission peak to ∼390–430 nm with a light yield of 200 Nhits/MeV. The final choice will depend on the timing optical properties, the light yield, and the scattering length of the full scintillator mixture.

2.3. Emission Timing Profiles and Optical Properties. The emission timing profile and the optical properties of the LAB-PPO and the Te-loaded scintillator have been thoroughly investigated. The timing profile of scintillation pulses depends on the ionization density of the charged particles, with signals caused by electrons being faster than those from protons or alpha particles. This property allows the discrimination among particle types, which is very important for background rejection. The timing profile of electron and alpha particles in the unloaded scintillator has been measured in [11]. Results show that, for a LAB-PPO sample, a peak-to-total ratio analysis allows us to reject $> 99.9\%$ of the alpha particles while retaining $> 99.9\%$ of the electron signal.

The measurement of the timing profiles in the 0.3% Te-loaded scintillator is described in [12]. The presence of water and the surfactant in the cocktail reduces the long tail of the alpha decay (slow component) with respect to the unloaded scintillator, resulting in a poorer discrimination between α-like and β-like signals.

The light yield of the unloaded LAB-PPO scintillator has been measured in bench top tests and extrapolated for the full SNO+ volume using Monte Carlo (MC) simulations, leading to 520 Nhits/MeV.

The energy response to the electron energy deposition, the index of refraction, and the absorption length

of the LAB-PPO liquid scintillator are investigated in [13, 14]. The energy response is linear in the region from 0.4 MeV to 3.0 MeV, while below 0.4 MeV the linearity is lost due to reemission effects and the loss of Cherenkov light (threshold of ∼0.2 MeV).

Finally, the quenching of proton and alpha particles for the unloaded scintillator and the Te-loaded cocktail has been measured in [15, 16]. The nonlinear energy–dependent proton/alpha light output is typically parametrized by Birks' parameter kB [17]. Its measurement is extremely important for the development of background rejection techniques as described in Section 4. For protons in the unloaded SNO+ scintillator, the value measured in [15] is $kB = 0.0098 \pm 0.0003\,\mathrm{cm{\cdot}MeV^{-1}}$. The measured value for alpha particles is $kB = 0.0076 \pm 0.0003\,\mathrm{cm{\cdot}MeV^{-1}}$, corresponding, approximately, to a quenching factor of 10 for energies between 5 MeV and 9 MeV.

2.4. Process Plant. The scintillator purification plant of SNO+ is fully described in [18, 19]. It will use the same techniques and has the same cleanliness requirements as the Borexino experiment, by which we expect to reach a purity level of about 10^{-17} g/g$_{LAB}$ for both the ^{238}U and ^{232}Th chain [20], corresponding to 9 counts per day (cpd) for the ^{238}U chain and 3 cpd for the ^{232}Th chain. Similar background levels have also been achieved by the KamLAND experiment [21]. A multistage distillation (to remove heavy metals and optical impurities) and a high temperature flash vacuum distillation are initially used to separately purify LAB and PPO. Then the PPO is combined with the LAB, and the scintillator is further purified by a N_2/steam gas stripping process to remove gases, such as Rn, Ar, Kr, O_2, and residual water.

After the detector fill, the entire scintillator volume can be recirculated in about 4 days to enable quasi-batch repurification and *ex-situ* radio-assaying. A rotating-stage liquid-liquid extraction column (water-LAB) and metal scavengers are used to effectively remove metals (K, Pb, Bi, Th, and Ra). Finally, microfiltration is used for removal of suspended fine particles.

During the neutrinoless double-beta decay phase, the tellurium, the water, and the surfactant will be purified prior to addition to the LAB-PPO scintillator. The purification technique for tellurium is described in [22]. It has been designed to remove both the U- and Th-chain impurities and the isotopes produced by cosmogenic neutron and proton spallation reactions while handling and storing tellurium on surface. It consists of a double-pass acid-recrystallization on the earth's surface, for which the overall purification factor reached in U/Th and cosmogenic induced isotopes is $>10^4$. Since the tellurium purification is expected to happen at the above ground facilities and some isotopes can be cosmogenically replenished even with short time exposures, a second purification stage is needed underground. In this stage telluric acid is dissolved in water at 80°C and left to cool to recrystallize without further rinsing. A further purification of about a factor 100 is obtained. Currently, we are investigating the possibility of moving the above ground purification underground, in order to reduce potential recontamination.

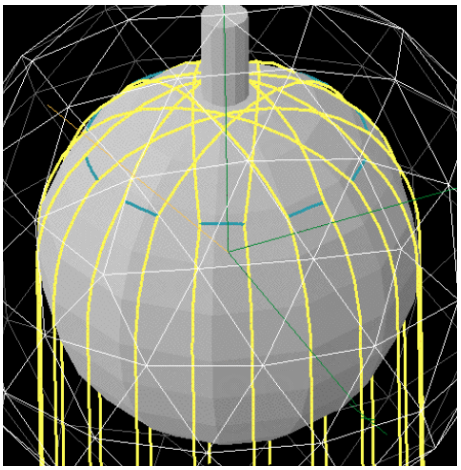

FIGURE 2: Sketch of the hold down rope system on the top of the acrylic vessel to compensate for the buoyant force that the scintillator produces on the AV.

The water purification plant at the SNOLAB underground laboratory is based on the SNO light water purification plant, which has been upgraded to improve its performance.

Spike tests have shown that some of the isotopes produced by cosmogenic activation of the surfactant are harder to remove by purification than in the case of telluric acid. The procedure to obtain pure surfactant will therefore be based on its chemical synthesis in a dedicated underground plant.

2.5. AV Rope System. The SNO+ liquid scintillator has a lower density ($\rho = 0.86$ g/cm^3 for LAB-PPO at T = 12°C) compared to the surrounding light water, requiring a new hold-down rope system (see Figure 2) to compensate the buoyant force, anchoring the acrylic vessel to the cavity floor. The new hold-down rope system consists of very high purity, high-performance polyethylene fiber (Tensylon) ropes of 38 mm diameter. The original hold-up rope system has also been replaced with new Tensylon ropes of 19 mm diameter in order to reduce the radioactivity contamination.

2.6. PMTs and Electronics. SNO+ uses the original 8 inch SNO photomultiplier tubes (Hamamatsu R1408). Each PMT is equipped with a 27 cm diameter concentrator, increasing the effective photocathode coverage to about 54%. Faulty PMT bases have been repaired and replaced, and about 9400 PMTs (90 of which are facing

Table 1: Calibration sources that are considered for use by the SNO+ experiment.

Source	AmBe	^{60}Co	^{57}Co	^{24}Na	^{48}Sc	^{16}N	^{220}Rn/^{222}Rn
Radiation	n, γ	γ	γ	γ	γ	γ	α, β, γ
Energy [MeV]	2.2, 4.4 (γ)	2.5 (sum)	0.122	4.1 (sum)	3.3 (sum)	6.1	various

outwards) are expected to be in operation at the start of the SNO+ experiment data taking.

In SNO+ the use of liquid scintillator as target volume greatly increases the light yield in contrast to the SNO heavy water, allowing the measurement of very low energy signals, like *pp* solar neutrinos (0.4 MeV end-point energy). Moreover, some of the background event types have high rates of several hundred Hz. For these reasons, the SNO readout boards and the data acquisition system were replaced with new ones capable of a higher bandwidth. New utilities have been added to the SNO+ trigger system which will allow for a more sophisticated use, a flexible calibration interface, and new background cuts to improve the physics sensitivity. The SNO+ trigger window is 400 ns long, during which time information and charge are collected from every PMT that fired. A dead-time of 30–50 ns separates two trigger windows [9].

In 2012 and 2014, the new electronics and trigger system were tested in runs with the detector empty and nearly half-filled with ultra-pure water (UPW).

2.7. Cover Gas System. As long-lived radon daughters are a potential background for the physics goals of SNO+ (see Section 4), the original SNO cover gas system has been upgraded to prevent radon ingress in the detector during operation. It consists of a sealed system filled with high purity nitrogen gas which acts as a physical barrier between the detector and the $\sim$130 Bq/m^3 of radon in the laboratory air. A new system of radon tight buffer bags has been designed and installed to accommodate the mine air pressure changes, with the aim of reaching a factor 10^5 in radon reduction.

2.8. Calibration Systems. The SNO+ detector will be calibrated using both optical sources (LEDs and lasers coupled to optical fibers) and radioactive sources (beta, gamma, alpha, and neutron). The optical sources are used to verify the PMT response and to measure *in situ* the optical properties of the detector media, while the radioactive sources are used to check the energy scale, the energy resolution, the linearity of the response, and the detector asymmetries, and to determine the systematic uncertainties and the efficiency of all reconstructed quantities (i.e., energy, position, and direction). Additionally, a system of cameras in underwater enclosures will be used to monitor the position of the acrylic vessel and the hold-down rope system, and to triangulate the positions of the calibration sources inserted into the detector.

The SNO+ calibration hardware has been designed to match the purity requirements of SNO+ and the need to have materials compatible with LAB. The calibration sources will be attached to an umbilical and moved by a system of high purity ropes in order to scan the detector off the central axis in two orthogonal planes.

The set of radioactive sources that are considered for the SNO+ experiment is shown in Table 1, covering the energy range from 0.1 MeV to 6 MeV. In addition, the internal radioactivity can be used to calibrate the detector and check any energy shift or variation of the response during data taking. Typical calibration references are ^{210}Po-alpha, ^{14}C-beta, delayed ^{214}Bi-Po (^{238}U chain) and ^{212}Bi-Po (^{232}Th chain) coincidences and muon followers.

The optical calibration hardware consists of internally deployable sources – a laserball (light diffusing sphere) and a Cherenkov source for absolute efficiency measurements – and an external system consisting of sets of optical fibers attached to the PSUP in fixed positions, sending pulses from fast LEDs or lasers into the detector. This system allows frequent calibrations of the PMTs response, time, and gain [23], and measuring the scattering and attenuation length of the scintillator without the need for source insertion.

2.9. Simulation and Analysis. A Geant4-based software package RAT (RAT is an Analysis Tool) has been developed to simulate the physics events in the SNO+ detector in great detail, and to perform analyses such as vertex reconstruction. The RAT simulation includes full photon propagation, from generation via scintillation and Cherenkov processes, through to absorption and detection on the PMTs. The detailed data acquisition and trigger systems are also part of the simulation. Several particle generators have been developed to simulate $0\nu\beta\beta$-decay events, solar neutrinos, geoneutrinos, reactor antineutrinos, supernova neutrinos and antineutrinos. The decay schemes of all relevant background isotopes are also part of the simulation tool. RAT communicates with a database that contains calibration constants and parameters describing the detector status during each run. This includes the optical properties of the various components of the scintillator cocktail, PMT calibration constants, and detector settings such as channel thresholds. Algorithms have been developed to reconstruct event in-

formation such as the vertex position, event direction (where relevant), and deposited energy. The SNO+ MC tool is continuously tuned to match newly available measurements.

For all SNO+ physics topics we have run a full Monte Carlo simulation to predict the fraction of background events in the corresponding region of interest (ROI), from which we have evaluated our sensitivities.

3. Physics Goals, Current Status, and Run Plan

The primary goal of SNO+ is to search for the neutrinoless double-beta decay of ^{130}Te. However, it has the potential to explore other physics including the following.

- *Low Energy pep and CNO Solar Neutrinos*
  The *pep*-neutrinos can be used to constrain new physics scenarios on how neutrinos couple to matter, while the CNO-neutrino flux can shed light on unresolved questions regarding solar metallicity.
- *Geoneutrinos*
  They are produced by the decay of U and Th chains in the Earth's crust and mantle. They can help to understand the heat production mechanisms of the Earth itself.
- *Reactor Antineutrinos*
  These can be used to better constrain the Δm^2_{12} neutrino oscillation parameter.
- *Supernova Neutrinos and Antineutrinos*
  The ability to detect a galactic supernova provides the potential for improving models of supernova explosions.
- *Exotic Physics*
  The low background expected in SNO+ allows searches for processes predicted by physics beyond the standard model (other than $0\nu\beta\beta$-decay), like invisible nucleon decay, and solar axion or axion-like particle searches.

Currently, the SNO+ cavity is partially filled with ultra-pure water. The upgrades to the SNO+ detector are nearly completed with a few items to be finished before the start of data taking. The detector parts that need to be finalized are the installation of the calibration system, underwater cameras, and the calibration optical fibers in most of the positions above the SNO+ equator, and the replacement of the PMTs. The installation will proceed along with the rise of the water level in the cavity. The scintillator plant is nearly completed. The newly installed electronic and trigger system and part of the optical calibration system have been tested in air and with the partially water-filled detector.

The data taking period of SNO+ will be divided into three main phases:

Water phase: In this phase, the acrylic vessel will be filled with about 905 tonnes of ultra-pure water and data taking will last for a few months. The main physics goals will be a search for exotic physics, including solar axion-like particles and invisible nucleon decay in ^{16}O, the watch for supernova neutrinos, and the detection (potentially) of reactor antineutrinos. During this phase, the detector performance, the PMT response and the data acquisition system characteristics will be tested. Optical calibrations to test the response of the PMT concentrators and the attenuation of the external water and the acrylic will be performed. The backgrounds coming from external sources, like external water, PMT array, hold-down ropes, and the acrylic vessel, will be characterized.

Pure scintillator phase: In this phase, the detector will be filled with about 780 tonnes of LAB-PPO liquid scintillator and data taking will last for a few months. The physics topics covered are the measurement of the low energy solar neutrinos, the measurement of geo and reactor antineutrinos, and the supernova neutrino watch. This phase will also be used to verify the optical model and the detector response and to characterize the backgrounds due to internal and external radioactive sources.

Te-loading phase: This phase is foreseen to start in 2017 and last for about 5 years. In this phase, also called Phase I, about 2.3 tonnes of natural tellurium (0.3% loading by weight) will be added to the detector for the search for the $0\nu\beta\beta$-decay of ^{130}Te. Simultaneously, geo and reactor neutrinos can be observed, and the detector will be live to a potential supernova.

The physics program and capabilities of SNO+ will be discussed in Sections 5 to 9.

4. Backgrounds

The background sources of the SNO+ experiment can be divided into two main categories: internal and external. Internal backgrounds are all the non-signal interactions that occur inside the AV ($R < 6\,m$). External backgrounds are the interactions that are produced in the region outside the target volume but that can propagate or are reconstructed within it. Full Monte Carlo simulations, along with *ex-situ* assays are used to explore the different background sources and develop rejection techniques.

In the following subsections the various background sources are presented: internal ^{238}U chain (Section 4.1), ^{210}Bi and ^{210}Po decays (Section 4.2), internal ^{232}Th chain

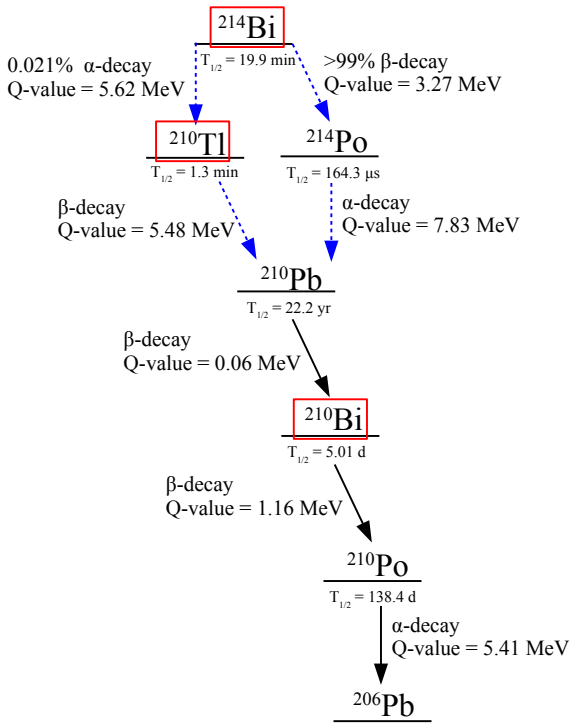

FIGURE 3: Part of ^{238}U-decay chain relevant for SNO+ with Q-values (total kinetic energy released in the ground state - ground state transition), half-life and decay modes [24]. The red squares highlight the nuclides of most concern: ^{214}Bi, ^{210}Tl, and ^{210}Bi. The decays used for α-β and β-α coincidence techniques are shown with a blue arrow (dash-dotted line).

(Section 4.3), internal ^{40}K, ^{39}Ar, and ^{85}Kr decays (Section 4.4), cosmogenically induced isotopes (Section 4.5), (α,n) reactions (Section 4.6), pile-up events (Section 4.7), and external backgrounds (Section 4.8).

4.1. Internal ^{238}U Chain. ^{238}U ($T_{1/2} = 4.47\times10^{9}$ yr) is a naturally occurring radioisotope present in the liquid scintillator. The part of the decay chain relevant for SNO+ is shown in Figure 3. The ^{238}U daughters of most concern are ^{214}Bi, ^{210}Tl, and ^{210}Bi (see Section 4.2). Secular equilibrium with the top part of the chain is assumed through the paper unless otherwise noted.

^{214}Bi ($T_{1/2} = 19.9$ min) beta-decays to ^{214}Po with a Q-value of 3.27 MeV in 99.979% of the cases. This decay can be tagged using the ^{214}Po alpha-decay ($T_{1/2} = 164.3\,\mu$s, $E_\alpha = 7.7$ MeV), during both the pure scintillator and the Te-loaded phase. In the pure scintillator phase, the β–α delayed coincidence will be used to measure the concentration of the ^{238}U-chain contaminants. ^{214}Bi is expected to be in secular equilibrium with the top part of the ^{238}U chain for most of the data taking period. This equilibrium can be broken by radon ingress into the detector during calibration campaigns, or from emanation by the calibration hardware materials. However, for non-continuum sources of radon, due to the short half-life of ^{214}Bi, equilibrium will be restored in a few weeks' time. In SNO+ the presence of the cover gas on the top of the detector provides an efficient barrier against laboratory air, highly reducing the radon ingress into the detector (see Section 2.7). Additionally, most of the radon short-lived daughters decay in the cover gas region or in the detector neck; thus they do not reach the fiducial volume.

During the Te phase, the delayed coincidence technique will be used to reject ^{214}Bi events that fall into the region of interest (ROI) for the $0\nu\beta\beta$-decay search.

Usually, Bi-β and Po-α are separated by more than 250 ns and the SNO+ detector records them as two separate events. The secondary events (alpha candidates) are identified by applying an energy cut around the alpha energy, shifted due to quenching to ~0.8 MeV electron equivalent energy, and by the short time separation from the preceding event. To reduce the misidentification of the events due to other decays occurring in the same energy region during the coincidence window, a position cut can also be applied. An α-β classification algorithm has been developed to further reduce the misidentification by classifying the events as α-like or β-like based on the hit-time distribution.

Occasionally, the beta and the alpha decays are separated by less than 250 ns and they may be recorded as a single event by the SNO+ detector. These events are important for the $0\nu\beta\beta$-decay phase as they may fall into the ROI. In this case, the rejection technique is based on the distortion in the time distribution of the light detected by the PMTs compared to the case of a single interaction. This rejection technique is enhanced if a pulse shape analysis can be applied to distinguish beta from alpha events.

In 0.021% of the cases ^{214}Bi alpha-decays to ^{210}Tl ($T_{1/2} = 1.3$ min), which beta-decays to ^{210}Pb with a Q-value of 5.5 MeV. Due to the small branching ratio this route is less important than the previous one. An α–β delayed coincidence, similar to the β–α one, can be applied. However, due to the longer half-life of ^{210}Tl, the mis-tagging probability is larger with respect to the ^{214}Bi-Po one which may result in a larger signal sacrifice.

Based on Borexino Phase-I achievements [20], the purity level aimed (target level) in the LAB-PPO scintillator for the ^{238}U chain is 1.6×10^{-17} g/g (see Table 2). During the Te-loaded phase, the addition of the isotope, the water, and the surfactant to LAB will worsen the mixture purity, but we will maintain a strict target level of 2.5×10^{-15} g/g (see Table 2).

4.2. ^{210}Bi and ^{210}Po Backgrounds. The ingress of ^{222}Rn into the SNO+ detector can break the secular-equilibrium in the ^{238}U chain at ^{210}Pb ($T_{1/2} = 22.2$ yr,

TABLE 2: Target levels, in g/g, and corresponding decay rates for the internal ^{238}U- and ^{232}Th-chain contaminants in the various SNO+ phases. Secular equilibrium has been assumed for all the isotopes except ^{210}Pb, ^{210}Bi, and ^{210}Po. The levels of ^{210}Bi and ^{210}Po during the pure scintillator phase and the Te-loaded phase are expected to be out of secular equilibrium due to the intrinsic scintillator contamination and the leaching off of the AV surface. For the 0.3% Te-loaded scintillator the tellurium/polonium affinity component is also included in the ^{210}Po decays/yr (see text).

Source	Target [g/g]	Decays/yr
Internal H_2O, water phase		
^{238}U chain	3.5×10^{-14}	1.2×10^{7}
^{232}Th chain	3.5×10^{-15}	4.1×10^{5}
LAB-PPO, pure scintillator phase		
^{238}U chain	1.6×10^{-17}	4900
^{232}Th chain	6.8×10^{-18}	700
^{210}Bi	-	7.6×10^{8} [a]
^{210}Po	-	7.8×10^{8} [a]
0.3% Te-loaded scintillator, Te phase		
^{238}U chain	2.5×10^{-15}	7.6×10^{5}
^{232}Th chain	2.8×10^{-16}	2.8×10^{4}
^{210}Bi	-	7.9×10^{9} [b]
^{210}Po	-	9.5×10^{9} [b]

[a]Expected number of events in the first year after 9 months of water phase.

[b]Expected number of events in the first year after 9 months of water phase followed by 6 months of pure scintillator phase.

Q-value = 0.06 MeV), resulting in a higher concentration of this isotope. Even if ^{210}Pb is not a direct background for the SNO+ experiment, its daughters ^{210}Bi ($T_{1/2}$ = 5.0 d, Q-value = 1.16 MeV) and ^{210}Po ($T_{1/2}$ = 138.4 d, E_α = 5.3 MeV, shifted to ∼0.5 MeV electron equivalent energy) are potentially relevant for the various physics searches. ^{210}Bi-beta decays are the main background for the CNO-ν measurement, as they have similar spectral shapes, while the ^{210}Po-alpha decay is a background for the β–α and α–β delayed coincidences, resulting in mistagging and potential signal sacrifice. Additionally, the emitted alphas can interact with the atoms in the scintillator producing neutrons as described in Section 4.6. The cover gas system placed at the top of the acrylic vessel greatly reduces the radon ingress into the detector. Furthermore, the majority of short-lived daughters decay before reaching the fiducial volume. However, due to its long half-life, ^{210}Pb is not attenuated by the presence of the detector neck and reaches the target volume.

^{210}Pb and its daughters may also leach from materials that are in contact with the liquid scintillator. Radon daughters deposited on the material's surface can implant by alpha recoil to a depth of a few hundred nm, where they eventually decay to ^{210}Pb. ^{210}Pb, ^{210}Bi, and ^{210}Po atoms might then leach off when the liquid scintillator mixture is in contact with the surface. This process can happen, for instance, during the handling and storing of the liquid scintillator, resulting in rates of ^{210}Pb, ^{210}Bi, and ^{210}Po out of equilibrium with the ^{238}U chain. Concentrations of ^{210}Bi and ^{210}Po different from each other and the rest of the ^{238}U chain have been seen by the Borexino experiment [25]. The levels initially measured by Borexino for these two isotopes are included in Table 2.

An additional source of ^{210}Pb, ^{210}Bi, and ^{210}Po is leaching from the internal surface of the AV, where radon daughters have implanted during the construction of SNO and when the detector was empty after draining the heavy water. This may create a continuous source of ^{210}Pb, ^{210}Bi, and ^{210}Po during the data taking period for all SNO+ phases. Leaching rates depend on several factors, like temperature, implantation depth, type of liquid in contact with the surface, and initial surface activity. The leaching rate of ^{210}Pb and its daughters for all the scintillator mixtures and the ultra pure water at different temperatures have been measured in bench top tests. With a measured activity of about 1 kBq on the inner AV surface, the activity of ^{210}Pb daughters leached in the scintillator media might be as high as a few hundred Bq depending on the duration of the data taking period. The activity of the backgrounds leached in the scintillator is expected to increase with time, while that of inner surface events is expected to decrease.

In the Te-loaded phase, an additional source of ^{210}Po is the tellurium itself. The CUORE collaboration has shown [26] that due to the chemical affinity between tellurium and polonium this element may still be present in tellurium after the crystal production process. In our background estimations we assume an additional ^{210}Po activity of 0.06 Bq/kg$_{Te}$, based on CUORE measurements. These decays, however, are not supported by ^{210}Pb and are considerably reduced, to about 16% of the initial activity, in a year after tellurium production. This contribution is included in the purity levels of ^{210}Po shown in Table 2.

4.3. Internal ^{232}Th Chain. ^{232}Th ($T_{1/2} = 1.4\times10^{10}$ yr) is also a naturally occurring radioisotope present in the liquid scintillator. The daughters of most concern are ^{212}Bi and ^{208}Tl (see Figure 4).

^{212}Bi ($T_{1/2}$ = 60.6 min) beta-decays to ^{212}Po ($T_{1/2}$ = 300 ns) with a Q-value of 2.25 MeV in 64% of the cases. As for the ^{214}Bi→^{214}Po decay, many events can be selected using a β–α delayed coincidence, which is used to extract the concentration of the ^{232}Th-chain contaminants in equilibrium in the pure scintillator. Nearly 45% of the ^{212}Bi→^{212}Po decays fall in the same trigger window and are a potential background for the $0\nu\beta\beta$-decay search. These can be rejected using the PMT timing distribution.

In the remaining 36% of the cases ^{212}Bi alpha-decays

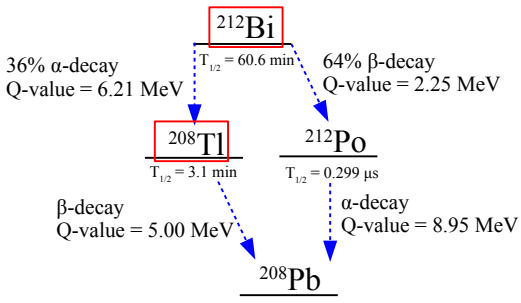

FIGURE 4: Part of ^{232}Th-decay chain relevant for SNO+ with Q-values, half-life and decay modes [24]. The red squares highlight the most important nuclides: ^{212}Bi and ^{208}Tl. The decays used for α-β and β-α coincidence techniques are shown with a blue arrow (dash-dotted line).

to ^{208}Tl ($T_{1/2}$ = 3.0 min), which beta-decays to ^{208}Pb with a Q-value of 5.0 MeV. An α–β delayed coincidence can be applied to identify the ^{208}Tl events as for the ^{210}Tl case.

The LAB-PPO scintillator target level for the ^{232}Th chain is 6.8×10^{-18} g/g (based on [20]), while the target level for the Te-loaded scintillator is 2.8×10^{-16} g/g (see Table 2).

4.4. Internal ^{40}K, ^{39}Ar, and ^{85}Kr Backgrounds. Other internal backgrounds are important for solar neutrino and other measurements.

^{40}K ($T_{1/2}$ = 1.248×10^{9} yr) has a very distinctive energy spectrum, having both a beta component and a gamma peak at 1.46 MeV. Due to the long half-life, it is naturally present in the scintillator and detector materials.

^{39}Ar ($T_{1/2}$ = 269 yr), and ^{85}Kr ($T_{1/2}$ = 10.8 yr) decay with a Q-value of 0.565 MeV and of 0.687 MeV, respectively. The amount of these isotopes can be reduced by minimising the contact time of LAB with air and thoroughly degassing the scintillator.

4.5. Cosmogenically Induced Backgrounds. Besides the natural radioactivity present in the scintillator, LAB can be activated by cosmic ray neutrons and protons while it is above ground. The main expected background is ^{7}Be ($T_{1/2}$ = 53.2 d, EC-decay with a 0.48 MeV gamma), with a maximum production rate at sea level (neutron and proton flux from [27, 28]) of about 1 kHz for 780 t of liquid scintillator. More than 99% of the produced ^{7}Be can be efficiently removed by the scintillator purification plant.

^{14}C ($T_{1/2}$ = 5700 yr, Q-value = 0.16 MeV) is naturally present in the liquid scintillator. It is a direct background for the very low energy *pp* neutrino measurements and may contribute to pile-up backgrounds (see Section 4.7). In SNO+, we expect a ^{14}C/^{12}C ratio of the order of 10^{-18}, similar to what was observed in the Borexino test facility [29], corresponding to a decay rate of a few hundred Hz. This is a reasonable assumption as in both cases the liquid scintillator is obtained from old oil fields, in which most of the ^{14}C has decayed away. The amount of ^{14}C produced by cosmogenic activation of LAB during transport to site is negligible in comparison.

^{11}C ($T_{1/2}$ = 20 min, Q-value = 1.98 MeV) is mainly produced by muon interactions with the carbon nuclei of the liquid scintillator. We expect a total of $(1.14\pm0.21)\times10^{3}$ decays/kt/yr during operation, extrapolated from KamLAND data in [30]. This is about a factor 100 less than what was observed in Borexino [31] due to the deeper underground location. A threefold coincidence tagging technique, like the one developed by Borexino [32], together with an electron-positron discrimination analysis [33], will further reduce these events.

Other muon induced backgrounds are generally very short lived (milliseconds to seconds half-life) and can be rejected by vetoing the detector for a few minutes after each muon event.

Important cosmogenic-induced backgrounds are isotopes produced by spallation reactions on tellurium while it is stored on surface [34], like ^{124}Sb ($T_{1/2}$ = 60.2 d, Q-value = 2.90 MeV), ^{22}Na ($T_{1/2}$ = 950.6 d, Q-value = 2.84 MeV), ^{60}Co ($T_{1/2}$ = 1925 d, Q-value = 2.82 MeV), ^{110m}Ag ($T_{1/2}$ = 249.8 d, Q-value = 2.89 MeV, E_{parent}(level) = 0.118 MeV) and ^{88}Y ($T_{1/2}$ = 106.6 d, Q-value = 3.62 MeV). We have developed a purification technique [22] (see Section 2.4) that, together with underground storage, reduces the cosmogenic-induced background on tellurium to a negligible level.

4.6. (α,n) Backgrounds. Neutrons can be produced in the liquid scintillator by (α,n) reactions on ^{13}C or ^{18}O atoms, muon interactions in the scintillator volume, ^{238}U fission, and (γ,n) reactions for $E_\gamma > 3$ MeV. Excluding the muon induced neutrons, the most prominent neutron source inside the scintillator volume is the $\alpha+{}^{13}\mathrm{C}\rightarrow{}^{16}\mathrm{O}+n$ reaction ($E_{thr.}$ = 0.0 keV), which is a potential background for both the $0\nu\beta\beta$-decay search and the antineutrino measurement. The main source of alpha particles in the various scintillator mixtures is ^{210}Po. Other U- and Th- chain's alpha emitters form a negligible contribution, as they are expected to be ~4 orders of magnitude less abundant.

Neutrons produced in (α,n) reactions will scatter from protons during the thermalization process, resulting in recoils emitting scintillation light. The visible proton energy together with the energy lost by the alphas before interaction is the prompt signal. If the isotope is in an excited state, the emitted deexcitation gammas are also part of the prompt signal. The thermalized neutrons in >99% of the cases are eventually captured by hydrogen atoms with the emission of the characteristic 2.22 MeV γ. In the remaining ~1% of the cases the

TABLE 3: ^{238}U- and ^{232}Th-chain levels for external background sources. Shown are measured levels and expected decay rates.

Source	Measured levels	Decays/yr
Internal ropes	^{214}Bi: $(2.8 \pm 5.4) \times 10^{-10}g_U$/g [37]	4966
	^{208}Tl: $< 2.0 \times 10^{-10}$g$_{Th}$/g [37]	< 418
Hold-down ropes	^{214}Bi: $(4.7 \pm 3.2) \times 10^{-11}g_U$/g [37]	4.06×10^6
	^{208}Tl: $(2.27 \pm 1.13) \times 10^{-10}g_{Th}$/g[37]	2.32×10^6
Hold-up ropes	^{214}Bi: $(4.7 \pm 3.2) \times 10^{-11}g_U$/g [37]	8.34×10^5
	^{208}Tl: $(2.27 \pm 1.13) \times 10^{-10}g_{Th}$/g [37]	4.78×10^5
Water Shielding	^{214}Bi: 2.1×10^{-13}g$_U$/g [38]	1.32×10^8
	^{208}Tl: 5.2×10^{-14}g$_{Th}$/g [38]	3.92×10^6
Acrylic Vessel	^{214}Bi: $< 1.1 \times 10^{-12}$g$_U$/g [a][7]	1.28×10^7
	^{208}Tl: $< 1.1 \times 10^{-12}$g$_{Th}$/g [a][7]	1.50×10^6
Acrylic Vessel External Dust[b]	^{214}Bi: $(1.1 \pm 0.1) \times 10^{-6}g_U$/g [39]	7.8×10^5
	^{208}Tl: $(5.6 \pm 0.5) \times 10^{-6}g_{Th}$/g [39]	4.6×10^5
Acrylic Vessel Internal Dust	^{214}Bi: $(1.1 \pm 0.1) \times 10^{-6}g_U$/g [39]	4.15×10^4
	^{208}Tl: $(5.6 \pm 0.5) \times 10^{-6}g_{Th}$/g [39]	2.48×10^4
PMTs	^{214}Bi: $100 \times 10^{-6} g_U$/PMT [7]	3.7×10^{11}
	^{208}Tl: $100 \times 10^{-6} g_{Th}$/PMT [7]	4.4×10^{10}

[a]Assumed 1.0×10^{-12}g/g

[b]It is assumed that the top hemisphere of the external AV surface is not cleaned, while the bottom hemisphere is at target level

thermal neutron is captured either on tellurium isotopes, producing mainly a 0.6 MeV gamma, or on ^{12}C, producing a 4.95 MeV gamma. The prompt and the delayed signal can be used to reject the (α,n) background using a delayed coincidence technique similar to that of β-α events.

4.7. Pile-Up Backgrounds. A pile-up event occurs when two or more decays (signal or background or a mixture) happen in the same trigger window and thus are potentially detected as a single event with energy equal to the sum of the single energies. Pile-up events become important when the event rate of one or all of the contributing decays is very high (hundreds of Hz), like ^{14}C decays or ^{210}Bi or ^{210}Po. A rejection technique, using the distortion of the timing, is used to efficiently reduce these backgrounds [35, 36].

4.8. External Backgrounds. Sources of external background include the hold-down and hold-up ropes, the PMT array, the AV bulk, and the external water (see Table 3). Radioactive decays occur outside the scintillator volume, so the main concerns for the signal extraction analysis are the high energy gammas and betas emitted by ^{214}Bi, ^{208}Tl, and ^{40}K decays. External background events reconstructing inside the AV can be greatly reduced by applying a fiducial volume cut. Events can be further reduced using the PMT time distribution. *In-situ* analysis during the water phase and the pure liquid scintillator phase will help to constrain the external backgrounds for the Te-loaded phase.

5. ^{130}Te Neutrinoless Double-Beta Decay

The main goal of the SNO+ experiment is the search for neutrinoless double-beta decay of ^{130}Te (Q-value = 2527.518 ± 0.013 keV [40]) by loading large quantities of the isotope into the liquid scintillator volume. This approach has several advantages: (1) external backgrounds can be removed by fiducialization, (2) internal and external background levels can be measured before and after the isotope deployment, allowing identification and removal of possible contamination, (3) internal backgrounds can be tagged by coincidences or particle identification, (4) the detector response can be tested with and without the isotope, (5) the spatial distribution of most background isotopes in a liquid is known to be uniform, (6) the loading can be easily and affordably scaled up or (7) changed to another isotope, and (8) tellurium and scintillator can be removed and repurified if high levels of backgrounds are found.

The choice of ^{130}Te as the preferred $0\nu\beta\beta$ candidate is the result of an extensive investigation by the SNO+ collaboration. The decision was based on several factors, including the following points:

1. ^{130}Te has a large natural abundance of 34.08%, which allows loading of several tonnes of isotope without enrichment.
2. The measured half-life of the ^{130}Te $2\nu\beta\beta$ decay is $(7.0 \pm 0.9(\text{stat}) \pm 1.1(\text{syst})) \times 10^{20}$ yr [41], one of the longest of all the $0\nu\beta\beta$ isotopes. This is particularly

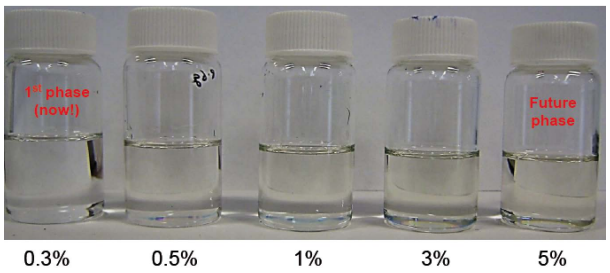

FIGURE 5: TeLS samples from the investigation of higher tellurium loading in LAB scintillator. The samples increase in loading from 0.3% (by weight) on the left to 5% on the right.

important for liquid scintillator-based experiments, as the energy resolution is usually some hundreds of keV.

3. An innovative loading technique has been developed, which enables deployment of up to 5% (by weight) of natural tellurium while maintaining good light transmission, minimal scattering, and an acceptable light yield (see Section 2.2). The 0.3% tellurium scintillator cocktail (TeLS) has been proven to be stable for a period of over two years. In Figure 5, various SNO+ loaded cocktails are shown. Cocktails with higher loading still maintain good optical transparency.
4. The TeLS does not present inherent optical absorption lines in the visible wavelength range, such that a secondary wavelength shifter may be added to the cocktail to better match the SNO+ PMT response.

5.1. Backgrounds. For the ^{130}Te $0\nu\beta\beta$ search, an asymmetric region of interest (ROI) is defined, which extends from -0.5σ to 1.5σ around the Gaussian signal peak. For the 0.3% Te-loaded cocktail with a light yield of 200 Nhits/MeV (see Section 2.2) the energy resolution at 2.5 MeV is ~270 keV (FWHM), while the averaged position resolution at the same energy is ~15 cm at the detector's center. An asymmetric ROI retains most of the $0\nu\beta\beta$ decays but considerably reduces the backgrounds from $2\nu\beta\beta$ and low energy ^{238}U- and ^{232}Th-chain decays. Most external backgrounds are rejected by a 3.5 m fiducial radius cut, which preserves 20% of signal events. Inside the 3.5 m fiducial volume (FV) and 2.47 MeV to 2.70 MeV energy ROI, the main background sources are:

^{8}B solar neutrinos: flat continuum background from the elastically scattered (ES) electrons normalized using the total ^{8}B flux and published solar mixing parameters [42].

$2\nu\beta\beta$: irreducible background due to the $2\nu\beta\beta$ decays of ^{130}Te: these events appear in the ROI due to the energy resolution of SNO+.

External Backgrounds: ^{208}Tl and ^{214}Bi nuclides contained in the AV, hold-down rope system, water shielding, and PMT glass are the major contributors in the defined ROI. The FV cut of 20% reduces these background events by several orders of magnitude. The PMT hit-time distribution cut reduces the external background events falling in the FV by an additional factor of two.

Internal ^{238}U- and ^{232}Th-chain Backgrounds: the dominant backgrounds in the signal ROI are due to ^{214}Bi-Po and ^{212}Bi-Po decays. Currently, we have achieved approximately 100% rejection of separately triggered ^{214}Bi-Po and ^{212}Bi-Po decays falling inside the ROI and FV using the β-α delayed coincidence. For ^{212}Bi-Po and ^{214}Bi-Po pile-up decays, cuts based on PMT hit timing achieve a rejection factor of ~50 for events that fall in the ROI and FV. Other minor contributions in the ROI are due to ^{234m}Pa (^{238}U chain), ^{210}Tl (^{238}U chain) and ^{208}Tl (^{232}Th chain).

Cosmogenic Backgrounds: The most relevant isotopes are ^{60}Co, ^{110m}Ag, ^{88}Y, and ^{22}Na (see Section 4.5). The developed purification techniques together with a long period of underground storage will reduce the cosmogenically induced background to less than one event per year in the FV and ROI.

(α,n) Backgrounds: both the prompt signal and the delayed 2.22 MeV-γ produced by (α,n) reactions can leak into the $0\nu\beta\beta$ ROI. Coincidence-based cuts have been developed that remove more than 99.6% of the prompt and ~90% of delayed events that fall in the FV and ROI.

Pile-up Backgrounds: the most important pile-up backgrounds for the $0\nu\beta\beta$ search are due to high-rate ^{210}Po+$2\nu\beta\beta$ and ^{210}Bi+$2\nu\beta\beta$, with bismuth and polonium coming from both the TeLS and the vessel surface. Timing-based cuts have been developed that reduce the pile-up backgrounds to a negligible level.

We have estimated the fraction of each background that falls in the ROI and FV based on our Monte Carlo simulations. A summary of the various background sources in the ROI and FV is shown in Table 4. The main contributions are due to ^{8}B ν ES and to $2\nu\beta\beta$. A total of about 22 events/yr in the FV and ROI are expected. The scale of the external background events within the ROI can be checked by fitting events outside the fiducial volume. Internal U- and Th-chain residuals can be checked via the ^{214}Bi-Po and the ^{212}Bi-Po delayed coincidences, whose tagging efficiency can be tested during the pure LAB-PPO scintillator phase. In addition, some of the cosmogenic-induced backgrounds, like ^{124}Sb and ^{88}Y, can be constrained using their relatively short half-life, while ^{8}B-ν and $2\nu\beta\beta$ decays can be constrained by

Table 4: Expected background counts in the signal ROI and 3.5 m FV in SNO+ for the first year (Year 1) and in 5 years of the 0.3% Te loading phase. A light yield of 200 Nhits/MeV has been assumed. Cuts have been applied as decribed in the text.

Isotope	1 Year	5 Years
$2\nu\beta\beta$	6.3	31.6
^{8}B ν ES	7.3	36.3
Uranium Chain	2.1	10.4
Thorium Chain	1.7	8.7
External	3.6	18.1
(α, n)	0.1	0.8
Cosmogenics	0.7	0.8
Total	21.8	106.8

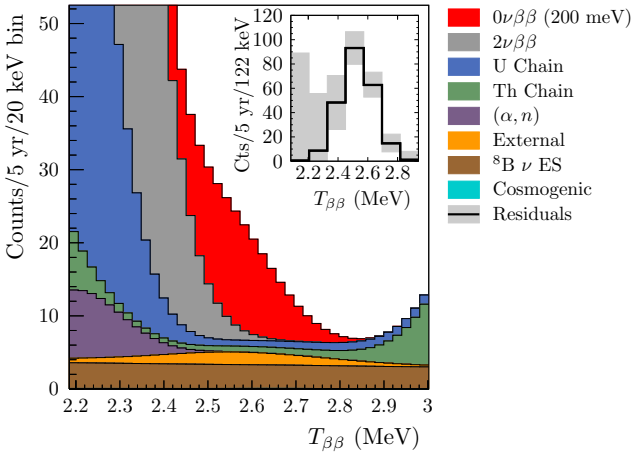

Figure 6: Summary stacked plot of all backgrounds and a hypothetical $0\nu\beta\beta$ signal corresponding to a mass $m_{\beta\beta} = 200$ meV for 5 years data taking. Events are shown in the FV of 3.5 m, for 0.3% natural tellurium loading and 200 Nhits/MeV light yield. $T_{\beta\beta}$ is the effective kinetic energy.

their known value. Furthermore, the detector response will be tested through a detailed calibration (see Section 2.8).

The expected signal and background spectrum for a five-year live-time is shown in Figure 6 for the 0.3% loading. A fiducial volume cut is applied at 3.5 m, >99.99% rejection for ^{214}Bi-Po and >98% for ^{212}Bi-Po are assumed, and the light yield is 200 Nhits/MeV. The $0\nu\beta\beta$ signal shown is for a $m_{\beta\beta} = 200$ meV, which corresponds to $T^{0\nu\beta\beta}_{1/2} \sim 1 \times 10^{25}$ yr using the IBM-2 nuclear matrix element [43].

5.2. Sensitivity. The expected number of $0\nu\beta\beta$ events occurring in the SNO+ detector is given by:

$$S = \epsilon \cdot N_{130} \cdot \ln 2 \cdot \frac{t}{T^{0\nu\beta\beta}_{1/2}} \quad (1)$$

where ϵ is the signal detection efficiency, N_{130} is the number of ^{130}Te atoms in the detector, t is the live-time, and $T^{0\nu\beta\beta}_{1/2}$ is the half-life of ^{130}Te $0\nu\beta\beta$. To compute the SNO+ sensitivity, we assume that the number of observed events in the FV and ROI is equal to the expected backgrounds. In this case the numerical value of the derived bound on the number of signal events is similar for either a Bayesian or a frequentist definition of 90% confidence level. With the natural tellurium concentration of 0.3% (by weight) in Phase I, corresponding to about 800 kg of ^{130}Te, a 20% FV cut, and five years of data taking, SNO+ can set a lower limit on the half-life of $T^{0\nu\beta\beta}_{1/2} > 9 \times 10^{25}$ yr at 90% CL ($T^{0\nu\beta\beta}_{1/2} > 4.8 \times 10^{25}$ yr at 3σ level). This corresponds to a limit on the effective Majorana neutrino mass, $m_{\beta\beta}$, of 55 – 133 meV, using a phase space factor $G = 3.69 \times 10^{-14}\,\mathrm{yr}^{-1}$ [44] and $g_A = 1.269$; the range is due to differences in nuclear matrix element calculation methods [43, 45, 46, 47, 48].

5.3. Higher Tellurium Concentration in the Future. One of the main advantages of the SNO+ technique is the possibility of moving toward higher sensitivities by increasing the loading. R&D efforts have demonstrated that, with 3% (by weight) tellurium loading, a light yield of 150 Nhits/MeV can be achieved using perylene as a secondary wavelength shifter. In SNO+ Phase II, this loss in light yield will be compensated by an upgrade to high quantum efficiency PMTs and improvements to PMT concentrators. These improvements will increase the light yield by a factor of ~ 3. A preliminary study shows that SNO+ Phase II can set a lower limit on the $0\nu\beta\beta$ half-life of $T^{0\nu\beta\beta}_{1/2} > 7 \times 10^{26}$ years (90% CL), for a $m_{\beta\beta}$ range of 19–46 meV.

6. Solar Neutrino Physics

SNO+ has the opportunity to measure low energy solar neutrinos with unprecedented sensitivity. This is due to the reduced production rate of cosmogenic isotopes at the SNOLAB depth and requires that the intrinsic background sources are low enough.

At scintillator purity levels similar to that of Borexino Phase I [20, 25], the unloaded scintillator phase of SNO+ provides excellent sensitivity to CNO, *pep*, and low energy ^{8}B neutrinos. With the scintillator sourced from a supply low in ^{14}C, SNO+ could also measure *pp* neutrinos with a sensitivity of a few percent. Due to the relatively high end-point of the spectrum, ^{8}B νs with energy above the ^{130}Te end-point can also be measured during the $0\nu\beta\beta$-decay phase.

The first measurement of the flux of neutrinos from the subdominant CNO fusion cycle would constrain the metallicity of the solar interior and thus provide critical input to the so-called solar metallicity problem: the

current disagreement between helioseismological observations of the speed of sound and model predictions, due to uncertainties in the heavy element (metal) content of the Sun. Historically, model predictions for the speed of sound were in excellent agreement with observation, one of the primary reasons for confidence in the Standard Solar Model during the period of uncertainty surrounding the solar neutrino problem. However, recent improvements in solar atmospheric modeling, including transitioning from one-dimensional to fully three-dimensional models, and including effects such as stratification and inhomogeneities [49], produced a lower value for the heavy element abundance of the photosphere and, thus, changed the prediction for the speed of sound. The theoretical prediction for the CNO flux depends linearly on the core metallicity and can be further constrained by a precision measurement of the ^{8}B flux, due to their similar dependence on environmental factors. A measurement of CNO neutrinos would thus resolve this uncertainty and also advance our understanding of heavier mass main-sequence stars, in which the CNO cycle dominates over the *pp* fusion chain.

Precision measurements of the *pep* flux and the low energy ^{8}B spectrum offer a unique opportunity to probe the interaction of neutrinos with matter and to search for new physics. The shape of the ν_e survival probability in the transition region between vacuum oscillation ($\leq 1\,\text{MeV}$) and matter-enhanced oscillation ($\geq 5\,\text{MeV}$) is particularly sensitive to new physics effects, such as flavor changing neutral currents or mass-varying neutrinos, due to the resonant nature of the MSW interaction. The *pep* neutrinos are a line source at 1.44 MeV, thus offering the potential for a direct disappearance measurement partway into this vacuum-matter transition region. However, due to their production region closer to the core of the Sun, the effect of new physics on the ^{8}B neutrino spectrum is significantly more pronounced. Thus, the most powerful search combines a precision measurement of the *pep* flux with a ^{8}B spectral measurement.

Borexino has published the first evidence for *pep* neutrinos [33], with a significance of just over 2σ from zero. In order to distinguish different models, a precision of at least 10% is required. A number of experiments have extracted the ^{8}B spectrum [42, 50, 51, 52, **?**], and there is some weak evidence for non-standard behaviour in the combined data set [54] but the significance is low (roughly 2σ). The theoretical uncertainty on *pep* neutrinos is very small, and well constrained by solar luminosity measurements. The ^{8}B flux is well measured by the SNO experiment [42]. Precise oscillation measurements are therefore possible.

Should the SNO+ scintillator be sourced from a supply naturally low in ^{14}C, similar to or within an order of magnitude or so of the level observed in Borexino, there also exists the potential for a precision measurement of *pp* neutrinos. Borexino has produced the first direct detection of these neutrinos, with a precision of a little over 10% [55]. A percent level measurement would allow a test of the so-called luminosity constraint, thus testing for additional energy loss or generation mechanisms in the Sun, and allowing us to monitor the Sun's output using neutrinos.

6.1. Backgrounds. The sensitivity of the SNO+ solar phase will depend critically on the leaching rate of ^{210}Bi. As described in Section 4.2, radon daughters, implanted on the internal AV surface, are expected to leach off during the various SNO+ phases with a rate that depends both on the temperature and on the liquid in contact with the vessel. We will be able to evaluate the levels of these backgrounds both during the initial water fill and during the scintillator fill itself. We are also investigating mitigation techniques to be applied in case the background levels are initially too high to perform the solar measurement. These techniques include *in–situ* recirculation, further purification, and the use of a balloon to shield from external backgrounds.

Other backgrounds for the measurements of *pep* and CNO neutrinos are the levels of ^{214}Bi (^{238}U chain), ^{212}Bi (^{232}Th chain), and ^{11}C in the pure scintillator. ^{238}U and ^{232}Th levels in the scintillator can be effectively constrained using the β-α delayed coincidence, as described in Section 4. ^{11}C decays, which were the main background for the measurement of *pep* neutrinos in Borexino [32], can be identified by a three-fold coincidence algorithm (see Section 4.5).

Another muon-induced isotope that is a potential background for low energy ^{8}B neutrino searches is ^{10}C ($T_{1/2} = 19.3\,\text{s}$, Q-value $= 3.65\,\text{MeV}$). However, due to the isotope's short half-life and the low cosmic muon rate at SNOLAB depth, it can be removed by cutting events that occur within a few minutes from each muon event.

6.2. Sensitivity. Sensitivity studies were performed assuming one year of unloaded scintillator data, which could be either prior to or following the Te-loaded phase. An extended maximum likelihood fit was performed in energy, with a conservative 50% fiducial volume, in order to reduce external background contributions to negligible levels. A two-dimensional fit would allow an increase in fiducial volume and thus improve sensitivity. Thirty-four signals were included in the fit: the four neutrino signals (^{8}B, ^{7}Be, CNO, and *pep*) as well as thirty background event types. Backgrounds expected to be in equilibrium were constrained to a single fit parameter; ^{210}Po, ^{210}Pb, and ^{210}Bi were treated independently, that is, not assumed to be in equilibrium with the parent decays. Background parameters included in the fit were the normalisations of: ^{7}Be, ^{39}Ar, ^{40}K, ^{85}Kr, ^{210}Po, ^{210}Pb, ^{14}C, ^{238}U chain, and ^{232}Th chain. ^{210}Bi was linked to CNO in the fit due to the similarity of the energy spectra; the sepa-

ration is best achieved by imposing an *ex–situ* constraint on the level of ^{210}Bi decays, or by using observables other than energy.

The nominal background levels assumed were those achieved by Borexino during their initial running. It was assumed that purification techniques (in particular, distillation) can reduce ^{7}Be contamination to negligible levels. Gaussian constraints were applied to backgrounds where an *ex–situ* or independent *in–situ* measurement of the rate is anticipated. α tagging is expected to reduce the ^{210}Po peak by 95%, with an uncertainty of 20% on the remaining 5% of the events. Coincidence decays provide a 50% constraint on ^{85}Kr, 25% on the ^{232}Th-chain backgrounds, and 7% on the portion of the ^{238}U chain that is treated as being in equilibrium.

The fit range was between 0.2 MeV and 6.5 MeV, with 10 keV bins in visible energy. Extending the fit to higher energies would improve the accuracy on the ^{8}B flux measurement. Bias and pull tests show that the fit is stable and accurate, and robust to changes in bin size or energy range (to within changes in statistics, e.g. ^{8}B flux accuracy is reduced if the energy range of the fit is reduced).

The simulations suggest that, with one year of data, the uncertainty on the *pep* flux will be less than 10%. The uncertainty on the linked CNO+^{210}Bi flux is 4.5%, into which we fold a conservative uncertainty for separating the two signals, resulting in a 15% predicted uncertainty on the CNO flux. The ^{7}Be flux can be measured to 4%, and ^{8}B to better than 8%. The uncertainty on the neutrino flux measurements is dominated by statistics, and by correlations between the neutrino signals themselves. A study of energy scale and resolution systematics shows that these parameters can be floated as nuisance parameters in the fit, and the data will constrain them to better than the required precision, with sub-percent level impact on the neutrino flux uncertainties. Calibration sources will be deployed in order to measure effects such as any non-Gaussianity of the resolution function, and any potential nonlinearity in the energy scale. In Figure 7 the full solar neutrino signals as detected by SNO+ are shown together with the main background sources for the LAB-PPO scintillator. A fiducial volume cut is applied at 5.5 m.

^{214}Bi-^{214}Po events are reduced by 95% using the β-α delayed coincidence as described in Section 4.1. A 95% rejection is applied to the ^{210}Po events and the remaining ^{214}Po events via alpha tagging. There is no rejection applied to the ^{212}Bi and ^{212}Po events. This is a conservative approach as we expect to reject the majority of these events using a β-α delayed coincidence as for the $0\nu\beta\beta$ search (see Section 5).

Studies show that the precision with which the *pp* neutrinos could be observed depends critically on the levels of backgrounds such as ^{14}C and ^{85}Kr in the scintillator. If these backgrounds are low, within 10–50 times that seen in Borexino, SNO+ could achieve a few-percent level measurement of the *pp* neutrino flux with just 6 months of solar neutrino data.

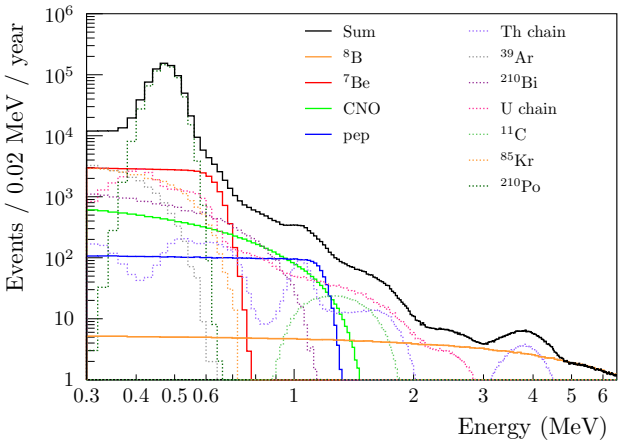

FIGURE 7: Expected solar neutrino fluxes as detected by SNO+ and the corresponding main backgrounds. Backgrounds levels are assumed to be equal to those initially achieved by Borexino [20, 25] (see text). Events are shown for the LAB-PPO scintillator, 400 Nhits/MeV light yield, and a fiducial volume cut of 5.5 m. A 95% reduction is applied to the ^{214}Bi-Po backgrounds via delayed coincidence tagging, and a 95% reduction on the ^{210}Po and the remaining ^{214}Po events via alpha tagging.

7. Antineutrino Studies

Antineutrino events in SNO+ will include geoneutrinos from the Earth's radioactive chains of uranium and thorium, antineutrinos from nuclear reactors, and the antineutrinos emitted by a supernova burst (which are considered in detail in Section 8). The measurement of geoneutrinos will constrain the radiogenic heat flow of the Earth for geophysics studies, while the measurement of reactor antineutrinos, with a known energy spectrum and a precise propagation distance, can better constrain the neutrino oscillation parameters [56].

7.1. Signal Detection. Antineutrinos are detected in SNO+ via inverse beta decay (IBD): $\overline{\nu}_e$s with energy greater than 1.8 MeV interact with the protons in the liquid scintillator, producing a positron and a neutron. The antineutrino energy is measured by the scintillation light emitted by the positron as it slows down and annihilates:

$$E_{\bar{\nu}_e} \simeq E_{prompt} + (M_n - M_p) - m_e \simeq E_{prompt} + 0.8\,\text{MeV} \quad (2)$$

where M_n, M_p, and m_e are the neutron, proton, and electron masses. The neutron emitted in the reaction will first thermalize and then be captured by hydrogen, leading to the characteristic 2.22 MeV delayed gamma from

the deuterium formation. The prompt+delayed signal allows the identification of the antineutrino event. The coincidence time interval is defined by the period elapsed from neutron emission to its capture, generally about 200 μs, while the spatial separation between the prompt and the delayed event depends on the distance travelled by the delayed gamma before scintillation light is emitted. The exact values to use for the time and distance coincidence tag, to identify the $\overline{\nu}_e$ events, depend crucially on the correct simulation of the neutron propagation in the scintillator mixture being used (unloaded or Te-loaded scintillator). Neutron propagation in each of the scintillator cocktails planned by SNO+ will be checked with a detailed calibration program using an AmBe source. This source, already extensively used by SNO, has a well-known neutron energy spectrum, extending to energies higher than those of the expected antineutrino signals. The calibration results will be cross-checked with a detailed Monte Carlo simulation.

7.2. Backgrounds. As the antineutrino signal is identified as a delayed coincidence in SNO+, the main backgrounds are true or random coincidences in the detector with the identified neutron capture. Most of the background neutrons are expected to come from external background sources and are therefore captured and reconstructed in the external regions of the detector. Events that reach the region inside the vessel can be mitigated by a fiducial volume cut, or by a radius-dependent analysis. The major source of neutrons inside the scintillator is the (α,n) reactions, which are mainly caused by ^{210}Po-alpha leached off the vessel surface and are expected to increase with time, as described in Section 4.6. The associated prompt signal, mainly due to the proton recoil, will be at energies lower than 3.5 MeV or, in case the product nucleus is in an excited state, in definite gamma peaks which will allow the study of the (α,n) background's time evolution.

7.3. Reactor Antineutrinos and Oscillations. In SNO+ we expect around 90 reactor antineutrino events per year. The total flux is obtained summing 3 components: (1) 40% of it comes from one reactor complex in Canada at a baseline of 240 km, (2) 20% is from two other complexes at baselines of around 350 km, and (3) 40% is divided between reactors in the USA and elsewhere at longer baselines. The signals from the first two sources (1 and 2) induce a very clear oscillation pattern (see Figure 8), which lead to a high sensitivity to the Δm^2_{12} neutrino oscillation parameter. For $E < 3.5$ MeV the geoneutrino signals and reactor signals overlap. Most of the backgrounds are concentrated in the energy region of the geoneutrinos. For a preliminary study of the reactor neutrino oscillation sensitivity, we conservatively exclude the region below 3.5 MeV. Assuming a light yield of 300 Nhits/MeV,

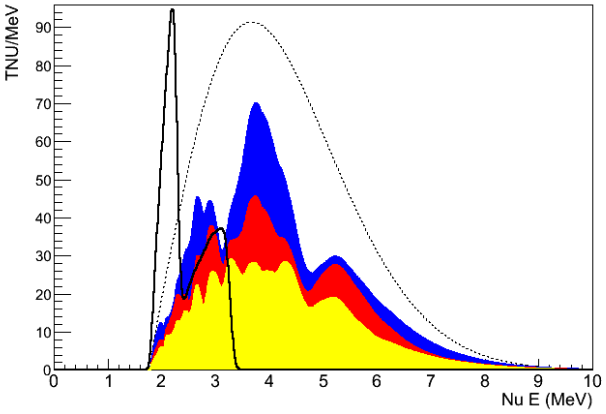

FIGURE 8: Expected visible antineutrino energy spectrum in SNO+, for 10^{32} proton-years per MeV. The nonoscillated reactor spectrum (dashed line) is shown together with the geoneutrino spectrum (solid line, arbitrary normalization). The stacked oscillated reactor spectrum is shown with different colors, each corresponding to a reactor complex: reactor at 240 km in blue (top), reactors at 350 km in red (middle), and other reactors in yellow (bottom). See text for details.

expected for the Te-loaded phase with perylene as secondary wavelength shifter, and a 5.5 m FV cut, we expect to reach a sensitivity in Δm^2_{12} of 0.2×10^{-5} eV^2, similar to the KamLAND result [56] in about 7 years of data taking. The full analysis will take the complete antineutrino spectrum into account, using constraints for the backgrounds, and measuring simultaneously the geoneutrino flux.

Generally, the $\overline{\nu}_e$ flux from the Canadian reactors (CANDU-type) is expected to be stable in time due to the continuous refuelling process. However, in the next few years there are expected upgrades in which different reactor cores will be turned off, with only one reactor core switched off at a time in each of the complexes. This will cause changes in the reactor spectrum, with an expected total flux reduction below 10% at each moment. This time evolution can be used to identify the very clear oscillation pattern in the reactor spectrum for each of the two identified baselines (240 km and 350 km) and to distinguish them from other antineutrino sources.

The oscillation patterns from the more distant reactors are less evident after they are combined. There is still a visible feature at antineutrino energies of 4.5 MeV from an accumulation of reactors at distances of the order of 550 km. A detailed description of the spectrum at this energy is still under discussion [57]. A preliminary study shows that the combined systematic uncertainties associated with the unoscillated spectrum description are below 5%. These uncertainties can be reduced using, for the distant reactors (source 3), direct measurements at

close-by detectors, like those of Daya Bay [57].

7.4. Geoneutrinos and Earth Studies. Interest in geoneutrinos has increased in the last few years with significant collaborations between neutrino physicists and geo-physicists. Joint results may finally explain the radiogenic heat flow of the Earth.

In SNO+ the geoneutrinos from the uranium and thorium chains can be detected. These antineutrinos come mainly from thick continental crust, with increases due to variations in local crust components [58].

The energy spectra of geoneutrinos are well-known for each of the standard decay chains [59]. The effect of neutrino oscillations is largely averaged out due to the long range in production distances, leading to a total survival probability of:

$$\langle P_{ee} \rangle = \cos^4\theta_{13} \cdot \left(1 - \frac{\sin^2(2\theta_{12})}{2}\right) + \sin^4\theta_{13} \simeq 0.547 \quad (3)$$

where $\theta_{13} = 9.1°$and $\theta_{12} = 33.6°$[42]. Detailed studies of the impact of the MSW effects on the energy spectrum are in progress.

As a first analysis step, we will fix the total U/Th ratio according to standard geological models [60], and fit for the total flux assuming a precise shape for the energy spectrum of geoneutrinos. The possible effect of local variations of this ratio is being quantified together with that from the low energy reactor spectrum. Systematic uncertainties in the energy scale and energy resolution and from the constraints on the alpha-n backgrounds will vary for each of the data taking phases. Overall, the SNO+ sensitivity to the total flux is expected to be dominated by statistical uncertainties. The accuracy will be close to that of Borexino for similar data-taking periods: the larger volume of the SNO+ detector compensates for the higher rate reactor background. We expect a similar rate of geoneutrinos and reactor antineutrinos in the 1.8 MeV–3.5 MeV energy region. However, the reactor spectrum extends up to much higher energies and contains features that can help in establishing the oscillation parameters. The time evolution analysis will also help to separate the reactor background (Section 7.3). In the Te-loaded phase the low energy backgrounds are expected to be about 50–150 times higher than in the pure scintillator phase, which can make the extraction of the geoneutrino signal more difficult.

We aim to additionally separate both the uranium and thorium contributions and the mantle and crust contributions in a global analysis of the geoneutrino spectrum including data from KamLAND [61] and Borexino [62].

Table 5: Supernova neutrino interaction channels in LAB-PPO. The event rates, per 780 tonnes of material, assume the incoming neutrino time-integrated flux described in the text. No flavor changing mechanisms are considered. The uncertainties on the event rates only include the cross section uncertainties [16].

Reaction	Number of Events
NC: $\nu + p \to \nu + p$	429.1 ± 12.0 [a]
CC: $\bar{\nu}_e + p \to n + e^+$	194.7 ± 1.0
CC: $\bar{\nu}_e$ + ^{12}C $\to$ ^{12}B$_{g.s.}$ + e$^+$	7.0 ± 0.7
CC: ν_e + ^{12}C $\to$ ^{12}N$_{g.s.}$ + e$^-$	2.7 ± 0.3
NC: ν + ^{12}C $\to$ ^{12}C*(15.1 MeV) + ν'	43.8 ± 8.7
CC/NC: ν + ^{12}C $\to$ ^{11}C or ^{11}B + X	2.4 ± 0.5
ν–electron elastic scattering	13.1 [b]

[a] 118.9 ± 3.4 above a trigger threshold of 0.2 MeV visible energy.
[b] The Standard Model cross section uncertainty is $< 1\%$.

8. Supernova Neutrino Observation

The era of neutrino astronomy commenced with the observation of 24 events, all associated with the inverse beta decay of $\bar{\nu}_e$, from the collapse of supernova SN 1987A at $\sim$ 50 kpc [63]. SNO+, with its large high purity liquid scintillator volume and the deep location underground, is one of the most promising experiments for the detection of neutrinos from core collapse supernovae (CCSNe), offering a rich sample of detection channels, low backgrounds, and a large number of target particles and nuclei. CCSNe are an exceptional source of neutrinos of all flavors and types, and a measurement is expected to shed light on the explosion mechanism. The shape of the individual supernova (SN) ν_α ($\nu_\alpha = \nu_e$, $\bar{\nu}_e$, ν_x, where in this context ν_x is the sum of ν_μ, $\overline{\nu}_\mu$, ν_τ and $\overline{\nu}_\tau$) energy spectra is expected to approximate a thermal spectrum [64] in the absence of neutrino flavor changing mechanisms. At postbounce times t < 1 s, before shock revival, the flavor changes are expected to be reduced to those induced by the well–known MSW effect in a quasi–static environment [65, 66]. At later times, many further effects interfere, significantly modifying the spectral shape. These effects are nontrivial and still lack a full understanding and a consistent analytical treatment. At present, sensitivity studies to thermal spectral parameters are only meaningful for at most the first second of the burst. It is estimated that half of all neutrinos are emitted in this time span [67].

8.1. Signal Detection in SNO+. For the detection potential of SNO+ presented in this paper, we assume that the distance from the SN to Earth is $d = 10$ kpc – known from, for example, the detection of the electromagnetic

radiation released in the SN event, and that 3×10^{53} erg of binding energy (ϵ_ν) are released in the form of neutrinos, equally partitioned amongst all six flavors and types. The mean energies used are 12 MeV for ν_e, 15 MeV for $\bar{\nu}_e$ and 18 MeV for ν_x [68], which are generic mean SN neutrino energies [69] consistent with the findings from SN 1987A.

The possible SN neutrino interaction channels during the SNO+ pure scintillator phase are listed in Table 5 together with the expected event rates. Several events due to $\bar{\nu}_e$ are expected, because of the comparatively large cross section for the IBD reaction [70]. This process, seen during SN 1987A, is the only interaction of SN neutrinos observed to date. Additionally, SNO+ can measure the flux of ν_x and ν_e. As the mean neutrino energy is below about 30 MeV, ν_es and $\overline{\nu}_e$s will be detected mainly by the charged current (CC) interactions, while supernova ν_xs can only be detected by the more challenging neutral current (NC) reactions. One NC reaction is neutrino–proton elastic scattering (ES), $\nu + p \to \nu + p$ [71], which is the only channel that provides spectral information about the ν_xs. The total cross section of this process [72] is about a factor of three smaller than the cross section of IBD; however, the reaction is possible for all six neutrino types yielding a similar number of events for a detector threshold above ∼0.2 MeV.

8.2. SNO+ Sensitivity to the ν_x Spectral Shape. In the preliminary estimations of the SNO+ sensitivity to ν_x spectral shape through ν–p ES we conservatively assume a spatial radius cut of 5 m and a 0.2 MeV threshold, corresponding to a minimal neutrino energy of $E_\nu^{\min} \approx 21.9$ MeV. This is close to the threshold we expect to use for events that will be permanently stored. We are currently discussing other settings for the trigger thresholds to avoid any loss of potential low energy supernova events.

In Figure 9 the reconstructed energy spectrum of all neutrinos emitted in the first second of the SN (ν_e, $\bar{\nu}_e$ and ν_x) and detected in SNO+ via the ν–p ES reaction is shown together with the true neutrino spectrum. The reconstructed energy spectrum is obtained from the detected proton energy unfolded using the TUnfold algorithm [73], on the basis of binned data. The strongly nonlinear quenching of the proton energy, which shifts most of the scattering events below ∼ 0.5 MeV electron–equivalent energy, and the finite detector resolution are taken into account. The number of events in the lowest bin is slightly overestimated, due to bin–to–bin migrations caused by the finite energy resolution. The statistical and total systematic uncertainties are also shown. A fit to the ν_x spectrum is only possible if the ν_e and $\bar{\nu}_e$ spectra are measured independently. SNO+ is sensitive to the spectral shape of $\overline{\nu}_e$s via the IBD reaction, while in the case of ν_es it has to be assumed that an independent detector, with, for example, a Pb target like HALO [74] or a LAr target [70], provides the necessary spectral information.

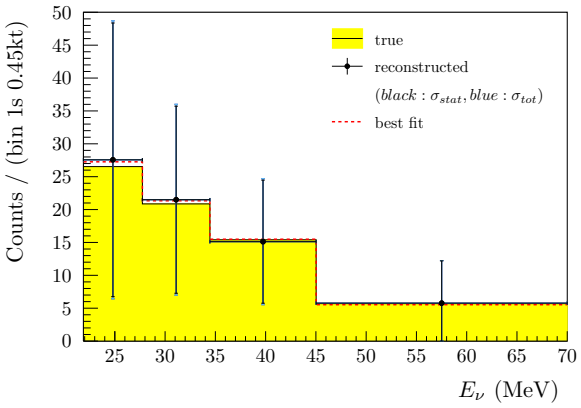

FIGURE 9: True, reconstructed, and best fit SN neutrino energy distribution of the ν–p ES detection channel within the FV and above the detector threshold [16]. Shown are the sum of the ν_e, $\bar{\nu}_e$ and ν_x spectra, considering their time-integrated flux in the first second of the reference SN. The statistical uncertainties are shown in black, while the total uncertainties are shown in blue. The contribution from systematic uncertainty is too small to be resolved.

The resulting best fit E_ν spectrum is also shown in Figure 9 and is in excellent agreement. The systematic uncertainties propagated within the fit are the ν-p ES cross section, the number of target protons, N_p, the ionization quenching parameter, the spectral ν_e and $\bar{\nu}_e$ parameters, and the energy resolution of the detector. The corresponding best fit values are $\langle E_{\nu_x} \rangle = 17.8^{+3.5}_{-3.0}(\text{stat.})^{+0.2}_{-0.8}(\text{syst.})$ MeV and $\varepsilon_{\nu_x} = (102.5^{+82.3}_{-42.2}(\text{stat.})^{+16.2}_{-13.0}(\text{syst.})) \times 10^{51}$ erg [16], while the respective expectation values are 18 MeV and 100×10^{51} erg.

8.3. SNEWS. SNO+ is preparing to participate in the inter–experiment Supernova Early Warning System (SNEWS) [75], which has the goal to provide a fast and reliable alert using the coincident observation of burst signals in several operating detectors. As neutrinos escape from the SN tens of minutes up to several hours before the first photons, their detection offers the possibility of alerting the astronomical community to the appearance of the next SN light signal.

9. Exotic Physics Searches

Due to its location deep underground, which significantly reduces the cosmogenic background, and the high radiopurity of the materials used, SNO+ has a unique sensitivity to search for exotic physics, including certain modes of nucleon decay and axion or axion-like particle searches.

TABLE 6: Expected backgrounds in the 5.4–9 MeV energy region during six months of water fill. A fiducial volume cut of 5.5 m is applied to all events. The events after the $\cos\theta_{sun} > -0.8$ cut are also shown. ϵ(n) and ϵ(p) are the neutron and proton decay-mode detection efficiencies in the 5.5 m FV and energy window.

Decay source	Events in six months	
		$\cos\theta_{sun} > -0.8$ Cut
^{214}Bi	0	0
^{208}Tl	0.6	0.6
Solar-neutrinos	86.4	17.7
Reactor antineutrinos	1.5	1.3
External ^{214}Bi-^{208}Tl	9.2	8.9
Total	97.7	28.5
ϵ(n)	0.1089	0.1017
ϵ(p)	0.1264	0.1129

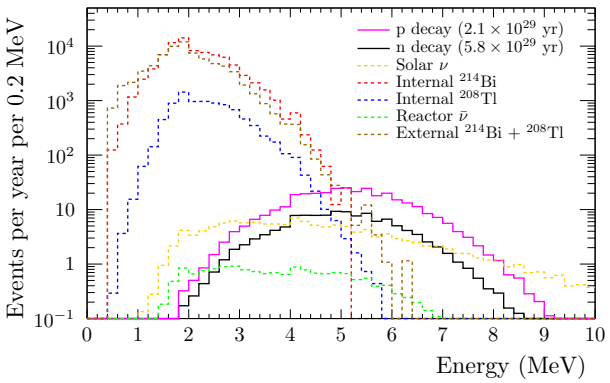

FIGURE 10: Expected energy spectrum for the water phase backgrounds. The signal from invisible proton [76] and neutron [78] decay is also shown. A fiducial radius cut of 5.5 m and a cut on $\cos\theta_{sun} > -0.8$ are applied.

9.1. Invisible Nucleon Decay. Nucleon decay modes to a final state undetected by the experiment, for example, n$\rightarrow 3\nu$, can be searched for by detecting the decay products of the remaining unstable nucleus as it deexcites. This process has been previously investigated by some experiments such as SNO [76] by searching for the decay of ^{16}O nuclei, and Borexino [77] and KamLAND [78] by looking for the decay of ^{12}C nuclei. We plan to search for the invisible nucleon decay of ^{16}O during the initial water phase of the experiment. In the case of a decaying neutron, the resulting ^{15}O will deexcite emitting a 6.18 MeV gamma 44% of the time. For a decaying proton, the nucleus is left as ^{15}N which in 41% of the decays de-excites emitting a 6.32 MeV gamma [79]. Both these signals will be in a favorable region of the SNO+ energy spectrum (5.4 MeV–9 MeV) in which few backgrounds are expected. These are: (1) internal and external ^{208}Tl and ^{214}Bi decays, (2) solar neutrinos, and (3) reactor and atmospheric antineutrinos. The expected contribution of each background in the 5.4–9 MeV energy region, in six months of running, is shown in Table 6. The targeted purity for the SNO+ internal water is the average of the SNO collaboration's H_2O and D_2O levels (see Table 2). The purity can be measured *in situ* using events below 5 MeV and cross checked using water assays. Solar neutrino events can be reduced by placing a cut on the direction of the event, which is reconstructed using the topology of the detected Cherenkov light. Reactor antineutrino events can be tagged using a delayed coincidence. The background due to atmospheric neutrinos is expected to be small based on SNO data [76].

The events in Table 6 are given for a fiducial volume cut of 5.5 m, which helps in reducing the external backgrounds. An additional cut at $\cos\theta_{sun} > -0.8$ relative to the solar direction further reduces the dominant solar background, removing ~80% of the events with a sacrifice of ~10% on the signals and the isotropic backgrounds. Figure 10 shows the energy spectrum of the water phase backgrounds: solar neutrinos, reactor antineutrinos, and radioactive decays from the uranium and thorium chains, after the two cuts are applied. It also shows the shapes based on the current best limits of the signal gammas from invisible proton [76] and neutron [78] decay.

Using a Poisson method [80] we can set the lower limit, at 90% C.L., on the invisible nucleon decay lifetime τ by:

$$\tau > \frac{N_{nucleons} \times \epsilon \times f_T}{S_{90\%}} \quad (4)$$

where $N_{nucleons} = 2.4 \times 10^{32}$, ϵ is the efficiency of detecting the decay in the signal window from Table 6, $S_{90\%}$ is the expected signal events at 90% C.L., and f_T is the live-time of 0.5 years. Assuming we reach the expected background, a limit of $\tau_n > 1.25 \times 10^{30}$ and $\tau_p > 1.38 \times 10^{30}$ years for the decay of neutrons and protons, respectively, can be set. This is an improvement over the existing limit set by KamLAND, $\tau > 5.8 \times 10^{29}$ years, by a factor of ~2 with just six months of running time. A likelihood approach is in development which is expected to provide a further improvement on the limit.

9.2. Axion-Like Particle Search. An axion-like particle (ALP) is defined as a neutral pseudoscalar particle that exists as an extension to the QCD Lagrangian [81].

A possible reaction channel for ALP production in the Sun is p+d$\rightarrow^3$He+A, where A is the ALP with an energy of 5.5 MeV [82]. In SNO+ the couplings of ALPs to electrons, g_{Ae}, photons, $g_{A\gamma}$, and nucleons, g_{AN}, can be observed mainly through Compton conversion (A+$e^- \rightarrow e^- + \gamma$) and the axioelectric effect (A + e^-+ $Z_X \rightarrow e^-$ + Z_X, with Z_X the charge of the involved nucleus X). In

both cases, for low ALP masses the signature is monoenergetic at ∼5 MeV electromagnetic energy deposition.

Different strategies for different phases of SNO+ are used for the detection of ALPs. In the water phase, the most likely interaction is the Compton conversion, which produces a Cherenkov ring with topology similar to that of ^{8}B-neutrinos. The main background events are very similar to those described for the invisible nucleon decay search (see Section 9.1) as the two signals have similar energies. However, since the Compton conversion has a strong directional bias, we expect to remove a significant amount of isotropic backgrounds, leaving ^{8}B-neutrinos as the dominant one. With 6 months of water data, due to the deeper location and larger fiducial volume, we expect to approach the current limit set by Borexino [82].

The BGO collaboration proposed a separate limit on the ALP couplings without having to assume axions interact via Compton conversion [83]. In this case, the detection of solar ALPs via the axioelectric effect, which depends on the nucleus charge as Z_X^5, could be particularly interesting during the Te-loaded phase. Due to the significantly large tellurium mass, SNO+ has the possibility of improving the limit on the axion-electron coupling constant set by the BGO collaboration by several orders of magnitude.

10. Conclusions

In this paper the broad physics program of the SNO+ experiment is presented. Three main data taking phases are planned: one with the detector filled with ultra-pure water, one with unloaded liquid scintillator, and one with 2.34 tonnes of tellurium loaded into the detector.

The primary physics goal of SNO+ is a sensitive search for $0\nu\beta\beta$-decay of ^{130}Te. We expect to set a lower limit on the half-life of this process of $T_{1/2}^{0\nu\beta\beta} > 9{\times}10^{25}$ yr (90% CL) in 5 years of data taking. This limit corresponds to an effective Majorana mass ranging from 55 to 133 meV, at the top of the inverted neutrino mass hierarchy. The possibility of loading 10 times more tellurium in order to cover the majority of the inverted hierarchy region is under investigation.

Along with the $0\nu\beta\beta$-decay search, SNO+ also has the potential to measure the low energy solar neutrinos, like *pep*-neutrinos. If the same purity levels as initially achieved by Borexino are reached, SNO+ can measure the *pep*-neutrinos with an uncertainty less than 10% in one year of data taking with pure liquid scintillator. Additionally, if the background is low enough SNO+ can measure CNO neutrinos.

Another physics topic that can be explored by SNO+ is the measurement of geoneutrinos in a geologically interesting location, which will be complementary to the measurements done by Borexino and KamLAND. Furthermore, SNO+ can measure reactor antineutrinos, which will help in reducing the uncertainty on the oscillation parameters.

With its depth and low background, SNO+ has an extraordinary opportunity to measure the supernova ν_x energy spectrum for the first time. This measurement provides valuable information in order to probe and constrain supernova dynamics. Participation in SNEWS will further support a reliable early warning to the astronomical community in the event of a nearby supernova.

During the water fill, SNO+ can search for exotic physics and set competitive limits in the invisible nucleon decay of ^{16}O.

We expect to start operation with the water fill phase soon, followed by the liquid scintillator fill phase after a few months of data taking. The Te-loaded phase is foreseen in 2017.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgements

Capital construction funds for the SNO+ experiment is provided by the Canada Foundation for Innovation (CFI). This work has been in part supported by the Science and Technology Facilities Council (STFC) of the United Kingdom (Grants no. ST/J001007/1 and ST/K001329/1), the Natural Sciences and Engineering Research Council of Canada, the Canadian Institute for Advanced Research (CIFAR), the National Science Foundation, national funds from Portugal and European Union FEDER funds through the COMPETE program, through FCT – Fundação para a Ciência e a Tecnologia (Grant no. EXPL/FIS-NUC/1557/2013), the Deutsche Forschungsgemeinschaft (Grant no. ZU123/5), the European Union's Seventh Framework Programme (FP7/2007-2013, under the European Research Council (ERC) grant agreement no. 278310 and the Marie Curie grant agreement no: PIEF-GA-2009-253701), the Director, Office of Science, of the U.S. Department of Energy (Contract no. DE-AC02-05CH11231), the U.S. Department of Energy, Office of Science, Office of Nuclear Physics (Award Number de-sc0010407), the U.S. Department of Energy (Contract No. DE-AC02-98CH10886), the National Science Foundation (Grant no. NSF-PHY-1242509) and the University of California, Berkeley. The authors acknowledge the generous support of the Vale and SNOLAB staff.

References

[1] K. Zuber, "Double-beta decay", *Contemporary Physics*, vol. 45, no. 6, pp. 491–502, 2004.

[2] A. M. Serenelli, W. C. Haxton, and C. Peña-Garay, "Solar Models with Accretion. I. Application to the Solar Abundance Problem", *The Astrophysical Journal*, vol. 743, no. 1, Article ID 24, 2011.

[3] A. Friedland, C. Lunardini and C. Peña-Garay, "Solar neutrinos as probes of neutrino-matter interactions", *Physics Letters B*, vol. 594, no. 3–4, pp. 347–354 , 2004.

[4] W. C. Haxton and A. M. Serenelli, "CN-cycle solar neutrinos and the Sun's primordial core metallicity", *The Astrophysical Journal*, vol. 687, no. 1, pp. 678–691, 2008.

[5] M. C. Chen, "The SNO Liquid Scintillator Project", *Nuclear Physics B - Proceedings Supplements*, vol. 145, pp. 65–68, 2005.

[6] SNO Collaboration, B. Aharmim et al., "Measurement of the cosmic ray and neutrino-induced muon flux at the Sudbury Neutrino Observatory", *Physical Review D*, vol. 80, no. 1, Article ID 012001, 2009.

[7] SNO Collaboration, J. Boger et al., "The Sudbury Neutrino Observatory", *Nuclear Instruments and Methods in Physics Research A*, vol. 449, no. 1–2, pp. 172–207, 2000.

[8] N. Jelley, A. B. McDonald, and R.G. H. Robertson, "The Sudbury Neutrino Observatory", *Annual Review of Nuclear and Particle Science*, vol. 59, pp. 431–465, 2009.

[9] P. G. Jones, "Background rejection for the neutrinoless double-beta decay experiment SNO+", *Ph.D. Thesis*, Lincoln College, Oxford, UK, 2011, `http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.559770`.

[10] M. Yeh et al, "A new water-based liquid scintillator and potential applications", *Nuclear Instruments and Methods in Physics Research A*, vol. 660, no. 1, pp. 51–56, 2011.

[11] H. M. O'Keeffe, E. O'Sullivan, and M. C. Chen, "Scintillation decay time and pulse shape discrimination in oxygenated and deoxygenated solutions of linear alkylbenzene for the SNO+ experiment", *Nuclear Instruments and Methods in Physics Research A*, vol. 640, no. 1, pp. 119-122, 2011.

[12] S. Grullon, "Light yield and scintillation decay time constants of Te-loaded liquid scintillator for the SNO+ experiment", *Neutrino 2014*, Boston University , 1–7 June 2014.

[13] H. Wan Chan Tseung, J. Kaspar, and N. Tolich, "Measurement of the dependence of the light yields of linear alkylbenzene-based and EJ-301 scintillators on electron energy", *Nuclear Instruments and Methods in Physics Research A*, vol. 654, no. 1, pp. 318–323, 2011.

[14] H. Wan Chan Tseung and N. Tolich, "Ellipsometric measurements of the refractive indices of linear alkylbenzene and EJ-301 scintillators from 210 to 1000 nm", *Physica Scripta*, vol, 84, no. 3, Article ID 035701, 2011.

[15] B. von Krosigk, L. Neumann, R. Nolte, S. Röttger, and K. Zuber, "Measurement of the proton light response of various LAB based scintillators and its implication for supernova neutrino detection via neutrino-proton scattering", *The European Physical Journal C*, vol. 73, Article ID 2390, 2013.

[16] B.von Krosigk, "Measurement of proton and α–particle quenching in LAB based scintillators and determination of spectral sensitivities to supernova neutrinos in the SNO+ detector", *Ph.D. Thesis*, Technische Universität Dresden, Germany, 2015.

[17] J. B. Birks, "The theory and practice of scintillation counting", *Pergamon, New York, USA*, 1964.

[18] R. Ford, M. Chen, O. Chkvorets, D. Hallman, and E. Vázquez-Jáuregui, "SNO+ scintillator purification and assay", *AIP Conference Proceedings*, vol. 1338, pp. 183–194, 2011.

[19] R. Ford for the SNO+ Collaboration, "A scintillator purification plant and fluid handling system for SNO+", *AIP Conference Proceedings*, vol. 1672, Article ID 080003, 2015.

[20] Borexino Collaboration, C. Arpesella et al., "Direct measurement of the ^{7}Be solar neutrino flux with 192 days of Borexino data", *Physical Review Letters*, vol. 101, no. 9, Article ID 091302, 2008.

[21] KamLAND Collaboration, K. Eguchi et al., "First results from KamLAND: Evidence for reactor antineutrino disappearance", *Physical Review Letters*, vol. 90, no. 2, Article ID 021802, 2003.

[22] S. Hans, et al., "Purification of telluric acid for SNO+ neutrinoless double-beta decay search", *Nuclear Instruments and Methods in Physics Research A*, vol. 795, pp. 132–139, 2015.

[23] R. Alves et al., "The calibration system for the photomultiplier array of the SNO+ experiment", *Journal of Instrumentation*, vol. 10, Article ID P03002, 2015.

[24] National Nuclear Data Center, "Nuclear structure & decay Data", `http://www.nndc.bnl.gov/nudat2/`

[25] Borexino Collaboration, G. Alimonti et al., "The liquid handling systems for the Borexino solar neutrino detector", *Nuclear Instruments and Methods in Physics Research A*, vol. 609, no. 1, pp. 58–78, 2009.

[26] F. Alessandria, et al., "CUORE crystal validation runs: Results on radioactive contamination and extrapolation to CUORE background", *Astroparticle Physics*, vol. 35, no. 12, pp. 839–849, 2012.

[27] T.W. Armstrong, K.C. Chandler, and J. Barish, "Calculations of neutron flux spectra induced in the Earth's atmosphere by galactic cosmic rays", *Journal of Geophysical Research*, vol. 78, no. 16, pp. 2715–2726, 1973.

[28] N. Gehrels, "Instrumental background in balloon-borne gamma-ray spectrometers and techniques for its reduction", *Nuclear Instruments and Methods in Physics Research A*, vol. 239, no. 2, pp. 324–349, 1985.

[29] Borexino Collaboration, G. Alimonti et al., "Measurement of the ^{14}C abundance in a low-background liquid scintillator", *Phisics Letters B*, vol. 422, no. 1–4, pp. 349–358, 1998.

[30] KamLAND Collaboration, S. Abe et al., "Production of radioactive isotopes through cosmic muon spallation in KamLAND", *Physical Review C*, vol. 81, no. 2, Article ID 025807, 2010.

[31] Borexino Collaboration, G. Bellini et al., "Precision measurement of the ^{7}Be solar neutrino interaction rate in Borexino", *Physical Review Letters*, vol. 107, no. 14, Article ID 141302, 2011.

[32] C. Galbiati, A. Pocar, D. Franco, A. Ianni, L. Cadonati, and S. Schönert, "Cosmogenic ^{11}C production and sensitivity of organic scintillator detectors to pep and CNO neutrinos", *Physical Review C*, vol. 71, no. 5, Article ID 055805, 2005.

[33] Borexino Collaboration, G. Bellini et al., "First evidence of *pep* solar neutrinos by direct detection in Borexino" , *Physical Review Letters*, vol. 108, no. 5, Article ID 051302, 2012.

[34] V. Lozza, and J. Petzoldt, "Cosmogenic activation of a natural tellurium target", *Astroparticle Physics*, vol. 61, pp. 62–71, 2015.

[35] E. Arushanova, "Pileup background rejection in SNO+ experiment", Talk at IoP meeting, Manchester, April 2015.

[36] E. Arushanova and A. R. Back, "Probing neutrinoless double beta decay with SNO+", arXiv:1505.00247 [physics.ins-det], 2015.

[37] Ge-detector measurement performed at SNOLAB: I. Lawson and B. Cleveland, "Low background counting at SNOLAB", *AIP Conference Proceedings*, vol. 1338, pp. 68, 2011.

[38] SNO Collaboration, B. Aharmim et al., "Electron energy spectra, fluxes, and day-night asymmetries of ^{8}B solar neutrinos from measurements with NaCl dissolved in the heavy-water detector at the Sudbury Neutrino Observatory", *Physical Review C*, vol. 72, no. 5, Article ID 055502, 2005.

[39] N. J. T. Smith, "Facility and science developments at SNOLAB", Talk at ASPERA Workshop: The next generation projects in Deep Underground Laboratories, Zaragoza, Spain, June 2011, `http://indico.cern.ch/event/130734/contribution/21/material/slides/0.pdf`

[40] M. Redshaw, B. J. Mount, E. G. Myers, and F. T. Avignone III, "Masses of ^{130}Te and ^{130}Xe and double–β–decay Q value of ^{130}Te", *Physical Review Letters*, vol. 102, no. 21, Article ID 212502, 2009.

[41] NEMO–3 Collaboration, R. Arnold et al., "Measurement of the $\beta\beta$ decay half-life of ^{130}Te with the NEMO-3 Detector", *Physical Review Letters*, vol. 107, no. 6, Article ID 062504, 2011.

[42] SNO Collaboration, B. Aharmim et al., "Combined analysis of all three phases of solar neutrino data from the Sudbury Neutrino Observatory", *Physical Review C*, vol. 88, no. 2, Article ID 025501, 2013.

[43] J. Barea, J. Kotila, and F. Iachello, "Nuclear matrix elements for double-β decay", *Physical Review C*, vol. 87, no. 1, Article ID 014315, 2013.

[44] J. Kotila and F. Iachello, "Phase-space factors for double-β decay", *Physical Review C*, vol. 85, no. 3, Article ID 034316, 2012.

[45] F. Šimkovic, V. Rodin, A. Faessler, and P. Vogel, "$0\nu\beta\beta$-decay and $2\nu\beta\beta$-decay nuclear matrix elements, quasiparticle random-phase approximation, and isospin symmetry restoration", *Physical Review C*, vol. 87, no. 4, Article ID 045501, 2013.

[46] J. Menéndez, A. Poves, E. Caurier, and F. Nowacki, "Disassembling the nuclear matrix elements of the neutrinoless $\beta\beta$ decay", *Nuclear Physics A*, vol. 818, no. 3–4, pp. 139–151, 2009.

[47] J. Hyvärinen and J. Suhonen, "Nuclear matrix elements for $0\nu\beta\beta$ decays with light or heavy Majorana-neutrino exchange", *Physical Review C*, vol. 91, no. 2, Article ID 024613, 2015.

[48] T. R. Rodríguez and G. Martínez-Pinedo, "Energy density functional study of nuclear matrix elements for neutrinoless $\beta\beta$ decay", *Physical Review Letters*, vol. 105, no. 25, Article ID 252503, 2010.

[49] M. Asplund, N. Grevesse, A. J. Sauval, and P. Scott, "The chemical composition of the Sun", *Annual Review of Astronomy and Astrophysics*, vol. 47, pp. 481–522, 2009.

[50] KamLAND Collaboration, S. Abe et al., "Measurement of the ^{8}B solar neutrino flux with the KamLAND liquid scintillator detector", *Physical Review C*, vol. 84, no. 3, Article ID 035804, 2011.

[51] Borexino Collaboration, G. Bellini et al., "Measurement of the solar ^{8}B neutrino rate with a liquid scintillator target and 3 MeV energy threshold in the Borexino detector", *Physical Review D*, vol. 82, no. 3, Article ID 033006, 2010.

[52] M. Smy, Talk at XXIV international conference on neutrino physics and astrophysics, Neutrino 2012, in Proceedings of the 25th International Conference on Neutrino Physics and Astrophysics (Neutrino '12), Kyoto, Japan, June 2012.

[53] A. Renshaw, "Solar neutrino results from Super-Kamiokande", arXiv:1403.4575 [hep-ex], 2014.

[54] R. Bonventre, A. LaTorre, J. R. Klein, G. D. Orebi Gann, S. Seibert, and O. Wasalski, "Nonstandard models, solar neutrinos, and large θ_{13}", *Physical Review D*, vol. 88, no. 5, Article ID 053010, 2013.

[55] Borexino Collaboration, G. Bellini et al., "Neutrinos from the primary proton-proton fusion process in the Sun", *Nature*, vol. 512, pp. 383–386, 2014.

[56] KamLAND Collaboration, A. Gando et al., "Constraints on θ_{13} from a three-flavor oscillation analysis of reactor antineutrinos at KamLAND", *Physical Review D*, vol. 83, no.5, Article ID 052002, 2011.

[57] Daya Bay Collaboration, F. P. An et al., "Spectral measurement of electron antineutrino oscillation amplitude and frequency at Daya Bay", *Physical Review Letters*, vol. 112, no. 6, Article ID 061801, 2014.

[58] H.K.C. Perry, J.-C. Mareschal, and C. Jaupart, "Enhanced crustal geo-neutrino production near the Sudbury Neutrino Observatory, Ontario, Canada", *Earth and Planetary Science Letters*, vol. 288, pp. 301–308, 2009.

[59] S. Enomoto, "Using Neutrinos to study the Earth: Geo-Neutrinos", *NeuTel 2009 Conference*, Venice, Italy, March 2009.

[60] A. M. Dziewonski and D. L. Anderson, "Preliminary reference Earth model", *Physics of the Earth and Planetary Interiors*, vol. 25, no. 4, pp. 297–356, 1981.

[61] KamLAND Collaboration, A. Gando et al., "Partial radiogenic heat model for Earth revealed by geoneutrino measurements", *Nature Geoscience*, vol. 4, pp. 647–651, 2011.

[62] Borexino Collaboration, G. Bellini et al., "Measurement of geo-neutrinos from 1353 days of Borexino", *Physics Letters B*, vol. 722, no. 4–5, pp. 295–300, 2013

[63] I. V. Krivosheina, "SN 1987A Historical view about registration of the neutrino signal with BAKSAN, KAMIOKANDE II and IMB detectors", *International Journal of Modern Physics D*, vol. 13, no. 10, Article ID 2085, 2004, and references therein.

[64] M. T. Keil, G. G. Raffelt, and H. -T. Janka, "Monte Carlo study of supernova neutrino spectra formation", *The Astrophysics Journal*, vol. 590, no. 2, pp. 971–991, 2003.

[65] X. Jing, H. Ming-Yang, H. Li-Jun, G. Xin-Heng, and Y. Bing-Lin, "Detection of Supernova neutrinos on the Earth for large θ_{13}", *Communication in Theoretical Physics*, vol, 61, no. 2, pp. 226–234, 2014, and references therein.

[66] S. Sarikas, G. G. Raffelt, L. Hüdepohl, and H.-T. Janka, "Suppression of self-induced flavor conversion in the Supernova accretion phase", *Physical Review Letters*, vol. 108, no. 6, Article ID 061101, 2012.

[67] G. Pagliaroli, F. Vissani, M. L. Costantini, and A. Ianni, "Improved analysis of SN1987A antineutrino events", *Astroparticle Physics*, vol. 31, no. 3, pp. 163–176, 2009.

[68] B. Dasgupta and J. F. Beacom, "Reconstruction of supernova ν_μ, ν_τ, anti-ν_μ, and anti-ν_τ neutrino spectra at scintillator detectors", *Physical Review D*, vol. 86, no. 11, Article ID 113006, 2011.

[69] T. Lund and J. P. Kneller, "Combining collective, MSW, and turbulence effects in supernova neutrino flavor evolution", *Physical Review D*, vol. 88, no. 2, Article ID 023008, 2013.

[70] K. Scholberg, "Supernova Neutrino Detection", *Annual Review of Nuclear and Particle Science*, vol. 62, pp. 81–103, 2012.

[71] J. F. Beacom, W. M. Farr, and P. Vogel, "Detection of supernova neutrinos by neutrino proton elastic scattering", *Physical Review D*, vol. 66, no. 3, Article ID 033001, 2002.

[72] L. A. Ahrens, et al., "Measurement of neutrino - proton and antineutrino - proton elastic scattering", *Physical Review D*, vol. 35, no. 3, pp. 785–809, 1987.

[73] S. Schmitt, "TUnfold: an algorithm for correcting migration effects in high energy physics", *Journal of Instrumentation*, vol. 7, Article ID T10003, 2012.

[74] D. Väänänen and C. Volpe, "The neutrino signal at HALO: learning about the primary supernova neutrino fluxes and neutrino properties", *Journal of Cosmology and Astroparticle Physics*, vol. 2011, Article ID 019, 2011.

[75] P. Antonioli et al., "SNEWS: The Supernova Early Warning System", *New Journal of Physics*, vol. 6, Article ID 114, 2004.

[76] SNO Collaboration, S. N. Ahmed et al., "Constraints on nucleon decay via invisible modes from the Sudbury Neutrino Observatory", *Physical Review Letters*, vol. 92, no. 10, Article ID 102004, 2004.

[77] Borexino Collaboration, H.O. Back et al., "New limits on nucleon decays into invisible channels with the

BOREXINO counting test facility", *Physics Letters B*, vol. 563, no. 1–2, pp. 23–34, 2003.

[78] KamLAND Collaboration, T. Araki et al., "Search for the invisible decay of neutrons with KamLAND", *Physical Review Letters*, vol. 96, no. 10, Article ID 101802, 2006.

[79] H.Ejiri, "Nuclear deexcitations of nucleon holes associated with nucleon decays in nuclei", *Physical Review C*, vol. 48, no. 3, pp. 1442–1444, 1993.

[80] O. Helene, "Upper limit of peak area", *Nuclear Instruments and Methods in Physics Research*, vol. 212, no. 1–3, pp. 319–322, 1983.

[81] P. W. Graham, et al., "Experimental searches for the axion and axion-like particles', *Annual Review of Nuclear and Particle Science*, vol. 65, pp. 485–514, 2015.

[82] Borexino Collaboration, G. Bellini et al.,"Search for solar axions produced in p(d,^{3}He)A reaction with Borexino detector", *Physical Review D*, vol. 85, no. 9, Article ID 092003, 2012.

[83] A. V. Derbin et al., "Search for axioelectric effect of solar axions using BGO scintillating bolometer", *The European Physical Journal C*, vol. 74, Article ID 3035, 2014

Constraints on Gravitino Decay and the Scale of Inflation using CMB spectral distortions

Emanuela Dimastrogiovanni[a], Lawrence M. Krauss[a,b] and Jens Chluba[c]
[a]*Department of Physics and School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85827, USA*
[b]*Research School of Astronomy and Astrophysics, Mt. Stromlo Observatory, Australian National University, Canberra, Australia 2611*
[c]*Institute for Astronomy, K30, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*
(Dated: February 4, 2016)

If local supersymmetry is the correct extension of the standard model of particle physics, then following Inflation the early universe would have been populated by gravitinos produced from scatterings in the hot plasma during reheating. Their abundance is directly related to the magnitude of the reheating temperature. The gravitino lifetime is fixed as a function of its mass, and for gravitinos with lifetimes longer than the age of the universe at redshift $z \simeq 2 \times 10^6$ (or roughly 6×10^6s), decay products can produce spectral distortion of the cosmic microwave background. Currently available COBE/FIRAS limits on spectral distortion can, in certain cases, already be competitive with respect to cosmological constraints from primordial nucleosynthesis for some gravitino decay scenarios. We show how the sensitivity limits on μ and y distortions that can be reached with current technology would improve constraints and possibly rule out a significant portion of the parameter space for gravitino masses and Inflation reheating temperatures.

I. INTRODUCTION

The cosmic microwave background (CMB) temperature and polarization anisotropies represent an invaluable source of information about the origin and evolution of the Universe. They are and have been, for the past few decades, one of the main targets of investigation for cosmology [1]. The CMB, however, presents us with an additional and independent cosmological probe: its energy/frequency spectrum. The frequency spectrum is compatible with a blackbody distribution with an average temperature of 2.726 K [2]. Deviations from a blackbody distribution are potentially generated by any physical process that entails an exchange of energy between matter and radiation [3] or the modification of the CMB photon number [4]. As such, spectral distortions allow to constrain mechanisms that are within the standard framework of cosmology (including, for instance, recombination [5], reionization and structure formation [6], Silk damping of small-scale fluctuations [7]) as well as more exotic possibilities, including ones inherent to beyond-the-standard model particle physics (see [8] for some examples).

At redshifts $z > 2 \times 10^6$, any produced distortion is quickly erased: double Compton emission, Bremsstrahlung and Compton scattering are efficient enough to immediately restore thermal equilibrium in the primordial plasma. At lower redshifts, $2 \times 10^6 \gtrsim z \gtrsim 5 \times 10^4$, Compton scattering between photons and electrons is still very rapid whereas double Compton and Bremsstrahlung are no longer efficient. As a result, a distortion is predominantly produced in the form of a non-vanishing chemical potential (μ distortion) at high frequencies. Moving down to $z \lesssim 5 \times 10^4$, Compton scattering also becomes inefficient at restoring kinetic equilibrium and a *y-type* distortion is created. The latter can be pictured as a high-z version of the Sunyaev-Zeldovish (SZ) effect in galaxy clusters [9]. During the transition between μ and y eras, additional (*r-type*) spectral distortions are produced that cannot be described as a superposition of μ and y distortions [10]. The r-type distortion is crucial if one wants to constrain the time-dependence of phenomena generating spectral distortion around $z \approx 10^4 - 10^5$ [11].

Current observational bounds on spectral distortions date back to the COBE/FIRAS measurements: these placed upper bounds $|\mu| \lesssim 9 \times 10^{-5}$ and $|y| \lesssim 1.5 \times 10^{-5}$ [12]. Modern technology could yield an improvement of more than three orders of magnitude in sensitivity [13], a threshold that would allow one to place meaningful bounds on a vast ensemble of processes of relevance for astrophysics and cosmology [14].

Spectral distortions from particle annihilations or decays at $z \lesssim 2 \times 10^6$ can be used to place constraints on their masses, abundance and interactions [15]. One such particle, the gravitino, is of particular interest and for this reason we focus on gravitinos in this paper.

Gravitinos are spin 3/2 superpartners of the graviton, predicted in the context of supergravity theories (see e.g. [17] for a review). They are expected to acquire a mass ($m_{3/2}$) via the super-Higgs mechanism. Because of their gravitational strength interactions one might expect that, if inflation [19] occurred, any initial abundance of gravitinos would be diluted by the expansion. Nevertheless, gravitinos can be produced after the end of inflation: thermal production from interactions in the hot plasma during reheating as well as non-thermal effects related to the rapid oscillations of the inflaton, can efficiently replenish the gravitino population [20]. Because their interactions are fixed to be of gravitational strength, their post-

Inflation abundance is fixed by the reheating scale alone [16].

Severe constraints on gravitino abundances arise from cosmology [18]. Gravitinos (or their decay products) surviving until the present time, for instance, would overclose the Universe if their mass density were to exceed the critical density. For unstable gravitinos, decays occurring after big-bang nucleosynthesis (BBN) may ruin the successful predictions of BBN. Decays of gravitinos into neutrinos may affect the effective number ($N_{\rm eff}$) of relativistic degrees of freedom (see e.g. [21]), which is also constrained by large scale structure and CMB polarization observations [22]. Because gravitinos lifetime is determined in terms of their mass, sufficiently light gravitinos will decay after $z \approx 10^6$. If gravitino decays occur during the μ or the y distortion eras and result into transfer of energy into the photon plasma, they will produce a distortion of the CMB spectrum, in addition to affecting BBN. Both BBN and spectral distortion bounds would result in exclusion regions in the $(T_{\rm rh}, m_{3/2})$ plane. Constraining gravitino abundances in this way can therefore put important new constraints on both the scale of supersymmetry breaking and the scale of Inflation in supersymmetric scenarios.

In this paper, we employ current CMB spectral distortion bounds and sensitivity limits for future measurements such as the ones proposed with a PIXIE-like experiment to place upper limits on the reheating temperature, in connection with supersymmetry and the thermal production of unstable gravitinos. This is especially important given that the reheating temperature is otherwise poorly constrained. Reheating only leaves indirect imprints on cosmological observables, which are often dependent on inflationary and reheating model-dependent uncertainties (as e.g. in [23]). The only model-independent bounds on $T_{\rm rh}$ arise from the requirement that reheating should precede BBN [24] ($T_{\rm rh} > 1$ MeV). Current bounds on the energy scale of inflation indicate that $T_{\rm rh} \lesssim O(10^{16})$ GeV [25], leaving a very large unconstrained parameter space in general.

This paper is organized as follows: in Sec. II we briefly review gravitinos thermal production during reheating along with the implications of an unstable gravitino for BBN; in Sec. III we compute the effects of gravitino decay on the CMB frequency spectrum; in Sec. IV we offer our conclusions and propose possible future improvements.

II. RELIC GRAVITINOS FROM REHEATING

In local supersymmetric theories (supergravity), when SUSY is spontaneously broken, the gravitino acquires a mass by absorbing the Goldstino (Goldstone fermion associated with the broken symmetry). The gravitino mass and the scale (F) of SUSY breaking are related by $F \approx \sqrt{m_{3/2}\, M_{\rm P}}$, where $M_{\rm P} \approx 2.4 \times 10^{18}$ GeV is the Planck mass.

During reheating, interactions in the hot plasma lead to gravitino production. The relic density is given by [26]

$$n_{3/2} = Y_{3/2}\, s(T)\,, \qquad Y_{3/2} \approx 10^{-12}\, \frac{T_{\rm rh}}{10^{10}\,{\rm GeV}}\,. \tag{1}$$

Here $s(T) \equiv (2\pi^2/45) g_*(T) T^3$ is the entropy density of the plasma, with g_* the total number of relativistic degrees of freedom. The numerical coefficient in $Y_{3/2}$ can vary depending on the specific value of the cross sections for production processes, but typically results in variations at most of a few in the overall value of the gravitino number density after inflation.

The gravitino decay rate is fixed as a function of the mass and of the effective number of decay channels ($N_{\rm dec}$) [27]:

$$\Gamma_{3/2} = \frac{N_{\rm dec}}{(2\pi)}\, \frac{m_{3/2}^3}{M_{\rm P}^2}\,. \tag{2}$$

The beginning of the μ distortion era ($z \simeq 2 \times 10^6$) is subsequent to the time frame of BBN (ranging from an initial temperature of 1 MeV down to 10 keV). The transfer of energy from decaying gravitinos into the CMB photons is most efficient if the decay products are energetic photons [36] or charged particles. As a result, for unstable gravitinos whose decays are relevant to spectral distortion, one also expects important effects on BBN [28]. Predictions of BBN theory for the current abundances of the light elements (mainly D, T, ^{3}He, ^{4}He) involve a main parameter, the baryon-to-photon ratio ($\eta_{\rm B}$). Agreement between theory and the observed abundances calls for $\eta_{\rm B} \approx 3 \times 10^{-10}$. The outcome of BBN may be entirely different in the presence of gravitinos (or other relic particles decaying after $T \approx 1$ MeV). Specifically, there are three main consequences on BBN: (i) the presence of massive gravitinos may affect the expansion rate, leading to an overproduction of ^{4}He; (ii) radiative decays of gravitinos may lead to a suppression of $\eta_{\rm B}$; (iii) energetic decay products (such as photons or charged particles) may destroy the light elements. The third class of processes has been shown to represent the dominant effect on BBN for relic particles in the range of masses that we will be concerned with in this work.

In the next section we will compute the CMB distortion from gravitino decay and present our results along with some constraints from the literature on (iii).

III. IMPLICATIONS FOR SPECTRAL DISTORTION

The first step is to compute the rate of energy release from gravitino decay

$$\frac{{\rm d}E_{\rm dec}}{{\rm dt}} = \epsilon_{3/2} m_{3/2} \frac{1}{a(t)^3} \frac{d}{dt} \left[a(t)^3 N_{3/2}(t) \right]\,. \tag{3}$$

Here $\epsilon_{3/2}$ is a dimensionless parameter quantifying the fraction of energy from the decay products that contributes to heating of the CMB photon bath via Comptonization, $m_{3/2}$ is the gravitino mass, a the scale factor and $N_{3/2}$ is the number density of gravitinos. Following the parametrization in [29], Eq. (3) can be rewritten as

$$\left\| \frac{{\rm d}E_{\rm dec}}{{\rm dt}} \right\| = f_{3/2}\, N_{\rm H}\, \Gamma_{3/2}\, e^{-\Gamma_{3/2} t}\,, \tag{4}$$

where the quantity $f_{3/2}$ collects all of the information about the decaying particle (its mass/lifetime as well as its abun-

dance) and about the decay process (e.g. the number of channels or the number of effective relativistic degrees of freedom produced from the decay). In Eq. (4), $N_{\rm H}$ is the number density of hydrogen nuclei and $\Gamma_{3/2}$ the decay rate of gravitinos. It can be shown that $f_{3/2}$ (or $f_{3/2}/z_{3/2}$, where $z_{3/2}$ indicates the redshift at which the decay occurs) factors out of the integrals in the definitions of the effective μ and y distortion parameters, thus proving a very convenient choice for parametrizing spectral distortion signals from decaying particles. With the above definitions, and using Eqs. (1) and (2), one arrives at

$$\frac{f_{3/2}}{z_{3/2}} \simeq \epsilon_{3/2}\,\frac{10^{-6}}{N_{\rm dec}^{1/2}}\left(\frac{T_{\rm rh}}{\rm GeV}\right)\left(\frac{m_{3/2}}{\rm GeV}\right)^{-1/2}{\rm eV}\,, \tag{5}$$

where we used $N_{\rm H} \approx 1.9\times10^{-7}\,(1+z)^3\,{\rm cm}^{-3}$ for the number density of hydrogen atoms. The relation between gravitino mass and lifetime is $t_{3/2} \simeq (2.4\times10^{13}/N_{\rm dec})\,(m_{3/2}/{\rm GeV})^{-3}$ s. We use a conventional time-temperature relation, $t \approx [\sqrt{45}\,M_{\rm P}]/[\sqrt{2\pi^2\,g_*(T)}T^2]$, and time-redshift relation, $z \approx 4.9\times10^9/\sqrt{t/{\rm s}}$ throughout.

The r-distortion appearing at intermediate redshifts ($10^4 \lesssim z \lesssim$ few $\times\,10^5$) normally requires a numerical treatment [10]. In this paper, we wish to retain analytic control over our calculations. We will therefore focus on μ and y distortions, for which simple analytic approximations have been found, leaving the study of any intermediate-type distortion for future work. It is also well known that, at smaller redshifts, another contribution to the y-distortion arises, namely that due to the inverse Compton scattering of CMB photons off free electrons (thermal Sunyaev-Zeldovich, tSZ, effect) [6, 30]. Thus, limits derived from the y-parameter should be interpreted as conservative upper limits.

We write the fractional variation of the energy density of CMB photons as the sum of μ and y distortion contributions

$$\frac{\Delta\rho_\gamma}{\rho_\gamma} \approx \left[\frac{\Delta\rho_\gamma}{\rho_\gamma}\right]_\mu + \left[\frac{\Delta\rho_\gamma}{\rho_\gamma}\right]_y\,, \tag{6}$$

where the effective distortion parameters are given by [31]

$$\left[\frac{\Delta\rho_\gamma}{\rho_\gamma}\right]_\mu \equiv \frac{\mu}{1.401}\,, \qquad \left[\frac{\Delta\rho_\gamma}{\rho_\gamma}\right]_y \equiv 4\,y\,. \tag{7}$$

For pure μ and y distortion one has

$$\mu \approx 1.4\int \mathcal{J}_{\rm bb}\,\mathcal{J}_\mu\frac{1}{\rho_\gamma}\left(\frac{{\rm d}E}{{\rm d}t}\right)dt\,,\;\; y \approx \frac{1}{4}\int \mathcal{J}_{\rm bb}\,\mathcal{J}_y\frac{1}{\rho_\gamma}\left(\frac{{\rm d}E}{{\rm d}t}\right)dt\,. \tag{8}$$

where the thermal response of the medium to the energy injection has been parametrized with the *visibility functions* [32]

$$\mathcal{J}_{\rm bb}(z) \approx \exp\left[-(z/z_\mu)^{5/2}\right]\,, \tag{9}$$

$$\mathcal{J}_y(z) \approx \left[1+\left(\frac{1+z}{6.0\times10^4}\right)^{2.58}\right]^{-1}\,,\;\; \mathcal{J}_\mu(z) \approx 1-\mathcal{J}_y(z)\,.$$

Here $\mathcal{J}_{\rm bb}$ accounts for the fact that thermal equilibration processes are highly efficient at $z_\mu \approx 2\times10^6$. The definition of $\mathcal{J}_\mu$ follows from enforcing energy conservation according to

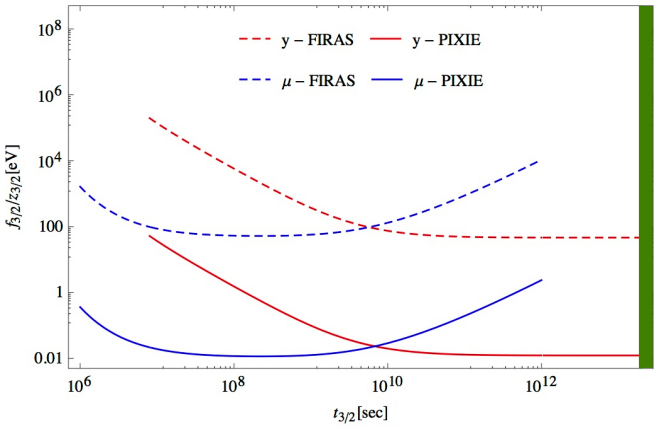

FIG. 1: Limits placed by FIRAS (dashed lines) and sensitivity projections for PIXIE (solid lines) from y and μ distortions on the effective energy input per hydrogen atom normalized to the redshift at decay, $f_{3/2}/z_{3/2}$, and the lifetime, $t_{3/2}$. The range of values of $t_{3/2}$ encompasses the whole μ-distortion era and the y distortion-era until around recombination (green band). The areas above the dashed lines have been excluded by FIRAS.

Eq. (6) and $\mathcal{J}_y$ was found to approximate the branching of energy eventually appearing as y-distortion [32]. In Eq. (8), we set the lower bound in redshift in the integral for y-distortion to $z \approx 1000$. The y-distortion is in principle produced down to $z \approx 200$ (at $z < 200$ the rate of baryons-photons interactions becomes too low for thermodynamic equilibrium to be maintained). However, at $z < 10^3$ a more detailed treatment of the energy exchange between matter and radiation may be required as the plasma recombines.

From the observed limits on μ and y from COBE/FIRAS and from the forecast sensitivity of an experiment like PIXIE ($|\mu| \lesssim 2\times10^{-8}$ and $|y| \lesssim 4\times10^{-9}$), one obtains the bounds in Fig. 1 in the $(f_{3/2}/z_{3/2}\,, t_{3/2})$ plane. Red lines are from y, blue lines from μ distortions. Dashed lines denote the lower contours of the exclusion regions defined by FIRAS observations. Solid lines show the sensitivity limits of PIXIE.

The joint bounds on $T_{\rm rh}$ and $m_{3/2}$ are derived from the ones on $(f_{3/2}/z_{3/2},\, t_{3/2})$

$$\frac{T_{\rm rh}}{\rm GeV} \simeq \left(\frac{f_{3/2}}{z_{3/2}\,{\rm eV}}\right)\frac{10^6\,N_{\rm dec}^{1/2}}{\epsilon_{3/2}}\left(\frac{m_{3/2}}{\rm GeV}\right)^{1/2}\,, \tag{10}$$

$$\frac{m_{3/2}}{\rm GeV} \simeq \left(\frac{2.4\times10^{13}}{N_{\rm dec}}\right)^{1/3}\left(\frac{t_{3/2}}{\rm s}\right)^{-1/3}\,. \tag{11}$$

For given values of the gravitino mass, $N_{\rm dec}$ and $\epsilon_{3/2}$, the temperature given in Eq. (10) represents: (i) the maximum reheating temperature currently allowed by FIRAS y and μ distortion bounds, for $f_{3/2}/z_{3/2}$ given by the FIRAS lines in Fig. 1; (ii) the smallest value for the upper bound that an experiment like PIXIE would be able to place on the reheating temperature, for $f_{3/2}/z_{3/2}$ given by the PIXIE lines in Fig. 1. Notice that, as one approaches $\epsilon_{3/2} \to 0$ (limit of no energy transfer to the photons), $T_{\rm rh}$ in Eq. (10) becomes larger, i.e. the constraint from spectral distortion weakens. The other extreme is

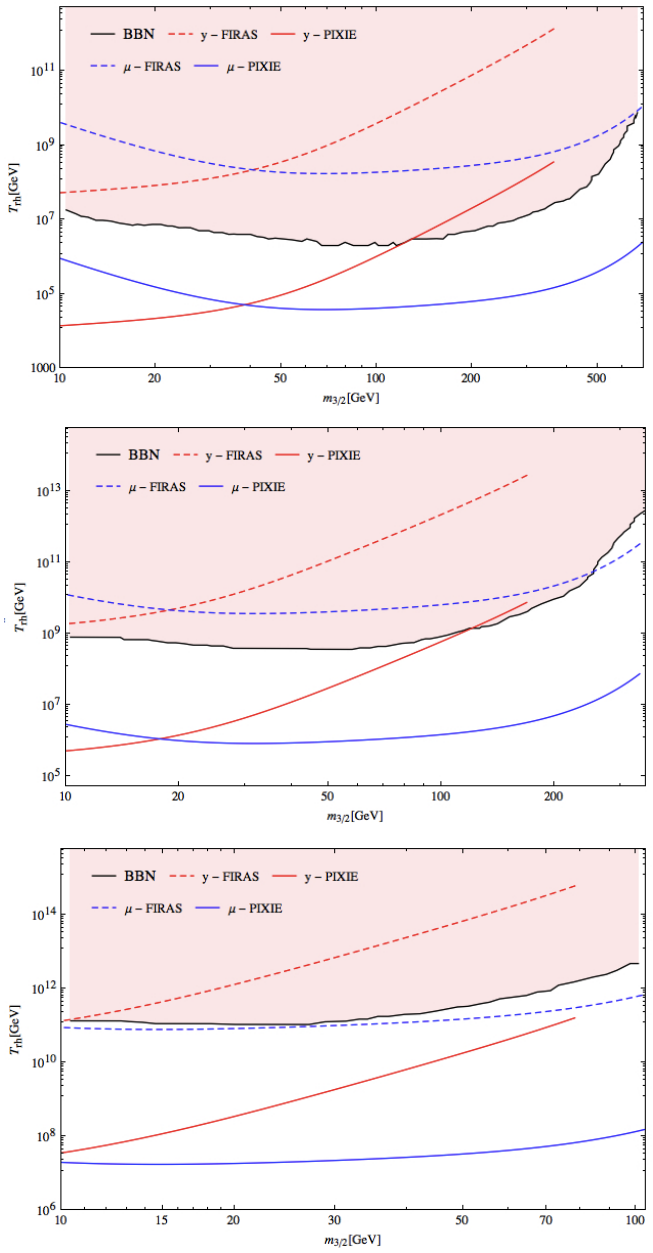

FIG. 2: Collection of FIRAS and BBN exclusion regions in the reheating temperature - gravitino mass plane, along with PIXIE sensitivity limits (solid lines). The shaded area is ruled out by BBN [33], dashed lines define the lower boundaries of the areas ruled out by FIRAS. Color codes are as in Fig. 1 for y and μ distortions. The energy release scenario is that of a decay of a gravitino into a photon + photino. The three figures correspond to decreasing values of the branching ratio for this process: $B_{[\mathrm{G}\to\gamma+\tilde{\gamma}]} = 1,\ 0.1,\ 0.01$, respectively moving from the figure at the top to the one at the bottom.

$\epsilon_{3/2} \to 1$, which provides the most stringent bounds that can be obtained for a given process.

Gravitinos decay through a variety of channels. Several studies have been proposed that quantify the effects of the decay on the BBN predictions for light elements primordial abundances [28]. We will refer to the studies carried out in

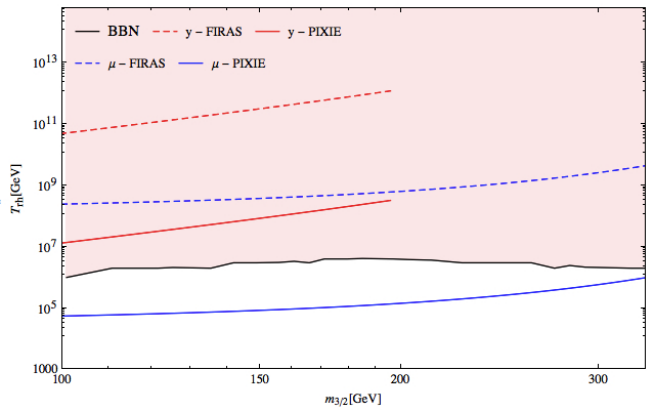

FIG. 3: Limits from spectral distortion (blue and red lines) and from BBN (black line, reproducing results in [34]) for gravitinos decaying entirely into hadrons.

[33] and [34] and include some of their results in our plots, alongside our spectral distortion sensitivity lines, so as to visualize the different cosmological constraints simultaneously.

Unsurprisingly, both in the context of BBN and of spectral distortion the most severely constrained scenarios involve direct decays into charged particles and/or into photons.

We will first consider the case of a gravitino decaying into photon + photino. If the photino mass is $m_{\tilde{\gamma}} \ll m_{3/2}$, the total decay rate reads

$$\Gamma_{3/2} \simeq \frac{m_{3/2}^3}{32\,\pi\,M_{\mathrm{P}}^2}\,. \tag{12}$$

Our results for this case are shown in the first plot of Fig. 2. Here we set $\epsilon_{3/2} = 1/2$: the fraction of gravitino initial energy effectively converted into photons is the one ultimately responsible for CMB distortion.

The effect of such a decay on BBN are well-studied. Photons emitted during the decay initiate an electromagnetic cascade. They can scatter off background photons transferring energy to the latter or producing, for example, electron-positron pairs. Photons can also interact with matter, scattering with background electrons or producing pair creation in the presence of nuclei.

We consider the results obtained in [33]; here the spectrum of high energy photons and electrons was computed and from it photodissociation effects on the light elements were quantified. The BBN lines shown in Fig. 2 arise primarily from D and ^{3}He abundances: the shaded region corresponds to an overproduction of these elements and is therefore excluded. Notice that the bounds from BBN are at least one order of magnitude stronger than the FIRAS limits in most of the mass range reported in the figure, with spectral distortion limits approaching the ones from BBN only near the extreme ends of the range, i.e. around $m_{3/2} \approx 10\,\mathrm{GeV}$ and towards $m_{3/2} \approx 700\,\mathrm{GeV}$. The latter value corresponds to gravitinos decaying at the onset of the μ era. In this plot (the same will apply to the remaining plots of Fig. 2), in order to draw a comparison with the limits from BBN studies in the literature, we considered $m_{3/2} \approx 10\,\mathrm{GeV}$ as the lowest value of our mass

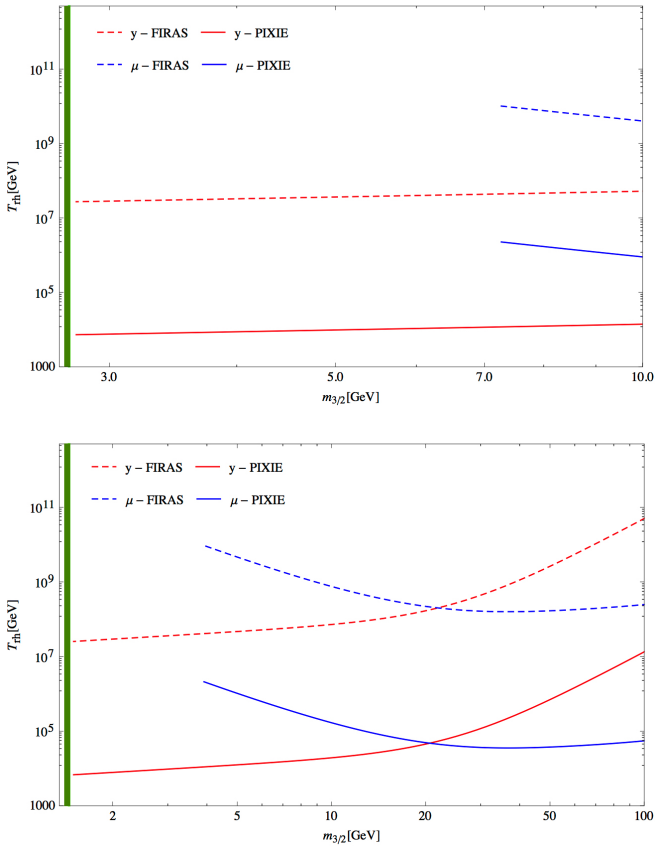

FIG. 4: FIRAS exclusion regions (dashed lines) and PIXIE sensitivity limits (solid lines) for gravitinos decaying before recombination (marked by the green band). Color codes are as in Fig. 1 for y and μ distortions. The energy release scenario is that of a decay of a gravitino into photon + photino (upper panel) and of a decay into hadrons (lower panel), both with unitary branching ratios.

range. However, our spectral distortion bounds also cover the $O(1-10)$ GeV range, which we show in Fig. 4 (upper panel).

Our plot shows that an experiment like PIXIE has the potential to provide highly competitive constraints on the reheating temperature: for masses $10\,\mathrm{GeV} \lesssim m_{3/2} \lesssim 100\,\mathrm{GeV}$, $T_{\rm rh}^{\rm max} \approx 10^6\,\mathrm{GeV}$ from BBN, whereas PIXIE may be able to probe reheating temperatures as low as $6\times10^3\,\mathrm{GeV}$.

In supergravity models where R-parity is preserved, the lightest supersymmetric particle (in our case the photino) is stable. One then should also require that the energy density of relic photinos does not overcome the critical density today. This results in an additional bound on the reheating temperature [33]: $T_{\rm rh} \lesssim 10^{11}\,(m_{\tilde\gamma}/100\,\mathrm{GeV})^{-1}h^2\,\mathrm{GeV}$, where h is the Hubble rate in units of $100\,(\mathrm{km/sec})/\mathrm{Mpc}$. The photino mass is strictly model-dependent. If $m_{3/2}$ is viewed as an upper bounds for $m_{\tilde\gamma}$ (condition for Eq. (12) to apply), then in the above range for the gravitino mass, the bound derived on the reheating temperature from the relic density of photinos is much weaker than both BBN and spectral distortion bounds.

If gravitinos only partially decay into photons and photinos, as will in general be the case, one may describe this by introducing an additional parameter, the branching ratio $B_{[\mathrm{G}\to\gamma+\tilde\gamma]} \equiv \Gamma_{[\mathrm{G}\to\gamma+\tilde\gamma]}/\Gamma_{\rm total}$. A value $B_{[\mathrm{G}\to\gamma+\tilde\gamma]} = 1$ would then correspond to the results just discussed and represented in the upper panel of Fig. 2. BBN bounds can be derived for different values of $B_{[\mathrm{G}\to\gamma+\tilde\gamma]}$ in [33]. Being blind to the effect of the remaining decay channels, some of which will likely have cosmological implications, the bounds derived with this procedure will therefore be conservative. For the sake of comparison between spectral distortion and BBN constraints, we adopt this simplified approach here. For consistency we assume that the ratio of the initial gravitino energy that is transferred to the CMB bath is simply reduced by a factor equal to the branching ratio w.r.t. the case where gravitinos entirely decay into photons and photino, i.e. $\epsilon_{3/2} = B_{[\mathrm{G}\to\gamma+\tilde\gamma]}/2$.

Our results are represented in the second and third panels of Fig. 2, respectively for $B_{[\mathrm{G}\to\gamma+\tilde\gamma]} = 0.1$ and $B_{[\mathrm{G}\to\gamma+\tilde\gamma]} = 0.01$. Notice that for these more realistic scenarios, the μ distortion bounds from FIRAS are now comparable to or slightly stronger than BBN bounds for the largest plotted values of the gravitino mass in the $B_{[\mathrm{G}\to\gamma+\tilde\gamma]} = 0.1$ case, and in the whole mass range for $B_{[\mathrm{G}\to\gamma+\tilde\gamma]} = 0.01$. As for $B_{[\mathrm{G}\to\gamma+\tilde\gamma]} = 1$, our plots show that also for these smaller values of the branching ratio PIXIE would be able to rule out a substantial portion of the currently allowed parameter space. One could gain access to temperature values of $T_{\rm rh}$ down to $10^5 - 10^6$ GeV, as opposed to the $T_{\rm rh} \gtrsim 10^9 - 10^{11}$ GeV range one can probe with BBN and current spectral distortion bounds for gravitino masses $10\,\mathrm{GeV} \lesssim m_{3/2} \lesssim 100 - 300\,\mathrm{GeV}$.

We can also consider exclusively hadronic decay channels for gravitinos, where one finds

$$\Gamma_{3/2} \approx \frac{m_{3/2}^3}{5\,\pi\,M_{\rm P}^2}\,. \tag{13}$$

We refer to the nucleosynthesis bounds obtained for this case in [34], which were derived for a gravitino with mass $m_{3/2} \gtrsim 100\,\mathrm{GeV}$. The heaviest gravitinos that μ distortion can constrain in this case have masses $m_{3/2} \lesssim 300\,\mathrm{GeV}$. In Fig. 3 we present our spectral distortion results (for which we set $\epsilon_{3/2} \approx 1$) and also display the primordial nucleosynthesis bounds from [34] (mostly due to ^{3}He/D and to ^{6}Li/H measurements) in the overlapping mass range. The μ distortion bounds from an experiment like PIXIE would be more stringent than nucleosynthesis bounds in most of the mass range, nearing BBN for the heaviest mass values. In Fig. 4 (lower panel) we present the FIRAS exclusion regions and the PIXIE sensitivity bounds for the range $O(1-100)\,\mathrm{GeV}$ of gravitino masses.

Our results in Eqs. (5), (10) and (11) are completely general and therefore applicable to any scenario for gravitino decay, simply by varying $N_{\rm dec}$ and $\epsilon_{3/2}$. We should also mention that the spectral distortion constraints are derived assuming that only a small fraction of all the energy transferred to the medium by the decay products is absorbed by light elements, and also ignoring a variety of other possible energy injection mechanisms beyond decay into photons. A more complete analysis should simultaneously follow the effective fraction of energy used up by destroying light elements, and also more complete energy injection cascades. The former effect might

at best slightly weaken our bounds. The latter is likely to strengthen them. We leave such an analysis to a future paper.

IV. CONCLUSIONS AND OUTLOOK

Cosmological datasets offer various routes to uncovering beyond-the-standard model particle physics. Supersymmetry is a promising candidate for extending the standard model and searching for SUSY-induced effects is one of the main goals in collider experiments. If SUSY is the correct description of nature, it must broken at some energy scale. We do not know what the SUSY breaking mechanism is, nor the scale at which it occurs. Local supersymmetry (supergravity) theories predict the existence of the gravitino, spin $3/2$ superpartner of the graviton. The gravitino mass is related to the SUSY breaking scale and its interactions are fixed in a nearly model-independent way. Being able to obtain stringent constraints on gravitinos is then invaluable for testing supergravity.

Moreover, in the early Universe, gravitinos will be generated thermally from interactions in the thermal bath during reheating following Inflation. In this case, their abundance would be a function of the reheating temperature. If gravitinos decay after $z \approx 2 \times 10^6$, they may produce observable distortions of the cosmic microwave background frequency spectrum. Thus constraints on gravitino decays can provide important constraints on the scale of Inflation.

In this paper we have analyzed how, with current technology, spectral distortions in the CMB constrain a sizable region of the reheating temperature - gravitino mass parameter space. We have derived general analytic expressions for computing μ and y distortions from gravitino decays occurring between the beginning of the μ distortion era and recombination. Our results are expressed in terms of a few parameters describing the number and types of decay channels.

We have plotted the exclusion regions in ($T_{\rm rh}$, $m_{3/2}$) space for COBE/FIRAS along with the sensitivity limits of a PIXIE-like experiment for various simplified assumptions regarding gravitino decay, considering energy injection purely by direct photon or hadron decay products. We show that, when compared with the bounds from primordial nucleosynthesis, a PIXIE-like experiment would be able to constrain a much larger region of parameter space and that the bounds from FIRAS can be competitive or exceed those derived from BBN considerations (see Fig. 2 and Fig. 3). We find that a PIXIE-like experiment will be able to constrain inflationary reheating temperatures as low as 6×10^3 GeV, which will cover most of the allowed parameter range for inflation, and therefore interestingly might allow a detection of SUSY-related Inflation, rather than a simple constraint on models.

Our study can be extended in several ways. It would be interesting to also include r-distortion, by performing a numerical analysis of the thermal response of CMB photons to gravitino decay. We also limited our study to gravitinos decaying before recombination. Numerical tools have been developed for extending our work to include lighter gravitino masses corresponding to later decays. These are predicted in a large number of supersymmetric models and are far from being ruled out by other cosmological probes. Finally, more stringent bounds could likely be derived with calculations of full charged particle cascades beyond simple initial photon or hadron decay products. As our analysis suggests, such improvements are worth considering, given the great reach of CMB spectral distortions for constraining gravitino physics and the physics of Inflation.

Acknowledgments

E.D. is grateful to Yue Zhao for fruitful discussions. J.C. is supported by the Royal Society as a Royal Society University Research Fellow at the University of Cambridge, UK. E.D. and L.M.K. acknowledge support from the DOE under grant No. de-sc0008016.

[1] C. L. Bennett *et al.*, Astrophys. J. **464**, L1 (1996) [astro-ph/9601067]; G. Hinshaw *et al.* [WMAP Collaboration], Astrophys. J. Suppl. **208**, 19 (2013) [arXiv:1212.5226 [astro-ph.CO]]; R. Adam *et al.* [Planck Collaboration], arXiv:1502.01582 [astro-ph.CO].

[2] D. J. Fixsen, Astrophys. J. **707**, 916 (2009) [arXiv:0911.1955 [astro-ph.CO]].

[3] Zeldovich Y. B., Sunyaev R. A., 1969, Astrophysics and Space Science, 4, 301; Sunyaev R. A., Zeldovich Y. B., 1970, Astrophysics and Space Science, 7, 20; Illarionov A. F., Sunyaev R. A., 1975, Soviet Astronomy, 18, 413; Illarionov A. F., Sunyaev R. A., 1975, Soviet Astronomy, 18, 691; Danese L., de Zotti G.. 1977. Nuovo Cimento Rivista Serie, 7, 277; Burigana C., Danese L., de Zotti G., 1991, A&A, 246, 49; W. Hu and J. Silk, Phys. Rev. D **48**, 485 (1993).

[4] W. T. Hu, astro-ph/9508126; J. Chluba, Mon. Not. Roy. Astron. Soc. **454**, 4182 (2015) doi:10.1093/mnras/stv2243 [arXiv:1506.06582 [astro-ph.CO]].

[5] P. J. E. Peebles, Astrophys. J. **153**, 1 (1968); Y. B. Zeldovich, V. G. Kurt and R. A. Sunyaev, J. Exp. Theor. Phys. **28**, 146 (1969) [Zh. Eksp. Teor. Fiz. **55**, 278 (1968)]; Dubrovich V. K., 1975, Soviet Astronomy Letters, 1, 196; J. A. Rubino-Martin, J. Chluba and R. A. Sunyaev, Mon. Not. Roy. Astron. Soc. **371**, 1939 (2006) [astro-ph/0607373]; J. A. Rubino-Martin, J. Chluba and R. A. Sunyaev, Astron. Astrophys. **485**, 377 (2008) [arXiv:0711.0594 [astro-ph]]; J. Chluba and R. A. Sunyaev, Astron. Astrophys. **503**, 345 (2009) [arXiv:0904.2220 [astro-ph.CO]].

[6] R. A. Sunyaev and Y. B. Zeldovich, Astron. Astrophys. **20**, 189 (1972); W. Hu, D. Scott and J. Silk, Phys. Rev. D **49**, 648 (1994) [astro-ph/9305038]; R. Cen and J. P. Ostriker, Astrophys. J. **514**, 1 (1999) [astro-ph/9806281]; A. Refregier, E. Komatsu, D. N. Spergel and U. L. Pen, Phys. Rev. D **61**, 123001 (2000) [astro-ph/9912180]; F. Miniati, D. Ryu, H. Kang, T. W. Jones, R. Cen and J. P. Ostriker, Astrophys. J. **542**, 608 (2000) [astro-ph/0005444]; R. Barkana and A. Loeb,

Phys. Rept. **349**, 125 (2001) [astro-ph/0010468]; S. P. Oh, A. Cooray and M. Kamionkowski, Mon. Not. Roy. Astron. Soc. **342**, L20 (2003) [astro-ph/0303007]; P. J. Zhang, U. L. Pen and H. Trac, Mon. Not. Roy. Astron. Soc. **355**, 451 (2004) [astro-ph/0402115]; J. C. Hill, N. Battaglia, J. Chluba, S. Ferraro, E. Schaan and D. N. Spergel, arXiv:1507.01583 [astro-ph.CO]; K. Dolag, E. Komatsu and R. Sunyaev, arXiv:1509.05134 [astro-ph.CO].

[7] R. A. Sunyaev and Y. B. Zeldovich, Astrophys. Space Sci. **7**, 3 (1970); Barrow, J. D. and Coles, P. 1991, MNRAS, 248, 52; Daly R. A., 1991, ApJ, 371, 14; W. Hu, D. Scott and J. Silk, Astrophys. J. **430**, L5 (1994) [astro-ph/9402045]; W. Hu and N. Sugiyama, Astrophys. J. **436**, 456 (1994) [astro-ph/9403031]; Y. F. Cai, J. B. Dent and D. A. Easson, Phys. Rev. D **83**, 101301 (2011) [arXiv:1011.4074 [hep-th]]; J. Chluba, R. Khatri and R. A. Sunyaev, Mon. Not. Roy. Astron. Soc. **425**, 1129 (2012) [arXiv:1202.0057 [astro-ph.CO]]; E. Pajer and M. Zaldarriaga, Phys. Rev. Lett. **109**, 021302 (2012) [arXiv:1201.5375 [astro-ph.CO]]; J. Chluba, A. L. Erickcek and I. Ben-Dayan, Astrophys. J. **758**, 76 (2012) [arXiv:1203.2681 [astro-ph.CO]]; J. Chluba and D. Grin, Mon. Not. Roy. Astron. Soc. **434**, 1619 (2013) [arXiv:1304.4596 [astro-ph.CO]]: M. Biagetti, H. Perrier, A. Riotto and V. Desjacques, Phys. Rev. D **87**, 063521 (2013) [arXiv:1301.2771 [astro-ph.CO]]; R. Emami, E. Dimastrogiovanni, J. Chluba and M. Kamionkowski, Phys. Rev. D **91**, no. 12, 123531 (2015) [arXiv:1504.00675 [astro-ph.CO]].

[8] J. L. Feng, A. Rajaraman and F. Takayama, Phys. Rev. D **68**, 063504 (2003) [hep-ph/0306024]; J. L. Feng, S. f. Su and F. Takayama, Phys. Rev. D **70**, 063514 (2004) [hep-ph/0404198]; J. L. Feng, S. Su and F. Takayama, Phys. Rev. D **70**, 075019 (2004) [hep-ph/0404231]; R. Lamon and R. Durrer, Phys. Rev. D **73**, 023507 (2006) [hep-ph/0506229]; K. Kohri, T. Moroi and A. Yotsuyanagi, Phys. Rev. D **73**, 123511 (2006) [hep-ph/0507245]; L. Zhang, X. Chen, M. Kamionkowski, Z. g. Si and Z. Zheng, Phys. Rev. D **76**, 061301 (2007) [arXiv:0704.2444 [astro-ph]]; J. L. Feng, Ann. Rev. Astron. Astrophys. **48**, 495 (2010) [arXiv:1003.0904 [astro-ph.CO]].

[9] Zeldovich Y. B., Sunyaev R. A., 1969, Astrophysics and Space Science, 4, 301.

[10] J. Chluba and R. A. Sunyaev, Mon. Not. Roy. Astron. Soc. **419**, 1294 (2012) [arXiv:1109.6552 [astro-ph.CO]]; R. Khatri and R. A. Sunyaev, JCAP **1209**, 016 (2012) [arXiv:1207.6654 [astro-ph.CO]]; J. Chluba, Mon. Not. Roy. Astron. Soc. **434**, 352 (2013) [arXiv:1304.6120 [astro-ph.CO]].

[11] J. Chluba, Mon. Not. Roy. Astron. Soc. **436**, 2232 (2013) [arXiv:1304.6121 [astro-ph.CO]]; J. Chluba and D. Jeong, Mon. Not. Roy. Astron. Soc. **438**, no. 3, 2065 (2014) [arXiv:1306.5751 [astro-ph.CO]].

[12] E. L. Wright *et al.*, Astrophys. J. **420**, 450 (1994); J. C. Mather *et al.*, Astrophys. J. **420**, 439 (1994). doi:10.1086/173574; D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer and E. L. Wright, Astrophys. J. **473**, 576 (1996) [astro-ph/9605054].

[13] A. Kogut *et al.*, JCAP **1107**, 025 (2011) [arXiv:1105.2044 [astro-ph.CO]]; P. Andr *et al.* [PRISM Collaboration], JCAP **1402**, 006 (2014) [arXiv:1310.1554 [astro-ph.CO]].

[14] J. Chluba and R. A. Sunyaev, Mon. Not. Roy. Astron. Soc. **419**, 1294 (2012) doi:10.1111/j.1365-2966.2011.19786.x [arXiv:1109.6552 [astro-ph.CO]]; R. A. Sunyaev and R. Khatri, Int. J. Mod. Phys. D **22**, 1330014 (2013) doi:10.1142/S0218271813300140 [arXiv:1302.6553 [astro-ph.CO]]; H. Tashiro, PTEP **2014**, no. 6, 06B107 (2014). doi:10.1093/ptep/ptu066.

[15] W. Hu and J. Silk, Phys. Rev. D **48**, 485 (1993); W. Hu and J. Silk, Phys. Rev. Lett. **70**, 2661 (1993); P. McDonald, R. J. Scherrer and T. P. Walker, Phys. Rev. D **63**, 023001 (2001) [astro-ph/0008134]; J. Chluba, Mon. Not. Roy. Astron. Soc. **402**, 1195 (2010) [arXiv:0910.3663 [astro-ph.CO]]; J. Chluba and R. A. Sunyaev, Mon. Not. Roy. Astron. Soc. **419**, 1294 (2012) [arXiv:1109.6552 [astro-ph.CO]]; J. Chluba, Mon. Not. Roy. Astron. Soc. **436**, 2232 (2013) [arXiv:1304.6121 [astro-ph.CO]]; J. Chluba and D. Jeong, Mon. Not. Roy. Astron. Soc. **438**, no. 3, 2065 (2014) [arXiv:1306.5751 [astro-ph.CO]].

[16] J. R. Ellis, A. D. Linde and D. V. Nanopoulos, Phys. Lett. B **118**, 59 (1982); L. M. Krauss, Phys. Lett. B **128**, 37 (1983); L. M. Krauss, Nucl. Phys. B **227**, 556 (1983); D. V. Nanopoulos, K. A. Olive and M. Srednicki, Phys. Lett. B **127**, 30 (1983); J. R. Ellis, J. E. Kim and D. V. Nanopoulos, Phys. Lett. B **145**, 181 (1984); M. Y. Khlopov and A. D. Linde, Phys. Lett. B **138**, 265 (1984); R. Juszkiewicz, J. Silk and A. Stebbins, Phys. Lett. B **158**, 463 (1985); G. F. Giudice, A. Riotto and I. Tkachev, JHEP **9911**, 036 (1999) [hep-ph/9911302].

[17] S. P. Martin, Adv. Ser. Direct. High Energy Phys. **21**, 1 (2010) [Adv. Ser. Direct. High Energy Phys. **18**, 1 (1998)] [hep-ph/9709356].

[18] H. Pagels and J. R. Primack, Phys. Rev. Lett. **48**, 223 (1982); S. Weinberg, Phys. Rev. Lett. **48**, 1303 (1982); J. R. Ellis, G. B. Gelmini, J. L. Lopez, D. V. Nanopoulos and S. Sarkar, Nucl. Phys. B **373**, 399 (1992). J. L. Feng, eConf C **0307282**, L11 (2003) [hep-ph/0405215].

[19] A. H. Guth and S. Y. Pi, Phys. Rev. Lett. **49**, 1110 (1982); A. A. Starobinsky, Phys. Lett. B **117**, 175 (1982); J. M. Bardeen, P. J. Steinhardt and M. S. Turner, Phys. Rev. D **28**, 679 (1983).

[20] J. R. Ellis, A. D. Linde and D. V. Nanopoulos, Phys. Lett. B **118**, 59 (1982); L. M. Krauss, Phys. Lett. B **128**, 37 (1983); L. M. Krauss, Nucl. Phys. B **227**, 556 (1983); D. V. Nanopoulos, K. A. Olive and M. Srednicki, Phys. Lett. B **127**, 30 (1983); J. R. Ellis, J. E. Kim and D. V. Nanopoulos, Phys. Lett. B **145**, 181 (1984); M. Y. Khlopov and A. D. Linde, Phys. Lett. B **138**, 265 (1984); R. Juszkiewicz, J. Silk and A. Stebbins, Phys. Lett. B **158**, 463 (1985); G. F. Giudice, A. Riotto and I. Tkachev, JHEP **9911**, 036 (1999) [hep-ph/9911302]; J. Ellis, M. A. G. Garcia, D. V. Nanopoulos, K. A. Olive and M. Peloso, arXiv:1512.05701 [astro-ph.CO].

[21] T. Kanzaki, M. Kawasaki, K. Kohri and T. Moroi, Phys. Rev. D **76**, 105017 (2007) doi:10.1103/PhysRevD.76.105017 [arXiv:0705.1200 [hep-ph]]; L. M. Krauss, Nucl. Phys. B **227**, 556 (1983);

[22] K. N. Abazajian *et al.* [Topical Conveners: K.N. Abazajian, J.E. Carlstrom, A.T. Lee Collaboration], Astropart. Phys. **63**, 66 (2015) doi:10.1016/j.astropartphys.2014.05.014 [arXiv:1309.5383 [astro-ph.CO]].

[23] L. Dai, M. Kamionkowski and J. Wang, Phys. Rev. Lett. **113**, 041302 (2014) [arXiv:1404.6704 [astro-ph.CO]]; J. Martin, C. Ringeval and V. Vennin, Phys. Rev. Lett. **114**, no. 8, 081303 (2015) [arXiv:1410.7958 [astro-ph.CO]]; J. B. Munoz and M. Kamionkowski, Phys. Rev. D **91**, no. 4, 043521 (2015) [arXiv:1412.0656 [astro-ph.CO]]; J. L. Cook, E. Dimastrogiovanni, D. A. Easson and L. M. Krauss, JCAP **1504**, 047 (2015) [arXiv:1502.04673 [astro-ph.CO]].

[24] G. Steigman, Ann. Rev. Nucl. Part. Sci. **57**, 463 (2007) [arXiv:0712.1100 [astro-ph]].

[25] P. A. R. Ade *et al.* [BICEP2 Collaboration], Phys. Rev. Lett. **112**, no. 24, 241101 (2014) [arXiv:1403.3985 [astro-ph.CO]]; P. A. R. Ade *et al.* [BICEP2 and Planck Collaborations], Phys. Rev. Lett. **114**, 101301 (2015) [arXiv:1502.00612 [astro-

ph.CO]].

[26] J. R. Ellis, J. E. Kim and D. V. Nanopoulos, Phys. Lett. B **145**, 181 (1984); M. Kawasaki and T. Moroi, Prog. Theor. Phys. **93**, 879 (1995) [hep-ph/9403364, hep-ph/9403061].

[27] L. M. Krauss, Nucl. Phys. B **227**, 556 (1983); L. M. Krauss, A. J. Long and S. Sabharwal, Phys. Rev. D **89**, no. 4, 043503 (2014) [arXiv:1309.1454 [hep-ph]].

[28] L. M. Krauss, Nucl. Phys. B **227**, 556 (1983); J. R. Ellis, J. E. Kim and D. V. Nanopoulos, Phys. Lett. B **145**, 181 (1984); J. A. Frieman and G. F. Giudice, Phys. Lett. B **224**, 125 (1989); M. Kawasaki and T. Moroi, Prog. Theor. Phys. **93**, 879 (1995) [hep-ph/9403364, hep-ph/9403061]; R. H. Cyburt, J. R. Ellis, B. D. Fields and K. A. Olive, Phys. Rev. D **67**, 103521 (2003) [astro-ph/0211258]; K. Kohri, T. Moroi and A. Yotsuyanagi, Phys. Rev. D **73**, 123511 (2006) [hep-ph/0507245]; F. D. Steffen, JCAP **0609**, 001 (2006) [hep-ph/0605306]; M. Kawasaki, K. Kohri, T. Moroi and A. Yotsuyanagi, Phys. Rev. D **78**, 065011 (2008) [arXiv:0804.3745 [hep-ph]]; M. Pospelov and J. Pradler, Ann. Rev. Nucl. Part. Sci. **60**, 539 (2010) [arXiv:1011.1054 [hep-ph]]; Y. Ema, R. Jinno and T. Moroi, JHEP **1410**, 150 (2014) [arXiv:1408.1745 [hep-ph]].

[29] X. L. Chen and M. Kamionkowski, Phys. Rev. D **70**, 043502 (2004) [astro-ph/0310473].

[30] N. Taburet, C. Hernandez-Monteagudo, N. Aghanim, M. Douspis and R. A. Sunyaev, Mon. Not. Roy. Astron. Soc. **418**, 2207 (2011) [arXiv:1012.5036 [astro-ph.CO]].

[31] R. A. Sunyaev and Y. B. Zeldovich, Astrophys. Space Sci. **7**, 20 (1970).

[32] R. A. Sunyaev and Y. B. Zeldovich, Astrophys. Space Sci. **7**, 3 (1970); J. Chluba and R. A. Sunyaev, Mon. Not. Roy. Astron. Soc. **419**, 1294 (2012) [arXiv:1109.6552 [astro-ph.CO]]; J. Chluba, Mon. Not. Roy. Astron. Soc. **434**, 352 (2013) [arXiv:1304.6120 [astro-ph.CO]].

[33] L. M. Krauss, Nucl. Phys. B **227**, 556 (1983); T. Moroi, hep-ph/9503210.

[34] M. Kawasaki, K. Kohri and T. Moroi, Phys. Rev. D **71**, 083502 (2005) [astro-ph/0408426].

[35] J. R. Ellis, J. E. Kim and D. V. Nanopoulos, Phys. Lett. B **145**, 181 (1984).

[36] For general photon injection scenarios, the distortion exhibits a richer phenomenology than the standard μ and y distortions which are caused by pure heating (see second reference in [4]).

Implications of a Primordial Magnetic Field for Magnetic Monopoles, Axions, and Dirac Neutrinos

Andrew J. Long[a*] and Tanmay Vachaspati[a,b†]
[a] *Physics Department, Arizona State University, Tempe, Arizona 85287, USA.*
[b] *Physics Department and McDonnell Center for the Space Sciences, Washington University, St. Louis, MO 63130, USA.*
(Dated: September 6, 2017)

We explore some particle physics implications of the growing evidence for a helical primordial magnetic field (PMF). From the interactions of magnetic monopoles and the PMF, we derive an upper bound on the monopole number density, $n(t_0) < 1 \times 10^{-20}\,\mathrm{cm}^{-3}$, which is a "primordial" analog of the Parker bound for the survival of galactic magnetic fields. Our bound is weaker than existing constraints, but it is derived under independent assumptions. We also show how improved measurements of the PMF at different redshifts can lead to further constraints on magnetic monopoles. Axions interact with the PMF due to the $g_{a\gamma}\varphi\mathbf{E}\cdot\mathbf{B}/4\pi$ interaction. Including the effects of the cosmological plasma, we find that the helicity of the PMF is a source for the axion field. Although the magnitude of the source is small for the PMF, it could potentially be of interest in astrophysical environments. Earlier derived constraints from the resonant conversion of cosmic microwave background (CMB) photons into axions lead to $g_{a\gamma} \lesssim 10^{-9}$ GeV^{-1} for the suggested PMF strength $\sim 10^{-14}\,\mathrm{G}$ and coherence length ~ 10 Mpc. Finally we apply constraints on the neutrino magnetic dipole moment that arise from requiring successful big bang nucleosynthesis in the presence of a PMF and we find $\mu_\nu \lesssim 10^{-16}\mu_B$.

I. INTRODUCTION

There is growing evidence for the existence of an intergalactic magnetic field from the observation of high energy gamma rays. It is likely that a magnetic field in intergalactic space would have been created in the early universe, since astrophysics alone is not expected to generate fields on such large length scales. (For a recent review on cosmic magnetic fields, see Ref. [1].) The discovery of a primordial magnetic field (PMF) has important ramifications for cosmology as it allows one to test models of magnetogenesis, which are often tied to the physics of inflation [2] cosmological phase transitions [3, 4], and baryogenesis [5–7]. The presence of a PMF after cosmological recombination can also aid in the formation of first stars [8] and provide the seed field for the galactic dynamo [9]. Additionally, the existence of a PMF in our universe opens the opportunity to place indirect constraints on exotic particle physics models where the new physics couples to electromagnetism. In this paper we will investigate the consequences of a PMF for models that contain magnetic monopoles, axions, and Dirac neutrinos with a magnetic moment.

Blazars that emit TeV gamma rays are expected to produce an electromagnetic cascade of lower energy gamma rays due to electron-positron pair production and the subsequent inverse Compton up-scattering of cosmic microwave background (CMB) photons [10–14]. In the presence of an intergalactic magnetic field, electrons and positrons directed toward the Earth can be deflected off of the line of sight, and those that are directed away from the Earth can be deflected back toward the line of sight. As a result the point source flux is depleted in the GeV band, and the blazar acquires a halo of GeV gamma rays. The non-observation of these GeV gamma rays was used to place a lower bound on the magnetic field strength at the level of $B \gtrsim 10^{-16}\,\mathrm{G}$ [15–17]. This bound depends on modeling of the blazar flux stability and also the plasma instabilities during propagation, and it may weaken substantially depending on these assumptions [18–21]. The search for the GeV halo extended emission has been ongoing [22–26]. Most recently, Chen et al. [27] have found evidence for the halo in a stacked analysis of known blazars at ~ 1 GeV energies and interpret it to be due to a field with strength $B \sim 10^{-17} - 10^{-15}\,\mathrm{G}$. For reference, measurements of the cosmic microwave background place an upper bound on the magnetic field strength at the level of $B \lesssim 10^{-9}\,\mathrm{G}$ [28].

There are theoretical motivations for considering the possibility that the PMF is *helical*, *i.e.* there is an excess of power in either right- or left-circular polarization modes. Helical magnetic fields emerge in many models of magnetogenesis [5–7, 29], and helicity conservation dramatically impacts the evolution of the PMF, aiding in its survival and growth [30]. Recently, Tashiro et al. [31, 32] analyzed the diffuse gamma ray sky at 10-60 GeV energies to look for the parity-violating signature [33, 34] of a helical magnetic field. They find evidence for an intergalactic magnetic field with strength $B \sim 10^{-14}\,\mathrm{G}$ on coherence scales $\lambda_B \sim 10\,\mathrm{Mpc}$ and with left-handed helicity. Although the results of Refs. [31, 32] and [27] appear inconsistent, it may be possible to reconcile them by noting that the weak bending approximation breaks down for $B \sim 10^{-14}\,\mathrm{G}$ for gamma rays at ~ 1 GeV energies [35].

*Electronic address: andrewjlong@asu.edu
†Electronic address: tvachasp@asu.edu

Motivated by these recent results, we investigate if the existence of a helical PMF can be used to constrain other particle physics ideas in a cosmological setting. Our analysis is sufficiently general that our results will remain relevant even if the gamma ray observation results should change or go away, *e.g.* with more data. However, for the purpose of numerical estimates we will use $B \sim 10^{-14}\,\mathrm{G}$ and $\lambda_B \sim 10\,\mathrm{Mpc}$ as the fiducial field strength and coherence length scale, and we will take the magnetic field to have maximal (left-handed) helicity.

In Sec. II we consider the interaction of a hypothetical abundance of cosmic magnetic monopoles and a PMF. The magnetic field does work on the monopoles, and its field strength is thereby depleted. The constraints we obtain in this way are generally weaker than existing bounds but are obtained under a different set of assumptions. These results are summarized in Fig. 1. We also discuss how heavy magnetic monopoles can lead to anomalous scaling of the energy density in the PMF. As observations of the PMF improve, they can be sensitive to the anomalous scaling and thus become a tool for further constraining magnetic monopoles.

In Sec. III we consider the interaction of an axion (φ) with the PMF through the coupling $g_{a\gamma}\varphi F\tilde{F}$. In this analysis we include the cosmological plasma, and thus we study the equations of magnetohydrodynamics coupled to an axion. Although we find that the axion has a negligible effect on the spectrum and evolution of the PMF, it is interesting to note that this conclusion is not sensitive to the assumed scale of Peccei-Quinn symmetry breaking, f_a, as long as $g_{a\gamma} \propto 1/f_a$. In turn, the PMF leaves the evolution of the axion condensate largely unaffected. In principle the PMF damps the axion oscillations and the helicity of the PMF shifts the equilibrium point, but these effects are quadratic in the already-small magnetic field strength.

In Sec. IV we consider the interaction of the neutrino magnetic dipole moment, μ_ν, with the PMF. Enqvist et al. [36, 37] have shown that this interaction induces a spin-flip transition, which cannot be in equilibrium in the early universe without running afoul of constraints on the number of relativistic neutrino species. Using their result with our fiducial value of $B \sim 10^{-14}\,\mathrm{G}$, we evaluate an upper bound on μ_ν, which is shown in Fig. 2.

We work in the CGS system with $\hbar = c = 1$. The unit of electric charge is $e = \sqrt{\alpha} \simeq 0.085$ with $\alpha \simeq 1/137$ the fine structure constant, and the unit of magnetic charge is $e_m = 1/2e \simeq 5.9$. The magnetic field is measured in Gauss, and $1\,\mathrm{G} \simeq 6.93 \times 10^{-20}\,\mathrm{GeV}^2$. The reduced Planck mass is denoted by $M_P \simeq 2.4 \times 10^{18}\,\mathrm{GeV}$. The metric signature is $(+---)$, and the antisymmetric tensor normalization is $\epsilon^{0123} = +1$.

II. MAGNETIC MONOPOLES

A conservative cosmological bound on the energy density of magnetic monopoles is $\Omega_m \equiv \rho_m/\rho_{\mathrm{cr}} < 0.3$ where ρ_m is the energy density in monopoles and $\rho_{\mathrm{cr}} \simeq 10^{-29}\,\mathrm{gm}\,c^2/\,\mathrm{cm}^3$ is the critical cosmological energy density. The number density of nonrelativistic monopoles is $n_m = \rho_m/m$ with m the monopole mass, and the cosmological bound implies

$$n_m < 0.3\frac{\rho_{\mathrm{cr}}}{m} \simeq (2 \times 10^{-23}\,\mathrm{cm}^{-3})\left(\frac{m}{10^{17}\,\mathrm{GeV}}\right)^{-1}. \quad (1)$$

The bound grows weaker for lighter monopoles since they contribute less to the energy density for the same number density.

The existence of the galactic magnetic field leads to another indirect bound. Magnetic monopoles tend to deplete a magnetic field in the same way that free electrons short out a conductor. The survival of the micro-Gauss galactic magnetic field implies an upper bound on the directed flux, $\mathcal{F}$, of magnetic charge onto the Milky Way. Requiring that the time scale for B-field depletion is longer than the dynamo time scale of B-field regeneration ($\tau_{\mathrm{dyn}} \simeq 10^8\,\mathrm{yr}$), leads to the so called Parker bound [38]

$$\mathcal{F} < 0.9 \times 10^{-16} e_m\,\mathrm{cm}^{-2}\,\mathrm{sec}^{-1}\,\mathrm{sr}^{-1}. \quad (2)$$

Assuming that monopoles have unit charge and travel with velocity v, the Parker bound can be expressed as an upper bound on the monopole number density:

$$n_m \approx \frac{(4\pi\ \mathrm{sr})\,\mathcal{F}}{e_m\,v} < (4 \times 10^{-23}\,\mathrm{cm}^{-3})\left(\frac{v}{10^{-3}c}\right)^{-1}. \quad (3)$$

Just as the Parker bound is predicated on the existence of a magnetic field in the Milky Way, we expect that a similar bound can be inferred from the existence of a primordial magnetic field in the early universe.

We study a gas of monopoles and antimonopoles immersed in a magnetic field that permeates the cosmological medium. The monopoles have mass m, magnetic charge e_m, and they are homogeneously distributed with number density $n_m(t)$. The magnetic field $\mathbf{B}(t, \mathbf{x})$ induces a Lorentz force of

$$\mathbf{F}_{\mathrm{B}} = e_m \mathbf{B} \quad (4)$$

on a monopole at $(t, \mathbf{x})$, and it begins to drift along the field line with a velocity $\mathbf{v}$. The field does work by pushing the monopole, and in this way the monopole extracts energy from the magnetic field at a rate

$$\dot{E}_{\mathrm{B}} = \mathbf{F}_{\mathrm{B}} \cdot \mathbf{v}\,. \quad (5)$$

To solve for the evolution of the magnetic field strength we must know the monopole velocity. Prior to electron-positron annihilation, the monopole's velocity is restricted by elastic scattering with the cosmological medium, but afterward it can free stream. We will consider each of these cases in turn.

A. Friction Dominated Regime

In the epoch prior to e^+e^- annihilation the cosmological medium was dense with electromagnetically charged particles. In this regime, monopoles interact with the medium through elastic scattering such as $M + e^\pm \to M + e^\pm$ with M the monopole. It is safe to assume that the monopole's rest mass is much larger than the kinetic energy of particles in the plasma. This allows us to characterize the effective interaction with a drag force that takes the form [60]

$$\mathbf{F}_{\rm drag} = -f_{\rm drag}\mathbf{v}\,. \tag{6}$$

At the time of interest, the scatterers are relativistic with energy comparable to the temperature T of the plasma, and they are in thermal equilibrium with number density $n \sim T^3$. For such a system the drag coefficient takes the form [39, 40]

$$f_{\rm drag} \approx \beta\, e^2 e_m^2\, g_{em} T^2 \tag{7}$$

where $g_{em}(t)$ is the number of relativistic, charged degrees of freedom in thermal equilibrium at time t and β is an $O(1)$ number related to the spin character and charge of the scatterers. We will drop β from this point onward since it is parametrically redundant with g_{em}. Also note that g_{em} and T depend on time, but this can be ignored on short time scales as compared to the Hubble time scale.

The monopole's equation of motion is

$$m\,\dot{\mathbf{v}} = e_m\mathbf{B} - f_{\rm drag}\mathbf{v}\,. \tag{8}$$

We assume that the distance traveled by the monopole is small compared to the correlation length λ_B of the magnetic field, and we treat $\mathbf{B}$ as uniform. Eq. (8) immediately gives the terminal velocity of the monopoles,

$$\mathbf{v}_{\rm term} = \frac{e_m\mathbf{B}}{e^2 e_m^2 g_{em} T^2}\,, \tag{9}$$

which is achieved on a time scale

$$\tau_{\rm term} = \frac{m}{e^2 e_m^2 g_{em} T^2}\,. \tag{10}$$

Comparing with the Hubble time $t_H \sim M_P/T^2$ (radiation era) we have $\tau_{\rm term} \ll t_H$ provided that $m < g_{em} M_P$, and thus the cosmological expansion is negligible.

At the present cosmic epoch, the photon temperature is $\sim 10^{-4}$ eV and $B \sim 10^{-14}\,$G. Assuming $B \propto T^2$, we get $B \simeq 10^6\,$G at $T \simeq$ MeV when $g_{em} \simeq 10$. These estimates give $v_{\rm term} \simeq 10^{-8}$, which validates our uses of the non-relativistic equation of motion. As a consequence, the distance traveled by a monopole during $\tau_{\rm term}$ is quite small, $d_{\rm term} < 10^{-8}\tau_{\rm term}$. We shall assume that the correlation length of the magnetic field is larger, $\lambda_B > d_{\rm term}$, thus justifying our treatment of the magnetic field as being uniform.

The magnetic field's response to this current is given by the magnetic analog of Ampere's law,

$$\dot{\mathbf{B}} = -4\pi\,\mathbf{j}_M = -4\pi e_m n_m \mathbf{v}\,. \tag{11}$$

where n_m is the number density of monopoles (assumed equal to the number density of antimonopoles). Here we have used $\mathbf{E} = 0$ since electric fields are screened due to the high electrical conductivity of the cosmological medium. For typical parameters, the inter-monopole spacing is small compared to the correlation length of the magnetic field, $n_m^{-1/3} \ll \lambda_B$, and we can interpret n_m and $\mathbf{B}$ as coarse grained quantities on this length scale. Then we insert $\boldsymbol{v} = \boldsymbol{v}_{\rm term}$ from Eq. (9) into Eq. (11) to get the solution,

$$\mathbf{B}(t) = \mathbf{B}(t_i)\, e^{-(t-t_i)/\tau_{\rm decay}} \tag{12}$$

where the decay time scale of the magnetic field is given by

$$\tau_{\rm decay} = \frac{e^2 e_m^2 g_{em} T^2}{4\pi e_m^2 n_m}. \tag{13}$$

In obtaining this solution, we have assumed $\boldsymbol{v} = \boldsymbol{v}_{\rm term}$ which is justified if the monopoles reach terminal velocity much more quickly than the decay time scale, *i.e.* $\tau_{\rm decay} \gg \tau_{\rm term}$. As we will see below, this condition is satisfied for the range of parameters of interest to us.

To ensure survival of the magnetic field, we require that $\tau_{\rm decay}$ is much larger than the Hubble time at temperature T,

$$t_H = \frac{1}{2H} \simeq 1.5\frac{M_P}{g_* T^2} \tag{14}$$

with H the Hubble parameter and g_* the effective number of relativistic degrees of freedom. Substitution of Eq. (13) into $\tau_{\rm decay} > t_H$ now leads to a constraint on the number density of monopoles,

$$n_m < \frac{e^2 e_m^2}{6\pi e_m^2}\frac{g_{em} g_* T^4}{M_P}\,, \tag{15}$$

when the universe had temperature T.

The strongest bound is obtained when T is smallest. Since our calculation assumes that monopoles scatter on relativistic, charged particles with a thermal abundance, the last time at which this is possible is the epoch of e^+e^- annihilation. At this time $T_{\rm ann} \simeq 1\,$MeV, $g_{em} \approx g_* \approx g_{*S} \simeq 10.75$. To translate this into a bound on the monopole number density today, denoted by n_0, we multiply by $(a_{\rm ann}/a_0)^3 = (g_{*S,0}T_0^3/g_{*S}T_{\rm ann}^3)$ with $g_{*S,0} \simeq 3.91$ and $T_0 \simeq 2.3\times10^{-4}\,$eV the temperature of the microwave background photons today. We find an upper bound on the monopole number density today,

$$n_0 < \frac{e^2 e_m^2}{6\pi e_m^2}\frac{g_{*S,0} g_* g_{em}}{g_{*S}}\frac{T_{\rm ann} T_0^3}{M_p} \simeq 1\times10^{-20}\,{\rm cm}^{-3}\,. \tag{16}$$

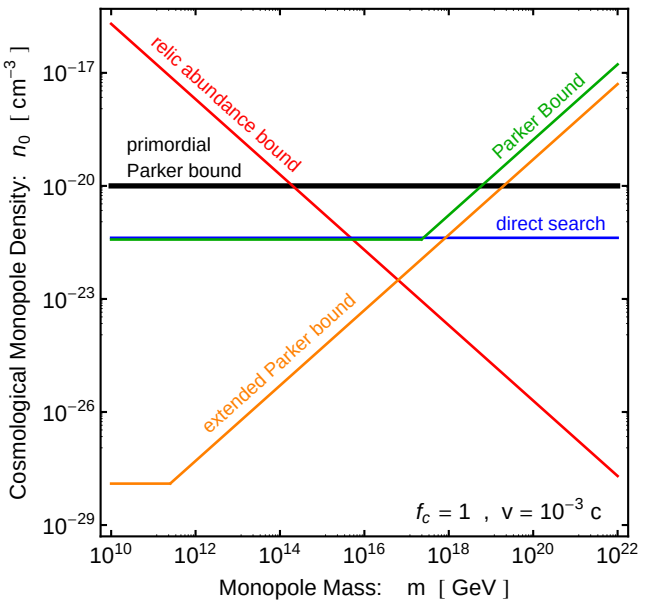

FIG. 1: A summary of upper bounds on the magnetic monopole abundance from this work and the literature. *Black:* the requirement of survival of the primordial magnetic field ("primordial Parker bound"), derived here in Eq. (16). *Red:* the cosmological abundance bound in Eq. (1). *Blue:* direct search constraints [42] (see the magnetic monopole review). *Green:* the requirement of survival of the Galactic magnetic field (Parker bound), given by Ref. [43]. *Orange:* the requirement of survival of the Galactic seed field ("extended Parker bound"), given by Ref. [41]. We take $v \simeq 10^{-3}$ and assume that monopoles are unclustered, $f_c \simeq 1$. If the monopoles are clustered then the Parker bound, extended Parker bound, and direct search limits move down by a factor of $f_c \sim 10^5$.

If this bound is not satisfied then any primordial magnetic field would have been exponentially depleted by the time of electron-positron annihilation. Due to the close connection with the Parker bound for survival of galactic magnetic fields, we will refer to Eq. (16) as the "primordial Parker bound."

In Fig. 1 we compare the primordial Parker bound in Eq. (16) with other constraints derived previously in the literature. Since these constraints are typically expressed as a bound on the monopole flux, we translate into a bound on the number density using $(4\pi\,\mathrm{sr})\mathcal{F} \approx f_c n v$. Here v is the average monopole velocity and $f_c = n_{\rm galaxy}/n_{\rm cosmo.}$ is the enhancement factor that accounts for clustering of monopoles in the galaxy. For clustered monopoles $f_c \sim 10^5$, but otherwise $f_c \sim 1$. For the extended Parker bound calculation of Ref. [41], we take $B_{\rm seed} = 10^{-11}\,\mathrm{G}$. For the direct search constraint we show a relatively conservative and robust limit of $\mathcal{F} < 10^{-15}\,\mathrm{cm}^{-2}\,\mathrm{sec}^{-1}\,\mathrm{sr}^{-1}$, but stronger constraints are available for specific monopole parameters [42]. From the figure, one can see that the primordial Parker bound becomes stronger than the cosmological bound for light monopoles, $m \lesssim 5 \times 10^{13}\,\mathrm{GeV}$, but it always remains weaker than the direct search constraints.

B. Free Streaming Regime

After cosmological electron-positron annihilation the number density of these scatterers decreases by a factor of $\sim 10^{-10}$. The monopoles experience very little drag force, and they can be accelerated freely by the magnetic field. For a uniform and static magnetic field, the solution of Eq. (8) with $f_{\rm drag} = 0$ is simply $v = e_m B t/m$, or for an inhomogenous field with domains of size λ_B we find

$$v(t) \sim \frac{e_m B \lambda_B}{m}\sqrt{\frac{t}{\lambda_B}} \tag{17}$$

if the motion is diffusive. The monopole becomes relativistic when $v(t_{\rm rel}) \sim 1$ and comparing this time with the present age of the universe gives

$$\frac{t_{\rm rel}}{t_0} \sim \left(\frac{m}{e_m B t_0}\right)^2 \frac{t_0}{\lambda_B} \sim 10^{18}\left(\frac{m}{M_P}\right)^2 \frac{t_0}{\lambda_B}. \tag{18}$$

With $\lambda_B \sim$ Mpc and $t_0 \sim 10\,\mathrm{Gpc}$, we find that monopoles are relativistic today if $m \lesssim 10^8\,\mathrm{GeV}$, and they are non-relativistic otherwise.

The above estimate ignores backreaction of the monopoles on the PMF. To check for consistency, we compare the kinetic energy in monopoles $\rho_{\rm kin}$ to the energy density available in the PMF $\rho_B = B^2/8\pi$. For relativistic monopoles we should have $\rho_B > \rho_{\rm kin} \gg m\,n_m$, and this provides an upper bound on the number density of monopoles for which the velocity estimate in Eq. (17) can be expected to hold. Taking $B \sim 10^{-14}\,\mathrm{G}$ we find

$$n_m \ll 10^{-35}\left(\frac{10^8\ \mathrm{GeV}}{m}\right)\ \mathrm{cm}^{-3}\,. \tag{19}$$

Since are interested in much larger number densities, as indicated by the bound in Eq. (16), we cannot use the velocity relation in Eq. (17), but instead the monopole and magnetic field equations will need to be evolved simultaneously.

Without friction to provide a means of energy dissipation, the monopoles cannot deplete the magnetic field strength. Instead there is a conservative exchange of energy between the magnetic field and the kinetic energy of the monopoles. This co-evolution can lead to an anomalous departure from the usual power law scaling behavior of the magnetic field energy density if the monopoles are non-relativistic. This can be seen from the following argument. In the absence of the monopole gas, the energy density in the magnetic field redshifts like radiation $\rho_B \sim (1+z)^4$ where z is the cosmological redshift. Meanwhile the kinetic energy stored in a gas of non-relativistic particles redshifts more quickly. We can write the kinetic energy density as $\rho_{\rm kin} = n_m p^2/(2m)$ where p is the typical momentum and n_m is the monopole number density.

Since $p \sim (1+z)$ and $n_m \sim (1+z)^3$, the kinetic energy density redshifts like $\rho_{\rm kin} \sim (1+z)^5$. If energy is transferred quickly between the monopoles and magnetic field then we might expect $\rho_B \sim \rho_{\rm kin.} \sim (1+z)^{9/2}$ where 9/2 is the average of 4 and 5. This result is confirmed by the full calculation, which we will now present.

Once again we consider a gas of monopoles coupled to a magnetic field, but we now include the effects of cosmological expansion. Let $X^\lambda(\tau)$ be the world line of a monopole, and let $U^\lambda(\tau) = dX^\lambda/d\tau$ be its 4-velocity. The monopole equation of motion, given previously by Eq. (8), is now replaced by

$$m\left[\frac{dU_m^\lambda}{d\tau} + \Gamma^\lambda_{\mu\nu} U_m^\mu U_m^\nu\right] = e_m \widetilde{F}^\lambda{}_\mu\, U_m^\mu \tag{20}$$

where $\Gamma^\lambda_{\mu\nu}$ is the Christoffel symbol. The magnetic analog of Ampere's law in Eq. (11) is now replaced by

$$\nabla_\alpha \widetilde{F}^{\alpha\beta} = 4\pi j_M^\beta \tag{21}$$

where ∇_α is the covariant derivative. The magnetic current density j_M^μ arises from monopoles with velocity U_m^μ and antimonopoles with velocity $U_{\bar m}^\mu$. It can be written as

$$j_M^\mu = e_m\left(n_m\, U_m^\mu - n_{\bar m}\, U_{\bar m}^\mu\right), \tag{22}$$

and it satisfies the conservation law

$$\nabla_\mu j_M^\mu = 0\,. \tag{23}$$

The spatial component of the 4-velocity is the *comoving* peculiar velocity $(\mathbf{U}_m)^i = U_m^i$, and with an additional factor of a we form the *physical* peculiar velocity $(\mathbf{v}_m)^i = aU_m^i$. Neglecting the electric field and spatial gradients, the system of equations can be put into the form

$$\partial_\eta(a\mathbf{v}_m) = \frac{e_m}{m}(a^2\mathbf{B}) \tag{24a}$$

$$a\partial_\eta(a^2\mathbf{B}) = -4\pi e_m(a^3 n_m)\,(a\mathbf{v}_m) \tag{24b}$$

$$\partial_\eta\left(a^3 n_m\right) = 0 \tag{24c}$$

where $d\eta = dt/a = da/Ha^2$ is the conformal time coordinate. The third equation implies that the number of monopoles per comoving volume is conserved; $n_0 = a^3 n_m$ is the number density of monopoles today. Then if not for the additional factor of a in the second equation, the solutions would simply be oscillatory with angular frequency

$$\omega_{\rm pl} = \sqrt{\frac{4\pi e_m^2 n_0}{m}}\,. \tag{25}$$

This is just the usual formula for plasma frequency, but instead of electron charge, mass, and density, here we find the corresponding parameters for the monopole gas.

To solve these equations we must relate a to η. During the radiation era we have $\eta = \eta_i + (a-a_i)/H_i a_i^2 \approx a/H_i a_i^2$, and the solution is

$$\mathbf{B} = \left(\frac{a}{a_i}\right)^{-9/4}\left[\frac{J_0(\phi)\;\mathbf{B}_1 + Y_0(\phi)\;\mathbf{B}_2}{(a/a_i)^{-1/4}}\right] \tag{26a}$$

$$\mathbf{v}_m = \left(\frac{a}{a_i}\right)^{-3/4}\left[\frac{\phi\; J_1(\phi)\;\mathbf{v}_1 + \phi\; Y_1(\phi)\;\mathbf{v}_2}{(a/a_i)^{1/4}}\right] \tag{26b}$$

where

$$\phi \equiv \frac{2\tilde\omega\,\eta_i}{a_i}\sqrt{a}\,. \tag{27}$$

At late times $\phi \gg 1$, and all of the Bessel functions go to zero with an envelop $\sim \phi^{-1/2} \sim a^{-1/4}$. Then the terms in square brackets do not scale with a and we find $B \sim a^{-9/4}$ and $v_m \sim a^{-3/4}$. One can check that the energy densities scale in the same way

$$\rho_B = \frac{|\mathbf{B}|^2}{8\pi} \propto a^{-9/2} \tag{28a}$$

$$\rho_{\rm kin.} = \frac{m}{2}|\mathbf{v}_m|^2\, n_m \propto a^{-9/2}\,, \tag{28b}$$

which confirms our earlier argument.

During the matter era we have $\eta = \eta_i + 2(\sqrt{a} - \sqrt{a_i})/H_i a_i^{3/2} \approx 2\sqrt{a}/H_i a_i^{3/2}$, and the solution is

$$\mathbf{B} = \left(\frac{a}{a_i}\right)^{-9/4}\sum_{s=\pm1}\left(\frac{a}{a_i}\right)^{\frac{i}{4}s\sqrt{4\tilde\omega^2\eta_i^2/a_i - 1}}\mathbf{B}_s \tag{29a}$$

$$\mathbf{v}_m = \left(\frac{a}{a_i}\right)^{-3/4}\sum_{s=\pm1}\left(\frac{a}{a_i}\right)^{\frac{i}{4}s\sqrt{4\tilde\omega^2\eta_i^2/a_i - 1}}\mathbf{v}_s\,. \tag{29b}$$

For typical parameters we have $4\tilde\omega^2\eta_i^2/a_i \gg 1$, and the solution is oscillatory with a power law envelope. We find the same anomalous scaling as in the radiation era, *cf.* Eq. (28). If we remove the monopoles from the problem by sending $n_m, \tilde\omega \to 0$ then the would-be oscillatory factors become a power law decay, and we regain the usual scaling $B \sim a^{-2}$ and $v_m \sim a^{-1}$.

This anomalous scaling does not provide constraints on the monopole number density. However, it does affect the way that we translate constraints on the magnetic field in the early universe into the value of the magnetic field today. Measurements of the cosmic microwave background restrict the magnetic field energy density to be less than $\sim 10^{-5}$ of the photon energy density at the time of recombination [28]

$$\rho_B(z_{\rm rec}) \lesssim 10^{-5}\rho_\gamma(z_{\rm rec})\,. \tag{30}$$

Using the scaling relations, $\rho_B \sim (1+z)^{9/2}$ and $\rho_\gamma \sim (1+z)^4$ this inequality implies that the magnetic field energy density today is bounded by

$$\rho_{B,0} \lesssim 10^{-5}\rho_{\gamma,0}(1+z_{\rm rec})^{-1/2} \simeq (3\times10^{-10}\,{\rm G})^2\left(\frac{1+z_{\rm rec}}{1300}\right)^{-1/2} \tag{31}$$

where $\rho_{\gamma,0} \approx 2\pi^2 T_0^4/30$ is the CMB energy density today. If it were not for the anomalous redshifting, the constraint on the B-field strength would be weaker by a factor of $(1+z_{\rm rec})^{1/4} \simeq 6$.

The anomalous scaling of the field strength in Eq. (28) can become a tool in the future as measurements of the PMF improve. By measuring the magnetic field strength at different redshifts, say by using TeV blazars at different distances, we can directly probe the anomalous scaling, and hence obtain a new handle on the relic density of non-relativistic magnetic monopoles.

III. AXIONS

Consider an axion $\varphi(x)$ coupled to the electromagnetic field $A_\mu(x)$. The Lagrangian takes the form [44]

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\varphi)^2 - \frac{m_a^2}{2}\varphi^2 - \frac{1}{16\pi}F_{\mu\nu}F^{\mu\nu} - \frac{g_{a\gamma}}{16\pi}\varphi F_{\mu\nu}\widetilde{F}^{\mu\nu} - A_\mu j^\mu \tag{32}$$

where $\widetilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}$ is the dual field strength tensor, and $j^\mu = (\rho, \mathbf{j})$ is the electromagnetic current arising from the charged Standard Model fields. Our analysis is sufficiently general to apply to any axion or axion-like particle described by Eq. (32), but as a fiducial reference point we will consider a QCD axion with Peccei-Quinn scale of $f_a \simeq 10^{10}\,\text{GeV}$, an axion mass of $m_a \approx \Lambda_{\rm QCD}^2/f_a \simeq 1\,\text{meV}$, and a photon-axion coupling constant $g_{a\gamma} \approx \alpha/(2\pi f_a) \simeq 10^{-13}\,\text{GeV}^{-1}$.

The classical axion condensate obeys the field equation

$$\Box\varphi + m_a^2\varphi = \frac{g_{a\gamma}}{4\pi}\mathbf{E}\cdot\mathbf{B} \tag{33}$$

where we have used $F_{\mu\nu}\widetilde{F}^{\mu\nu} = -4\mathbf{E}\cdot\mathbf{B}$. The electromagnetic field evolves according to the constraint equation

$$\partial_\mu\widetilde{F}^{\mu\nu} = 0 \tag{34}$$

and the modified field equation

$$\partial_\mu F^{\mu\nu} + g_{a\gamma}\,\partial_\mu\varphi\widetilde{F}^{\mu\nu} = \frac{4\pi}{c}j^\nu\,. \tag{35}$$

In terms of the electric and magnetic vector fields we have

$$\boldsymbol{\nabla}\cdot\mathbf{B} = 0 \tag{36a}$$

$$\boldsymbol{\nabla}\times\mathbf{E} + \frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} = 0 \tag{36b}$$

$$\boldsymbol{\nabla}\cdot\mathbf{E} + g_{a\gamma}\boldsymbol{\nabla}\varphi\cdot\mathbf{B} = 4\pi\rho \tag{36c}$$

$$\boldsymbol{\nabla}\times\mathbf{B} - \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t} - g_{a\gamma}\dot{\varphi}\mathbf{B} - g_{a\gamma}\boldsymbol{\nabla}\varphi\times\mathbf{E} = \frac{4\pi}{c}\mathbf{j}\,. \tag{36d}$$

Note the presence of the additional terms arising from the spatio-temporal variation of the axion field.

We seek to study the coevolution of the coupled axion and electromagnetic fields. Eqs. (33) and (36) describe a non-dissipative system. Dissipation is introduced as the electromagnetic field couples to charged particles in the cosmological medium, which opens an avenue for energy to be lost in the form of heat. This coupling is parametrized by the conductivity σ which appears in Ohm's law

$$\boldsymbol{j} = \sigma\,(\mathbf{E} + \mathbf{v}\times\mathbf{B}) \tag{37}$$

where $\mathbf{v}(t,\mathbf{x})$ is the local velocity of the plasma. Prior to the epoch of e^+e^- annihilation, free charge carriers were abundant and the cosmological medium had a high conductivity [45]

$$\sigma \approx T/\alpha \tag{38}$$

where $\alpha \simeq 1/137$ is the fine structure constant. Ohm's law allows us to eliminate $\mathbf{j}$ and thereby reduce the system of equations in four unknowns $\{\mathbf{E},\mathbf{B},\mathbf{j},\varphi\}$ to a set of equations describing only three unknowns:

$$\ddot{\varphi} - \nabla^2\varphi + m_a^2\varphi = \frac{g_{a\gamma}}{4\pi}\mathbf{E}\cdot\mathbf{B} \tag{39a}$$

$$\dot{\mathbf{B}} = -\boldsymbol{\nabla}\times\mathbf{E} \tag{39b}$$

$$\dot{\mathbf{E}} = \boldsymbol{\nabla}\times\mathbf{B} - g_{a\gamma}\dot{\varphi}\mathbf{B} - g_{a\gamma}\boldsymbol{\nabla}\varphi\times\mathbf{E} - 4\pi\sigma\mathbf{E} - 4\pi\sigma\mathbf{v}\times\mathbf{B}\,. \tag{39c}$$

In the MHD approximation (nonrelativistic flow) we can neglect the displacement current since it is negligible compared to the curl of the magnetic field, $|\dot{\mathbf{E}}|/|\boldsymbol{\nabla}\times\mathbf{B}| \sim (v/c)^2 \ll 1$ [46]. Then Eq. (39c) becomes algebraic in $\mathbf{E}$, and we can solve it to eliminate $\mathbf{E}$ from the remaining equations. Focusing now on a homogenous axion field, the system of equations reduces to

$$\ddot{\varphi} + g_{a\gamma}^2\frac{\eta_{\rm d}|\mathbf{B}|^2}{4\pi}\dot{\varphi} + m_a^2\varphi = \frac{g_{a\gamma}\eta_{\rm d}}{4\pi}\mathbf{B}\cdot\boldsymbol{\nabla}\times\mathbf{B} \tag{40a}$$

$$\dot{\mathbf{B}} = \boldsymbol{\nabla}\times(\mathbf{v}\times\mathbf{B}) + \eta_{\rm d}\,\nabla^2\mathbf{B} + g_{a\gamma}\eta_{\rm d}\dot{\varphi}\,\boldsymbol{\nabla}\times\mathbf{B} \tag{40b}$$

where

$$\eta_{\rm d} \equiv \frac{1}{4\pi\sigma} \approx \frac{\alpha}{4\pi T} \tag{41}$$

is the magnetic diffusivity, assumed to be homogenous. Eqs. (40a) and (40b) together with the Navier-Stokes equation for the plasma velocity $\mathbf{v}$ are the final equations to be solved. We first solve Eq. (40b) to determine the effect of the axion on the magnetic field, and afterward we will consider the evolution of the axion according to Eq. (40a).

A. Effect of Axion on Magnetic Field

In order to solve Eq. (40b) for the B-field we must know the fluid velocity, which appears in the advection term,

$\boldsymbol{\nabla} \times (\mathbf{v} \times \mathbf{B})$. Since our primary interest is in the co-evolution of the magnetic field and the axion, not in the magnetohydrodynamics, we will neglect this term [61]. Since Eq. (40b) is linear in $\mathbf{B}$, and we assume that φ is homogenous, we can solve the equation by first performing a Fourier transform. For a given mode $\mathbf{k}$ let $(\mathbf{e}_1(\mathbf{k}), \mathbf{e}_2(\mathbf{k}), \mathbf{e}_3(\mathbf{k}))$ form a right-handed, orthonormal triad of unit vectors with $\mathbf{e}_3(\mathbf{k}) = \mathbf{k}/|\mathbf{k}|$. It is convenient to introduce the right- and left-circular polarization vectors by

$$\mathbf{e}^{\pm}(\mathbf{k}) = \frac{\mathbf{e}_1(\mathbf{k}) \pm i\mathbf{e}_2(\mathbf{k})}{\sqrt{2}} \,. \tag{42}$$

Note that $i\mathbf{k} \times \mathbf{e}^{\pm}(\mathbf{k}) = \pm|\mathbf{k}|\mathbf{e}^{\pm}(\mathbf{k})$. The mode decomposition is given by

$$\mathbf{B}(t, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}} \sum_{\mathrm{s}=\pm} b_{\mathrm{s}}(t, |\mathbf{k}|)\mathbf{e}^{\mathrm{s}}(\mathbf{k}) \,. \tag{43}$$

With this replacement, Eq. (40b) becomes

$$\dot{b}_{\pm}(t,k) = -\eta_{\mathrm{d}} k^2 b_{\pm}(t,k) \pm g_{a\gamma}\eta_{\mathrm{d}}\, k\, \dot{\varphi}\, b_{\pm}(t,k) \tag{44}$$

where we have written $k = |\mathbf{k}|$. The last term of this equation, essentially the chiral-magnetic effect [47], has been studied previously in the context of axions [48] and cosmology [49, 50]. The solution is

$$b_{\pm}(t,k) = b_{\pm}(t_i,k)\; e^{-k^2(t-t_i)\eta_{\mathrm{d}}}\; e^{\pm k/k_{\mathrm{ax}}(t)} \tag{45}$$

where we have defined the wavenumber

$$k_{\mathrm{ax}}(t) \equiv \frac{2}{g_{a\gamma}\Delta\varphi(t)\eta_{\mathrm{d}}} \tag{46}$$

and $\Delta\varphi(t) = \varphi(t) - \varphi(t_i)$ is the change in the axion field. The prefactor in Eq. (45) is the initial spectrum of the magnetic field which will depend on the PMF generation mechanism. The first exponential is the usual diffusive decay term, which exponentially suppresses modes on a length scale shorter than $k_{\mathrm{diff}}^{-1} = \sqrt{(t-t_i)\eta_{\mathrm{d}}} \simeq 10^{-1}\sqrt{t/T}$. The second exponential only kicks in at small length scales where $k > k_{\mathrm{ax}}$. Then it leads to a suppression of one polarization mode and enhancement of the other, depending on the sign of $\Delta\varphi(t)$.

The value of $\Delta\varphi(t)$ depends on the solution for the axion field as well as the initial time t_i. We expect that the misalignment mechanism sets the initial condition $\varphi(t_i) \sim \pm f_a$. The subsequent evolution is determined by solving Eq. (40a), which we will turn to in the next section. For the moment we will assume that the axion evolution is not significantly affected by the presence of the magnetic field, and the solution is the standard one: the axion remains "frozen" at $\varphi(t_i)$ until the time of the QCD phase transition when it begins to oscillate around $\varphi = 0$ with angular frequency $\omega = m_a$ [51]. Then we can approximate

$$\Delta\varphi(t) \approx \begin{cases} 0 & t < t_{\mathrm{QCD}} \\ \mathrm{s} f_a & t > t_{\mathrm{QCD}} \end{cases} \tag{47}$$

where $\mathrm{s} = \mathrm{sign}[\varphi(t_i)]$. Using this approximation we can estimate k_{ax}. Prior to the QCD phase transition, $\Delta\varphi$ is small and k_{ax} is large, meaning that none of the modes receive the enhancement or suppression from the axion coupling. This is reasonable since the axion is derivatively coupled, and as long as it is stationary there will be no effect on the magnetic field. After the QCD transition, we can estimate

$$k_{\mathrm{ax}} \approx \frac{4\pi\sigma}{g_{a\gamma}f_a} \approx \frac{4\pi T}{\alpha g_{a\gamma} f_a} \tag{48}$$

using Eqs. (38) and (41). Note that this result is insensitive to the Peccei-Quinn scale, and as long as $g_{a\gamma} = \alpha/(2\pi f_a)$ we have $k_{\mathrm{ax}}^{-1} \approx \alpha^2/(8\pi^2 T) \simeq 10^{-6} T^{-1}$.

The solution in Eq. (45) can also be written as

$$b_{\pm}(t,k) = b_{\pm}(t_i,k)\; e^{K^2(t-t_i)\eta_{\mathrm{d}}}\; e^{-(k\mp K)^2(t-t_i)\eta_{\mathrm{d}}} \tag{49}$$

where

$$K(t) \equiv \frac{1}{k_{\mathrm{ax}}(t-t_i)\eta_{\mathrm{d}}} = \frac{g_{a\gamma}\Delta\varphi(t)}{2(t-t_i)} \,. \tag{50}$$

This representation of the solution is convenient, because all the spectral information is contained in the second factor. One of the helicity modes has a Gaussian spectrum peaked at $|K(t)| > 0$ with width $\sqrt{1/(t-t_i)\eta_{\mathrm{d}}}$, and the other helicity mode peaks at $k = 0$. Estimating k_{ax} as above, we find that the associated length scale of the spectral peak corresponds to $K^{-1} \approx k_{\mathrm{ax}} t\eta_{\mathrm{d}} \simeq 600t$, which is larger than the scale of the cosmological horizon $d_H \sim t$.

It appears that the presence of an axion condensate coupled to electromagnetism has a negligible impact on the evolution of a primordial magnetic field, unless there are situations in which $\Delta\varphi$ can be much larger than f_a.

We note that our analysis ignores the possibility of turbulence in the primordial plasma. It would be of interest to include both turbulence and the axion coupling in future studies.

B. Effect of Magnetic Field on Axion

Next we will investigate the effect of a background magnetic field on the axion condensate. We have seen that the magnetic field is approximately unmodified on length scales larger than the diffusion length, $k_{\mathrm{diff}}^{-1} \sim \sqrt{\eta_{\mathrm{d}} t}$ (*cf.*, Eq. (45)). In this regime Eq. (40a) can be rewritten as

$$\ddot{\varphi} + 2\frac{\dot{\varphi}}{\tau_{\mathrm{decay}}} + \frac{\varphi}{\tau_a^2} = \mathcal{H} \tag{51}$$

where

$$\tau_{\mathrm{decay}} \equiv \frac{8\pi}{g_{a\gamma}^2 \eta_{\mathrm{d}} \langle|\mathbf{B}|^2\rangle} \,, \tag{52}$$

$$\tau_a \equiv \frac{1}{m_a} \,, \tag{53}$$

$$\mathcal{H} \equiv \frac{g_{a\gamma}\eta_{\mathrm{d}}}{4\pi} \langle \mathbf{B} \cdot \boldsymbol{\nabla} \times \mathbf{B} \rangle, \tag{54}$$

and the angled brackets $\langle\cdot\rangle$ denote spatial averaging. The axion condensate evolves like a damped and driven harmonic oscillator, where the damping and driving forces are induced by the magnetic field background. As we discuss below, it is interesting that the driving force is associated with the *helicity* of the magnetic field.

The magnetic-induced damping of axion oscillations is parametrized by the time scale $\tau_{\rm decay}$. To determine when this damping will be relevant for the evolution of the axion, we compare it with the cosmological time scale, given by Eq. (14). To express $\langle|\mathbf{B}|^2\rangle = B^2$ in terms of the magnetic field strength today, B_0, we use $B = B_0(a_0/a)^2 \simeq 10B_0(T/T_0)^2$ where the factor of 10 is related to the number of relativistic degrees of freedom in the early universe and today. Then the ratio is found to be

$$\frac{\tau_{\rm decay}}{t_H} \approx \frac{16\pi^2 g_*}{75}\frac{T_0^4}{\alpha g_{a\gamma}^2 M_P B_0^2 T} \simeq 10^{16}\frac{(10^{-13}\,{\rm GeV}^{-1})^2}{g_{a\gamma}^2}\frac{(10^{10}\,{\rm GeV})}{T}\frac{(10^{-14}\,{\rm G})^2}{B_0^2}\,. \tag{55}$$

This estimate suggests that the magnetic-induced decay of the axion field is negligible for a typical Peccei-Quinn scale and B-field strength. If the B-field strength today were as large as $B_0 \sim 10^{-9}\,{\rm G}$ and the Peccei-Quinn scale was as low as $f_a \sim$ TeV, then $\tau_{\rm decay}$ would be comparable to the Hubble time at $T \approx f_a$. As the temperature decreases, the magnetic-induced decay becomes less relevant.

It is interesting that the magnetic field also induces a driving force, parametrized by $\mathcal{H}$. The pseudoscalar $\mathcal{H}$ is related to the helicity of the magnetic field. This is perhaps more evident from the initial form of the axion field equation, Eq. (33), where $\mathbf{E}\cdot\mathbf{B}$ is equal to the rate of change of the helicity density $-(1/2)d(\mathbf{A}\cdot\mathbf{B})/dt$ plus a divergence, which vanishes upon spatial averaging. If the power in the magnetic field is localized on a particular length scale λ_B we can estimate $\langle\mathbf{B}\cdot\boldsymbol{\nabla}\times\mathbf{B}\rangle \sim B^2/\lambda_B \sim 300(B_0^2/\lambda_{B_0})(T/T_0)^5$ where we used $\lambda_B \sim 3\lambda_{B,0}(T_0/T)$.

Prior to the QCD phase transition we can neglect the mass and drag terms in Eq. (51), and the solution is simply $\varphi = \mathcal{H}t^2/2$. Since the axion is massless, there is no restorative potential, and the helical magnetic field leads to an unbounded growth of the axion condensate. Although this analysis neglects the Hubble drag, we can estimate the maximum field excursion in one one Hubble time to be $\mathcal{H}t_H^2/2$. Comparing with the Peccei-Quinn breaking scale, the corresponding angular excursion is

$$\Delta\theta \approx \frac{\mathcal{H}t_H^2}{f_a} \approx \frac{75}{16\pi^2}\frac{\alpha g_{a\gamma} M_P^2 B_0^2}{f_a g_*^2 T_0^5 \lambda_{B,0}} \simeq 10^{-35}\left(\frac{B_0}{10^{-14}\,{\rm G}}\right)^2\left(\frac{\lambda_{B,0}}{10\,{\rm Mpc}}\right)^{-1} \times\left(\frac{g_{a\gamma}}{10^{-13}\,{\rm GeV}^{-1}}\right)\left(\frac{f_a}{10^{10}\,{\rm GeV}}\right)^{-1}. \tag{56}$$

We are led to conclude that for realistic parameters, the helical PMF does not significantly impact the evolution of the axion condensate prior to the QCD phase transition.

After the QCD phase transition, the axion mass reaches its asymptotic value, and the source term displaces the minimum of the axion potential from $\varphi = 0$ to $\varphi_{\rm min} = \mathcal{H}\tau_a^2 = \mathcal{H}/m_a^2$. In terms of the angular coordinate:

$$\theta_{\rm min} \approx \frac{\mathcal{H}\tau_a^2}{f_a} \approx \frac{25}{12\pi^2}\frac{\alpha g_{a\gamma} B_0^2 T^4}{f_a m_a^2 \lambda_{B,0} T_0^5} \simeq 10^{-47}\left(\frac{B_0}{10^{-14}\,{\rm G}}\right)^2\left(\frac{\lambda_{B,0}}{10\,{\rm Mpc}}\right)^{-1}\left(\frac{T}{200\,{\rm MeV}}\right)^4 \times\left(\frac{f_a}{10^{10}\,{\rm GeV}}\right)^{-1}\left(\frac{g_{a\gamma}}{10^{-13}\,{\rm GeV}^{-1}}\right)\left(\frac{m_a}{1\,{\rm meV}}\right)^{-2}. \tag{57}$$

The temperature dependence enters through $B \sim T^2$, $\lambda_B \sim 1/T$, and $\eta_{\rm d} \sim 1/T$, and the fractional shift is largest at high temperature where B is large and λ_B is small. Immediately after the QCD phase transition, $T_{\rm QCD} \sim 200\,{\rm MeV}$, the fractional shift is already extremely small. Moreover if the PMF is not helical then $\mathcal{H} = 0$ and there is no shift in the axion potential.

It is interesting that the estimate of Eq. (57) is insensitive to the Peccei-Quinn scale; as long as $g_{a\gamma} = \alpha/(2\pi f_a)$ and $m_a = \Lambda_{\rm QCD}^2/f_a$ we have $g_{a\gamma}/f_a m_a^2 = \alpha/(2\pi\Lambda_{\rm QCD}^4)$. Then the primarily challenge toward obtaining a large effect is the smallness of the magnetic field strength. Although unrelated to primordial magnetic fields, which is the motivation for this work, it would be interesting to study the axion condensate in an astrophysical system where the magnetic field is both helical and strong. For instance, the field strength in a magnetar can grow as large as $B \sim 10^{15}\,{\rm G}$ and the magnetic field in some astrophysical jets is known to be helical [52].

C. Axion-Photon Interconversion

Until this point we have focused our attention on the interplay between the axion condensate and the primordial magnetic field, and we now turn our attention to the quanta of these fields. In the presence of a background magnetic field, the interaction in Eq. (32) yields a mixing between axion particles and photons. Typically the conversion is inefficient, but in the presence of a plasma the photon acquires an effective mass, and the conversion probability experiences a resonance when $m_\gamma = m_a$ [53]. In the cosmological context, the conversion of photons into axions may lead to a dimming of the cosmic microwave background across frequencies. Then measurements of the spectrum of the CMB can be used to place constraints on the axion-photon coupling and the magnetic field strength.

Bounds were obtained from the COBE / FIRAS measurement of the CMB spectrum in Ref. [54], and recently a second group [55] has extended the calculation

to include forecasts for next-generation CMB telescopes, namely PIXIE and PRISM. The latter references finds an upper bound on the product of the axion-photon coupling and the r.m.s. magnetic field strength today:

$$\begin{aligned} g_{a\gamma} B_0 &< 10^{-14}\,\mathrm{GeV}^{-1}\,\mathrm{nG} \quad \text{(COBE-FIRAS data)} \\ g_{a\gamma} B_0 &< 10^{-16}\,\mathrm{GeV}^{-1}\,\mathrm{nG} \quad \text{(PIXIE/PRISM forecast)} \end{aligned} \tag{58}$$

for a light axion $m_a < 10^{-14}\,\mathrm{eV}$. For larger axions masses the bound weakens. Using our fiducial value for the magnetic field $B_0 \simeq 10^{-14}\,\mathrm{G}$, we can write

$$\begin{aligned} g_{a\gamma} &< 10^{-9}\,\mathrm{GeV}^{-1} \left(\frac{B_0}{10^{-14}\,\mathrm{G}}\right)^{-1} \quad \text{(COBE-FIRAS)} \\ g_{a\gamma} &< 10^{-11}\,\mathrm{GeV}^{-1} \left(\frac{B_0}{10^{-14}\,\mathrm{G}}\right)^{-1} \quad \text{(PIXIE/PRISM)}\,. \end{aligned} \tag{59}$$

These bounds are comparable to the direct search limits from the CAST helioscope [56],

$$g_{a\gamma} < 8.8 \times 10^{-11}\,\mathrm{GeV}^{-1} \tag{60}$$

for $m_a \lesssim 0.02\,\mathrm{eV}$.

IV. DIRAC NEUTRINOS

While the neutrinos are known to be massive particles, the nature of their mass remains a mystery. If neutrinos are Dirac particles then the theory contains four light states per generation: an active neutrino ν_L, an active antineutrino, $\bar{\nu}_R$, a sterile neutrino ν_R, and a sterile antineutrino $\bar{\nu}_L$. The active states interact through the weak force, and this allows them to come into thermal equilibrium in the early universe. The sterile states, on the other hand, interact only via the Yukawa interaction with the Higgs boson, and because of the smallness of the Yukawa coupling $y_\nu \sim m_\nu/v \sim 10^{-12}$, these states are not expected to be populated.

This story is modified if a strong magnetic field permeated the early universe. The nonzero neutrino mass implies that the neutrino will also have a nonzero magnetic moment μ_ν. From Standard Model physics alone one expects [57, 58]

$$\mu_\nu^{\rm SM} \simeq (3 \times 10^{-20}\mu_B)\frac{m_\nu}{0.1\,\mathrm{eV}}\,, \tag{61}$$

where $\mu_B \equiv e/2m_e \simeq 83.6\,\mathrm{GeV}^{-1}$ is the Bohr magneton, but new physics can increase this value appreciably. The magnetic field couples to μ_ν and induces the spin-flip transitions $\nu_L \to \nu_R$ and $\bar{\nu}_R \to \bar{\nu}_L$, which can be viewed as the absorption or emission of a photon. If the spin-flip occurs rapidly in the early universe, the sterile states would be populated, and the effective number of relativistic neutrino species would double from $N_\nu = 3$ to 6. However, this is not consistent with measured abundances of the light elements, which imply $N_\nu \approx 3$ at the time of nucleosynthesis [59]. We must therefore require that the spin-flip transition goes out of equilibrium prior to the QCD epoch, $T_{\rm QCD} \simeq 200\,\mathrm{MeV}$, so that the subsequent entropy injection at the QCD phase transition can suppress the relative abundance of sterile states to acceptable levels [51]. This translates into an upper bound on the neutrino magnetic moment and magnetic field strength, which was originally discussed by Enqvist et al. [36, 37].

In the rest of this section we apply the results of Ref. [37]. The spin-flip transition occurs with a rate

$$\Gamma_{L\to R} = \langle P_{\nu_L\to\nu_R}\rangle \Gamma_W^{\rm tot} \tag{62}$$

where $\langle P_{\nu_L\to\nu_R}\rangle$ is the average conversion probability and $\Gamma_W^{\rm tot}$ is the total weak scattering rate. The active neutrinos scatter via the weak interaction which leads to $\Gamma_W^{\rm tot} \simeq 30 G_F^2 T_{\rm QCD}^5$ at the QCD epoch. The conversion probability depends on the magnetic moment and field strength as $\langle P_{\nu_L\to\nu_R}\rangle \propto \mu_\nu^2 B^2$, since the interaction Hamiltonian is $H_{\rm int} = -\boldsymbol{\mu}_\nu \cdot \mathbf{B}$. The coefficient takes different values depending on the relative scale of the magnetic field domains λ_B and the weak collision length $L_W \approx (\Gamma_W^{\rm tot})^{-1}$. At the QCD epoch $L_W \simeq 1.6 \times 10^{-2}\,\mathrm{cm}$, which corresponds to a length scale of $L_{W,0} \simeq 3 \times 10^{10}\,\mathrm{cm}$ today. It is safe to assume that the magnetic field of interest is much larger than this length scale, and therefore $\lambda_B \gg L_W$. To ensure that the spin-flip transition is out of equilibrium one must impose $\Gamma_{L\to R} < H$ with H the Hubble parameter. This inequality resolves to the bound (see Eq. (37) of Ref. [37])

$$\mu_\nu B(t_{\rm QCD}) < (3.5 \times 10^2 \mu_B\,\mathrm{G})\sqrt{\frac{L_W}{\lambda_B}}\,. \tag{63}$$

To express this inequality in terms of the B-field strength and correlation length today, we use $B \simeq 6B_0(T_{\rm QCD}/T_0)^2$ and $L_W/\lambda_B = L_{W,0}/\lambda_{B,0}$. This leads to an upper bound on the neutrino magnetic moment:

$$\mu_\nu < (3 \times 10^{-16}\mu_B)\left(\frac{B_0}{10^{-14}\,\mathrm{G}}\right)^{-1}\left(\frac{\lambda_{B,0}}{10\,\mathrm{Mpc}}\right)^{-1/2}\,. \tag{64}$$

If this bound is not satisfied, the sterile neutrino states will still be thermalized with the active neutrino states at the time of BBN leading to $N_\nu \approx 6$, which is inconsistent with the data. If the neutrinos are Majorana particles, then the sterile states are much heavier, and this bound does not apply.

The bound in Eq. (64) is represented graphically in Fig. 2. For comparison we show the SM prediction from Eq. (61) and the direct search limits. The strongest laboratory constraints arise from elastic $\nu - e$ scattering. The limits are flavor-dependent, but they are typically at the level of [42]

$$\mu_\nu \lesssim 10^{-10}\mu_B \qquad \text{(direct)}\,. \tag{65}$$

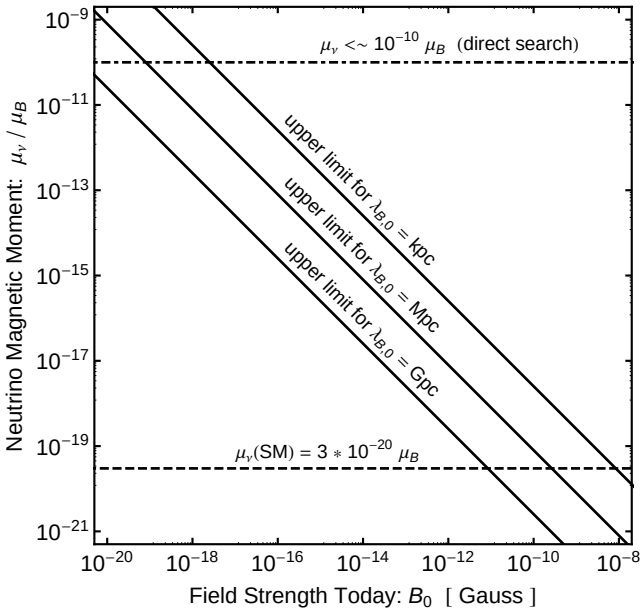

FIG. 2: The requirement that spin-flip transitions are out of equilibrium at the QCD epoch leads to an upper bound the neutrino magnetic moment given by Eq. (64).

From the figure we see that the indirect early universe constraint is significantly stronger than the direct constraint for $B \gtrsim 10^{-18}$ G. This provides the exciting opportunity to constrain extensions of the SM that predict an enhancement to the magnetic moment of Dirac neutrinos.

V. SUMMARY

Growing evidence for the existence of an intergalactic magnetic field has motivated us to consider the effects of a primordial magnetic field on models of exotic particle physics in the early universe. We have focused our study on magnetic monopoles, axions, and Dirac neutrinos with a magnetic moment. We summarize our results here.

In the context of a universe containing relic magnetic monopoles, we have derived a "primordial Parker bound" by requiring the survival of a primordial magnetic field until the time of electron-positron annihilation. The bound, which appears in Eq. (16), gives an upper limit on the cosmological monopole number density today: $n_0 < 1 \times 10^{-20}\,\mathrm{cm}^{-3}$. This translates into an upper bound on the monopole flux in the Milky Way; if the monopoles are unclustered then $\mathcal{F} < 3{\times}10^{-14}\,\mathrm{cm}^{-2}\,\mathrm{sec}^{-1}\,\mathrm{sr}^{-1}(v/10^{-3})$, and if they are clustered the bound weakens by a factor of $\sim 10^5$. In Fig. 1 we compare the primordial Parker bound with other constraints on relic monopoles. If the primordial magnetic field is not generated prior to $T \simeq$ MeV, then this bound does not apply.

After e^+e^- annihilation the monopoles are able to free stream, and they evolve along with the magnetic field as described by the system of equations in Eq. (24). The solution is an analog of the familiar plasma oscillations ("Langmuir oscillations") seen in an electron-ion plasma. In the regime where the plasma oscillations are fast compared to the cosmological expansion, the coupling of the monopoles to the magnetic field leads to an anomalous scaling with redshift such that $B \sim a^{-9/4}$, $v_m \sim a^{-3/4}$, and $\rho_B \sim \rho_{\mathrm{kin.}} \sim a^{-9/2}$. The behavior of the coupled system is effectively the average of the usual scalings for radiation $\rho_B \sim a^{-4}$ and the kinetic energy of a non-relativistic gas $\rho_{\mathrm{kin.}} \sim a^{-5}$. If the strength of the intergalactic magnetic field could be measured over a range of redshifts, this would allow for a direct test of the anomalous scaling, and thereby probe relic magnetic monopoles.

We have also studied the effect of a primordial magnetic field on the evolution of an axion condensate in the early universe. We obtain an exact solution to the MHD equations for the magnetic field in the limit where the advection term is negligible and the axion is homogenous. After Peccei-Quinn breaking but prior to the QCD phase transition, the axion field is frozen, because its mass is smaller than the Hubble scale, and since the axion is derivatively coupled, this leads to no effect on the magnetic field. Below the QCD scale the axion field begins to oscillate, and the spectrum of the magnetic field is distorted as in Eq. (45). One helicity mode of the magnetic field is enhanced while the other is suppressed; this CP-violation is a consequence of the axion's pseudoscalar nature. However, the spectral shape of the magnetic field is only affected on extremely large length scales, as given by Eq. (50), except in situations where there can be significant axion evolution prior to the QCD epoch.

We next study the evolution of the homogenous axion condensate in the presence of a background magnetic field. The axion behaves as a damped and driven harmonic oscillator, as seen from its equation of motion Eq. (51). The damping time scale depends on the strength of the magnetic field and the photon-axion coupling. For typical parameters it is generally larger than the cosmological time scale, and therefore irrelevant for the evolution of the axion. It is interesting that the driving force (source term $\mathcal{H}$) is only operative when the magnetic field has a helicity. This can be seen directly from the interaction $\mathcal{L} \ni \varphi F\widetilde{F}$ where $F\widetilde{F} \sim \mathbf{E}\cdot\mathbf{B} \sim \dot{h}$ is related to the rate of change of magnetic helicity $h = \mathbf{A}\cdot\mathbf{B}$. Prior to the QCD phase transition when the axion was effectively massless, the axion field equation reduces to $\ddot{\varphi} = \mathcal{H}$. In principle a very strong magnetic field could cause the axion to grow without bound as $\varphi(t) = \mathcal{H}t^2/2$, by drawing energy from the magnetic field. For typical parameters, however, this growth occurs on a time scale that is much longer than the cosmological time. It may still be the case that helical magnetic fields occurring in astrophysical environments are strong enough to lead to observable signatures.

We have also considered the resonant conversion of

CMB photons into axions, which leads to a distortion of the CMB blackbody spectrum [53–55]. Using constraints on spectral distortions from current and anticipated future CMB telescopes, Ref. [55] obtained an upper bound on the axion-photon coupling. For our fiducial magnetic field strength this translates into $g_{a\gamma} \lesssim 10^{-9}\,\mathrm{GeV}^{-1}$ with current data, and a forecast of $g_{a\gamma} \lesssim 10^{-11}\,\mathrm{GeV}^{-1}$ for experiments presently under discussion (see Eq. (59)).

Finally we turn to the effect of the primordial magnetic field on Dirac neutrinos, which carry a magnetic moment. In the presence of a magnetic field, left-handed neutrinos can be converted into right-handed neutrinos. If this spin-flip process is in equilibrium in the early universe, the right-handed states would be populated, and the effective number of relativistic neutrino species would double from 3 to 6, which is inconsistent with observations. Requiring that this process is out of equilibrium at the time of the QCD phase transition leads to an upper bound on the neutrino magnetic moment and magnetic field strength. Drawing on the work of Ref. [37], we find the limit in Eq. (64), which implies $\mu_\nu < 3 \times 10^{-16}\mu_B$ for our fiducial magnetic field parameters $B_0 = 10^{-14}\,\mathrm{G}$ and $\lambda_B = 10\,\mathrm{Mpc}$. As seen in Fig. 2, this bound is significantly stronger than the direct search limits over most of the parameter space.

Acknowledgments

We are grateful to David Marsh and Hiroyuki Tashiro for comments. TV gratefully acknowledges the Clark Way Harrison Professorship at Washington University during the course of this work. This work was supported by the DOE at ASU. A.J.L. was also supported in part by the National Science Foundation under grant number PHY-1205745.

[1] R. Durrer and A. Neronov, Astron.Astrophys.Rev. **21**, 62 (2013), 1303.7121.
[2] M. S. Turner and L. M. Widrow, Phys.Rev. **D37**, 2743 (1988).
[3] T. Vachaspati, Phys.Lett. **B265**, 258 (1991).
[4] J. Ahonen and K. Enqvist, Phys.Rev. **D57**, 664 (1998), hep-ph/9704334.
[5] J. M. Cornwall, Phys.Rev. **D56**, 6146 (1997), hep-th/9704022.
[6] T. Vachaspati, Phys.Rev.Lett. **87**, 251302 (2001), astro-ph/0101261.
[7] A. J. Long, E. Sabancilar, and T. Vachaspati, JCAP **1402**, 036 (2014), 1309.2315.
[8] M. J. Rees, Quarterly Journal, Royal Astronomical Society **28**, 197 (1987).
[9] Y. B. Zeldovich, A. A. Ruzmaikin, and D. D. Sokoloff, *Magnetic Fields in Astrophysics* (Gordon and Breach Science Pub, 1990).
[10] F. A. Aharonian, P. S. Coppi, and H. J. Voelk, Ap. J. Lett. **423**, L5 (1994), astro-ph/9312045.
[11] R. Plaga, Nature **374**, 430 (1995).
[12] A. Neronov and D. V. Semikoz, JETP Lett. **85**, 473 (2007), astro-ph/0604607.
[13] A. Elyiv, A. Neronov, and D. Semikoz, Phys.Rev. **D80**, 023010 (2009), 0903.3649.
[14] K. Dolag, M. Kachelriess, S. Ostapchenko, and R. Tomas, Astrophys.J. **703**, 1078 (2009), 0903.2842.
[15] A. Neronov and I. Vovk, Science **328**, 73 (2010), 1006.3504.
[16] F. Tavecchio, G. Ghisellini, L. Foschini, G. Bonnoli, G. Ghirlanda, et al., Mon.Not.Roy.Astron.Soc. **406**, L70 (2010), 1004.1329.
[17] K. Dolag, M. Kachelriess, S. Ostapchenko, and R. Tomas, Astrophys.J. **727**, L4 (2011), 1009.1782.
[18] C. D. Dermer, M. Cavadini, S. Razzaque, J. D. Finke, and B. Lott, Astrophys.J. **733**, L21 (2011), 1011.6660.
[19] A. E. Broderick, P. Chang, and C. Pfrommer, Astrophys.J. **752**, 22 (2012), 1106.5494.
[20] F. Miniati and A. Elyiv, Astrophys.J. **770**, 54 (2013), 1208.1761.
[21] R. Schlickeiser, D. Ibscher, and M. Supsar, Astrophys.J. **758**, 102 (2012).
[22] S. Ando and A. Kusenko, Astrophys.J. **722**, L39 (2010), 1005.1924.
[23] A. Neronov, D. Semikoz, P. Tinyakov, and I. Tkachev, Astron.Astrophys. **526**, A90 (2011), 1006.0164.
[24] J. Aleksic, L. Antonelli, P. Antoranz, M. Backes, C. Baixeras, et al., Astron.Astrophys. **524**, A77 (2010), 1004.1093.
[25] M. F. Alonso (VERITAS) (2014), 1406.4764.
[26] A. Abramowski et al. (HESS), Astron.Astrophys. **562**, 145 (2014), 1401.2915.
[27] W. Chen, J. H. Buckley, and F. Ferrer (2014), 1410.7717.
[28] P. Ade et al. (Planck) (2015), 1502.01594.
[29] C. Caprini and L. Sorbo, JCAP **1410**, 056 (2014), 1407.2809.
[30] T. Kahniashvili, A. G. Tevzadze, A. Brandenburg, and A. Neronov, Phys.Rev. **D87**, 083007 (2013), 1212.0596.
[31] H. Tashiro, W. Chen, F. Ferrer, and T. Vachaspati, Monthly Notices of the Royal Astronomical Society: Letters 2014 pp. 445 (1): L41–L45 (2013), 1310.4826.
[32] W. Chen, B. D. Chowdhury, F. Ferrer, H. Tashiro, and T. Vachaspati (2014), 1412.3171.
[33] H. Tashiro and T. Vachaspati, Phys.Rev. **D87**, 123527 (2013), 1305.0181.
[34] H. Tashiro and T. Vachaspati, Mon.Not.Roy.Astron.Soc. **448**, 299 (2015), 1409.3627.
[35] B. D. Chowdhury, A. J. Long, and T. Vachaspati (in preparation).
[36] K. Enqvist, P. Olesen, and V. Semikoz, Phys.Rev.Lett. **69**, 2157 (1992).
[37] K. Enqvist, A. Rez, and V. Semikoz, Nucl.Phys. **B436**, 49 (1995), hep-ph/9408255.
[38] E. N. Parker, Astrophys.J. **160**, 383 (1970).
[39] J. T. Goldman, E. W. Kolb, and D. Toussaint, Phys.Rev. **D23**, 867 (1981).

[40] A. Vilenkin and Shellard, *Cosmic Strings and Other Topological Defects* (Cambridge University Press, Cambridge, UK, 1994).

[41] F. C. Adams, M. Fatuzzo, K. Freese, G. Tarle, R. Watkins, et al., Phys.Rev.Lett. **70**, 2511 (1993).

[42] K. Olive et al. (Particle Data Group), Chin.Phys. **C38**, 090001 (2014).

[43] M. S. Turner, E. N. Parker, and T. Bogdan, Phys.Rev. **D26**, 1296 (1982).

[44] J. E. Kim and G. Carosi, Rev.Mod.Phys. **82**, 557 (2010), 0807.3125.

[45] G. Baym and H. Heiselberg, Phys.Rev. **D56**, 5254 (1997), astro-ph/9704214.

[46] A. R. Choudhuri, *The Physics of Fluids and Plasmas: An Introduction for Astrophysicists* (Cambridge University Press, 1998).

[47] A. Vilenkin, Phys.Rev. **D22**, 3080 (1980).

[48] G. B. Field and S. M. Carroll, Phys.Rev. **D62**, 103008 (2000), astro-ph/9811206.

[49] M. Joyce and M. E. Shaposhnikov, Phys.Rev.Lett. **79**, 1193 (1997), astro-ph/9703005.

[50] H. Tashiro, T. Vachaspati, and A. Vilenkin, Phys.Rev. **D86**, 105033 (2012), 1206.5549.

[51] E. W. Kolb and M. S. Turner, *The Early Universe* (Westview Press, 1990).

[52] R. E. Pudritz, M. J. Hardcastle, and D. C. Gabuzda, Space Science Reviews **169**, 27 (2012), 1205.2073.

[53] T. Yanagida and M. Yoshimura, Phys.Lett. **B202**, 301 (1988).

[54] A. Mirizzi, G. G. Raffelt, and P. D. Serpico, Phys.Rev. **D72**, 023501 (2005), astro-ph/0506078.

[55] H. Tashiro, J. Silk, and D. J. E. Marsh, Phys.Rev. **D88**, 125024 (2013), 1308.0314.

[56] S. Andriamonje et al. (CAST), JCAP **0704**, 010 (2007), hep-ex/0702006.

[57] B. W. Lee and R. E. Shrock, Phys.Rev. **D16**, 1444 (1977).

[58] K. Fujikawa and R. Shrock, Phys.Rev.Lett. **45**, 963 (1980).

[59] Y. Izotov and T. Thuan, Astrophys.J. **710**, L67 (2010), 1001.4440.

[60] We obtain identical results when characterizing the interaction with a magnetic analog of Ohm's law, and the conductivity is taken to be $\sigma_M = e_m^2 n_m / f_{\rm drag}$.

[61] Strictly speaking the advective term is negligible compared to the diffusive term, $\eta_{\rm d}\nabla^2\mathbf{B}$, only when $v < \eta_{\rm d}/\lambda_B \sim (10^3 T\lambda_B)^{-1}$ where λ_B is the typical length scale of the magnetic field. Since $\lambda_B \gg T^{-1}$ is a macroscopic scale, this imposes a very strong upper bound on v, which is not easily satisfied. We expect that inclusion of the advection term, perhaps in the context of a turbulent medium, will lead to a more rich and interesting solution, but that analysis is beyond the scope of this work.

PREPARED FOR SUBMISSION TO JCAP

Morphology of blazar-induced gamma ray halos due to a helical intergalactic magnetic field

Andrew J. Long and Tanmay Vachaspati

Physics Department and School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287, USA.

E-mail: andrewjlong@asu.edu, tvachasp@asu.edu

Abstract. We study the characteristic size and shape of idealized blazar-induced cascade halos in the $1 - 100$ GeV energy range assuming various non-helical and helical configurations for the intergalactic magnetic field (IGMF). While the magnetic field creates an extended halo, the helicity provides the halo with a twist. Under simplifying assumptions, we assess the parameter regimes for which it is possible to measure the size and shape of the halo from a single source and then to deduce properties of the IGMF. We find that blazar halo measurements with an experiment similar to Fermi-LAT are best suited to probe a helical magnetic field with strength and coherence length today in the ranges $10^{-17} \lesssim B_0/\text{Gauss} \lesssim 10^{-13}$ and $10\,\text{Mpc} \lesssim \lambda \lesssim 10\,\text{Gpc}$ where $\mathcal{H} \sim B_0^2/\lambda$ is the magnetic helicity density. Stronger magnetic fields or smaller coherence scales can still potentially be investigated, but the connection between the halo morphology and the magnetic field properties is more involved. Weaker magnetic fields or longer coherence scales require high photon statistics or superior angular resolution.

Keywords: intergalactic magnetic field, helicity, gamma ray, blazar

Contents

1 Introduction

Several lines of reasoning suggest that the voids between galaxies and galaxy clusters contain a large-scale intergalactic magnetic field (IGMF). The presence of even a weak IGMF is sufficient to explain the observed micro-Gauss galactic and cluster magnetic fields [1–3] through the dynamo amplification. Moreover, the IGMF may be a remnant from the early universe as it could have been generated during cosmological phase transitions [4, 5], in the epoch of matter-genesis [6–8], or in certain inflationary scenarios [9]. An observation of the IGMF and measurement of its energy and helicity spectra, therefore, could serve as a powerful new probe of astrophysics, particle physics, and early universe cosmology. (For recent reviews on cosmic magnetic fields, see Ref. [10, 11].)

On the observational side, TeV blazars offer one of the best strategies for measurements of the IGMF at redshifts out to $z \sim 1$ [12–16]. The blazar initiates an electromagnetic cascade as its TeV gamma rays produce electron and positron pairs (leptons) upon scattering on extragalactic background light (EBL). The cascade develops as the leptons inverse-Compton scatter on cosmic microwave background (CMB) photons producing secondary GeV gamma rays. Since the initial TeV gamma ray has a mean free path of $\gtrsim 10$ Mpc, the cascade develops outside of the host halo where it probes the IGMF. In the presence of an IGMF, the charged leptons are deflected by the magnetic field and the blazar acquires a halo of GeV gamma rays. In the weak field regime, the leptons experience a gentle deflection and the angular extent of the magnetically broadened cascade (MBC) goes as $\Theta \propto B$. In the strong field regime, the leptons are so dramatically deflected that the secondary emission is isotropized into an extended pair-halo (PH).

The hunt for cascade halos is ongoing at a number of gamma ray observatories [17–19] as well as independent collaborations [20–24]. Most recently Ref. [24] reports evidence for GeV halos around 24 low redshift blazars, which are revealed in a stacked analysis of the Fermi-LAT gamma

ray data. Assuming a magnetically broadened cascade, *i.e.* weak bending approximation, they infer the IGMF field strength to be $B_0 \sim 10^{-17} - 10^{-15}$ G where B_0 is the magnetic field strength at the present cosmological epoch.

While there has been extensive analytic and numerical work on the relationship between halo size and magnetic field strength, the literature contains little discussion of what information could be extracted from the halo shape. In principle the halo size and shape together (morphology) may encode not only the magnetic field strength but also its helicity[1].

There are two compelling reasons for considering a *helical* IGMF. First, magnetic helicity is a prediction of many models of magnetogenesis from the matter-genesis epoch [6–8], cosmological inflation [25–30], and other early universe scenarios [31–33]. Second, the evolution of a magnetic field within the magneto-hydrodynamic (MHD) approximation is known to be more robust to dissipation if it is helical [34]. So, if a magnetic field was generated by causal processes in the early universe, it has a much better chance of surviving if it is helical.

The helicity of the IGMF can be observed if charged particles propagate in this field, such as in the case of cosmic rays [35] and cascade gamma rays [36, 37]. The parity violating IGMF helicity leads to certain non-trivial parity-odd correlation functions of the arrival directions of cosmic rays and cascade gamma rays. An advantage of seeking a parity odd signature is that other sources of noise are expected to be parity even and hence do not contribute to the signature on average. The parity-odd correlation, called Q, has been evaluated for diffuse gamma ray data obtained by the Fermi Gamma Ray Telescope, and provides evidence for an IGMF of strength $\sim 10^{-14}$ G on distance scales ~ 10 Mpc with left-handed helicity [38, 39].

In this work we endeavor to develop an understanding of how a helical IGMF leads to parity-violating features in the shape of blazar-induced cascade halos. To that end, we study an idealized system: we focus on a few toy models of the IGMF, we do not model astrophysical sources, we do not include stochasticity in the development of the cascade, and we do not include foreground (noise). In this setting we can focus on the relationships between the parameters of the magnetic field model, *i.e.* the field strength, coherence length, and helicity, and the resulting size and shape of the cascade halo. We find that there are several regimes in the range of interesting physical parameters that each lead to qualitatively different halo morphology. We identify the region of parameter space where measurements of cascade halo size and shape are best suited to probe the helical IGMF.

In Sec. 2 we review the physics giving rise to the cascade halo. In Sec. 3 we study the propagation of particles through the cascade and derive a set of equations that can be solved to find the size and shape (morphology) of the halo. In Sec. 4 we apply the results of Sec. 3 to calculate the halo shape for five specific magnetic field configurations, three non-helical and two helical. In Sec. 5 we analyze which regions of the magnetic field parameter space could be probed by measurements of cascade halos. We conclude in Sec. 6 with a discussion of potential directions for future work.

2 The Blazar-Induced Cascade Halo

As depicted in Fig. 1, a blazar's TeV gamma rays initiate an electromagnetic cascade when they scatter on extragalactic background light. If the cascade occurs in the presence of a magnetic field, the blazar acquires a halo of GeV photons. Then the shape and angular extent of the halo are related to the magnetic field in the neighborhood of the blazar. This section is a review of the physics giving rise to the halo following Ref. [40].

[1] A helical magnetic field has a larger amplitude in either left- or right-circularly polarized modes.

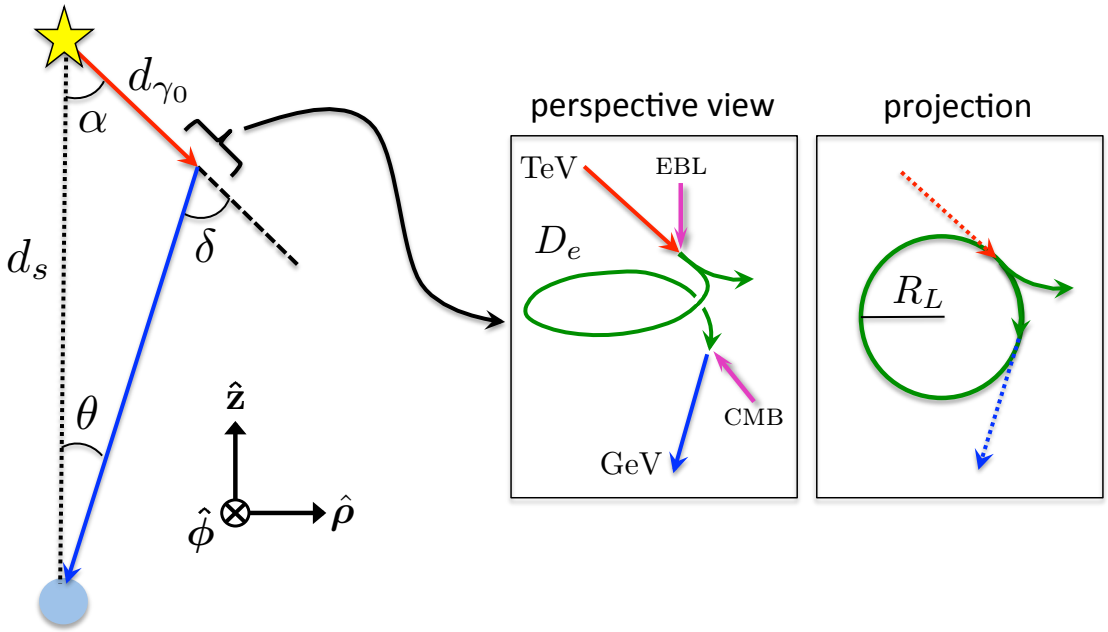

Figure 1. The geometry of gamma ray propagation. A TeV gamma ray (red arrow) travels off of the line of sight (black dashed line) connecting the blazar (yellow star) to Earth (pale blue dot). Pair production occurs and either the electron or positron (green arrow) is deflected back toward the line of sight. Inverse Compton scattering creates a GeV photon (blue arrow) that eventually reaches Earth. The lepton need not make a complete orbit as we have shown here.

The comoving distance from the Earth to a blazar (source) at redshift z_s is

$$d_s = \frac{c}{a_0 H_0}\int_0^{z_s}\frac{dz}{\sqrt{\Omega_m(1+z)^3+\Omega_\Lambda}} \simeq (1\ \mathrm{Gpc})\frac{z_s}{0.24} \tag{2.1}$$

where we have used the measured values of the cosmological parameters [41] and assumed $z_s \ll 1$ in the last equality. The blazar emits $O(1-10\ \mathrm{TeV})$ gamma rays into a jet or pair of jets [42]. In the simplest model of the jet, radiation is uniform within a cone of half opening angle $\theta_{\rm jet}$, and typically $\theta_{\rm jet} \simeq 5^\circ$ corresponding to a solid angle of $\Omega_{\rm jet} \simeq 0.024$ sr.

A gamma ray with energy $E_{\gamma_0} \sim$ TeV at redshift $z_{\gamma\gamma}$ is likely to scatter on optical and infrared extragalactic background light (EBL) and produce an electron-positron pair. The mean free path is given by $D_{\gamma_0} = \langle\sigma_{\gamma\gamma} n_{\rm EBL}\rangle^{-1}$ where $\sigma_{\gamma\gamma}$ is the pair production cross section and $dn_{\rm EBL}(\epsilon, z_{\gamma\gamma})/d\epsilon$ is the spectrum of EBL photons at redshift $z_{\gamma\gamma}$. Assuming $n_{\rm EBL}(z_{\gamma\gamma}) = (1+z_{\gamma\gamma})^{-2} n_{\rm EBL}(z=0)$ it was argued in Ref. [40] that the mean free path can be reliably approximated as

$$D_{\gamma_0} \simeq (80\ \mathrm{Mpc})\frac{\kappa}{(1+z_{\gamma\gamma})^2}\left(\frac{E_{\gamma_0}}{10\ \mathrm{TeV}}\right)^{-1}. \tag{2.2}$$

Ref. [40] estimates a range of values $0.3 < \kappa < 3$ for the dimensionless coefficient, and we will take $\kappa = 1$ hereafter. The comoving mean free path is given by $d_{\gamma_0} = (1+z_{\gamma\gamma})D_{\gamma_0}$. As long as $d_{\gamma_0} \ll d_s$ we can approximate $z_{\gamma\gamma} \approx z_s$ and write the comoving mean free path as

$$d_{\gamma_0} \simeq (80\ \mathrm{Mpc})\frac{1}{(1+z_s)}\left(\frac{E_{\gamma_0}}{10\ \mathrm{TeV}}\right)^{-1}. \tag{2.3}$$

After pair production the charged leptons acquire an energy $E_e \approx E_{\gamma_0}/2$, and, to accuracy $m_e/E_e \sim 10^{-6}$ (the inverse boost factor), they travel in the same direction as the initial TeV gamma ray. The leptons produce secondary γ-rays through inverse Compton (IC) scattering on cosmic microwave background (CMB) photons. If $E_{\rm CMB}$ is the energy of a CMB photon at redshift $z_{\gamma\gamma}$ then energy conservation gives the energy of the corresponding IC photon to be

$$E'_\gamma = \frac{4}{3} E_{\rm CMB} \frac{E_e^2}{m_e^2} \tag{2.4}$$

where $m_e \simeq 0.511\ {\rm MeV}/c^2$ is the electron mass. When this photon reaches Earth ($z = 0$) it will have been redshifted to an energy of $E_\gamma = (1 + z_{\gamma\gamma})^{-1} E'_\gamma$. If the spectrum of the TeV blazar is known, one can use the spectra of the EBL and CMB to calculate the spectrum of GeV gamma rays arriving at Earth. For our purposes only the average relationships are required. At redshift $z_{\gamma\gamma}$ the average energy of a CMB photon is $\langle E_{\rm CMB}(z_{\gamma\gamma})\rangle = \epsilon_{\rm CMB} \simeq (6 \times 10^{-4}\ {\rm eV})(1 + z_{\gamma\gamma})$, and the average energy of a GeV photon arriving at Earth is then

$$E_\gamma = \frac{4}{3}(1 + z_{\gamma\gamma})^{-1} \epsilon_{\rm CMB} \frac{E_e^2}{m_e^2} \simeq (77\ {\rm GeV}) \left(\frac{E_{\gamma_0}}{10\ {\rm TeV}}\right)^2 . \tag{2.5}$$

Evidently an initial spectrum of gamma rays from $E_{\gamma_0} \sim 1 - 10$ TeV are transferred into cascade photons with $E_\gamma \sim 1 - 100$ GeV energies. The scattering of leptons on CMB photons is a stochastic process that occurs with a typical mean free path $l_{\rm mfp}$. On average the leptons lose their energy via IC within the electron cooling distance

$$D_e = \frac{3 m_e^2 c^4}{4 \sigma_T U_{\rm CMB} E_e} \simeq (31\ {\rm kpc}) \left(\frac{E_e}{5\ {\rm TeV}}\right)^{-1} \left(\frac{1 + z_{\gamma\gamma}}{1.24}\right)^{-4} , \tag{2.6}$$

where $\sigma_T \simeq 6.65 \times 10^{-25}\ {\rm cm}^2$ is the Thomson scattering cross section and $U_{\rm CMB} \simeq (1+z_{\gamma\gamma})^4 (0.26\ {\rm eV/cm}^3)$ is the CMB energy density at redshift $z_{\gamma\gamma}$.

The halo emerges because TeV gamma rays directed off of the line of sight can induce a cascade that is deflected back toward the line of sight by the magnetic field. At the point of pair production (redshift $z_{\gamma\gamma}$) let $\mathbf{B}$ and $\mathbf{v}$ be the magnetic field and the velocity of the lepton, respectively. Assuming that the coherence length of the magnetic field $\lambda \approx |\boldsymbol{B}|/|\nabla \boldsymbol{B}|$ is much larger than D_e, the electron or positron will probe an effectively homogeneous magnetic field. In this background the lepton follows a helical trajectory with gyroradius (Larmor radius)

$$R_L = R \frac{|\mathbf{v}_\perp|}{c} \qquad \text{where} \qquad R \equiv \frac{E_e}{e|\mathbf{B}|} , \tag{2.7}$$

and $-e$ is the charge of the electron. The component of $\mathbf{v}$ perpendicular to $\mathbf{B}$ is $\mathbf{v}_\perp = \mathbf{v} - (\mathbf{v} \cdot \hat{\mathbf{B}})\hat{\mathbf{B}}$ where unit vectors are denoted by a hat. Using this expression we can write

$$R_L = R \sqrt{1 - (\hat{\mathbf{v}} \cdot \hat{\mathbf{B}})^2} . \tag{2.8}$$

If the magnetic field is frozen into the plasma, then its energy density redshifts like radiation, and we have $|\mathbf{B}| = B_0 (1 + z_{\gamma\gamma})^2$ with B_0 the field strength today. Then we can we estimate

$$R \simeq (3.5\ {\rm Mpc}) \left(\frac{E_e}{5\ {\rm TeV}}\right) \left(\frac{B_0}{10^{-15}\ {\rm G}}\right)^{-1} \left(\frac{1 + z_s}{1.24}\right)^{-2} \tag{2.9}$$

where we have also used $z_{\gamma\gamma} \approx z_s$.

By approximating $z_s \approx z_{\gamma\gamma}$ throughout our analysis we assume that $d_{\gamma_0}/d_s \ll 1$. Using Eqs. (2.1) and (2.3) this ratio is

$$\frac{d_{\gamma_0}}{d_s} \simeq 0.18 \left(\frac{E_\gamma}{10\ \mathrm{GeV}}\right)^{-1/2} \left(\frac{1+z_s}{1.24}\right)^{-1} \left(\frac{d_s}{1\ \mathrm{Gpc}}\right)^{-1} \tag{2.10}$$

where d_s and z_s are related by Eq. (2.1). We will primarily be interested in $E_\gamma \gtrsim 10$ GeV and $d_s \gtrsim 0.7$ Gpc where the approximation is well-justified. For lower energy gamma rays or closer sources, the approximation begins to break down as the Earth is located inside of the developing cascade.

It is also useful to compare the electron cooling distance D_e and gyroradius R. Using Eqs. (2.6) and (2.9) and approximating $z_{\gamma\gamma} \approx z_s$ this ratio is

$$\frac{D_e}{R} \simeq 0.067 \left(\frac{B_0}{10^{-15}\ \mathrm{G}}\right) \left(\frac{E_\gamma}{10\ \mathrm{GeV}}\right)^{-1} \left(\frac{1+z_s}{1.24}\right)^{-2} . \tag{2.11}$$

Since $2\pi R_L$ is the circumference of the lepton's orbit, this ratio indicates whether the lepton travels around the orbit many times $D_e/R \gg 1$ or whether it makes only a small arc $D_e/R \ll 1$. In the former case, called the "pair halo regime," the cascade gamma rays are spread out over large angles; in the latter case, called the "magnetically broadened cascade regime," the cascade photons are slightly spread out around the source.

3 Morphology of the Cascade Halo

We are interested in the size and shape of the cascade halo as it appears from Earth. In this section we first establish an analytic formalism for calculating halo maps, *i.e.* the orientation $\hat{\mathbf{n}}(E_\gamma)$ of GeV gamma rays reaching Earth. Next we introduce parameters that quantify the halo size and shape, which can be extracted from the halo map.

3.1 Formalism and Assumptions

We move between spherical, cylindrical, and Cartesian coordinates. The origin is located at the Earth and $\hat{\mathbf{z}}$ is oriented along the line of sight to the blazar. The polar and azimuthal angles are denoted by (θ, ϕ), and the spherical and cylindrical radial coordinates are r and ρ. The sets of unit vectors $\{\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}\}$, $\{\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\phi}}, \hat{\mathbf{z}}\}$, and $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$ form right-handed orthonormal coordinate systems. It will be useful to note the relationships

$$\begin{aligned} \hat{\boldsymbol{\rho}} &= \cos\phi\, \hat{\mathbf{x}} + \sin\phi\, \hat{\mathbf{y}} \\ \hat{\boldsymbol{\phi}} &= -\sin\phi\, \hat{\mathbf{x}} + \cos\phi\, \hat{\mathbf{y}} \end{aligned} \quad , \quad \begin{aligned} \hat{\mathbf{x}} &= \cos\phi\, \hat{\boldsymbol{\rho}} - \sin\phi\, \hat{\boldsymbol{\phi}} \\ \hat{\mathbf{y}} &= \sin\phi\, \hat{\boldsymbol{\rho}} + \cos\phi\, \hat{\boldsymbol{\phi}} \end{aligned} . \tag{3.1}$$

For plotting gamma ray arrival directions, it is convenient to introduce the lateral and transverse angular extent as

$$\vartheta_{\mathrm{lat}} = \theta\, \cos\phi \qquad \text{and} \qquad \vartheta_{\mathrm{trans}} = \theta\, \sin\phi\, . \tag{3.2}$$

The distinction between the lateral and transverse directions is arbitrary. Although other mappings are possible, this one has the convenient feature that the Euclidean norm is equal to the polar angle $\theta = \sqrt{\vartheta_{\mathrm{lat}}^2 + \vartheta_{\mathrm{trans}}^2}$. A gamma ray arriving at Earth is specified by an energy E_γ and an orientation, which can be expressed as the pair (θ, ϕ) or equivalently $(\vartheta_{\mathrm{lat}}, \vartheta_{\mathrm{trans}})$ or $\hat{\mathbf{n}} = \hat{\mathbf{r}}$.

We seek to study the development of the cascade *semi-analytically* by deriving a set of trigonometric equations that can be solved (analytically if possible, numerically if not) to find the orientation of GeV gamma rays reaching Earth. In this sense our approach differs from a purely numerical simulation in which TeV gamma rays are ejected from the blazar in random directions, the paths of the lepton and GeV gamma rays are calculated, and trajectories that do not intersect the Earth are discarded. As we will see, inspection of Fig. 1 leads to a set of three equations: the first arises from the trigonometry of the triangle, the second arises from the geometry of the lepton's orbit, and the third ensures that the GeV gamma ray intersects with the line of sight. We solve these three equations for a given magnetic field configuration and gamma ray energy E_γ to obtain the orientation of the GeV gamma ray at Earth (θ, ϕ) as well as the bending angle δ. This semi-analytic approach is applicable thanks to the following two well-justified approximations.

First, we assume that the lepton samples a homogeneous magnetic field, and thus its path is a simple helix. If λ is the coherence length of the magnetic field, this condition is expressed as $D_e \ll \lambda$. We saw in Eq. (2.6) that typically $D_e \sim 100$ kpc, and since we will be interested in $\lambda > 1$ Mpc this assumption is well-justified. If we were interested in smaller coherence scales, where the lepton motion is diffusive, then our approach would not be applicable.

Second, we assume that the displacement of the lepton can be neglected. This ensures that, to a good approximation, the two gamma rays lie in the same plane and that they form the legs of a triangle as in Fig. 1. To verify that the longitudinal displacement along the field line is negligible, we need $|\mathbf{v}_\parallel| D_e/|\mathbf{v}| \ll d_s, d_{\gamma_0}$; to ensure that the transverse displacement around the orbit is negligible we need $\mathrm{Min}[D_e, R_L] \ll d_s, d_{\gamma_0}$. In Eqs. (2.1), (2.3), and (2.9) we saw that typically $d_s \sim 1$ Gpc, $d_{\gamma_0} \sim 100$ Mpc, and $D_e, R \sim 100$ kpc. Then over all of the relevant parameters space we have $D_e \ll d_{\gamma_0}$, and the assumption is very well-justified.

Although our approach only requires the above two approximations, we also use the following two assumptions as a matter of convenience. First, it is important to remark that many aspects of the cascade are stochastic in nature: the spectrum of TeV gamma rays emitted by the blazar, the distance traveled by the TeV gamma rays before pair production, the spectrum of EBL photons, the distance traveled by the leptons before IC, the number of IC photons emitted before electron cooling becomes appreciable, and the spectrum of CMB photons that are up-scattered. For simplicity we neglect the stochastic spread in each of these various parameters, and we fix them equal to their average values given in Sec. 2. As a result, we obtain a deterministic relationship between the energy of gamma rays reaching Earth and their orientations on the sky; we call this function the halo map $\hat{\mathbf{n}}(E_\gamma)$. These halo maps are useful tools for studying the connection between halo morphology and the underlying magnetic field since they can be calculated quickly, without sophisticated numerical simulation, and they capture the characteristic features of the halo size and shape. It is expected that properly taking account of the stochasticity will lead to a significantly "smeared" version of the halo maps shown here, since the variance in the random parameters is typically $O(1)$. We discuss the stochastic smearing further in the conclusions, Sec. 6, as a direction for future work.

As a second simplification, we neglect the blazar's jet structure and assume that emission from the blazar is isotropic. If we were to properly treat the angular distribution of radiation from the blazar, then only a portion of our halo maps would be visible. We return to this point in the conclusions, Sec. 6, where we also show a few halo maps that have been calculated with the jet restriction in place.

3.2 Constraint Equations

We obtain the first in the set of three equations by applying the law of sines to the triangle in Fig. 1:

$$\sin\theta = \frac{d_{\gamma_0}}{d_s}\ \sin\delta\ . \tag{3.3}$$

Here θ is the polar angle that a GeV gamma ray arriving at Earth makes with the line of sight to the blazar, and the bending angle δ is the angle between the orientation of the initial TeV gamma ray and the final GeV gamma ray. The second equation is a relationship for the bending angle δ:

$$1 - \cos\delta = \Big(1 - (\hat{\mathbf{v}}_i \cdot \hat{\mathbf{B}})^2\Big)\Big(1 - \cos(D_e/R)\Big) \,. \tag{3.4}$$

To derive Eq. (3.4), we write the magnetic field at the point of pair production as $\mathbf{B} = B\,\hat{\mathbf{n}}_\parallel$, and we write the initial lepton velocity as $\mathbf{v}_i = v_\parallel\,\hat{\mathbf{n}}_\parallel + v_\perp\,\hat{\mathbf{n}}_\perp$ where $\hat{\mathbf{n}}_\parallel \cdot \hat{\mathbf{n}}_\perp = 0$. Then $v_\parallel = \mathbf{v}_i \cdot \hat{\mathbf{B}}$ and $v_\parallel^2 + v_\perp^2 = v^2 \approx c^2$. The lepton velocity at time t after pair production is

$$\mathbf{v}(t) = v_\parallel\,\hat{\mathbf{n}}_\parallel + v_\perp\cos(\omega t)\,\hat{\mathbf{n}}_\perp \mp v_\perp \sin(\omega t)\,\hat{\mathbf{n}}_\parallel \times \hat{\mathbf{n}}_\perp \tag{3.5}$$

where $\omega = v_\perp/R_L = c/R$ is the angular frequency of the orbital motion. The sign ambiguity in the last term is related to the charge of the lepton; the $(-)$ is for positrons and $(+)$ is for electrons. The lepton travels an electron cooling distance[2], D_e, in time $\tau = D_e/c$. Since the gamma rays are approximately tangential to the lepton trajectory, the orientation of the initial TeV gamma ray is $\hat{\mathbf{v}}(0)$, the orientation of the final GeV gamma ray is $\hat{\mathbf{v}}(D_e/c)$, and the bending angle satisfies $\cos\delta = \hat{\mathbf{v}}(0) \cdot \hat{\mathbf{v}}(D_e/c)$, which gives Eq. (3.4).

Before discussing the third equation, it is useful to consider the limits of small and large lepton deflection. Recall that $D_e/R = \omega\tau$ is the angular deflection of the lepton as it travels a distance D_e around the gyro-circle, and we saw in Eq. (2.11) that $D_e/R \propto B_0/E_\gamma$. For sufficiently weak magnetic field or high gamma ray energy we have $D_e/R \ll 1$ and Eqs. (3.3) and (3.4) reduce to

$$\theta \approx \sqrt{1 - (\hat{\mathbf{v}}_i \cdot \hat{\mathbf{B}})^2}\,\Theta_{\rm ext} \qquad \text{and} \qquad \delta \approx \sqrt{1 - (\hat{\mathbf{v}}_i \cdot \hat{\mathbf{B}})^2}\,\frac{D_e}{R} \tag{3.6}$$

where

$$\Theta_{\rm ext} \equiv \frac{d_{\gamma_0} D_e}{d_s R} \,. \tag{3.7}$$

In $\Theta_{\rm ext}$ we find the familiar expression for the angular extent of the halo [14]. Using the numerical estimates from Sec. 2 we have

$$\Theta_{\rm ext} \simeq (0.68^\circ)\left(\frac{B_0}{10^{-15}\ {\rm G}}\right)\left(\frac{E_\gamma}{10\ {\rm GeV}}\right)^{-3/2}\left(\frac{d_s}{1\ {\rm Gpc}}\right)^{-1}\left(\frac{1+z_s}{1.24}\right)^{-3} \,. \tag{3.8}$$

Since $\Theta_{\rm ext} \propto B_0$ the angular extent of the halo grows larger as the field strength is increased, which is characteristic of the MBC regime.

In the opposite regime where the lepton deflection is large we have $D_e/R \gg 1$, and the lepton makes multiple orbits before IC occurs. If $(\hat{\mathbf{v}}_i \cdot \hat{\mathbf{B}})^2 < 1/2$ then Eq. (3.4) has a discrete set of solutions at energies $E_\gamma^{(n)}$ where the halo reaches a maximum angular extent

$$\theta \approx \Theta_{\rm max} \qquad \text{and} \qquad \delta \approx \frac{\pi}{2} \tag{3.9}$$

[2]Throughout our analysis we assume that the lepton always travels an electron cooling distance, and that IC scattering only occurs once, at this point. In reality the mean free path of the lepton is shorter than D_e and multiple IC photons are emitted. These multiple emissions are an example of the stochastic effects, discussed previously, that we neglect for simplicity. We perform a preliminary investigation of the stochasticity in Sec. 6."

with

$$\Theta_{\rm max} \equiv \arcsin\left(\frac{d_{\gamma_0}}{d_s}\right) \simeq (10^\circ)\left(\frac{E_\gamma^{(n)}}{10\ {\rm GeV}}\right)^{-1/2}\left(\frac{1+z_s}{1.24}\right)^{-1}\left(\frac{d_s}{1\ {\rm Gpc}}\right)^{-1} . \tag{3.10}$$

In this limit, known as the PH regime, the angular extent of the halo is not proportional to the magnetic field strength, but instead it is restricted only by the geometry of the TeV gamma ray propagation.

We finally turn to the third constraint equation, which needs to enforce that the TeV and GeV gamma rays approximately lie in a plane of constant ϕ, as shown in Fig. 1. To ensure that the charged lepton is not deflected out of this plane, the Lorentz force $\mathbf{F} = (e/c)\,\mathbf{v}\times\mathbf{B}$ must be normal to $\hat{\boldsymbol{\phi}}$. Of course, as the lepton follows its helical path, the direction of its velocity changes and so too does the direction of the Lorentz force. Then we must average the Lorentz force over the trajectory of the lepton.

Let $\mathbf{x}(t)$ be the helical trajectory of the lepton, $\mathbf{v}(t) = d\mathbf{x}/dt$ be its velocity, $t_{\gamma\gamma}$ be the time of pair production, and $\tau = D_e/c$ be the time elapsed before IC up-scattering. Neglecting the cosmological expansion, which is not relevant on such short time scales, the impulse imparted on the charged lepton at redshift $z_{\gamma\gamma}$ is given by

$$\mathbf{J} = \pm\frac{e}{c}\int_{t_{\gamma\gamma}}^{t_{\gamma\gamma}+\tau} {\rm d}t\,\mathbf{v}(t)\times\mathbf{B}(\mathbf{x}(t),t) \tag{3.11}$$

where the $\pm$ is related to the charge on the lepton. If the magnetic field is static and homogeneous over the path of the lepton, we can pull it out of the integral, and the integrand contains only $\mathbf{v}(t)$. Defining the time averaged electron velocity as

$$\mathbf{v}_{\rm avg} \equiv \frac{1}{\tau}\int_{t_{\gamma\gamma}}^{t_{\gamma\gamma}+\tau} {\rm d}t\,\mathbf{v}(t)\ , \tag{3.12}$$

we can write the impulse as

$$\mathbf{J} = \pm\frac{e\tau}{c}\,\mathbf{v}_{\rm avg}\times\mathbf{B}\ . \tag{3.13}$$

Since the magnetic field does no work, the magnitude $|\mathbf{v}(t)| \approx c$ is fixed. Then using the geometry shown in Fig. 1, we see that $\mathbf{v}_{\rm avg}$ must bisect the angle δ, and it can be written as

$$\hat{\mathbf{v}}_{\rm avg} = \sin\left(\frac{\delta}{2}-\theta\right)\hat{\boldsymbol{\rho}} - \cos\left(\frac{\delta}{2}-\theta\right)\hat{\mathbf{z}}\ . \tag{3.14}$$

Writing also

$$\hat{\mathbf{B}} = b_\rho\hat{\boldsymbol{\rho}} + b_\phi\hat{\boldsymbol{\phi}} + b_z\hat{\mathbf{z}} \tag{3.15}$$

we have

$$\hat{\mathbf{v}}_{\rm avg}\times\hat{\mathbf{B}}\cdot\hat{\boldsymbol{\phi}} = -b_\rho\cos\left(\frac{\delta}{2}-\theta\right) - b_z\sin\left(\frac{\delta}{2}-\theta\right) = 0\ . \tag{3.16}$$

Typically b_ρ or b_z will depend on the azimuthal angle ϕ, and then this equation fixes the plane (normal to $\hat{\boldsymbol{\phi}}$) in which lie the two gamma rays and the line of sight.

Now we summarize the three constraint equations. To ensure that the motion remains in the plane, we impose Eq. (3.16):

$$b_\rho \cos\left(\frac{\delta}{2} - \theta\right) + b_z \sin\left(\frac{\delta}{2} - \theta\right) = 0 \tag{3.17}$$

where Eq. (3.15) gives the decomposition of the magnetic field into cylindrical coordinates. From the geometry of the gamma ray trajectories, we have the law of sines in Eq. (3.3),

$$\sin\theta = \frac{d_{\gamma_0}}{d_s}\ \sin\delta\ , \tag{3.18}$$

and finally the bending angle is given by Eq. (3.4),

$$1 - \cos\delta = \Big(1 - \big(b_\rho \sin(\delta-\theta) - b_z\cos(\delta-\theta)\big)^2\Big)\Big(1 - \cos(D_e/R)\Big)\ , \tag{3.19}$$

where we have written the initial lepton velocity as

$$\hat{\mathbf{v}}_i = \sin(\delta-\theta)\hat{\boldsymbol{\rho}} - \cos(\delta-\theta)\hat{\mathbf{z}}\ . \tag{3.20}$$

The constraints in Eqs. (3.17)-(3.19) can also be derived from the single vector equation

$$d_{\gamma 0}\hat{\boldsymbol{v}}_i + \Delta\boldsymbol{x} + L\hat{\boldsymbol{v}}_f + d_s\hat{\boldsymbol{z}} = 0\ , \tag{3.21}$$

which ensures that the cascade photon reaches Earth. Here L is the distance from the IC scattering to the observation point and $\Delta\boldsymbol{x}$ is the displacement of the lepton between pair production and IC scattering. To find $\Delta\boldsymbol{x}$, we can integrate Eq. (3.5). However, in the limit that $D_e \ll d_\gamma, d_s$, the $\Delta\boldsymbol{x}$ term can be dropped from the equation. Trigonometric manipulation of Eq. (3.21) then once again leads to the above constraint equations.

3.3 Shape Parameter

In the next section we consider various static magnetic field configurations $\hat{\mathbf{B}}(\mathbf{x})$, and we solve Eqs. (3.17)-(3.19) for θ, ϕ, and δ to determine the halo map $\hat{\mathbf{n}}(E_\gamma)$. Having solved Eqs. (3.17)-(3.19) we construct the halo map as the radial unit vector

$$\hat{\mathbf{n}}(E_\gamma) = \sin\theta\ \cos\phi\,\hat{\mathbf{x}} + \sin\theta\ \sin\phi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}} \tag{3.22}$$

that points from the Earth toward the arriving gamma ray of energy E_γ. Note that $\hat{\mathbf{n}}(E_\gamma)$ may be multi-valued meaning that gamma rays of a particular energy may appear from multiple directions in the sky. All of the information about the halo size and shape is contained in the function $\hat{\mathbf{n}}(E_\gamma)$. For instance, the size of the halo is given by $\Theta(E_\gamma) = \arccos[\hat{\mathbf{n}}(E_\gamma)\cdot\hat{\mathbf{z}}]$.

There are many ways to quantify the halo shape and orientation. Since we are interested in probing magnetic helicity, we are motivated to consider a parity-odd Q-statistic [36, 37], which is sensitive to the sign of the magnetic helicity. The statistic is defined as the triple product of vectors sampled from the halo map $\hat{\mathbf{n}}(E_\gamma)$ at three different energies. For illustrative purposes, we consider a few energy combinations:

$$Q_{10,30,50} = \hat{\mathbf{n}}_{10}\times\hat{\mathbf{n}}_{30}\cdot\hat{\mathbf{n}}_{50} \tag{3.23a}$$
$$Q_{25,30,35} = \hat{\mathbf{n}}_{25}\times\hat{\mathbf{n}}_{30}\cdot\hat{\mathbf{n}}_{35} \tag{3.23b}$$
$$Q_{8,10,12} = \hat{\mathbf{n}}_{8}\times\hat{\mathbf{n}}_{10}\cdot\hat{\mathbf{n}}_{12} \tag{3.23c}$$
$$Q_{38,40,42} = \hat{\mathbf{n}}_{38}\times\hat{\mathbf{n}}_{40}\cdot\hat{\mathbf{n}}_{42} \tag{3.23d}$$

where $\hat{\mathbf{n}}_{\#}$ is shorthand notation for $\hat{\mathbf{n}}(E_\gamma = \#\ \mathrm{GeV})$. We will see that different energy combinations are sensitive to different ranges of parameters, namely magnetic field strength and coherence length. Since $\hat{\mathbf{n}}(E_\gamma)$ may be multi-valued in general, one can calculate Q by first averaging over the multiple arrival directions for a given energy.

It is also useful to write the triple product $Q_{abc} = \hat{\mathbf{n}}_a \times \hat{\mathbf{n}}_b \cdot \hat{\mathbf{n}}_c$ as

$$Q_{abc} = \sin\vartheta_{ab} \sin\varphi_{abc} \tag{3.24}$$

where $0 \leq \vartheta_{ab} \leq \pi$ is the angle between $\hat{\mathbf{n}}_a$ and $\hat{\mathbf{n}}_b$, and $-\pi/2 \leq \varphi_{abc} \leq \pi/2$ is the angle between $\hat{\mathbf{n}}_c$ and its projection onto the plane spanned by $\hat{\mathbf{n}}_a$ and $\hat{\mathbf{n}}_b$. If $\hat{\mathbf{n}}_a$ and $\hat{\mathbf{n}}_b$ are collinear then $\vartheta_{ab} = 0$ and Q_{abc} vanishes; if $\hat{\mathbf{n}}_c$ lies in the same plane as $\hat{\mathbf{n}}_a$ and $\hat{\mathbf{n}}_b$ then $\varphi_{abc} = 0$ and Q_{abc} also vanishes. The sign of Q_{abc} is controlled by the sign of φ_{abc}, which depends on whether $\hat{\mathbf{n}}_c$ is "in front of" or "behind" the plane normal to $\hat{\mathbf{n}}_a \times \hat{\mathbf{n}}_b$. Since it is odd under reflections one sees that Q_{abc} is a measure of parity violation in the halo map.

4 Halo Morphology for Specific Magnetic Field Configurations

In this section we study the size and shape of the GeV halo for various specific magnetic field configurations. The configurations we consider are simplified and do not realistically model the intergalactic magnetic field. However, these examples serve to illustrate the parametric relationships between the field configuration and the halo morphology. Previous studies have focused on the size information alone, and we will see that the shape information provides insight into the magnetic field's orientation and helicity. We consider five different non-helical and helical magnetic field configurations, as shown in Fig. 2.

4.1 Case 1: Uniform Magnetic Field Parallel to Line of Sight

The simplest configuration is a homogeneous magnetic field oriented along the line of sight with the blazar,

$$\hat{\mathbf{B}} = -\hat{\mathbf{z}}\ . \tag{4.1}$$

The three constraint equations, Eqs. (3.17)-(3.19) , reduce to

$$\sin(\delta/2 - \theta) = 0 \tag{4.2a}$$

$$\sin\theta = \frac{d_{\gamma_0}}{d_s}\sin\delta \tag{4.2b}$$

$$1 - \cos\delta = \sin^2(\delta - \theta)\Big(1 - \cos(D_e/R)\Big)\ . \tag{4.2c}$$

There is a trivial solution with $\delta = \theta = 0$ corresponding to gamma rays oriented along the line of sight, $\hat{\mathbf{v}} = -\hat{\mathbf{z}}$, that are not deflected by the magnetic field. There is also a nontrivial solution,

$$\delta = 2\theta = 2\cos^{-1}\left(\frac{d_s}{2d_\gamma}\right) \quad , \quad \frac{D_e}{R} = (2n+1)\pi \tag{4.3}$$

where n is an integer. Since the azimuthal angle ϕ does not appear in these equations, the solution will be rotationally symmetric about the line of sight to the blazar. Also, since D_e/R and d_{γ_0}/d_s depend on energy, there will only be a discrete set of energies for which a solution exists. The solution can be understood in physical terms: the velocity component along the magnetic field remains constant, and the velocity component perpendicular to the magnetic field gets reflected, and so the triangle in Fig. 1 is an isosceles triangle. Apart from these solutions, the magnetic field deflects other gamma rays away from the line of sight, and they do not reach Earth.

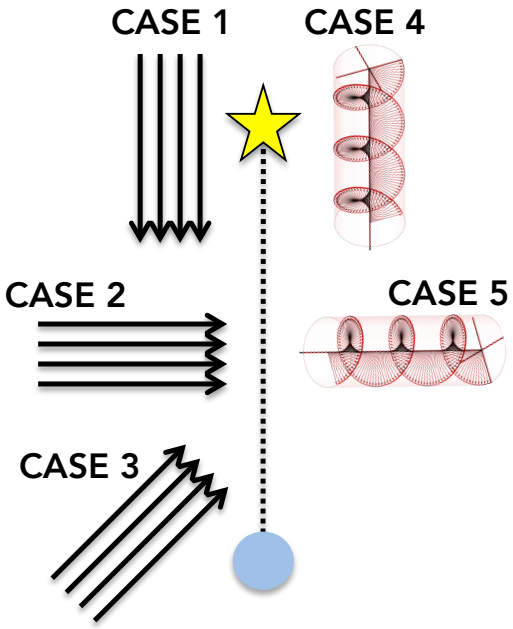

Figure 2. The five magnetic field configurations that we consider. Cases 1-3 are homogeneous field configurations that have different orientations with respect to the line of sight with the blazar. Cases 4-5 are helical field configurations with their wavevectors oriented either along or normal to the line of sight.

4.2 Case 2: Uniform Magnetic Field Normal to Line of Sight

Next we consider a homogeneous magnetic field that is oriented normal to the line of sight with the blazar (see Fig. 2). Without loss of generality we can align the Cartesian coordinate system with the magnetic field such that

$$\hat{\mathbf{B}} = \hat{\mathbf{y}} = \sin\phi\,\hat{\boldsymbol{\rho}} + \cos\phi\,\hat{\boldsymbol{\phi}}\,. \tag{4.4}$$

and Eqs. (3.17)-(3.19) reduce to

$$\sin\phi\,\cos(\delta/2-\theta) = 0 \tag{4.5a}$$

$$\sin\theta = \frac{d_{\gamma_0}}{d_s}\sin\delta \tag{4.5b}$$

$$1-\cos\delta = \Big(1-\sin^2(\delta-\theta)\,\sin^2\phi\Big)\Big(1-\cos(D_e/R)\Big)\,. \tag{4.5c}$$

For a given gamma ray energy E_γ there is a solution

$$\phi = 0,\pi \quad , \qquad \sin\theta = \frac{d_{\gamma_0}}{d_s}\sin\delta \quad , \quad \text{and} \quad \cos\delta = \cos\frac{D_e}{R} \tag{4.6}$$

where E_γ enters through d_{γ_0}/d_s and D_e/R, see Eqs. (2.10) and (2.11). Recall that $D_e/R > 0$ is unbounded from above but $0 \le \delta \le \pi$ and $0 \le \theta \le \pi/2$. For this magnetic field configuration, the trajectories of all the gamma rays lie in the $y = 0$ plane where $\phi = 0, \pi$. In the limit of small lepton deflection, $D_e/R \ll 1$, the solution further simplifies to

$$\theta \approx \Theta_{\rm ext} = \frac{d_{\gamma_0} D_e}{d_s R} \tag{4.7}$$

as in Eqs. (3.6) and (3.7).

The halo map corresponding to the solution in Eq. (4.6) is shown in Fig. 3. We have mapped the polar and azimuthal angles (θ, ϕ) into the lateral and transverse angular extent $(\vartheta_{\rm lat}, \vartheta_{\rm trans})$ using Eq. (3.2). The y-component of velocity is conserved for $\hat{\mathbf{B}} = \mathbf{y}$, and only gamma rays lying in the plane $y = 0$ normal to the magnetic field are deflected back toward the line of sight. The halo map $\hat{\mathbf{n}}(E_\gamma)$ is double-valued as there are precisely two gamma ray arrival directions for each energy. Along one branch the electron was deflected back toward Earth and generated the secondary IC gamma ray, and on the other branch it was the positron.

The energy dependence of the polar angle is shown in Fig. 4 for various parameter combinations. The highest energy gamma rays experience the smallest deflection, and they arrive closest to the line of sight ($\theta = 0$). In this regime, the small angle approximation is valid, and we have the scaling $\theta \approx \Theta_{\rm ext} \propto E_\gamma^{-3/2}$ from Eq. (3.7). Lower energy gamma rays are found farther from the line of sight, and there is an energy gradient. As the energy decreases further, the bending angle $\delta = D_e/R \propto 1/E_\gamma$ continues to grow, as per Eq. (2.11), and eventually $\delta = \pi/2$ where the electron experiences a deflection of 90°. This corresponds to an energy

$$E_\gamma^{(\rm crit)} \simeq (0.43\ {\rm GeV}) \left(\frac{B_0}{10^{-15}\ {\rm G}}\right) \left(\frac{1+z_s}{1.24}\right)^{-2}, \tag{4.8}$$

which also serves to indicate where the small bending approximation breaks down. Since $\sin\delta = 1$ is maximized at this point, the halo achieves a maximum angular extent (*cf.* Eq. (3.10))

$$\Theta_{\rm max} = \arcsin\left(\frac{d_{\gamma_0}}{d_s}\right) \simeq (59^\circ) \left(\frac{d_s}{1\ {\rm Gpc}}\right)^{-1} \left(\frac{B_0}{10^{-15}\ {\rm G}}\right)^{-1/2}. \tag{4.9}$$

Still lower energy gamma rays arrive closer to the line of sight, because the lepton is bent more than 90°.

For this case the halo map is wide in lateral extent and narrow in transverse extent as seen in Fig. 3. In principle one could measure the orientation of the magnetic field by measuring the orientation of the halo map. This is an example of how shape information can probe additional aspects of the IGMF beyond just its field strength. Of course, we have assumed that the magnetic field is uniform over the scale probed by the TeV gamma rays, *i.e.* if the magnetic field coherence length is λ, then we have implicitly assumed $\lambda \gg d_{\gamma_0} \sim 100$ Mpc. In a realistic setting it is more likely that the magnetic field forms domains smaller than d_{γ_0}. If the magnetic field is statistically isotropic across domains, then different leptons probe random orientations of the field, and the halo will resemble a more familiar, rotationally symmetric halo map. Apart from this isotropization, the discussion of this section is largely unchanged, and specifically Eq. (4.6) still gives the relationship between the gamma ray energy and polar angle.

4.3 Case 3: Uniform Magnetic Field with Arbitrary Orientation

Next we consider a homogeneous magnetic field that has a component along the line of sight to the blazar. This is a generalization of the previous two cases. Without loss of generality we can write the magnetic field configuration as

$$\hat{\mathbf{B}} = (\cos\beta\,\hat{\mathbf{y}} - \sin\beta\,\hat{\mathbf{z}}) = \cos\beta\sin\phi\,\hat{\boldsymbol{\rho}} + \cos\beta\cos\phi\,\hat{\boldsymbol{\phi}} - \sin\beta\,\hat{\mathbf{z}} \tag{4.10}$$

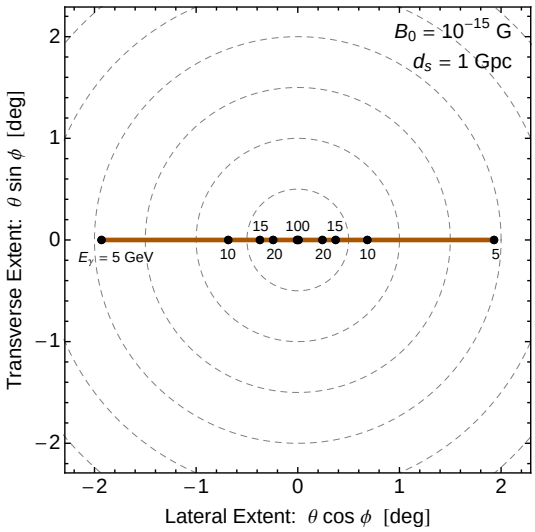

Figure 3. The halo map for Case 2. Gamma rays arrive at Earth collimated into a line with zero transverse extent. The black dots indicate $E_\gamma = 100, 20, 15, 10,$ and 5 GeV with the higher energy gamma rays arriving near the origin. The gray dashed lines are curves of constant polar angle θ.

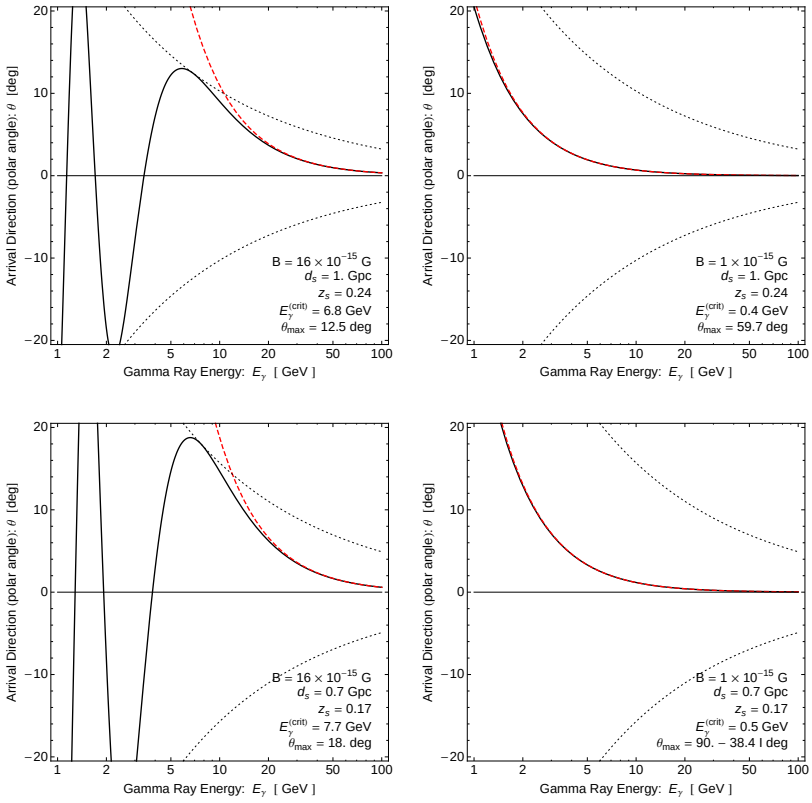

Figure 4. Angular extent of the halo as a function of gamma ray energy for Case 2. Each panel shows the exact solution (black line) from Eq. (4.6), the small bending approximation (red dashed) from Eq. (4.7), and the geometric limit (black dashed) from Eq. (3.10). The top-right panel shows the same parameters as in Fig. 3.

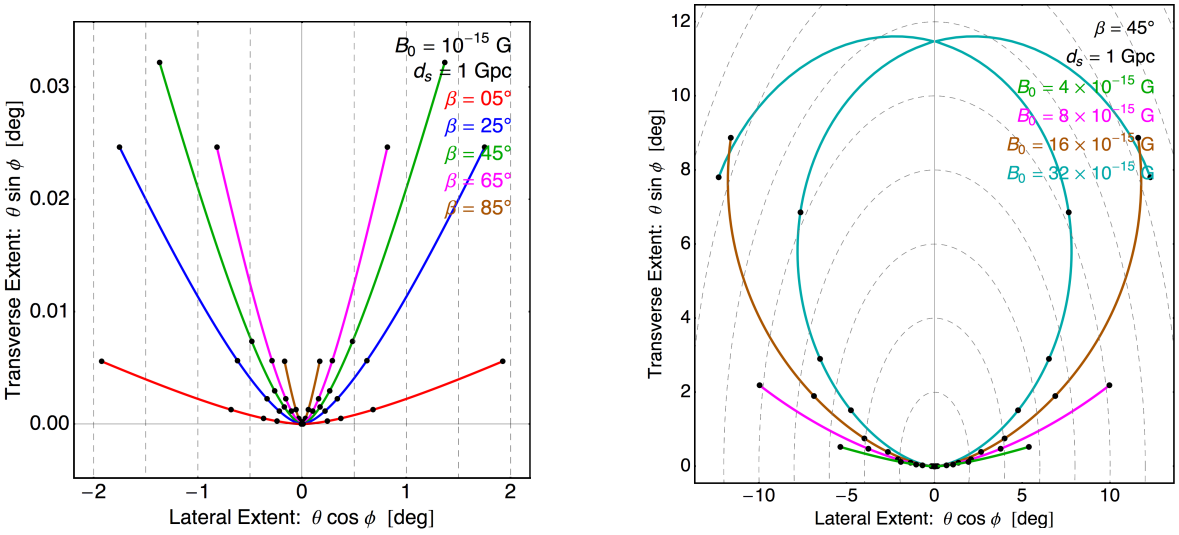

Figure 5. The halo map for Case 3. If we had taken $\beta < 0$ then the halo maps would be reflected in the vertical direction. The black dots and gray dashes have the same meaning as in Fig. 4.

where the skew angle β controls the component of $\hat{\mathbf{B}}$ along the line of sight. Eqs. (3.17)-(3.19) reduce to

$$\sin\phi = \tan\beta \, \tan(\delta/2 - \theta) \tag{4.11a}$$

$$\sin\theta = \frac{d_{\gamma_0}}{d_s} \sin\delta \tag{4.11b}$$

$$1 - \cos\delta = \left(1 - \sin^2\beta \frac{\cos^2(\delta/2)}{\cos^2(\delta/2 - \theta)}\right)\left(1 - \cos(D_e/R)\right) \tag{4.11c}$$

where we have used the equation for $\sin\phi$ to simplify the third equation.

Numerically solving Eq. (4.11) leads to the halo maps shown in Fig. 5. In comparing with Case 2 from Sec. 4.2 we see that the halo map is no longer restricted to a line, but instead the halo acquires a transverse extent. We vary the skew angle β in the left panel of Fig. 5. In the limit that β goes to zero, we regain the line-like halo map of Case 2, and in the limit that β goes to 90°, we regain point-like halo map of Case 1. We vary the field strength in the right panel of Fig. 5. In the PH regime where $B_0 \lesssim 10 \times 10^{-15}$ G, the halo size is proportional to the field strength (green and magenta curves), while in the MBC regime where $B_0 \gtrsim 10 \times 10^{-15}$ G, the halo size is limited by the geometry (brown and teal curves).

Our analytic approach to calculating the halo map has an advantage over numerical shooting techniques insofar as we can solve for the halo map analytically in certain limiting regimes. To demonstrate this point, first consider the limit of small skew angle $\beta \ll 1$ in which Case 3 reduces to Case 2. In this limit, Eq. (4.11) becomes

$$\phi \approx \beta \, \tan(\delta/2 - \theta) \quad , \qquad \theta \approx \arcsin\left(\frac{d_{\gamma_0}}{d_s} \sin\frac{D_e}{R}\right) \quad , \quad \text{and} \quad \delta \approx \frac{D_e}{R} \, , \tag{4.12}$$

and it follows that

$$\vartheta_{\rm lat} \approx \arcsin\left(\frac{d_{\gamma_0}}{d_s}\sin\frac{D_e}{R}\right) \tag{4.13}$$

$$\vartheta_{\rm trans} \approx \beta\,\tan\left(\frac{D_e}{2R} - \arcsin\left(\frac{d_{\gamma_0}}{d_s}\sin\frac{D_e}{R}\right)\right)\arcsin\left(\frac{d_{\gamma_0}}{d_s}\sin\frac{D_e}{R}\right)\,. \tag{4.14}$$

This behavior is seen in the left panel of Fig. 5. Second, consider the small bending angle regime $D_e/R \ll 1$ (weak magnetic field) where Eq. (4.11) becomes

$$\phi \approx \frac{\sin\beta}{2\cos^2\beta}\frac{D_e}{R} \quad , \qquad \theta \approx \frac{\Theta_{\rm ext}}{\cos\beta} \quad , \quad \text{and} \quad \delta \approx \frac{D_e}{R}\cos\beta\,, \tag{4.15}$$

and $\Theta_{\rm ext} \propto B_0$ was given by Eq. (3.7). Then the lateral and transverse extents are

$$\vartheta_{\rm lat} \approx \frac{\Theta_{\rm ext}}{\cos\beta} \tag{4.16}$$

$$\vartheta_{\rm trans} \approx \frac{\sin\beta}{2\cos^2\beta}\frac{D_e}{R}\frac{\Theta_{\rm ext}}{\cos\beta}\,. \tag{4.17}$$

Observe that $\vartheta_{\rm trans}$ is suppressed with respect to $\vartheta_{\rm lat}$ by an additional factor of $(D_e/R) \propto B_0$.

We quantify the halo shape using the triple product Q-statistic, given by Eq. (3.23). Since the magnetic field configuration under consideration is not helical, we expect $Q = 0$. In fact this is immediately evident from the symmetry of Fig. 5: the gamma rays on the branch in the first quadrant contribute $Q < 0$ while those in the second quadrant contribute $Q > 0$, and upon summing the two branches, they cancel. However, this cancellation is possible in part because we have assumed isotropic emission from the blazar. In practice, the jet may only illuminate a small patch of the halo map. Thus, to demonstrate the parametric dependence and typical scale of Q it is illustrative to calculate the statistic using only the gamma rays in one of the two branches.

We evaluate $Q_{10,30,50}$ from Eq. (3.23) and show the results in Fig. 6. Similar results are obtained for the other energy combinations, and we do not show them here. The parameters are chosen to correspond with the halo maps in Fig. 5. The statistic $Q_{10,30,50}$ becomes small (i) in the limit $\beta \to 0$ where the halo map approaches a straight line, (ii) in the limit $\beta \to 90^\circ$ where the halo map approaches a point, and (iii) in the limit $B_0 \to 0$ where the angular extent of the halo decreases. Increasing the field strength grows $Q_{10,30,50}$ until the crossover from the MBC to the PH regime at $B_0 \sim 50 \times 10^{-15}$ G. For larger B_0 the statistic first decreases and then begins to oscillate, similar to the behavior of the halo size, seen in Fig. 4.

4.4 Case 4: Helical Magnetic Field with Wave Vector Parallel to Line of Sight

We now turn our attention to helical magnetic field configurations. The simplest configuration consists of a single circular polarization mode with wavelength λ and wavevector $\mathbf{k} = (2\pi/\lambda)\hat{\mathbf{z}}$ oriented along the line of sight with the blazar:

$$\hat{\mathbf{B}} = \cos(\psi + 2\pi z/\lambda)\,\hat{\mathbf{y}} + \sigma\,\sin(\psi + 2\pi z/\lambda)\,\hat{\mathbf{x}}\,. \tag{4.18}$$

The spatial coordinate z should not be confused with the redshift z_s. The three parameters are the coherence length λ, the handedness index $\sigma = \pm 1$, and the phase shift ψ. The handedness index controls the sign of the magnetic helicity density,

$$\mathcal{H} \equiv \mathbf{B}\cdot\boldsymbol{\nabla}\times\mathbf{B} = \sigma\frac{2\pi}{\lambda}|\mathbf{B}|^2\,. \tag{4.19}$$

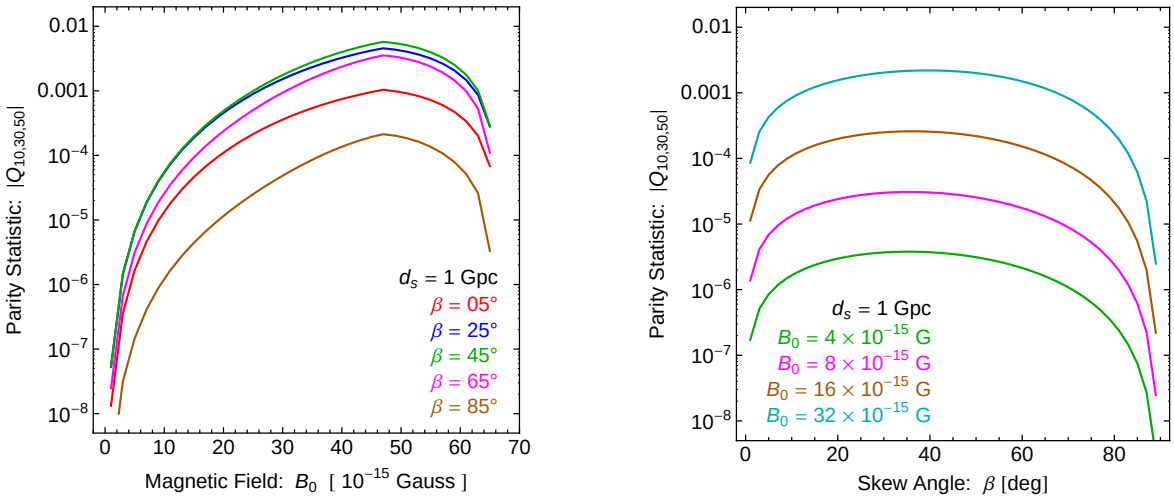

Figure 6. The magnitude of the parity statistic, given by Eq. (3.23), for Case 3. If we had averaged over all gamma rays, we would find $Q_{10,30,50} = 0$ since the gamma rays in the first and second quadrants of Fig. 5 cancel. Here we just show $Q_{10,30,50}$ for the gamma rays in the second quadrant.

The case $\sigma = 0$ corresponds to a non-helical, linearly polarized plane wave, while $\sigma = +1$ corresponds to left-circular polarization and -1 to right. In the subsequent analysis, one should bear in mind that varying λ at fixed $|\mathbf{B}|$ implies that the magnetic helicity is being varied.

It is convenient to define the angle $\beta_{\rm eff} = \psi + 2\pi z/\lambda$. Then using the geometry of Fig. 1 the longitudinal coordinate at the point of pair production is $z = d_s - d_{\gamma_0}\cos(\delta - \theta)$, and

$$\beta_{\rm eff}(\delta,\theta) = \psi + \frac{2\pi}{\lambda}\left(d_s - d_{\gamma_0}\cos(\delta-\theta)\right) \tag{4.20}$$

is the effective skew angle. For this case, Eqs. (3.17)-(3.19) reduce to the set of equations

$$\tan\phi = -\sigma\,\tan\beta_{\rm eff}(\delta,\theta) \tag{4.21a}$$

$$\sin\theta = \frac{d_{\gamma_0}}{d_s}\sin\delta \tag{4.21b}$$

$$\cos\delta = \cos\frac{D_e}{R}\,, \tag{4.21c}$$

which can be solved analytically. The solutions are shown in Fig. 7. The most striking feature in these figures is that the halo map forms a spiral pattern. The handedness of the spiral is controlled by the helicity of the magnetic field, parametrized here by $\sigma = \pm 1$. As the phase shift ψ is varied, the halo map is uniformly rotated clockwise or counterclockwise.

As we vary the coherence length λ the spiral becomes flatter or tighter. In the limit of large coherence length $\lambda \gg d_{\gamma_0} \sim 100$ Mpc, the cascade takes place in an effectively homogeneous magnetic field, $\hat{\mathbf{B}} \approx \hat{\mathbf{y}}$. Then we regain the behavior of Case 2 from Sec. 4.2 in which the gamma rays propagate in a plane and arrive at Earth collimated into a line with small transverse extent. In the opposite limit of small coherence length, $\lambda \ll d_{\gamma_0} \sim 100$ Mpc, the TeV gamma rays sample the magnetic field at a random phase. In terms of the halo map, this translates into a tightly wound spiral with multiple cycles. Varying the magnetic field strength has the same effect as in the previous cases. In the PH regime where the field is weak, the angular size of the halo grows with increasing field

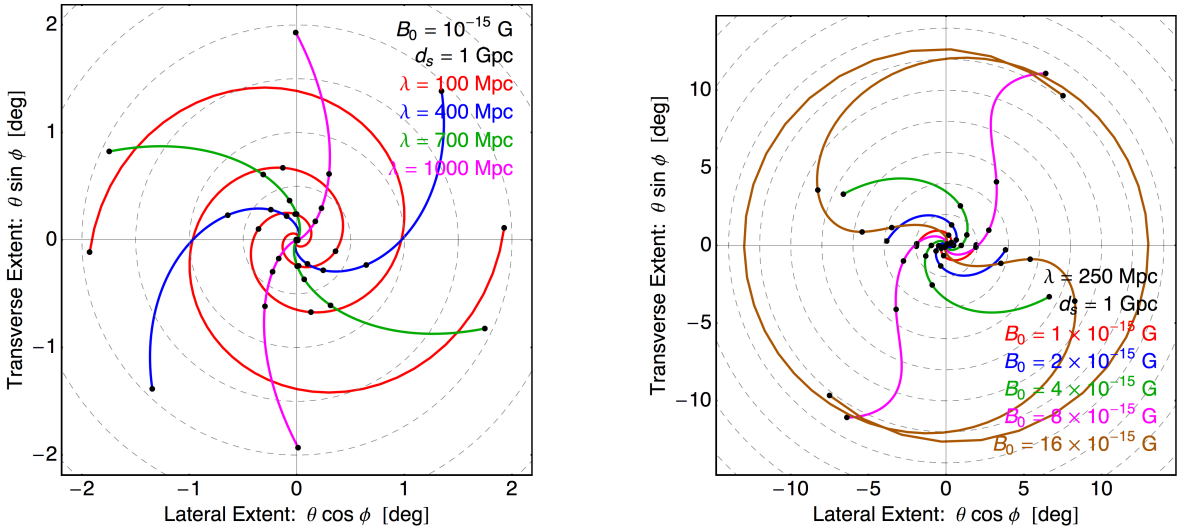

Figure 7. The halo map for Case 4. We have taken $\sigma = +1$ and $\psi = 0$. For $\sigma = -1$ the handedness of the spiral is reversed, and for a phase shift $\psi \neq 0$ the halo map is uniformly rotated, but otherwise the structure remains unchanged. In the left panel, if λ is further decreased below 100 Mpc the spiral becomes tighter, and if λ is further increased above 1 Gpc the halo map asymptotes to a straight line.

strength, while in the MBC regime where the field is strong, the angular size of the halo is limited by the geometry (brown curve in right panel).

The halo maps in Fig. 7 display a clear parity-violation that should be captured by the Q-statistics. However, the halo map $\hat{\mathbf{n}}(E_\gamma)$ is double-valued, *i.e.* there are two gamma rays at each energy, which correspond to the two branches of the spiral in Fig. 7. If we first average $\hat{\mathbf{n}}(E_\gamma)$ over its multiple solutions, we would obtain $\mathrm{Avg}[\hat{\mathbf{n}}(E_\gamma)] = 0$ and therefore $Q = 0$. This result is a consequence of the high symmetry of the system under consideration: we have assumed that the wavevector of the magnetic field is oriented along the line of sight, and we have allowed for isotropic emission from the blazar. As we discussed for Case 3 of Sec. 4.3, under more realistic conditions the blazar's jet will only illuminate a part of the full halo map, and then the cancellation is disrupted. To model this effect, we calculate Q along a single branch of the spiral.

We calculate the Q-statistics using Eq. (3.23) and show the parametric dependence on coherence length λ and field strength B_0 in Fig. 8. The statistic has the same qualitative behavior for each energy combination: (i) $|Q|$ decreases for small B_0, (ii) $|Q|$ decreases for large λ, and (iii) Q oscillates rapidly from positive to negative values for small λ. When the magnetic field is weak, the halo has a small angular extent, as shown in the right panel of Fig. 7, and $Q_{abc} = \hat{\mathbf{n}}_a \times \hat{\mathbf{n}}_b \cdot \hat{\mathbf{n}}_c$ decreases as $\hat{\mathbf{n}}_a, \hat{\mathbf{n}}_b$, and $\hat{\mathbf{n}}_c$ become approximately collinear (see also Eq. (3.24)). When the coherence length is large the halo map resembles a straight line, as seen in the left panel of Fig. 7, and Q_{abc} decreases as $\hat{\mathbf{n}}_a, \hat{\mathbf{n}}_b$, and $\hat{\mathbf{n}}_c$ become approximately coplanar. In this regime, the magnetic field appears uniform, and therefore non-helical, on the scale probed by the TeV gamma rays. When the coherence length is small the halo map resembles a tight spiral, as seen in the left panel of Fig. 7, and the three gamma rays used to construct Q may not lie on the same cycle of the spiral. In this case, Q may take positive or negative values depending on the energies at which the spiral is sampled, and as λ decreases further and the spiral becomes more tightly wound, the sign of Q oscillates. We will use λ_{osc} to denote the coherence length at which Q begins to oscillate.

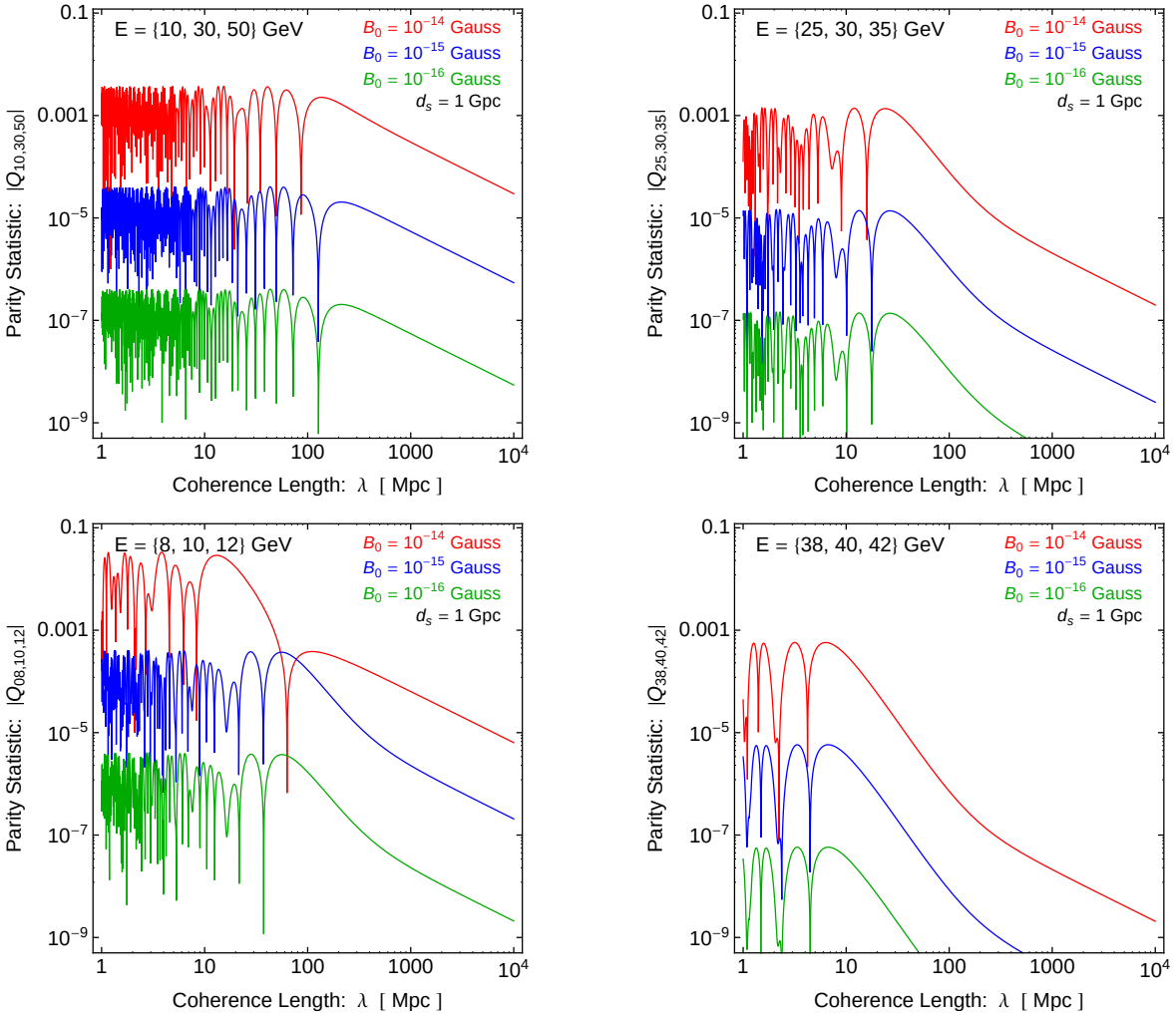

Figure 8. The parity statistics, given by Eq. (3.23), for Case 4. For large λ the sign of the Q-statistic is given by $\mathrm{sign}[Q] = -\sigma = -\mathrm{sign}[\mathcal{H}]$. As the coherence length is lowered, the halo map begins to spiral around the line of sight, as seen in Fig. 7, and the sign of Q oscillates. Statistics calculated from different energy combinations have the same qualitative behavior, but they differ in the magnitude of Q and the value of λ at which the oscillations begin.

The four cases in Fig. 8 display important quantitative differences between Q-statistics constructed from different energy combinations. Comparing the bottom two panels, $|Q_{8,10,12}| > |Q_{38,40,42}|$, we see that lower energy gamma rays leads to a larger value for $|Q|$. This is simply because lower energy gamma rays are more easily deflected and lead to a larger halo in the MBC regime, see Eq. (3.8). Comparing the top two panels, $\lambda_{\rm osc}(\{10, 30, 50\}\ \mathrm{GeV}) > \lambda_{\rm osc}(\{25, 30, 35\}\ \mathrm{GeV})$, we see how the energy spacing affects the coherence length scale below which Q begins to oscillate. As λ decreases and the halo map becomes a more tightly wound spiral, and the gamma rays from more closely spaced energies, in this case $\{25, 30, 35\}$ GeV, remain on the same cycle of the spiral longer than more widely spaced energies. In the two lower panels the energy spacings are identical but nevertheless $\lambda_{\rm osc}(\{38, 40, 42\}\ \mathrm{GeV}) < \lambda_{\rm osc}(\{8, 10, 12\}\ \mathrm{GeV})$. This is because the spiraling

behavior becomes more pronounced as the gamma ray energy is lowered; see Fig. 7 and note that the separation between the 5 and 10 GeV gamma rays is much larger than the separation between the 10 and 15 GeV gamma rays.

4.5 Case 5: Helical Magnetic Field with Wave Vector Normal to Line of Sight

Finally we consider a more generic helical magnetic field configuration. We do not require that the wave vector is oriented along the line of sight, but for simplicity we don't allow it to have a general orientation either, and instead we require that $\mathbf{k}$ is normal to the line of sight with the blazar. Without further loss of generality we can write $\mathbf{k} = (2\pi/\lambda)\hat{\mathbf{x}}$. The magnetic field configuration is

$$\begin{aligned}\hat{\mathbf{B}} &= \cos(\psi + 2\pi x/\lambda)\,\hat{\mathbf{y}} - \sigma\,\sin(\psi + 2\pi x/\lambda)\,\hat{\mathbf{z}} \\ &= \cos(\psi + 2\pi x/\lambda)\,\sin\phi\,\hat{\boldsymbol{\rho}} + \cos(\psi + 2\pi x/\lambda)\,\cos\phi\,\hat{\boldsymbol{\phi}} - \sigma\,\sin(\psi + 2\pi x/\lambda)\,\hat{\mathbf{z}}\,. \end{aligned} \tag{4.22}$$

In this case the phase shift ψ plays a nontrivial role. For specific values $\psi = 0, \pi/2, \pi, \cdots$ the magnetic field is either symmetric or antisymmetric when reflected across the line of sight ($x \to -x$ and $\phi \to -\phi$), but for general ψ there is no such symmetry. Once again the handedness index σ controls the magnetic helicity via Eq. (4.19).

It is convenient to introduce the effective skew angle

$$\beta_{\rm eff}(\delta, \theta, \phi) = \psi + \frac{2\pi d_{\gamma_0}}{\lambda}\sin(\delta - \theta)\cos\phi \tag{4.23}$$

where we have used $x = d_{\gamma_0}\sin(\delta - \theta)\,\cos\phi$. Then Eqs. (3.17)-(3.19) reduce to

$$\sin\phi = \sigma\,\tan\beta_{\rm eff}(\delta, \theta, \phi)\,\tan\left(\delta/2 - \theta\right) \tag{4.24a}$$

$$\sin\theta = \frac{d_{\gamma_0}}{d_s}\sin\delta \tag{4.24b}$$

$$1 - \cos\delta = \left(1 - \sin^2\beta_{\rm eff}(\delta, \theta, \phi)\frac{\cos^2(\delta/2)}{\cos^2(\delta/2 - \theta)}\right)\left(1 - \cos(D_e/R)\right)\,. \tag{4.24c}$$

These equations have a rich and interesting family of solutions, but as a result they cannot be solved analytically as in Case 4. Instead we solve Eq. (4.24) numerically.

The halo maps are shown in Fig. 9, and we have chosen the same parameters as in Fig. 7 to facilitate comparison with Case 4. In each of the three panels we have taken the handedness index $\sigma = +1$, and the associated parity-violation is evident in the "S"-like shape of the halo maps. Choosing $\sigma = -1$ reflects the halo maps across the vertical axis, which flips the handedness.

In the first panel, the coherence length is reduced from $\lambda = 1000$ Mpc to 50 Mpc. The behavior in the large λ regime is the same as in Case 4: when the coherence length is much larger than the scale of the cascade, $d_{\gamma_0} \sim 100$ Mpc, the gamma rays probe an effectively homogeneous magnetic field, and we regain the line-like halo map that was originally seen in Case 2 of Sec. 4.2. The small λ behavior is distinctly different than in Case 4 where we encountered a spiral-shaped halo map. Now the wavevector crosses the line of sight to the blazar. As λ is reduced, the surfaces of constant phase become compressed in the lateral direction, and the "S"-like halo map becomes "squeezed."

In the second panel, the magnetic field strength is increased from $B_0 = 2$ to 16×10^{-15} G. In the small B_0 regime, we once again regain the behavior of the previous cases: a smaller field strength translates into a smaller halo. In the large B_0 regime, on the other hand, a new phenomenon emerges: the halo map acquires multiple disconnected branches. In this panel only, all of the black dots denote $E_\gamma = 5$ GeV, and we see that for $B_0 = 8$ and 16×10^{-15} G the halo map $\hat{\mathbf{n}}(E_\gamma)$ has six distinct

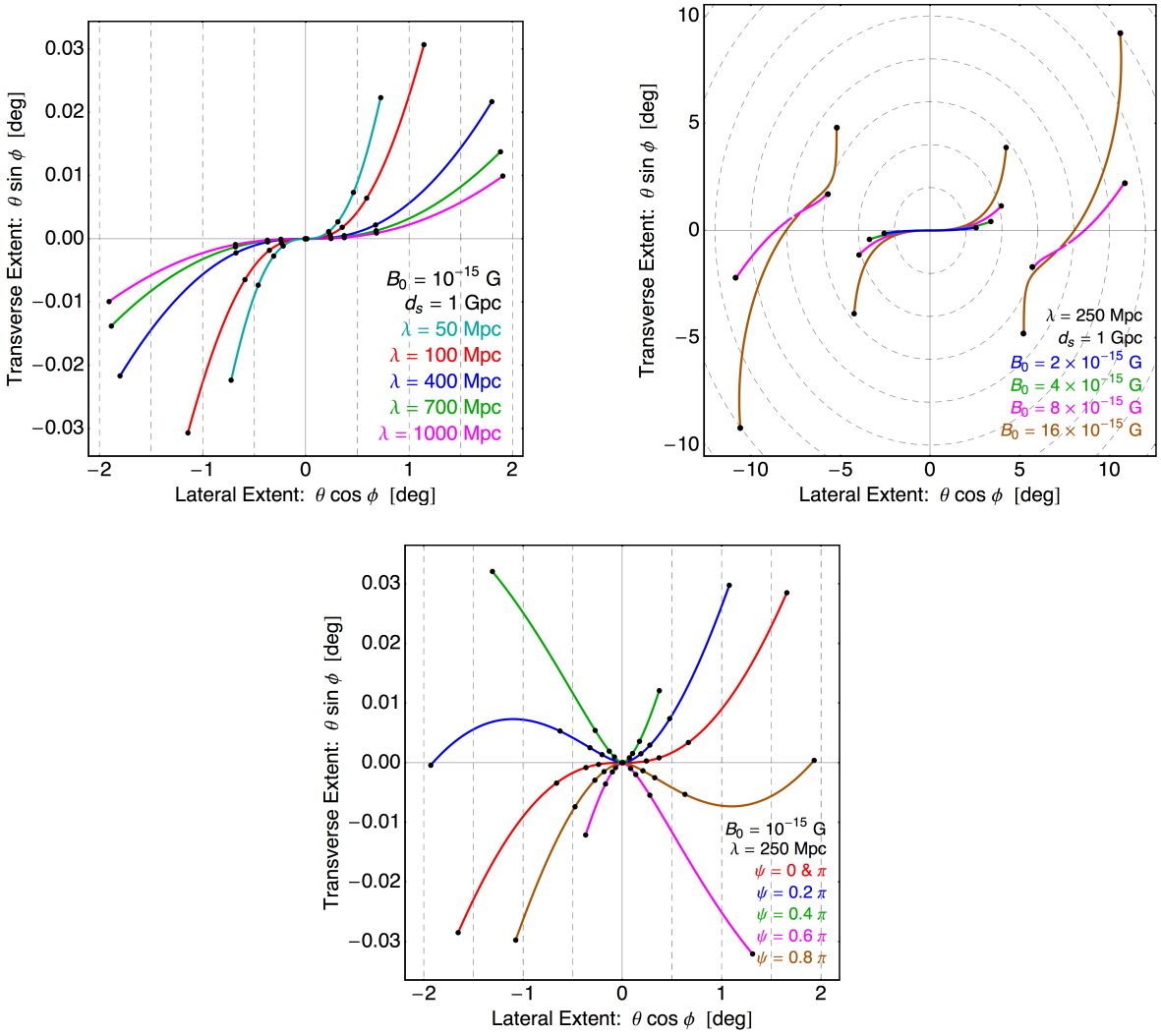

Figure 9. The halo map in Case 5. We have taken $\sigma = +1$ in all panels, and for $\sigma = -1$ the halo maps are reflected across the vertical axis. In the first and second panels we fix $\psi = 0$. In the first and third panels, the black dots have the same meaning as in the previous halo maps: they indicate $E_\gamma = 100, 20, 15, 10$, and 5 GeV. In the second panel, the black dots show only $E_\gamma = 5$ GeV, the black square indicates $E_\gamma = 7$ GeV for $B_0 = 8 \times 10^{-15}$ G, and the black triangle indicates $E_\gamma = 11$ GeV for $B_0 = 16 \times 10^{-15}$ G.

values. In the previous cases, the halo map was only double-valued. The new branches contain only low energy gamma rays, which are more easily deflected, and the bifurcation points are denoted by a black square and black triangle for $B_0 = 8$ and 16×10^{-15} G, respectively. As the field strength is further increased or the coherence length lowered, additional branches will emerge.

In the third panel, we vary the phase parameter ψ from 0 to π. If $\psi \neq (0 \mod \pi)$ then the reflection symmetry of the halo map is disrupted. This was not the situation in Case 4 where varying ψ simply lead to a uniform rotation of the spiral-like halo map.

To quantify the parity-violating features in the halo maps, we calculate the Q-statistics using Eq. (3.23) and show the results in Fig. 10. As we have seen, the halo map becomes multi-valued for large B_0 and small λ, and it displays multiple branches. To generate Fig. 10 we calculate Q

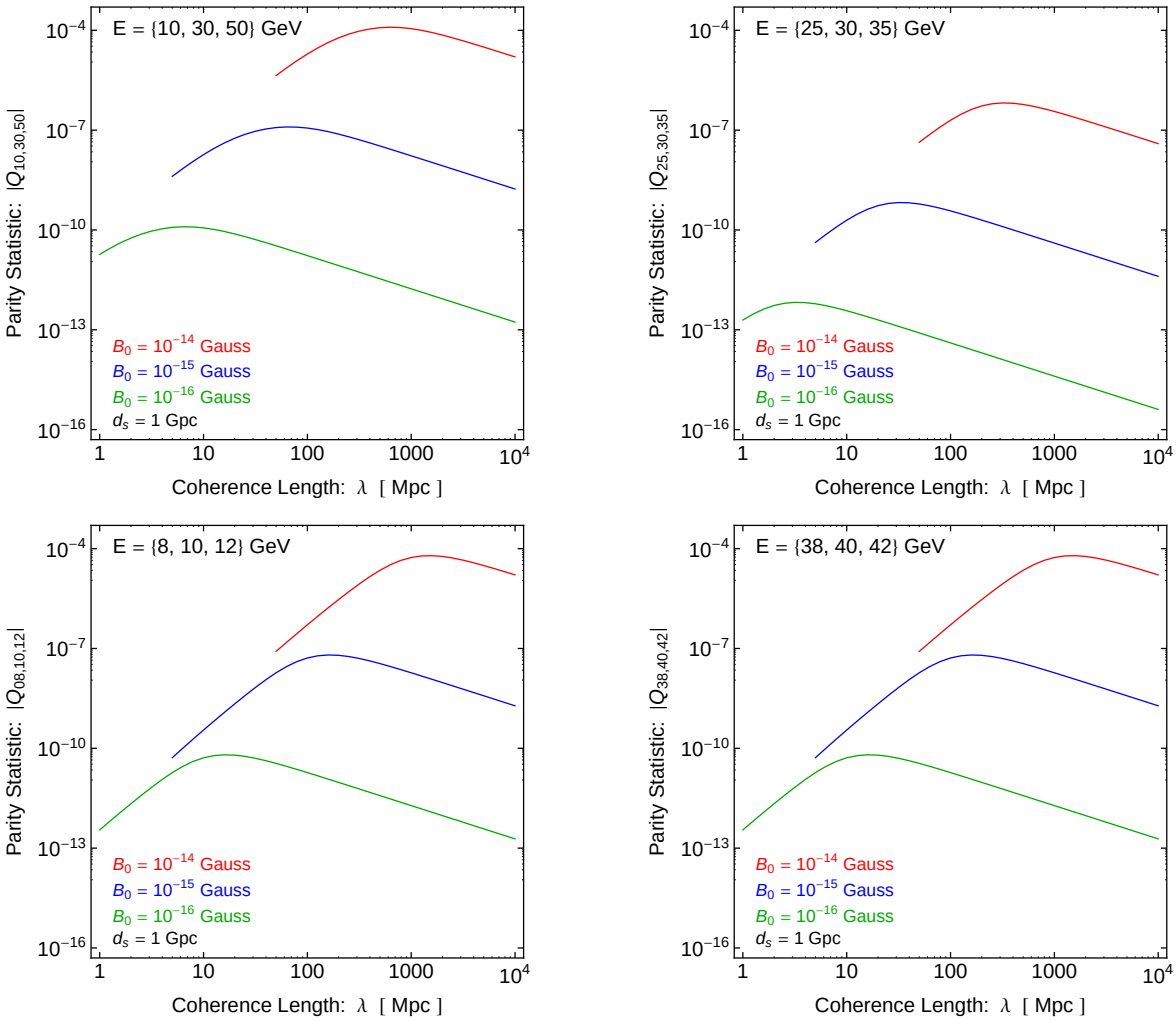

Figure 10. The parity statistics for Case 5. We calculate Q using only the main solution branch. The plots are truncated at small λ and large B_0 where numerical issues arise. For small λ and large B_0 where there can be multiple branches, this approach probably underestimates Q.

using only the main branch, which is continuously connected to the line of sight. By neglecting the branches with larger angular separation from the line of sight, we would presumably underestimate Q in this regime.

Over the parameter space shown Fig. 10 we have $\mathrm{sign}[Q] = -\sigma = -\mathrm{sign}[\mathcal{H}]$, where $\mathcal{H} > 0$ for the left-circular polarization mode. In the limit of large λ we have the same behavior as in Case 4, see Fig. 8. In the limit of small λ the Q-statistics decrease because the main branch is squeezed, and the angular extent of the halo is smaller. Also in comparing with Case 4, we see that the scaling with B_0 is different: in Fig. 8 we found $|Q| \sim B_0^2$ but Fig. 10 implies that $|Q| \sim B_0^4$ for Case 5. As such, the magnitude of Q very quickly decreases with decreasing field strength.

5 Implications for Helicity Measurement

By considering particular realizations of the IGMF, our analysis reveals the parametric behavior of the halo size and shape on the various magnetic field variables, specifically the field strength B_0, the coherence length λ, and the magnetic helicity density $\mathcal{H} = \pm(2\pi/\lambda)B_0^2$. In the "corners" of the parameter space, where either B_0 or λ is very large or very small, the halo map behaves in a way that may make a measurement of the magnetic helicity more challenging than in the "central" parameter regime. For instance, small B_0 implies a small halo, which may not be distinguishable from a point source with a telescope's finite angular resolution. In this section we will demarcate the various parametric regimes and discuss the challenges posed by each.

The quantitative analysis in this section uses the simplified field configurations considered in Sec. 4, but we expect qualitatively similar results (parameter space boundaries) even if the intervening magnetic field configuration is not of one of the forms we have discussed, *e.g.* if the field is stochastically homogeneous and isotropic. The inclusion of stochastic variables in the development of the cascade, *e.g.* spectrum of the EBL or scattering probabilities, will smear out the halo patterns we have seen, and non-cascade photons introduce noise in our signal and dilute the halo. A rigorous evaluation of experimental sensitivities, even for a given set of experimental parameters, will require more information on blazar sources (gamma ray flux and spectrum, jet orientation and structure) and the background noise.

Strong Field Regime

For a strong magnetic field, at energies such that the gyroradius R is smaller than the typical distance traveled by the charged lepton D_e, see Eq. (2.11), we have $D_e/R > 1$. This has two consequences. First, the maximum angular extent of the halo is no longer tied to the magnetic field strength, but instead it is fixed by the geometry as in Eq. (3.10). Then if a halo is seen, one can infer the presence of an IGMF, but one cannot measure the field strength from the halo size alone. Second, since the charged leptons can make a complete orbit around the gyrocircle, the GeV gamma rays will be emitted isotropically. This reduces the flux by roughly $\Omega_{\rm jet}/4\pi$, where $\Omega_{\rm jet}$ is the solid angle of the blazar's jet, and makes it more difficult to see the halo.

Therefore in the strong field regime, measurements of the cascade halo have a reduced capacity to probe the parameters of the IGMF. To avoid this regime, we require

$$\frac{D_e}{R} \lesssim 1 \,, \tag{5.1}$$

which leads to an upper bound on the magnetic field strength,

$$B_0 \lesssim (15 \times 10^{-15}\ \mathrm{G}) \left(\frac{E_\gamma}{10\ \mathrm{GeV}}\right) \left(\frac{1+z_s}{1.24}\right)^2 \tag{5.2}$$

upon using the expression for D_e/R in Eq. (2.11). We plot this boundary in Fig. 11. It divides the PH regime $D_e/R > 1$ from the MBC regime $D_e/R < 1$.

Weak Field Regime

A gamma ray telescope cannot pinpoint the arrival direction of a gamma ray with arbitrary precision. If the angular resolution is too poor, then the halo cannot be distinguished from a point source, and one cannot use halo size and shape measurements to probe the IGMF. This issue becomes especially relevant for weak magnetic fields, which do not induce much bending and lead to a smaller halo (*cf.* Eq. (3.8)).

Angular resolution is quantified by the point-spread function (PSF), which is the probability distribution function for the angle between the true and reconstructed arrival direction of a gamma ray

of given energy. The 68% confinement radius is the angular radius containing 68% of the probability. The confinement radius of the Fermi-LAT (Large Area Telescope) [43] is well-approximated by the following empirical formula:[3]

$$\delta\theta_{68}(E_\gamma) \simeq (0.11^\circ)\sqrt{1+\left(\frac{E_\gamma}{7.9\text{ GeV}}\right)^{-1.62}}\,. \tag{5.3}$$

Although our analysis is not specific to the Fermi-LAT instrument, we use this confinement radius as fiducial point of reference.

In the weak field regime, the angular extent of the halo is given approximately by $\theta \sim \Theta_{\rm ext}(E_\gamma)$ from Eq. (3.8), regardless of the specific field configuration under consideration. If $\Theta_{\rm ext}$ is sufficiently large compared to $\delta\theta_{68}$, then the halo can be distinguished from a point source and its angular extent measured. In fact, if the detector response is known very well (negligible systematic error), then halos as small as $\delta\theta_{68}/\sqrt{N_\gamma}$ can be probed when a large number N_γ of halo gamma rays are visible. Thus we assess when the telescope will be able to distinguish the halo from a point source using

$$\Theta_{\rm ext}(E_\gamma) \gtrsim \frac{\delta\theta_{68}(E_\gamma)}{\sqrt{N_\gamma(E_\gamma)}} \tag{5.4}$$

where $N_\gamma(E_\gamma)$ is the number of gamma rays collected at energy E_γ.

The requirement of sufficient angular resolution in Eq. (5.4) leads to a lower bound on the magnetic field strength,

$$B_0 \gtrsim \frac{(0.16\times10^{-15}\text{ G})}{\sqrt{N_\gamma(E_\gamma)}}\left[1+\left(\frac{E_\gamma}{7.9\text{ GeV}}\right)^{-1.62}\right]^{1/2}\left(\frac{E_\gamma}{10\text{ GeV}}\right)^{3/2}\left(\frac{d_s}{1\text{ Gpc}}\right)\left(\frac{1+z_s}{1.24}\right)^3 \tag{5.5}$$

which is shown in Fig. 11. For smaller B_0 the cascade halo is too small to distinguish from a point source given an angular resolution comparable to the Fermi-LAT. We show a scenario with small photon counts, $N_\gamma = 1$, and large photon counts, $N_\gamma = 1+10^3(E_\gamma/\text{GeV})^{-2}$, which could potentially be achieved in a stacked halo analysis [22, 24]. The E_γ^{-2} dependence is included as a crude model of the gamma ray flux: lower energy gamma rays are more abundant.

The preceding discussion yields a range of field strengths where measurements of cascade halo sizes are best suited to probing the IGMF. This range depends on the energy of the observed gamma rays, as illustrated in Fig. 11, and taking the extremal values $1\text{ GeV} < E_\gamma < 100\text{ GeV}$ leads to $10^{-18}\text{ G} < B_0 < 10^{-13}\text{ G}$. However, the range also depends on photon fluxes. For instance, probing $B_0 \gtrsim 10^{-14}$ G requires a sufficient abundance of high energy photons with $E_\gamma \gtrsim 10$ GeV, which remain in the MBC regime, and probing $B_0 \lesssim 10^{-17}$ G requires a large number of low energy photons $E_\gamma \lesssim 10$ GeV so as to beat down the effective angular resolution. More conservatively speaking, the range of magnetic field strengths that can be probed using cascade halo size measurements lies between $10^{-18}-10^{-17}$ G and $10^{-14}-10^{-13}$ G. This range is indicated on the parameter space plots of Fig. 12 as the blue shaded region. The gradient in the blue color indicates where the boundaries depend on the details of the analysis, just discussed. This range is independent of the magnetic field coherence length λ as long as λ is sufficiently large that the charged lepton probes an effectively homogeneous magnetic field, *i.e.*, $\lambda \gtrsim D_e \sim 100$ kpc. For smaller coherence length, the charged leptons do not follow helical arcs; instead their motion is that of a random walk, and our analysis breaks down.

[3] This formula matches Fig. 2 of [43] for the case of on-orbit data (P6_V11) and front-converting events.

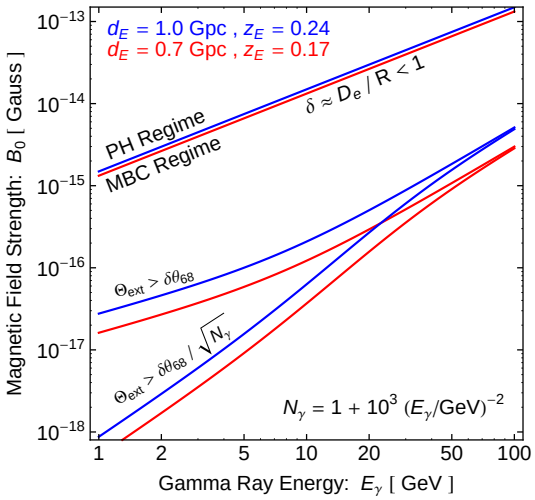

Figure 11. The reach of a gamma ray telescope like Fermi-LAT to probing an IGMF of strength B_0 using cascade gamma rays of energy E_γ. We consider blazars located at both $d_s = 1$ Gpc (blue) and 0.7 Gpc (red) from Earth. The two upper lines labeled $\delta \approx D_e/R < 1$ demarcate the boundary between the MBC regime, where a measurement of the halo size furnishes a measurement of the field strength, and the PH regime, where the halo size is insensitive to the field strength. The two middle curves labeled $\Theta_{\rm ext} > \delta\theta_{68}$ indicate the field strength below which the halo size is smaller than the Fermi-LAT 68% confinement radius. Finally, we suppose that many cascade photons are observed $N_\gamma = 1 + 10^3 (E_\gamma/\mathrm{GeV})^{-2}$, and the two lower curves, $\Theta_{\rm ext} > \delta\theta_{68}/\sqrt{N_\gamma}$, indicate the approximate field strength below which the halo cannot be resolved.

Short Coherence Length Regime

When the coherence length is small, information about spatial inhomogeneities is encoded in features of the halo at small angular scales. It is challenging to measure this halo substructure given limitations on angular resolution and photon flux. If one is only able to measure the large scale halo morphology, then gamma ray observations have a diminished capacity to probe magnetic helicity on small length scales.

As a specific example, recall the study of helical magnetic fields in Case 4 of Sec. 4.4. In the left panel of Fig. 7, the halo of GeV gamma rays form a spiral around the blazar. For small λ the angular scale of the spiral (separation between subsequent cycles) is smaller than the angular scale of the full halo. If we coarse grain on the scale of the halo, *e.g.* to model the finite detector resolution, the map would appear rotationally symmetric, but when the small scale behavior is resolved, the spiral can be seen.

To study the small scale structure of the halo, we require not only good angular resolution, but also closely spaced gamma ray energies. Recall from Eqs. (2.3) and (2.5) that the mean free path of the TeV gamma ray is

$$d_{\gamma_0} \simeq (180\ \mathrm{Mpc}) \left(\frac{1+z_s}{1.24}\right)^{-1} \left(\frac{E_\gamma}{10\ \mathrm{GeV}}\right)^{-1/2} \tag{5.6}$$

where E_γ is the energy of the GeV gamma ray reaching Earth. Each TeV gamma ray emitted by the blazar can be viewed as a local probe of the magnetic field at the point where pair production occurs. If we want to probe the magnetic field on the length scale λ, then we should employ gamma rays with energies $E_1 = E_\gamma$ and $E_2 = E_\gamma + \Delta E_\gamma$ such that $d_{\gamma_0}(E_1) - d_{\gamma_0}(E_2) \approx \lambda$. The optimal energy

separation is therefore

$$\Delta E_\gamma \approx 1\ \mathrm{GeV} \left(\frac{E_\gamma}{10\ \mathrm{GeV}}\right)^{3/2} \left(\frac{\lambda}{10\ \mathrm{Mpc}}\right) \left(\frac{1+z_s}{1.24}\right) \tag{5.7}$$

assuming $\Delta E_\gamma \ll E_\gamma$ for which $\lambda \approx (\mathrm{d}d_s/\mathrm{d}E_\gamma)\Delta E_\gamma$. Considerations of gamma ray flux prohibit one from taking ΔE_γ arbitrarily small to probe arbitrarily small λ.

Since the Q-statistics are constructed from three different gamma ray energies, as in Eq. (3.23), the associated ΔE_γ and E_γ for a given statistic lead to a corresponding optimal coherence length $\lambda_{\rm opt}$ via Eq. (5.7). The effects of inhomogeneities on smaller length scales, $\lambda < \lambda_{\rm opt} \approx d_{\gamma_0}(E_\gamma) - d_{\gamma_0}(E_\gamma + \Delta E_\gamma)$, are washed out. We have seen this behavior explicitly for Case 4 in Fig. 8 and Case 5 in Fig. 10: as the coherence length decreases below $\lambda \sim \lambda_{\rm opt}$, the Q-statistic stops growing and turns over or flattens out. To ensure that the optimal coherence length of a particular Q-statistic is small enough to probe magnetic field inhomogeneities on the scale λ, we impose

$$\lambda \gtrsim 10\ \mathrm{Mpc} \left(\frac{\Delta E_\gamma}{1\ \mathrm{GeV}}\right) \left(\frac{E_\gamma}{10\ \mathrm{GeV}}\right)^{-3/2} \left(\frac{1+z_s}{1.24}\right)^{-1}\ , \tag{5.8}$$

While a non-zero measurement of Q is of interest even for smaller λ, it becomes difficult to draw a connection between the value of Q and the parameters of the underlying magnetic field. This is particularly evident in Fig. 8 where Q oscillates with varying λ, and even $\mathrm{sign}[Q]$ does not necessarily correspond to the sign of the magnetic helicity. Presumably detailed modeling of the cascade development would be required to infer properties of the IGMF at such small λ.

In the parameter space plots of Fig. 12 we indicate the coherence length below which oscillations set in by writing "Q oscillations." This boundary is drawn at the point where each of the Q-statistics has its first zero-crossing, determined numerically, but it is well-approximated by Eq. (5.8). For the energies and energy spacings that we have shown, the Q-statistics are best suited to probe coherence lengths in the range $10 - 100$ Mpc. In order to probe shorter coherence lengths, Eq. (5.8) indicates that we need to consider Q-statistics built from larger E_γ and smaller ΔE_γ. However, this limit is technically challenging: higher energy gamma rays are less abundant and this problem is exacerbated by small energy bins.

Long Coherence Length Regime

We have already seen in the analyses of Secs. 4.4 and 4.5 that the Q-statistic becomes small as the coherence length is made large. In this limit, the magnetic field is effectively homogeneous on the scale probed by the cascade, and information about the magnetic helicity, which is encoded in gradients, becomes inaccessible. In terms of the halo map, the characteristic "S"-like curve becomes flattened out (see the left panel of Fig. 7 or 9), and the transverse angular extent of the halo becomes small. As a consequence angular resolution is a challenge for a measurement of small Q since one needs not only to distinguish the halo from a point, but also distinguish the flattened S from a line.

Recall that the triple product $Q_{abc} = \hat{\mathbf{n}}_a \times \hat{\mathbf{n}}_b \cdot \hat{\mathbf{n}}_c$ can be written as in Eq. (3.24)

$$Q_{abc} = \sin\vartheta_{ab}\ \sin\varphi_{abc} \tag{5.9}$$

where ϑ_{ab} is the angle between $\hat{\mathbf{n}}_a$ and $\hat{\mathbf{n}}_b$, and φ_{abc} is the angle between $\hat{\mathbf{n}}_c$ and the plane spanned by the other two vectors. Because of the detector's finite angular resolution, the unit vectors can only be localized on the sphere with a precision of $\delta\theta_{68}$. This translates into a comparable uncertainty on ϑ_{ab} and φ_{abc}, and we estimate the precision with which the Q-statistic can be measured as

$$\delta Q_{abc} \approx \sqrt{\cos^2\vartheta_{ab}\ \sin^2\varphi_{abc} + \sin^2\vartheta_{ab}\ \cos^2\varphi_{abc}}\ \delta\theta_{68}\ . \tag{5.10}$$

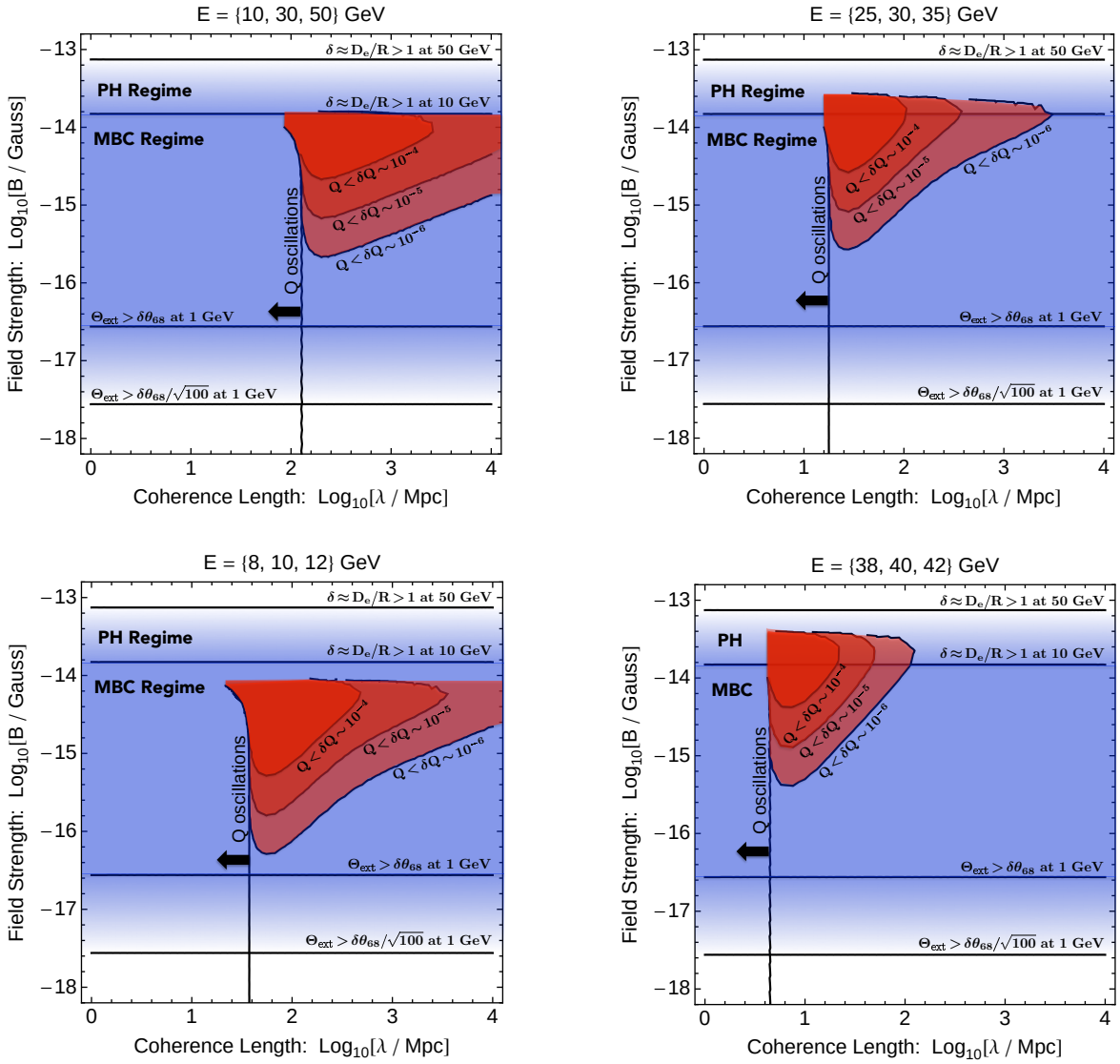

Figure 12. This sensitivity plot gives a rough indication of the region of magnetic field parameter space that can be probed with a gamma ray telescope like the Fermi-LAT using four different Q-statistics, Eq. (3.23). In the blue shaded regions, the magnetic field strength can be inferred from a measurement of the halo *size*. For smaller field strength the halo cannot be distinguished from a point source, and for larger field strength (PH regime), one may still be able to measure the halo size, but its connection to the field strength is obscured, see Fig. 4. In the red shaded region, the magnetic helicity can be inferred from a measurement of the halo *shape* via the Q-statistic. (These curves are defined with respect to the helical field configuration in Case 4 of Sec. 4.4, and we expect similar boundaries for more general helical configurations.) For larger λ and smaller B the parity-odd signal is too small to measure given the assumed angular resolution, and for smaller λ one may still be able to measure the halo shape via Q, but its connection to the coherence length is obscured, see Fig. 8.

The factor on the right side depends on the choice of gamma ray energies used to form the triplet as well as the specific magnetic field configuration being considered. As a rough estimate, we assume the small angle approximation and take $\vartheta_{ab} \sim \varphi_{abc} \sim \Theta_{\rm ext}(10\ {\rm GeV})$ using Eq. (3.8). Thus we infer that the Fermi-LAT angular resolution is satisfactory to measure a Q-statistic satisfying

$$Q > \delta Q \sim \Theta_{\rm ext}(10\ {\rm GeV})\,\delta\theta_{68} \sim 3\times 10^{-5}\frac{B_0}{10^{-15}\ {\rm G}}\ , \tag{5.11}$$

which is analogous to Eq. (5.4) for the halo size. As per the discussion surrounding Eq. (5.4), we expect that large photon counts can weaken the bound. Since this is a rough estimate, we consider $\delta Q \simeq 10^{-4}, 10^{-5}$, and 10^{-6} in the following analysis.

In Fig. 12 we also show the boundaries corresponding to the inequality in Eq. (5.11) for each of the energy combinations considered previously. The heavy red shaded region corresponds to $Q > \delta Q \sim 10^{-4}$ where gamma ray observations of cascade halos should be able to measure the strength and helicity of the magnetic field. Moving outside of the red shaded region, the helicity measurement becomes more challenging, because either Q is too small to measure given an angular resolution comparable to the Fermi-LAT (large λ regime) or Q has a complicated dependence on the magnetic helicity (small λ regime). Outside of the blue shaded region, even the field strength measurement becomes challenging, because either the halo is too small to distinguish from a point source given the angular resolution (small B_0 regime) or the halo size is insensitive to the field strength (large B_0 regime). With better statistics one can presumably beat down the angular resolution issues to probe weaker fields.

6 Summary and Discussion

One day gamma ray observations may provide measurements of both the size and the shape of cascade halos. Whereas the halo size is tied to the average magnetic field strength, the halo shape can depend on other properties of the magnetic field, particularly its helicity. In this work we have attempted to illuminate the connection between magnetic helicity and halo shape by studying the development of the electromagnetic cascade for simplified magnetic field configurations. We have taken an analytic approach in which the three equations, (3.17), (3.18), and (3.19), are solved to find the trajectory of gamma rays reaching Earth.

By acting as local probes of the magnetic field, blazar-induced cascade halos offer an interesting opportunity to measure properties of the IGMF. The physical parameters of the cascade and the detector resolution come together to determine the region of magnetic field parameter space that can be probed with this technique. By studying a particular helical magnetic field configuration with its wavevector oriented along the line of sight to the blazar, we have estimated the boundaries of this region in the two dimensional parameter space consisting of field strength B_0 and coherence length λ; see Fig. 12. The boundaries are partly determined by the anticipated detector resolution, and an improvement in angular resolution would allow access to smaller B_0 and larger λ. The boundaries are also depend on the measurement strategy, namely the energy combinations used to form the parity-odd Q-statistics via Eq. (3.23). For the four statistics considered here, we find that blazar-induced cascade halos are well-suited to probing helical magnetic fields with coherence length $\lambda \gtrsim 10\ {\rm Mpc}$, see Eq. (5.8). Reaching smaller length scales would require more closely spaced energy combinations, and then one runs into issues with angular resolution and photon flux.

There are a number of avenues for future work. First, we have considered only the simplest magnetic field configurations in order to draw the connection between halo shape and magnetic helicity. These configurations are not likely to be realistic models of the true IGMF. It would be straightforward to generalize our analysis and consider field configurations built from multiple modes in

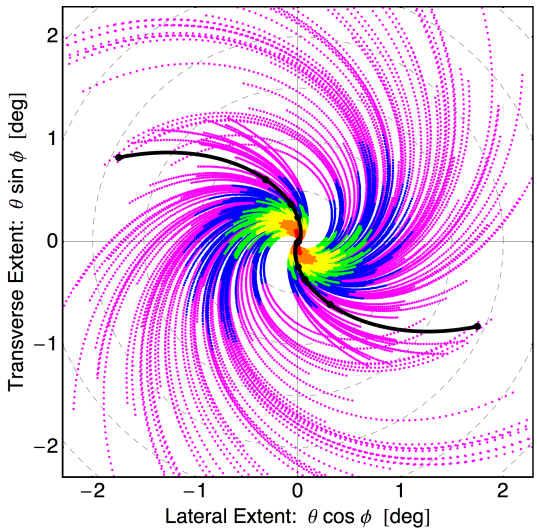

Figure 13. A "smeared" halo map for Case 4 with parameters $B_0 = 10^{-15}$ G, $\lambda = 700$ Mpc, and $d_s = 1$ Gpc. The same parameters correspond to the the green curve in Fig. 7 (left panel), which is also shown on this figure as a black curve. The colors denote different gamma ray energies: $5 < E_\gamma/\text{ GeV} < 10$ (magenta), $10 < E_\gamma/\text{ GeV} < 15$ (blue), $15 < E_\gamma/\text{ GeV} < 20$ (green), $20 < E_\gamma/\text{ GeV} < 30$ (yellow), $30 < E_\gamma/\text{ GeV} < 50$ (orange), and $50 < E_\gamma/\text{ GeV} < 100$ (red).

random orientations. In this case, it may become too difficult to solve Eqs. (3.17)-(3.19), and instead one would want to simulate the cascade using a shooting algorithm, such as in the analysis of Ref. [15].

Second, the parity-odd test statistic Q is not an automatic proxy for the magnetic helicity. We have seen in Sec. 4.4 that even the sign of Q won't necessarily equal the sign of the magnetic helicity if the coherence length is small. (This was also pointed out in [37, 39] in the case of statistically homogeneous and isotropic magnetic fields.) The statistic Q can still be useful in this regime if the three energies at which it is evaluated are finely spaced, though then the photon counts may be small and the resulting error bars will be large. Alternately, now that we know qualitatively the effects of magnetic field helicity in the gamma ray pattern, it may be possible to devise improved statistics to detect these patterns.

Third, our analysis neglects the stochastic nature of the cascade's development. This simplification allows us to calculate the halo map $\hat{\mathbf{n}}(E_\gamma)$ that gives a deterministic connection between the energy of a gamma ray and its arrival direction on the sky. A more realistic model would account for the stochasticity in the gamma ray propagation distances, spectra, and so on. As a crude model of this effect, we calculate a halo map for Case 4 by replacing d_{γ_0} and D_e in Eqs. (3.17)-(3.19) with $r_1 d_{\gamma_0}$ and $r_2 D_e$ where r_1 and r_2 are randomly selected from a uniform probability distribution over the range $0.5 < r_1, r_2 < 1.5$. We calculate 100 such halo maps in this way for the same parameter set and show the results in Fig. 13. Although the degree of smearing is significant, the halo map retains the qualitative features that we have seen previously, namely it angular extent and parity-odd spiraling shape.

Fourth, our analysis neglects the jet structure of the blazar, and assumes that the emission is isotropic. It is straightforward to include the effect of the jet in our semi-analytic formalism. Let

$$\hat{\mathbf{n}}_{\text{jet}}(\theta_{\text{off}}, \phi_{\text{off}}) = \sin(\theta_{\text{off}})\hat{\boldsymbol{\rho}}(\theta_{\text{off}}, \phi_{\text{off}}) - \cos(\theta_{\text{off}})\hat{\mathbf{z}}(\theta_{\text{off}}, \phi_{\text{off}}) \tag{6.1}$$

be the orientation of the jet where $\theta_{\rm off}$ and $\phi_{\rm off}$ give the polar and azimuthal offset angles from the line of sight. The orientation of the initial TeV gamma ray, $\hat{\mathbf{v}}_i$, is given by Eq. (3.20). Then the jet criterion,

$$\arccos\left[\hat{\mathbf{v}}_i \cdot \hat{\mathbf{n}}_{\rm jet}\right] < \theta_{\rm jet} \ , \tag{6.2}$$

ensures that the TeV gamma ray is emitted from within the cone of the jet. As a consequence of the jet criterion, only a portion of the full halo map becomes illuminated. This behavior is shown in Fig. 14 where we have reevaluated the halo maps from Case 4 using Eq. (6.2), and there is a similar effect for the other cases. When $\theta_{\rm off} > \theta_{\rm jet}$ the Earth is not contained within the cone of the jet, an the blazar itself may not be seen. In this regime, the cascade photons would appear to contribute to the diffuse gamma ray flux.

Acknowledgments

We acknowledge illuminating discussions with Sean Bryan, James Buckley, Daniel Chung, Manel Errando, Francesc Ferrer, and Angela Olinto. We are especially grateful to Borun Chowdhury for discussion through the course of this project. TV gratefully acknowledges the Clark Way Harrison Professorship at Washington University during the course of this work. This work was supported by the Department of Energy at ASU. A.J.L. was also supported in part by the National Science Foundation under grant number PHY-1205745.

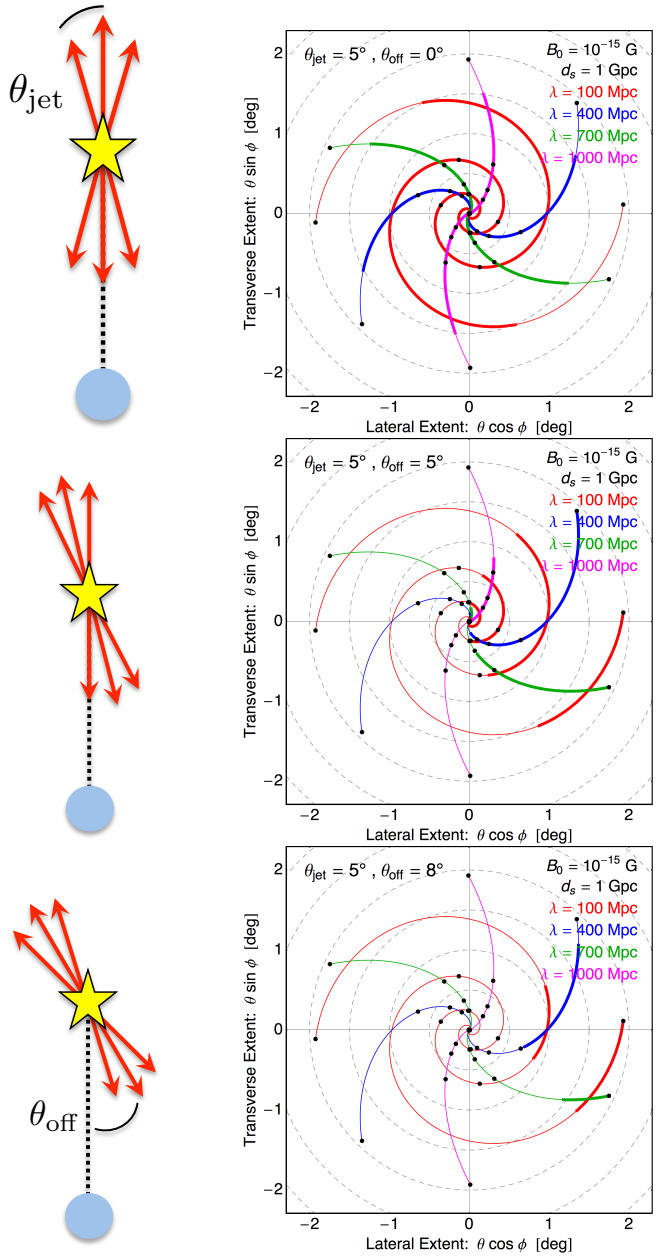

Figure 14. The effect of the finite jet angle, $\theta_{\rm jet}$, and offset from the line of sight, $\theta_{\rm off}$, where $\phi_{\rm off} = 0$. The thin colored curves show the halo map for an isotropically emitting blazar, $\theta_{\rm jet} = 180°$, and the thick curves show the halo map for a blazar jet with $\theta_{\rm jet} = 5°$ and different offsets $\theta_{\rm off} = 0°, 5°$, and $8°$ in the three rows, respectively. For $\theta_{\rm off} > 10°$ none of the halo is visible with $E_\gamma > 5$ GeV.

References

[1] P. P. Kronberg, *Extragalactic magnetic fields*, Rept.Prog.Phys. **57** (1994) 325–382.

[2] L. M. Widrow, *Origin of galactic and extragalactic magnetic fields*, Rev.Mod.Phys. **74** (2002) 775–823, [astro-ph/0207240].

[3] L. M. Widrow, D. Ryu, D. R. Schleicher, K. Subramanian, C. G. Tsagas, et al., *The First Magnetic Fields*, Space Sci.Rev. **166** (2012) 37–70, [arXiv:1109.4052].

[4] T. Vachaspati, *Magnetic fields from cosmological phase transitions*, Phys.Lett. **B265** (1991) 258–261.

[5] J. Ahonen and K. Enqvist, *Magnetic field generation in first order phase transition bubble collisions*, Phys.Rev. **D57** (1998) 664–673, [hep-ph/9704334].

[6] J. M. Cornwall, *Speculations on primordial magnetic helicity*, Phys.Rev. **D56** (1997) 6146–6154, [hep-th/9704022].

[7] T. Vachaspati, *Estimate of the primordial magnetic field helicity*, Phys.Rev.Lett. **87** (2001) 251302, [astro-ph/0101261].

[8] A. J. Long, E. Sabancilar, and T. Vachaspati, *Leptogenesis and Primordial Magnetic Fields*, JCAP **1402** (2014) 036, [arXiv:1309.2315].

[9] M. S. Turner and L. M. Widrow, *Inflation Produced, Large Scale Magnetic Fields*, Phys.Rev. **D37** (1988) 2743.

[10] R. Durrer and A. Neronov, *Cosmological Magnetic Fields: Their Generation, Evolution and Observation*, Astron.Astrophys.Rev. **21** (2013) 62, [arXiv:1303.7121].

[11] K. Subramanian, *The origin, evolution and signatures of primordial magnetic fields*, arXiv:1504.0231.

[12] F. A. Aharonian, P. S. Coppi, and H. J. Voelk, *Very high energy gamma rays from active galactic nuclei: Cascading on the cosmic background radiation fields and the formation of pair halos*, Ap. J. Lett. **423** (Mar., 1994) L5–L8, [astro-ph/9312045].

[13] R. Plaga, *Detecting intergalactic magnetic fields using time delays in pulses of γ-rays*, Nature **374** (Mar., 1995) 430–432.

[14] A. Neronov and D. V. Semikoz, *A method of measurement of extragalactic magnetic fields by TeV gamma ray telescopes*, JETP Lett. **85** (2007) 473–477, [astro-ph/0604607].

[15] A. Elyiv, A. Neronov, and D. Semikoz, *Gamma-ray induced cascades and magnetic fields in intergalactic medium*, Phys.Rev. **D80** (2009) 023010, [arXiv:0903.3649].

[16] K. Dolag, M. Kachelriess, S. Ostapchenko, and R. Tomas, *Blazar halos as probe for extragalactic magnetic fields and maximal acceleration energy*, Astrophys.J. **703** (2009) 1078–1085, [arXiv:0903.2842].

[17] J. Aleksic, L. Antonelli, P. Antoranz, M. Backes, C. Baixeras, et al., *Search for an extended VHE gamma-ray emission from Mrk 421 and Mrk 501 with the MAGIC Telescope*, Astron.Astrophys. **524** (2010) A77, [arXiv:1004.1093].

[18] **HESS** Collaboration, A. Abramowski et al., *Search for Extended gamma-ray Emission around AGN with H.E.S.S. and Fermi-LAT*, Astron.Astrophys. **562** (2014) 145, [arXiv:1401.2915].

[19] **VERITAS** Collaboration, M. F. Alonso, *Search for extended gamma ray emission in Markarian 421 using VERITAS observations*, arXiv:1406.4764.

[20] A. Neronov and I. Vovk, *Evidence for strong extragalactic magnetic fields from Fermi observations of TeV blazars*, Science **328** (2010) 73–75, [arXiv:1006.3504].

[21] F. Tavecchio, G. Ghisellini, L. Foschini, G. Bonnoli, G. Ghirlanda, et al., *The intergalactic magnetic field constrained by Fermi/LAT observations of the TeV blazar 1ES 0229+200*, Mon.Not.Roy.Astron.Soc. **406** (2010) L70–L74, [arXiv:1004.1329].

[22] S. Ando and A. Kusenko, *Evidence for Gamma-Ray Halos Around Active Galactic Nuclei and the First Measurement of Intergalactic Magnetic Fields*, Astrophys.J. **722** (2010) L39, [arXiv:1005.1924].

[23] W. Essey, S. Ando, and A. Kusenko, *Determination of intergalactic magnetic fields from gamma ray data*, Astroparticle Physics **35** (Oct., 2011) 135–139, [arXiv:1012.5313].

[24] W. Chen, J. H. Buckley, and F. Ferrer, *Evidence for GeV Pair Halos around Low Redshift Blazars*, arXiv:1410.7717.

[25] C. Caprini and L. Sorbo, *Adding helicity to inflationary magnetogenesis*, JCAP **1410** (2014), no. 10 056, [arXiv:1407.2809].

[26] K. Atmjeet, T. R. Seshadri, and K. Subramanian, *Helical cosmological magnetic fields from extra-dimensions*, arXiv:1409.6840.

[27] N. Bartolo, S. Matarrese, M. Peloso, and M. Shiraishi, *Parity-violating and anisotropic correlations in pseudoscalar inflation*, JCAP **1501** (2015), no. 01 027, [arXiv:1411.2521].

[28] S.-L. Cheng, W. Lee, and K.-W. Ng, *Inflationary dilaton-axion magnetogenesis*, arXiv:1409.2656.

[29] T. Fujita, R. Namba, Y. Tada, N. Takeda, and H. Tashiro, *Consistent generation of magnetic fields in axion inflation models*, JCAP **1505** (2015), no. 05 054, [arXiv:1503.0580].

[30] L. Campanelli, *Lorentz-violating inflationary magnetogenesis*, arXiv:1503.0741.

[31] R. Jackiw and S.-Y. Pi, *Creation and evolution of magnetic helicity*, Phys.Rev. **D61** (2000) 105015, [hep-th/9911072].

[32] M. Joyce and M. E. Shaposhnikov, *Primordial magnetic fields, right-handed electrons, and the Abelian anomaly*, Phys.Rev.Lett. **79** (1997) 1193–1196, [astro-ph/9703005].

[33] A. Diaz-Gil, J. Garcia-Bellido, M. Garcia Perez, and A. Gonzalez-Arroyo, *Magnetic field production during preheating at the electroweak scale*, Phys.Rev.Lett. **100** (2008) 241301, [arXiv:0712.4263].

[34] T. Kahniashvili, A. G. Tevzadze, A. Brandenburg, and A. Neronov, *Evolution of Primordial Magnetic Fields from Phase Transitions*, Phys.Rev. **D87** (2013) 083007, [arXiv:1212.0596].

[35] T. Kahniashvili and T. Vachaspati, *On the detection of magnetic helicity*, Phys.Rev. **D73** (2006) 063507, [astro-ph/0511373].

[36] H. Tashiro and T. Vachaspati, *Cosmological magnetic field correlators from blazar induced cascade*, Phys.Rev. **D87** (2013), no. 12 123527, [arXiv:1305.0181].

[37] H. Tashiro and T. Vachaspati, *Parity-odd correlators of diffuse gamma rays and intergalactic magnetic fields*, Mon.Not.Roy.Astron.Soc. **448** (2015) 299, [arXiv:1409.3627].

[38] H. Tashiro, W. Chen, F. Ferrer, and T. Vachaspati, *Search for CP Violating Signature of Intergalactic Magnetic Helicity in the Gamma Ray Sky*, arXiv:1310.4826.

[39] W. Chen, B. D. Chowdhury, F. Ferrer, H. Tashiro, and T. Vachaspati, *Intergalactic magnetic field spectra from diffuse gamma rays*, arXiv:1412.3171.

[40] A. Neronov and D. Semikoz, *Sensitivity of gamma-ray telescopes for detection of magnetic fields in intergalactic medium*, Phys.Rev. **D80** (2009) 123012, [arXiv:0910.1920].

[41] **Planck Collaboration** Collaboration, P. Ade et al., *Planck 2015 results. XIII. Cosmological parameters*, arXiv:1502.0158.

[42] L. Costamante, *Gamma-rays from Blazars and the Extragalactic Background Light*, Int.J.Mod.Phys. **D22** (2013), no. 13 1330025, [arXiv:1309.0612].

[43] **Fermi-LAT** Collaboration, *Determination of the Point-Spread Function for the Fermi Large Area Telescope from On-orbit Data and Limits on Pair Halos of Active Galactic Nuclei*, arXiv:1309.5416.

Magnetic monopole - domain wall collisions

Micah Brush[1], Levon Pogosian[1], and Tanmay Vachaspati[2,3]
[1]*Department of Physics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada*
[2]*Department of Physics, Arizona State University, Tempe, Arizona 85287, USA and*
[3]*Department of Physics, Washington University, St. Louis, MO 63130, USA*

Interactions of different types of topological defects can play an important role in the aftermath of a phase transition. We study interactions of fundamental magnetic monopoles and stable domain walls in a Grand Unified theory in which SU(5) × Z_2 symmetry is spontaneously broken to SU(3) × SU(2) × U(1)/Z_6. We find that there are only two distinct outcomes depending on the relative orientation of the monopole and the wall in internal space. In one case, the monopole passes through the wall, while in the other it unwinds on hitting the wall.

I. INTRODUCTION

Grand Unified Theories (GUTs) are based on large symmetry groups, the smallest of which is an SU(5) model with an additional, possibly approximate, Z_2 symmetry. When such large symmetries are broken in a cosmological setting, several kinds of topological defects can be produced. The ensuing cosmology will depend critically on the interactions of the different defects. In particular, the SU(5) × Z_2 symmetry breaking leads to the existence of magnetic monopoles and domain walls in the aftermath of the phase transition. We expect the magnetic monopoles to interact with domain walls, potentially resolving the magnetic monopole over-abundance problem [1]. To investigate this idea further, we study the interactions of SU(5) monopoles and Z_2 domain walls in this paper.

The interaction of monopoles and domain walls was also studied in [2] with the domain wall structure given by

$$\Phi = \tanh\left(\frac{z}{w}\right)\Phi_0 \tag{1}$$

where the order parameter Φ is in the adjoint representation of SU(5), Φ_0 is its constant vacuum expectation value (VEV), and w is the width of the domain wall. By numerical evaluation it was found that monopoles hitting this domain wall will unwind and spread on the wall. Subsequently, however, it was found [3–6] that the model actually has several domain wall solutions, including the one in Eq. (1), and that the lightest (stable) wall has a different structure (see Sec. II B). Hence the interaction of the stable wall and the monopole needs to be revisited.

In Sec. II we provide details of the SU(5) × Z_2 model, the monopole solution, the wall solutions, and finally our scheme for setting up a configuration with a monopole and a domain wall together. This provides us with initial conditions that we numerically evolve in Sec. III. The complexity of the field equations and the problem requires some special numerical techniques that we briefly describe in Sec. III.

Our results are summarized in Sec. IV. Essentially we find that there are two internal space polarizations for the monopole with respect to the wall. One of the polarizations is able to pass through the wall with only some kinematic changes. The monopole with the other polarization is unable to pass through the domain wall and unwinds on the wall, radiating away its gauge fields. The disappearance of this monopole is further explained in Sec. IV.

II. THE MODEL

The SU(5) model we consider is given by the Lagrangian:

$$L = -\frac{1}{4}X^a_{\mu\nu}X^{a\mu\nu} + \frac{1}{2}D_\mu\phi^a D^\mu\phi^a - V(\Phi) \tag{2}$$

where $\Phi = \phi^a T^a$ ($a = 1, ..., 24$), $X^a_{\mu\nu}$ are the gauge field strengths defined as

$$X_{\mu\nu} = \partial_\mu X_\nu - \partial_\nu X_\mu - ig[X_\mu, X_\nu]\,, \tag{3}$$

$X_\mu = X^a_\mu T^a$ are the gauge fields and g is the coupling constant. T^a are the generators of SU(5) normalized by $\mathrm{Tr}(T^aT^b) = \delta_{ab}/2$. The covariant derivative is given by

$$D_\mu\phi^a = \partial_\mu\phi^a - ig[X_\mu, \Phi]^a\,. \tag{4}$$

The most general renormalizable SU(5) potential is

$$V(\Phi) = -m^2\mathrm{Tr}\Phi^2 + \gamma\mathrm{Tr}\Phi^3 + h(\mathrm{Tr}\Phi^2)^2 + \lambda\mathrm{Tr}\Phi^4 - V_0\,, \tag{5}$$

and we will assume that γ vanishes, giving the model an additional Z_2 symmetry. For $\lambda \geq 0$ and $h + 7\lambda/30 \geq 0$, the potential has its global minimum at [7]

$$\Phi_0 = \frac{\eta}{2\sqrt{15}}\mathrm{diag}(2, 2, 2, -3, -3), \tag{6}$$

with $\eta = m/\sqrt{h + 7\lambda/30}$. The VEV, Φ_0, spontaneously breaks the SU(5) symmetry to SU(3) × SU(2) × U(1)/Z_6.

In what follows, the four diagonal generators of SU(5) are chosen to be

$$\begin{aligned}
\lambda_3 &= \frac{1}{2}\mathrm{diag}(1, -1, 0, 0, 0),\\
\lambda_8 &= \frac{1}{2\sqrt{3}}\mathrm{diag}(1, 1, -2, 0, 0),\\
\tau_3 &= \frac{1}{2}\mathrm{diag}(0, 0, 0, 1, -1),\\
Y &= \frac{1}{2\sqrt{15}}\mathrm{diag}(2, 2, 2, -3, -3).
\end{aligned} \tag{7}$$

We use $a = 1,2,3$ to denote generators $T^a = \tau_a = \mathrm{diag}(0,0,0,\sigma_a/2)$ where σ_a are the Pauli spin matrices.

A. The monopole

Let us consider a magnetic monopole whose winding lies in the 4-5 block of Φ. This is possible [8] if we take the VEV along one of the radial directions far away from the monopole to be

$$\begin{aligned}\Phi_\infty &= \frac{\eta}{2\sqrt{15}}\mathrm{diag}(2,-3,2,2,-3)\\ &= \eta\sqrt{\frac{5}{12}}(\lambda_3+\tau_3)+\frac{\eta}{6}(Y-\sqrt{5}\lambda_8).\end{aligned} \tag{8}$$

The monopole ansatz for the scalar field can be written as [3]

$$\Phi_M(r) = P(r)\sum_{a=1}^{3}x^a\tau_a + M(r)\left(\frac{\sqrt{3}}{2}\lambda_3-\frac{1}{2}\lambda_8\right)+N(r)Y, \tag{9}$$

while the non-zero gauge fields can be written as

$$X_i^a = \epsilon_{ij}^a\frac{x^j}{gr^2}(1-K(r)),\ (a=1,2,3) \tag{10}$$

and $P(r), M(r), N(r)$, and $K(r)$ are profile functions that depend only on the spherical radial coordinate $r = \sqrt{x^2+y^2+z^2}$ and satisfy the boundary conditions:

$$\begin{aligned}\lim_{r\to\infty} rP(r) = \eta\sqrt{\frac{5}{12}},\ M(\infty) = \eta\frac{\sqrt{5}}{3},\\ N(\infty) = \frac{\eta}{6},\quad K(\infty) = 0.\end{aligned} \tag{11}$$

The profile functions for the monopole alone were evaluated numerically and are shown in Fig. 1.

The non-Abelian magnetic field can be defined as [9]

$$B^k = -\frac{1}{2}\epsilon^{ijk}X_{ij}$$

with the associated energy density given by $\mathrm{Tr}(B_kB^k)$. Far away from the centre, the monopole field becomes $B^k \to Qx^k/(gr^3)$, with $Q = \tau^j x^j/r$.

The monopole charge Q includes a component along the generator of the unbroken U(1) symmetry (Φ_∞ of Eq. (8)), as well as SU(2) and SU(3) magnetic charges. The U(1) part of the magnetic field, which is a defining feature of a topological SU(5) monopole, is given by

$$B_Y^k = -\frac{1}{2}\epsilon^{ijk}X_{ij}^a\hat{\phi}^a \tag{12}$$

where $\hat{\phi}^a \equiv \phi^a/\sqrt{\phi^b\phi^b}$. As discussed in [10], other definitions of the Abelian magnetic field are possible, and these differ from our definition but only within the core of the monopole. Since we only use our definition to plot the long range Abelian magnetic field (see Fig. 3) the definition in Eq. (12) is sufficient.

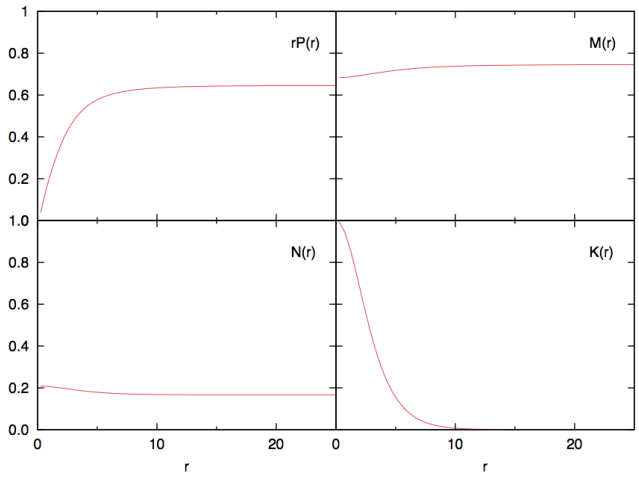

FIG. 1. The profile functions for the monopole alone, evaluated numerically, for a model with $\eta = 1$, $h/\lambda = -0.2$ and $\lambda = 0.5$.

B. The wall

Without loss of generality [4], the domain wall solution can be taken to be diagonal at all z and written in terms of the diagonal generators of SU(5) as

$$\Phi_{DW}(z) = a(z)\lambda_3 + b(z)\lambda_8 + c(z)\tau_3 + d(z)Y\,. \tag{13}$$

In each of the two disconnected parts of the vacuum manifold $\mathcal{M}$ there are a total of 10 different diagonal VEVs corresponding to all possible permutations of 2's and 3's in Eq. (6). Topology dictates that there must be a domain wall separating any pair of VEVs from the two disconnected parts of $\mathcal{M}$. However, not every such pair of VEVs corresponds to a stable domain wall solution. For instance, as shown in [3], the wall across which Φ_0 goes to $-\Phi_0$ is unstable and will decay into a lower energy stable wall. The stable domain walls are obtained when both 3's in Eq. (6) change into 2's across the wall.

Let us choose the boundary condition at $z = -\infty$ to be

$$\begin{aligned}\Phi_- = \Phi(z=-\infty) &= \frac{\eta}{2\sqrt{15}}\mathrm{diag}(2,-3,2,2,-3)\\ &= \eta\sqrt{\frac{5}{12}}(\lambda_3+\tau_3)+\frac{\eta}{6}(Y-\sqrt{5}\lambda_8)\,.\end{aligned} \tag{14}$$

For this choice of Φ_-, there are three different choices of $\Phi(z=+\infty)$, proportional to

$$\begin{aligned}&\mathrm{diag}(3,-2,-2,3,-2)\\ &\mathrm{diag}(-2,-2,3,3,-2)\\ &\mathrm{diag}(3,-2,3,-2,-2)\,,\end{aligned} \tag{15}$$

that lead to stable domain walls. For the purpose of understanding the monopole-wall interactions, it is sufficient to consider only two of the above, corresponding to

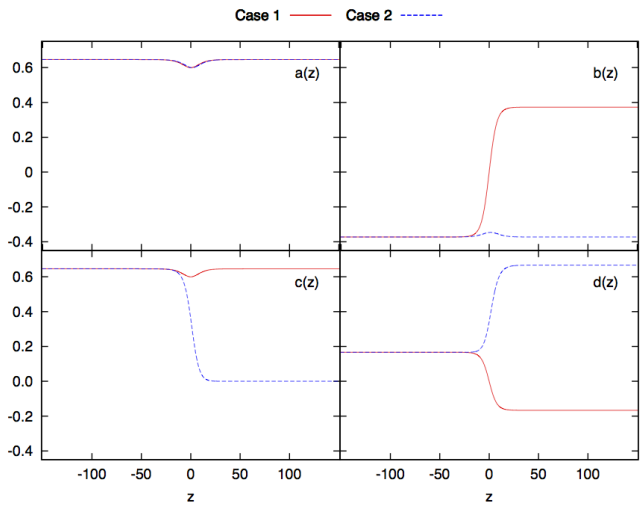

FIG. 2. The wall profile functions for Cases 1 and 2 for a model with $\eta = 1$, $h/\lambda = -0.2$ and $\lambda = 0.5$. Note that the profile function $c(z)$ goes to zero in Case 2, which gives an unbroken SU(2) $\subset$ SU(3) symmetry in the 4-5 block of $\Phi_+^{(2)}$. The profile function $a(z)$ is the same for Cases 1 and 2.

the two distinct entries in the 4-5 block of Φ. We take the first to be the same as in [3], subsequently referred to as Case 1:

$$\Phi_+^{(1)} = \frac{\eta}{2\sqrt{15}}\mathrm{diag}(3,-2,-2,3,-2) = \eta\sqrt{\frac{5}{12}}(\lambda_3+\tau_3) - \frac{\eta}{6}(Y-\sqrt{5}\lambda_8)\,. \quad (16)$$

The value of the field in the core of this wall is proportional to $\mathrm{diag}(1,-1,0,1,-1)$. The other case, subsequently referred to as Case 2, has

$$\Phi_+^{(2)} = \frac{\eta}{2\sqrt{15}}\mathrm{diag}(3,-2,3,-2,-2) = \eta\frac{\sqrt{15}}{6}\lambda_3 + \frac{\eta}{6}(4Y-\sqrt{5}\lambda_8)\,, \quad (17)$$

with the field in the wall being proportional to $\mathrm{diag}(1,-1,1,0,-1)$. A novel feature of these walls is that the unbroken symmetry groups on either side of the wall are isomorphic to each other but they are realized along different directions of the initial SU(5) symmetry group. Hence the wall is the location of a clash of symmetries [12].

Note that the symmetry within the wall is $[\mathrm{SU}(2)\times\mathrm{U}(1)]^2$. The SU(2)'s correspond to rotations in the 1-3 and 2-5 blocks and the U(1)'s to rotations along σ_3 in the 1-2 and 3-5 blocks. Therefore the symmetry group within the wall is 8-dimensional, and is smaller than the 12-dimensional symmetry outside the wall[1]. Also note that the symmetry in the 4-5 block is different for the $\Phi_+^{(1)}$ and $\Phi_+^{(2)}$ vacua. This is going to be of direct relevance for the fate of the monopoles.

The profile functions $a(z)$, $b(z)$, $c(z)$ and $d(z)$ for both cases are shown in Fig. 2. In each case, they are linear combinations of two functions $F_+(z)$ and $F_-(z)$ defined by the alternative way of writing the domain wall solution [3]

$$\Phi_{DW} = \frac{\Phi_+(z)-\Phi_-(z)}{2}F_-(z) + \frac{\Phi_+(z)+\Phi_-(z)}{2}F_+(z)\,, \quad (18)$$

where $F_+(\pm\infty) = 1$, $F_-(\pm\infty) = \pm 1$. For a general choice of parameters, functions $F_\pm(z)$ must be found numerically. For $h/\lambda = -3/20$, they are known in closed form [3]: $F_+(z) = 1$, $F_-(z) = \tanh(mz/\sqrt{2})$. Correspondingly, for this value of h/λ, the four functions $a(z)$, $b(z)$, $c(z)$, and $d(z)$ are either constant or describe a transition from one constant value to another. For $h/\lambda \neq -3/20$ the "constant" functions develop a small bump around $z = 0$ as can be seen in Fig. 2.

C. Monopole and Wall

As our initial configuration, we take the monopole to be on the $z = -\infty$ side, far away from the wall. In this case, the ansatz for the initial combined field configuration of the wall and the monopole can be written as [3]

$$\Phi_{M+DW} = P(r)\frac{c(z')}{c(-\infty)}\sum_{a=1}^{3}x^a\tau_a + N(r)\frac{d(z')}{d(-\infty)}Y + M(r)\left(\frac{\sqrt{3}}{2}\frac{a(z')}{a(-\infty)}\lambda_3 - \frac{1}{2}\frac{b(z')}{b(-\infty)}\lambda_8\right) \quad (19)$$

where $z' = \gamma(z-z_0)$, $\gamma = 1/\sqrt{1-v^2}$ is the boost factor, v is the wall velocity and z_0 is the initial position of the wall. The monopole is at $x = 0 = y = z$. It is easy to check that, far away from the monopole, the profile functions take on the values in Eq. (11) and $\Phi_{M+DW} \to \Phi_{DW}$. Close to the monopole, $z' \to -\infty$, since the monopole is initially very far from the wall, and $\Phi_{M+DW} \to \Phi_M$ as desired. We work in the temporal gauge, $X_0^a = 0$, and with the initial ansatz for the gauge fields given by Eq. (10) for both cases.

It is instructive to examine the difference in the nature of the magnetic field in Cases 1 and 2. As mentioned in Sec. II A, the charge of our monopole along the z-direction, $Q = (1/2)\mathrm{diag}(0,0,0,1,-1)$, is a combination of the U(1), the SU(2), and the SU(3) magnetic charges. Since the VEV of Φ in our model is along the generator

[1] For simplest domain walls, such as kinks in $\lambda\Phi^4$, the full symmetry of the Lagrangian is restored inside the core. However, the symmetry inside stable domain walls in $\mathrm{SU}(N)\times\mathrm{Z}_2$ is always lower than that of the vacuum [4].

of (hypercharge) U(1), the magnetic field, as defined in Eq. (12), corresponds solely to the U(1) component of the charge. In Case 1, $\mathrm{Tr}(Q\Phi)$ is the same on both sides of the wall and the U(1) magnetic field is unaffected by the presence of the DW. In Case 2, however, $\mathrm{Tr}(Q\Phi_+^{(2)}) = 0$ and there is no magnetic field corresponding to the unbroken U(1) on the $z = +\infty$ side of the wall. Instead, the gauge field on that side is associated with an SU(2) subgroup of the unbroken SU(3). We note that, while the magnetic *energy density* associated with the gauge field is unaffected by the presence of the wall, it is specifically the U(1) magnetic field that is a defining feature of a topologically stable monopole.

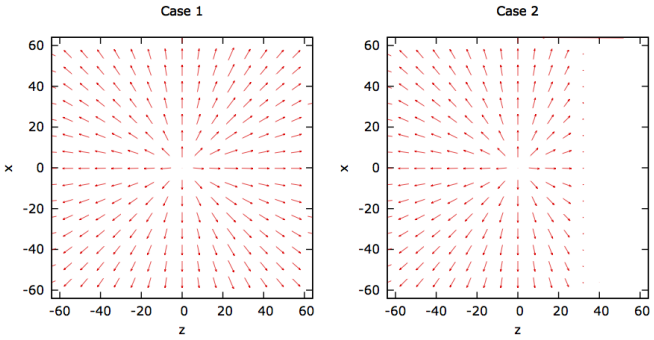

FIG. 3. The magnetic field $\mathbf{B}_Y$ (defined in Eq. (12)) multiplied by r^2 for Cases 1 and 2, where at each point $r^2 B_Y^z$ and $r^2 B_Y^x$ are plotted as a vector. In Case 1, there is a magnetic field associated with the unbroken U(1) symmetry on both sides of the wall. In Case 2, the magnetic field becomes associated with the SU(2) $\subset$ SU(3) on the $z = +\infty$ side on the wall, while its U(1) component vanishes. Note that it is the U(1) magnetic field that characterizes a topologically stable monopole.

The magnetic field, as defined in Eq. (12), is plotted for both cases in Fig. 3, where the vectors have components $r^2 B_Y^z$ and $r^2 B_Y^x$. This plot shows that, in Case 1, there is a U(1) magnetic field on both sides of the wall falling off as r^2 as expected, while in Case 2 the U(1) magnetic field is zero on the $z = +\infty$ side of the wall.

III. EVOLUTION

Let us consider an initial monopole-wall configuration given by Eq. (19) in which VEV at $z = -\infty$ is given by Φ_- in Eq. (14). As mentioned in the previous Section, there are 2 types of boundary conditions at $z = +\infty$, given by Eqs. (16) and (17), dubbed Case 1 and Case 2, leading to 2 different outcomes of the monopole-wall collision.

Before considering the two cases in detail, let us note that initially, when the monopole and the wall are very far away from each other, the field configuration has just three non-zero gauge fields and six scalar fields corresponding to the generators that appear in Eq. (19). Because these six generators form a closed algebra, it follows from the equations of motion that the subsequent evolution does not involve fields corresponding to the other 18 generators. Namely, the scalar and the gauge field equations are

$$D_\mu D^\mu \phi^a = -\partial V/\partial \phi^a \tag{20}$$

$$D_\mu X^{\mu\nu a} = g f_{abc} (D^\nu \Phi)^b \phi^c \tag{21}$$

where f_{abc} are the SU(5) structure constants defined by $[T^a, T^b] = i f_{abc} T^c$. Let $\mathcal{C}$ be the set of indices of the 6 generators that appear in the initial field configuration given by Eq. (19). Since the 6 generators form a closed algebra, $f_{abc} = 0$ for $a \notin \mathcal{C}$ and $b, c \in \mathcal{C}$. Now let ϕ^a and X_μ^a be fields corresponding to any $a \notin \mathcal{C}$. If ϕ^a and X_μ^a are zero at the initial time, they will remain zero if $f_{abc} = 0$ for $b, c \in \mathcal{C}$ and $\partial V/\partial \phi^a \neq 0$. The former condition is satisfied as mentioned above, while the latter holds since $\mathrm{Tr}[T^a T^b] \propto \delta_{ab}$ and $\mathrm{Tr}[T^a T^b T^c T^d] = 0$ for $b, c, d \in \mathcal{C}$, as we have checked by explicit evaluation. Thus, for our purposes, it is sufficient[2] to consider only $a \in \mathcal{C}$.

Our numerical implementation is based on techniques developed in [2]. First, the DW and the monopole profile functions are found via numerical relaxation. The monopole is initially located at the center of the lattice. We give the DW a velocity towards the monopole and boosted profiles are inserted into the initial configuration given by Eq. (19). With the initial time derivatives simply determined from the Lorentz boost factor, this initial configuration is evolved forward in time using a staggered leapfrog code. The boundary conditions require special care since the wall extends all the way across the lattice. We have implemented boundary conditions in which the field is extrapolated across the boundary. We have numerically tested that this boundary condition leads to a smoothly evolving domain wall, without any spurious incoming radiation. Even though our problem has axial symmetry, we work in Cartesian coordinates as this offers superior stability. However, as discussed in [13], we take advantage of the axial symmetry of our configuration to restrict the lattice to just three lattice spacings along the y direction. We then use a 256×256 lattice grid for the x and z coordinates. Additionally, the axial symmetry allows us to solve only for positive x and use reflection to find the fields at negative x. The units of length are set by $\eta = 1$ and we take each lattice spacing to correspond to half of a length unit. In these units, the range of x and z axis for a 256×256 grid is $[-64, 64]$. Note that in some figures we do not plot the entire lattice. The radius of the monopole core is about 10 length units and is about the same as a half of the domain wall width. At the initial time, the wall is 30 length units away from the center of the monopole.

[2] Although the field components for $a \notin \mathcal{C}$ continue to vanish during evolution if they vanish initially, we cannot exclude the possibility that the fields in these other directions may grow unstably if they did not vanish initially.

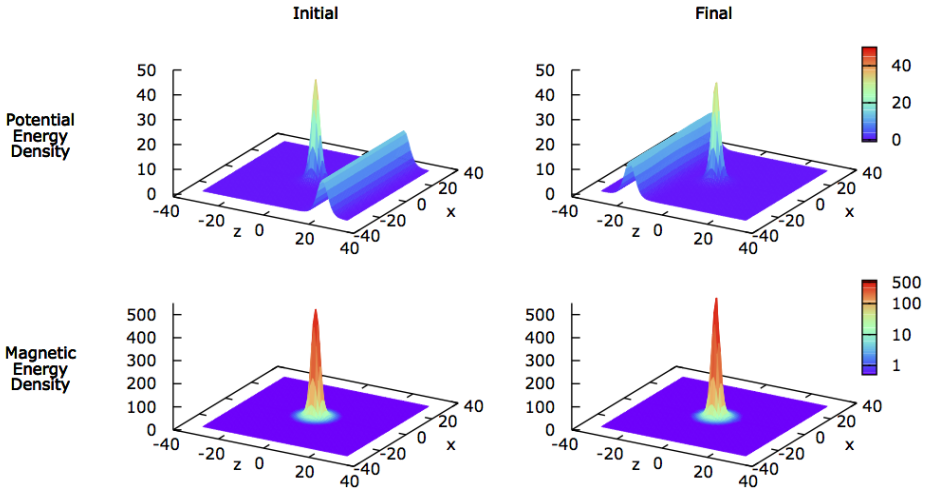

FIG. 4. The potential and magnetic energy densities in the xz plane for the colliding monopole and wall in Case 1. We see that the monopole passes through the wall and the energy densities remain localized. Additionally, we see the magnetic energy density is unchanged before and after the collision.

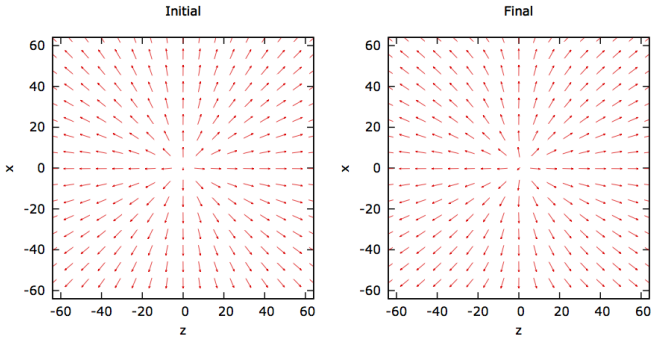

FIG. 5. The scalar field ϕ^a in the xz plane for the colliding monopole and wall in Case 1, where at each point ϕ^3 and ϕ^1 are plotted as a vector. In this case the scalar field arrangement in direction and magnitude remains virtually unchanged.

A. Case 1: the monopole passes through

It is not difficult to predict that the monopole in Case 1 will pass through the wall. The monopole winding is due to the fields in the SU(2) subgroup corresponding to generators τ_a, $a = 1,..,3$. In Eq. (19), these fields are multiplied by the function $c(z)$ which has the same value at $z = \pm\infty$ and, as known from [3], is approximately constant across the domain wall. Only $b(z)$ and $d(z)$ change signs across the wall, but these are irrelevant for the winding of the monopole. Thus, the presence of the wall is of no qualitative consequence to the winding of the monopole or its profile functions. The only effect is the small change in $c(z)$ around $z = 0$ (note that, as mentioned earlier, $c(z)$ is strictly a constant when $h/\lambda = -3/20$).

We numerically collide the monopole and the wall by giving the wall an initial velocity of 0.8 (in speed of light units) and choosing parameters $\eta = 1$, $h = -\lambda/5$, and $\lambda = 0.5$ for $V(\Phi)$.

Fig. 4 shows the potential and magnetic energy densities as the wall hits the monopole in Case 1. In addition, we plot the scalar field configuration in Fig. 5, where each point is a vector with components ϕ^3 and ϕ^1. These figures show that the magnetic energy density and the scalar field configuration remain unchanged after the collision, and that the potential energy densities corresponding to the monopole and the domain wall remain localized. This does not imply a complete absence of interaction between the wall and the monopole – some interaction is expected due the non-linearity of the scalar field potential.

To see if the monopole gains momentum due to the interaction, we have evaluated the centre of energy (COE) defined as

$$z_{\rm COE}(t) = \frac{\int_V d^3x\ z\ \rho(t,\mathbf{x})}{\int_V d^3x\ \rho(t,\mathbf{x})}\ , \tag{22}$$

where V is the volume of a finite cylindrical region centred at the origin and extending 1/8th of the lattice size in the x- and z-directions, while ρ is the energy density. For $h/\lambda = 0$, we give the wall a velocity of $v = 0.9$ towards the monopole and compare the initial $z_{\rm CEO}$ to the one after the wall passes away. We see a very slow drift of the COE in the direction of the wall velocity. We performed the same procedure using different model parameters and wall velocities and the outcome was qualitatively the same. In all cases, while the direction of the drift is clear, the magnitude is extremely small and too

close to the numerical uncertainties to allow a definitive quantitative analysis.

B. Case 2: the monopole unwinds

As in Case 1, it is possible to guess the outcome of the monopole-wall collision without doing numerical simulations. For this, we note that $\Phi_+^{(2)}$ has an SU(2) symmetry in the 4-5 block, which means that there is no topology that can support the winding. Thus, the monopole cannot exist in that corner of the matrix. An equivalent way to see this is to note that the function $c(z)$, which multiplies the three relevant monopole scalar fields, goes to zero at $z = +\infty$ (see Fig. 2), effectively erasing the monopole.

Additional insight can be gained by noting that the long range magnetic field of the monopole transforms into an SU(3) magnetic field on the far side of the wall. More explicitly, the U(1) magnetic field is given by Eq. (12) with X_{ij}^a determined using the solution in Eq. (10). Since X_{ij} only has components in the τ^a directions, it lies in the 4-5 block. However, the 4-5 block is entirely within the unbroken SU(3) on the right-hand side of the wall. Thus the long range magnetic field of the monopole is purely SU(3) on the right-hand side of the wall and, from the vantage point of someone there, there is no U(1) magnetic field emerging from the left-hand side of the wall. However, a U(1) magnetic field is an essential feature of a topological monopole. Thus, from the right-hand side of the wall, there is no magnetic monopole in the system, only some source of SU(3) magnetic flux.

Doing the numerical simulation with the parameters chosen as before, we plot the potential and magnetic energy densities as the wall hits the monopole in Fig. 6. This figure shows that the potential energy for the monopole disappears as the wall and monopole collide, and the magnetic energy that was stored in the monopole radiates away in a hemispherical wave. The collision was simulated with initial wall velocities ranging from 0.1 to 0.99 for $h/\lambda = -1/5$, and initial wall velocities of 0.6, 0.8 and 0.99 for $h/\lambda = -3/20$ and $1/5$. In all of these cases, the result of the collision was unchanged.

In Fig. 7, we show the $a = 1, 2, 3$ components of the scalar field using two different representations. In the first row, the fields ϕ^3 and ϕ^1 are plotted as a vector. The plot shows that the components of the field that are responsible for the winding vanish on the $z = +\infty$ side of the wall. In the second row of Fig. 7, the color represents the magnitude $|\phi| \equiv \sqrt{\phi^a \phi^a}$, $a = 1, 2, 3$, while vectors are drawn of fixed length and direction given by $\tan^{-1}(\phi^3/\phi^1)$. Even though $|\phi|$ becomes very small, it is not strictly zero at a finite distance from the wall, and so one can still define the direction of the arrow in this way. One can see that initially the field has a hedgehog configuration across the wall. However, as the wall sweeps along, the fields on the $z = +\infty$ side of the wall rotate around in such a way as to unwind the monopole. In the final step, all fields that are non–zero are pointing in one direction, and therefore the monopole winding is gone.

IV. CONCLUSIONS

In a Grand Unified model there can be several types of defects, including magnetic monopoles and domain walls. In the aftermath of the cosmological phase transition in which the Grand Unified symmetry is spontaneously broken to the standard model symmetry, the monopoles and walls will interact[3]. We have studied these interactions explicitly in an $SU(5) \times Z_2$ GUT, taking into account that the model has several different types of domain walls, and that only the lowest energy wall is expected to be cosmologically relevant. Even this stable wall has several different orientations in internal space, two of which are distinct for the purposes of monopole-wall interaction.

The first wall (Case 1 above) is found to be transparent to the monopole. This is simply because the domain wall mainly resides in a certain block of field space, while the winding of the monopole resides in a different non-overlapping block. The interactions between the monopole and the wall are very weak, and only affect the dynamics of the monopole as it passes through the wall. Depending on the parameters, the monopole might be attracted or repelled by the wall leading to a time delay or advance as the monopole goes through.

The second wall (Case 2 above) is opaque to the monopole. When the monopole hits the wall its energy is transformed into radiation on the other side of the wall, as seen in Fig. 6. A useful way to picture this system is to consider a magnetic monopole that is located inside a spherical domain wall. Now there is a topological magnetic monopole inside the wall, but only an SU(3) magnetic flux from the outside. In particular, there is no topological magnetic monopole as seen from the outside. Therefore the spherical wall itself must carry the topological charge of an antimonopole[4]. If the spherical wall shrinks, either it can annihilate the magnetic monopole within it and radiate away the energy, or the monopole can escape the wall, in which case the wall would then collapse into an antimonopole so that the total topological charge of the system continues to vanish. Our explicit numerical evolution shows that annihilation occurs for the parameter ranges we have considered. We note that the unwinding of the monopole in the Case 2 may be related to the mechanism of formation of non-Abelian clouds (massless monopoles) [15].

Our results have bearing on cosmology as they explicitly show the possible destruction of magnetic monopoles. In the case where the Z_2 symmetry is approximate, the walls will eventually decay away, and it is possible that these interactions could lead to a universe that is free of

[3] Scattering of fermions and GUT domain walls was studied in [14]

[4] The correspondence between spherical domain walls and global monopoles in SU(N) has previously been noted in [5].

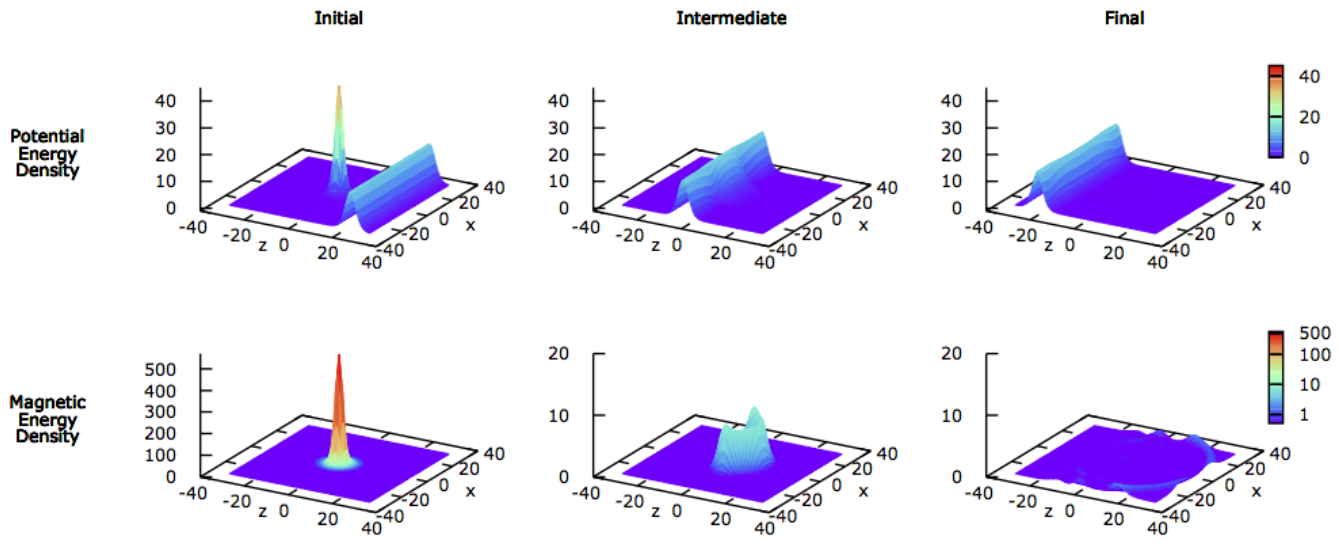

FIG. 6. The potential and magnetic energy densities in the xz plane for the colliding monopole and wall in Case 2. We can see that as the domain wall and monopole collide, the potential energy contained by the monopole disappears and the monopole begins to radiate away its magnetic energy in a hemispherical wave. Note that the middle and final plots for the magnetic energy density have a much smaller scale as the ripples are not visible at the original scale.

magnetic monopoles. Estimates in [1] indicate that this possibility is worth investigating in more detail. With several types of domain walls and monopoles simultaneously forming in a phase transition [16–18], and with the complex nature of both the inter-wall [19] and monopole-wall interaction, the fate of the monopoles will remain uncertain until a comprehensive simulation of the GUT phase transition is performed. We leave this for a future study.

ACKNOWLEDGMENTS

LP and MB are supported by the National Sciences and Engineering Research Council of Canada. TV gratefully acknowledges the Clark Way Harrison Professorship at Washington University during the course of this work, and was supported by the DOE at ASU.

[1] G. R. Dvali, H. Liu and T. Vachaspati, Phys. Rev. Lett. **80**, 2281 (1998) [hep-ph/9710301].
[2] L. Pogosian and T. Vachaspati, Phys. Rev. D **62**, 105005 (2000) [hep-ph/9909543].
[3] L. Pogosian and T. Vachaspati, Phys. Rev. D **62**, 123506 (2000) [hep-ph/0007045].
[4] T. Vachaspati, Phys. Rev. D **63**, 105010 (2001) [hep-th/0102047].
[5] L. Pogosian and T. Vachaspati, Phys. Rev. D **64**, 105023 (2001) [hep-th/0105128].
[6] T. Vachaspati, Phys. Rev. D **67**, 125002 (2003) [hep-th/0303137].
[7] H. Ruegg, Phys. Rev. D **22**, 2040 (1980).
[8] D. Wilkinson and A. S. Goldhaber, Phys. Rev. D **16**, 1221 (1977).
[9] C. P. Dokos and T. N. Tomaras, Phys. Rev. D **21**, 2940 (1980).
[10] G. 't Hooft, Nucl. Phys. B **79**, 276 (1974).
[11] H. Liu and T. Vachaspati, Phys. Rev. D **56**, 1300 (1997) [hep-th/9604138].
[12] A. Davidson, B. F. Toner, R. R. Volkas and K. C. Wali, Phys. Rev. D **65**, 125013 (2002) [hep-th/0202042].
[13] M. Alcubierre, S. Brandt, B. Bruegmann, D. Holz, E. Seidel, R. Takahashi and J. Thornburg, Int. J. Mod. Phys. D **10**, 273 (2001) [gr-qc/9908012].
[14] D. A. Steer and T. Vachaspati, Phys. Rev. D **73**, 105021 (2006) [hep-th/0602130].
[15] K. M. Lee, E. J. Weinberg and P. Yi, Phys. Rev. D **54**, 6351 (1996) [hep-th/9605229].
[16] L. Pogosian and T. Vachaspati, Phys. Rev. D **67**, 065012 (2003) [hep-th/0210232].
[17] N. D. Antunes, L. Pogosian and T. Vachaspati, Phys. Rev. D **69**, 043513 (2004) [hep-ph/0307349].
[18] N. D. Antunes and T. Vachaspati, Phys. Rev. D **70**, 063516 (2004) [hep-ph/0404227].
[19] L. Pogosian, Phys. Rev. D **65**, 065023 (2002) [hep-th/0111206].

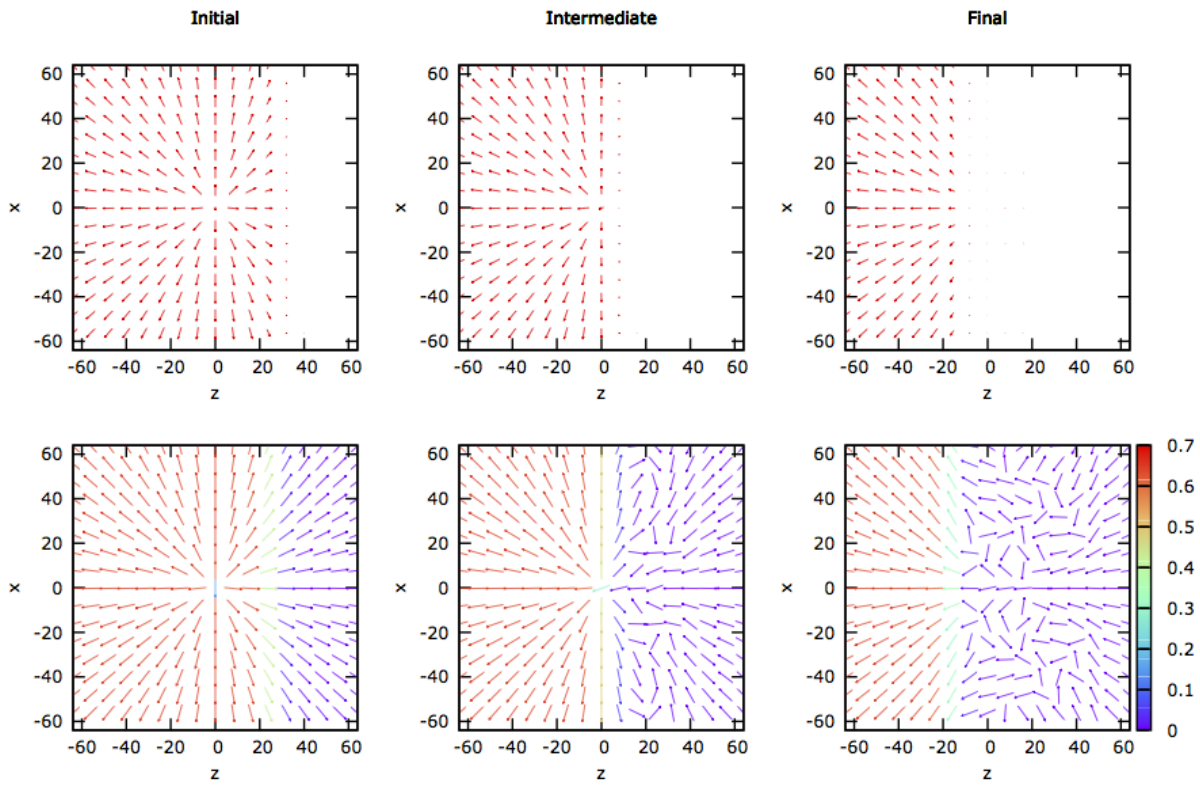

FIG. 7. The scalar field ϕ^a in the xz plane for the colliding monopole and wall in Case 2. At each point in the first row, ϕ^3 and ϕ^1 are plotted as a vector. In the second row, the length of the arrow is fixed, while the direction of the arrow is given by $\tan^{-1}(\phi^3/\phi^1)$ and the color represents the magnitude of the field $|\phi| = \sqrt{\phi^a \phi^a}$ for $a = 1, 2, 3$. The first row shows the monopole unwinding as the wall sweeps past it, and the second shows how the fields arrange themselves to unwind the monopole.

Monopoles on strings

T W B Kibble[1], **Tanmay Vachaspati**[2]

[1] Blackett Laboratory, Imperial College, London SW7 2AZ, UK

[2] Department of Physics, Arizona State University, Tempe, AZ 85287

E-mail: kibble@ic.ac.uk,tvachasp@asu.edu

Abstract. In cosmological scenarios based on grand unification, string theory or braneworlds, many kinds of topological or non-topological defects, including monopoles and cosmic strings, are predicted to be formed in the early universe. Here we review specifically the physics of composite objects involving monopoles tied to strings. There is a wide variety of these, including for example "dumbbells" and "necklaces", depending on how many strings attach to each monopole and on the extent to which the various fluxes are confined to the strings. We also briefly survey the prospects for observing such structures, the existing observational limits, and potential evidence for a cosmological role.

PACS numbers: 11.27.+d,14.80.Hv,98.80.Cq,

Submitted to: *J. Phys. G: Nucl. Part. Phys.*

1. Introduction

Symmetry-breaking phase transitions at which topological defects form are ubiquitous in low-temperature condensed-matter systems. In fundamental particle physics too, there are good reasons for thinking that similar phenomena occur at vastly higher energy scales. If so, the defects formed may have significant effects in the early universe.

Electroweak unification is now well-established. It is natural to suppose that further unification may occur at even higher energies, in a Grand Unified Theory (GUT). Some evidence in support of this idea comes from the running of the three independent coupling constants in the $SU(3) \times SU(2) \times U(1)$ standard model. In the basic standard model, they do not quite seem to meet. But in its supersymmetric extension, the parameters can be chosen so that they do all come together at an energy of around 10^{15} GeV [1, 2], suggesting the possibility of a high-temperature phase with a larger symmetry represented by a simple group such as $SO(10)$. In such a model, the universe would be expected to go through one or more phase transitions with decreasing symmetry as it cooled after the Big Bang. Even without supersymmetry, models with multiple phase transitions may be viable [3].

Depending on the pattern of symmetry breaking, such transitions could create topological defects of various types. These include point defects (monopoles), linear defects (cosmic strings, analogous to vortices in condensed matter) and domain walls. It has long been known that because the standard model group contains a $U(1)$ factor, monopoles are generic in GUT models that start with a simple gauge group. Avoiding the resulting over-abundance of monopoles was one of the prime original motivations for the introduction of the theory of inflation; if the monopoles are generated *before* the inflationary era, they will be diluted to insignificance by the rapid expansion. Nevertheless, inflation is compatible with the existence of defects, which can be formed during the reheating phase that terminates that era [4]. Moreover, essentially all realistic GUTs predict the existence of cosmic strings, though not always stable ones [5].

Cosmological scenarios derived from fundamental string theory or M-theory, such as braneworld models, also frequently predict the appearance of defects of similar types [6, 7, 4, 8]. These strings can have somewhat different properties. In particular, the probability of exchanging partners when strings intersect, can be much less than one [9], in contrast to the situation for cosmic strings in gauge theories [10]. Moreover, there can be strings of different tension, fundamental strings (F-strings) as well as Dirichlet D1 branes (D-strings) and (p, q)-strings, composites of p F-strings and q D-strings [11, 12]. There may be junctions where three strings meet. The evolution of a network of such strings is a more complicated problem, but the final result may not be so very different [13]. Analytic and numerical studies have shown that a network of ordinary cosmic strings generally evolves, at least on large scales, to a scaling regime in which the strings form a roughly constant fraction of the energy density of the Universe. Though the analysis is less clear cut, this appears also to be true for a multi-tension network. In that case the lightest strings come to dominate [14, 15]. Here we shall not discuss these

added complications in any detail. Another important topic that lies outside the scope of this review is the relevance of monopole and string networks to QCD confinement. See for example [16].

There has been extensive discussion in the literature of the characteristics and effects of these various defects, and of ways in which they might be detected. But in addition to the simple defects there may also be composite defects such as domain walls bounded by strings and strings connected to monopoles, and there has been less discussion of the effects of these more exotic structures. In this paper, we shall concentrate on the composites of strings and monopoles. As we aim to make clear, even within this category there are many different types of structures.

In Sec. 2, we briefly review the topological requirements for the different types of defects to form at a phase transition, where the symmetry is broken from a group G to a subgroup H. These are governed by the topology, in particular the homotopy groups, of the manifold $\mathcal{M}$ of degenerate vacuum states, which may be identified with the quotient space G/H. Then in Sec. 3, we discuss the case where the system undergoes two successive phase transitions, with the symmetry first broken from G to H and then to a smaller subgroup K. Such scenarios often lead to the formation of composite defects. In the remainder of the section, we discuss a number of different models that illustrate the wide range of possible defect structures. The interactions between the various defects that can form can be quite complex [17, 18]. Here we concentrate only on a few examples.

Sec. 4 is devoted to a discussion of the strings and monopoles that appear in the standard electroweak model. Electroweak monopoles and strings are not strictly speaking topological defects. Electroweak monopoles are confined while electroweak strings are known to be unstable, but configurations of these electroweak defects still can play an important role in cosmology, especially perhaps in connection with baryon number violation and cosmological magnetic field generation.

Possible means of observing composite defects of various kinds are discussed in Sec. 5, where we also discuss the observational constraints arising from existing observations. The conclusions are briefly summarized in Sec. 6.

2. Simple defects

2.1. Topological conditions for defects

We first recall the conditions for the appearance of topological defects of various types at a symmetry-breaking phase transition (see for example [19]). When the system is cooled through the transition temperature, there is some order parameter field multiplet ϕ that acquires a vacuum expectation value, say $\langle 0|\phi|0\rangle = \phi_0$, lying somewhere on a manifold $\mathcal{M}$ of minima of the potential $\mathcal{V}(\phi)$. If the symmetry group is G, then any operation $g \in G$ will transform this vacuum state into another degenerate vacuum state, with expectation value $g\phi_0$. If $H = \{h \in G | h\phi_0 = \phi_0\} \subset G$ is the subgroup leaving ϕ_0

invariant, then the vacuum manifold may be identified with the quotient $\mathcal{M} = G/H$, the set of left cosets $\{gH\}$ of H in G. The types of defects that may be formed are governed by the topology of $\mathcal{M}$.

Cosmic strings can form if $\mathcal{M}$ is not simply connected, *i.e.* its fundamental group or first homotopy group $\pi_1(\mathcal{M}) \neq 1$, where 1 stands for the group comprising the identity alone. That means there are closed loops in $\mathcal{M}$ that cannot be continuously shrunk to a point. The value of ϕ at points on a large loop in space surrounding a cosmic string will follow such a path. In the simplest case, where ϕ is a complex scalar and G comprises the phase rotations $\phi \to \phi e^{i\alpha}$, then $\mathcal{M}$ is a circle, $|\phi| = \eta$, *i.e.* $\phi = \eta e^{i\alpha}$ with arbitrary phase α. In this case, $\pi_1(\mathcal{M}) = \mathbb{Z}$. Thus the strings are labelled by an integer *winding number* n; on a loop around a string of winding number n, the phase α changes by $2n\pi$ (see section 2.2).

There is a simple general criterion for the existence of strings in a model where the symmetry group G is connected and simply connected, *i.e.* $\pi_0(G) = \pi_1(G) = 1$. Then a standard theorem tells us that $\pi_1(G/H) = \pi_0(H)$. Here the zeroth homotopy group counts the number of disconnected pieces of H; $\pi_0(H) = H/H_0$, where H_0 is the connected component of H containing the identity. So strings exist if and only if H is disconnected. The theorem may still be applied even if G is not simply connected, merely by replacing it by its simply connected universal covering group. For example, we may replace $U(1)$ by the additive group of real numbers $G = \mathbb{R}$, in which case $H = \mathbb{Z}$, the set of transformations with $\alpha = 2n\pi$.

Similarly, *monopoles* exist if $\pi_2(\mathcal{M}) \neq 1$, that is, if there are non-shrinkable two-dimensional surfaces in $\mathcal{M}$. Surrounding a monopole, the value of ϕ will lie on such a surface. The simplest example here is when $G = SU(2)$, with ϕ in the three-dimensional adjoint representation. Then $H = U(1)$, and $\mathcal{M}$ is a 2-sphere, $|\phi| = \eta$, so $\pi_2(\mathcal{M}) = \mathbb{Z}$. The monopoles are again labelled by an integer (see section 2.3). There is also a similar theorem, applicable when $\pi_0(G) = \pi_1(G) = 1$, namely $\pi_2(G/H) = \pi_1(H)$. So monopoles exist when H is not simply connected.

For completeness, we mention two other topological objects. *Domain walls* occur when $\mathcal{M}$ itself is disconnected, $\pi_0(\mathcal{M}) \neq 1$. For example, we may take a real scalar field ϕ with a double-well potential and $\mathbb{Z}_2$ symmetry under $\phi \to -\phi$. The domain wall separates regions where the vacuum expectation value lies in one well or the other. Finally *textures* occur if $\pi_3(\mathcal{M}) \neq 1$. Here there is no actual defect, in the sense of a compact region of concentrated energy. However, a non-trivial texture cannot be smoothly eliminated and converted to the vacuum state; it represents excess energy, albeit spread out rather than concentrated. Unwinding of the texture occurs in a restricted region of spacetime. Textures in the universe could have real physical consequences [20].

In the remainder of this section, we discuss specific models that illustrate a variety of different types of strings and monopoles.

2.2. Simple string models

a. The simplest model that leads to cosmic strings is the $U(1)$-symmetric *Abelian Higgs model*, comprising a complex scalar field ϕ interacting with a gauge field A_μ, described by the Lagrangian

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

where

$$D_\mu\phi = \partial_\mu\phi + ieA_\mu\phi, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{2}$$

and

$$\mathcal{V} = \tfrac{1}{4}\lambda(\phi^*\phi - \eta^2)^2. \tag{3}$$

Here λ and η are real positive constants and we set $c = \hbar = 1$. The potential $\mathcal{V}$ has a maximum at $\phi = 0$, so the $U(1)$ symmetry is broken in the vacuum. There is a degenerate family of vacua labelled by the phase angle: $\langle 0|\phi|0\rangle = \eta e^{i\alpha}$. This is essentially scalar electrodynamics but with a symmetry-breaking potential. The masses of the scalar and vector particles in the theory are $m_{\rm s} = \sqrt{\lambda}\eta$, $m_{\rm v} = \sqrt{2}|e|\eta$.

A static string with winding number n along the z axis is described in cylindrical polars (ρ, φ, z) by a field configuration

$$\phi = \eta f(\rho)e^{ni\varphi}, \qquad A_0 = 0, \qquad A_k = -\frac{n}{e\rho}h(\rho)\partial_k\varphi, \tag{4}$$

where the dimensionless functions f and h satisfy the boundary conditions

$$f(0) = h(0) = 0, \qquad f(\infty) = h(\infty) = 1. \tag{5}$$

The magnetic field along the string carries a total magnetic flux $2n\pi/e$. This is the *Nielsen–Olesen string* solution [21].

The solution with $n = 1$ is always stable, but the stability of strings with $n > 1$ depends on the value of the ratio $\beta = m_{\rm s}^2/m_{\rm v}^2 = \lambda/2e^2$. For Type-II strings, with $\beta > 1$, close parallel strings repel, and any string with $n > 1$ is unstable to break-up into $n = 1$ strings. Type-I strings, with $\beta < 1$, are stable for all values of n, and can form three-string junctions where for example strings with winding numbers m and n meet to form an $(m + n)$ string. For the critical case of $\beta = 1$, there is no force between parallel strings.

For these strings, the tension is equal to the energy per unit length, μ, and is given by $\mu = 2\pi g(\beta)\eta^2$, where g is a slowly varying, monotonically increasing function with the value $g(1) = 1$ for the critical coupling.

b. As a second example, we consider the symmetry group $G = SU(2)$ with *two* scalar fields in the adjoint representation, $\phi = \phi^a\sigma^a$ and $\psi = \psi^a\sigma^a$, where the σ^a are Pauli matrices. If we take

$$\mathcal{V} = \tfrac{1}{4}\lambda(\vec{\phi}^2 - \eta^2)^2 + \tfrac{1}{4}\lambda(\vec{\psi}^2 - \eta^2)^2 + \tfrac{1}{4}\mu(\vec{\phi}\cdot\vec{\psi})^2, \tag{6}$$

where $\vec{\phi}^2 = \phi^a\phi^a = \frac{1}{2}\,\mathrm{tr}(\phi^2)$, then it is clear that in the vacuum we will have $|\vec{\phi}| = |\vec{\psi}| = \eta$ and $\vec{\phi}\cdot\vec{\psi} = 0$. This breaks the symmetry down to the centre of $SU(2)$, namely

$H = \{1, -1\} \cong \mathbb{Z}_2$. Since H is discrete, there is no remaining massless gauge field. But there are strings, because $\pi_1(\mathcal{M}) = \pi_0(H) = \mathbb{Z}_2$. For these "$\mathbb{Z}_2$ strings", stable strings with higher winding numbers may exist but are not stable for topological reasons. We shall return to this point below.

In fact, it is easy to construct similar models of "$\mathbb{Z}_n$ strings", by taking the symmetry group to be $SU(n)$ with n fields in the adjoint representation, and choosing a potential that constrains them all to be non-zero and mutually orthogonal (see Section 3.2d).

c. Interesting possibilities occur when some of the symmetries are local and others global [22]. For example, suppose that $G = U(2) \cong SU(2) \times U(1)/\mathbb{Z}_2$, with a scalar field in the fundamental (spinor) representation. (The $\mathbb{Z}_2$ factor is required because the centre of $SU(2)$, comprising the two elements $\{1, -1\}$ is also contained in $U(1)$.) Moreover, suppose that only the Abelian factor $U(1)$ is gauged, so there is just one gauge boson, while $SU(2)$ is a *global* symmetry group.

If we looked only at the local symmetry, we might expect the appearance of strings because of the breaking of $U(1)$. However the vacuum manifold $\psi^\dagger\psi = \eta^2$ is $\mathcal{M} = S^3$, and $\pi_1(\mathcal{M}) = 1$, so there are no *topologically stable* strings. It is easy to construct a string solution by embedding the Nielsen–Olesen string solution (4); we take

$$\phi = \eta f(\rho) e^{ni\varphi} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad A_0 = 0, \qquad A_k = -\frac{n}{e\rho} h(\rho) \partial_k \varphi. \tag{7}$$

Stability of this *semi-local* solution is not guaranteed by any topological argument. Using the other component of ϕ, it can be smoothly deformed into a configuration lying entirely in $\mathcal{M}$, so that the potential energy vanishes, but at the cost of increasing the gradient energy. Despite the absence of a topological guarantee of stability, detailed analysis shows that it is indeed dynamically stable in the Type-I regime $\beta < 1$, though not when $\beta > 1$ [23, 24].

2.3. 't Hooft–Polyakov monopoles

In this model, $G = SU(2)$, and ϕ belongs to the three-dimensional adjoint representation. We can write $\phi = \phi^a \sigma^a$, $A_\mu = A_\mu^a \sigma^a$. Here we take

$$D_\mu \phi = \partial_\mu \phi + \tfrac{1}{2} ie [A_\mu, \phi], \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + \tfrac{1}{2} ie [A_\mu, A_\nu], \tag{8}$$

or equivalently

$$D_\mu \phi^a = \partial_\mu \phi^a - e \epsilon^{abc} A_\mu^b \phi^c, \qquad F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - e \epsilon^{abc} A_\mu^b A_\nu^c. \tag{9}$$

With $\mathcal{V} = \frac{1}{4}(\vec{\phi}^2 - \eta^2)^2$, we find that the vacuum manifold $\mathcal{M}$ is a two-sphere, $\vec{\phi}^2 = \eta^2$, and $H = U(1)$. In this case there is a scalar particle of mass $m_\mathrm{s} = \sqrt{\lambda}\eta$ and three vector particles, one massless (identified with the photon) and two with masses $m_\mathrm{v} = \sqrt{2}|e|\eta$. This is essentially the Weinberg-Salam model with vanishing weak mixing angle, $\theta_\mathrm{w} = 0$.

Here a static monopole at the origin is described by the solution [25, 26]

$$\phi^a = \eta f(r)\frac{x^a}{r}, \qquad A_0^a = 0, \qquad A_k^a = -h(r)\frac{\epsilon^{akj}x^j}{er^2}, \tag{10}$$

where $r = \sqrt{x^k x^k}$ and the functions f and h obey the boundary conditions

$$f(0) = h(0) = 0, \qquad f(\infty) = h(\infty) = 1. \tag{11}$$

At large values of r, the gauge field is found to be

$$F_{0k}^a = 0, \qquad F_{ij}^a = \frac{x^a}{r}\frac{\epsilon_{ijk}x^k}{er^3}. \tag{12}$$

This field is in the direction of the unbroken symmetry generator, corresponding to the electromagnetic field. It represents a radial magnetic field

$$B^k \equiv -\frac{1}{2}\epsilon^{kij}\hat{\phi}^a F_{ij}^a = -\frac{x^k}{er^3}. \tag{13}$$

Hence the total outward magnetic flux, the magnetic charge of the monopole, is

$$q = -\frac{4\pi}{e}. \tag{14}$$

In any model, for the field A_k^a to be single-valued, the magnetic charge for any monopole must always satisfy the condition

$$eq = 2n\pi \tag{15}$$

for some integer n. Note that for this particular monopole solution, the charge is twice the minimal value.

It can be shown that the mass of the monopole obeys the Bogomol'nyi bound [27],

$$m_{\rm mon} \geq \frac{4\pi\eta}{e}, \tag{16}$$

which is saturated in the Prasad-Sommerfeld limit of small scalar coupling, $\lambda/e^2 \to 0$ [28].

3. Composite defects

3.1. Defects formed at multiple phase transitions

There are many field-theory models that predict more than one phase transition in the early universe. In such cases, composite defects may form [29].

Suppose we start with a theory with symmetry group G, and that it goes through a phase transition where a field ϕ acquires a non-zero expectation value, breaking the symmetry to a subgroup $H \subset G$, and then subsequently a second phase transition, where another field ψ gets a non-zero, but generally smaller, vacuum expectation value,

breaking the symmetry further to $K \subset H$. After the first breaking, we have a vacuum manifold $\mathcal{M} = G/H$. If this is topologically non-trivial, defects will form. In the second transition another set of defects may form if the manifold H/K has non-trivial homotopy groups. However, the existence of stable defects in the final phase is actually controlled by the topology of $\mathcal{M}' = G/K$.

The simplest example here is a $U(1)$ gauge model with two scalar fields, ϕ of charge $2e$, and ψ of charge e. We assume that the potential contains an interaction term of the form $-m(\phi^*\psi^2 + \psi^{*2}\phi)$. The absolute minimum of the potential occurs when ϕ and ψ have fixed magnitudes, say $|\phi| = \eta, |\psi| = \zeta$, and the phase of ψ^2 is the same as that of ϕ. After the first stage of symmetry breaking, when $\langle\phi\rangle$ becomes non-zero, the symmetry is reduced from $U(1)$ to $H = \mathbb{Z}_2$, comprising the transformation $\psi \to -\psi$. Here $\mathcal{M}$ is a circle S^1, and $\pi_1(\mathcal{M}) = \mathbb{Z}$. Therefore strings are formed, labelled by an integer winding number n, with the phase of ϕ changing by $2n\pi$ around the string.

Now when the second transition occurs, the remaining $\mathbb{Z}_2$ symmetry is broken, because ψ has to choose between the two degenerate vacuum values. Breaking this discrete symmetry would be expected to create domain walls, separate regions where opposite choices are made. But considering the overall symmetry breaking, from $U(1)$ to 1, no discrete symmetry breaking is involved, and there are no truly stable domain walls. In fact, $\mathcal{M}'$ is also a circle, but each point of $\mathcal{M}$ corresponds to two diametrically opposite points of $\mathcal{M}'$.

Consider a string along the z axis, where outside the core, $\phi = \eta e^{in\varphi}$. To minimize the potential we must then have $\psi = \pm\zeta e^{in\varphi/2}$. But note that for $n = 1$ or any odd number, that would not lead to a continuous solution. The strings with even winding number survive, but around one with odd winding number there must be a point where ψ changes sign over a short distance. In other words, the string becomes attached as the boundary of a domain wall.

Unlike fully stable domain walls, these are potentially unstable to the formation of holes surrounded by new loops of string, though such a decay has to overcome an energy barrier. The hole has to attain a minimum size before its creation becomes energetically favorable.

3.2. Monopoles joined by strings

We now discuss several examples where the first stage of the symmetry breaking leads to the formation of 't Hooft–Polyakov monopoles, followed by a second stage where strings form.

a. One simple example is provided by the $SU(2)$ model above with two adjoint fields ϕ and ψ, but where the constants λ and η in the first two terms of (6) are different, say λ, λ' and η, η', with $\sqrt{\lambda'}\eta' \ll \sqrt{\lambda}\eta$. Then in the first stage of symmetry breaking, when ϕ becomes non-zero, the symmetry will break to $H = U(1)$, while after the second stage it will break further to $K = \mathbb{Z}_2$. At the first stage, monopoles will form, because $\pi_2(G/H) = \mathbb{Z}$. In the second breaking, since $\pi_1(H/K) = \mathbb{Z}$, we expect strings, classified

as usual by an integer winding number. Overall, however, since $\pi_1(G/K) = \mathbb{Z}_2$, the only topologically stable strings are $\mathbb{Z}_2$ strings.

Moreover, there are no truly stable monopoles. It is easy to see what happens. Around an $n = 1$ monopole the field $\vec{\phi}$ may be chosen to point radially outwards. When $\vec{\psi}$ becomes nonzero it needs to be orthogonal to $\vec{\phi}$, so around a sphere it should lie in a tangential direction. But it is not possible to choose such a direction everywhere. There have to be points where it vanishes. For example, we could take it everywhere in the azimuthal φ direction, but to maintain continuity it must then vanish at the north and south poles. In fact, there have to be two strings attached to the monopole. The monopoles are like beads on the string. The configuration is often called a *necklace*. Similar structures can appear very naturally in string-theory models [30].

It is useful to consider the fields around a string. If the first field $\vec{\phi}$ is taken to be along the string, then $\vec{\psi}$ must wind around it, either clockwise or anticlockwise. Thus the string has a direction; a string is not identical to an anti-string, in spite of the fact that they are topologically equivalent. A string can be converted to an antistring, but it takes energy to do so. In fact, what it takes is the creation of a pair of monopoles.

Similarly, strings with higher winding numbers (in the Type-I case $\beta < 1$) may exist, but are not truly stable; an $n = 2$ string can terminate on a monopole.

b. Now let us consider another model, this time with symmetry group $G = U(2)$, as in the example of the semi-local string, but here with all symmetries gauged. A scalar field ϕ in the adjoint representation breaks the symmetry, here to $H = U(1) \times U(1)$. The manifold of degenerate vacua is $\mathcal{M} = S^2$, and monopoles can form. Now suppose there is another scalar field ψ in the fundamental (spinor) representation, and that there are extra terms in the potential:

$$\mathcal{V} = \tfrac{1}{4}\lambda(\phi^a\phi^a - \eta^2)^2 + \tfrac{1}{2}\lambda'(\psi^\dagger\psi - \eta'^2)^2 + g\psi^\dagger\sigma^a\psi\phi^a, \tag{17}$$

where again $\sqrt{\lambda'}\eta' \ll \sqrt{\lambda}\eta$. As the system cools further, it will go through a second transition at which $\langle\psi\rangle$ becomes non-zero. If, for example, $\langle\phi^a\rangle = \eta\delta^a_3$, then clearly, to minimize the potential, $\langle\psi\rangle$ should be proportional to the eigenvector $\binom{0}{1}$ of σ^3. This breaks the symmetry down to $K = U(1)$, generated by $\frac{1}{2}(1 + \sigma^3) = \left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$.

This model is very different from the previous one, in that there remains a massless vector field in the final phase; indeed the gauge-field structure is the same as in the bosonic sector of the standard electroweak model. Since $\pi_1(H/K) = \mathbb{Z}$, strings labelled by an integer winding number will be formed in the second transition. The manifold of vacua becomes $\mathcal{M}' = G/K = S^3$. Thus there are no truly stable strings or monopoles in the final phase. What happens is that each monopole becomes attached to a string; each string is either a closed loop or connects a monopole to an antimonopole.

It is easy to see what a monopole configuration looks like. At large distance from the monopole, we can take $\vec{\phi}$ radially outwards, so that

$$\phi = \eta\frac{x^k}{r}\sigma^k = \eta\begin{pmatrix}\cos\theta & \sin\theta e^{-i\varphi}\\ \sin\theta e^{i\varphi} & -\cos\theta\end{pmatrix}. \tag{18}$$

All around the sphere, ψ must be proportional to the eigenvector with eigenvalue -1, so we can take

$$\psi = \eta' \begin{pmatrix} \sin\frac{\theta}{2} e^{-i\varphi} \\ -\cos\frac{\theta}{2} \end{pmatrix}. \tag{19}$$

But it is impossible to make this choice continuous everywhere. Here it is singular at the south pole, where a string must be attached, around which the phase of ψ changes by 2π.

Note that here the strings carry a magnetic flux $(2\pi/e)$, equal to the magnetic charge on the monopole, so there is only one string attached to each, not two.

c. Very different behaviour can be seen in a model based on the symmetry group $G = SU(3)$ with three fields ϕ, ψ_1, ψ_2, all in the 8-dimensional adjoint representation [31]. In the first stage of symmetry breaking ϕ acquires a non-zero expectation value, satisfying $|\phi| = \eta$, where $|\phi|^2 = \frac{1}{2}\,\mathrm{tr}(\phi^2)$. The vacuum manifold is then $\mathcal{M} = SU(3)/U(2)$, which may be identified with the complex projective space $\mathbb{C}P^2$. Points in this space may be labelled by triples of complex numbers $Z^T = (z_1, z_2, z_3)$, where (z_1, z_2, z_3) and $(\kappa z_1, \kappa z_2, \kappa z_3)$ represent the same point for any non-zero $\kappa \in \mathbb{C}$. The point in $\mathcal{M}$ corresponding to $Z \in \mathbb{C}P^2$ is

$$\phi = \frac{\eta}{\sqrt{3}} \left(1 - 3\frac{ZZ^\dagger}{Z^\dagger Z} \right). \tag{20}$$

For example, we may choose the value

$$\phi_0 = \eta T^8 \equiv \frac{\eta}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}, \qquad Z_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{21}$$

Here $T_1, \ldots, T_8$ are the generators of $SU(3)$, the Gell-Mann matrices [32].

After this first symmetry breaking the remaining symmetry group is $H = U(2) \cong SU(2) \times U(1)/\mathbb{Z}_2$. There are non-trivial loops in H, and $\pi_1(H) = \mathbb{Z}$, so there are monopoles, labelled by an integer n. But this is somewhat misleading. Homotopically non-trivial loops in H corresponding to odd values of n cannot lie solely in the $U(1)$ factor; they must include a path in $SU(2)$ from the identity element $\mathbf{1}$ to $(-\mathbf{1})_2 = \mathrm{diag}(-1, -1, 1)$. A monopole with $n = 1$ must in a sense carry a "$\mathbb{Z}_2$ charge" as well as the monopole charge $2\pi/e$. Note however that the $\mathbb{Z}_2$ charge can only have the values 0 or 1; it obeys the $\mathbb{Z}_2$ addition rule, $1 + 1 \equiv 0$. Repeated twice this path in $SU(2)$ is trivial, so for even n the paths can be confined to $U(1)$. For even n, the monopoles do not carry a $\mathbb{Z}_2$ charge.

Next, we introduce two more adjoint fields, $\psi_{1,2}$, and choose the potential so that all three fields have definite magnitude and are orthogonal, in the sense that $\mathrm{tr}(\phi\psi_{1,2}) = \mathrm{tr}(\psi_1\psi_2) = 0$ and also so that at the minimum ψ_1 and ψ_2 commute with ϕ. For example, with the choice (21) for ϕ, we may take $\psi_{1,2} = \eta' T^{1,2}$. This then breaks the $SU(2)$ symmetry down to $\mathbb{Z}_2$, so the final symmetry group is merely $K = U(1)$.

A typical solution representing the field around a minimal-charge monopole, in spherical polars, is

$$\phi = \frac{\eta}{2\sqrt{3}} \begin{pmatrix} 3\cos\theta - 1 & 0 & -3\sin\theta\, e^{i\varphi} \\ 0 & 2 & 0 \\ -3\sin\theta\, e^{-i\varphi} & 0 & -3\cos\theta - 1 \end{pmatrix}, \qquad Z = \begin{pmatrix} \sin\frac{\theta}{2}\, e^{i\varphi} \\ 0 \\ \cos\frac{\theta}{2} \end{pmatrix}. \tag{22}$$

Suitable forms for the other two fields can be found by starting with $\psi_{1,2} = \eta' T^{1,2}$ at the north pole $\theta = 0$, and applying $SU(3)$ transformations $U(\theta, \varphi)$ that perform the transformation $\phi(\theta, \varphi) = U(\theta, \psi)\phi_0 U^\dagger(\theta, \varphi)$. A simple choice is

$$U(\theta, \varphi) = \begin{pmatrix} \cos\frac{\theta}{2} & 0 & -\sin\frac{\theta}{2}\, e^{-i\varphi} \\ 0 & 1 & 0 \\ \sin\frac{\theta}{2}\, e^{i\varphi} & 0 & \cos\frac{\theta}{2} \end{pmatrix}. \tag{23}$$

This means that around the south pole

$$U(\pi, \varphi) = \begin{pmatrix} 0 & 0 & -e^{-i\varphi} \\ 0 & 1 & 0 \\ e^{i\varphi} & 0 & 0 \end{pmatrix}. \tag{24}$$

Evidently, the configuration of the fields $\psi_{1,2}$ is singular at the south pole. This singularity cannot be removed by a gauge transformation (though it could of course be moved to a different location), because the path as φ ranges from 0 to 2π is non-contractible in $SU(3)$, whereas it would be contractible if φ ranged from 0 to 4π. A $\mathbb{Z}_2$ string must be attached at the south pole of the monopole configuration.

Every monopole of charge $n = 1$ must be attached to a string. The strings may terminate on monopoles or antimonopoles. A string may join a pair of equal-charge monopoles or a monopole-antimonopole pair. However, numerical simulations show that typically the second possibility is much more probable than the first. If the dynamics leads to the string shortening and disappearing, then in the first case this would lead to charge-2 monopoles, but in the second case to complete annihilation. The charge-2 monopoles have no $\mathbb{Z}_2$ charge, but are pure $U(1)$ monopoles.

d. A different choice of potential in the $SU(3)$ model can lead to an alternative symmetry breaking pattern, again with very different behaviour [33]. The first stage can proceed as before, with ϕ typically given by (21), breaking the symmetry down to $H = U(2)$ and again generating monopoles. But then we can choose the potential so that the minimum typically occurs when $\psi_{1,2} = \eta' T^{4,6}$, generators that do not commute with T^8 and so do not belong to H. This choice has the effect of breaking the symmetry down to $K = \mathbb{Z}_3$, the centre of $SU(3)$, comprising the matrices $\{e^{2\pi n i/3}\mathbf{1} | n = 0, 1, 2\}$. Consequently, this produces $\mathbb{Z}_3$ strings. Since K is purely discrete, no massless gauge fields remain.

Around a typical $\mathbb{Z}_3$ string, the fields $\psi_{1,2}$ at large distance behave as

$$\psi_1 = \eta'(T^4 \cos n\varphi + T^5 \sin n\varphi), \qquad \psi_2 = \eta'(T^6 \cos n\varphi + T^7 \sin n\varphi). \tag{25}$$

This configuration can be induced by applying the gauge rotation

$$U(\varphi) = e^{inT^8\varphi/\sqrt{3}} = \mathrm{diag}(e^{ni\varphi/3}, e^{ni\varphi/3}, e^{-2ni\varphi/3}). \tag{26}$$

What then happens around a pre-existing monopole? As in the previous example, we may expect that at the second symmetry breaking, the two new fields $\psi_{1,2}$ may be frustrated from finding the vacuum configuration everywhere around it. So we may expect strings to be attached. But there is an important difference this time. Here the gauge rotation around a string, Eq. (26), does not constitute a closed loop in $SU(3)$ unless $n \equiv 0 \mod 3$, since $U(2\pi) = e^{2\pi ni/3}\mathbf{1}$. Consequently, we cannot attach just one $n = 1$ string for example to the monopole. We need three of them, and if we are in the region of parameter space in which forces between identical strings are repulsive, the three will tend to spread out around the monopole. So this symmetry breaking pattern yields a quite different type of string network, with junctions where three strings meet at a monopole.

Another point should be noted here. An $n = 2$ string may or may not be unstable to splitting into two $n = 1$ strings. But in any case it is topologically equivalent to a $n = -1$ string, *i.e.* an $n = 1$ string in the opposite direction, so if it is stable to splitting, it is indeed in principle unstable to turning into an $n = -1$ string. But it may nevertheless be *locally* stable, because this transformation can only happen via the creation of a monpole-antimonopole pair, which requires energy.

4. Monopoles and strings in the standard electroweak model

We have already discussed semilocal strings in Sec. 2.2 (see Eq. (7)) in an $SU(2) \times U(1)/\mathbb{Z}_2$ model where the $SU(2)$ is global and the $U(1)$ is local. This model coincides with the standard model of the electroweak interactions whose symmetry group is denoted $[SU(2)_L \times U(1)_Y]/\mathbb{Z}_2$, but with the important difference that the $SU(2)_L$ factor is gauged. If the $SU(2)_L$ and $U(1)_Y$ coupling constants are denoted by g and g' respectively, the relative strength of the two coupling constants is given by the "weak mixing angle", θ_w, defined by

$$\tan\theta_w = \frac{g'}{g} \tag{27}$$

with the measured value $\sin^2\theta_w = 0.23$.

Since the semilocal model is the $\sin^2\theta_w \to 1$ limit of the standard model, we expect the semilocal string solution to also be present in the standard model. Thus the standard model has an electroweak string solution given by

$$\phi = \eta f(\rho) e^{ni\varphi} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad Z_0 = 0, \qquad Z_k = -\frac{n}{e\rho} h(\rho) \partial_k \varphi. \tag{28}$$

Note that only the Z-gauge field of the standard model is non-vanishing; the charged $W^\pm$ and the electromagnetic gauge fields vanish. Thus this solution is sometimes called

a "Z-string" and also distinguishes it from other embedded electroweak strings called "W-strings" in which the $W^{\pm}$ gauge fields are non-vanishing.

As in the semilocal case, there is no topological reason for the existence of the electroweak string solution; nor is its existence protected by a topological winding number. Hence we expect the Z-string to be unstable under small perturbations. A detailed stability analysis of the electroweak string shows that it is metastable if $m_H < m_Z$ and for $\sin^2\theta_w \gtrsim 0.95$, and is unstable for other parameters, including the physical values: $m_H = 125$ GeV, $m_Z = 91$ GeV, $\sin^2\theta_w = 0.23$.

Since the Z-string solution is not topological, a particular Z-string can terminate. To understand the properties of the terminus, we decompose the Z-magnetic flux inside the string into a linear combination of $SU(2)_L$ flux and $U(1)_Y$ flux. When the Higgs has the conventional vacuum expectation value: $\phi = \eta(0,1)^T$, the decomposition is

$$Z_\mu \equiv \cos\theta_w W^3_\mu - \sin\theta_w Y_\mu. \tag{29}$$

The W^3 magnetic flux is non-Abelian and can terminate, but the Y magnetic flux is Abelian and divergenceless, and cannot terminate. Then the Y magnetic flux must extend beyond the terminus of the Z-string and can only do so in the form of massless electromagnetic (A) magnetic flux defined by

$$A_\mu \equiv \sin\theta_w W^3_\mu + \cos\theta_w Y_\mu. \tag{30}$$

Thus the terminus of the Z-string is a source of A magnetic flux *i.e.* a magnetic monopole. Note that the Z and A gauge fields are orthogonal, so the magnetic monopole has $\mathrm{div}(\mathbf{B}_A) \neq 0$, where $\mathbf{B}_A$ is the electromagnetic magnetic field, while it is confined by a string that has nothing to do with electromagnetism. (The situation is very similar to the dual case where electrically charged quarks are confined by QCD color strings.)

Before describing the properties of the electroweak magnetic monopole and Z-string, we will provide another way of seeing the existence of the monopole, more in line with the original paper by Nambu [34]. Essentially one constructs a composite adjoint field

$$n^a(x) = -\frac{\phi^\dagger \tau^a \phi}{\phi^\dagger \phi}. \tag{31}$$

Once ϕ gets a vacuum expectation value, we will have $n^a \neq 0$. Note that n^a transforms trivially under $U(1)_Y$ and, as far as its properties under $SU(2)_L$ are concerned, it is exactly like the field ϕ^a in Sec. 2.3. Thus, as in the 't Hooft-Polyakov monopole, we can write down a "hedgehog" configuration

$$\eta n^a = \eta f(r)\frac{x^a}{r}, \qquad W^a_0 = 0, \qquad W^a_k = -h(r)\frac{\epsilon^{akj}x^j}{gr^2}. \tag{32}$$

However, this configuration is disallowed in the underlying model because the relation in Eq. (31) cannot be inverted to obtain a non-singular ϕ. Instead there has to be a string attached to the hedgehog on which $\phi = 0$. This is exactly the location of the Z-string.

More explicitly, the asymptotic Higgs and gauge field configurations for an electroweak monopole with a semi-infinite Z-string along the $-z$ axis are given by

$$\phi = \eta \begin{pmatrix} \cos\theta/2 \\ \sin\theta/2 e^{i\varphi} \end{pmatrix} \tag{33}$$

$$gW^a_\mu = -\epsilon^{abc} n^b \partial_\mu n^c + i\cos^2\theta_w n^a(\phi^\dagger\partial_\mu\phi - \partial_\mu\phi^\dagger\phi) \tag{34}$$

$$g'Y_\mu = -i\sin^2\theta_w(\phi^\dagger\partial_\mu\phi - \partial_\mu\phi^\dagger\phi). \tag{35}$$

A finite segment of Z-string will have an electroweak monopole on one end and an antimonopole on the other end. A field configuration for such a finite energy "dumbbell" configuration can be written as [34]

$$\phi = \begin{pmatrix} \cos(\Theta/2) \\ \sin(\Theta/2) e^{i\varphi} \end{pmatrix} \tag{36}$$

where

$$\cos\Theta \equiv \cos\theta_m - \cos\theta_{\bar{m}} + 1 \tag{37}$$

and θ_m and $\theta_{\bar{m}}$ are the spherical polar angles with the axes origin located at the monopole and the antimonopole respectively. Nambu also considered the lifetime of rotating dumbbells, though only accounting for decay by emission of electromagnetic radiation. In particular, decay by fragmentation and other instabilities were not considered and remain to be investigated.

The magnetic flux of the electroweak monopole can be shown to be

$$F = \frac{4\pi}{e}\sin^2\theta_w. \tag{38}$$

Seemingly this does not obey the Dirac quantization condition but this is not a contradiction because of the Z-string that is attached to the monopole.

The mass of the electroweak monopole cannot be defined because it is always confined. If the mass is measured in terms of the energy barrier to the breaking of Z-strings, it would turn out to be negative because the Z-string is unstable.

Finally we discuss the electroweak "sphaleron" [35] in terms of a bound state of an electroweak monopole and antimonopole. Since a monopole and an antimonopole carry opposite magnetic charges, there is an attractive Coulomb force that tends to bring them together so that they can annihilate. However, a monopole and an antimonopole have an extra degree of freedom, namely a relative phase between them. To see this in the context of the electroweak model [36], consider the asymptotic Higgs field configuration

$$\phi_{m\bar{m}} = \eta \begin{pmatrix} \sin(\theta_m/2)\sin(\theta_{\bar{m}}/2)e^{i\gamma} + \cos(\theta_m/2)\cos(\theta_{\bar{m}}/2) \\ \sin(\theta_m/2)\cos(\theta_{\bar{m}}/2)e^{i\varphi} - \cos(\theta_m/2)\sin(\theta_{\bar{m}}/2)e^{i(\varphi-\gamma)} \end{pmatrix} \tag{39}$$

where θ_m and $\theta_{\bar{m}}$ are spherical polar angles measured from the location of the monopole at $z = +a$ on the $z-$axis and the location of the antimonopole at $z = -a$ respectively. The phase angle γ will be explained in a moment. Note that $|\phi_{m\bar{m}}| = \eta$.

Away from the antimonopole and close to the monopole we can take $\theta_{\bar{m}} \approx 0$ and the configuration reduces to

$$\phi_{m\bar{m}} \to \eta \begin{pmatrix} \cos(\theta_m/2) \\ \sin(\theta_m/2)e^{i\varphi} \end{pmatrix} \tag{40}$$

which is the configuration around an electroweak monopole (see Eq. (33)). On the other hand, if we take $\theta_m \to \pi$,

$$\phi_{m\bar{m}} \to \eta e^{i\gamma} \begin{pmatrix} \sin(\theta_{\bar{m}}/2) \\ \cos(\theta_{\bar{m}}/2)e^{i(\varphi-\gamma)} \end{pmatrix}. \tag{41}$$

which is the Higgs configuration around an antimonopole up to an irrelevant overall phase factor. The $\varphi - \gamma$ phase shows that the antimonopole has a relative rotation compared to the monopole. Thus γ is a relative "twist" between the monopole and the antimonopole. Further, the twist provides a repulsive force between the monopole and the antimonopole. By adjusting the twist parameter and the monopole-antimonopole separation, a static solution can be found. The solution was first found in an $O(3)$ model in Ref. [37] and then in the electroweak model (in the $\theta_w = 0$ limit) in Ref. [35] using very elegant mathematical techniques. The solution, now called a "sphaleron", plays an important role in anomalous baryon number violation, and may play a critical role in explaining the cosmic matter-antimatter asymmetry, and may also provide a mechanism to generate cosmological magnetic fields (see below).

5. Observational constraints

As depicted in Fig. 1, there are three distinct cases relevant to monopoles connected by strings that need to be considered in a cosmological setting. The three cases correspond to whether a monopole is connected by 1 or 2 or many (≥ 3) strings. In addition, in all three cases, we can consider the possibility that *all* the monopole magnetic flux has been confined to the string, or only *some* of the flux is confined while the remaining flux is unconfined. For example, in Sec. 3.2 we have discussed the case of $SU(2) \to U(1) \to 1$ and there all the monopole flux gets confined to a string. On the other hand, for the electroweak monopole discussed in Sec. 4, the Z-flux is confined to a string, but the monopole still carries an unconfined electromagnetic flux.

First consider the case when a monopole is connected to an antimonopole by a single string and forms a "dumbbell". Simulations find that the length distribution of dumbbells is exponential: $\exp(-l/\xi)$ where ξ is set by the average distance between monopoles at the time of string formation[38].

If dumbbells are produced at some cosmological epoch, the strings will quickly shrink and bring the monopole and antimonopole together. The acceleration of the monopole and antimonopole will lead to electromagnetic radiation, whereby the system will lose energy, as given by the classical electromagnetic radiation formula $\dot{E} = g^2a^2/6\pi$

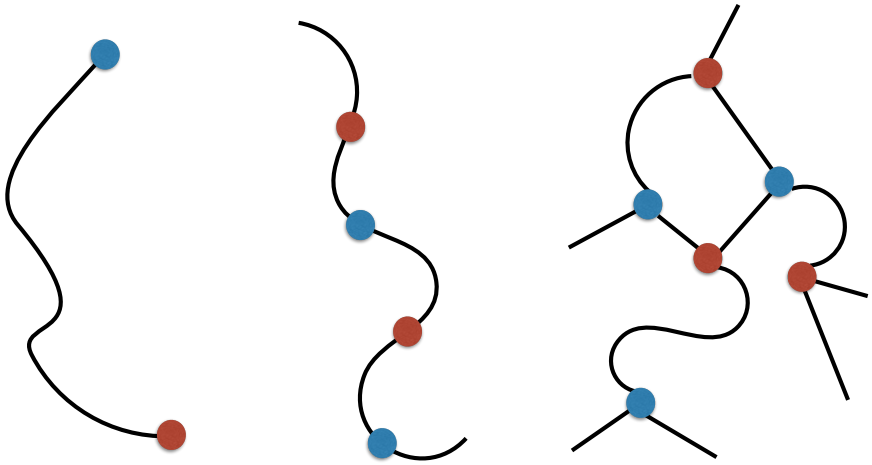

Figure 1. Three cases of monopoles on strings; blue and red dots represent monopoles and antimonopoles respectively.

where the acceleration $a = \mu/M$, M is the monopole mass, g the magnetic charge, and μ the string tension. Once the monopole and antimonopole collide, they annihilate and the time scale for a given dumbbell to dissipate will be given by its initial length. If a dumbbell is very long initially, there is also a probability that it will be chopped into shorter segments when other strings collide or when the long string crosses on itself, through a process called intercommutation. Since the length distribution of dumbbells is dominated by the smallest length, and the smallest length is typically much shorter than the cosmic horizon, most of the energy in dumbbells is dissipated within a Hubble time. Then observational signatures can only arise if a telltale remnant is produced during the decay process. We will shortly discuss three possible remnants.

In the second case shown in Fig. 1, the monopole is like a "bead on a string" and we expect the formation of "cosmic necklaces" [39, 30]. Monopoles and antimonopoles can slide along the string, collide, and annihilate, and produce high energy particles that can potentially be observed as cosmic rays [40, 41, 42]. However, closer scrutiny of the process [42] finds that the monopoles annihilate very rapidly after formation and the network soon resembles a network of ordinary cosmic strings. Then the observational constraints on ordinary cosmic strings (discussed below) also apply to beads-on-strings, independent of whether the monopole (beads) carry unconfined magnetic flux.

In the third case of Fig. 1, a string web is formed with monopoles at the junctions of the web. Then the web stretches with the expansion of the universe, and dilutes due to monopole-antimonopole annihilation. The resulting network scales self-similarly

in time, *i.e.* the statistical properties of the web do not depend on time but the characteristic overall scale (distance between monopoles) grows in proportion to cosmic time [43]: $d(t) \sim \chi t$ with $\chi \sim \mu/(24\pi M^2)$, where μ is the string tension and M is the monopole mass. We now need to distinguish between the case when the monopoles carry unconfined flux and the case when all the flux is confined, since these two scenarios lead to very different cosmological scenarios [43].

If the monopoles carry unconfined flux, their rapid acceleration under the pull of the strings leads to the emission of very high energy gamma rays whose spectrum peaks at ~ 100 TeV. The energy density in such gamma rays divided by the critical energy density of the universe is estimated to be [43]

$$\Omega_{\gamma\ \mathrm{TeV}} \sim \frac{30G\mu}{\chi^2}\Omega_\gamma \tag{42}$$

where Ω_γ is the fractional cosmic radiation energy density. The observed gamma ray flux dies off very rapidly at such high energies. Using the numerical values in Ref. [44], the relative energy density in cosmic gamma rays at energies above say 100 GeV is $\Omega_{\gamma>100\mathrm{GeV}} \lesssim 10^{-11}$, thus leading to the constraint

$$\frac{M^2}{m_P\sqrt{\mu}} \lesssim 10^{-6} \tag{43}$$

where $m_P = 1.2 \times 10^{19}$ GeV is the Planck mass. The constraint will be stronger if we restrict to gamma rays with energy greater than ~ 100 TeV where observations indicate a sharp cutoff in the gamma ray flux. It has been suggested that under some circumstances even particles of trans-Planckian energy could have been generated [45].

If all the magnetic flux of the monopoles is confined to the string, the monopoles do not radiate high energy photons even as they are accelerated by the connecting strings to relativistic energies. In this case, the web of strings and monopoles does not have an efficient way to dissipate its energy. The energy density in the web then dilutes due to Hubble expansion and due to occasional rearrangements when monopoles annihilate or when strings intercommute. As a result the relative energy density in the web compared to the matter density grows with time [43, 46]. Eventually the web dominates the cosmological matter energy density. Once the cosmological evolution of the web is understood in detail‡, the growth of the web relative energy density potentially leads to a constraint on the parameters of the fundamental model but a rigorous constraint has not been derived so far.

Current cosmological constraints on ordinary cosmic strings as derived from the millisecond pulsar timing observations limit the mass per unit length of the string, μ, to be less than $\sim 10^{-9}$ in Planck units, *i.e.* $\mu \lesssim 10^{19}$ gm/cm [47, 48]. This constraint depends on the gravitational radiation from strings, which is turn depends on the dynamics of strings, and in particular on the loop distribution. In the cases of beads on strings, the dynamics is expected to be similar to ordinary cosmic strings and so

‡ And with the inclusion of dark energy.

this constraint also applies. However, in the case of a web of strings, the dynamics is very different and loop formation is suppressed. In this case, the constraints from the non-detection of string gravitational lensing, and non-observation of string induced distortions of the angular power spectrum of the cosmic microwave background (CMB), can still be applied. These provide the bound $\mu \lesssim 10^{-7}$ (for a summary of observational bounds on cosmic strings, see Ref. [49]). If we combine this bound on the string tension with Eq. (43) for the case of string webs in which monopoles have unconfined gauge flux, we obtain a constraint on the monopole mass

$$M \lesssim 10^{-5} m_P. \tag{44}$$

There are two remnants that can arise from dumbbells that are created at some cosmological epoch that can potentially lead to an observable signature. The first remnant is simply the energy resulting from the decay of dumbbells provided they decay at cosmological redshifts between $z \approx 10^4$ and 10^6. In this case there is not enough time left until hydrogen recombination for the energy injected into the cosmological medium to get thermalized. As a result, the decay of dumbbells can distort the spectrum of the CMB. No such distortions have been measured so far and this limits the amount of energy deposition in the medium. However, the cosmic temperature at these redshifts is < 1 keV and the cosmic time is ~ 1 yr, and from the particle physics side, we think we know that there are no dumbbells that can survive for this long a period. The exception is if plasma effects can somehow play a role as discussed in the case of "embedded defects" in Ref. [50, 51] or if quantum effects are important and stabilize the dumbbells [52].

The second remnant produced by decaying dumbbells is a magnetic field that can be trapped in the cosmological medium, which can then survive until the present epoch [53, 54]. Indeed, primordial magnetic fields may also help explain the ubiquity of magnetic fields seen in galaxies and clusters of galaxies (for a recent review, see [55]).

We have already related twisted dumbbells to the electroweak sphaleron in Sec. 4 (see around Eq. (39)). If we assume that the cosmic matter-antimatter asymmetry is generated dynamically via sphaleron processes, then sphaleron decay will leave behind twisted or "helical" magnetic fields [56, 57]. Such magnetic fields violate parity since the handedness of the field is related to the preference of matter over antimatter. Evidence for helical cosmological magnetic fields has recently been discovered [58, 59], suggesting that they may have been produced during the decay of monopoles-on-strings.

If the monopoles on dumbbells do not carry unconfined flux, they will still lose energy by emitting gravitational waves, thus providing a third cosmological remnant from dumbbells. Further, if the dumbbells are sufficiently long at production, as can happen if the strings are produced after an inflationary epoch or with certain string theory cosmic strings, the distribution of dumbbells will produce a gravitational wave background [61] and gravitational wave bursts [62]. Upcoming gravitational wave detectors can be sensitive to the bursts and can potentially provide constraints at the level $G\mu \lesssim 10^{-12}$.

6. Conclusions

Field theories admit a wide variety of topological defects of which monopoles, strings, and domain walls are commonly discussed. In this review we have focussed on a type of "hybrid" or "composite" defect, namely monopoles connected by strings. We have discussed field theories in which monopoles are connected to 1, 2, or 3 strings. The case of one string per monopole is relevant to the electroweak model, and also to a proposed explanation of the observed absence of cosmological magnetic monopoles [60]. For more than 1 string per monopole, we have considered the symmetry breaking pattern $SU(N) \to SU(N-1) \times U(1) \to \mathbb{Z}_N$. The first stage of symmetry breaking gives monopoles and the second connects the monopoles to N strings. We have also described the monopoles connected by a single string arising in the symmetry breaking $SU(2) \times U(1) \to U(1) \times U(1) \to U(1)$ and this is directly relevant to the standard electroweak model.

Monopoles-on-strings can have observable effects in cosmology and ongoing observational efforts constrain their abundance. If monopoles are connected by 2 or more strings, a string network should exist in the universe. The strongest bounds on a string network arise from gravitational radiation from loops of strings and lead to $\mu \lesssim 10^{-9}$, where μ is the string tension in Planck units. The bound may not apply to the string web in which monopoles are connected by more than 2 strings, since the loop distribution will likely be suppressed. Gravitational lensing constraints still imply $\mu \lesssim 10^{-7}$. Non-gravitational constraints due to particle emission have also been derived in the literature and are summarized in Sec. 5.

The case when a monopole is connected by a single string is special because the strings then bring monopoles and antimonopoles together, and the whole system can rapidly annihilate. In this case, cosmological observables can only be sensitive to the decay products of the system. Since the annihilation of monopoles and antimonopoles releases magnetic fields, the growing evidence for cosmological magnetic fields may indeed indicate a role for monopoles-on-strings in the early universe.

Acknowledgements

This work was supported by the Department of Energy at Arizona State University.

References

[1] U. Amaldi, W. de Boer, P. H. Frampton, H. Furstenau and J. T. Liu, Phys. Lett. B **281**, 374 (1992).
[2] H. E. Haber, Nucl. Phys. Proc. Suppl. **62**, 469 (1998), [arXiv:hep-ph/9709450].
[3] Y. Mambrini, N. Nagata, K. A. Olive, J. Quevillon and J. Zheng, Phys. Rev. D **91**, no. 9, 095010 (2015) [arXiv:1502.06929 [hep-ph]].
[4] S. Sarangi and S. H. H. Tye, Phys. Lett. B **536**, 185 (2002) [hep-th/0204074].
[5] R. Jeannerot, J. Rocher and M. Sakellariadou, Phys. Rev. D **68** (2003) 103514 [hep-ph/0308134].
[6] N. T. Jones, H. Stoica and S. H. H. Tye, JHEP **0207** (2002) 051 [hep-th/0203163].

[7] M. Majumdar and A.-C. Davis, JHEP **0203** (2002) 056 [hep-th/0202148].
[8] K. Becker, M. Becker and A. Krause, Phys. Rev. D **74** (2006) 045023 [hep-th/0510066].
[9] M. G. Jackson, N. T. Jones and J. Polchinski, JHEP **0510** (2005) 013 [hep-th/0405229].
[10] K. Hashimoto and D. Tong, JCAP **0509** (2005) 004 [hep-th/0506022].
[11] J. Polchinski, Lectures presented at the 2004 Cargèse Summer School [hep-th/0412244].
[12] G. Dvali and A. Vilenkin, JCAP **0403** (2004) 010 [hep-th/0312007].
[13] A. Avgoustidis and E. P. S. Shellard, Phys. Rev. D **73** (2006) 041301 [astro-ph/0512582].
[14] S.-H. H. Tye, I. Wasserman and M. Wyman, Phys. Rev. D **71** (2005) 103508 [Phys. Rev. D **71** (2005) 129906] [astro-ph/0503506].
[15] M. Hindmarsh and P. M. Saffin, JHEP **0608** (2006) 066 [hep-th/0605014].
[16] M. A. C. Kneipp, Phys. Rev. D **76**, 125010 (2007) [arXiv:0707.3791 [hep-th]].
[17] N. Sakai and D. Tong, JHEP **0503** (2005) 019 [hep-th/0501207].
[18] D. Tong, TASI lectures on solitons: *Instantons, monopoles, vortices and kinks* [hep-th/0509216].
[19] T. W. B. Kibble, in *Patterns of Symmetry Breaking*, ed. H. Arodz, J. Dziarmaga & W.H. Zurek, NATO Science II, **127**, pp. 3–36 (2002). [arXiv:cond-mat/0211110].
[20] N. Turok, Phys. Rev. Lett. **63**, 2625 (1989).
[21] H. B. Nielsen and P. Olesen, Nucl. Phys. B **61** (1973) 45.
[22] A. Achucarro and T. Vachaspati, Phys. Rept. **327**, 347 (2000), arXiv:hep-ph/9904229.
[23] M. Hindmarsh, Phys. Rev. Lett. **68**, 1263 (1992).
[24] M. James, L. Perivolaropoulos and T. Vachaspati, Nucl. Phys. B **395**, 534 (1993) [hep-ph/9212301].
[25] G. 't Hooft, Nucl. Phys. B **79** (1974) 276.
[26] A. M. Polyakov, JETP Lett. **20** (1974) 194 [Pisma Zh. Eksp. Teor. Fiz. **20** (1974) 430].
[27] E. B. Bogomol'nyi, Sov. J. Nucl. Phys. **24**, 449 (1976).
[28] M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35** (1975) 760.
[29] T. W. B. Kibble, G. Lazarides and Q. Shafi, Phys. Rev. D **26** (1982) 435.
[30] L. Leblond and M. Wyman, Phys. Rev. D **75** (2007) 123522 [astro-ph/0701427].
[31] Y. Ng, T. W. B. Kibble and T. Vachaspati, Phys. Rev. D **78** (2008) 046001 [arXiv:0806.0155 [hep-th]].
[32] H. Georgi, Front. Phys. **54** (1982) 1.
[33] J. Heo and T. Vachaspati, Phys. Rev. D **58** (1998) 065011 [hep-ph/9801455].
[34] Y. Nambu, Nucl. Phys. B **130**, 505 (1977).
[35] N. S. Manton, Phys. Rev. D **28**, 2019 (1983).
[36] T. Vachaspati and G. B. Field, Phys. Rev. Lett. **73**, 373 (1994) [hep-ph/9401220].
[37] C. H. Taubes, Commun. Math. Phys. **86**, 257 (1982).
[38] E. J. Copeland, D. Haws, T. W. B. Kibble, D. Mitchell and N. Turok, Nucl. Phys. B **298**, 445 (1988).
[39] M. Hindmarsh and T. W. B. Kibble, Phys. Rev. Lett. **55**, 2398 (1985).
[40] V. Berezinsky and A. Vilenkin, Phys. Rev. Lett. **79**, 5202 (1997) [astro-ph/9704257].
[41] X. Siemens, X. Martin and K. D. Olum, Nucl. Phys. B **595**, 402 (2001) [astro-ph/0005411].
[42] J. J. Blanco-Pillado and K. D. Olum, JCAP **1005**, 014 (2010) [arXiv:0707.3460 [astro-ph]].
[43] T. Vachaspati and A. Vilenkin, Phys. Rev. D **35**, 1131 (1987).
[44] M. Ackermann *et al.* [The Fermi LAT Collaboration], Astrophys. J. **799**, no. 1, 86 (2015) [arXiv:1410.3696 [astro-ph.HE]].
[45] V. Berezinsky, X. Martin and A. Vilenkin, Phys. Rev. D **56**, 2024 (1997) [astro-ph/9703077].
[46] C. J. A. P. Martins, Phys. Rev. D **82**, 067301 (2010) [arXiv:1009.1707 [hep-ph]].
[47] F. A. Jenet, G. B. Hobbs, W. van Straten, R. N. Manchester, M. Bailes, J. P. W. Verbiest, R. T. Edwards and A. W. Hotan *et al.*, Astrophys. J. **653**, 1571 (2006) [astro-ph/0609013].
[48] J. J. Blanco-Pillado, K. D. Olum and B. Shlaer, Phys. Rev. D **89**, no. 2, 023512 (2014) [arXiv:1309.6637 [astro-ph.CO]].
[49] Tanmay Vachaspati, Levon Pogosian, and Danièle A. Steer, (2015) *Cosmic strings*. Scholarpedia, 10(2):31682. `http://www.scholarpedia.org/article/Cosmic_strings`

[50] M. Nagasawa and R. Brandenberger, Phys. Rev. D **67**, 043504 (2003) [hep-ph/0207246].
[51] J. Karouby and R. Brandenberger, Phys. Rev. D **85**, 107702 (2012) [arXiv:1203.0073 [hep-th]].
[52] H. Weigel, M. Quandt and N. Graham, [arXiv:1505.07631 [hep-th]].
[53] T. Vachaspati, Phys. Lett. B **265**, 258 (1991).
[54] T. Vachaspati, Sintra Electroweak 1994:171-184 [hep-ph/9405286].
[55] K. Subramanian, *The origin, evolution and signatures of primordial magnetic fields*, arXiv:1504.02311 [astro-ph.CO].
[56] C. J. Copi, F. Ferrer, T. Vachaspati and A. Achucarro, Phys. Rev. Lett. **101**, 171302 (2008) [arXiv:0801.3653 [astro-ph]].
[57] Y. Z. Chu, J. B. Dent and T. Vachaspati, Phys. Rev. D **83**, 123530 (2011) [arXiv:1105.3744 [hep-th]].
[58] H. Tashiro, W. Chen, F. Ferrer and T. Vachaspati, Monthly Notices of the Royal Astronomical Society: Letters 2014 445 (1): L41-L45 [arXiv:1310.4826 [astro-ph.CO]].
[59] W. Chen, B. D. Chowdhury, F. Ferrer, H. Tashiro and T. Vachaspati, *Intergalactic magnetic field spectra from diffuse gamma rays*, arXiv:1412.3171 [astro-ph.CO].
[60] P. Langacker and S. Y. Pi, Phys. Rev. Lett. **45**, 1 (1980).
[61] X. Martin and A. Vilenkin, Phys. Rev. Lett. **77**, 2879 (1996) [astro-ph/9606022].
[62] L. Leblond, B. Shlaer and X. Siemens, Phys. Rev. D **79**, 123519 (2009) [arXiv:0903.4686 [astro-ph.CO]].

Cosmic strings*

Tanmay Vachaspati,[1,†] Levon Pogosian,[2,‡] and Danièle A. Steer[3,4,5,§]

[1]*Physics Department, Arizona State University, Tempe, Arizona 85287, USA.*
[2]*Department of Physics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada*
[3]*AstroParticule & Cosmologie, UMR 7164-CNRS,*
Université Denis Diderot-Paris 7, CEA, Observatoire de Paris,
10 rue Alice Domon et Léonie Duquet, F-75205 Paris Cedex 13, France
[4]$\mathcal{GR}\varepsilon\mathcal{CO}$*, Institut d'Astrophysique de Paris, CNRS, UMR 7095, et Sorbonne Universités,*
UPMC Univ Paris 6, 98bis boulevard Arago, F-75014 Paris, France
[5]*LAPTH, Université Savoie Mont Blanc, CNRS,*
B.P.110, F-74941 Annecy-le-Vieux Cedex, France

(Dated: June 22, 2015)

This article, written for Scolarpedia [1], provides a brief introduction into the subject of cosmic strings, together with a review of their main properties, cosmological evolution and observational signatures.

Contents

---

* Review published in Scholarpedia [1], `http://www.scholarpedia.org/article/Cosmic_strings`
†Electronic address: tvachasp@asu.edu
‡Electronic address: levon@sfu.ca
§Electronic address: steer@apc.univ-paris7.fr

I. INTRODUCTION

"Strings" are solutions of certain field theories, whose energy is concentrated along an infinite line. Strings exist in many field theories motivated by particle physics, and this suggests that they may exist in the universe — hence the name "cosmic strings". String solutions are also present in condensed matter systems where they are called "vortices". In cosmological applications, strings are generally curved, dynamical, and may form closed loops. The energy of a string remains concentrated along a time-dependent curve for a duration that is very long compared to the dynamical time of the string.

II. ROLE OF TOPOLOGY

The topological properties of a field theory may be used to motivate the existence of string solutions. If a field theory has certain symmetries and symmetry breaking patterns, the vacuum state (the state of lowest energy) may not be unique. The collection of possible vacua form a manifold, M, which may have "holes" i.e. there may be closed paths on M that cannot be continuously shrunk to a point. In this case, the field theory has topology that is suitable for the existence of string solutions. In mathematical terms, the topology relevant for strings is described by the first homotopy group of the vacuum manifold, $\pi_1(M)$. The relevance of topology is best understood with an example.

Consider a complex scalar field, Φ, in three spatial dimensions, with potential energy function,

$$V(|\Phi|) = (|\Phi|^2 - \eta_v^2)^2$$

The minimum energy configuration has $|\Phi| = \eta_v$ but the phase of Φ is undetermined and labels the points on the vacuum manifold which is a circle. A closed path that wraps around the circle cannot be continuously contracted to a point and hence there can be strings in this field theory. If, as one goes around a closed path in physical space, one also wraps around around the circle on the vacuum manifold $n = \pm 1, \pm 2, \ldots$ times, then there will be possibly n strings going through the closed path in physical space. (See Fig. 1.) Notice that at the center of the string $|\Phi| = 0$ and hence the energy density is non-zero at the string core.

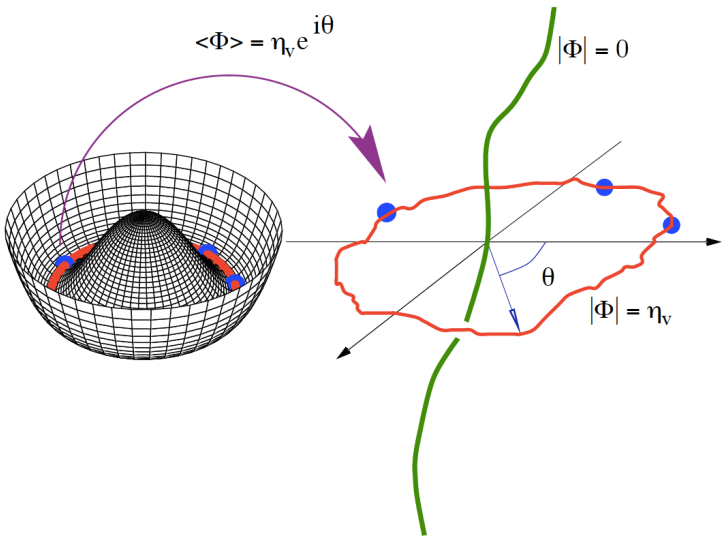

FIG. 1: String formation in the "Mexican-hat" potential $V(|\Phi|)$. The potential is shown on the left-hand-side, with its circular manifold (red) on which 3 points (blue) have been chosen at random. The right-hand-side shows (in red) a closed path in "physical space", along which $|\Phi| = \eta_v$; the blue points in physical space are the points that map on to the blue points shown on the vacuum manifold. On going around the path in physical space, the field wraps once around the vacuum manifold . By continuity of the field, Φ must vanish somewhere within the circle in physical space. This is the center of the string drawn in green. Fig. from [2].

Caution: Non-trivial topology of a field configuration does not necessarily imply the existence of a *static* solution. For example, if two $n = 1$ strings repel at all separations, then an $n = 2$ configuration will split into two $n = 1$ strings that move apart. In this case, there is non-trivial topology since $n = 2$, but no static solution exists.

III. SOLUTION AND PROPERTIES

A. Straight global string

For a straight, static string, it is sufficient to look for a solution of the equations of motion in two spatial dimensions, and then use translation invariance to extend the solution to three dimensions. For example, if the solution in two dimensions is $\Phi_0(x, y)$, then the solution in three dimensions is $\Phi(x, y, z) = \Phi_0(x, y)$.

The simplest theory which gives rise to string solutions is described by the Lagrangian

$$L = |\partial_\mu \Phi|^2 - \frac{\lambda}{4}(|\Phi|^2 - \eta^2)^2$$

where λ is a dimensionless coupling constant, η is the vacuum expectation value of the field Φ, and the metric has signature $(+, -, -, -)$. We will also use natural units throughout so that $\hbar = c = 1$. This Lagrangian is invariant under a global $U(1)$ symmetry, $\Phi \to \Phi e^{i\Lambda}$ (for any constant Λ), and the corresponding equations of motion are

$$\partial_\mu \partial^\mu \Phi = -\lambda(|\Phi|^2 - \eta^2)\Phi.$$

The static string solution in this model is

$$\Phi(x, y) = \eta\, f(m\rho) e^{in\theta}$$

where (ρ, θ) are polar coordinates on the xy-plane, $m^2 = \lambda\eta^2$, and n is the (integer) winding number of the string. On substituting into the equations of motion, the function $f(\tilde{\rho})$, $\tilde{\rho} \equiv m\rho$, has the features

$$f(\tilde{\rho}) = f_0 \tilde{\rho}^{|n|}(1 + \mathcal{O}(\tilde{\rho}^3)), \qquad \tilde{\rho} \ll 1$$
$$f(\tilde{\rho}) = 1 - \mathcal{O}\left(\frac{1}{\tilde{\rho}^2}\right), \qquad \tilde{\rho} \gg 1.$$

The energy density $\mathcal{E} = |\vec{\nabla}\Phi|^2 + V(\Phi)$ is peaked within $\rho \sim m$, and falls off as $1/\rho^2$ at large distances. The total energy per unit length of the string diverges weakly (logarithmically). In a physical setting when there are lots of strings or in a condensed matter sample of finite volume, the divergence gets cut off. This string solution is known as a "global" string because there are no gauge fields in the model.

B. The Nielsen-Olesen string

The model can be extended to include gauge fields

$$L = |D_\mu \Phi|^2 - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{\lambda}{4}(|\Phi|^2 - \eta)^2$$

where $D_\mu = \partial_\mu - ieA_\mu$ is the gauge covariant derivative, and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. This "Abelian Higgs model" was considered by Nielsen and Olesen in their discovery paper on string solutions in relativistic field theories [3]. The Lagrangian is now invariant under a local $U(1)$ symmetry in which $\Phi \to \Phi e^{i\Lambda(x)}$ and $A_\mu \to \partial_\mu \Lambda / e$. In the static string configuration, the asymptotic properties of the scalar field differs from the global case. In particular as $\rho \to \infty$, the scalar and gauge fields both contribute to the energy density and make it fall off exponentially fast, and the energy per unit length μ of the string is finite.

The string also contains a flux of magnetic field that is quantized as can be seen by noting that $D_\mu \Phi \to 0$ outside the string,

$$\text{Magnetic flux} = \oint dx^\mu A_\mu = \frac{2\pi n}{e}$$

1. *Type I and Type II strings*

The properties of strings in the local $U(1)$ model depend on the ratio of coupling constants $\beta = \lambda/2e^2$. In the limit $\beta = 1$ the equations of motion simplify and an important method to find string solutions, also often used in supersymmetric field theories, was developed by Bogomol'nyi [4]. Here the energy per unit length μ_n of a string with

winding number n is exactly equal to that of n strings each with winding number 1. In this case $\mu_1 = \pi\eta^2$ and $\mu_n = n\mu_1$.

In the limit $\lambda < 1$, often called the type I regime in analogy with superconductors, $\mu_n < n\mu_1$. In particular, two $n = 1$ strings can merge to form an $n = 2$ string [5], see Fig. 2.

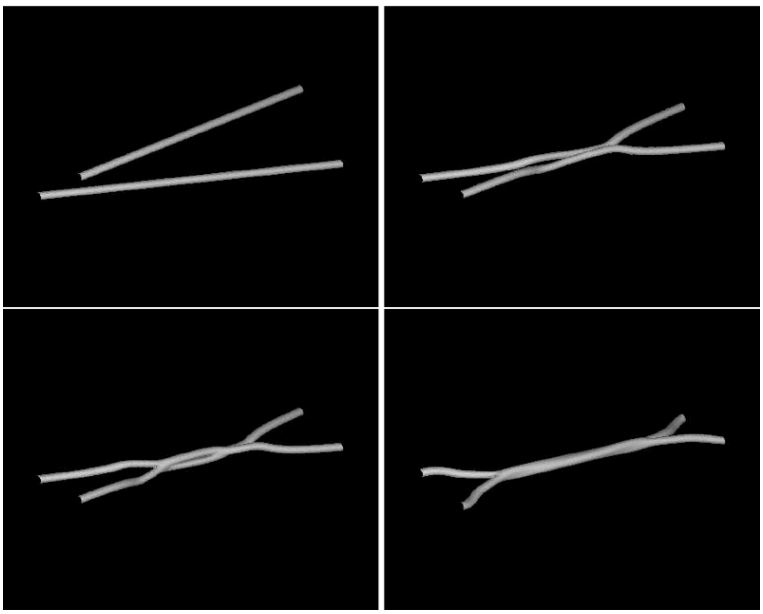

FIG. 2: Snapshots (from left to right and top to bottom) showing constant energy density isosurfaces in a simulation of two strings with $n = 1$ in the type I regime ($\beta = 0.125$) colliding to form an $n = 2$ string . From [6].

In the type II regime, $\lambda > 1$ and only the winding 1 strings are stable. Usually when discussing the cosmological properties of cosmic strings, the strings being considered are those of the local $U(1)$ model in the type II regime. These are referred to as "gauge strings" or "local strings". When two type II strings collide, for essentially all angles and collision velocities, they "intercommute": that is, they exchange partners (Fig. 3). Thus gauge strings have an intercommutation probability $P = 1$ [7, 8], except at very high incoming velocities [9].

C. Other types of strings

If there are fermions in the model that couple to the scalar field that winds around the string, "fermion zero modes" may exist [10]. These are solutions of the Dirac equation that are localized on the string and have zero energy. If the fermions also carry electromagnetic charge, the cosmic strings can carry electric currents, leading to interesting astrophysical signatures in the cosmological context. In some models, charged scalar fields can also be localized on the string. Current-carrying strings are also known as "superconducting strings" [11].

Many other types of strings (e.g. semi-local strings, Alice strings etc) can form depending on the topology and coupling to other fields (see [12–16]).

In summary, the basic structure of a string is a scalar field that winds around the location of the string, where there is a concentration of energy density. Gauge fields that interact with the scalar field provide the string with a quantized magnetic flux. Fermion zero modes can be localized on the string and be responsible for currents that run along the string.

IV. BULK PROPERTIES

If the mass of the scalar field that winds around the string is m and all dimensionless coupling constants are O(1), the width of a local string is $\approx m^{-1}$. In most cosmological applications, the width of the string is very small compared to the other length scales in the problem, and the thin string limit is commonly adopted. Then the string is simply modeled as a line with mass per unit length $\mu \approx m^2$. If the string is not superconducting, its tension T, *i.e.* the longitudinal component of the string energy-momentum tensor, is also μ. In the zero-width approximation,

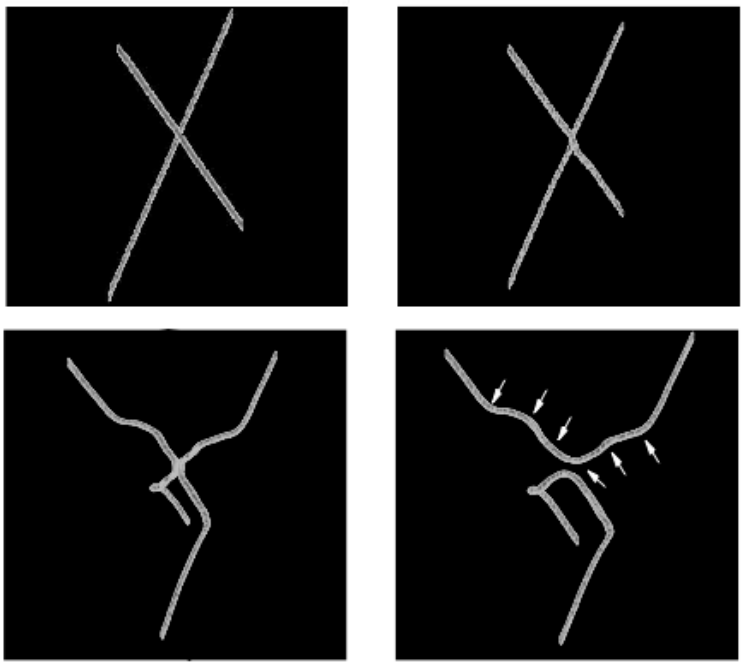

FIG. 3: Snapshots (from left to right and top to bottom) showing constant energy density isosurfaces in a simulation of two strings with $n = 1$ in the type II regime ($\beta = 32$) colliding and intercommuting with the formation of kinks (as indicated by arrows). From [9].

the strings are referred to as "Nambu-Goto" strings as their dynamics is obtained by solving the Nambu-Goto action which minimises the area swept out by the worldsheet of the string. Numerically, $\mu \approx 10^{22}$ gms/cm and $G\mu \approx 10^{-6}$ when $m \approx 10^{16}$ GeV.

An important feature of Nambu-Goto strings is that they contain "kinks" and "cusps". A kink is a point at which the tangent vector of the string changes discontinuously, and kinks are formed when strings intercommute (Fig. 3). Kinks travel along the string at the speed of light. At a cusp, the string instantaneously travels at the speed of light. Kinks and cusps give rise to important observational signatures of strings (see below).

Superconducting strings can carry an electric current j which can be timelike, spacelike or lightlike, and leads to an equation of state of the string $T = T(\mu)$. In general, the maximum current allowed on the string is $O(m^2)$ though, depending on the detailed particle interactions, it can be substantially weaker [17, 18]. The effective action for superconducting strings is no longer the Nambu-Goto action.

The metric around a static infinitely straight Nambu-Goto string lying along the z-axis can be obtained by solving the Einstein equations. It is "conical" on the plane transverse to the string, and the line element is [19]

$$ds^2 = dt^2 - dz^2 - d\rho^2 - \rho^2 d\theta^2 \ , \quad 0 \leq \theta < 2\pi(1 - 4G\mu)$$

where G is Newton's gravitational constant. As is apparent, the metric is locally flat and the only non-trivial feature is that the angular coordinate θ lies in an interval that is less than 2π. This particular form of the metric is central to many of the observational signatures of cosmic strings described below.

In physical applications, a whole network of strings is formed when the symmetry is broken, and individual strings can be infinitely long or in the shape of closed loops, and the network evolves in time. A curved string is a dissipative solution of the equations of motion. Loops will eventually decay into various forms of radiation including the scalar and gauge fields of which the string is formed, and gravitational waves through the coupling of the string to gravity; infinite curved strings will tend to straighten out. The dissipation time-scale is generally very long compared to the dynamical time of loops for long loops, so the string picture is useful. For example, a loop will oscillate $\sim (100G\mu)^{-1}$ times, (that is $\sim 10^5$ times for the string tension of $G\mu \sim 10^{-7}$), before losing an $O(1)$ fraction of its energy to gravitational waves.

In certain field theories, strings networks can also have junctions — namely points at which three strings meet. In particular this occurs in the Abelian-Higgs model when $\beta < 1$ as is shown in Fig. 2. Junctions also occur in more complicated models in which non-abelian symmetries are broken. Cosmic superstring networks, predicted in fundamental superstring theories, also have junctions. There they are located at the meeting point between

fundamental F-strings, Dirichlet D-strings and a bound states of these two. Cosmic superstrings are known as (p, q)-strings, an abbreviation for p F-strings and q D-strings [20, 21]. The effective action for strings with junctions is a set of three coupled Nambu-Goto actions [22], but it should be noted that the intercommutation probability of superstrings $P < 1$.

Cosmic strings also interact with the ambient cosmological medium. Particles scatter off the string with differential cross-section per unit length [23]

$$\frac{d\sigma}{d\theta} = \frac{\pi}{2k(\log(k\delta))^2},$$

where k is the momentum of the particle transverse to the string and δ is the string width. The cross-section is larger if there are particles that scatter by the "Aharonov-Bohm" interaction [24–26],

$$\frac{d\sigma}{d\theta} = \frac{\sin^2(\pi\nu)}{2\pi k \sin^2(\theta/2)}$$

where $2\pi\nu$ is the Aharonov-Bohm phase. Note that the scattering cross-sections only depend on the momentum of the incoming particle, and are insensitive to the mass scale of the string. The interaction of strings with ambient particles plays an important role in the early stages after a string network forms as it over-damps the string dynamics. However, as the universe expands, the density of ambient matter falls and particle interactions cease to be an important factor.

V. COSMOLOGY

A. Formation

Based on our current understanding of particle physics, the vacuum structure may have topology that is suitable for the existence of string solutions. The mathematical existence of string solutions in a field theory, however, does not imply that they will be realized in a physical setting and additional arguments are needed to make the case that strings can be present in the universe [27]. Essentially, during spontaneous symmetry breaking, different vacua are chosen in different spatial domains, and the non-trivial topology of the vacuum manifold then inevitably implies the presence of strings in cosmology. At formation, a large fraction of the string network (roughly 80% in the simplest formation models) is in infinite strings and the rest is in loops with a scale-invariant distribution [28].

Subsequently, the network relaxes under several forces that include the string tension, frictional forces due to ambient matter, cosmic expansion, and the process of intercommuting. In particular when a loop or an infinite string intercommutes with itself, it chops off a loop. This means that at any given time the network will contain many loops: those formed at time t as well as at all previous times. In addition, a Nambu-Goto loop evolves periodically in time and hence loses energy to gravitational and other forms of radiation. As a result a loop of initial length ℓ_0 formed at time t_0 has a length $\ell(t)$ at time t given by

$$\ell(t) = \ell_0 - \Gamma G\mu(t - t_0)$$

where Γ is a constant which, in the case of gravitational radiation, is of order 100 [29] (the precise value depends on the shape of the loop). A typical loop will have a number of kinks and cusps, and the spectrum of high-frequency gravitational radiation emitted from a string depends on these features.

B. Evolution

The evolution of the network from its formation until today is an extremely complex problem involving very disparate length scales. Several groups have tackled this problem by using numerical simulations of Nambu-Goto (zero thickness limit) strings, starting with the pioneering work in [30–32]. Other groups have performed field theory simulations in which the strings have structure. And yet others have built analytical models to describe the evolution of the network. These analyses show that the network reaches a self-similar attractor solution on large scales in which all the properties and length scales describing the network scale with time. In this 'scaling solution', the physical number density of loops $n(\ell, t)d\ell$ with length between ℓ and $\ell + d\ell$ at cosmic time t is characterized by different power-law behaviors. For example, recent simulations by [33], give

$$n(\ell, t) = \frac{0.18}{t^{3/2}(\ell + \Gamma G\mu t)^{5/2}}, \quad \ell < 0.1t \quad \text{(radiation era)}$$

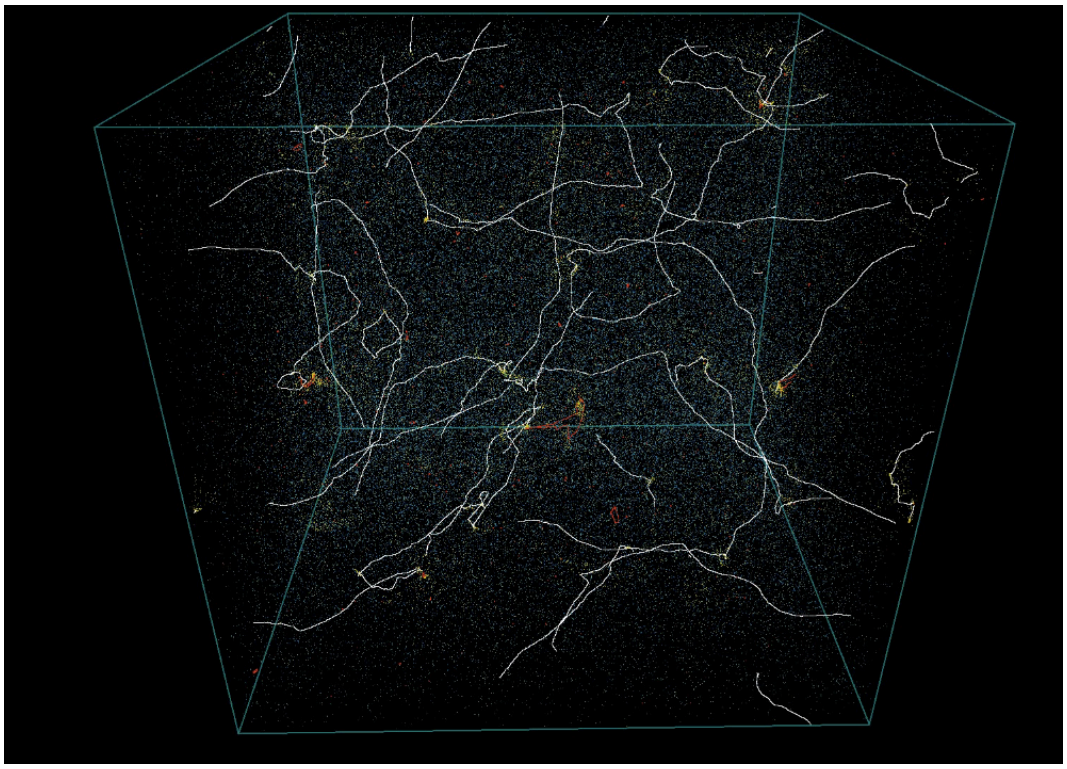

FIG. 4: Nambu-Goto strings in the matter era. Infinite strings (defined as having a size bigger than the conformal horizon, which is the box size in this figure) are colored in white. The red loops are those whose average density has reached scaling. The remaining loops are not (yet) in scaling, and they are color-coded according to their age: freshly formed loops are yellow and older loops are bluer. The bluest ones are therefore those coming from the initial conditions and the greenish ones have been formed in between. Figure given with kind permission by Christophe Ringeval.

$$n(\ell,t) = \frac{0.27 - 0.45(\ell/t)^{0.31}}{t^2(\ell + \Gamma G\mu t)^2}, \quad \ell < 0.18t \quad \text{(matter era)}.$$

Nambu-Goto simulations by other groups (see for instance [34]) find similar loop distributions though there is disagreement about very small loops that, however, carry only a very small fraction of the total energy in the network. In Abelian-Higgs simulations, many fewer loops are seen and the string network energy is mostly dissipated directly into particle radiation [35].

Notice that the above are the pure scaling distributions, meaning that the dimensionless number density of loops $t^4 n(\ell, t)$ only depends on the dimensionless ratio ℓ/t for all times. At formation though, the loops are not in the scaling distribution: they relax towards scaling after a time which can be estimated. Numerical simulations, however, observe a population of non-scaling loops. Some of these are a remnant of the initial loop distribution formed at the phase transition, and others are small loops freshly formed from small scale structure on long strings (see Fig. 4). Similarly, on entering the matter era, the radiation era scaling distribution relaxes to the matter era scaling distribution. The timescale for this process depends on the length of the loop, and is longer for shorter loops.

In addition to loops the network contains many infinite strings with typical separation $\sim 0.15 d_h$ where d_h is the horizon distance [36]. This corresponds to ~ 40 infinite strings in any horizon volume. A typical distribution of strings is show in Fig. 4.

VI. HOW WE CAN LOOK FOR THEM

The presence of strings in the universe can be deduced from their gravitational effects and other non-gravitational signatures if they happen to couple to other forces. For example, cusps on cosmic string loops emit bursts of gravitational waves [37]. Moving strings produce wakes in matter and line discontinuities in the cosmic microwave background (CMB). They also induce characteristic patterns of lensed images of background light sources. Superconducting strings, in addition to the above effects, emit electromagnetic radiation that can potentially be detected as radio

bursts. At present, the strongest bounds on the string tension come from constraints on the stochastic gravitational wave background from pulsar timing measurements and the LIGO interferometer. However, these bounds are sensitive to the details of the string network evolution. On the other hand, bounds from CMB are weaker but also less model-dependent. Different types of cosmic string signatures and their current status are reviewed below.

A. Gravitational signatures

1. *Cosmic Microwave Background*

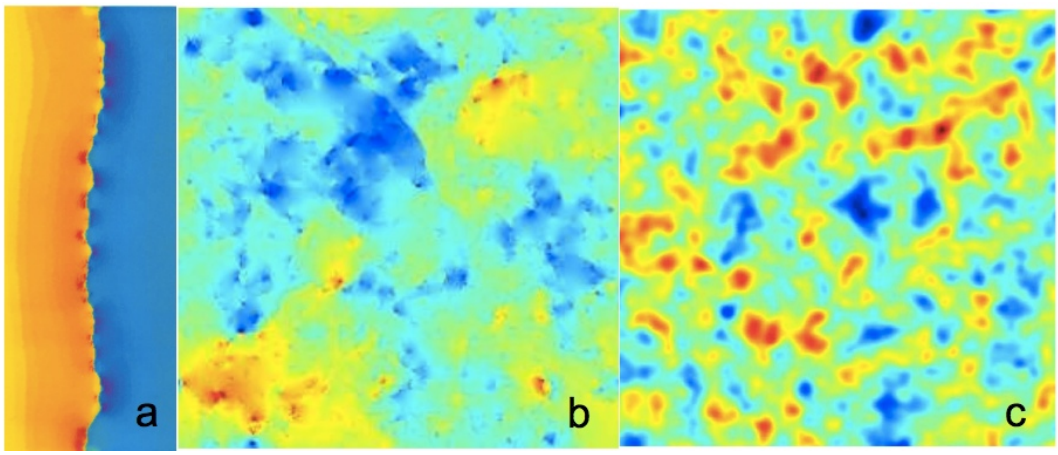

FIG. 5: CMB anisotropy sourced by strings: a) a line-discontinuity in CMB temperature caused by a single string on a uniform background (image provided by Proty Wu and Paul Shellard, (J.H.P.Wu PhD thesis, U. of Cambridge, 2000)); b) anisotropy caused by a network of strings via the Kaiser-Stebbins-Gott effect alone (image from [38]); c) anisotropy caused by a network of strings with full recombination physics taken into account. (Image from [39].)

Cosmic string networks persist throughout the history of the universe and actively source metric perturbations at all times. Prior to cosmic recombination, density and velocity perturbations of baryon-photon fluid are produced in the wakes of moving cosmic strings, which then remain imprinted on the surface of last scattering. After recombination, strings crossing our line of sight generate line-like discontinuities in the CMB temperature, which is the so-called Kaiser-Stebbins-Gott (KSG) effect [40–42]. Both, wakes and the KSG effect, are induced by the deficit angle in the metric around a string. In addition, matter particles experience gravitational attraction to the string if it is not perfectly straight. The search for cosmic string signatures in the CMB can be broadly divided into attempts to *directly* detect line discontinuities in the temperature or polarization patterns, and *statistical methods* based on calculations of various correlation functions.

Direct Searches. The spacetime around a straight cosmic string is locally flat, but globally conical, with a deficit angle determined by the string tension. Thus, a string passing across our line of sight would produce a discrete step in the CMB temperature proportional to $G\mu\,|\vec{v}\times\hat{n}|$, where $\vec{v}$ is the velocity of the string and $\hat{n}$ is the direction of the line of sight (see Fig. 5). Several groups have tried searching for such line-like features in the existing CMB maps and to forecast the prospects for future observations. The tightest existing bound on the string tension based on direct signatures is $G\mu < 3.7\times10^{-6}$ at the 95% confidence level [43] and is based on the assumption that the number density of strings is approximately known. The most optimistic forecast, based on Canny algorithm, claims that direct searches with future CMB experiments can achieve bounds of $G\mu < 3\times10^{-8}$ [44]. Detectable sharp edges can be present not only in CMB temperature maps, but also in *polarization* maps. The primary limitation in these types of studies comes not so much from the instrumental noise and angular resolution of the experiment, but from the fact that CMB is dominated by the Gaussian fluctuations on scales comparable to the size of the horizon at decoupling. Also, the above mentioned forecasts assume idealized line discontinuities produced by straight string segments. Actual strings are not straight, and contain both infinite strings and string loops. It remains to be seen how well these methods perform under more realistic assumptions.

Statistical Methods. The most reliable bounds on cosmic strings are derived from the angular power spectrum of CMB temperature anisotropies measured by the WMAP and Planck satellites. Calculating the spectrum of CMB anisotropies sourced by strings requires tracking their evolution from just before recombination until today over a

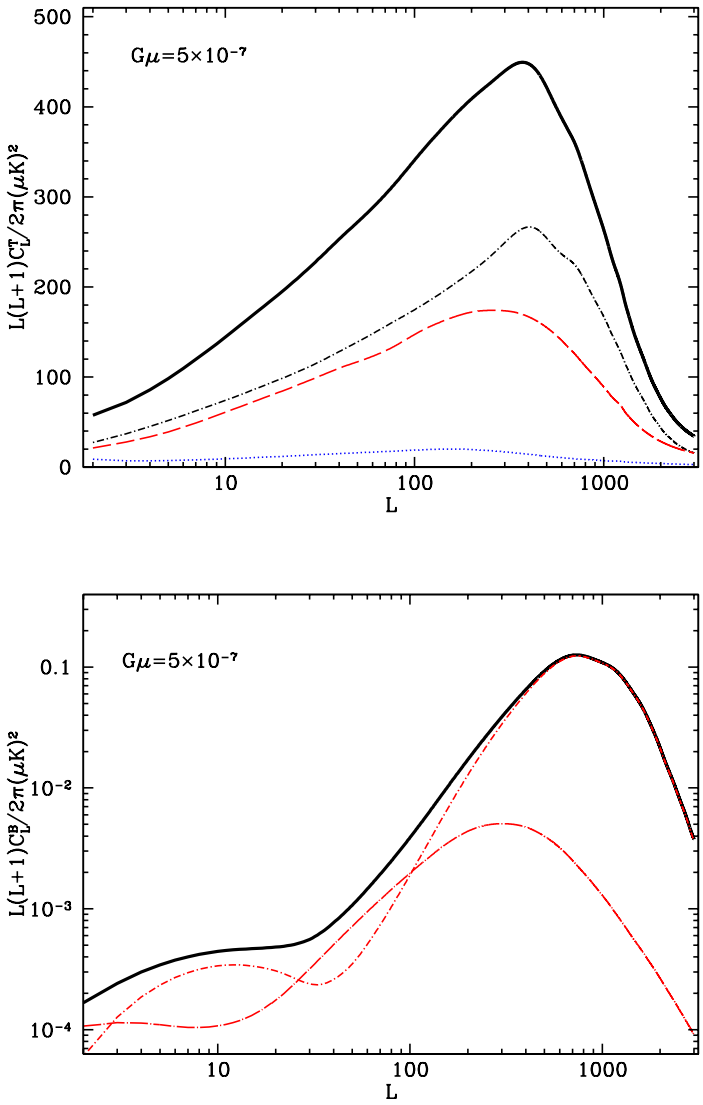

FIG. 6: The CMB temperature (top) and B-mode polarization (bottom) spectra sourced by strings. Contributions from scalar (black dash-dot), vector (red dash) and tensor modes (blue dot) and their sum (solid black) are shown separately.

large range of scales. Because obtaining an exact solution is quite challenging numerically, several approximate and semi-analytical methods have been used to evaluate the string CMB spectra. There is broad agreement between results from different approaches on the general shape of CMB spectra sourced by "local" cosmic strings, with the allowed fraction of the string contribution to the total CMB temperature anisotropy currently limited to under 3% [45]. For "conventional" strings, i.e. those with order unity inter-commutation probability, this implies $G\mu \lesssim few \times 10^{-7}$.

While most approaches adopt the Nambu-Goto approximation, [46], evolved the cosmic string field configurations in the Abelian-Higgs model. To make the calculation of CMB spectra numerically feasible, the fields were separately evolved over limited time ranges in the radiation and matter eras, and an interpolation scheme was used to connect the two scaling regimes. Perfect scaling was then assumed to extend their range to later times. To circumvent the problem of resolving the fixed width core of strings in an expanding background, the core size was allowed to grow with the expansion in a prescribed fashion. This approach is designed to directly calculate the spectra and cannot predict CMB maps.

Another method predicting the spectra but not the map, is the so-called Unconnected Segment Model implemented in the publicly available code CMBACT [47, 48]. In the unconnected segment model, the string network is represented as a collection of uncorrelated straight string segments.

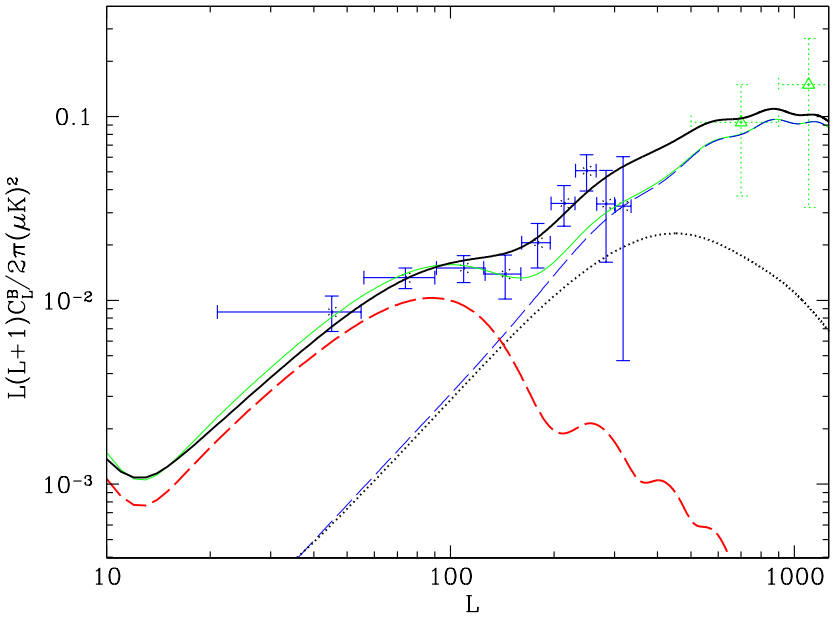

FIG. 7: B-mode spectra from cosmic strings (black dot), weak lensing (blue long dash), inflationary gravity waves with r=0.15 (red dash) and their sum (black solid) along with the BICEP2 [49] and POLARBEAR [50] data points. The faint solid line is the BICEP2 best fit model with r=0.2.

Although strings cannot be the main source of the CMB temperature anisotropy, they can generate observable B-mode polarization. The B-mode from strings is primarily generated by vector modes, with a spectrum that is different from the one generically produced from tensor modes arising in inflationary scenarios. Future CMB polarization experiments should be able to reveal the presence of cosmic strings through their B-mode signature even if strings contribute as little as 0.1% to the CMB temperature anisotropy. Constraints comparable to those previously obtained from CMB temperature spectra were obtained with the POLARBEAR [50] and BICEP2 [49] B-mode spectra in [51–54]. Given the uncertainty regarding the potentially significant unresolved foreground contributions to the BICEP2 signal, some of these results and conclusions may have to be revisited in the future.

Typical CMB temperature and polarization spectra sourced by local cosmic strings are shown in Fig. s 6 and 7. The temperature spectrum features a broad peak at multipole moment $L \sim 300 - 500$ and no acoustic peaks. The B-mode spectrum has a peak at $L \sim 600$ and lower peak at $L \sim 10$ due to polarization generated at the epoch of reionization. The position of the main peak is determined by the most dominant Fourier mode stimulated at last scattering. It primarily depends on the string correlation length and the average string velocity at last scattering. Measuring the location of the main peak would provide valuable insights into fundamental physics. For example, in the case of cosmic superstrings the position of the peak of the B-mode spectrum constrains the value of the fundamental string coupling g_s in string theory [55].

Fluctuations sourced by strings are intrinsically non-Gaussian and hence their statistical signatures are not limited to power spectra. Several groups have made predictions for various non-Gaussian estimators that could be sensitive to cosmic strings. So far, the resulting bounds on strings are not competitive with those derived from power spectra.

The CMB bispectrum and trispectrum induced by strings on small angular scales is generally suppressed by symmetry considerations, but the trispectrum can be large [56–58]. The trispectrum parameter τ_{NL} can be as large as 104 for strings, hence one can anticipate strong constraints on cosmic strings as observational estimates of the trispectrum improve.

2. 21-cm

The prospects of future observations of the 21 cm cosmological background motivated investigations of their ability to constrain cosmic strings. Neutral hydrogen absorbs or emits 21 cm radiation at all times after recombination. Cosmic strings would stir the hydrogen as they move around and create wakes, leading to 21 cm brightness fluctuations.

Under certain optimistic assumptions, future experiments with a collecting area of $10^4 - 10^6$ km^2, can in principle constrain $G\mu$ in the $10^{-10} - 10^{-12}$ range [59]. The same strings that create wakes would also perturb the CMB via the KSG effect, leading to potentially observable spatial correlations between the 21 cm and CMB anisotropies [60]. Also, the ionization fraction in the cosmic string wake is enhanced, leading to an excess 21 cm radiation confined to a wedge-shaped region [61]. It remains to be seen if terrestrial and galactic foregrounds (which become very bright at low frequencies) can be overcome to use 21 cm for mapping the high redshift distribution of matter.

3. Gravitational waves

Oscillating loops of cosmic strings generate a stochastic gravitational wave background that is strongly non-Gaussian, and includes occasional sharp bursts due to cusps and kinks [37]. In the case when loops are large at formation, which is the case favored by latest simulations, pulsars currently provide the tightest bounds of $G\mu \lesssim 10^{-9}$ for "conventional" strings, i.e. those with order unity intercommutation probability [33, 62]. (Notice that these bounds do not apply to global strings as they decay primarily through Goldstone boson radiation.)

In the case of cosmic *superstrings*, the probability of cusp formation is expected to be reduced due to the extra spatial dimensions, and there is some smoothing of the cusps. This can significantly damp the gravity waves emitted by cusps, and to a lesser extent by kinks, and relax pulsar timing bounds on cosmic superstrings. On the other hand, junctions on superstring loops give rise to a proliferation of sharp kinks that can amplify the gravitational wave footprint of cosmic superstrings [63].

4. Lensing

The peculiar form of the metric around a cosmic strings can result in characteristic lensing patterns of distant light sources. For instance, a straight long string passing across our line of sight to a distant galaxy can produce two identical images of the same galaxy [64]. In the more general case of loops and non-straight strings, the image patterns will be more complicated, but still have a characteristic stringy signature.

When strings bind and create junctions, as in the case of F-D superstring networks, the resulting configurations can lead to novel gravitational lensing patterns, like tripling of images when lensed by a Y-junction [65, 66].

The existence of cosmic strings can be strongly constrained by the next generation of gravitational lensing surveys at radio frequencies. LOFAR and SKA can give an upper bound of $G\mu < 10^{-9}$ [67]. Microlensing surveys are less constraining [68]. Effects of loop clustering on microlensing [69], gravitational lensing due to a moving string string on pulsar timing, and quasar variability [70] have also been considered with an aim to derive constraints. Cosmic string loops within the Milky Way can micro-lens background point sources and this offers a potentially powerful methodology for searching for cosmic strings [71].

Vector perturbations sourced by strings or other topological defects can generate a curl-like (or B-mode) component in the weak lensing signal which is not produced by standard density perturbations at linear order [72]. Future large scale weak lensing surveys should be able to detect this signal even for string tensions an order of magnitude lower than current CMB constraints.

B. Non-gravitational signatures

In the simplest cases, such as the Abelian Higgs model, the sole impact of cosmic strings on their surroundings is through their gravity. In extended models, in which cosmic string solutions occur within a more complete particle theory, it is quite common for strings to interact via forces present in the Standard Model. However, since the precise nature of the coupling is unknown, the non-gravitational signatures of strings are more model-dependent than those discussed in earlier sections.

1. Cosmic rays

If strings couple to other forces, cusps and kinks can emit beams of a variety of forms of radiation which can potentially be detected on Earth as cosmic rays. For example, high energy gamma rays can be emitted from superconducting strings [73].

Several authors have calculated the emission of particles from strings and the possibility of detecting them as cosmic rays (for a review see [74]. An important feature for certain particle-string interactions is that the flux of particles

on Earth is "inversely" related to the string tension, at least for strings that are not too light. Thus lighter strings produce larger cosmic ray fluxes. The reason is simply that the density of string loops is greater if the strings are lighter, and the larger number of strings give a larger cosmic ray flux. Hence, if there are cosmic strings that emit cosmic rays, the constraints imply a "lower" bound on the string tension. Although, at very low tension, the constraint again gets weaker because then the fractional energy loss in particles is very large and this reduces the loop number density [75, 76]. Superconducting strings can also emit high energy cosmic rays with different dependencies on the string parameters [77].

Another important constraint on the cosmic string scenario arises because the particles emitted by strings generally include protons and also very high energy ($\sim 10^{20}$ eV) photons [78]. Even though the nature of the ultra-high energy cosmic rays is not clear at present - they could be protons or heavy nuclei or an admixture - it is certain that they do not include a significant photon component. With particular interactions strings may be able to source the ultra-high energy cosmic rays without conflicting with the photon bounds [79].

In the case of cosmic superstrings, radiation may include dilaton and other moduli. The case when the dilaton has gravitational-strength coupling to matter has been discussed in [80], with constraints arising from a number of different experiments and observations. In the case of large volume and warped Type-IIB compactifications, the coupling of the moduli is stronger than gravitational-strength, and the resulting constraints in the three dimensional parameter space – cosmic string tension, moduli mass, coupling strength – have been analyzed in [81]. Cosmic superstrings can also be expected to provide distinctive cosmic ray signatures via the moduli emitted from cusps.

2. Radio bursts

The gravitational coupling between photons and cosmic strings leads to the emission of light from strings [82–84]. This particular emission is generic to cosmic strings but it is suppressed by two powers of the gravitational coupling and it is unclear if it can lead to an observable signature.

Superconducting cosmic strings — strings that carry electric currents — can give transient electromagnetic signatures ("radio bursts") that are most evident at radio frequencies [85]. The event rate is dominated by kink bursts in a range of parameters that are of observational interest, and can be quite high (several a day at 1 Jy flux) for a canonical set of parameters [86]. In the absence of events, the search for radio transients can place stringent constraints on superconducting cosmic strings, though additional recently discovered cosmological radio burst candidates are compatible with the superconducting string model [87].

3. CMB spectral distortions

If the strings are superconducting, they emit electromagnetic radiation that produces μ- and y-distortions of the black body spectrum of the CMB. This will allow future CMB experiments, such as PIXIE [88], to place tight constraints on $G\mu$ and the electric current on the string [89].

Acknowledgments

We are grateful to Ana Achucarro, Christophe Ringeval, Paul Shellard, Gerard Verbiest and Proty Wu for permission to use figures from their publications, and to Tom Kibble and Alex Vilenkin for their comments and helpful feedback. TV acknowledges support by the DOE at ASU. LP is supported by the Natural Sciences and Engineering Research Council of Canada. DAS acknowledges the support of the excellence cluster/Labex ENIGMASS.

[1] T. Vachaspati, L. Pogosian, and D. A. Steer, Scholarpedia **10**, 31682 (2015).
[2] C. Ringeval, Adv.Astron. **2010**, 380507 (2010), 1005.4842.
[3] H. B. Nielsen and P. Olesen, Nucl.Phys. **B61**, 45 (1973).
[4] E. Bogomolny, Sov.J.Nucl.Phys. **24**, 449 (1976).
[5] L. Jacobs and C. Rebbi, Phys.Rev. **B19**, 4486 (1979).
[6] P. Salmi, A. Achucarro, E. Copeland, T. Kibble, R. de Putter, et al., Phys.Rev. **D77**, 041701 (2008), 0712.1204.
[7] E. Shellard, Nucl.Phys. **B283**, 624 (1987).
[8] R. Matzner, Computers in Physics **2(5)**, 51 (1988).
[9] G. Verbiest and A. Achucarro, Phys.Rev. **D84**, 105036 (2011), 1106.4666.

[10] R. Jackiw and P. Rossi, Nucl.Phys. **B190**, 681 (1981).
[11] E. Witten, Nucl.Phys. **B249**, 557 (1985).
[12] A. Achucarro and T. Vachaspati, Phys.Rept. **327**, 347 (2000), hep-ph/9904229.
[13] E. J. Copeland, L. Pogosian, and T. Vachaspati, Class.Quant.Grav. **28**, 204009 (2011), 1105.0207.
[14] M. Hindmarsh and T. Kibble, Rept.Prog.Phys. **58**, 477 (1995), hep-ph/9411342.
[15] C. Ringeval, Adv.Astron. **2010**, 380507 (2010), 1005.4842.
[16] A. Vilenkin and E. P. S. Shellard (2000).
[17] S. M. Barr and A. Matheson, Phys.Lett. **B198**, 146 (1987).
[18] P. Peter, Phys.Rev. **D45**, 1091 (1992).
[19] A. Vilenkin, Phys.Rev. **D23**, 852 (1981).
[20] E. J. Copeland, R. C. Myers, and J. Polchinski, JHEP **0406**, 013 (2004), hep-th/0312067.
[21] G. Dvali and A. Vilenkin, JCAP **0403**, 010 (2004), hep-th/0312007.
[22] E. Copeland, T. Kibble, and D. A. Steer, Phys.Rev.Lett. **97**, 021602 (2006), hep-th/0601153.
[23] A. E. Everett, Phys.Rev. **D24**, 858 (1981).
[24] R. Rohm (1985).
[25] P. d. S. Gerbert, Phys. Rev. D **40**, 1346 (1989).
[26] M. G. Alford and F. Wilczek, Phys.Rev.Lett. **62**, 1071 (1989).
[27] T. Kibble, J.Phys. **A9**, 1387 (1976).
[28] T. Vachaspati and A. Vilenkin, Phys.Rev. **D30**, 2036 (1984).
[29] T. Vachaspati and A. Vilenkin, Phys.Rev. **D31**, 3052 (1985).
[30] A. Albrecht and N. Turok, Phys.Rev. **D40**, 973 (1989).
[31] D. P. Bennett and F. R. Bouchet, Phys.Rev. **D41**, 2408 (1990).
[32] B. Allen and E. Shellard, Phys.Rev.Lett. **64**, 119 (1990).
[33] J. J. Blanco-Pillado, K. D. Olum, and B. Shlaer, Phys.Rev. **D89**, 023512 (2014), 1309.6637.
[34] L. Lorenz, C. Ringeval, and M. Sakellariadou, JCAP **1010**, 003 (2010), 1006.0931.
[35] G. Vincent, N. D. Antunes, and M. Hindmarsh, Phys.Rev.Lett. **80**, 2277 (1998), hep-ph/9708427.
[36] J. J. Blanco-Pillado, K. D. Olum, and B. Shlaer, Phys.Rev. **D83**, 083514 (2011), 1101.5173.
[37] T. Damour and A. Vilenkin, Phys.Rev.Lett. **85**, 3761 (2000), gr-qc/0004075.
[38] A. A. Fraisse, C. Ringeval, D. N. Spergel, and F. R. Bouchet, Phys.Rev. **D78**, 043535 (2008), 0708.1162.
[39] M. Landriau and E. Shellard, Phys.Rev. **D83**, 043516 (2011), 1004.2885.
[40] N. Kaiser and A. Stebbins, Nature **310**, 391 (1984).
[41] I. Gott, J. Richard, Astrophys.J. **288**, 422 (1985).
[42] F. R. Bouchet, D. P. Bennett, and A. Stebbins, Nature **335**, 410 (1988).
[43] E. Jeong and G. F. Smoot, Astrophys.J. **624**, 21 (2005), astro-ph/0406432.
[44] R. J. Danos and R. H. Brandenberger, Int.J.Mod.Phys. **D19**, 183 (2010), 0811.2004.
[45] P. Ade et al. (Planck), Astron.Astrophys. **571**, A25 (2014), 1303.5085.
[46] N. Bevis, M. Hindmarsh, M. Kunz, and J. Urrestilla, Phys.Rev. **D75**, 065015 (2007), astro-ph/0605018.
[47] L. Pogosian and T. Vachaspati, Phys.Rev. **D60**, 083504 (1999), astro-ph/9903361.
[48] L. Pogosian, http://www.sfu.ca/ levon/cmbact.html (2014).
[49] P. Ade et al. (BICEP2), Phys.Rev.Lett. **112**, 241101 (2014), 1403.3985.
[50] P. Ade et al. (POLARBEAR), Astrophys.J. **794**, 171 (2014), 1403.2369.
[51] A. Moss and L. Pogosian, Phys.Rev.Lett. **112**, 171302 (2014), 1403.6105.
[52] J. Lizarraga, J. Urrestilla, D. Daverio, M. Hindmarsh, M. Kunz, et al., Phys.Rev.Lett. **112**, 171301 (2014), 1403.4924.
[53] J. Lizarraga, J. Urrestilla, D. Daverio, M. Hindmarsh, M. Kunz, et al., Phys.Rev. **D90**, 103504 (2014), 1408.4126.
[54] A. Lazanu and P. Shellard, JCAP **1502**, 024 (2015), 1410.5046.
[55] A. Avgoustidis, E. Copeland, A. Moss, L. Pogosian, A. Pourtsidou, et al., Phys.Rev.Lett. **107**, 121301 (2011), 1105.6198.
[56] M. Hindmarsh, C. Ringeval, and T. Suyama, Phys.Rev. **D80**, 083501 (2009), 0908.0432.
[57] M. Hindmarsh, C. Ringeval, and T. Suyama, Phys.Rev. **D81**, 063505 (2010), 0911.1241.
[58] D. Regan and E. Shellard, Phys.Rev. **D82**, 063527 (2010), 0911.2491.
[59] R. Khatri and B. D. Wandelt, Phys.Rev.Lett. **100**, 091302 (2008), 0801.4406.
[60] A. Berndsen, L. Pogosian, and M. Wyman, Mon.Not.Roy.Astron.Soc. **407**, 1116 (2010), 1003.2214.
[61] R. H. Brandenberger, R. J. Danos, O. F. Hernandez, and G. P. Holder, JCAP **1012**, 028 (2010), 1006.2514.
[62] F. A. Jenet, G. Hobbs, W. van Straten, R. Manchester, M. Bailes, et al., Astrophys.J. **653**, 1571 (2006), astro-ph/0609013.
[63] P. Binetruy, A. Bohe, T. Hertog, and D. Steer, Phys.Rev. **D82**, 126007 (2010), 1009.2484.
[64] A. Vilenkin, Astrophys.J. **282**, L51 (1984).
[65] B. Shlaer and M. Wyman, Phys.Rev. **D72**, 123504 (2005), hep-th/0509177.
[66] R. Brandenberger, H. Firouzjahi, and J. Karouby, Phys.Rev. **D77**, 083502 (2008), 0710.1636.
[67] K. J. Mack, D. H. Wesley, and L. J. King, Phys.Rev. **D76**, 123515 (2007), astro-ph/0702648.
[68] K. Kuijken, X. Siemens, and T. Vachaspati, Mon.Not.Roy.Astron.Soc. **384**, 161 (2008), 0707.2971.
[69] M. Pshirkov and A. Tuntsov, Phys.Rev. **D81**, 083519 (2010), 0911.4955.
[70] A. Tuntsov and M. Pshirkov, Phys.Rev. **D81**, 063523 (2010), 1001.4580.
[71] J. K. Bloomfield and D. F. Chernoff, Phys.Rev. **D89**, 124003 (2014), 1311.7132.
[72] D. B. Thomas, C. R. Contaldi, and J. Magueijo, Phys.Rev.Lett. **103**, 181301 (2009), 0909.2866.
[73] A. Vilenkin and T. Vachaspati, Phys.Rev.Lett. **58**, 1041 (1987).

[74] P. Bhattacharjee and G. Sigl, Phys.Rept. **327**, 109 (2000), astro-ph/9811011.
[75] P. Bhattacharjee, Phys.Rev. **D40**, 3968 (1989).
[76] A. J. Long and T. Vachaspati, JCAP **1412**, 040 (2014), 1409.6979.
[77] V. Berezinsky, K. D. Olum, E. Sabancilar, and A. Vilenkin, Phys.Rev. **D80**, 023014 (2009), 0901.0527.
[78] F. Aharonian, P. Bhattacharjee, and D. Schramm, Phys.Rev. **D46**, 4188 (1992).
[79] T. Vachaspati, Phys.Rev. **D81**, 043531 (2010), 0911.2655.
[80] T. Damour and A. Vilenkin, Phys.Rev.Lett. **78**, 2288 (1997), gr-qc/9610005.
[81] E. Sabancilar, Phys.Rev. **D81**, 123502 (2010), 0910.5544.
[82] J. Garriga, D. Harari, and E. Verdaguer, Nucl.Phys. **B339**, 560 (1990).
[83] K. Jones-Smith, H. Mathur, and T. Vachaspati, Phys.Rev. **D81**, 043503 (2010), 0911.0682.
[84] D. A. Steer and T. Vachaspati, Phys.Rev. **D83**, 043528 (2011), 1012.1998.
[85] T. Vachaspati, Phys.Rev.Lett. **101**, 141301 (2008), 0802.0711.
[86] Y.-F. Cai, E. Sabancilar, D. A. Steer, and T. Vachaspati, Phys.Rev. **D86**, 043521 (2012), 1205.3170.
[87] Y.-W. Yu, K.-S. Cheng, G. Shiu, and H. Tye, JCAP **1411**, 040 (2014), 1409.5516.
[88] A. Kogut, D. J. Fixsen, D. T. Chuss, J. Dotson, E. Dwek, M. Halpern, G. F. Hinshaw, S. M. Meyer, S. H. Moseley, M. D. Seiffert, et al., JCAP **7**, 025 (2011), 1105.2044.
[89] H. Tashiro, E. Sabancilar, and T. Vachaspati, Phys.Rev. **D85**, 103522 (2012), 1202.2474.

Monopole-Antimonopole Scattering

Tanmay Vachaspati
Physics Department, Arizona State University, Tempe, AZ 85287, USA.

We numerically study the head-on scattering of a 't Hooft-Polyakov magnetic monopole and antimonopole for a wide range of parameters. In contrast to the scattering of a $\lambda\phi^4$ kink and antikink in 1+1 dimensions, we find that the monopole and antimonopole annihilate even when scattered at relativistic velocities. If the monopole and antimonopole have a relative twist, there is a repulsive force between them and they can initially be reflected. However, in every case we have examined, the reflected monopoles remain bound and eventually annihilate. We also calculate the magnetic helicity in the aftermath of monopole-antimonopole annihilation and confirm the conversion of relative twist to magnetic helicity as discussed earlier in the electroweak case.

Beautiful results have been obtained on the scattering of monopoles on monopoles [1–3]. For example, analytical techniques show that head-on collision leads to 90° scattering for a certain value of the coupling constant (in the so-called Bogomolny-Prasad-Sommerfield (BPS) limit) [2]. Monopole-*antimonopole* scattering, though, has received less attention, perhaps because the process is less amenable to analysis.

A general expectation is that monopole-antimonopole ($\mathrm{M\overline{M}}$) scattering will lead to their annihilation and the energy will be dissipated in the form of radiation. However this is not the result obtained in the analogous process of Z_2 kink-antikink scattering in 1+1 dimensions. Numerical studies of kink-antikink scattering show annihilation at low kinetic energy, reflection at higher incoming energy, followed by annihilation at yet higher energies, *etc.*, yielding a band structure reminiscent of solutions of the Mathieu equation [4, 5]. One motivation for the present work is to check for chaotic behavior in $\mathrm{M\overline{M}}$ scattering.

A second motivation for studying $\mathrm{M\overline{M}}$ scattering comes from the recent interest in the possible existence and detection of a helical inter-galactic magnetic field [6]. Early work had speculated on the production of magnetic fields during monopole-antimonopole annihilation [7, 8]. The connection to baryogenesis was made when the electroweak sphaleron solution that mediates baryon number violation was interpreted in terms of electroweak $\mathrm{M\overline{M}}$ pairs [9, 10]. It is crucial for this connection that monopole-antimonopole pairs can have a relative "twist" and the (unstable) electroweak sphaleron solution is really an $\mathrm{M\overline{M}}$ pair that is prevented from annihilating by the presence of a twist. In sphaleron decay, the twist is believed to be the reason that the resultant magnetic field has non-zero helicity, h, defined by

$$h = \int d^3x \ \boldsymbol{A}\cdot\boldsymbol{B} \tag{1}$$

where $\boldsymbol{A}$ is the electromagnetic gauge potential and $\boldsymbol{B}$ is the magnetic field. In this paper, we will also study the scattering of twisted $\mathrm{M\overline{M}}$ pairs and confirm that magnetic field helicity originates in the relative twist of the $\mathrm{M\overline{M}}$.

The results of our investigations are easily summarized: numerical evolution for a wide range of $\mathrm{M\overline{M}}$ initial conditions show that untwisted $\mathrm{M\overline{M}}$ scattering always leads to annihilation. Thus we do not see any evidence for chaotic behavior similar to that seen in 1+1 dimensions. However, when the monopoles are initially twisted, there is a repulsive force between the monopoles. At low velocities, the monopoles slow down or may even reflect back. Yet this reflection is temporary and soon reversed, and the $\mathrm{M\overline{M}}$ then annihilate. Further, the annihilation of twisted $\mathrm{M\overline{M}}$ results in the production of a helical magnetic field.

We start out in Sec. I by defining the field theory, describing the magnetic monopoles and the twisted $\mathrm{M\overline{M}}$ ansatz in which the monopoles are also Lorentz boosted. The $\mathrm{M\overline{M}}$ field ansatz will form the initial conditions for the numerical evolution described in Sec. II, where we also show sample plots of the scattering, the trajectories of the $\mathrm{M\overline{M}}$, and the magnetic helicity generated during annihilation. We conclude in Sec. III.

I. SO(3) MODEL, MONOPOLES, AND $\mathrm{M\overline{M}}$ ANSATZ

A. SO(3)model

The model we study contains an SO(3)adjoint scalar and gauge field, $\{\phi^a, W^a_\mu\}$ $(a = 1, 2, 3)$ with the Lagrangian

$$L = \frac{1}{2}(D_\mu\phi)^a(D^\mu\phi)^a - \frac{1}{4}W^a_{\mu\nu}W^{a\mu\nu} - \frac{\lambda}{4}(\phi^a\phi^a - \eta^2)^2 \tag{2}$$

where,

$$(D_\mu\phi)^a = \partial_\mu\phi^a - igW^c_\mu(T^c)^{ab}\phi^b \tag{3}$$

and the SO(3)generators are $(T^a)^{bc} = -i\epsilon^{abc}$. The gauge field strengths are defined by

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^{abc}W^b_\mu W^c_\nu. \tag{4}$$

The scalar field equations of motion are

$$\begin{aligned}\partial_t^2\phi^a &= \partial_i\partial_i\phi^a + igW^{\mu c}(T^c)^{ab}\partial_\mu\phi^b \\ &\quad + igW^{\mu c}(T^c)^{ab}(D_\mu\phi)^b - \lambda(\phi^b\phi^b - \eta^2)\phi^a. \end{aligned}\tag{5}$$

We will work in the Lorenz gauge given by the equation

$$\partial_t W_0^a = \partial_i W_i^a \tag{6}$$

and then the gauge field equations are

$$\begin{aligned}\partial_t^2 W_\mu^a &= \partial_i\partial_i W_\mu^a - g\epsilon^{abc}W^{\nu b}\partial_\nu W_\mu^c - g\epsilon^{abc}W_\nu^b W^{\nu c}{}_\mu \\ &\quad -g\epsilon^{abc}\phi^b(D_\mu\phi)^c. \end{aligned}\tag{7}$$

By rescaling the coordinates and the fields, as we shall do from now on, we can set $g = 1$ and $\eta = 1$. Then λ is the only free parameter left in the model. The BPS case is when $\lambda = 0$. We will numerically evolve the 15 second order partial differential equations (PDE) in (5) and (7).

The energy density for the model is given by

$$\begin{aligned}\mathcal{E} &= \frac{1}{2}(D_t\phi)^a(D_t\phi)^a + \frac{1}{2}(D_i\phi)^a(D_i\phi)^a \\ &\quad +\frac{1}{2}(W_{0i}^a W_{0i}^a + W_{ij}^a W_{ij}^a) + \frac{\lambda}{4}(\phi^a\phi^a - 1)^2 \end{aligned}\tag{8}$$

where the sum over the repeated index j is restricted to $j > i$.

Once ϕ^a acquires its vacuum expectation value, the model contains two massive gauge fields and one massless gauge field. The massless gauge field is

$$A_\mu = n^a W_\mu^a \tag{9}$$

where $n^a \equiv \phi^a/\sqrt{\phi^b\phi^b}$ is a unit vector at all spatial points. The field strength corresponding to the gauge field A_μ is defined as [11]

$$\begin{aligned}A_{\mu\nu} &= n^a W_{\mu\nu}^a - \epsilon^{abc}n^a(D_\mu n)^b(D_\nu n)^c \\ &= \partial_\mu A_\nu - \partial_\nu A_\mu - \epsilon^{abc}n^a\partial_\mu n^b\partial_\nu n^c. \end{aligned}\tag{10}$$

This field strength definition correspond to the usual Maxwell electric and magnetic field only when the magnitude $|\phi|$ is constant. We shall apply them at late times after the monopoles have annihilated and when $|\phi|$ is approximately constant and non-zero everywhere.

B. Monopoles

The monopole solution takes the form

$$\phi^a = P(r)\hat{x}^a \tag{11}$$

$$W_i^a = \frac{(1-K(r))}{r}\epsilon^{aij}\hat{x}^j \tag{12}$$

where $\hat{x} = \boldsymbol{x}/r$ and r is the (rescaled) spherical radial distance centered on the monopole. The profile functions $P(r)$, $K(r)$ are not known in closed form except in the BPS ($\lambda = 0$) case [12, 13]

$$P_{\rm BPS}(r) = \frac{1}{\tanh(r)} - \frac{1}{r}, \tag{13}$$

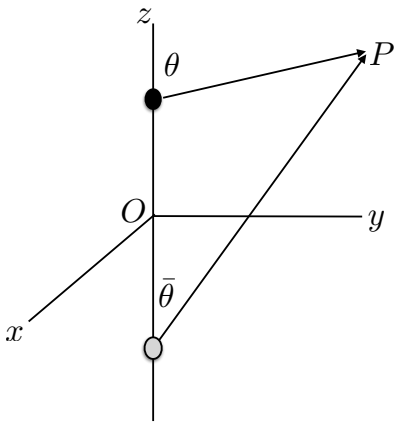

FIG. 1: Monopole and antimonopole are chosen to be on the $z-$axis with some initial separation $2z_0$. The spherical angles θ and $\bar{\theta}$ are defined as shown.

$$K_{\rm BPS}(r) = \frac{r}{\sinh(r)}. \tag{14}$$

We will be studying the evolution of monopoles for a range of λ. A functional form that reduces to the BPS profile functions for $\lambda = 0$ and has the correct asymptotic properties is

$$P(r) = \frac{1}{\tanh(r)} - (1+mr)\frac{e^{-mr}}{r} \tag{15}$$

$$K(r) = \frac{r}{\sinh(r)} \tag{16}$$

where $m = \sqrt{2\lambda}$ is the scalar particle mass (in $\eta = 1$ units).

Next we will need to patch together a monopole and an antimonopole, with a relative twist, and also boost the monopole and antimonopole towards each other.

C. M$\overline{\rm M}$ Ansatz

A twisted monopole-antimonopole ansatz is known in the context of the electroweak model where the scalar field is an SU(2) doublet [9]. The form is

$$\Phi = \begin{pmatrix} \sin(\theta/2)\sin(\bar{\theta}/2)e^{i\gamma} + \cos(\theta/2)\cos(\bar{\theta}/2) \\ \sin(\theta/2)\cos(\bar{\theta}/2)e^{i\varphi} - \cos(\theta/2)\sin(\bar{\theta}/2)e^{i(\varphi-\gamma)} \end{pmatrix} \tag{17}$$

where θ and $\bar{\theta}$ are the spherical angles centered on the monopole and antimonopole respectively (see Fig. 1), φ is the azimuthal angle, and γ is the twist. A little algebra shows that $\Phi^\dagger\Phi = 1$.

From Φ, we construct the corresponding unit vector field n^a using

$$n^a = \Phi^\dagger\sigma^a\Phi \tag{18}$$

where σ^a are the Pauli spin matrices. The result, with the replacement $\varphi \to \varphi - \gamma/2$ to make the expressions more symmetrical, is

$$\begin{aligned} n^1 &= (\sin\theta\cos\bar{\theta}\cos\gamma - \sin\bar{\theta}\cos\theta)\cos(\varphi-\gamma/2) \\ &\quad - \sin\theta\sin\gamma\sin(\varphi-\gamma/2) \qquad (19) \\ n^2 &= (\sin\theta\cos\bar{\theta}\cos\gamma - \sin\bar{\theta}\cos\theta)\sin(\varphi-\gamma/2) \\ &\quad + \sin\theta\sin\gamma\cos(\varphi-\gamma/2) \qquad (20) \\ n^3 &= \cos\theta\cos\bar{\theta} + \sin\theta\sin\bar{\theta}\cos\gamma \qquad (21) \end{aligned}$$

Close to the monopole, we have $\bar{\theta} \to 0$ and then

$$\begin{aligned} n^1 &\to \sin\theta\cos(\varphi+\gamma/2) \qquad (22) \\ n^2 &\to \sin\theta\sin(\varphi+\gamma/2) \qquad (23) \\ n^3 &\to \cos\theta \qquad (24) \end{aligned}$$

as we would expect around a monopole. Close to the antimonopole, we have $\theta \to \pi$ and then

$$\begin{aligned} n^1 &\to \sin\bar{\theta}\cos(\varphi-\gamma/2) \qquad (25) \\ n^2 &\to \sin\bar{\theta}\sin(\varphi-\gamma/2) \qquad (26) \\ n^3 &\to -\cos\bar{\theta} \qquad (27) \end{aligned}$$

which corresponds to an antimonopole (because of the minus sign in n^3). Also note the relative twist along φ of the monopole and antimonopole.

Our ansatz has the nice feature that $\hat{n} \propto \hat{z}$ far away from the M$\overline{\text{M}}$ in all directions when the twist vanishes. To check this we set $\gamma = 0$, $\bar{\theta} \to \theta$ and obtain $\hat{n} = (0,0,1)$.

Now we are ready to write down the scalar field for a twisted monopole-antimonopole pair:

$$\phi^a(x,y,z) = P(r_m)P(r_{\bar{m}})n^a \qquad (28)$$

where $P(r)$ is the profile function in Eq. (15) and r_m, $r_{\bar{m}}$ are the distances of the spatial point (x,y,z) from the monopole and antimonopole respectively.

At this stage the monopole-antimonopole are at rest. To boost the monopole along the $-z$ direction and the antimonopole along the $+z$ direction we first re-express r_m, $r_{\bar{m}}$ and n^a in Cartesian coordinates

$$r_m = |\boldsymbol{x} - \boldsymbol{x}_m|, \quad r_{\bar{m}} = |\boldsymbol{x} - \boldsymbol{x}_{\bar{m}}| \qquad (29)$$

where $\boldsymbol{x}_m = (0,0,z_0)$ and $\boldsymbol{x}_{\bar{m}} = (0,0,-z_0)$ are the locations of the monopole and the antimonopole respectively. The unit vector n^a is also expressed in Cartesian coordinates,

$$\begin{aligned} r_m r_{\bar{m}} n^1 &= (cx+sy)[(z+z_0)\cos\gamma - (z-z_0)] \\ &\quad -(cy-sx)r_{\bar{m}}\sin\gamma \qquad (30) \\ r_m r_{\bar{m}} n^2 &= (cy-sx)[(z+z_0)\cos\gamma - (z-z_0)] \\ &\quad +(cx+sy)r_{\bar{m}}\sin\gamma \qquad (31) \\ r_m r_{\bar{m}} n^3 &= (z-z_0)(z+z_0) + (x^2+y^2)\cos\gamma \qquad (32) \end{aligned}$$

where $c \equiv \cos(\gamma/2)$, $s \equiv \sin(\gamma/2)$. Here we have been careful to distinguish the $(z \pm z_0)$ factors coming from the monopole and antimonopole, since these will be boosted differently,

$$(z \pm z_0) \to (z \pm z_0)^{(b)} = \gamma_L((z \pm z_0) \mp v_z t) \qquad (33)$$

where, $\gamma_L = (1 - v_z^2)^{-1/2}$. Note that these boosts also have to be included in r_m and $r_{\bar{m}}$. (We will denote boosted quantities by a (b) superscript.) Then the scalar fields at $t = 0$ for a boosted, twisted monopole-antimonopole pair are:

$$\phi^a(x,y,z) = \left[P(r_m^{(b)})P(r_{\bar{m}}^{(b)})n^{(b)a}\right]_{t=0} \qquad (34)$$

We also need the first time derivative (denoted by an overdot) of the scalar field at $t = 0$ and this is given by

$$\dot{\phi}^a(x,y,z) = \left[\partial_t\left(P(r_m^{(b)})P(r_{\bar{m}}^{(b)})n^{(b)a}\right)\right]_{t=0} \qquad (35)$$

The partial time derivative can be expressed in terms of spatial derivatives as discussed below.

Now that we have the initial scalar fields, we move on to specify the initial gauge fields. This is most simply done numerically using the following scheme. We fix the internal space orientation of the gauge fields by minimizing the covariant derivative. The vacuum solution of $D_\mu \hat{n} = 0$ is

$$W_\mu^a|_{\text{vacuum}} = -\epsilon^{abc}\hat{n}^b\partial_\mu\hat{n}^c \qquad (36)$$

To this we attach profile functions so that the gauge fields are well defined at the locations of the monopole and antimonopole. So

$$\begin{aligned} W_\mu^a|_{t=0} &= -\Big[(1-K(r_m^{(b)}))(1-K(r_{\bar{m}}^{(b)})) \\ &\qquad \times\epsilon^{abc}\hat{n}^{(b)b}\partial_\mu\hat{n}^{(b)c}\Big]_{t=0} \qquad (37) \end{aligned}$$

Finally we need the initial time derivative of W_μ^a. We shall treat the spatial and temporal components differently to enforce the Lorenz gauge condition. For the spatial components, as in the case of the scalar field, the time derivative is given by

$$\begin{aligned} \dot{W}_i^a|_{t=0} &= -\Big[\partial_t\Big((1-K(r_m^{(b)}))(1-K(r_{\bar{m}}^{(b)})) \\ &\qquad \times\epsilon^{abc}\hat{n}^{(b)b}\partial_i\hat{n}^{(b)c}\Big)\Big]_{t=0} \qquad (38) \end{aligned}$$

For the time component of the gauge field, we use the Lorenz gauge condition

$$\dot{W}_0^a|_{t=0} = [\partial_i W_i^a]_{t=0} \qquad (39)$$

Although the form of the initial conditions is quite involved, they are not too difficult to implement since temporal derivatives can be related to spatial derivatives using

$$[(z \pm z_0)^{(b)}]_{t=0} = \gamma_L(z \pm z_0) \qquad (40)$$

$$[\partial_t(z \pm z_0)^{(b)}]_{t=0} = \mp\gamma_L v_z \qquad (41)$$

and spatial derivatives can be evaluated numerically.

II. EVOLUTION

We discretize the 15×2 first-order equations of motion and evolve the system using the iterated Crank-Nicholson method with two iterations [14]. Our code has the novelty that all field theory specific routines are generated symbolically and are then inserted into a PDE integrating routine. We have also implemented absorbing boundary conditions by assuming that all fields only depend on $t-r$ where r is the distance from the center of the lattice. For the specific problem at hand, all the non-trivial dynamics is well within the simulation volume and the choice of boundary conditions is not crucial.

The initial energy of our ansatz for $\gamma = 0$ matches the analytic result for untwisted BPS monopoles. During the numerical evolution we have checked energy conservation at the few percent level at early times, before energy can start leaving the simulation volume. The Lorenz gauge condition is also approximately satisfied at all times in the parameter space we have investigated.

The free parameters in the model are the coupling constant λ, the boost velocity v_z, and the twist γ. The initial separation is taken to be 0.3 times the semi-lattice size plus an offset that ensures that the magnitude of ϕ does not vanish on a lattice point at the initial time. (This simplifies some of the numerics.) We have also chosen $\lambda = 1$ for our runs, and experimentation with a few other values (including $\lambda = 0$) showed similar results. The initial boost velocity v_z was varied in the interval $(0.1, 0.9)$, and the twist angle was chosen to range from 0 to 2π in steps of $\pi/4$. The only runs where we do not explicitly see annihilation until the end of the simulation is in the case when $\gamma = \pi$ and for some low values of v_z. However, even in the cases when the M$\overline{\text{M}}$ do not annihilate, they form a bound system and do not escape to infinity. In some cases, we have let the system evolve much longer and always found that the M$\overline{\text{M}}$ eventually annihilate. In Fig. 2 we show snapshots of untwisted M$\overline{\text{M}}$ and they simply come together and annihilate. In Fig. 3 we show snapshots of twisted ($\gamma = \pi$) M$\overline{\text{M}}$ at the same times as for the untwisted case and we see that they have not yet annihilated.

We plot the location of the monopole as a function of time for a few sample parameters in Figs. 4 and 5. The monopole location is defined by the location of the minimum of $\phi^a\phi^a$ over the simulation volume for $z > 0$ provided $\min[\sqrt{\phi^a\phi^a}] < 0.25$. In Fig. 4, we hold the velocity fixed at 0.5 and vary the twist from 0 to π. (The dynamics for twist of γ is the same as that for a twist of $2\pi - \gamma$.) It is clear from the plot that the twist slows down the monopole and can even cause it to bounce back. In Fig. 5 we show $z(t)$ for the monopole when the twist is held fixed at π and $v_z = 0.25, 0.50, 0.75$. Here the bounce back is very apparent. However, the monopoles are still bound after they bounce back and will eventually annihilate.

The untwisting and annihilation of the M$\overline{\text{M}}$ is expected to radiate magnetic fields that are helical [15, 16]. To test this expectation, we have calculated the helicity defined in Eq. (1) using (9) and (10). The plot of the magnetic helicity as a function of time is shown in Fig. 6 where we hold the velocity fixed at 0.75 and vary the twist. The plot shows that the helicity vanishes if there is no twist ($\gamma = 0$). Also, we see that the $h(\gamma) = -h(2\pi - \gamma)$, and the helicity vanishes in the case $\gamma = \pi$ (though in this case the M$\overline{\text{M}}$ survive until the end of the simulation). These observations can be understood if the helicity is due to the untwisting motion of the M$\overline{\text{M}}$. If $\gamma < \pi$, the M$\overline{\text{M}}$ untwist in one direction and then annihilate, while if $\gamma > \pi$, the M$\overline{\text{M}}$ untwist in the other direction so that $\gamma \to 2\pi$. The opposite senses of untwisting lead to the production of magnetic fields with opposite helicity. The value $\gamma = \pi$ is an unstable point where the M$\overline{\text{M}}$ are unable to decide which way to untwist. Eventually numerical instabilities will cause untwisting in one way or the other. In Fig. 6 we also observe oscillations in the magnetic helicity, suggesting that there may be oscillations in the twist.

In Fig. 7 we plot the magnetic helicity for $\lambda = 1$, $\gamma = 3\pi/4$ and for $v_z = 0.25, 0.50, 0.75$. The plots are similar in shape but shifted to earlier times for higher velocities. This can be understood because the M$\overline{\text{M}}$ scatter at earlier times for higher velocities.

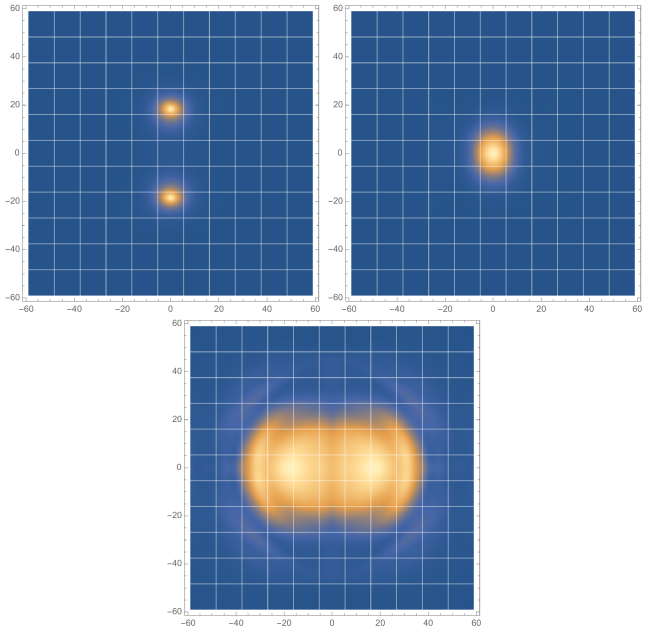

FIG. 2: Snapshots of a planar slice of annihilating monopole and antimonopole for $\lambda = 1$, $\gamma = 0$, and $v_z = 0.5$. The colors represent energy density.

III. CONCLUSIONS

We have studied M$\overline{\text{M}}$ scattering by numerical methods. Part of the challenge was to devise initial conditions that are suitable to describe boosted and twisted M$\overline{\text{M}}$. Our ansatz for initial conditions are given in Sec. I C but there

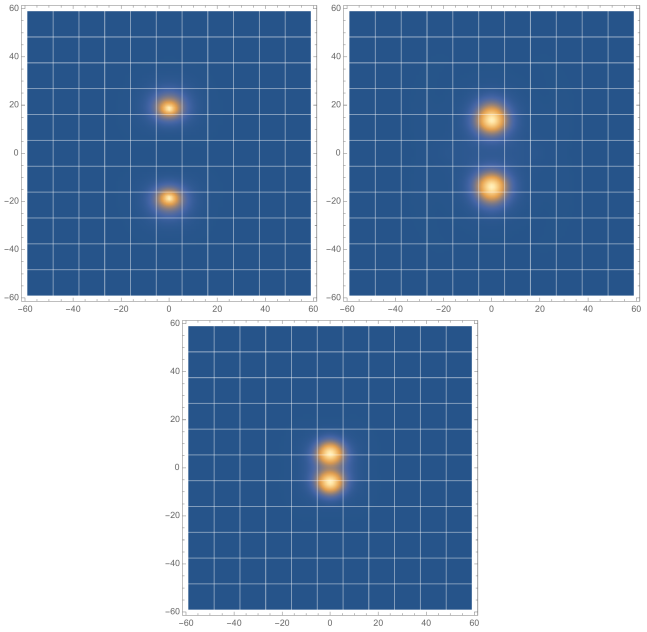

FIG. 3: Snapshots of a planar slice of non-annihilating monopole and antimonopole for $\lambda = 1$, $\gamma = \pi$, and $v_z = 0.5$. Except for the twist, all parameters, including snapshot times, are identical to those in Fig. 2. The colors represent energy density. At yet later times, the monopoles back-scatter but are still bound and return to annihilate as discussed in the text.

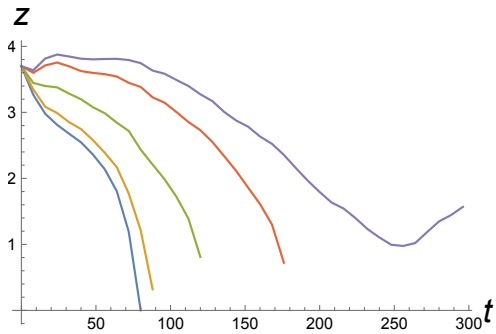

FIG. 4: The z-coordinate of the monopole as a function of time for $\lambda = 1$, $v_z = 0.50$ and $\gamma/(\pi/4) = 0, 1, 2, 3, 4$ (curves from left to right). The curves terminate once $\min[\sqrt{\phi^a\phi^a}] \geq 0.25$ (a condition that is met after the M$\overline{\text{M}}$ have annihilated) except in the $\gamma = \pi$ case, when the M$\overline{\text{M}}$ have not annihilated even by the end of the simulation run (300 time steps with $dt = dx/2 = 0.1$).

may be other choices.

The numerical evolution of M$\overline{\text{M}}$ shows that, unlike the scattering of kinks in 1+1 dimensions, M$\overline{\text{M}}$ scattering is not chaotic, as the M$\overline{\text{M}}$ are always found to annihilate over the wide range of parameters we have investigated. A twist in the initial conditions produces a repulsive force between the monopole and antimonopole that can have an important effect on the scattering dynamics. An interpretation of our results is that, as the M$\overline{\text{M}}$ approach

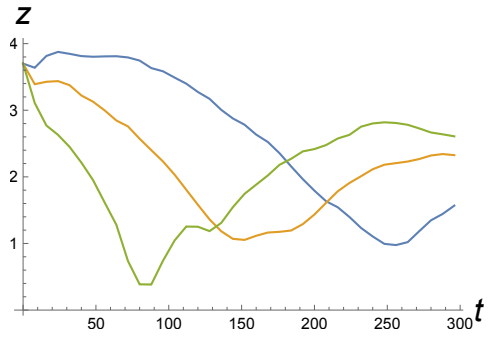

FIG. 5: The z-coordinate of the monopole as a function of time for $\lambda = 1$, $\gamma = \pi$, and $v_z = 0.25, 0.50, 0.75$ (blue, orange and green curves). The M$\overline{\text{M}}$ have not annihilated until the end of the simulation.

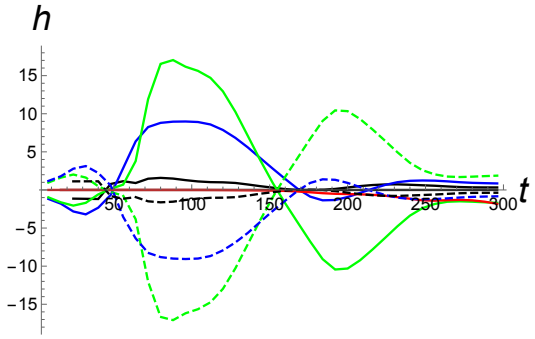

FIG. 6: Magnetic helicity in the aftermath of M$\overline{\text{M}}$ annihilation as a function of time for $\lambda = 1$, $v_z = 0.75$ and $\gamma/(\pi/4) = 0, 1, 2, 3, 4, 5, 6, 7$. The curves for $\gamma/(\pi/4) = 0, 4$ essentially coincide with $h = 0$ and are not visible. The dashed curves are for $\gamma/(\pi/4) = 5, 6, 7$ (green, blue, black), and mirror the solid curves for $\gamma/(\pi/4) = 3, 2, 1$ (green, blue, black). This shows that $h(\gamma) = -h(2\pi - \gamma)$.

each other, they also tend to untwist. The untwisting dynamics is damped due to radiation and eventually the M$\overline{\text{M}}$ can annihilate. However, damping of the untwisting dynamics leads to the production of helical magnetic fields and the sign of the magnetic helicity is related to the direction of untwisting.

Acknowledgments

I am grateful to Nick Manton and Daniele Steer for helpful comments. The computations were done on the A2C2 Saguaro Cluster at ASU. This work was supported by the U.S. Department of Energy, Office of High Energy Physics, under Award No. de-sc0013605 at ASU.

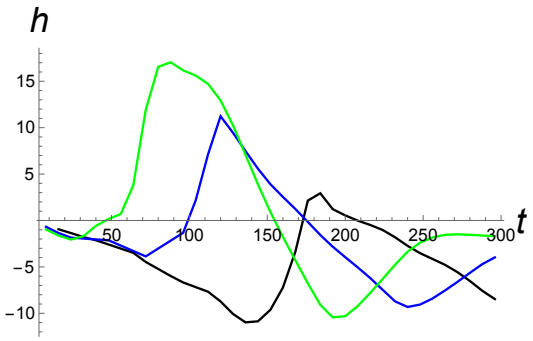

FIG. 7: Magnetic helicity in the aftermath of M$\overline{\text{M}}$ annihilation as a function of time for $\lambda = 1$, $\gamma = 3\pi/4$, and $v_z = 0.25, 0.50, 0.75$ (black, blue and green curves).

[1] N. S. Manton, Phys. Lett. B **110**, 54 (1982).

[2] "The Geometry and Dynamics of Magnetic Monopoles", M. F. Atiyah and N. J. Hitchin, Princeton University Press (1988).

[3] "Topological Solitons", N. S. Manton and P. Sutcliffe, Cambridge University Press (2004).

[4] D. K. Campbell, J. F. Schonfeld and C. A. Wingate, Physica **9D**, 1 (1983).

[5] P. Anninos, S. Oliveira and R. A. Matzner, Phys. Rev. D **44**, 1147 (1991).

[6] H. Tashiro, W. Chen, F. Ferrer and T. Vachaspati, Mon. Not. Roy. Astron. Soc. **445**, no. 1, L41 (2014) doi:10.1093/mnrasl/slu134 [arXiv:1310.4826 [astro-ph.CO]].

[7] T. Vachaspati, Phys. Lett. B **265**, 258 (1991).

[8] T. Vachaspati, in: J.C. Romão, F. Friere (Eds.), Proceedings of the NATO Workshop on "Electroweak Physics and the Early Universe", Sintra, Portugal, 1994; Series B: Physics Vol. 338, Plenum Press, New York (1994).

[9] T. Vachaspati and G. B. Field, Phys. Rev. Lett. **73**, 373 (1994) [hep-ph/9401220]; *Erratum* Phys. Rev. Lett. 74, 1258 (1995).

[10] M. Hindmarsh and M. James, Phys. Rev. D **49**, 6109 (1994) [hep-ph/9307205].

[11] G. 't Hooft, Nucl. Phys. B **79**, 276 (1974).

[12] M. K. Prasad and C. M. Sommerfield, Phys. Rev. Lett. **35**, 760 (1975).

[13] E. B. Bogomolny, Sov. J. Nucl. Phys. **24**, 449 (1976) [Yad. Fiz. **24**, 861 (1976)].

[14] S. A. Teukolsky, Phys. Rev. D **61**, 087501 (2000) [gr-qc/9909026].

[15] C. J. Copi, F. Ferrer, T. Vachaspati and A. Achúcarro, Phys. Rev. Lett. **101**, 171302 (2008) [arXiv:0801.3653 [astro-ph]].

[16] Y. Z. Chu, J. B. Dent and T. Vachaspati, Phys. Rev. D **83**, 123530 (2011) [arXiv:1105.3744 [hep-th]].

Stability of superfluid vortices in dense quark matter

Mark G. Alford, S. Kumar Mallavarapu and Andreas Windisch
Physics Department, Washington University,St. Louis, MO 63130, USA

Tanmay Vachaspati
Physics Department, Arizona State University, Tempe, AZ 85287, USA

(Dated: 16 Jan 2016)

Superfluid vortices in the color-flavor-locked (CFL) phase of dense quark matter are known to be energetically disfavored relative to well-separated triplets of "semi-superfluid" color flux tubes. However, the short-range interaction (metastable versus unstable) has not been established. In this paper we perform numerical calculations using the effective theory of the condensate field, mapping the regions in the parameter space of coupling constants where the vortices are metastable versus unstable. For the case of zero gauge coupling we analytically identify a candidate for the unstable mode, and show that it agrees well with the results of the numerical calculations. We find that in the region of the parameter space that seems likely to correspond to real-world CFL quark matter the vortices are unstable, indicating that if such matter exists in neutron star cores it is very likely to contain semi-superfluid color flux tubes rather than superfluid vortices.

PACS numbers: 12.38.-t, 25.75.Nq

I. INTRODUCTION

The densest phase of matter according to standard model physics is the color-flavor-locked (CFL) phase [1]. The CFL condensate breaks the baryon number symmetry of the theory, hence the CFL phase is a superfluid, and CFL matter in the core of a spinning neutron star will carry angular momentum in the form of superfluid vortices. Unlike the fermions in terrestrial superfluids, quarks interact via a non-Abelian gauge group; the structure of the vacuum manifold [2] is $SU(3) \times U(1)/Z_3$ and this permits the existence of a non-Abelian vortex configuration which is three times lower in energy than the usual superfluid vortex. This configuration consists of three widely-separated semi-superfluid flux tubes, each carrying color magnetic flux. At separations much larger than the size of the core any two semi-superfluid flux tubes strongly repel each other [3], and it has therefore been conjectured that CFL superfluid vortices will spontaneously decay into triplets of semi-superfluid flux tubes. However, the short-range interaction between the flux tubes has not been calculated [4], leaving open the possibility that the vortices might be metastable, in which case the decay rate, occurring via barrier penetration, could be extremely low. Such metastability has already been established for vortices in an analogous system, a three component Bose-Einstein condensate [5].

In this paper we address these unresolved questions. To probe the stability of the CFL superfluid vortices we solve the classical field equations for the CFL condensate on a two-dimensional lattice, analogously to previous calculations done for SU(2) Yang-Mills-Higgs theory [6],[7]. This approach gives us a full understanding of the decay process, far exceeding the insight that can be gained from asymptotic methods. The results presented in this paper are:

- We map the regions in the parameter space of the couplings where the superfluid vortex is metastable (Section IV).
- In the unstable regions we numerically extract the unstable mode (Section IV).
- We analytically construct an unstable mode arising in the case of zero gauge coupling (Section III), and show that it is very similar to the numerically extracted unstable mode.
- We clarify the nature of interaction of semi-superfluid vortices at short distances in the unstable region.

II. VORTICES AND FLUX TUBES

The non-Abelian Ginzburg-Landau Lagrangian used to describe semi-superfluid vortices [2] in the CFL phase of dense quark matter is given by

$$\mathcal{L} = \mathrm{Tr}\left[-\frac{1}{4}F_{ij}F^{ij} + D_i\Phi^\dagger D^i\Phi + m^2\Phi^\dagger\Phi - \lambda_2(\Phi^\dagger\Phi)^2\right] - \lambda_1(\mathrm{Tr}[\Phi^\dagger\Phi])^2 + \frac{3m^4}{4(3\lambda_1+\lambda_2)}\ , \quad (1)$$

where $D_i = \partial_i - igA_i$, $F_{ij} = \partial_i A_j - \partial_j A_i - ig\,[A_i, A_j]$. A_i represents the gluonic gauge field. We choose the normalization $\mathrm{Tr}[T^\alpha T^\beta] = \delta^{\alpha\beta}$ for the $SU(3)$ generators. In the CFL phase Φ represents the color-flavor-locked diquark condensate. It is a 3×3 complex matrix. An element of Φ may be denoted as $\phi_{\alpha a}$, where α is a color index and a is a flavor index. In the symmetry-breaking phase the field in the ground state has a non-zero vacuum expectation value,

$$A_i = 0\ , \quad \Phi = \bar{\phi}\mathbf{1}_{3\times3}\ , \quad \bar{\phi} = \sqrt{\frac{m^2}{2\lambda}}\ , \quad (2)$$

where

$$\lambda \equiv 3\lambda_1 + \lambda_2, \quad (3)$$

and the mass spectrum contains the Goldstone boson and two massive Higgs fields associated with perturbations along the singlet and adjoint directions. Their masses are $m_\phi = \sqrt{2}m$ and $m_\chi = 2\sqrt{\lambda_2}\bar{\phi}$ respectively. Stability of the ground state requires $\lambda > 0$ and $\lambda_2 > 0$. There are also massive gluons of mass $m_g = \sqrt{2}g\bar{\phi}$.

The superfluid vortex is

$$A_i = 0 \ , \quad \Phi_{\rm sf} = \bar{\phi}\,\beta(r)e^{i\theta}\,\mathbf{1}_{3\times 3} \ , \tag{4}$$

where $\beta(r)$ is the radial profile that satisfies

$$\beta'' + \frac{\beta'}{r} - \frac{\beta}{r^2} - m^2\beta(\beta^2-1) = 0 \ , \tag{5}$$

with boundary conditions

$$\begin{array}{rcl} \beta & \to & 0 \text{ as } r \to 0 \ , \\ \beta & \to & 1 \text{ as } r \to \infty \ . \end{array} \tag{6}$$

However, as noted above, the superfluid vortex is not the lowest energy configuration in the topological sector of configurations with net winding number 1 at radial infinity. The lowest energy configuration consists of three (red, green, and blue) semisuperfluid color flux tubes, each with global winding 1/3. A red semisuperfluid flux tube solution is

$$\Phi_{\rm ssft}(r,\theta) = \bar{\phi}\begin{pmatrix} f(r)\mathrm{e}^{i\theta} & 0 & 0 \\ 0 & g(r) & 0 \\ 0 & 0 & g(r) \end{pmatrix}, \tag{7}$$

$$A_\theta^{\rm ssft}(r) = -\frac{1}{gr}(1-h(r))\begin{pmatrix} -\frac{2}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}, \tag{8}$$

$$A_r^{\rm ssft} = 0 \ . \tag{9}$$

Green and blue flux tubes are obtained by permuting the diagonal elements. The profile functions $f(r)$, $g(r)$ and $h(r)$ obey

$$\begin{aligned} &f'' + \frac{f'}{r} - \frac{(2h+1)^2}{9r^2}f - \frac{m_\phi^2}{6}f\left(f^2+2g^2-3\right) \\ &\quad - \frac{m_\chi^2}{3}f(f^2-g^2) = 0 \ , \end{aligned} \tag{10}$$

$$\begin{aligned} &g'' + \frac{g'}{r} - \frac{(h-1)^2}{9r^2}g - \frac{m_\phi^2}{6}g\left(f^2+2g^2-3\right) \\ &\quad + \frac{m_\chi^2}{6}g(f^2-g^2) = 0 \ , \end{aligned} \tag{11}$$

$$h'' - \frac{h'}{r} - \frac{m_G^2}{3}\left(g^2(h-1)+f^2(2h+1)\right) = 0 \ , \tag{12}$$

with boundary conditions

$$f \to 0, \ g' \to 0, \ h \to 1 \text{ as } r \to 0, \tag{13}$$

$$f \to 1, \ g \to 1, \ h \to 0 \text{ as } r \to \infty. \tag{14}$$

To understand why a configuration of three well-separated semisuperfluid flux tubes has lower energy than a single superfluid vortex, compare the energy densities far from the core:

$$\begin{array}{rcl} \epsilon_{\rm sf} & = & 3\bar{\phi}^2/r^2, \\ \epsilon_{\rm ssft} & = & \frac{1}{3}\bar{\phi}^2/r^2. \end{array} \tag{15}$$

The energy density arises entirely from the scalar field gradient, which for each component is proportional to n^2/r^2, where n is the net winding of the field. In the superfluid vortex there is net winding of 1 in each of the three diagonal components, whereas in the semisuperfluid flux tube there is net winding of 1/3 in each component. So the energy density of a single semisuperfluid flux tube is 1/9 of the energy density of a superfluid vortex. This leads to a repulsive force between the flux tubes, since the further apart they are, the more of space is filled with the energetically cheaper semisuperfluid field configuration, as opposed to the energetically costlier superfluid vortex field configuration. To estimate the leading term in the resultant potential

$$V(l) = E_{3\,\rm ssft} - E_{\rm sf} \ , \tag{16}$$

we note that the energy density of the three semisuperfluid vortices can be approximated as being the same as a superfluid vortex at $r \gg l$, and being a superposition of the three individual energy densities at $r \lesssim l$. If we assume that the cores of the flux tubes have radius ρ, and neglect the core contributions (which are independent of l), we find

$$\begin{aligned} V(l) &\approx 2\pi\int_\rho^l r\mathrm{d}r\,(3\epsilon_{\rm ssft} - \epsilon_{\rm sf}) \\ &= -4\pi\bar{\phi}^2\ln(l/\rho) \ . \end{aligned} \tag{17}$$

For large separation l, there is a strong repulsive force decaying as $1/l$. This justifies our neglect of contributions that would be subleading in l, such as the core energies.

III. STABILITY ANALYSIS OF THE SUPERFLUID VORTEX

The Ginzburg-Landau energy functional for a two-dimensional static field configuration is

$$E = \int d^2x\,\mathcal{H}, \tag{18}$$

with the Hamiltonian density

$$\begin{aligned} \mathcal{H} = \mathrm{Tr}\left[\frac{1}{4}F_{ij}F^{ij} + D_i\Phi^\dagger D^i\Phi - m^2\Phi^\dagger\Phi + \lambda_2(\Phi^\dagger\Phi)^2\right] \\ +\lambda_1(\mathrm{Tr}[\Phi^\dagger\Phi])^2 + \frac{3m^4}{4\lambda}, \end{aligned} \tag{19}$$

where λ has been defined in equation (3). We define the spatial behavior of the superfluid vortex (4) by

$$\psi(r,\theta) \equiv \bar{\phi}\beta(r)\mathrm{e}^{i\theta} \ . \tag{20}$$

Up to an additive constant, the energy density of the vortex is

$$\mathcal{H}(\Phi_{\rm sf}) = 3\left(|\partial_i\psi|^2 - m^2|\psi|^2 + \lambda|\psi|^4\right) \ . \qquad (21)$$

To investigate the stability of the vortex, consider a perturbation $\delta\Phi$ which only affects the quark condensate, leaving the gauge field unperturbed,

$$\delta\Phi = \frac{f_0}{\sqrt{3}}\mathbf{1}_{3\times3} + f_\alpha T_\alpha \ . \qquad (22)$$

(We use the normalization $\mathrm{Tr}(T_\alpha T_\beta) = \delta_{\alpha\beta}$ for the $SU(3)$ generators.) To second order, the change in energy density due to the perturbation is given by,

$$\begin{aligned}
&\delta\mathcal{H} \equiv \mathcal{H}(\Phi_{\rm sf}+\delta\Phi) - \mathcal{H}(\Phi_{\rm sf}) = \delta\mathcal{H}_0 + \sum_{\alpha=1}^{8}\delta\mathcal{H}_\alpha \\
&= (|\partial_i f_0|^2 + \lambda(\psi f_0^* + f_0\psi^*)^2 + |f_0|^2(-m^2+2\lambda|\psi|^2)) + \\
&\sum_{\alpha=1}^{8}|\partial_i f_\alpha|^2 + \lambda_2(\psi f_\alpha^* + f_\alpha\psi^*)^2 + |f_\alpha|^2(-m^2+2\lambda|\psi|^2) \\
&+\mathcal{O}(f_{\{0,\alpha\}}^3) \ . \qquad (23)
\end{aligned}$$

Focusing on perturbations in the T_8 color direction, and decomposing $f_8 = f_{8R} + if_{8I}$, we can write

$$\begin{aligned}
\delta\mathcal{H}_8 &= (f_{8R} \quad f_{8I})\begin{pmatrix}\Omega_{RR} & \Omega_{RI}\\ \Omega_{RI} & \Omega_{II}\end{pmatrix}\begin{pmatrix}f_{8R}\\ f_{8I}\end{pmatrix}, \\
\Omega_{RR} &= -\nabla^2 + m^2\left(2\frac{\lambda_2}{\lambda}\beta^2\cos^2\theta + \beta^2 - 1\right), \\
\Omega_{II} &= -\nabla^2 + m^2\left(2\frac{\lambda_2}{\lambda}\beta^2\sin^2\theta + \beta^2 - 1\right), \\
\Omega_{RI} &= \frac{\lambda_2}{\lambda}m^2\beta^2\sin 2\theta.
\end{aligned} \qquad (24)$$

It can be shown that, for $\lambda_1 > 0$ (i.e. $\lambda > \lambda_2$) the vortex is unstable to a perturbation of the form

$$\delta\Phi^{(8)} = \epsilon\,\hat{n}\cdot\nabla\psi(r,\theta)\,T_8 \ , \qquad (25)$$

which corresponds to a translation of the red and green components of the vortex a small distance ϵ in the $\hat{n}$ direction, and translation of the blue component a distance 2ϵ in the opposite direction.

Using (24) we obtain the energy of the perturbation to order ϵ^2,

$$\delta E_8 = \epsilon^2(\lambda_2 - \lambda)\frac{\pi m^4}{\lambda^2}\int_0^\infty r\mathrm{d}r\beta'^2\beta^2 \ . \qquad (26)$$

This is the main result of this section. We see that if $\lambda > \lambda_2$ (i.e. $\lambda_1 > 0$) then the perturbation (25) lowers the energy of the vortex. At this point this is just a guess: there might be a lower-energy perturbation that involves the gauge field or has a different spatial profile or color structure. However, we will see in Sec. IV that the numerically obtained unstable mode matches (23) very closely.

IV. NUMERICAL RESULTS

To analyze the instability of the superfluid vortex and map the unstable/metastable boundary in the parameter space spanned by the three couplings g, λ, and λ_2, we solve the classical field equations on a two-dimensional lattice. For details see Appendix A. For all numerical calculations we chose the mass scale to be $m^2 = 0.25$ and use a lattice spacing $a = 1$. This provides an adequate resolution for the superfluid vortex solution on the lattice for all parameter values that we studied, since the size of the vortex depends only on m^2, not on any of the couplings.

We use the ket $|\Phi\rangle$ as a convenient notation for the actual lattice configurations of matrices in 2-dimensional position space. We define the following inner product and norm

$$\begin{aligned}
\langle A, B\rangle &\equiv \int d^2x\,\mathrm{Tr}\{B^\dagger A\} \ , \\
||A|| &\equiv \sqrt{\langle A, A\rangle} \ .
\end{aligned} \qquad (27)$$

To characterize the degree to which field configurations resemble each other, we introduce the notion of an angle ϑ between two lattice configurations A and B,

$$\vartheta \equiv \arccos\frac{|\langle A, B\rangle|}{||A||\ ||B||} \ , \qquad (28)$$

so when $\vartheta = 0$ the configurations are the same up to an overall multiplicative factor.

A. Numerical analysis of the unstable mode

In parameter regions where the superfluid vortex is unstable, the vortex solution has an unstable mode, and spontaneously decays to three semisuperfluid flux tubes that repel each other. Snapshots of this process are shown in Figure 1 where we plot the total energy density of the system along the vertical axis. The instability arises from a direction in the high-dimensional configuration space of perturbations to the superfluid vortex along which the potential possesses a negative curvature. This direction is the unstable mode $|\delta\Phi^{(u)}\rangle$, and its time evolution is

$$\begin{aligned}
\partial_{tt}|\delta\Phi^{(u)}\rangle &= \gamma^2|\delta\Phi^{(u)}\rangle \ , \\
|\delta\Phi^{(u)}\rangle_t &= \mathrm{e}^{\gamma t}|\delta\Phi^{(u)}\rangle_{t=0} \ .
\end{aligned} \qquad (29)$$

In parameter regions where the vortex is metastable, there is no such unstable direction, and all perturbations oscillate but remain small.

To find out whether, for a given set of parameter values g, λ, and λ_2, the superfluid vortex is unstable or metastable, we proceed as follows.

We first generate a lattice field configuration $|\Phi_{\rm sf}\rangle$ that is the exact superfluid vortex solution of the lattice field equations. We do this by transferring the continuum vortex solution (with radial profile obeying Eq. (5)) to the

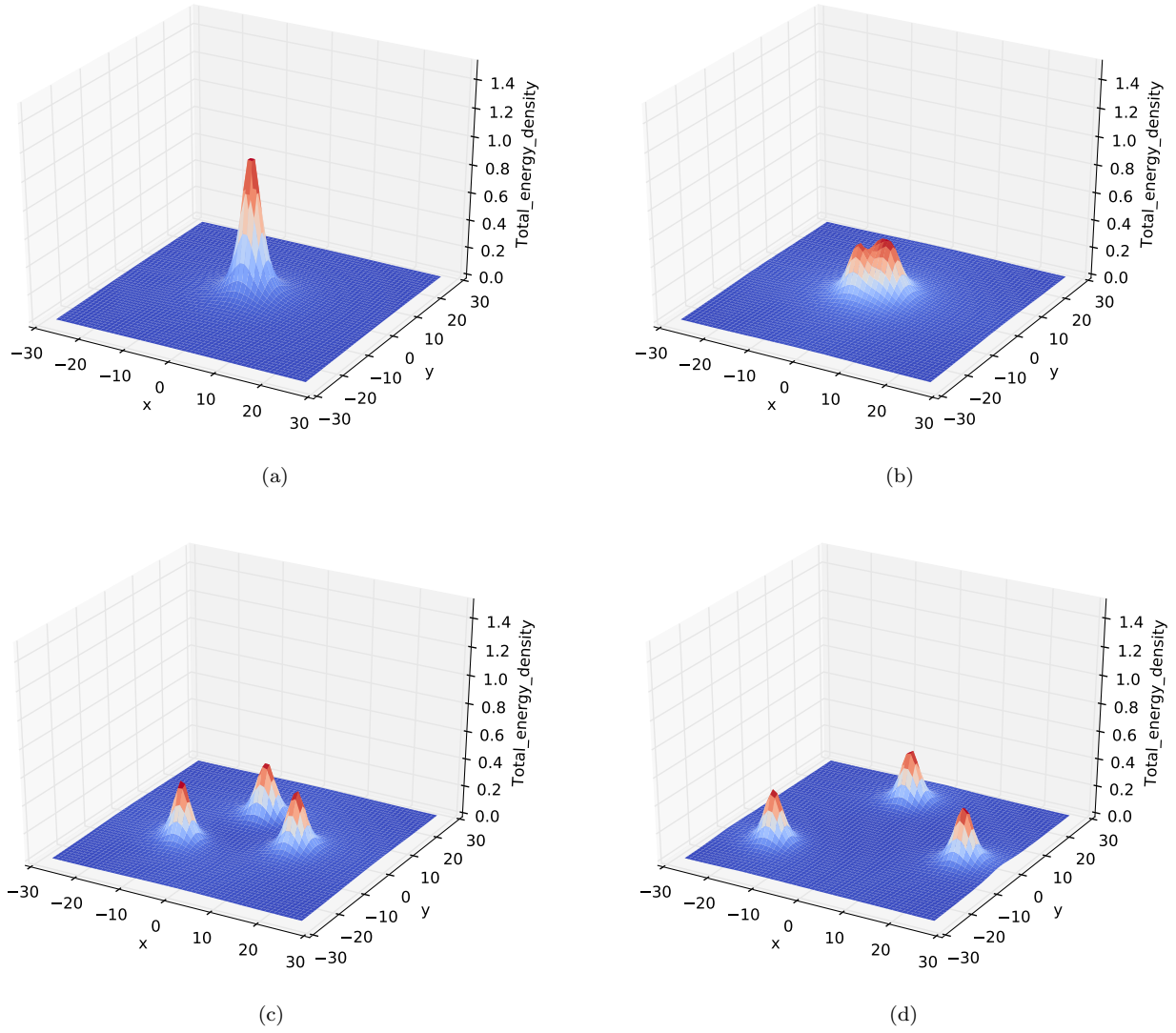

FIG. 1. Plots of the total energy density in position space for: (a) a superfluid vortex before decay; (b) fission of a superfluid vortex; (c) formation of well separated semi-superfluid vortices; (d) semi-superfluid vortices repelling each other.

lattice and then allowing it to relax to the lowest energy state by evolving it using the Langevin equation with damping but no noise. We can do this even when the superfluid vortex is unstable because the initial approximate vortex configuration is proportional to the unit matrix in the color-flavor space of complex 3×3 matrices, and the color-flavor symmetry of the Lagrangian guarantees that under time evolution it will remain proportional to the unit matrix, whereas decay would require the generation of non-singlet color components.

To probe the stability of the equilibrated vortex we add a small perturbation $|\delta\Phi^{(p)}\rangle$ to the superfluid vortex, and evolve it forward in time. The exact form of the perturbation is not important, as long as it has some overlap with the unstable mode (if any). If the vortex is unstable then even a tiny initial perturbation with some component along the unstable direction will grow exponentially as we evolve forward in time. It quickly dominates the other components of the initial perturbation. We tried three different perturbations for the Φ-field (initial gauge links put to unity without perturbation):

(a) a random configuration
Here we used 8 complex random numbers (uniformly distributed between zero and 10^{-16}) as entries for the 3×3 Φ-matrices at each lattice site.

(b) the analytically obtained unstable mode for $g = 0$
This mode is constructed using (25), which is controlled by two parameters, the direction $\hat{n}$ and magnitude of the displacement, ϵ. Using the set $\{\mathbf{1}_{3\times3}, T_\alpha\}$ as a complete basis, the radial profile serves as the position dependent coefficient of one or more of the basis

elements $\{T_\alpha\}$, since $\mathbf{1}_{3\times3}$ is a stable direction. We usually used T_8 only, but one can equally well choose any other basis element or combinations thereof.

(c) the numerically obtained unstable mode (see below).

During an initial transient period $t_{\rm min}$ (which lasts longest for the random initial perturbation) the unstable mode grows to become the dominant component. After $t_{\rm min}$ the overlap of the growing mode with the original perturbation grows exponentially in time, following (29),

$$A(t) \equiv \langle \delta\Phi^{(p)} | \Big(|\Phi\rangle_t - |\Phi_{\rm sf}\rangle \Big) \propto {\rm e}^{\gamma t}, \quad t_{\rm min} < t < t_{\rm max} \ . \tag{30}$$

The exponential growth ends after time $t_{\rm max}$ when the amplitude is so large that nonlinearities become non-negligible. Using (30) we can measure the growth rate γ without knowing the actual form of the unstable mode. If such an exponential growth is observed, the superfluid vortex is unstable. The growth rate is determined by the negative curvature of the energy in the unstable direction, and is independent of the initial "seed" perturbation.

For an unstable vortex we can numerically obtain the unstable mode up to an overall normalization factor, by computing

$$|\delta\Phi^{(u)}\rangle \approx |\Phi\rangle_{t+\Delta t} - |\Phi\rangle_t \ . \tag{31}$$

where $t_{\rm min} < t < t_{\rm max}$ and similarly for $t + \Delta t$. Once we know the unstable mode $|\delta\Phi^{(u)}\rangle$, we can go back and repeat the procedure described above using that mode as the initial seed perturbation, in which case the initial transient time $t_{\rm min}$ is very short and exponential growth starts immediately.

During the epoch of exponential growth we use (28) to compute the angle θ between the growing mode and the initial perturbation. When using the numerically constructed unstable mode as a seed we find that θ is very small, typically in the range 10^{-2} to 2 degrees. When using an unstable mode that has been constructed analytically by Eq. (25), θ is typically of the order of 10^{-3} degrees. Thus, the analytic mode discussed in Section (III) seems to be identical to the unstable mode, at least as far as the $g = 0$ case is concerned. Allowing for (larger) nonzero values of the gauge-coupling changes the picture, as indicated by Figure 4. The transition line between unstable and metastable configurations is not parallel to the g axis, suggesting that the gauge field plays a more prominent role in the decay process as g grows larger. The angle in the case of the random mode is slightly larger as compared to the case where we used the analytic mode to perturb the system. This discrepancy is most likely a numerical artifact, as we are operating with very small numbers to isolate and extract the unstable mode in this very high-dimensional space.

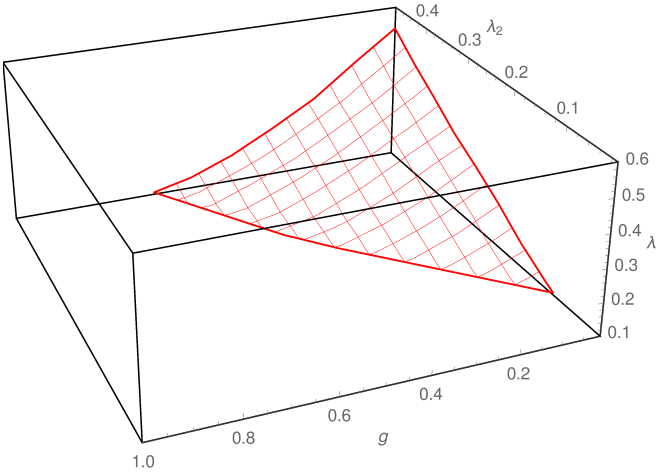

FIG. 2. The critical surface in the parameter space of the couplings g, λ and λ_2. The surface separates the metastable from the unstable regime, where the latter corresponds to the larger volume shown in the plot. An exploration in λ-direction up to a value of $\lambda = 6.0$ revealed nothing but unstable points.

B. Parameter space scan

Using the procedure described above to determine whether the vortices for a gives set of couplings (g, λ, λ_2) are unstable or metastable, we performed a scan of the parameter space for $g \in [0.01, 1]$, $\lambda \in [0.1, 6]$, $\lambda_2 \in [0.01, 0.5]$. The translation of these couplings to physical values is discussed in Section IV C below. Our main findings are summarized in Figures 2 and 3.

The critical surface separating the metastable from the unstable regime is depicted in Figure 2. We see that only a small subspace of the surveyed parameter space yields metastable superfluid vortices. The plot suggests that at higher values of the gauge coupling ($g \gtrsim 1$), which is the relevant region for QCD, the superfluid vortices are unstable, spontaneously decaying in a short time. Larger values of the coupling λ also yield unstable vortices.

In Figure 3 we sliced the three-dimensional parameter space along the physically most relevant coupling among the three, the gauge coupling g. With increasing g, the area of metastable solutions decreases rapidly. The dotted line in the plot corresponds to $\lambda = \lambda_2$, which is the line along which the analytic analysis of Section III predicts the stable/unstable transition. For small values of g, this seems to be approximately true. As the gauge coupling increases, however, the transition boundary deviates more and more from the predicted line, which is most likely due to the gauge field playing an increasingly important role in the decay process.

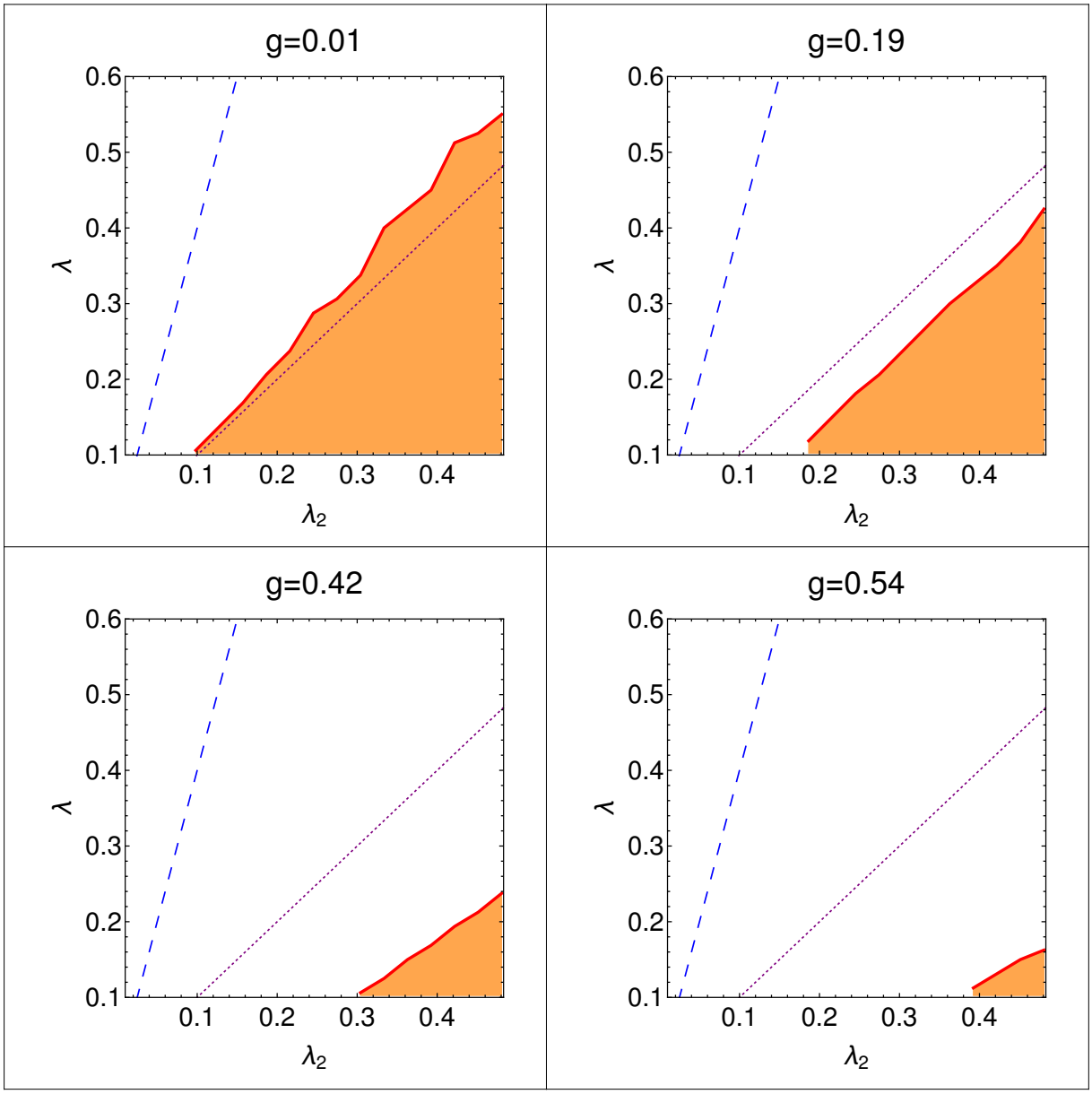

FIG. 3. Stability of the superfluid vortex for different values of g, λ and λ_2. White areas indicate unstable sectors, while shaded areas correspond to regions where the superfluid vortex has been found to be stable. The dashed line corresponds to the line $\lambda = 4\lambda_2$, along which the equations that relate the couplings to physical values of μ and T_C are valid, see discussion in Section IV C. According to the stability analysis discussed in Section III, the transition region between stable and unstable solutions should correspond to the dotted line $\lambda = \lambda_2$ in the case of vanishing coupling g. This prediction seems to hold approximately in the case of small g.

C. Relating stable and unstable regions to ratio of masses

In Ref. [8] it was suggested that the short range interaction between semi-superfluid flux tubes, and hence the unstable/metastable boundary for the superfluid vortex, is dictated by the hierarchy of the masses m_ϕ, m_χ and m_g (Section II). In Figure 4 we investigate this proposal by choosing one slice of the parameter space along the plane $\lambda = 0.1$ and plotting the different mass hierarchies along with the unstable/metastable boundary. The vertical line corresponds to the line along which $\lambda = \lambda_2$. We could not find a direct agreement between the unstable/metastable boundary and any of the boundaries derived from arranging the masses according to their mass hierarchy.

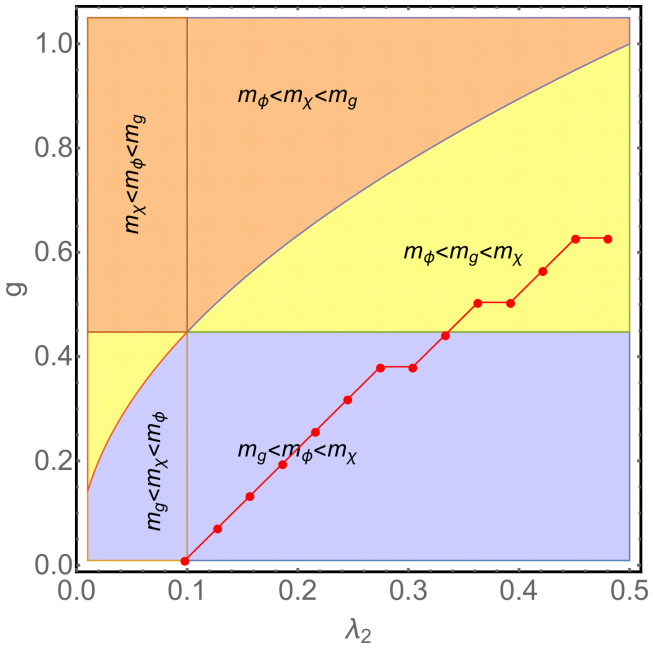

FIG. 4. Mass hierarchy in $g - \lambda_2$ plane for $\lambda = 0.1$. The different shaded regions correspond to the different mass hierarchies of the scalar and gauge field masses. The intersection of the vertical line ($\lambda_2 = \lambda$) with the horizontal axis is the point where the superfluid vortex changes from being unstable to being metastable at zero gauge coupling. The red points (connected by straight lines for better visual reference) denote the line along which the stable/unstable transition occurs. Points to the left of this line correspond to unstable configurations, points to the right to metastable superfluid solutions. The non-smooth nature of this line can be attributed to the finite resolution used while exploring the space of couplings.

D. Relation between the effective theory and QCD

At arbitrary density it is not possible to calculate the couplings in our effective theory in terms of the microscopic physics, namely QCD, because QCD is strongly coupled. However, in the ultra-high density regime, where the coupling becomes weak, the parameters λ_1 and λ_2 have been calculated using the mean field approximation [9], [10] in terms of baryon chemical potential μ and the transition temperature T_c of the CFL condensate,

$$\lambda_1 = \lambda_2 = \frac{\lambda}{4} = \frac{36}{7}\frac{\pi^4}{\zeta(3)}\left(\frac{T_c}{\mu}\right)^2 . \qquad (32)$$

Using this expression, the range of values of λ and λ_2 that we have explored in our study correspond to values of μ ranging from 400 MeV to 500 MeV and values of T_c ranging from 10 MeV to 15 MeV. As we can see, in the weak coupling mean field calculation $\lambda_1 = \lambda_2$, i.e. $\lambda = 4\lambda_2$. In our calculations the superfluid vortex is unstable for $\lambda = 4\lambda_2$. We illustrate this in Figure 3 where the dashed straight line is where $\lambda = 4\lambda_2$, and even at very small QCD coupling g this line is in the unstable region. Increasing the value of the coupling constant g shrinks the region of meta-stability away from the plane $\lambda = 4\lambda_2$, and increasing the value of T_c just takes us to larger λ and λ_2, so from Figure 3 it seems likely that the vortices will remain unstable at large g and T_c.

V. CONCLUSION AND DISCUSSION

We have studied the stability properties of the superfluid vortices in the CFL phase of dense quark matter. Using a Ginzburg-Landau effective theory of the condensate field discretized on a two-dimensional spatial lattice, we evolved the vortex configuration in time and looked for an exponentially growing unstable mode. We scanned the parameter space of the couplings, mapped the regions where the vortices are unstable as opposed to metastable, and in unstable regions we identified the unstable mode. We found that the region where superfluid vortices are metastable is rather small. Vortices are metastable when the gauge coupling g is sufficiently small. In that case, the transition line separating the metastable from the unstable direction is almost as predicted in equation (26), that is, $\lambda \approx \lambda_2$, see upper left panel in Figure 3. At larger values of g, the metastable region shrinks and vanishes around $g = 0.6$. We could not find any sign of metastability above values of $\lambda \approx 0.6$. If we use mean-field weak-coupling calculations to relate the Ginzburg-Landau couplings to QCD parameters such as g, μ, and T then it seems likely that CFL vortices in neutron stars would be unstable rather than metastable, but these calculations are not valid in the density range of interest for neutron stars. If better calculations of the effective theory couplings become available, the physical region in our plots could be more precisely identified.

It is very interesting that a superfluid vortex can be rendered unstable by solely perturbing the quark condensate, in spite of the fact that in CFL matter gauge fields play an essential role in the later states of the decay process, in which the vortex separates into three color flux tubes that repel each other. In a theory without gauge fields there would be no such repulsive force, and our preliminary calculations suggest that in that theory the superfluid vortex decays into a molecule-like configuration of three separate vortices which remain bound at a fixed spacial separation.

This work opens up several avenues of future enquiry. We only investigated the early stages of the process of vortex decay. For regions of parameter space where the vortices are metastable it would be interesting to measure their lifetime and evaluate the height of the energy barrier. Our Ginzburg-Landau theory did not include entrainment (current-current) interactions, and it would be interesting to study how that affect our results. The same methods that we used could be applied to vortices in the color-spin-locked phase of quark matter, and to study the stability of the proposed color-magnetic flux tubes in two-flavor color superconducting quark matter.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Award Number #DE-FG02-05ER41375. AW acknowledges support by the Schrödinger Fellowship J 3800-N27 of the Austrian Science Fund (FWF). TV thanks Washington University for hospitality. TV acknowledges support by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award number #DE-SC0013605.

Appendix A: Equations of motion on the lattice

We solved the equations of motion on a two-dimensional spacial lattice whose lattice spacing we denote by "a". The matter field and its conjugate momentum are represented by 3×3 complex matrices Φ and Π with one color and one flavor index which live on the lattice sites. The gauge fields are 3×3 unitary matrices U with two color indices living on the links, and their time derivative is encoded in the electric field E which is a Hermitian matrix. In what follows, i and j are discretized versions of spatial coordinates x and y, and t represents time. $U(t,i,j,+\hat{\mu})$ represents a gauge link at time t that emanates from the site (i,j) along the $\hat{\mu}$ direction. We perform our calculations in the temporal gauge, $A_t(t,i,j) = 0$. One can construct a plaquette at time t, denoted by $U_\Box(t,i,j,\hat{\mu})$, by starting at site (i,j) and going in $\hat{\mu}$ direction. After each step, the orientation is changed by a 90-degree turn to the left, such that the plaquette closes after four steps. At any given time t, starting at site (i,j), one can thus construct four different elementary plaquettes, depending on the direction of the initial step. The plaquette then corresponds to the product of the four link variables in the order of their appearance. For example, the plaquette $U_\Box(t,i,j,-\hat{y})$ becomes the matrix product $U^\dagger(t,i,j-1,+\hat{y})U(t,i,j-1,+\hat{x})U(t,i+1,j-1,+\hat{y})U^\dagger(t,i,j,+\hat{x})$.

The lattice energy functional is

$$E_{\text{lattice}} = \sum_{ij} (\mathcal{H}_{\text{magnetic}} + \mathcal{H}_{\text{electric}} + \mathcal{H}_{\text{potential}} + \mathcal{H}_{\text{kinetic}}), \tag{A1}$$

where

$$\mathcal{H}_{\text{magnetic}} = \frac{6}{g^2 a^2}\left(1 - \frac{1}{3}\Re\text{Tr}\left[U_\Box(t,i,j,+\hat{x})\right]\right), \tag{A2}$$

$$\mathcal{H}_{\text{electric}} = \frac{1}{2}\text{Tr}\left[E(t,i,j)^2\right], \tag{A3}$$

$$\begin{aligned}\mathcal{H}_{\text{potential}} = {} & \text{Tr}\left[-m^2\Phi(t,i,j)^\dagger\Phi(t,i,j)\right] \\ & +\text{Tr}\left[\lambda_2(\Phi(t,i,j)^\dagger\Phi(t,i,j))^2\right] \\ & +\lambda_1(\text{Tr}[\Phi(t,i,j)^\dagger\Phi(t,i,j)])^2 + \frac{3m^4}{4\lambda},\end{aligned} \tag{A4}$$

$$\mathcal{H}_{\text{kinetic}} = \text{Tr}\left[\Pi(t,i,j)^\dagger\Pi(t,i,j)\right]. \tag{A5}$$

The lattice equations of motion derived from the above energy functional are

$$\dot{\Pi}(t,i,j) = \nabla^2\Phi(t,i,j) + m^2\Phi(t,i,j) - 2\lambda_1\text{Tr}\left[\Phi(t,i,j)^\dagger\Phi(t,i,j)\right]\Phi(t,i,j) - 2\lambda_2\Phi(t,i,j)\Phi(t,i,j)^\dagger\Phi(t,i,j), \tag{A6}$$

where

$$\begin{aligned}\nabla^2\Phi(t,i,j) = {} & \frac{1}{a^2}\left(U(t,i,j,+\hat{x})\Phi(t,i+1,j) + U^\dagger(t,i-1,j,+\hat{x})\Phi(t,i-1,j) - 2\Phi(t,i,j)\right) \\ & + \frac{1}{a^2}\left(U(t,i,j,+\hat{y})\Phi(t,i,j+1) + U^\dagger(t,i,j-1,+\hat{y})\Phi(t,i,j-1) - 2\Phi(t,i,j)\right),\end{aligned} \tag{A7}$$

$$\dot{E}^\alpha_x(t,i,j) = \text{Im}\left(\text{Tr}\left[T^\alpha\left(\frac{2}{ga}\left(U_\Box(t,i,j,+\hat{x}) - U_\Box(t,i,j,-\hat{y})\right) + \frac{2g}{a}\left(U(t,i,j,+\hat{x})\Phi(t,i+1,j)\Phi^\dagger(t,i,j)\right)\right)\right]\right), \tag{A8}$$

$$\dot{E}^\alpha_y(t,i,j) = \text{Im}\left(\text{Tr}\left[T^\alpha\left(\frac{2}{ga}\left(U_\Box(t,i,j,+\hat{y}) - U_\Box(t,i,j,+\hat{x})\right) + \frac{2g}{a}\left(U(t,i,j,+\hat{y})\Phi(t,i,j+1)\Phi^\dagger(t,i,j)\right)\right)\right]\right), \tag{A9}$$

$$U(t+\delta t,i,j,+\hat{\mu}) = \exp\left[-\text{i}ga\,\delta t\,E^\alpha_\mu(t+\delta t,i,j)T^\alpha\right]U(t,i,j,+\hat{\mu}), \tag{A10}$$

where T^α's are the SU(3) generators. We choose the normalization $\text{Tr}[T^\alpha T^\beta] = \delta^{\alpha\beta}$.

$$\Phi(t+\delta t,i,j) = \Phi(t,i,j) + \delta t\,\Pi(t+\delta t,i,j). \tag{A11}$$

In order to obtain the initial superfluid vortex configuration on the lattice, we use the Langevin evolution

method. The continuum profile of the superfluid vortex is first evolved using the Langevin approach, with temperature set to zero. We then take the final relaxed configuration and use it as the input configuration for our numerical studies on stability. In the Langevin implementation, equations (A10) and (A11) are modified as follows :

$$U(t+\delta t,i,j,+\hat{\mu}) = \exp\left[-\mathrm{i}ga\,\delta t\,\left(E^{\alpha}_{\mu}(t,i,j) + \delta t\left(\dot{E}^{\alpha}_{\mu}(t,i,j) - \eta E^{\alpha}_{\mu}(t,i,j) + \zeta^{\alpha}(t,i,j)\right)\right)T^{\alpha}\right] U(t,i,j,+\hat{\mu}), \quad \text{(A12)}$$

$$\Phi(t+\delta t,i,j) = \Phi(t,i,j) + \delta t\left(\Pi(t,i,j) + \delta t\left(\dot{\Pi}(t,i,j) - \eta\Pi(t,i,j) + \zeta_{\mathrm{Re}}(t,i,j) + \mathrm{i}\,\zeta_{\mathrm{Im}}(t,i,j)\right)\right). \quad \text{(A13)}$$

where η is the coefficient of viscosity. By the fluctuation-dissipation theorem the stochastic noise terms ζ_{Re}, ζ_{Im}, ζ^{α} are independently drawn from the same Gaussian probability distribution

$$\zeta = \sqrt{\frac{2\eta\Theta}{a^2\delta t}}\,\xi(0,1) \quad \text{(A14)}$$

where $\xi(0,1)$ is a Gaussian random number of zero mean and unit variance. Θ is the temperature of the thermal bath that is coupled to the system.

We tested two types of boundary conditions (BC): fixed and Neumann. Fixed BC consisted of locking the matter and gauge fields at the edge of the lattice to the values they would take when there is a superfluid vortex at the center of the lattice. With fixed BC the boundary affects the later stages of the decay of a superfluid vortex because the edges repel the semisuperfluid flux tubes. For Neumann BC we fixed the matter and gauge fields at the boundary to be the same as their neighbors one lattice spacing in. This sets the gradient of the field configuration to zero at the edge. With Neumann BC the later stages of the decay behave correctly, with the three semisuperfluid flux tubes leaving the lattice and disappearing across the boundary. However, with Neumann BC the early states of the decay were affected by a very slight attraction of the vortex to the boundary, so if the vortex lives for long enough it starts moving slowly towards the boundary. We therefore used fixed BC in our calculations.

In order to study finite size effects, we used different lattice sizes for our calculations. Calculations that led to the main results of this paper were performed on a 61^2 lattice and were repeated on a lattice of size 101^2, where we found no significant discrepancy in the results. We furthermore varied the criteria in the code that are responsible for detecting the exponential growth of the unstable mode. We found our results to be robust under these changes as well.

[1] M. G. Alford, K. Rajagopal, and F. Wilczek, Phys. Lett. **B422**, 247 (1998), arXiv:hep-ph/9711395 [hep-ph].

[2] A. P. Balachandran, S. Digal, and T. Matsuura, Phys. Rev. **D73**, 074009 (2006), arXiv:hep-ph/0509276 [hep-ph].

[3] E. Nakano, M. Nitta, and T. Matsuura, Phys. Lett. **B672**, 61 (2009), arXiv:0708.4092 [hep-ph].

[4] M. Eto, Y. Hirono, M. Nitta, and S. Yasui, PTEP **2014**, 012D01 (2014), arXiv:1308.1535 [hep-ph].

[5] M. Cipriani and M. Nitta, Phys. Rev. **A88**, 013634 (2013), arXiv:1304.4375 [cond-mat.quant-gas].

[6] M. Gleiser and J. Thorarinson, Phys. Rev. **D79**, 025016 (2009), arXiv:0808.0514 [hep-th].

[7] U. W. Heinz, C. R. Hu, S. Leupold, S. G. Matinyan, and B. Muller, Phys. Rev. **D55**, 2464 (1997), arXiv:hep-th/9608181 [hep-th].

[8] M. Eto and M. Nitta, Phys. Rev. **D80**, 125007 (2009), arXiv:0907.1278 [hep-ph].

[9] K. Iida and G. Baym, Phys. Rev. **D63**, 074018 (2001), [Erratum: Phys. Rev.D66,059903(2002)], arXiv:hep-ph/0011229 [hep-ph].

[10] I. Giannakis and H.-c. Ren, Phys. Rev. **D65**, 054017 (2002), arXiv:hep-ph/0108256 [hep-ph].

Fundamental Implications of Intergalactic Magnetic Field Observations

Tanmay Vachaspati
Physics Department, Arizona State University, Tempe, AZ 85287, USA.

Helical intergalactic magnetic fields at the $\sim 10^{-14}$ G level on ~ 10 Mpc length scales are indicated by current gamma ray observations. The existence of magnetic fields in cosmic voids and their non-trivial helicity suggest that they must have originated in the early universe and thus have implications for the fundamental interactions. I combine present knowledge of the observational constraints and the dynamics of cosmological magnetic fields to derive characteristics that would need to be explained by the magnetic field generation mechanism. The importance of CP violation and a possible crucial role for chiral effects in the early universe are pointed out.

Several independent investigations of gamma rays from blazars indicate the presence of intergalactic magnetic fields [1–6]. Emission of TeV energy gamma rays from blazars and the subsequent electromagnetic cascade in the intergalactic medium is expected to distort the intrinsic blazar spectrum by depleting photons from the TeV range and adding photons in the GeV range. The lack of expected additional photons in the GeV range is explained by invoking an intergalactic magnetic field of strength $\gtrsim 10^{-16}$ GeV. As an intergalactic magnetic field disperses the additional GeV photons, the intergalactic magnetic field hypothesis also predicts a halo of GeV photons around the blazar. An analysis of stacked blazars provides evidence for such a halo and adds support to the derived lower bound on intergalactic magnetic fields [6].

An alternative approach developed in Refs. [7, 8] utilizes the *helical* nature of intergalactic magnetic fields. The reasoning is that intergalactic magnetic fields are measured in cosmic voids, $\sim$ 100 Mpc away from astrophysical sources, and thus were most likely generated in the early universe. (For a review of magnetic fields and some possible astrophysical generation mechanisms see Ref. [9].) Unless the magnetic fields are coherent on very long length scales or are helical at the time of production, they would dissipate and not survive until the present epoch. If the magnetic field generation mechanism was causal, the magnetic fields are not coherent on large length scales and helicity is essential for survival. Furthermore, the observation of helicity can help distinguish between cosmological and astrophysical magnetic fields as a globally preferred sign of the helicity would be indicative of a fundamental production mechanism.

In Refs. [7, 8] it was shown that the helicity of the intergalactic magnetic field leaves a parity odd imprint on the distribution of cascade gamma rays. Thus helicity can be deduced by calculating parity odd correlators of observed gamma ray arrival directions. (Simulations of the process can be found in [10, 11].) Using this technique, it becomes possible to *measure* – not jut bound – the power spectra of intergalactic magnetic fields. Applying this technique on current Fermi-LAT data, Refs. [4, 5] estimate the intergalactic magnetic field to be $\sim 10^{-14}$ G as measured on a length scale ~ 10 Mpc . The statistical significance of these measurements is at $\sim 3.5\sigma$ level in analysis with current data [12]. Further observations, especially using a variety of observational tools, will be able to confirm or refute these findings. For this paper we proceed on the assumption that the accumulating observational evidence is correct.

The existence of helical intergalactic magnetic fields points to an early universe origin and therefore is of interest to particle cosmology. As observational dataset gets larger, it will become possible to measure the magnetic field correlation functions over a range of scales. If the spectrum is flat or red, *i.e.* does not fall off at large length scales, the magnetic field would likely be a product of the big bang or inflation. In this case, the primordial magnetic field may shed light on cosmological initial conditions and it may also have important consequences for the origin of the matter-antimatter asymmetry [13, 14] and other theoretical ideas [15]. If the spectrum is measured to be blue, we expect the magnetic field to have been produced in high energy particle processes, and the helicity of the magnetic field points to an important role for CP violating interactions in the early universe.

For the rest of our discussion, we will assume that the intergalactic magnetic field is stochastic and isotropic, and is generated by a causal mechanism. (If the generation mechanism were acausal, the field may not even be stochastic within our cosmic horizon.) Then the spatial correlation function of the magnetic field is given by [16]

$$\langle B_i(\boldsymbol{x})B_j(\boldsymbol{x}+\boldsymbol{r})\rangle = M_N(r)P_{ij}+M_L(r)\hat{r}_i\hat{r}_j+\epsilon_{ijk}\hat{r}_k M_H(r) \quad (1)$$

where $P_{ij} = \delta_{ij} - \hat{r}_i\hat{r}_j$. $M_N(r)$ and $M_L(r)$ are the "normal" and "longitudinal" power spectra and are related by a differential equation [16]; $M_H(r)$ is the helical power spectrum and is what is measured by the parity odd gamma ray correlators.

Our first task is to relate the spatial helical correlation function to its counterpart in Fourier space because the magneto-hydrodynamic (MHD) evolution of the magnetic field is carried out in Fourier space while the field correlations are measured in physical space. The Fourier space correlation functions for a stochastic,

isotropic magnetic field are written as

$$\langle b_i(\boldsymbol{k}) b_j^*(\boldsymbol{k}')\rangle = \left[\frac{E_M(k)}{4\pi k^2} p_{ij} + i\epsilon_{ijl}k_l \frac{H_M(k)}{8\pi k^2}\right] \times (2\pi)^6 \delta^{(3)}(\boldsymbol{k}-\boldsymbol{k}') \quad (2)$$

where $p_{ij} = \delta_{ij} - \hat{k}_i\hat{k}_j$ and

$$\boldsymbol{b}(\boldsymbol{k}) = \int d^3x \boldsymbol{B}(\boldsymbol{x}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}}, \quad \boldsymbol{B}(\boldsymbol{x}) = \int \frac{d^3k}{(2\pi)^3} \boldsymbol{b}(\boldsymbol{k}) e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \quad (3)$$

We now use Eq. (3) in (1) to obtain

$$M_H(r) = \frac{1}{2}\int_0^\infty dk\ kH_M(k) \frac{d}{d\rho}\left(\frac{\sin\rho}{\rho}\right) \quad (4)$$

where $\rho = kr$.

Studies of the MHD equations show that a cosmological magnetic field with helicity evolves so that at late times [17–20]

$$E_M(k) = \frac{k}{2}|H_M(k)| = \begin{cases} E_0(k/k_d)^4, & 0 \le k \le k_d \\ 0, & k_d < k \end{cases} \quad (5)$$

where the first equality is the relation for maximal helicity, the functional dependence k^4 defines the "Batchelor spectrum", and k_d is a dissipation scale that will be discussed below. For $k > k_d$, the spectrum falls off rapidly and so we have set it to zero. Strictly, the Batchelor spectrum only applies for $k < k_I$ where $k_I < k_d$ is the "inertial scale" where the spectrum peaks. For $k_I < k < k_d$, the spectrum falls off as a power law and there is a sharper fall off for $k > k_d$ [50] [21]. For simplicity, we have taken $k_I \approx k_d$, which may also be justified if the magnetic field is generated on very small scales. We shall also assume $H_M(k) \ge 0$ to be concrete. Below we will estimate the power spectrum amplitude, E_0 in Eq. (5).

Gamma ray observations have been used to measure $M_H(r)$. So we use Eq. (5) in (4) to obtain $M_H(r)$

$$M_H(r) = \frac{E_0 k_d}{\rho_d^5}[(\rho_d^3 - 8\rho_d)\sin\rho_d + 4(\rho_d^2 - 2)\cos\rho_d + 8] \quad (6)$$

where $\rho_d \equiv k_d r$. One can check: $M_H(r) \propto r$ as $r \to 0$ and $M_H(r) \to \sin(k_d r)/r^2$ as $r \to \infty$, so $M_H(r)$ is well-behaved for all r.

Any observation will measure a "smeared" $M_H(r)$. For example, gamma ray observations in Refs. [4, 5] measure M_H on a certain distance scale r that is determined from the energies of observed gamma rays. However, for statistical purposes, the observed gamma rays are binned according to their energies – in 10 GeV wide bins in Refs. [4, 5]. This means that observations yield M_H that is smeared over a range, Δr, of r. With present day observations, r is typically on the order of Mpc, and $l_d = 2\pi/k_d$ is typically kpc, so that $\rho_d = k_d r \gg 1$. The precise smearing function depends on the binning procedure and experimental details (*e.g.* energy dependence of time exposure of the experiment), however, with current parameters $l_d \ll \Delta r \lesssim r$.

Let us write $\Delta\rho_d = k_d \Delta r$. Then, from Eq. (6), the smearing procedure will effectively replace the oscillating trigonometric functions by (weighted) averages. For example,

$$\frac{\sin\rho_d}{\rho_d^2} \to \frac{1}{\Delta\rho_d}\int_{\rho_d}^{\rho_d+\Delta\rho_d} d\rho \frac{\sin\rho}{\rho^2} \approx \frac{\mathcal{O}(1)}{\rho_d^2}. \quad (7)$$

Since $\rho_d \gg 1$, the ρ_d^3 term in the square bracket in Eq. (6) will dominate and we can write

$$M_H(r) \approx \frac{E_0 k_d}{\rho_d^2} \quad (8)$$

Therefore a measurement of $M_H(r)$ at $r = r_*$, denoted M_{H*}, will give

$$E_0 = \frac{\rho_*^2 M_{H*}}{k_d} \quad (9)$$

where $\rho_* = k_d r_*$, and the magnetic field energy and helicity spectra in Eq. (5) become,

$$E_M(k) = \frac{k}{2}|H_M(k)| = \rho_* r_* |M_{H*}| \left(\frac{k}{k_d}\right)^4 \quad (10)$$

From Eq. (50) of Ref. [7] [51] we have the estimate

$$|M_{H*}| \sim (10^{-14}\ \mathrm{G})^2 \quad (11)$$

and $r_* \sim 10$ Mpc. Subsequent (and ongoing) analyses [5] show rough agreement with these estimates and future observations should be able to pin down the values more accurately. Other analyses [1–3, 6] do not provide measurements of the field strength but they do provide lower bounds if they assume a coherence scale and a spectrum. These lower bounds on the field strength are on the order of 10^{-16} G (see Fig. 12 of Ref. [9]).

The energy density in the magnetic field is

$$\mathcal{E} = \frac{1}{2}\langle \boldsymbol{B}^2\rangle = \int dk\ E_M(k) \sim (10^{-14}\ \mathrm{G})^2 \frac{\rho_*^2}{5} \quad (12)$$

Similarly the helicity density is given by

$$H = \lim_{V\to\infty}\left\langle \frac{1}{V}\int_V d^3x \boldsymbol{A}\cdot\boldsymbol{B}\right\rangle = \int dk\ H_M(k) \sim (10^{-14}\ \mathrm{G})^2 \frac{\rho_*^2}{2k_d} \quad (13)$$

where $\boldsymbol{B} = \mathrm{curl}(\boldsymbol{A})$.

Next we discuss the dissipation length scale l_d. In Ref. [22], the authors considered a homogeneous magnetic field and calculated the damping rate of small perturbations on this background. The dominant dissipation

of the small perturbations is due to the damping of fast magnetosonic modes. Hence this mechanism sets the dissipation scale that then depends on the strength of the background uniform field.

The damping of a *stochastic, helical* magnetic field has been discussed in Ref. [23, 24]. The evolution of the dissipation scale, which roughly coincides with the coherence scale for the Batchelor spectrum, depends on properties of the magnetic field at the time it was generated. The result for the dissipation scale at the present epoch is (see Eqs. (4) and (5) of [25])

$$l_{d0} \;=\; 0.45 \text{ pc}\sqrt{n}\left(\frac{\Omega_{B\text{Rad}g}}{0.083}\right)^{1/2} x^{-2/(n+2)} \times \left(\frac{T_g}{100 \text{ MeV}}\right)^{-n/(n+2)} \tag{14}$$

where n is the spectral index for the magnetic field, $\Omega_{B\text{Rad}g}$ is the ratio of the energy density in magnetic fields to that in radiation (in all relativistic species), T_g is the temperature, and all quantities are taken at the time of magnetic field generation (denoted by subscript "g"). Also, $x = 2.3 \times 10^{-9}$ is a numerical factor. This formula yields

$$l_{d0} \approx 1 \text{ pc} - 1 \text{ kpc} \tag{15}$$

for magnetic field generation at the electroweak epoch ($T_g = 100$ GeV), for $n = 2 - 5$ – larger n gives smaller l_{d0} – and with $\Omega_{B\text{Rad}g} = 0.083$. The index n is defined in [25] by the relation $\rho_B \propto l^{-n}$ where ρ_B is the energy density in magnetic fields on a length scale l *at the epoch of magnetogenesis*. Translating this into our language with the relation in Eq. (12) we have $n = 5$ for the Batchelor spectrum, and $n = 3$ based on a model of processes that might have occured during a first order phase transition [26].

With $l_{d0} = 1$ kpc and $r_* = 10$ Mpc we get $\rho_* = 2\pi \times 10^4$. Inserting this estimate of ρ_* in Eq. (12) gives the magnetic field energy density at the present epoch,

$$\mathcal{E}_0 \sim (3 \times 10^{-10} \text{ G})^2 \left(\frac{1 \text{ kpc}}{l_{d0}}\right)^2 \tag{16}$$

and Eq. (13) gives

$$H_0 \sim 3 \times 10^{-20} \text{ G}^2 - \text{kpc} \left(\frac{1 \text{ kpc}}{l_{d0}}\right) \tag{17}$$

In natural units ($\hbar = 1 = c$), with the conversions 1 G $= 1.95 \times 10^{-20}$ GeV$^2 = 5 \times 10^7$ cm^{-2}, we can also write

$$H_0 \sim 2 \times 10^{17} \text{ cm}^{-3} \left(\frac{1 \text{ kpc}}{l_{d0}}\right) \tag{18}$$

To get a feel for these estimates, we compare the energy density in magnetic fields to that in photons,

$$\Omega_{B\gamma 0} = \frac{\mathcal{E}_0}{\rho_{\gamma 0}} \sim 10^{-8} \left(\frac{1 \text{ kpc}}{l_{d0}}\right)^2. \tag{19}$$

where $\rho_{\gamma 0} = 4.6 \times 10^{-34}$ gms/cm$^3 \approx (4 \times 10^{-6}$ G$)^2$ is the energy density in photons at the present epoch.

To proceed further we would like to estimate $\Omega_{B\gamma}$ at earlier times. The full details of the evolution are complicated because of episodes (*e.g.* e^+e^- annihilation), viscosity, finite electrical conductivity, and unknown factors (*e.g.* neutrino masses). However a simple approximate picture emerges from various studies within the context of conventional MHD Refs. [18, 23]. Most crucially, helicity is found to be conserved, so the helicity density $H \propto a^{-3}$ where $a(t)$ is the cosmic scale factor. The dissipation scale, also the scale where most of the magnetic energy is stored, grows as $l_d \propto a \times a^{2/3}$ in the radiation dominated era and as $l_d \propto a$ in the matter dominated era (*i.e.* for temperatures greater than the temperature at matter-radiation equality $T_{\text{eq}} \approx 1$ eV) as long as the helicity is maximal [18]. So, from the relations in Eqs. (12) and (13), the energy density scaling is $\mathcal{E} \propto a^{-4} \times a^{-2/3}$ in the radiation dominated era and $\mathcal{E} \propto a^{-4}$ in the matter dominated era. With these scalings, and with the cosmic cooling rate $T \propto a^{-1}$ and the temperature at big bang nucleosynthesis (BBN) $T_{\text{BBN}} \sim 1$ MeV, we get $\Omega_{B\gamma\text{BBN}} \sim 10^{-4}(1 \text{ kpc}/l_{d0})^2$. Requiring $\Omega_{B\gamma\text{BBN}} \lesssim 1$, this means that the magnetic dissipation scale today (also the coherence scale) is observationally constrained to be larger than ~ 10 pc.

Spectral distortions of the cosmic microwave background (CMB) also provide a means to probe small scale magnetic fields for cosmological redshift z between 10^3 and 10^6 [27–30]. As of now the bounds from COBE/FIRAS measurements of the CMB spectrum are not competitive with the BBN bound. Proposed experiments, such as PIXIE, can change this situation and be able to detect CMB $\mu-$distortions for $l_{d0} \sim 1$ kpc (see Figs. 2 and 3 of Ref. [30]). Small scale magnetic fields may also leave an imprint on the CMB anisotropies through non-linear effects [31–34].

The estimate in Eq. (16) shows that intergalactic magnetic fields that are indicated by gamma ray observations may be of $\sim 3 \times 10^{-10}$ G strength on 1 kpc scales. During structure formation, the field would get compressed within galaxies by a factor $(\rho_{\text{gal}}/\rho_{\text{c}})^{1/3}$, where $\rho_{\text{gal}} \approx 10^{-24}$ gm/cm^3 is the baryonic density in the galactic disk and $\rho_{\text{c}} \approx 10^{-31}$ gm/cm^3 is the cosmic baryon density. If we assume flux freezing during structure formation, the magnetic field strength will increase by $(\rho_{\text{gal}}/\rho_{\text{c}})^{2/3} \approx 10^5$ and the coherence scale will decrease by $(\rho_{\text{c}}/\rho_{\text{gal}})^{1/3} \approx 10^{-2}$. With these numbers, and $l_{d0} = 1$ kpc, a galaxy would inherit a magnetic field with strength $\sim 3 \times 10^{-5}$ G and coherence ~ 10 pc. A somewhat larger value of $l_{d0} \sim 10$ kpc would lead to estimates that are closer to observations of the random component of the Milky Way magnetic field, $4-6$ μG on $10-100$ pc [35]. This conclusion is in line with that of Ref. [25] where the authors argue that magnetic fields in galaxy clusters may arise directly from the intergalactic magnetic field.

We now turn to the helicity of the magnetic field, a quantity that is parity (P) odd and also odd under combined charge and parity (CP) transformations. Hence observed non-zero magnetic helicity indicates a period of CP violation in the early universe, as is also necessary for the generation of the observed cosmic matter-antimatter asymmetry. Thus it is natural to compare the observed magnetic helicity to the cosmic baryon number density, $n_{b0} \approx 10^{-7}$ cm^{-3},

$$\eta_{Bb0} \equiv \frac{H_0}{n_{b0}} \sim 2 \times 10^{24} \left(\frac{1 \text{ kpc}}{l_d}\right) \tag{20}$$

This estimate raises a challenge for fundamental physics – what processes can generate such a large helicity to baryon number ratio?

The simplest particle physics based scenarios of magnetogenesis are based on the evidence that a baryon number changing process via an electroweak sphaleron [36] also produces magnetic fields with $\sim 10^2$ helicity [37, 38]. Then the magnetic helicity is proportional to the baryon number and we get $\eta_{Bb0} \sim 10^2$ [39, 40]. Even in the unbroken phase of the electroweak model, where the electroweak sphaleron solution does not exist *per se*, we expect gauge field production to occur during changes of Chern-Simons number which is necessary for baryon number violation.

A more realistic view of the production of cosmic matter asymmetry is that baryon number violating processes occur so as to produce both baryons and antibaryons but with a slight excess of baryon production. In terms of magnetic fields this means that both left- and right-handed helical fields are produced but with a slight excess of left-handed helicity that is given by the fundamental CP violation [37]. Within the context of baryogenesis in the standard model, CP violation is extremely weak [41] and the total helicity is tiny compared to the energy density in the magnetic field [37]. If we assume energy equipartition, the energy density in magnetic fields will be comparable to that in other forms of radiation [42].

The above description shows that the energy density in magnetic fields may be much larger than that implied by magnetic helicity alone. However, the problem we are encountering based on observation, is that the initial magnetic *helicity* also needs to be much larger (see Eq. (20)). Is there some dynamics beyond standard MHD that could potentially increase the magnetic helicity and saturate the maximal helicity condition in the early universe?

A simple possibility is to look for a mechanism that selectively amplifies one handedness of the magnetic field. Then, if we start with a magnetic field, even with zero net helicity, the dynamics will amplify one of the two helicities, increase the magnetic field energy density, and also saturate the helicity at its maximal value. This has been the focus of earlier studies of the "chiral magnetic effect" [43], in which a magnetic field induces an electric current $\boldsymbol{j} \propto \boldsymbol{B}$, which results in the amplification of certain Fourier modes of only one handedness [44]. More importantly for us, however, the chiral magnetic effect also selectively *dissipates* one handedness of the magnetic field (see, for example, [45]). Thus, if baryon number violating interactions (or other dynamics) produce a large but non-helical magnetic field, the chiral magnetic effect can dissipate one of the two helicities – the handedness being determined by the sign of the chiral imbalance – and reduce the magnetic field energy by half, and saturate the helicity at its maximal value.

More quantitatively, ignoring the plasma velocity field, the equations satisfied by the difference of the two helical amplitudes of the magnetic field Fourier modes, $\Delta B \equiv |B^+(k)| - |B^-(k)|$, is given by (see Eq. (60) in [45]),

$$\partial_\eta \Delta B = +\frac{k_p}{\sigma}(|B^+| + |B^-|) + \mathcal{O}(\Delta B) \tag{21}$$

where η is the conformal time, σ is the electrical conductivity of the plasma, $k_p = e^2\Delta\mu/2\pi^2$, and the chemical potential $\Delta\mu = \mu_L - \mu_R$ is a measure of the chiral imbalance of the medium. Thus ΔB grows in proportion to the summed amplitudes of the two helicities of the magnetic field and the field tends to become maximally helical on a time scale set by the chirality of the medium. (A chiral imbalance might arise naturally above the electroweak scale since the weak interactions distinguish between left- and right-handed particles at a fundamental level.) Once the field becomes maximally helical, it stays maximally helical. The precise dynamics, however, needs further investigation since the analysis outlined above ignores the plasma velocity field. The joint evolution of the magnetic field and the plasma velocity is essential to see effects such as the inverse cascade of helical fields. In a chiral medium, it is also necessary to co-evolve the chemical potentials. The joint evolution of homogeneous chiral imbalance has started to receive attention [46] but even the equations necessary to describe dynamics with spatially varying chirality have not yet been established (recent attempts can be found in [47, 48]).

The main point of this paper is that current observational evidence for intergalactic magnetic fields has profound implications for fundamental interactions. The observed magnetic fields must have originated in the early universe since they are seen in voids and are helical. If we uncover a red spectrum of the magnetic field, we would know that they were generated by an acausal mechanism. The magnetic fields would then provide valuable information about the earliest moments of the universe. If the spectrum turns out to be blue, the properties of the magnetic field will give us important clues about particle physics beyond the standard model. The observation of magnetic helicity implies a strong role for fundamental CP violation in the early universe. Since helical magnetic fields are closely connected with baryon number violating

processes, the observation of helical magnetic fields can inform us about matter-genesis. But baryogenesis by itself is insufficient to explain the large helicity that is indicated by observations. We have suggested that there may be a role for the chiral magnetic effect to drive magnetic helicity to its maximal value. Then the standard model must be extended to allow for successful baryogenesis *and* the chiral magnetic effect should play a role in cosmology. This would have implications for particle physics close to the electroweak scale and may perhaps also be testable at the LHC or future accelerator experiments. Future observations (*e.g.* by the Cherenkov Telescope Array [49]) will further sharpen the case for intergalactic magnetic fields and allow for more precise measurements of the power spectra.

I thank Jens Chluba, Francesc Ferrer, Tina Kahniashvili, Andrew Long, Levon Pogosian and Alex Vilenkin for comments and suggestions. I am grateful to the Institute for Advanced Study, Princeton for hospitality while this work was being done. This work was supported by the U.S. Department of Energy, Office of High Energy Physics, under Award No. de-sc0013605 at ASU.

[1] A. Neronov and I. Vovk, Science **328**, 73 (2010) doi:10.1126/science.1184192 [arXiv:1006.3504 [astro-ph.HE]].

[2] S. Ando and A. Kusenko, Astrophys. J. **722**, L39 (2010) doi:10.1088/2041-8205/722/1/L39 [arXiv:1005.1924 [astro-ph.HE]].

[3] W. Essey, S. Ando and A. Kusenko, Astropart. Phys. **35**, 135 (2011) doi:10.1016/j.astropartphys.2011.06.010 [arXiv:1012.5313 [astro-ph.HE]].

[4] H. Tashiro, W. Chen, F. Ferrer and T. Vachaspati, Mon. Not. Roy. Astron. Soc. **445**, no. 1, L41 (2014) doi:10.1093/mnrasl/slu134 [arXiv:1310.4826 [astro-ph.CO]].

[5] W. Chen, B. D. Chowdhury, F. Ferrer, H. Tashiro and T. Vachaspati, Mon. Not. Roy. Astron. Soc. **450**, no. 4, 3371 (2015) doi:10.1093/mnras/stv308 [arXiv:1412.3171 [astro-ph.CO]].

[6] W. Chen, J. H. Buckley and F. Ferrer, Phys. Rev. Lett. **115**, 211103 (2015) doi:10.1103/PhysRevLett.115.211103 [arXiv:1410.7717 [astro-ph.HE]].

[7] H. Tashiro and T. Vachaspati, Mon. Not. Roy. Astron. Soc. **448**, 299 (2015) doi:10.1093/mnras/stu2736 [arXiv:1409.3627 [astro-ph.CO]].

[8] H. Tashiro and T. Vachaspati, Phys. Rev. D **87**, no. 12, 123527 (2013) doi:10.1103/PhysRevD.87.123527 [arXiv:1305.0181 [astro-ph.CO]].

[9] R. Durrer and A. Neronov, Astron. Astrophys. Rev. **21**, 62 (2013) doi:10.1007/s00159-013-0062-7 [arXiv:1303.7121 [astro-ph.CO]].

[10] A. J. Long and T. Vachaspati, JCAP **1509**, no. 09, 065 (2015) doi:10.1088/1475-7516/2015/09/065 [arXiv:1505.07846 [astro-ph.CO]].

[11] R.A. Batista, A. Saveliev, G. Sigl and T. Vachaspati, in preparation (2016).

[12] F. Ferrer and T. Vachaspati, in preparation (2016).

[13] M. M. Anber and E. Sabancilar, Phys. Rev. D **92**, no. 10, 101501 (2015) doi:10.1103/PhysRevD.92.101501 [arXiv:1507.00744 [hep-th]].

[14] T. Fujita and K. Kamada, Phys. Rev. D **93**, no. 8, 083520 (2016) doi:10.1103/PhysRevD.93.083520 [arXiv:1602.02109 [hep-ph]].

[15] A. J. Long and T. Vachaspati, Phys. Rev. D **91**, 103522 (2015) doi:10.1103/PhysRevD.91.103522 [arXiv:1504.03319 [hep-ph]].

[16] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics*, MIT Press, Cambridge, MA, (1975).

[17] K. Jedamzik and G. Sigl, Phys. Rev. D **83**, 103005 (2011) doi:10.1103/PhysRevD.83.103005 [arXiv:1012.4794 [astro-ph.CO]].

[18] T. Kahniashvili, A. G. Tevzadze, A. Brandenburg and A. Neronov, Phys. Rev. D **87**, no. 8, 083007 (2013) doi:10.1103/PhysRevD.87.083007 [arXiv:1212.0596 [astro-ph.CO]].

[19] A. Saveliev, K. Jedamzik and G. Sigl, Phys. Rev. D **87**, no. 12, 123001 (2013) doi:10.1103/PhysRevD.87.123001 [arXiv:1304.3621 [astro-ph.CO]].

[20] A. Brandenburg, T. Kahniashvili, A. Tevzadze and T. Vachaspati, in preparation (2016).

[21] T. Kahniashvili, A. G. Tevzadze and B. Ratra, Astrophys. J. **726**, 78 (2011) doi:10.1088/0004-637X/726/2/78 [arXiv:0907.0197 [astro-ph.CO]].

[22] K. Jedamzik, V. Katalinic and A. V. Olinto, Phys. Rev. D **57**, 3264 (1998) [astro-ph/9606080].

[23] R. Banerjee and K. Jedamzik, Phys. Rev. D **70**, 123003 (2004) doi:10.1103/PhysRevD.70.123003 [astro-ph/0410032].

[24] T. Kahniashvili, A. G. Tevzadze, S. K. Sethi, K. Pandey and B. Ratra, Phys. Rev. D **82**, 083005 (2010) doi:10.1103/PhysRevD.82.083005 [arXiv:1009.2094 [astro-ph.CO]].

[25] R. Banerjee and K. Jedamzik, Phys. Rev. Lett. **91**, 251301 (2003) Erratum: [Phys. Rev. Lett. **93**, 179901 (2004)] doi:10.1103/PhysRevLett.91.251301, 10.1103/PhysRevLett.93.179901 [astro-ph/0306211].

[26] Y. Ng and T. Vachaspati, Phys. Rev. D **82**, 023008 (2010) doi:10.1103/PhysRevD.82.023008 [arXiv:1001.4817 [astro-ph.CO]].

[27] K. Jedamzik, V. Katalinic and A. V. Olinto, Phys. Rev. Lett. **85**, 700 (2000) doi:10.1103/PhysRevLett.85.700 [astro-ph/9911100].

[28] K. Miyamoto, T. Sekiguchi, H. Tashiro and S. Yokoyama, Phys. Rev. D **89**, no. 6, 063508 (2014) doi:10.1103/PhysRevD.89.063508 [arXiv:1310.3886 [astro-ph.CO]].

[29] K. E. Kunze and E. Komatsu, JCAP **1401**, 009 (2014) doi:10.1088/1475-7516/2014/01/009 [arXiv:1309.7994 [astro-ph.CO]].

[30] J. M. Wagstaff and R. Banerjee, Phys. Rev. D **92**, no. 12, 123004 (2015) doi:10.1103/PhysRevD.92.123004 [arXiv:1508.01683 [astro-ph.CO]].

[31] S. K. Sethi and K. Subramanian, Mon. Not. Roy. Astron. Soc. **356**, 778 (2005) doi:10.1111/j.1365-2966.2004.08520.x [astro-ph/0405413].

[32] K. Jedamzik and T. Abel, JCAP **1310**, 050 (2013). doi:10.1088/1475-7516/2013/10/050

[33] K. E. Kunze and E. Komatsu, JCAP **1506**, no. 06, 027 (2015) doi:10.1088/1475-7516/2015/06/027 [arXiv:1501.00142 [astro-ph.CO]].

[34] J. Chluba, D. Paoletti, F. Finelli and J. A. Rubio-Martn, Mon. Not. Roy. Astron. Soc. **451**, no. 2, 2244 (2015) doi:10.1093/mnras/stv1096 [arXiv:1503.04827 [astro-ph.CO]].

[35] H. Ohno and S. Shibata, Mon. Not. Roy. Astron. Soc. **262**, 953 (1993).

[36] N. S. Manton, Phys. Rev. D **28**, 2019 (1983). doi:10.1103/PhysRevD.28.2019

[37] C. J. Copi, F. Ferrer, T. Vachaspati and A. Achúcarro, Phys. Rev. Lett. **101**, 171302 (2008) [arXiv:0801.3653 [astro-ph]].

[38] Y. Z. Chu, J. B. Dent and T. Vachaspati, Phys. Rev. D **83**, 123530 (2011) [arXiv:1105.3744 [hep-th]].

[39] J. M. Cornwall, Phys. Rev. D **56**, 6146 (1997) doi:10.1103/PhysRevD.56.6146 [hep-th/9704022].

[40] T. Vachaspati, Phys. Rev. Lett. **87**, 251302 (2001) doi:10.1103/PhysRevLett.87.251302 [astro-ph/0101261].

[41] J. M. Cline, "Baryogenesis," hep-ph/0609145 (2006).

[42] T. Vachaspati, Phys. Lett. B **265**, 258 (1991). doi:10.1016/0370-2693(91)90051-Q

[43] A. Vilenkin, Phys. Rev. D **22**, 3080 (1980). doi:10.1103/PhysRevD.22.3080

[44] M. Joyce and M. E. Shaposhnikov, Phys. Rev. Lett. **79**, 1193 (1997) doi:10.1103/PhysRevLett.79.1193 [astro-ph/9703005].

[45] H. Tashiro, T. Vachaspati and A. Vilenkin, Phys. Rev. D **86**, 105033 (2012) doi:10.1103/PhysRevD.86.105033

[46] J. Schober, "First Numerical Simulations of the Dynamo Effect in Chiral MHD", talk presented at 28th Texas Symposium, Geneva, `https://indico.cern.ch/event/336103/contributions/786594/at` (2015).

[47] A. Boyarsky, J. Frohlich and O. Ruchayskiy, Phys. Rev. D **92**, 043004 (2015) doi:10.1103/PhysRevD.92.043004 [arXiv:1504.04854 [hep-ph]].

[48] E. V. Gorbar, I. A. Shovkovy, S. Vilchinskii, I. Rudenok, A. Boyarsky and O. Ruchayskiy, Phys. Rev. D **93**, no. 10, 105028 (2016) doi:10.1103/PhysRevD.93.105028 [arXiv:1603.03442 [hep-th]].

[49] M. Meyer, J. Conrad and H. Dickinson, arXiv:1603.03431 [astro-ph.HE].

[50] I thank Tina Kahniashvili for emphasizing this point.

[51] The definition of M_H we are using in Eq. (1) differs from that in [7] by a factor of r.

Probing Intergalactic Magnetic Fields with Simulations of Electromagnetic Cascades

Rafael Alves Batista[1,*], Andrey Saveliev[2,3,†], Günter Sigl[2,‡], Tanmay Vachaspati[4,§]
[1] *University of Oxford, Department of Physics - Astrophysics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, United Kingdom*
[2] *Universität Hamburg, II. Institute for Theoretical Physics, Luruper Chaussee 149, 22761 Hamburg, Germany*
[3] *Russian Academy of Sciences, Keldysh Institute of Applied Mathematics, Miusskaya sq. 4, 125047 Moscow, Russian Federation*
[4] *Arizona State University, Physics Department, 650 E. Tyler Mall, Tempe, AZ 85287, United States of America*

We determine the effect of intergalactic magnetic fields on the distribution of high energy gamma rays by performing three-dimensional Monte Carlo simulations of the development of gamma-ray-induced electromagnetic cascades in the magnetized intergalactic medium. We employ the so-called "Large Sphere Observer" method to efficiently simulate blazar gamma ray halos. We study magnetic fields with a Batchelor spectrum and with maximal left- and right-handed helicities. We also consider the case of sources whose jets are tilted with respect to the line of sight. We verify the formation of extended gamma ray halos around the source direction, and observe spiral-like patterns if the magnetic field is helical. We apply the Q-statistics to the simulated halos to extract their spiral nature and also propose an alternative method, the S-statistics. Both methods provide a quantative way to infer the helicity of the intervening magnetic fields from the morphology of individual blazar halos for magnetic field strengths $B \gtrsim 10^{-15}\,\mathrm{G}$ and magnetic coherence lengths $L_\mathrm{c} \gtrsim 100\,\mathrm{Mpc}$. We show that the S-statistics has a better performance than the Q-statistics when assessing magnetic helicity from the simulated halos.

I. INTRODUCTION

The origin, strength and structure of intergalactic magnetic fields (IGMF) remain a mystery up to the present day. Possible mechanisms to explain cosmic magnetogenesis may be divided into two main categories: cosmological scenarios predict that magnetic fields were generated through processes taking place in the early universe, such as inflation [1–4], electroweak [5–8] or QCD phase transitions [9–12], and leptogenesis [13], among others; in astrophysical scenarios the fields would be created during the later stages of evolution of the universe, for example during structure formation [14] or even thereafter [15].

Measurements of IGMF are rather difficult due to their low magnitude. Common methods to estimate the strength of IGMF are indirect and include the well-known Faraday rotation measurements which yield upper limits of the order of a few nG [18]. Lower bounds, $B \gtrsim 10^{-17}\,\mathrm{G}$, have been obtained by several authors using gamma-ray-induced electromagnetic cascades in the intergalactic space [18–25]. These lower bounds are controversial because of the claims [26–30] that the development of the cascade is suppressed by plasma instabilities that arise from interactions with the intergalactic medium. On the other hand, recent direct observations of cascades [31] suggest that plasma instabilities are not operative and that the original bounds hold. We expect that future analyses will clarify the role, if any, of plasma instabilities in the development of the electromagnetic cascade.

Magnetic fields can carry helicity ($\mathcal{H}$), which is defined as

$$\mathcal{H} = \int \mathbf{A}\cdot\mathbf{B}\,\mathrm{d}^3 r\,, \tag{1}$$

where $\mathbf{A}$ is the magnetic vector potential and $\mathbf{B} = \nabla\times\mathbf{A}$ is the magnetic field. Since magnetic helicity affects the dynamical evolution of magnetic fields, an indirect way to measure magnetic helicity is to measure the magnetic field power spectrum and compare it with the evolution seen in magnetohydrodynamical (MHD) simulations [32–34]. Ther are also some proposals to *directly* measure magnetic helicity based on the propagation of cosmic rays [35]. More recently, it has been proposed that helicity can leave characteristic parity-odd imprints on the arrival directions of gamma rays that are the result of gamma-ray-induced electromagnetic cascades [31, 36–39]. In particular, Long & Vachaspati [39] have carried out a thorough analysis of the morphology of the arrival directions of gamma rays using a semi-analytical approach, but without including the stochasticity of the magnetic field or the cascade process. Hence, a full Monte Carlo approach and three-dimensional simulations are needed in order to confirm or refute their findings and provide a solid basis for further analyses.

The observation of helical primordial magnetic fields has profound implications for particle physics and the early universe. Scenarios in which the cosmological matter-antimatter asymmetry is generated dynamically are found to also produce helical magnetic fields [40]. The handedness of the field is related to details of the

*E-mail: rafael.alvesbatista@physics.ox.ac.uk
†E-mail: andrey.saveliev@desy.de, *corresponding author*
‡E-mail: guenter.sigl@desy.de
§E-mail: tvachasp@asu.edu

matter-genesis scenario [13, 40]. If the observed magnetic fields are coherent on very large scales, they may have been produced at the initial epoch, perhaps during an inflation [1, 2]. Helicity on these scales would indicate the presence of certain parity violating interactions in the fundamental Lagrangian [41].

In the present work we perform simulations of the propagation of gamma rays in both helical and non-helical IGMF. This paper is structured as follows: first, we discuss the theory and implementation of simulations of electromagnetic cascades in Sec. II; in Sec. III we apply our approach to different magnetic field configurations, focusing in particular on the role of magnetic helicity (Sec. III C - III D); in Sec. IV we discuss the results, draw our conclusions and give a short outlook.

II. SIMULATIONS OF ELECTROMAGNETIC CASCADES IN THE INTERGALACTIC MEDIUM

A. Interactions and Energy Losses

The basic physics underlying the development of electromagnetic cascades induced by high energy gamma rays from blazars is well-known [42, 43]. A gamma ray emitted by a blazar interacts with photons from the diffuse extragalactic background radiation fields producing an electron-positron pair. The electrons[1] then upscatter photons of the cosmic microwave background to high energies in a process known as inverse Compton scattering (ICS). The electrons continue to upscatter photons until their energy diminishes. The upscattered photons can produce yet more electron-positron pairs until the energy of the photon drops below the threshold for pair production. We should therefore observe the blazar source as well as gamma rays originating from the cascade process, unless magnetic fields bend the electron trajectories sufficiently away from the line of sight.

To perform three-dimensional simulations of the development of gamma-ray-induced electromagnetic cascades in the IGM, we have modified the CRPropa 3 [44] code, commonly used for ultra-high energy cosmic ray propagation. Taking advantage of the modular structure of the code and the flexibility to handle custom magnetic field configurations, we have implemented relevant interactions for gamma rays and electrons in the energy range of interest ($1\,\mathrm{GeV} \lesssim E \lesssim 1\,\mathrm{PeV}$). Relevant interactions are pair production by gamma rays and inverse Compton scattering by electrons. Adiabatic losses due to the expansion of the universe are also taken into account. Synchrotron losses, albeit small in this energy range, are considered as well, for the sake of completeness.

Particles are propagated step-by-step. Within each step the probability of a given interaction to occur is computed using tabulated values for the interaction rate. If the particle is charged, deflections due to magnetic fields are calculated by integrating the equations of motion. By doing so, we are adopting a three-dimensional Monte Carlo approach for the propagation.

Interaction rates for pair production and inverse Compton scattering are calculated following the implementation used in the Elmag code [45], and defined as the inverse of the mean free path λ. They are tabulated for the CMB and various models of extragalactic background light (EBL) at different redshifts as follows[2] [45]:

$$R(E,z) \equiv \lambda^{-1}(E,z) = \frac{1}{8E^2}\int_0^\infty \mathrm{d}\varepsilon \int_{s_{\min}}^{s_{\max}} \mathrm{d}s \frac{n(\varepsilon,z)}{\varepsilon^2} F_{\mathrm{int}}(s)\,, \tag{2}$$

where E is the energy of the interacting particle (electron, positron or photon), $n(\varepsilon, z)$ is the comoving spectral density distribution of photons with energy ε at redshift z, s denotes the center of mass energy in the kinematic range $s_{\min} \leq s \leq s_{\max}$, and F_{int} is a function that depends on the interaction in question.

In the case of pair production $F_{\mathrm{int}} = F_{\mathrm{PP}}$ is

$$F_{\mathrm{PP}}(s) = s\sigma_{\mathrm{PP}}(s)\,, \tag{3}$$

where $\sigma_{\mathrm{PP}}(s)$ is the cross section for pair production, and $s = 2E\varepsilon(1-\cos\theta)$, with $0 \leq \theta \leq \pi$ being the angle between the gamma ray of energy E and the background photon of energy ε. The values of s range from $s_{\min} = 4m_e^2$ to $s_{\max} = 4E\varepsilon_{\max}$, where $\varepsilon_{\max}$ is the cutoff energy for the photon field, assumed to be approximately 0.1 eV for the CMB and 15 eV for the EBL.

For inverse Compton scattering $F_{\mathrm{int}} = F_{\mathrm{ICS}}$ is given by

$$F_{\mathrm{ICS}}(s) = \frac{1}{\beta}\sigma_{\mathrm{ICS}}(s - m_e^2)\,, \tag{4}$$

with $\beta = (1 - m_e^2/E^2)^{\frac{1}{2}}$. The center of mass energies in this case are $s = m_e^2 + 2E\epsilon(1-\beta\cos\theta)$, for $s_{\min} \leq s \leq s_{\max}$, with $s_{\min} = m_e^2$ and $s_{\max} = m_e^2 + 2E\epsilon_{\max}(1+\beta)$.

Cross sections for these interactions are well-known (see e.g. [45, 46]). The spectral density distribution of the cosmic microwave background (CMB) can be described as a black-body. The EBL is model-dependent. For this particular work we adopt the lower limit EBL model of Kneiske & Dole [47].

Synchrotron losses are given by

$$\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{m_e^2\chi^2}{(1 + 4.8(1+\chi)\ln(1+1.7\chi) + 3.4\chi^2)^{2/3}}\,, \tag{5}$$

[1] Hereafter we will collectively refer to electrons and positrons simply as "electrons".

[2] Unless otherwise stated, in this section we use "natural units" in which $\hbar = c = 1$.

following Ref. [45]. Here m_e is the electron mass, χ is

$$\chi \equiv \frac{|\mathbf{p}\times\mathbf{B}|}{m_e B_0}\,, \tag{6}$$

with $B_0 = 4.1\times10^{13}\,\mathrm{G}$, and $\mathbf{B}$ the magnetic field vector acting on an electron with momentum $\mathbf{p}$.

Adiabatic losses due to the expansion of the universe are given by

$$-\frac{1}{E}\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{H(t)}{c} = \frac{H_0}{c}\sqrt{\Omega_\mathrm{m}(1+z)^3+\Omega_\Lambda}\,, \tag{7}$$

with $H_0 \equiv H(0) \simeq 70\,\mathrm{km/s/Mpc}$ designating the Hubble constant at present time, $\Omega_\mathrm{m} \simeq 0.3$ being the density of matter, and $\Omega_\Lambda \simeq 0.7$ being the density of dark energy, assuming the standard ΛCDM cosmological model.

In our simulations we consider a monochromatic source and all emitted gamma rays are assumed to have an energy of 10 TeV. Photons from the source with energies much smaller than this will be below the threshold for creating a cascade, while photons with much higher energies will have a diminished flux.

B. Sampling of Helical Magnetic Fields

In order to run a simulation for a given magnetic field scenario or, more specifically, for a given magnetic field (and magnetic helicity) spectrum, one has to sample a magnetic field grid which then may be used as input. This procedure is explained in the following using the formalism of [48].

The aim is to decompose the magnetic field into modes of the divergence-free eigenfunctions $\mathbf{K}^\pm$ of the Laplace operator which for a specific wave vector $\mathbf{k}$ are given by[3]

$$\mathbf{K}^\pm(\mathbf{k}) = \mathbf{e}^\pm(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}} \equiv \frac{\mathbf{e}_1(\mathbf{k}) \pm i\mathbf{e}_2(\mathbf{k})}{\sqrt{2}}e^{i\mathbf{k}\cdot\mathbf{x}}\,, \tag{8}$$

where $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ is a right-handed orthonormal system of real unit vectors with $\mathbf{e}_3 = \mathbf{k}/k \equiv \hat{\mathbf{k}}$. In order to obtain $\mathbf{e}_1$ and $\mathbf{e}_2$ we chose a fixed arbitrary vector $\mathbf{n}_0 \nparallel \mathbf{k}$ with which we calculate

$$\mathbf{e}_1 \equiv \frac{\mathbf{n}_0\times\hat{\mathbf{k}}}{\left|\mathbf{n}_0\times\hat{\mathbf{k}}\right|}\,,\ \mathbf{e}_2 \equiv \frac{\hat{\mathbf{k}}\times\mathbf{e}_1}{\left|\hat{\mathbf{k}}\times\mathbf{e}_1\right|}\,. \tag{9}$$

With these definitions the $\mathbf{K}^\pm$ fullfil the following relations [48]:

$$\nabla\cdot\mathbf{K}^\pm = 0\,,\ \nabla\times\mathbf{K}^\pm = \pm k\mathbf{K}^\pm\,. \tag{10}$$

Considering these relations the magnetic field with $\nabla\cdot\mathbf{B} = 0$ or, in Fourier space, $\mathbf{k}\cdot\tilde{\mathbf{B}}(\mathbf{k}) = 0$ may be decomposed as

$$\mathbf{B}(\mathbf{x}) = \int\left[\tilde{B}^+(\mathbf{k})\mathbf{K}^+(\mathbf{k}) + \tilde{B}^-(\mathbf{k})\mathbf{K}^-(\mathbf{k})\right]\frac{\mathrm{d}^3k}{(2\pi)^3}\,, \tag{11}$$

for which, in order for $\mathbf{B}(\mathbf{x})$ to be real, the condition

$$\begin{aligned}&\tilde{B}^+(\mathbf{k})\mathbf{e}^+(\mathbf{k}) + \tilde{B}^-(\mathbf{k})\mathbf{e}^-(\mathbf{k})\\ &= \tilde{B}^+(-\mathbf{k})^*\mathbf{e}^+(-\mathbf{k})^* + \tilde{B}^-(-\mathbf{k})^*\mathbf{e}^-(-\mathbf{k})^*\end{aligned} \tag{12}$$

must hold. A possible realization of this condition is

$$\tilde{B}^\pm(\mathbf{k})\mathbf{e}^\pm(\mathbf{k}) = \tilde{B}^\pm(-\mathbf{k})^*\mathbf{e}^\pm(-\mathbf{k})^*\,, \tag{13}$$

which can be fulfilled by setting

$$\tilde{B}^\pm(\mathbf{k}) = \tilde{B}^\pm(-\mathbf{k})^*\,, \tag{14}$$

which we are going to use in the following. Together with (8) and (13) this leads to

$$\mathbf{e}^\pm(\mathbf{k}) = \mathbf{e}^\pm(-\mathbf{k})^*, \tag{15}$$

and thus

$$\mathbf{e}_1(\mathbf{k}) = -\mathbf{e}_1(-\mathbf{k})\,,\ \mathbf{e}_2(\mathbf{k}) = \mathbf{e}_2(-\mathbf{k})\,. \tag{16}$$

The $\tilde{B}^\pm$ may be obtained from the given spectra using the relations [48]

$$\begin{aligned}\frac{1}{8\pi}\langle|\mathbf{B}(\mathbf{x})|^2)\rangle &= \int\left[\left|\tilde{B}^+(\mathbf{k})\right|^2 + \left|\tilde{B}^-(\mathbf{k})\right|^2\right]\frac{k^2\,\mathrm{d}k}{16\pi^3}\\ &\equiv \int E_B(k)\,\mathrm{d}\ln k\end{aligned} \tag{17}$$

and

$$\begin{aligned}\langle\mathbf{A}(\mathbf{x})\cdot\mathbf{B}(\mathbf{x})\rangle &= \int\left[\left|\tilde{B}^+(\mathbf{k})\right|^2 - \left|\tilde{B}^-(\mathbf{k})\right|^2\right]\frac{k\mathrm{d}k}{2\pi^2}\\ &\equiv \int H_B(k)\,\mathrm{d}\ln k\,,\end{aligned} \tag{18}$$

where $\mathbf{A}$ is the vector potential and E_B and H_B are the spectra of the magnetic energy density and the magnetic helicity density, respectively. E_B and H_B are related to each other through the inequality [49]

$$\frac{k}{8\pi}\,|H_B(k)| \le E_B(k) \tag{19}$$

which may be also expressed as

$$H_B(k) = f_H(k)\frac{8\pi}{k}E_B(k) \tag{20}$$

with $-1 \le f_H(k) \le 1$.

Numerical and analytical analyses [34, 50] show that E_B is a power-law for small k, *i.e.*

$$E_B \propto k^\alpha\,, \tag{21}$$

[3] We adopt CGS units in this section.

with $\alpha = 5$. This power-law behavior for E_B is also known as the Batchelor spectrum.

In our numerical analysis with stochastic magnetic fields of Sec. III C, we will use magnetic fields with the spectrum

$$E_B \propto \begin{cases} k^5\,,\ k \leq 2\pi/L_{\rm min}\,, \\ 0\,,\ k > 2\pi/L_{\rm min}\,, \end{cases} \tag{22}$$

where, for a correlation length $L_{\rm c} = 120\,{\rm Mpc}$, $L_{\rm min} = 8L_{\rm c}/5 = 192\,{\rm Mpc}$ is the cutoff scale (cf. Eq. (28) below). Finally, solving (17) and (18) for $\left|\tilde{B}^{\pm}\right|^2$ gives

$$\begin{aligned} \left|\tilde{B}^{\pm}\right|^2 &= \frac{8\pi^3}{k^3}\left[E_B(k) \pm \frac{k}{8\pi}H_B(k)\right] \\ &= \left(\frac{2\pi}{k}\right)^3 \left[1 \pm f_H(k)\right] E_B(k)\,. \end{aligned} \tag{23}$$

With these considerations the procedure for sampling a magnetic field for given spectra E_B and H_B on a grid in x-space is the following: first, for each $\mathbf{k}$ in the Fourier-transformed k-space a value for the norm of $\tilde{B}(k)$ is generated from a normal distribution with mean value $\mu = 0$ and standard deviation $\sigma = 2(2\pi/k)^3 E_B(k)$ as follows from (23) with $f_{\rm H} = \pm 1$. Next, we include a random phase factor

$$\tilde{B}^{\pm}(\mathbf{k}) = \left|\tilde{B}^{\pm}(\mathbf{k})\right| \left[\cos\theta^{\pm}(\mathbf{k}) + i\sin\theta^{\pm}(\mathbf{k})\right], \tag{24}$$

where $\theta^{\pm}(\mathbf{k})$ are random phases distributed uniformly on $[0; 2\pi)$. Once we have $\tilde{B}^{\pm}(\mathbf{k})$, we use Eq. (14) to find $\tilde{B}^{\pm}(-\mathbf{k})$. These $\tilde{B}^{\pm}(\mathbf{k})$ can then be plugged into (11) to obtain the value for $\mathbf{B}(\mathbf{x})$ at a given $\mathbf{x}$.

As sometimes it is more convenient to have $\tilde{\mathbf{B}}(\mathbf{k})$ given in terms of the real and imaginary parts, we use Eq. (8) to write it down in the form

$$\begin{aligned} \tilde{\mathbf{B}}(\mathbf{k}) &= \tilde{B}^{+}(\mathbf{k})\mathbf{e}^{+}(\mathbf{k}) + \tilde{B}^{-}(\mathbf{k})\mathbf{e}^{-}(\mathbf{k}) \\ &= \frac{1}{\sqrt{2}}\Big\{ \left[\left(\left|\tilde{B}^{+}(\mathbf{k})\right|\cos\theta^{+} + \left|\tilde{B}^{-}(\mathbf{k})\right|\cos\theta^{-}\right)\mathbf{e}_1 + \left(-\left|\tilde{B}^{+}(\mathbf{k})\right|\sin\theta^{+} + \left|\tilde{B}^{-}(\mathbf{k})\right|\sin\theta^{-}\right)\mathbf{e}_2\right] \\ &+ i\left[\left(\left|\tilde{B}^{+}(\mathbf{k})\right|\sin\theta^{+} + \left|\tilde{B}^{-}(\mathbf{k})\right|\sin\theta^{-}\right)\mathbf{e}_1 + \left(\left|\tilde{B}^{+}(\mathbf{k})\right|\cos\theta^{+} - \left|\tilde{B}^{-}(\mathbf{k})\right|\cos\theta^{-}\right)\mathbf{e}_2\right]\Big\}. \end{aligned} \tag{25}$$

III. RESULTS

In this section we present the results. Some preliminary considerations regarding the setup of simulations should first be made.

We use the Large Sphere Observer approach which is a computationally efficient method for studying cosmic and gamma rays from a single source [51, 52]. It is defined by the fact that this source is located in the center of a sphere which has a radius equal to $D_{\rm s}$, the distance from the source to the observer. Hence, if a particle crosses the sphere from the inside to the outside, it is flagged 'detected'. This will henceforth be called a 'hit' and it corresponds to the particle reaching the observer.

The source can emit gamma rays either within a jet or isotropically. Due to the choice of a large sphere as an observer, all events above a given energy threshold (here we use 1.5 GeV) are detected. Moreover, we can easily select a subset of the events and consider an arbitrary emission pattern, such as a jet of arbitrary half-opening angle Ψ, or an emission around an arbitrary direction tilted with respect to the line of sight.

Simple geometrical considerations allow us to correct the arrival directions on the large sphere to mimic Earth's field of view. In the sky maps presented in this work, for each hit, the corresponding coordinate system of the observer is placed such that its origin is located at the position of the hit while the z axis points towards the source, *i.e.* in the direction of the center of the sphere. In order to determine the directions of the x and y axes, we take a "global" reference frame at a fixed point of the sphere and parallel-transport it along a geodesic to the location of the hit. Then the spherical angles of the event are measured in the local frame located at the hit point.

While the "Large Sphere Observer" method is economical as no photons are wasted, one possible concern is that in a realistic set-up most photons would indeed be wasted and the actual halo morphology would be sensitive to the absent photons. However, our results in the test case of a uniform magnetic field correlate well with analytic simulations [39], giving us confidence in the method.

The magnetic field (except for the uniform case) is sampled in a grid with 1000^3 cells, where each cell has a size of ~ 10 Mpc.

A. Comparison with Analytic Estimates

For a gamma ray emitted at TeV energy $E_{\rm TeV}$ and observed at an energy E_γ, originating from a source

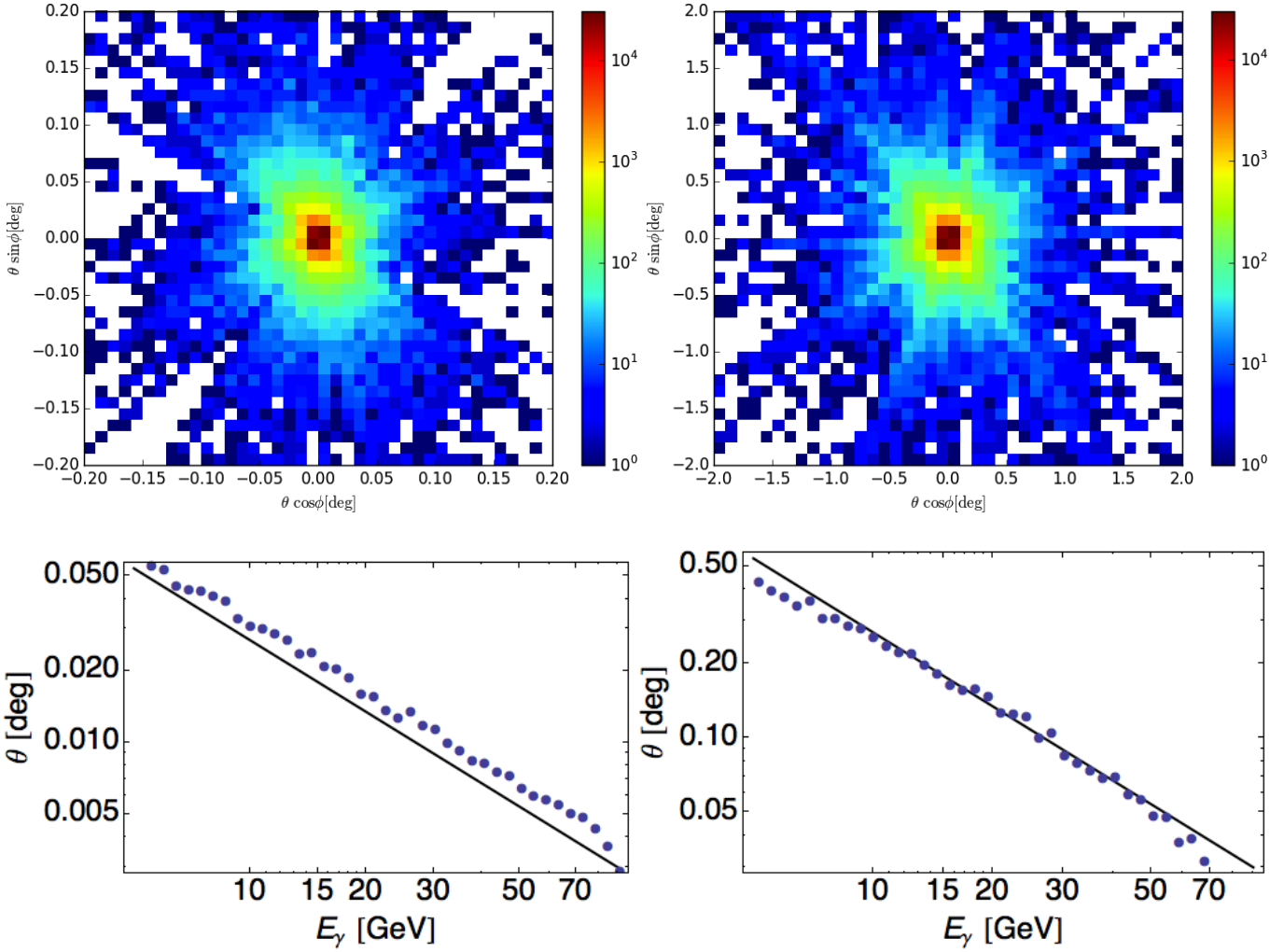

FIG. 1: Arrival directions of photons from a monochromatic TeV blazar emitting gamma rays with energies $E_{\rm TeV} = 10$ TeV in a collimated jet, in the energy range of 1-100 GeV, projected onto a plane, are shown in the upper row; the color scale indicates the number of photons per bin. The deflection angle (θ) of observed gamma rays as a function of the energy are presented in the bottom row. The magnetic field is stochastic with a spectrum according to (21) and a mean field strength of 10^{-15} G (left column) and 10^{-16} G (right column); blue dots correspond to simulation results and the black line represents the analytical prediction using Eq. (26).

(blazar in our case) located at redshift $z_{\rm s}$ and distance $D_{\rm s}$, traversing a magnetic field of strength B, the expected average angular arrival direction is [18]

$$\theta(E_\gamma) \simeq 0.05^\circ \kappa (1+z_{\rm s})^{-4} \times \left(\frac{B}{\rm fG}\right)\left(\frac{E_\gamma}{0.1\,\rm TeV}\right)^{-1}\left(\frac{D_{\rm s}}{\rm Gpc}\right)^{-1}\left(\frac{E_{\rm TeV}}{10\,\rm TeV}\right)^{-1}. \quad (26)$$

This formula is only a rough estimate where κ is a factor close to unity, $\kappa \simeq 1$, which varies slightly with the EBL model chosen. Furthermore, this equation is only valid if the coherence length ($L_{\rm c}$) of the field is much larger than the propagation length of electrons before they upscatter photons via inverse Compton. This is always true in our simulations because the propagation length is of the order of 30 kpc, whereas the minimum coherence length is 10 Mpc.

In order to compare our results with Eq. (26), we simulate the propagation of gamma rays with initial energies $E_{\rm TeV} = 10\,\rm TeV$, distance $D_{\rm s} = 1\,\rm Gpc$ ($z_s \simeq 0.25$), emitted in a collimated jet along the line of sight assuming stochastic magnetic fields with strength of $B = 10^{-16}\,\rm G$ and $B = 10^{-15}\,\rm G$. The maps containing the arrival directions are shown in the top panel of Fig. 1.

We have compared the deflections obtained from the simulations with the theoretical prediction of Eq. (26). This is shown in the bottom panel of Fig. 1. The results show a good agreement with the expected deflections. Differences are due to the nature of the analytic formula itself, which has been derived in [18] using various simplifying assumptions. Furthermore, as has been pointed out in Ref. [18], the deflection angle is highly sensitive to the particular EBL model used. In particular, for the

EBL model used here (Kneiske & Dole [47]), we expect $\kappa \approx 2.3$. As pointed out in Ref. [18], $0.3 \lesssim \kappa \lesssim 3.0$ for typical EBL models found in the literature.

B. Uniform Magnetic Fields

We now consider a simple scenario with a uniform magnetic field. By definition, a uniform magnetic field has a preferred direction, and therefore one has to distinguish among three general cases depending on the orientation of the magnetic field with respect to the axis of the jet: parallel, perpendicular, and intermediate orientation. The jet direction is assumed to be along the line of sight.

In Fig. 2 these different cases are shown for a magnetic field of strength 10^{-15} G, assuming that the gamma rays are emitted in a jet with a half-opening angle of 5° and with energy $E_{\rm TeV} = 10\,{\rm TeV}$. The results for the three cases with a specific focus on their energy dependence are shown in Fig. 2.

The results for the first case, in which the magnetic field is parallel to the jet axis, are rather intuitive and are shown in the upper left panel of Fig. 2. One can see that there is only one possible arrival direction, face-on, *i.e.* $\theta = 0°$, which means that only electrons created with momenta parallel to the magnetic field lines, and thus not influenced by the Lorentz force, can reach the observer. Any electron that deviates from the line of sight will have a trajectory that leaves the plane spanned by the line of sight and the velocity direction of the initial TeV photon and will not reach the observer.

The second case, shown in the top right panel of Fig. 2, has a magnetic field perpendicular to the line of sight. Here only photons arriving in a plane perpendicular to the magnetic field are detected. This means that the parent-electrons of these photons describe circular motion in this same plane. If an electron has a velocity component *parallel* to the magnetic field, it is initially directed away from the line of sight, and there is no component of the Lorentz force that can bend it back towards the observer.

In the case of an intermediate orientation of the magnetic field, illustrated through the bottom panel of Fig. 2 for a tilt angle of 45° (left) and 75° (right), we obtain results between the two extreme cases previously discussed, as expected. It is interesting to notice that the patterns are now smeared out since electrons from a range of directions can be directed towards the observer. Still, the dilution of the signal is small compared to the actual deflection, and hence this can be observed. Therefore, relevant information can still be extracted from sky maps by using the morphology of the arrival directions.

C. Stochastic Helical Magnetic Fields

Now we introduce magnetic helicity to the simulations. The source is assumed to have a half opening angle $\Psi = 5°$. We take the field to be stochastic with a Batchelor spectrum as in Eq. (22). As we are assuming the maximal helical case, *i.e.* $f_H = \pm 1$, this also fixes the spectrum of $H_B(k)$ according to Eq. (20). The field has an average field strength of $B_{\rm rms} = 10^{-15}\,{\rm G}$ and a correlation length of $L_{\rm c} \simeq 120\,{\rm Mpc}$. Here, $B^2_{\rm rms}$ can be extracted from Eq. (17) by setting

$$B^2_{\rm rms} \equiv \langle |\mathbf{B}(\mathbf{x})|^2 \rangle = \frac{1}{(2\pi)^3} \int \left| \tilde{B}(\mathbf{k}) \right|^2 {\rm d}^3 k = 8\pi \int E_B(k) {\rm d}\ln k\,, \tag{27}$$

while $L_{\rm c}$ is defined by [53]

$$L_c = \frac{1}{(2\pi)^3} \frac{\pi}{B^2_{\rm rms}} \int |\tilde{B}(\mathbf{k})|^2 k^{-1}\, {\rm d}^3 k = \frac{8\pi^2}{B^2_{\rm rms}} \int E_B(k) k^{-1} {\rm d}\ln k\,, \tag{28}$$

such that for the E_B defined in (22) we have $L_{\rm c} \simeq 5L_{\rm min}/8$, where $L_{\rm min}$ is the cutoff scale.

We have simulated the propagation of gamma rays with initial energy $E_{\rm TeV} = 10\,{\rm TeV}$ in the presence of stochastic magnetic fields with maximally negative ($f_H = -1$), zero ($f_H = 0$) and maximally positive ($f_H = +1$) helicities[4]. To simulate 10^5 photons in our standard scenario described above, *i.e.* with $D_{\rm s} = 1\,{\rm Gpc}$ and $B = 10^{-15}\,{\rm G}$, the current version of the code takes ~ 8 hours on 64 cores at 2300 MHz.

The actual values for helicities for the whole simulation box as well as along the line sight are shown in Fig. 5 in order to illustrate to which extent statistics play a role. As one can see, both for the whole grid as well as just along the line of sight, which is more important to judge about the statistical significance for a given case, the helicity distribution corresponds to the sign it has been assigned. Furthermore, from the panel on the right, one can see that for these particular realizations, for $f_{\rm H} = +1$ the absolute magnitude of $\mathbf{B}\cdot(\nabla \times \mathbf{B})$ is high close to the source and low close to the observer, while for $f_{\rm H} = -1$ it stays roughly equal along the propagation path. We can understand the qualitatively similar patterns for $f_H = \pm 1$, *i.e.* both patterns are spirals with similar twist, by noting that pair production on average happens close to the source, and both cases have similar helicity measures in that region. As a confirmation of

[4] We could generate the $f_H = -1$ gamma ray distribution by a parity reversal of the $f_H = +1$ plot. However, we simulate the two cases independently to show two different stochastic realizations.

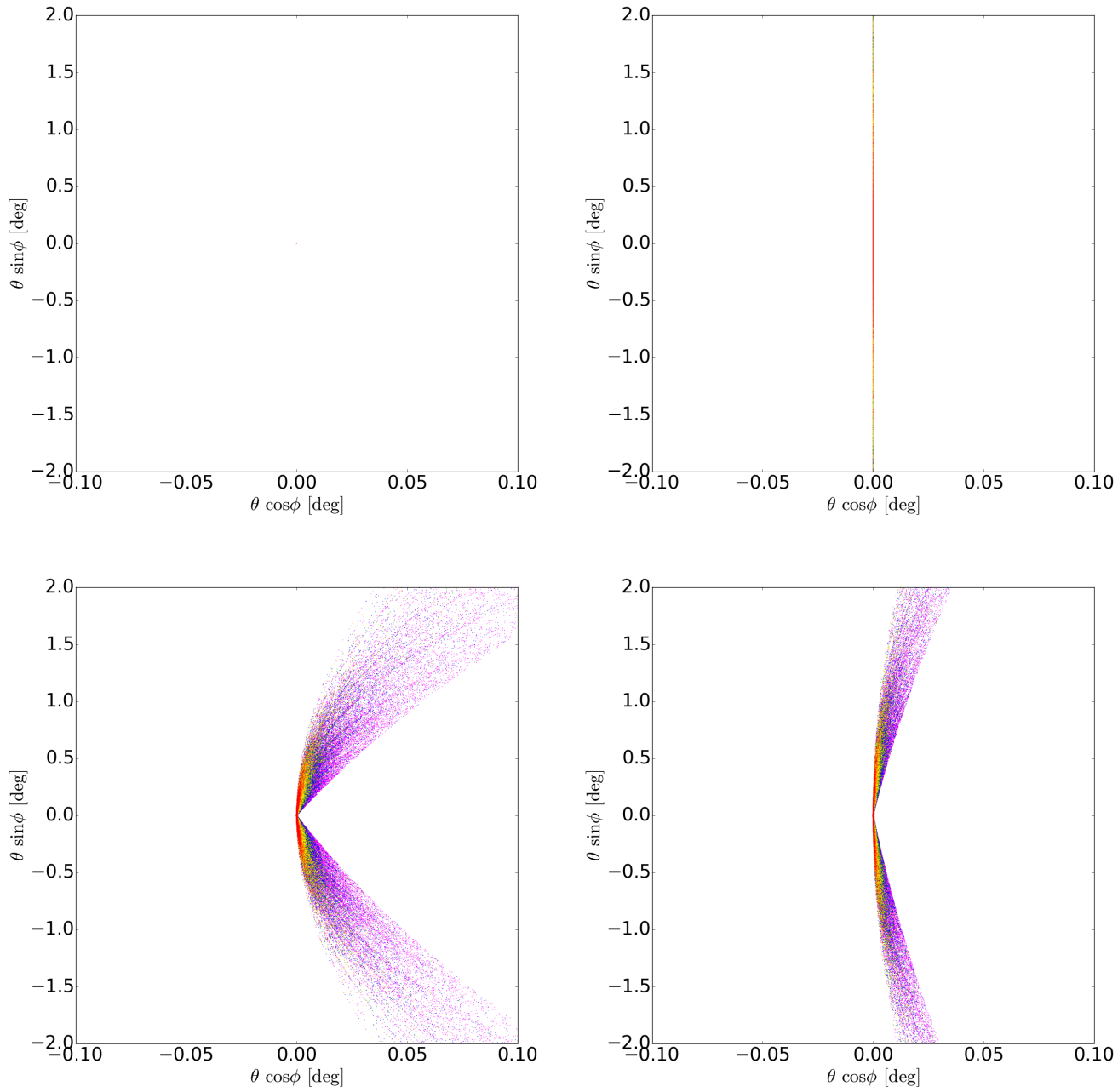

FIG. 2: Energy-dependent sky maps for a uniform magnetic field with $B = 10^{-15}$ G. We show the cases of a tightly collimated jet with magnetic field parallel (top left), perpendicular (top right), and tilted by 45 deg (bottom left) and 75 deg (bottom right) to the blazar jet direction which is taken to be along the line of sight. The different colors represent the following energy ranges: $5 - 10$ GeV (magenta), $10 - 15$ GeV (blue), $15 - 20$ GeV (green), $20 - 30$ GeV (yellow), $30 - 50$ GeV (orange), $50 - 100$ GeV (red).

this interpretation we found that in simulations in which the absolute value of $\mathbf{B} \cdot (\nabla \times \mathbf{B})$ is small close to the source, the spiral-like structures tend to be less distinct.

The sky map containing the arrival directions of gamma rays are shown in Fig. 3. We consider both the case for which the jet is directed along the line of sight (left column) and for which it is tilted by 5° (right column). For the former one can see the impact of magnetic helicity by comparing the top ($f_H = -1$) and bottom ($f_H = +1$) panels. A remarkable spiral-like pattern is clearly visible, being left- or right-handed depending on whether the helicity is negative or positive, respectively. For zero helicity ($f_H = 0$, middle panels), on the other hand, no clear orientation can be seen.

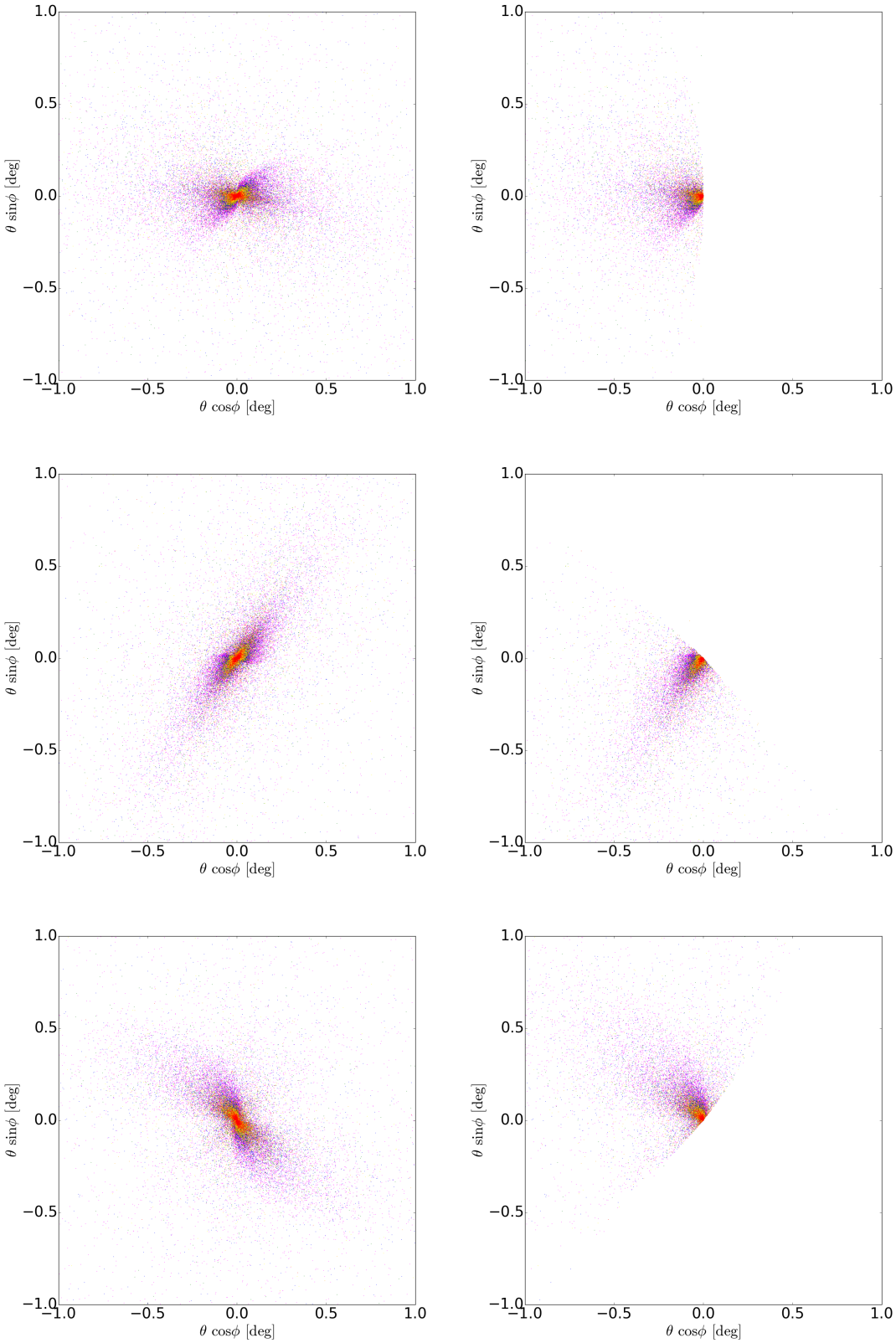

FIG. 3: Sky maps of arrival directions of photons from a blazar at a distance $D_{\rm s} = 1\,{\rm Gpc}$ emitting photons with energy $E_{\rm TeV} = 10\,{\rm TeV}$ in a jet with a half opening angle of $\Psi = 5^\circ$ directed at the observer (left column) and tilted by 5° with respect to the line of sight (right column), respectively. The magnetic field is assumed to be stochastic with RMS strength of $B = 10^{-15}\,{\rm G}$, coherence length $L_{\rm c} \simeq 120\,{\rm Mpc}$, and maximal negative (upper panels, $f_H = -1$), null (central, $f_H = 0$) and maximal positive (lower panels, $f_H = +1$) helicities, respectively. The colors represent the same energies as in Fig. 2.

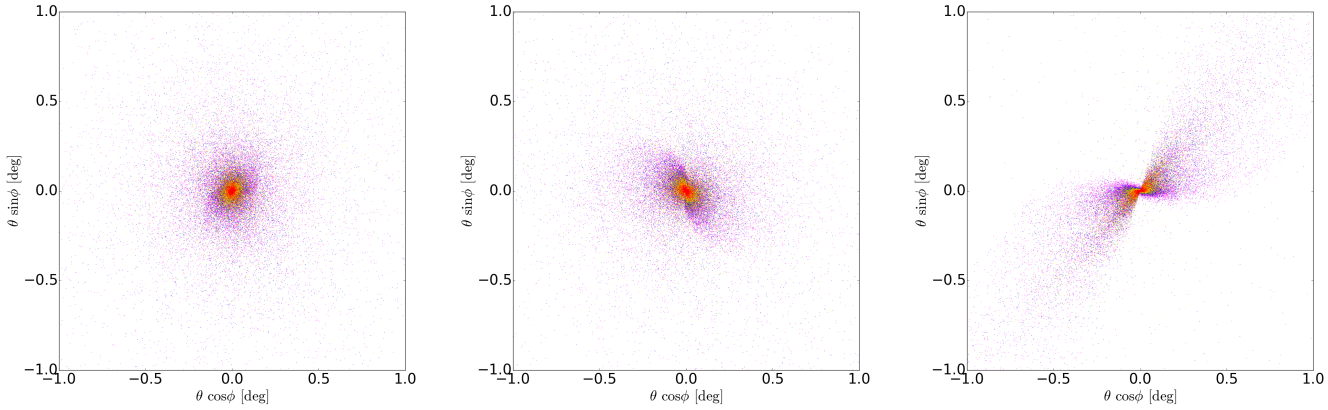

FIG. 4: Sky maps of arrival directions of photons from a blazar at a distance $D_{\rm s} = 1\,{\rm Gpc}$ emitting photons with energy $E_{\rm TeV} = 10\,{\rm TeV}$ in a jet with a half opening angle of $\Psi = 5^\circ$ directed at the observer. The magnetic field is assumed to be stochastic with RMS strength of $B = 10^{-15}\,{\rm G}$ and a coherence length of $L_{\rm c} \simeq 50\,{\rm Mpc}$ (left), $L_{\rm c} \simeq 150\,{\rm Mpc}$ (center) and $L_{\rm c} \simeq 250\,{\rm Mpc}$ (right) for $f_H = +1$. The colors represent the same energies as in Fig. 2.

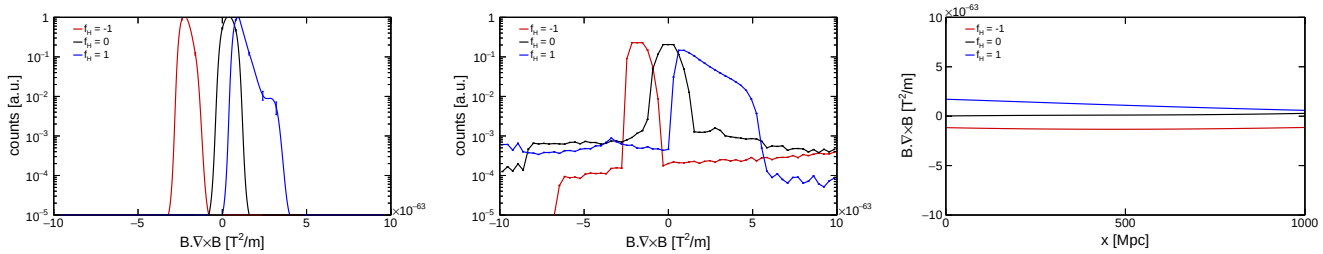

FIG. 5: Different magnetic helicity measures for the three cases shown in Fig. 3, *i.e.* negative helicity ($f_{\rm H} = -1$, red), zero helicity ($f_{\rm H} = 0$, black) and positive helicity ($f_{\rm H} = +1$, blue). The left panel shows the total distribution of "physical helicity", defined as $\mathbf{B} \cdot (\nabla \times \mathbf{B})$, in the whole simulation box, normalized to 1. In the center panel the same measure is shown, however restricted only to the line of sight and the neighboring cells. Finally, the right panel shows the helicity values along the line of sight from the source (at $x = 0\,{\rm Mpc}$) to the observer (at $x = 1000\,{\rm Mpc}$).

We show here the results for $L_{\rm c} \simeq 120\,{\rm Mpc}$. For lower coherence lengths ($L_{\rm c} \lesssim 50\,{\rm Mpc}$) and $B \lesssim 10^{-15}{\rm G}$ we find that the arrival direction pattern is washed out, and it is not possible to infer the presence of helicity, thus confirming the analytical predictions of Ref. [39] for this combination of parameters using simulations. This can be seen in Fig. 4 where the results for different $L_{\rm c}$ and $f_H = +1$ are shown. While for $L_{\rm c} = 250\,{\rm Mpc}$ a clear characteristic spiral in the arrival directions can be seen, it becomes less visible for $L_{\rm c} = 150\,{\rm Mpc}$ and disappears for $L_{\rm c} = 50\,{\rm Mpc}$. Therefore, $L_{\rm c} = 120\,{\rm Mpc}$ is a reasonable choice in order to show the effects of helicity discussed below. It is also a valid value in certain magnetogenesis scenarios [54].

To understand the dependence of the spiral pattern on the coherence scale, we note that, for small coherence lengths, the spirals become too tight to be resolved, *i.e.* their angular size becomes too small compared to the overall halo [39]. It seems, however, that the quality of the spiral might be highly sensitive to the specific values of the parameters of the setting such as B, $D_{\rm s}$ and $L_{\rm c}$ which we will further investigate in the future.

On the other hand, for larger coherence lengths the spirals tend to a straight line, similarly to the top right panel of Fig. 2, approaching the case of a simple uniform magnetic field. This, again, is rather intuitive, since if $L_{\rm c} \gtrsim D_{\rm s}$, the stochastic magnetic field will effectively be uniform on the length scales in question.

On the right hand side of Fig. 3 we show the same scenario described above, but this time the direction of the jet is tilted by 5° with respect to the line of sight. As one can see in the figure, this reduces the effective area of arrival directions and also the symmetry of the pattern. In our example, for instance, one of the "arms" of the spiral pattern or a part of it is removed. This enables us to apply the Q-statistics [37] (discussed below) to relate the helicity of the field with the arrival directions of gamma rays. It should be noted that all findings of this and the previous sections are in good agreement with the analytic predictions of Ref. [39].

D. Computing the Q-factors

One possibility to quantify the role of magnetic helicity is to use the Q-statistics, introduced in Refs. [36–38]. The key elements here are the observed energies and the arrival directions of gamma rays at Earth. For sets of photons with energies E_1, E_2 and E_3 with $E_1 < E_2 < E_3$, the Q-statistics is given by [37]

$$Q(E_1, E_2, E_3, R) = \frac{1}{N_3} \sum_{j=1}^{N_3} [\boldsymbol{\eta}_{1j}(R) \times \boldsymbol{\eta}_{2j}(R)] \cdot \mathbf{n}_j(E_3)\,, \tag{29}$$

where $\mathbf{n}_j(E_a)$ is the arrival direction of the j-th photon with energy E_a, N_a is the total number of photons of energy E_a, and $\boldsymbol{\eta}_{aj}(R)$ is given by

$$\boldsymbol{\eta}_{aj}(R) \equiv \frac{1}{N_a} \sum_{i \in \mathcal{D}_a(\mathbf{n}_j(E_3), R)} \mathbf{n}_i(E_a)\,, \tag{30}$$

where $\mathcal{D}_a(\mathbf{n}_j(E_3), R)$ represents the set of photons of energy E_a that are located in a disk of radius R centered on $\mathbf{n}_j(E_3)$. Essentially, the Q-statistics is the average value of the triple product of photon arrival vectors of energies E_1, E_2, E_3 that lie within an angle R of the highest energy photon (E_3).

As has been shown in Refs. [36–38], the calculation of the parity-odd statistics, or Q-statistics, should enable us, depending on the sign and general shape of the Q-factors for different values of E_1, E_2, E_3, and R, to draw conclusions about the helicity of the intervening helical magnetic field.

We now use Eq. (29) to calculate the Q-factors for the three helicity scenarios analyzed ($f_H = -1$, 0 and 1). We display the results for the case of tilted jets (*i.e.* the scenario shown in the right panel of Fig. 3) in Fig. 6. We consider triplets of energies (E_1, E_2, E_3) as needed for Eq. (29), where each energy E_i corresponds to an interval $[E_i, E_i + 10\,\mathrm{GeV}]$.

The reason we consider the scenario of *tilted* jets is that this is the most probable case – it is very unlikely for the blazar jet to be directed exactly along the line of sight. As discussed in Ref. [38], the function $Q(R)$ is expected to start at the origin since the angular deflections are small for small R. For larger R, the magnetic helicity causes Q to grow, and at much larger R, Q will approach a constant value (Q_∞) as there are no more photons to include at such large R. The large R behavior gets modified in a realistic setting where, in addition to the blazar photons, we also observe background photons from other sources. Then, for large R, the blazar contribution gets diluted by the background noise and Q decreases to zero. In this case, we would see a peak in $Q(R)$ whose position is set by the relative number of blazar to background photons. In our simulations, however, we do not include background photons and indeed find $Q \to Q_\infty$ at large R.

E. Computing the S-Statistics

As the last part of our results we present a new alternative way to quantify the pattern of gamma ray arrival directions and thus, indirectly, the helicity orientation. The idea underlying idea of this new method that we denote S-Statistics (for "Spiral") is that a gamma ray from a cascade that has a greater deflection away from the source direction due to the magnetic field will also have a greater azimuthal deflection if the magnetic field is helical. The pattern of observed gamma rays will have a spiral structure that can be measured by finding the average deflection of gamma rays, $\overline{\theta}(\phi,E_\gamma)$, as a function of the azimuthal angle ϕ and the considered gamma ray energy E_γ. We assume that there is at least one angle $\phi_{\rm max}$ for which $\overline{\theta}(\phi, E_\gamma)$ has a well-defined and significant maximum, *i.e.* a maximum from an average in a bin which has a reasonable number of photons *and* is statistically significant. We consider events inside a band around $\phi_{\rm max}$ with width $2\Delta\phi$, *i.e.* events with $\phi_{\rm max} - \Delta\phi \leq \phi \leq \phi_{\rm max} + \Delta\phi$. For a right-handed spiral there should be higher values of $\overline{\theta}(\phi, E_\gamma)$ for $\phi < \phi_{\rm max}$ than for $\phi > \phi_{\rm max}$ inside the band, while for a left-handed pattern $\overline{\theta}(\phi, E_\gamma)$ should be smaller for $\phi < \phi_{\rm max}$ than for $\phi > \phi_{\rm max}$. In other words, the peak of the function $\bar{\theta}(\phi)$ should be skewed to the right or to the left depending on whether the spiral is right- or left-handed, respectively. By finding a measure for this asymmetry or skew of the maximum one can deduce the orientation and subsequently the magnetic helicity.

More concretely, the calculation is performed in the following way: first, we subdivide the interval on which ϕ is defined, *i.e.* $\phi \in [0, 2\pi)$, in $n_{\rm bin}$ bins, such that each of the bins has a width $\delta\phi = 2\pi/n_{\rm bin}$. The jth bin, which corresponds to the interval $[(j-1)\delta\phi, j\delta\phi)$, $j = 1, ..., n_{\rm bin}$, will be labeled $\phi^{(j)} = (j-1)\delta\phi$. For each bin we calculate $\overline{\theta}$ by

$$\overline{\theta}(\phi^{(j)}, E_\gamma) = \frac{1}{N_j} \sum_{\{i | \phi^{(j)} \leq \phi_i < \phi^{(j+1)}\}} \theta_i\,, \tag{31}$$

where (ϕ_i, θ_i) are the coordinates of the ith event in the set $\{i | \phi^{(j)} \leq \phi_i < \phi^{(j+1)}\}$ and N_j is the total number of events in this set. If $\{i | \phi^{(j)} \leq \phi_i < \phi^{(j+1)}\}$ is empty, we set $\overline{\theta}(\phi^{(j)}, E_\gamma) = 0$. Furthermore, for real data it might be necessary to restrict the analysis to events with θ smaller than a certain value $\theta_{\rm max}$ as for $\theta > \theta_{\rm max}$ background photons might dominate and result in a false signal.

In this set of $\overline{\theta}(\phi^{(j)}, E_\gamma)$ one now has to identify the relevant and significant maxima as well as the corresponding bin number $j_{\rm max}$ and calculate the quantities

$$\Phi_- = \sum_{j=j_{\rm max}-\delta_{\rm bin}}^{j_{\rm max}-1} \overline{\theta}(\phi^{(j)}, E_\gamma),\ \Phi_+ = \sum_{j=j_{\rm max}+1}^{j_{\rm max}+\delta_{\rm bin}} \overline{\theta}(\phi^{(j)}, E_\gamma), \tag{32}$$

where $\delta_{bin} \geq 1$ is the number of bins we need to consider in order to include the width of the peak. Here one has

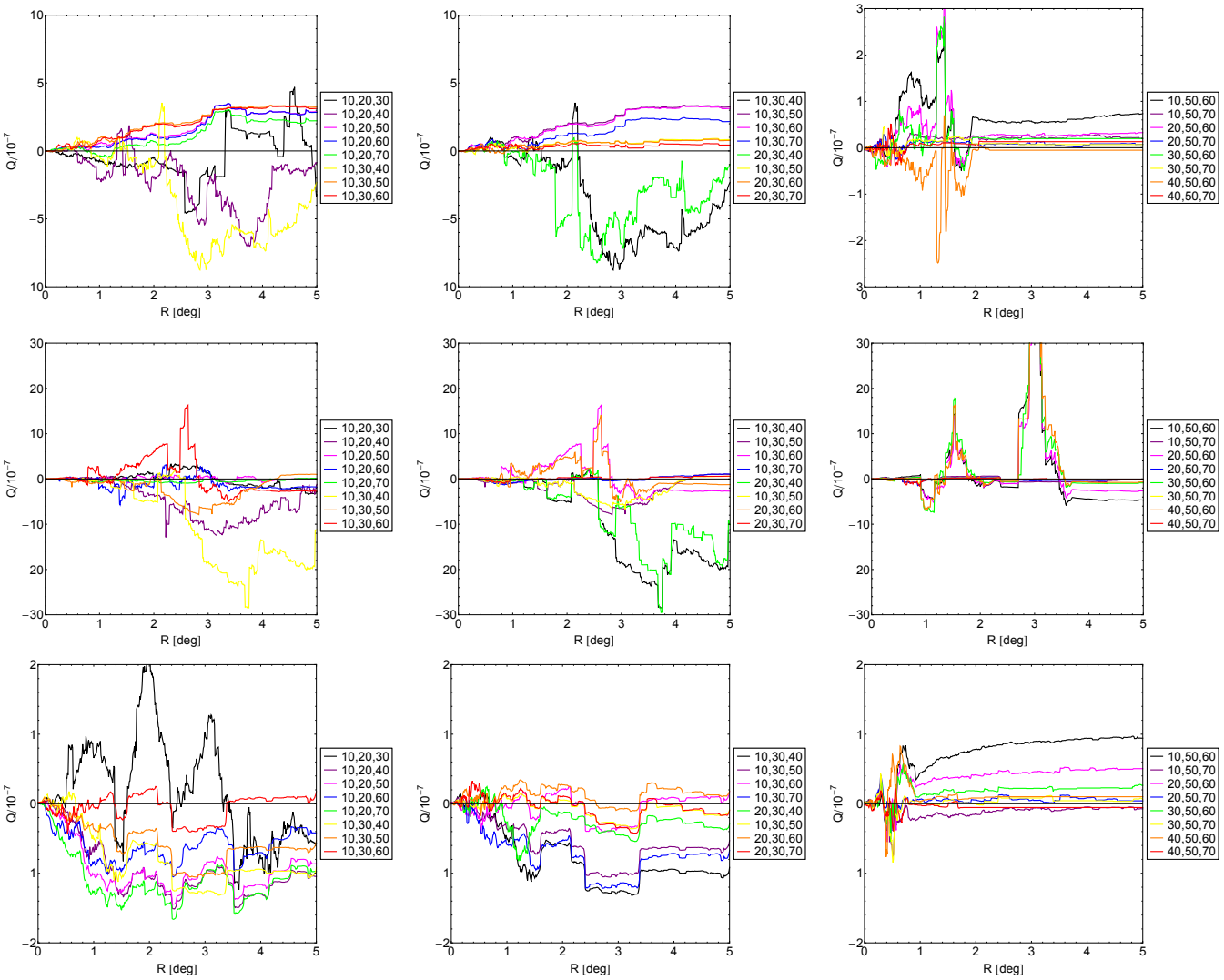

FIG. 6: Q-statistics for the case of a source tilted 5° with respect to the line of sight, for $D_{\rm s} = 1\,{\rm Mpc}$, $E_{\rm TeV} = 10\,{\rm TeV}$ and $\Psi = 5°$. All panels correspond to the three right hand panels of Fig. 3, *i.e.* $f_{\rm H} = -1$ at the top, $f_{\rm H} = 0$ in the middle $f_{\rm H} = +1$ at the bottom panel. The triplets in the legends correspond to E_1, E_2, E_3 in GeV, each in intervals of $[E_i, E_i + 10\,{\rm GeV}]$.

to assume periodicity, *i.e.* $\overline{\theta}(\phi^{(j+n_{\rm bin})}) = \overline{\theta}(\phi^{(j)}, E_\gamma)$. Essentially Φ_- corresponds to the average value of $\overline{\theta}$ to the left of the peak and Φ_+ to the right of the peak.

The final step is to define the S-statistics that measures the skewness of the peak

$$S \equiv \frac{\Phi_- - \Phi_+}{\Phi_- + \Phi_+} \tag{33}$$

For a right-handed spiral S will be positive, whereas for a left-handed spiral it will be negative.

We performed this computation for the data shown in Fig. 3. The plots for $\overline{\theta}(\phi, E_\gamma)$, shown for different energies, are presented in Fig. 7. Even without any further analysis one can see in this figure that the peaks for opposite orientations indeed show opposite skews – while on the left panel higher angles are achieved for $\phi > \phi_{\rm max}$, on the right panel, even more clearly, that is the case for $\phi > \phi_{\rm max}$. For the central panel, however, peaks of either skewness are found.

In order to support these qualitative considerations, one has to look at the S-values which have been calculated and are presented for the three cases in Tabs. I–III. The most clear case here is the one for $f_{\rm H} = -1$, where for all energy ranges we obtain $S < 0$ with two values even going as low as $S \simeq -0.5$. This means that the morphology of the arrival directions is solely right-handed which is also clearly seen in the top panel of Fig. 3. For the case of $f_{\rm H} = +1$ the situation is less clear as there is only one value being as high as $S \simeq +0.48$. Nevertheless, since S is positive for all energy ranges with three exceptions for which, however the absolute value of S is close to zero, this is strong evidence for a left-handed orientation. Finally, no clear statement can be made regarding the case with no helicity ($f_{\rm H} = 0$) – here one does not

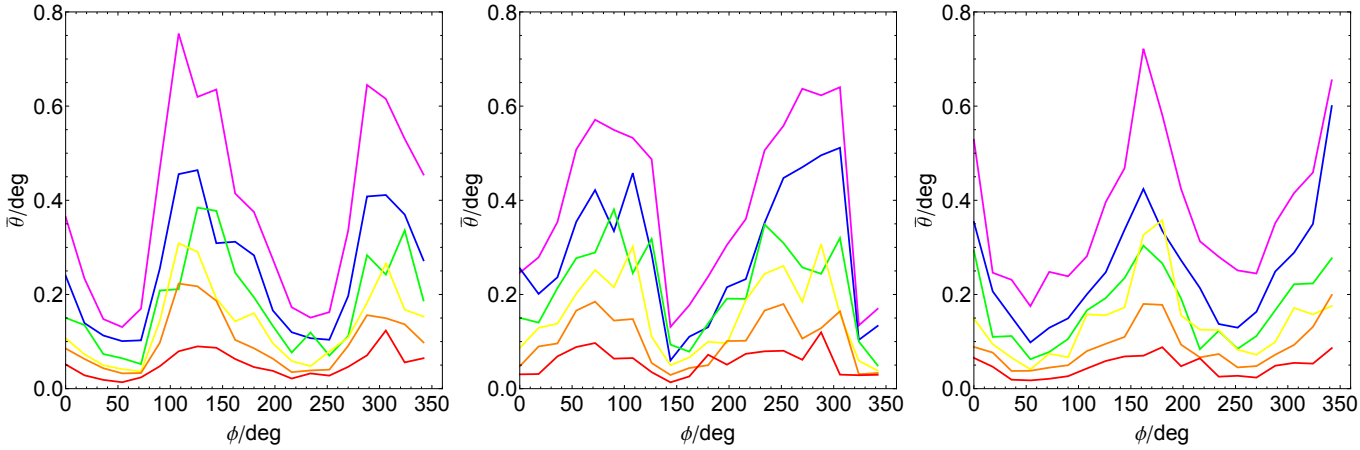

FIG. 7: The average polar angle $\overline{\theta}$ in dependence on the azimuthal angle ϕ, as calculated in (31) for the three cases $f_{\rm H} = -1$ (left), $f_{\rm H} = 0$ (center) and $f_{\rm H} = +1$ (right), corresponding to the three cases in Fig. 3. Here we have chosen $n_{\rm bin} = 20$. The colors correspond to the energy ranges of the arrival energies E_γ in the same way as in Fig. 2, *i.e.* $5 - 10\,\text{GeV}$ (magenta), $10 - 15\,\text{GeV}$ (blue), $15 - 20\,\text{GeV}$ (green), $20 - 30\,\text{GeV}$ (yellow), $30 - 50\,\text{GeV}$ (orange), $50 - 100\,\text{GeV}$ (red).

find significant negative or positive values for S.

As the last step, we need to connect the handedness of the arrival direction pattern with the sign of helicity. Their correlation has been found in Ref. [39], where an analysis has been carried out for a homogeneous magnetic field. In this reference the authors indeed find that for a positive helicity one expects a right-handed orientation, whereas for negative helicities a left-handed orientation should be observed, thus confirming our results for stochastic fields.

As a concluding remark it should be stated that for the results presented above we used simulations containing approximately 1.4×10^5 photons arriving at Earth in the energy range $1.5 \leq E_\gamma/\text{GeV} \leq 100$, which provided clear patterns with satisfactory statistical significance. The upcoming Cherenkov Telescope Array (CTA) [55] might be able to detect this kind of signature in the energy range $E \simeq 10 - 100$ GeV with $\gtrsim 10$ hours of observation. Fewer photons would distort the picture since, for example in the case of the S-statistics, the peaks would become less visible, such that a reliable calculation would no longer be possible. This is the case for high energy photons as their contribution to the total flux is rather small. On the other hand, for the lowest energies ($\sim$ a few GeV), even with as few as 10^4 photons relevant peaks can be seen, which, however, might be more difficult to construe in a more realistic case considering additionally diffuse gamma ray radiation and multiple sources.

IV. DISCUSSION AND OUTLOOK

We have performed three-dimensional Monte Carlo studies of the development of gamma-ray-induced electromagnetic cascades in the intergalactic medium in the presence of magnetic fields. We have used the "Large Sphere Observer" method for improved computational

TABLE I: Table for S for maximal negative helicity ($f_{\rm H} = -1$).

E_γ/GeV	$j_{\max}$	$\phi_{\max}$/deg	Φ_-	Φ_+	S
5–10	6	108.0	0.768	1.67	-0.37
	16	288.0	0.649	1.60	-0.42
10–15	7	126.0	0.813	0.904	-0.05
	17	306.0	0.709	0.882	-0.11
15–20	7	126.0	0.472	0.818	-0.27
	18	324.0	0.637	0.473	0.15
20–30	6	108.0	0.222	0.625	-0.48
	17	306.0	0.370	0.428	-0.07
30–50	6	108.0	0.163	0.507	-0.51
	16	288.0	0.170	0.385	-0.39
50–100	7	126.0	0.151	0.200	-0.13
	17	306.0	0.145	0.172	-0.09

TABLE II: Table for S for vanishing helicity ($f_{\rm H} = 0$).

E_γ/GeV	$j_{\max}$	$\phi_{\max}$/deg	Φ_-	Φ_+	S
5–10	4	72.0	1.69	12.12	-0.11
	17	306.0	2.99	1.69	+0.28
10–15	6	108.0	1.80	1.04	+0.27
	17	306.0	2.21	1.28	+0.27
15–20	5	90.0	1.12	1.07	+0.02
	13	234.0	1.01	1.28	-0.12
20–30	6	108.0	1.02	0.610	+0.25
	16	288.0	1.07	0.608	+0.28
30–50	4	72.0	0.463	0.470	-0.01
	14	252.0	0.491	0.511	-0.02
50–100	4	72.0	0.277	0.275	+0.00
	16	288.0	0.418	0.218	+0.31

TABLE III: Table for S for maximal positive helicity ($f_{\rm H} = +1$).

E_γ/GeV	$j_{\rm max}$	$\phi_{\rm max}$/deg	Φ_-	Φ_+	S
5–10	19	342.0	1.23	1.01	+0.10
	9	162.0	1.15	1.32	-0.07
10–15	19	342.0	0.888	0.713	+0.11
	9	162.0	0.785	0.822	-0.02
15–20	0	0.0	0.722	0.284	+0.44
	9	162.0	0.594	0.542	+0.05
20–30	19	342.0	0.428	0.309	+0.16
	10	180.0	0.655	0.405	+0.24
30–50	19	342.0	0.296	0.203	+0.19
	9	162.0	0.285	0.338	-0.08
50–100	19	342.0	0.157	0.131	+0.09
	10	180.0	0.120	0.138	+0.18

performance. In this case all cascade photons hitting the surface of the sphere are detected by the the observer. With a standard three-dimensional Monte Carlo simulation most cascade photons would not reach Earth, resulting in wasted computation and very low statistics. A simplification made in our treatment is that the magnetic field evolves adiabatically with redshift as $B(z) = B(z=0)(1+z)^2$. This is justified because the cascade development we have discussed occurs in cosmic voids where MHD amplification and contamination by sources is minimal. Also, the sources are at redshifts $z \lesssim 1$.

We first compared our computational setup with analytical approximations and then validated it in simple scenarios containing a uniform magnetic field oriented parallelly and perpendicularly to the line of sight of the blazar jet. As expected, for a magnetic field parallel to the direction of the jet of half-opening angle Ψ, assumed to be pointing toward Earth, effects of the field were not observed. For a magnetic field perpendicular to the direction of the jet, deflections were non-zero and in the expected direction. Similar results were obtained for stronger and weaker magnetic fields and other orientations. These results are in accordance with Ref. [39] and also with the predictions of Eq. (26).

We have also studied the particular case of a magnetic field with a Batchelor power spectrum with and without helicity. The effects of helicity can be clearly seen in Fig. 3, where arrival directions follow right- or left-handed spirals, depending on the sign of the helicity. For stochastic fields, in general, the results tend to converge toward the case of a uniform magnetic field in the limit of large coherence lengths. We have considered only large values of correlation length ($L_{\rm c} \simeq 120\,{\rm Mpc}$) since for much smaller coherence lengths, with the other parameters being held fixed, no clear signature of helicity can be seen, as shown in Fig. 4. Nevertheless, one should bear in mind that the current upper limits of coherence length of magnetic fields in voids range between a few and hundreds of Mpc [54], placing the chosen value of 120 Mpc well within the allowed bounds.

We have deployed the so-called Q-statistics, a powerful analysis tool that makes it possible to determine the properties of magnetic helicity directly from the observables of gamma rays measured at Earth. In this work we for the first time applied Q-statistics to realistic three-dimensional simulations of electromagnetic cascades. Our results for Q are shown in Fig. 6. The plots do not show a strong correlation between Q and the existence and sign of the helicity. At the moment we cannot clearly state whether averaging over several objects will show a stronger correlation. We plan on investigating this issue in a future work.

It is important to stress the fact that Q-statistics might not be the final method to quantify magnetic helicity, however it is a good initial approach and has been used in several works (Refs. [36, 37, 39]) with satisfactory results. In this work we have, for the first time, introduced the S-statistics, which is a direct measure of the handedness of a pattern with respect to the line of sight. We have shown that the orientation, represented by the sign of S, is directly correlated with the sign of helicity. This shows that the S measure is also a powerful tool to be used in the analysis of helicity of IGMF.

Backgrounds at the $\sim$10-100 GeV energy range are expected due to secondary photons from AGN halos whose jet opening angles do not encompass the Earth. Other astrophysical sources of photons in this energy range also exist and have to be taken into account. In this first work we have neglected these backgrounds, which will be included in future studies.

We found that it is probably necessary to analyze various sources in order to make a definite statement about the sign of the helicity, since a clear signature cannot always be seen. In the future we will extend our simulations to the case of multiple sources and diffuse gamma rays. We expect to be able to reproduce actual detections and consequently retrieve more precise information about IGMF, which can be used to infer their origin and evolution.

In addition, we will extend the analysis by further exploring the parameter space as varying quantities such as the magnetic field strength $B_{\rm rms}$, the magnetic correlation length $L_{\rm c}$ and source parameters such as its distance from the observer, its energy spectrum or its cutoff energy, as they may be important in order to obtain a complete picture of their influence as discussed above and to explain actual observations.

Acknowledgments

R. A. B. acknowledges the financial support from the John Templeton Foundation. The work of A. S. has been supported by the DAAD funded by the BMBF and the EU Marie Curie Actions. A. S. would like to thank the

Arizona State University for the hospitality during his stay. Furthermore, A. S. is grateful to the "Helmholtz Alliance for Astroparticle Physics" (HAP) and the Collaborative Research Center SFB 676 "Particles, Strings and the Early Universe" for providing generous travel funds which were important for the collaboration leading to this work. T. V. is supported by DOE, Office of High Energy Physics, under Award number #DE-SC0013605 at ASU. T. V. is grateful to the Institute for Advanced Study, Princeton for hospitality while this work was being done. Special thanks go to Andrew J. Long for his ideas and discussions which helped to complete this work and to NORDITA in Stockholm for organizing the workshop "Origin, Evolution, and Signatures of Cosmological Magnetic Fields" during which important aspects of this work have been discussed.

[1] M. S. Turner and L. M. Widrow, Phys. Rev. D **37**, 2743 (1988).
[2] B. Ratra, Astrophys. J. **391**, L1 (1992).
[3] C. T. Byrnes, L. Hollenstein, R. K. Jain, and F. R. Urban, J. Cosmol. Astropart. Phys. **1203**, 009 (2012).
[4] R. J. Z. Ferreira, R. K. Jain, and M. S. Sloth, J. Cosmol. Astropart. Phys. **1406**, 053 (2014).
[5] T. Vachaspati, Phys. Lett. B **265**, 258 (1991).
[6] K. Enqvist and P. Olesen, Phys. Lett. B **319**, 178 (1993).
[7] G. Baym, D. Bödeker, and L. McLerran, Phys. Rev. D **53**, 662 (1996).
[8] D. Grasso and A. Riotto, Phys. Lett. B **418**, 258 (1998).
[9] C. J. Hogan, Phys. Rev. Lett. **51**, 1488 (1983). doi:10.1103/PhysRevLett.51.1488
[10] J. M. Quashnock, A. Loeb, and D. N. Spergel, Astrophys. J. Lett. **344**, L49 (1989).
[11] G. Sigl, A. V. Olinto, and K. Jedamzik, Phys. Rev. D **55**, 4582 (1997).
[12] A. G. Tevzadze, L. Kisslinger, A. Brandenburg, and T. Kahniashvili, Astrophys. J. **759**, 54 (2012).
[13] A. J. Long, E. Sabancilar, and T. Vachaspati, J. Cosmol. Astropart. Phys. **1402**, 036 (2014).
[14] R. M. Kulsrud, R. Cen, J. P. Ostriker, and D. Ryu, Astrophys. J. **480**, 481 (1997).
[15] D. A. Leahy and A. Vilenkin, Astrophys. J. **248**, 13 (1981).
[16] S. Bertone, C. Vogt, and T. Enßlin, Mon. Not. R. Astron. Soc. **370**, 319 (2006).
[17] J. Donnert, K. Dolag, H. Lesch, and E. Muller, Mon. Not. R. Astron. Soc. **392**, 1008 (2009).
[18] A. Neronov and D. V. Semikoz, Phys. Rev. D **80**, 123012 (2009).
[19] A. Neronov and D. V. Semikoz, JETP Lett. **85**, 473 (2007).
[20] W. Essey, S. Ando, and A. Kusenko, Astropart. Phys. **35**, 135 (2011).
[21] S. Ando and A. Kusenko, Astrophys. J. Lett. **722**, L39 (2010).
[22] F. Tavecchio *et al.*, Mon. Not. R. Astron. Soc. **406**, L70 (2010).
[23] A. M. Taylor, I. Vovk, and A. Neronov, Astron. Astrophys. **529**, A144 (2011).
[24] K. Takahashi *et al.*, Astrophys. J. **771**, L42 (2013).
[25] W. Chen, J. H. Buckley, and F. Ferrer, Phys. Rev. Lett. **115**, 211103 (2015).
[26] A. E. Broderick, P. Chang, and C. Pfrommer, Astrophys. J. **752**, 22 (2012).
[27] R. Schlickeiser, D. Ibscher, and M. Supsar, Astrophys. J. **758**, 102 (2012).
[28] R. Schlickeiser, S. Krakau, and M. Supsar, Astrophys. J. **777**, 49 (2013).
[29] A. Saveliev, C. Evoli, and G. Sigl, arXiv:1311.6752 [astro-ph.HE] (2013).
[30] P. Chang *et al.*, Astrophys. J. **797**, 110 (2014).
[31] W. Chen *et al.*, Mon. Not. R. Astron. Soc. **450**, 3371 (2015).
[32] G. Sigl, Phys. Rev. D **66**, 123002 (2002).
[33] A. Alexakis, P. D. Mininni, and A. Pouquet, Astrophys. J. **640**, 335 (2006).
[34] A. Saveliev, K. Jedamzik, and G. Sigl, Phys. Rev. D **87**, 123001 (2013).
[35] T. Kahniashvili and T. Vachaspati, Phys. Rev. D **73**, 063507 (2006).
[36] H. Tashiro and T. Vachaspati, Phys. Rev. D **87**, 123527 (2013).
[37] H. Tashiro, W. Chen, F. Ferrer, and T. Vachaspati, Mon. Not. R. Astron. Soc. **445**, L41 (2013).
[38] H. Tashiro and T. Vachaspati, Mon. Not. R. Astron. Soc. **448**, 299 (2015).
[39] A. J. Long and T. Vachaspati, J. Cosmol. Astropart. Phys. **1509**, 065 (2015).
[40] T. Vachaspati, Phys. Rev. Lett. **87**, 251302 (2001).
[41] C. Caprini and L. Sorbo, J. Cosmol. Astropart. Phys. **1410**, 056 (2014).
[42] F. Aharonian, P. S. Coppi, and H. J. Völk, Astrophys. J. Lett. **423**, L5 (1994).
[43] R. Plaga, Nature **374**, 430 (1994).
[44] R. Alves Batista *et al.*, J. Cosmol. Astropart. Phys. **1605**, 038 (2016).
[45] M. Kachelrieß, S. Ostapchenko, and R. Tomàs, Comp. Phys. Commun. **183**, 1036 (2012).
[46] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons. The Relativistic Quantum Field Theory of Charged Particles with Spin One-Half*, 2nd ed. (Springer, New York, 1976).
[47] T. M. Kneiske and H. Dole, Astron. Astrophys. **515**, A19 (2010).
[48] H. Tashiro, T. Vachaspati, and A. Vilenkin, Phys. Rev. D **86**, 105033 (2012).
[49] A. Brandenburg and K. Subramanian, Phys. Rep. **417**, 1 (2005).
[50] T. Kahniashvili, A. G. Tevzadze, A. Brandenburg, and A. Neronov, Phys. Rev. D **87**, 083007 (2013).
[51] A. Elyiv, A. Neronov, and D. V. Semikoz, Phys. Rev. D **80**, 023010 (2009).
[52] A. Neronov *et al.*, Astrophys. J. Lett. **719**, L130 (2010).
[53] D. Harari, S. Mollerach, E. Roulet, and F. Sanchez, J. High Energy Phys. **03**, 045 (2002).
[54] R. Durrer and A. Neronov, Astron. Astrophys. Rev. **21**, 1 (2013).
[55] B. S. Acharya *et al.*, Astropart. Phys. **43**, 3 (2013).

Creation of Magnetic Monopoles in Classical Scattering

Tanmay Vachaspati
Physics Department, Arizona State University, Tempe, AZ 85287, USA.

We consider the creation of 't Hooft-Polyakov magnetic monopoles by scattering classical wave packets of gauge fields. An example with eight clearly separated magnetic poles created with parity violating helical initial conditions is shown. No clear separation of topological charge is observed with corresponding parity symmetric initial conditions.

Magnetic monopoles are of key interest in current research as they embody non-perturbative aspects of field theories. Their rich physical and mathematical properties have inspired continued investigations ever since Dirac first proposed their existence (*e.g.* [1–4]). Dualities that relate the spectra of particles and magnetic monopoles can be an important element in solving strongly coupled problems [5, 6] and may also help understand the spectrum of fundamental particles [7, 8] In particle physics, monopoles necessarily arise in grand unified models of particle physics, and the standard electroweak model contains field configurations that correspond to confined monopoles [9].

The current investigation involves the interpretation of magnetic monopoles in terms of particles. Can we create magnetic monopoles by assembling particles? This problem is difficult because particles are the quanta in a *quantum* field theory and magnetic monopoles are classical objects in that field theory. No perturbative expansion of the quantum field theory in powers of coupling constants can describe magnetic monopoles because properties of the magnetic monopole are proportional to *inverse* powers of the coupling constant. (Recent work on resurgence in quantum mechanics [10] offers a glimmer of hope that divergences in the perturbative expansion may hold non-perturbative information.) A more modest objective is to study the creation of magnetic monopoles by scattering *classical* waves, where the classical waves can themselves be thought of as quantum states containing high occupation numbers of quanta. This is the approach we shall take.

Past work on the creation of kinks in 1+1 dimensions [11–17], on the decay of electroweak sphalerons [18, 19], and on the scattering and annihilation of magnetic monopole-antimonopole [20], together with results from magneto-hydrodynamics (MHD) [21], offers some guidance on initial conditions that may be suitable for creating magnetic monopoles. We will further explain these motivations when describing our initial conditions.

We will work with an SO(3) field theory, as considered by 't Hooft [22] and Polyakov [23], that contains a scalar field in the adjoint representation, ϕ^a ($a = 1, 2, 3$), and gauge fields, W^a_μ, with the Lagrangian

$$L = \frac{1}{2}(D_\mu\phi)^a(D^\mu\phi)^a - \frac{1}{4}W^a_{\mu\nu}W^{a\mu\nu} - \frac{\lambda}{4}(\phi^a\phi^a - \eta^2)^2 \quad (1)$$

where,

$$(D_\mu\phi)^a = \partial_\mu\phi^a - igW^c_\mu(T^c)^{ab}\phi^b \quad (2)$$

and the SO(3) generators are $(T^a)^{bc} = -i\epsilon^{abc}$. The gauge field strengths are defined by

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^{abc}W^b_\mu W^c_\nu. \quad (3)$$

Our numerical methods are borrowed from Numerical Relativity [24]. We use temporal gauge $W^a_0 = 0$ and treat $\Gamma^a \equiv \partial_i W^a_i$ as new variables whose evolution ensures that the Gauss constraints are satisfied. The resulting classical equations of motion that we want to solve are written as

$$\begin{aligned}\partial_t^2\phi^a &= \nabla^2\phi^a - g\epsilon^{abc}\partial_i\phi^b W^c_i - g\epsilon^{abc}(D_i\phi)^b W^c_i \\ &\quad -\lambda(\phi^b\phi^b - \eta^2)\phi^a - g\epsilon^{abc}\phi^b\Gamma^c \end{aligned} \quad (4)$$

$$\begin{aligned}\partial_t W^a_{0i} &= \nabla^2 W^a_i + g\epsilon^{abc}W^b_j\partial_j W^c_i - g\epsilon^{abc}W^b_j W^c_{ij} \\ &\quad -D_i\Gamma^a - g\epsilon^{abc}\phi^b(D_i\phi)^c \end{aligned} \quad (5)$$

$$\begin{aligned}\partial_t\Gamma^a &= \partial_i W^a_{0i} - g_p^2[\partial_i(W^a_{0i}) + g\epsilon^{abc}W^b_i W^c_{0i} \\ &\quad +g\epsilon^{abc}\phi^b(D_t\phi)^c] \end{aligned} \quad (6)$$

where $W^a_{0i} = \partial_t W^a_i$ in the temporal gauge, $D_i\Gamma^a \equiv \partial_i\Gamma^a - g\epsilon^{abc}\Gamma^b W^c_i$, and g_p^2 is a free parameter. Analytically, the square bracket in Eq. (6) vanishes due to the Gauss constraints and the value of g_p^2 is irrelevant. However the square bracket does not vanish when we discretize the system and a non-zero value of g_p^2 is critical to ensure numerical stability [24]. After some experimentation we set $g_p^2 = 0.75$ in our runs. We also set $g = 0.5$, $\lambda = 1$ and $\eta = 1$ in our numerical work.

The fields are evolved using the explicit Crank-Nicholson method with two iterations [25]. We have used a new implementation of absorbing boundary conditions. Essentially, only the Laplacian of the fields on the lattice boundaries are replaced using radially outgoing boundary conditions. For example,

$$\nabla^2\phi^a \to -\hat{r}\cdot\nabla(\partial_t\phi^a) \quad (7)$$

at a boundary point with $\hat{r}$ the unit radial vector from the center of the box. The first order spatial derivatives throughout the equations of motion are evaluated using one-sided differences. We have found good stability and smooth evolution with this strategy.

The non-algorithmic part of this project is to devise initial conditions that are likely to result in monopole creation. As noted in Ref. [15], a crucial hint comes from

the conservation of helicity in MHD in plasmas with high electrical conductivity. (Magnetic helicity is defined as the volume integral of $\boldsymbol{A}\cdot\boldsymbol{B}$ where $\boldsymbol{A}$ is the electromagnetic gauge potential and $\boldsymbol{B}=\nabla\times\boldsymbol{A}$.) Combined with the observed conservation of electromagnetic helicity during sphaleron decay [18, 19] and the repulsive force between monopoles and antimonopoles that are twisted and that yield magnetic helicity on annihilation [20], it seems like a good idea to try initial conditions that are built from helical, *i.e.* circularly polarized gauge waves. Also, MHD simulations indicate that helicity causes magnetic fields to expand out to larger length scales ("inverse cascade"), so that by colliding helical waves, helicity will get compressed, causing tension against the natural tendency to expand. This tension can relax if helicity conservation is violated, either with a decrease in the plasma electrical conductivity or by producing magnetic monopoles.

The natural way to discuss initial conditions is to first specify the value of the scalar field since this determines the massless and massive components of the gauge fields. In the numerics, however, it is easier to specify the gauge field and then make various choices for the uniform value of the scalar field, and this is how we will present the initial conditions.

We choose only one of the 3 SO(3) gauge fields to be non-trivial in the initial conditions. Let this be W_i^3. Initially, at $t=0$, W_i^3 is given separately for waves propagating in the $+z$ and $-z$ direction in terms of scalar functions $f_1(x,y)$, $f_2(t+(z-z_0))$ and $f_3(t-(z+z_0))$ with $z_0>0$. For the waves that are functions of $t+(z-z_0)$, we have:

$$\begin{aligned}
W_x^3 &= \partial_y f_1(\omega f_2-\partial_z f_2)\cos(\omega(t+(z-z_0))) && (8)\\
W_y^3 &= \partial_x f_1(\omega f_2+\partial_z f_2)\sin(\omega(t+(z-z_0))) && (9)\\
W_z^3 &= \partial_x\partial_y f_1 f_2[\cos(\omega(t+(z-z_0)))\\
&\qquad -\sin(\omega(t+(z-z_0)))] && (10)
\end{aligned}$$

In this form it is easy to see that $\nabla\cdot\boldsymbol{W}^3=0$. Then $\partial_t W_i^3=+\partial_z W_i^3$ gives

$$\begin{aligned}
\partial_t W_x^3 &= \partial_y f_1[(\omega\partial_z f_2-\partial_z^2 f_2)\cos(\omega(t+(z-z_0))\\
&\qquad -(\omega f_2-\partial_z f_2)\omega\sin(\omega(t+(z-z_0)))] && (11)\\
\partial_t W_y^3 &= \partial_x f_1[(\omega\partial_z f_2+\partial_z^2 f_2)\sin(\omega(t+(z-z_0))\\
&\qquad +(\omega f_2+\partial_z f_2)\omega\cos(\omega(t+(z-z_0)))] && (12)\\
\partial_t W_z^3 &=\\
&\partial_x\partial_y f_1[\partial_z f_2(\cos(\omega(t+(z-z_0)))-\sin(\omega(t+(z-z_0)))\\
&+\omega f_2(-\sin(\omega(t+(z-z_0)))-\cos(\omega(t+(z-z_0))))] && (13)
\end{aligned}$$

Since $\nabla\cdot\boldsymbol{W}^3=0$, and the electric field $\boldsymbol{E}^3=-\partial_t\boldsymbol{W}^3$, the Gauss constraint is satisfied with vanishing charge density. We will arrange for a vanishing charge density by taking the scalar field to have vanishing time derivative initially

$$\partial_t\phi^a|_{t=0}=0. \qquad (14)$$

We will also take $\phi^a=$ constant initially, with different choices for the constant describing different physical situations as discussed below.

For a packet traveling in the opposite direction, we write the formulae in terms of $f_3(t-(z+z_0))$:

$$\begin{aligned}
W_x^3 &= \partial_y f_1(-\omega' f_3-\partial_z f_3)\cos(\omega'(t-(z+z_0))) && (15)\\
W_y^3 &= -\partial_x f_1(\omega' f_3-\partial_z f_3)\sin(\omega'(t-(z+z_0))) && (16)\\
W_z^3 &= \partial_x\partial_y f_1 f_3\ (\cos(\omega'(t-(z+z_0)))\\
&\qquad -\sin(\omega'(t-(z+z_0)))) && (17)
\end{aligned}$$

For these packets we use $\partial_t W_i^3=-\partial_z W_i^3$ to write

$$\begin{aligned}
\partial_t W_x^3 &= -\partial_y f_1[(-\omega'\partial_z f_3-\partial_z^2 f_3)\cos(\omega'(t-(z+z_0)))\\
&\qquad -(\omega' f_3+\partial_z f_3)\omega'\sin(\omega'(t-(z+z_0)))] && (18)\\
\partial_t W_y^3 &= \partial_x f_1[(\omega'\partial_z f_3-\partial_z^2 f_3)\sin(\omega'(t-(z+z_0)))\\
&\qquad -(\omega' f_3-\partial_z f_3)\omega'\cos(\omega'(t-(z+z_0)))] && (19)\\
\partial_t W_z^3 &= -\partial_x\partial_y f_1[\partial_z f_3(\cos(\omega'(t-(z+z_0)))\\
&\qquad -\sin(\omega'(t-(z+z_0))))\\
&+\omega' f_3(\cos(\omega'(t-(z+z_0)))+\sin(\omega'(t-(z+z_0))))] && (20)
\end{aligned}$$

The profile functions are taken such as to create a localized packet in all directions

$$f_1(x,y)=a\ \exp\left[-\frac{(x^2+y^2)}{2w^2}\right] \qquad (21)$$

$$f_2(t+(z-z_0))=\exp\left[-\frac{(t+(z-z_0))^2}{2w^2}\right] \qquad (22)$$

$$f_3(t-(z+z_0))=\exp\left[-\frac{(t-(z+z_0))^2}{2w^2}\right] \qquad (23)$$

where a is an amplitude and w is a width. The frequencies ω and ω' can be different in general but we only consider $\omega'=\pm\omega$. The case $\omega'=\omega$ corresponds to scattering of left- and right-handed circular polarizations, while $\omega'=-\omega<0$ corresponds to scattering of left- on left-handed circular polarization waves.

Now we linearly superpose the counterpropagating wave packets and set $t=0$ to get the initial conditions for the gauge fields for our scattering experiments.

Next we discuss the choice of the scalar field ϕ^a. The simplest choice is $\phi^1=0=\phi^2$, $\phi^3=\eta$ but this is too simple. In this case, $\boldsymbol{W}^3$ corresponds to the massless "photon" of the model, and in this classical evolution, the scattering of photons does not excite any other field. In other words, the dynamics lies in a subspace of the full field theory [26] and the classical dynamics is exactly as it would be in Maxwell theory. The next choice we considered is $\phi^1=\eta$, $\phi^2=0=\phi^3$. Now $\boldsymbol{W}^3$ is a massive boson of the theory. This too leads to dynamics in a subspace, namely that spanned by $\{\phi^1,\phi^2,\boldsymbol{W}^3\}$. So now the model is effectively the Abelian-Higgs U(1) model. It is interesting that when we performed some runs with these initial conditions, we did observe zeros of ϕ^a, suggesting that we had created loops of strings. We will postpone this investigation for the future since here we are focusing on the production of magnetic monopoles.

For the classical dynamics to explore the full model, we take

$$\phi^1 = \frac{\eta}{\sqrt{2}},\ \phi^2 = 0,\ \phi^3 = \frac{\eta}{\sqrt{2}} \tag{24}$$

at $t = 0$. Now the initial gauge field wave packet is a superposition of the photon and the massive gauge boson.

After the system has evolved for a while, we would like to know if monopoles have been created. Since monopoles are stable objects and the scalar field vanishes at their centers, the existence of a monopole can be detected by looking for peaks of the potential energy density that are close to the value $\lambda\eta^4/4 = 0.25$. We follow the potential energy diagnostic with a calculation of the topological winding which is defined as

$$W(S) = \frac{1}{8\pi}\oint_S d\hat{n}^i \epsilon_{ijk}\epsilon_{abc}\hat{\phi}^a\partial_j\hat{\phi}^b\partial_k\hat{\phi}^c \tag{25}$$

where $\hat{n}$ is the outward unit normal to a closed surface S and $\hat{\phi}^a = \phi^a/|\vec{\phi}|$. We replace the continuum formula for the winding with a discrete formula as follows. We first define the vector, $\vec{v}$, at every vertex of the lattice,

$$v_i = \epsilon_{ijk}\epsilon_{abc}\hat{\phi}^a\partial_j\hat{\phi}^b\partial_k\hat{\phi}^c. \tag{26}$$

Then the winding for a fundamental cell of the simulation lattice is given by

$$\overline{W}(S) = \frac{1}{8\pi}\sum_{\text{plaq.}}\left(\frac{1}{4}\sum_{\text{vertices}}\hat{n}^i v_i\right) \tag{27}$$

where the outside sum is over the 6 plaquettes bounding a cell, $\hat{n}$ is the unit vector normal to the plaquette, v_i is the vector in Eq. (26) evaluated at the vertices of the plaquette, and the $1/4$ is due to an averaging over the 4 corners of the plaquette. Even though $W(S)$ takes integer values, the discrete version $\overline{W}(S)$ may not be an integer. However, for large surfaces S, $\overline{W}(S)$ will also tend towards an integer value.

Our simulations are run on a 128^3 lattice with lattice spacing $dx = 0.1$ with field theory parameters: $g = 0.5$, $\lambda = 1$, $\eta = 1$. The initial condition parameters were chosen to be: $w = 0.4$, $z_0 = 1$, $a = 10$, $\omega = 4$. With this choice of parameters, the initial energy is $\sim 10^5$ and is much larger than the energy per monopole-antimonopole pair, which is $\sim 10^2$. Further exploration of parameters and choice of initial conditions is likely to yield monopoles even when we start with less energy, though intuitively the initial conditions will have to be more finely tuned or "coherent" if we take lower initial energy.

The first indication that monopoles have been produced during evolution is that we see zeros of the Higgs. This is shown in Fig. 1.

The presence of monopoles is confirmed by finding the topological winding, $\overline{W}$ for every cell of the lattice. In Fig. 2 we show the distribution of topological charge on xy-slices, *i.e.* on $z =$ constant slices of the lattice. Only

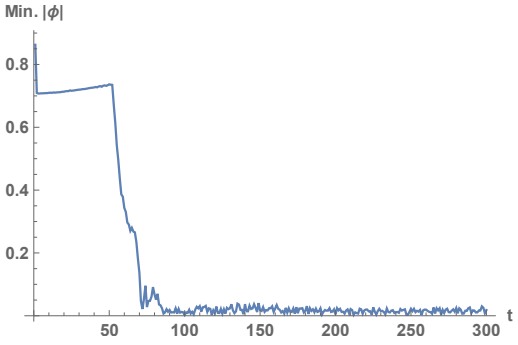

FIG. 1: Minimum value of $|\vec{\phi}|$ on the lattice as a function of time showing that zeros of the scalar field are produced after some evolution.

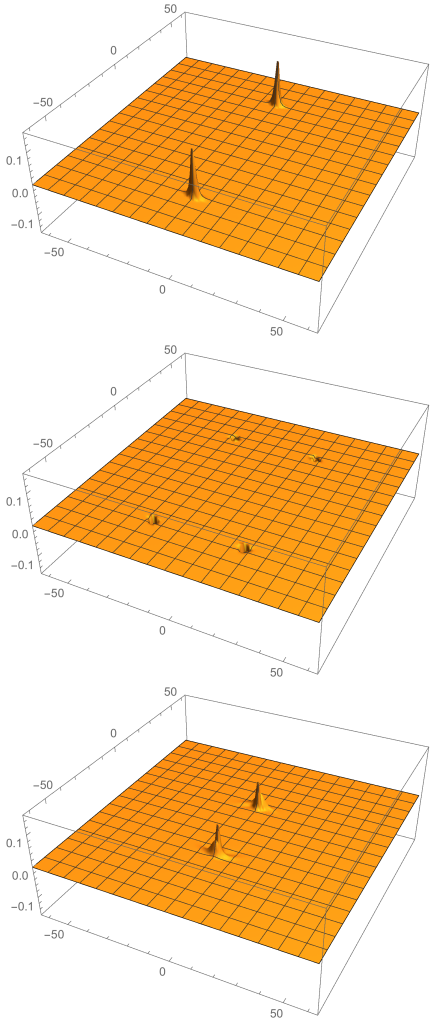

FIG. 2: Topological winding at late times on slices with $z = 9.2, 10.1$ and 12.1. The total topological charges on these slices are +2, -4, and +2 respectively.

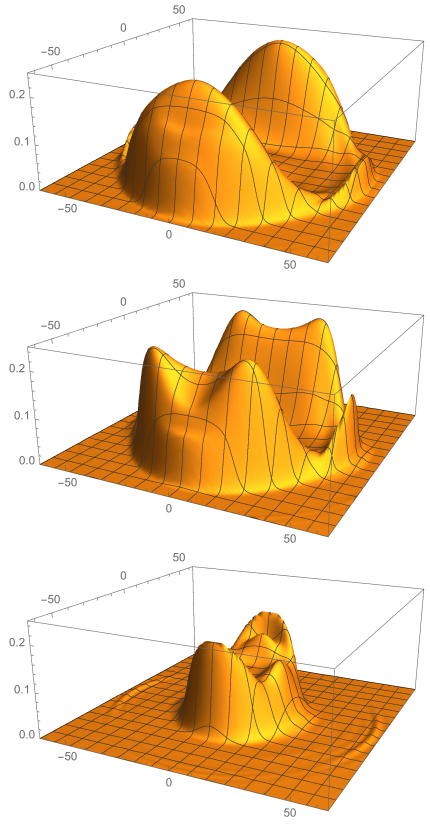

FIG. 3: Potential energy density distribution at the final time of the simulation on spatial slices with $z = 9.2, 10.1$ and 12.1. With $|\vec{\phi}| = 0$, the potential energy density is 0.25 for our parameters.

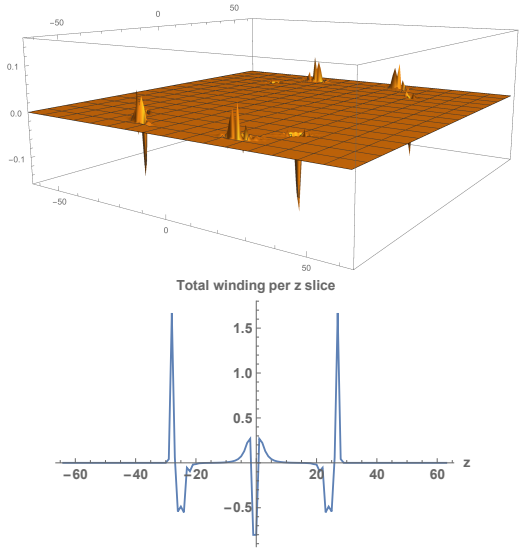

FIG. 4: Topological winding on the $z = 0$ slice for the $\omega = +\omega' = 4$ simulation. The plot does not show a clear separation of positive and negative winding. In the second panel we show the integrated winding per z slice as a function of z. Here too we do not see a clear separation of positive and negative charges.

slices with significant windings are shown and the total topological charge on the entire lattice vanishes. It is clear that the scattering has resulted in 4 monopoles and 4 antimonopoles. This is further confirmed by plotting the potential energy density on these slices, shown in Fig. 3. The peaks in the potential energy represent monopoles within which the scalar field has a zero. In the discrete simulation, the zero may lie within a cell of the lattice and the potential will not quite be its maximal value of 0.25.

The distances between monopoles and antimonopoles can be estimated and is on the order of 3 monopole widths where we take the monopole width to be the inverse scalar boson mass, $m_S = \sqrt{2\lambda}\eta$. We can estimate the velocities of the monopoles from Fig. 1 and our choice of time step $dt = dx/4$ where dx is the lattice spacing. We find that the monopoles are relativistic with $v \sim 1$. A simple estimate of the monopole-antimonopole escape velocity gives $v_{\rm esc} \sim 0.1$ when the separation of the pair is a few monopole widths. Since the monopole and antimonopole velocities are not aligned, the monopoles and antimonopoles are not bound and will continue to fly apart with time, as we observe directly during the later stages of the simulation.

A curious feature of the final configuration of monopoles is that they are all located at $z > 0$. However, this is not in contradiction with any symmetry, since our initial conditions for $\omega' = -\omega$ are not reflection symmetric under $z \to -z$.

We intend to automate the numerical program so that it can scan over a range of parameters and detect and record magnetic monopoles when they do occur. For the time being we have tried a few different values of the parameters and find monopole creation for larger values of the amplitude a and frequency ω. Of particular interest is the dependence on the sign of ω' that determines whether we are scattering left-on right-handed waves or left- on left-handed waves. The results discussed above are for $\omega = 4$, $\omega' = -4$ (left- on left-handed waves); so we also ran the code with $\omega' = +4$ and all other parameters kept the same. In Fig. 4 we show the topological winding distribution on the $z = 0$ slice. The sharp negative peaks signifying possible antimonopoles have positive peaks in their neighborhoods and the integrated charge vanishes. There are other peaks at non-zero z but these too have canceling charge distributions in their vicinity. The total topological charge per z slice is plotted in the second panel of Fig. 4 to further illustrate this feature. Hence, simply flipping the handedness of one of the initial waves results in evolution in which there is no clear separation of monopole and antimonopole charge.

The probability for creating monopoles depends on the probability measure on initial states and this depends on the human will to create such states. For example, the probability of creating a complex structure like the Large Hadron Collider by pure chance is incredibly low, nonetheless it exists. A more meaningful question is the

sensitivity of the outcome of the scattering to small errors in the initial conditions. Is the creation of monopoles a "stable" process? In the case of kinks in 1+1 dimensions, it is known that their scattering and annihilation is chaotic [27, 28]. This ties in with the chaotic behavior seen in the creation of kinks [11, 12] and it appears that the creation of kinks is very sensitive to the initial conditions, *i.e.* it is unstable. However, chaos seems to be absent in the annihilation of magnetic monopole and antimonopole, at least within the domain of scattering parameters that have been investigated [20]. This suggests that the creation of monopoles will also be a stable process but is something that needs to be investigated.

Acknowledgments

I thank Sourish Dutta, Jeff Hyde, Henry Lamm and Erick Weinberg for comments. I am grateful to the Institute for Advanced Study, Princeton, for hospitality during the course of this work. This work was also performed in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1066293. TV is supported by the U.S. Department of Energy, Office of High Energy Physics, under Award No. de-sc0013605 at ASU.

[1] C. Rebbi and G. Soliani, eds., *SOLITONS AND PARTICLES* (1985).

[2] J. Preskill, in *Architecture of Fundamental Interactions at Short Distances: Proceedings, Les Houches 44th Summer School of Theoretical Physics: Les Houches, France, July 1-August 8, 1985, pt1* (1987), pp. 235–338, URL http://alice.cern.ch/format/showfull?sysnb=0079532.

[3] A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and Other Topological Defects* (Cambridge University Press, 2000), ISBN 9780521654760, URL http://www.cambridge.org/mw/academic/subjects/physics/theoretical-physics-and-mathematical-physics/cosmic-strings-and-other-topological-defects?format=PB.

[4] N. S. Manton and P. Sutcliffe, *Topological solitons* (Cambridge University Press, 2007), ISBN 9780521040969, 9780521838368, 9780511207839, URL http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=0521838363.

[5] P. Goddard, J. Nuyts, and D. I. Olive, Nucl. Phys. **B125**, 1 (1977).

[6] N. Seiberg and E. Witten, Nucl. Phys. **B426**, 19 (1994), [Erratum: Nucl. Phys.B430,485(1994)], hep-th/9407087.

[7] T. Vachaspati, Phys. Rev. Lett. **76**, 188 (1996), hep-ph/9509271.

[8] H. Liu and T. Vachaspati, Phys. Rev. **D56**, 1300 (1997), hep-th/9604138.

[9] A. Achucarro and T. Vachaspati, Phys. Rept. **327**, 347 (2000), [Phys. Rept.327,427(2000)], hep-ph/9904229.

[10] G. V. Dunne and M. Unsal, Phys. Rev. **D87**, 025015 (2013), 1210.3646.

[11] S. Dutta, D. A. Steer, and T. Vachaspati, Phys. Rev. Lett. **101**, 121601 (2008), 0803.0670.

[12] T. Romanczukiewicz and Ya. Shnir, Phys. Rev. Lett. **105**, 081601 (2010), 1002.4484.

[13] S. V. Demidov and D. G. Levkov, Phys. Rev. Lett. **107**, 071601 (2011), 1103.0013.

[14] S. V. Demidov and D. G. Levkov, JHEP **06**, 016 (2011), 1103.2133.

[15] T. Vachaspati, Phys. Rev. **D84**, 125003 (2011), 1109.1065.

[16] H. Lamm and T. Vachaspati, Phys. Rev. **D87**, 065018 (2013), 1301.4980.

[17] S. V. Demidov and D. G. Levkov, JHEP **11**, 066 (2015), 1509.07125.

[18] C. J. Copi, F. Ferrer, T. Vachaspati, and A. Achucarro, Phys. Rev. Lett. **101**, 171302 (2008), 0801.3653.

[19] Y.-Z. Chu, J. B. Dent, and T. Vachaspati, Phys. Rev. **D83**, 123530 (2011), 1105.3744.

[20] T. Vachaspati, Phys. Rev. **D93**, 045008 (2016), 1511.05095.

[21] A. R. Choudhuri, *The physics of fluids and plasmas : an introduction for astrophysicists /* (1998).

[22] G. 't Hooft, Nucl. Phys. **B79**, 276 (1974).

[23] A. M. Polyakov, JETP Lett. **20**, 194 (1974), [Pisma Zh. Eksp. Teor. Fiz.20,430(1974)].

[24] T. W. Baumgarte and S. L. Shapiro, *Numerical Relativity: Solving Einstein's Equations on the Computer* (2010).

[25] S. A. Teukolsky, Phys. Rev. **D61**, 087501 (2000), gr-qc/9909026.

[26] M. Barriola, T. Vachaspati, and M. Bucher, Phys. Rev. **D50**, 2819 (1994), hep-th/9306120.

[27] D. K. Campbell, J. F. Schonfeld, and C. A. Wingate, Physica **9D**, 1 (1983).

[28] P. Anninos, S. Oliveira, and R. A. Matzner, Phys. Rev. **D44**, 1147 (1991).

Vortex structure in superfluid color-flavor locked quark matter

Mark G. Alford[1,⋆], *S. Kumar* Mallavarapu[1,⋆⋆], *Tanmay* Vachaspati[2,⋆⋆⋆], and *Andreas* Windisch[1,⋆⋆⋆⋆]

[1] *Physics Department, Washington University, St. Louis, MO 63130, USA*
[2] *Physics Department, Arizona State University, Tempe, AZ 85287, USA*

Abstract. The core region of a neutron star may feature quark matter in the color-flavor-locked (CFL) phase. The CFL condensate breaks the baryon number symmetry, such that the phenomenon of superfluidity arises. If the core of the star is rotating, vortices will form in the superfluid, carrying the quanta of angular momentum. In a previous study we have solved the question of stability of these vortices, where we found numerical proof of a conjectured instability, according to which superfluid vortices will decay into an arrangement of so-called semi-superfluid fluxtubes. Here we report first results of an extension of our framework that allows us to study multi-vortex dynamics. This will in turn enable us to investigate the structure of semi-superfluid string lattices, which could be relevant to study pinning phenomena at the boundary of the core.

1 Introduction

In its densest form, matter appears in the color-flavor locked (CFL) phase [1]. The CFL condensate breaks the baryon number symmetry, which renders this phase a superfluid. If this form of matter is present in the core region of a rotating neutron star, vortices will carry the angular momentum of the spinning core. In a recent study [2], we addressed the question of stability of these superfluid vortices, using a Ginzburg-Landau effective theory. There, from a topological point of view, stable vortex solutions are expected, since the first homotopy group of the vacuum manifold is non-trivial [3]. However, the global vortex solution does *not* possess the lowest energy, and it has been conjectured that the vortex undergoes a decay into a configuration of a triplet of well-separated semi-superfluid strings [4]. In our study [2], we not only observed and numerically confirmed this decay, but we also mapped out the stability-/metastability boundary in the parameter space of the couplings. We furthermore identified an analytically constructed mode that proved to be sufficient to trigger the decay of a global vortex. Let us briefly review our findings here. Assuming $m_u = m_d = m_s = 0$, the Ginzburg-Landau Lagrangian of the effective theory reads

$$\mathcal{L} = \mathrm{Tr}\left[-\frac{1}{4}F_{ij}F^{ij} + D_i\Phi^\dagger D^i\Phi + m^2\Phi^\dagger\Phi - \lambda_2(\Phi^\dagger\Phi)^2\right] - \lambda_1(\mathrm{Tr}[\Phi^\dagger\Phi])^2 + \frac{3m^4}{4\lambda}\,, \tag{1}$$

⋆ e-mail: alford@physics.wustl.edu
⋆⋆ e-mail: kumar.s@wustl.edu
⋆⋆⋆ e-mail: tvachasp@asu.edu
⋆⋆⋆⋆ e-mail: windisch@physics.wustl.edu

where $D_i = \partial_i - igA_i$ is the covariant derivative, $F_{ij} = \partial_i A_j - \partial_j A_i - ig\left[A_i, A_j\right]$ is the gauge field-strength tensor, and the A_i represent the gauge fields (gluons). The coupling λ is a linear combination of the original self-couplings of the condensate,

$$\lambda \equiv 3\lambda_1 + \lambda_2. \tag{2}$$

The matter field Φ represents the CFL condensate, and has a $3_c \times 3_f$ complex matrix structure. A general entry of the Φ matrix is thus characterized by a color index α and a flavor index a, $\phi_{\alpha a}$. In the broken phase, the vacuum expectation value (vev) is given by

$$A_i = 0\,, \quad \Phi = \bar{\phi}\mathbf{1}_{3\times 3}\,, \quad \bar{\phi} = \sqrt{\frac{m^2}{2\lambda}}\,. \tag{3}$$

A (global) *superfluid vortex* can then be written as

$$A_i = 0\,, \quad \Phi_{\rm sf} = \bar{\phi}\beta(r)e^{i\theta}\,\mathbf{1}_{3\times 3}\,, \tag{4}$$

where $\beta(r)$ is the radial profile obtained from solving

$$\beta^{''} + \frac{\beta^{'}}{r} - \frac{\beta}{r^2} - m^2\beta(\beta^2 - 1) = 0\,, \tag{5}$$

with the boundary conditions $\beta \to 0$ as $r \to 0$, and $\beta \to 1$ as $r \to \infty$. A (red) *semi-superfluid flux tube*, on the other hand, can be written as

$$\Phi_{\rm ssft}(r,\theta) = \bar{\phi}\begin{pmatrix} f(r)\mathrm{e}^{i\theta} & 0 & 0 \\ 0 & g(r) & 0 \\ 0 & 0 & g(r) \end{pmatrix}, \tag{6}$$

$$A_\theta^{\rm ssft}(r) = -\frac{1}{gr}(1 - h(r))\begin{pmatrix} -\frac{2}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}, \tag{7}$$

$$A_r^{\rm ssft} = 0\,. \tag{8}$$

and the solutions for the green and blue flux tubes follow from swapping the diagonal elements of the matrix. The profile functions $f(r)$, $g(r)$ and $h(r)$ obey a set of coupled differential equations, see equations (10)-(14) in [2]. As discussed in our paper, far from the core of the vortex, the energy density of one semi-superfluid flux tube is one ninth of the energy density of a global vortex,

$$\begin{aligned} \epsilon_{\rm sf} &= 3\bar{\phi}^2/r^2, \\ \epsilon_{\rm ssft} &= \tfrac{1}{3}\bar{\phi}^2/r^2, \end{aligned} \tag{9}$$

which is the cause of the instability. In the case of vanishing gauge coupling, we identified an unstable mode analytically,

$$\delta\Phi^{(8)} = \epsilon\,\hat{n}\cdot\nabla\psi(r,)\,T_8\,. \tag{10}$$

This mode establishes a distortion of the red and green components of the vortex by a small amount ϵ in direction of the unit vector $\hat{n}$, while shifting the blue component by an amount of 2ϵ in the opposite direction. This perturbation of the global vortex can be plugged into the Hamiltonian density, and the change in energy density evaluates to

$$\delta E_8 = \epsilon^2(\lambda_2 - \lambda)\frac{\pi m^4}{\lambda^2}\int_0^\infty r\mathrm{d}r\beta^{'2}\beta^2\,. \tag{11}$$

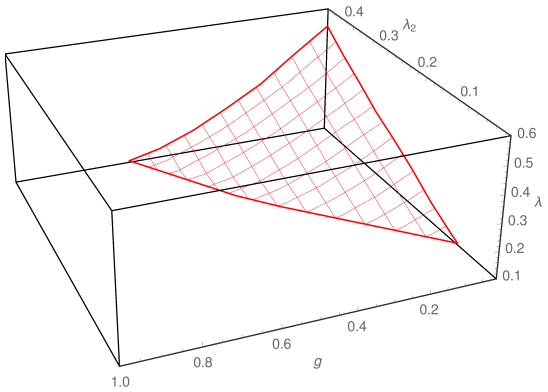

Figure 1. The parameter space of the couplings g, λ and λ_2. Points behind the surface form the region of meta-stability, all other points constitute the region of instability.

For $\lambda_2 > \lambda$, this clearly lowers the energy and is thus an unstable direction. Depending on the gauge coupling g and the condensate self-couplings λ_1 and λ_2, there are regions in parameter space where the global vortex solution is unstable and decays immediately, but we could also identify regions of meta-stability, see Figures 1 and 2. In Figure 2, the solid line corresponds to the instability boundary in parameter space as derived from the perturbation (10), which seems to hold approximately for small gauge couplings.

The dashed line in Figure 2 corresponds to the region in parameter space where $\lambda_1 = \lambda_2$. This has been identified as the physically relevant regime at ultra-high density, where the coupling is sufficiently small to allow for mean-field calculations, [5, 6],

$$\lambda_1 = \lambda_2 = \frac{\lambda}{4} = \frac{36}{7}\frac{\pi^4}{\zeta(3)}\left(\frac{T_c}{\mu}\right)^2 . \tag{12}$$

If this result can be extrapolated to densities where the system is strongly coupled, our study indicates that there are no regions of metastability for the physically relevant case. This is supported by the fact that the region of metastability dies away quickly with increasing coupling g.

2 Initial condition and boundary treatment for multi-vortex dynamics

So far we have explored the behavior of a single global vortex, as well as arrangements of three well-separated semi-superfluid flux tubes. Currently we are investigating two possible extensions to our numerical framework. On the one hand, we introduced strange-quark mass asymmetry, where we followed the work of [7]. Our findings will be presented in a future publication. On the other hand, we study the dynamics of multiple vortices on the lattice, which introduces several complications. In these proceedings, we discuss the first steps in the simulation of multiple vortices on the lattice. To set up many vortices on the lattice, we first obtain a radial profile for a winding-one global vortex by solving (5) using a relaxation method on a large grid with high precision. In the absence of a gauge field at initial time (that is, the link variables are unity), we construct the Φ-field in the presence of N

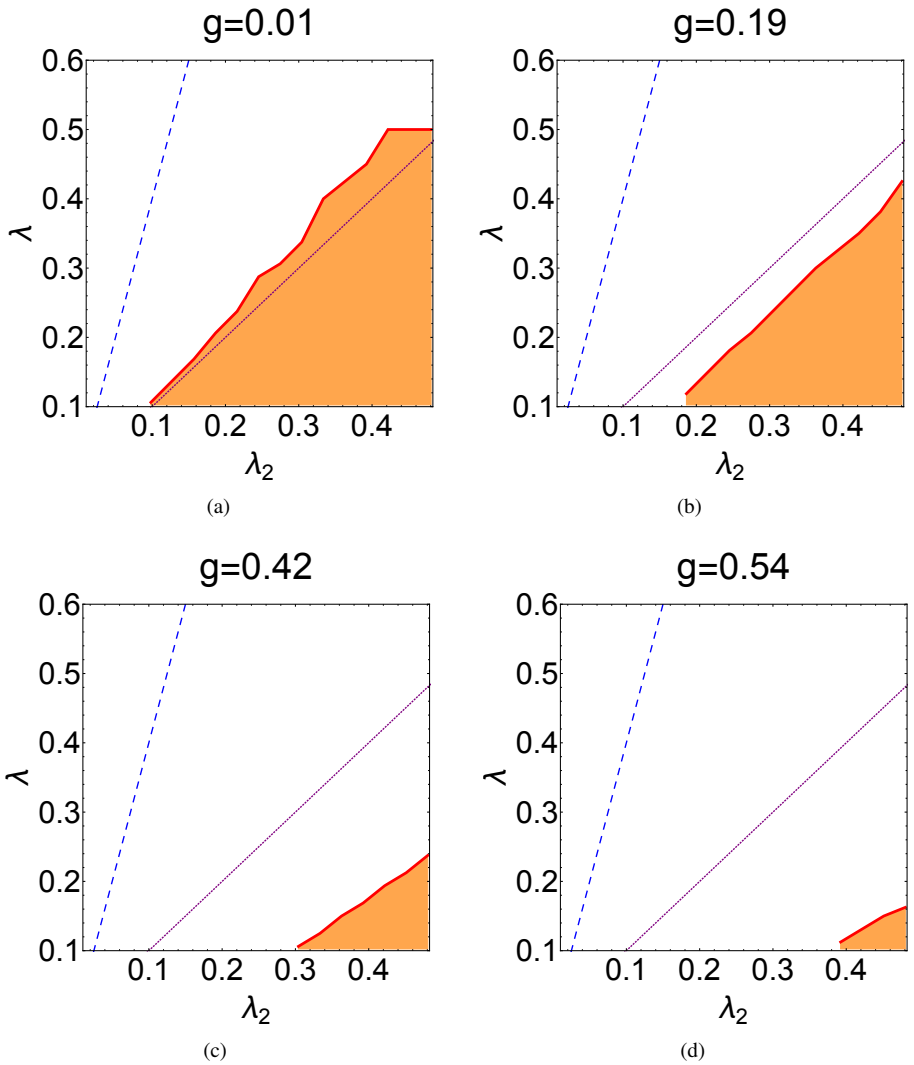

Figure 2. Slices through the parameter space of the couplings for different values of g. The shaded area is the region of meta-stability, its complement constitutes regions of instability. The dashed line is the projection of the plane of weak-coupling results, see text. The solid line represents $\lambda = \lambda_2$, which, in the case of vanishing gauge coupling, should yield the boundary of stability/meta-stability. Note that the metastability region can be characterized by $\lambda_1 \lesssim -0.16g$. For fixed values of g and λ_1, a change in λ_2 has almost no effect on the stability boundary.

vortices according to

$$\Phi = \left(\prod_{i=1}^{N} \beta(r_i)\right) \exp\left\{ i \sum_{i=1}^{N} n_i \theta_i \right\} \bar{\phi} \mathbf{1}_{3\times 3}, \tag{13}$$

which is just a superposition of the N vortices. At a given point (r, θ), the variable r_i indicates the distance to a vortex at position i, and θ_i is the angle with respect to the vortex at position i. Even though we plan to simulate systems of multiple vortices at non-zero gauge coupling in the unstable regime, we start with the most simple case of zero gauge coupling and in the meta-stable region of couplings. In this case, the Ansatz (13) is all we need. Since the superfluid vortices feature a repulsive behavior, they separate quickly once we start our simulation, and, if we use the boundary conditions of our previous study, approach the boundary and pin there. As far as semi-superfluid strings are concerned, the fixed boundary case of our initial work should serve our purpose, since the boundary is repulsive to the individual flux tubes. Together with the repulsion of the semi-superfluid vortices among themselves, this will allow us to find the lattice structure of the multi-fluxtube arrangement. However, at this point we have not yet observed the decay into semi-superfluid fluxtubes in the multi-vortex system, and in this first step we intend to study the numerically much cheaper approach of global vortices only. We thus have to use a modification of the first boundary condition. Using fixed boundary conditions without any modifications also introduces the problem that, far away from the center, many vortices constitute a large overall winding number, which in turn causes the energy density to increase with decreasing distance to the boundary. In order to cancel some of the winding, we introduced a grid of vortices which resides *beyond* the boundary, that is, outside of the simulation region. The presence of those vortices, however, is taken into account when the initial condition is computed, and, in particular, is frozen into the fixed boundary, see Figure 3.

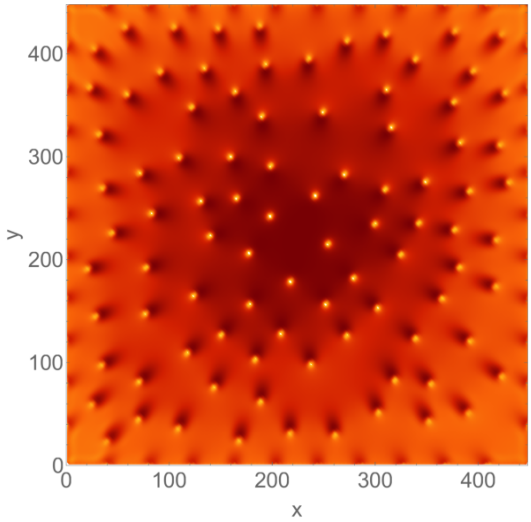

Figure 3. Example of the energy density of an initial configuration of 90 vortices. The vortex positions have been chosen randomly, with the additional constraint of a minimal vortex separation of 35 lattice units. In addition to the 90 vortices, one can see the vortices sitting *beyond* the simulation region. They appear as 'shadows' along the edges of the lattice. Those vortices have been frozen into the fixed boundary condition at initial time, and help in providing a better initial setting for large vortex number arrangements, see text.

3 Time evolution of a multi-superfluid-vortex arrangement

In this section we show the time evolution of a random setting of 90 vortices on a 448×448 lattice at vanishing gauge coupling in the meta-stable regime of the theory. In this case, the superfluid vortices won't undergo the decay to the semi-superfluid fluxtubes, but are expected to form a hexagonal lattice structure. Snapshots of the time evolution of a first simulation run are shown in Figure 4. In this run we have used a very high damping factor, in order to relax the system towards its lowest energy state as gentle as possible. This is necessary, since the initial state has a high vortex (winding) number, and since our Ansatz is not the lowest energy state, it has a lot of excess energy. The downside of a high damping factor is a very slow time evolution, which is evident from the comparably small changes in the configurations shown in Figure 4(a), despite the large time separation of the configurations. In this first simulation run we successfully managed to simulate the time evolution of a state with very high vorticity in a controlled and stable way. The vortices separate, and seem to prefer to occupy the boundary, which requires some improvement of the boundary condition. So far we did not observe the expected hexagonal lattice structure, however, this should be easily achievable by an appropriate choice of initial condition and simulation parameters.

4 Conclusions and outlook

In this progress report we have reviewed our findings from our previous study [1], where we have found numerical proof of the instability of superfluid vortices in CFL quark matter. We identified regions of stability and metastability in the parameter space of the couplings, and connected our result to the physically relevant regime by extrapolating weak coupling results. This indicates that superfluid vortices are unstable in CFL quark matter. An analysis of the coupling dependence of the stability/metastability boundary revealed that it seems to be independent of the coupling λ_2, which raises the question of the role of the condensate self-couplings in the decay process. We also constructed a mode that is sufficient to trigger the decay of a superfluid vortex into the triplet of semi-superfluid fluxtubes. The phenomenological consequences of the instability, for example on neutron star glitches, remain elusive for now, since that requires a thorough understanding of a pinning mechanism. As a first step towards a better understanding of the dynamics of a system with high vorticity we extended our framework and presented first results for the time evolution of a multi-superfluid-vortex system. We modified the boundary conditions slightly, which allows for a better initial setting of a multi-vortex state, and simulated a system of 90 superfluid vortices in a controlled and stable way. Because it is numerically less expensive, we started with the zero-gauge coupling case, and focused on the stable regime of condensate self-couplings. As a next step, we intend to study the non-zero gauge coupling case. This is numerically more expensive, since, on the lattice, gauge fields are represented as link variables corresponding to group elements, and their time evolution involves a matrix exponential. In order to speed up the code, we thus plan to use Graphics Processing Units (GPUs), which allows for much faster evolution on larger lattices.

Acknowledgments

We thank Andreas Schmitt for helpful discussions. This study has been supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Award number #DE-FG02-05ER41375. AW acknowledges support through the Austrian Science Fund (FWF) Schrödinger Fellowship J 3800-N27. TV acknowledges support by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award number #de-sc0013605.

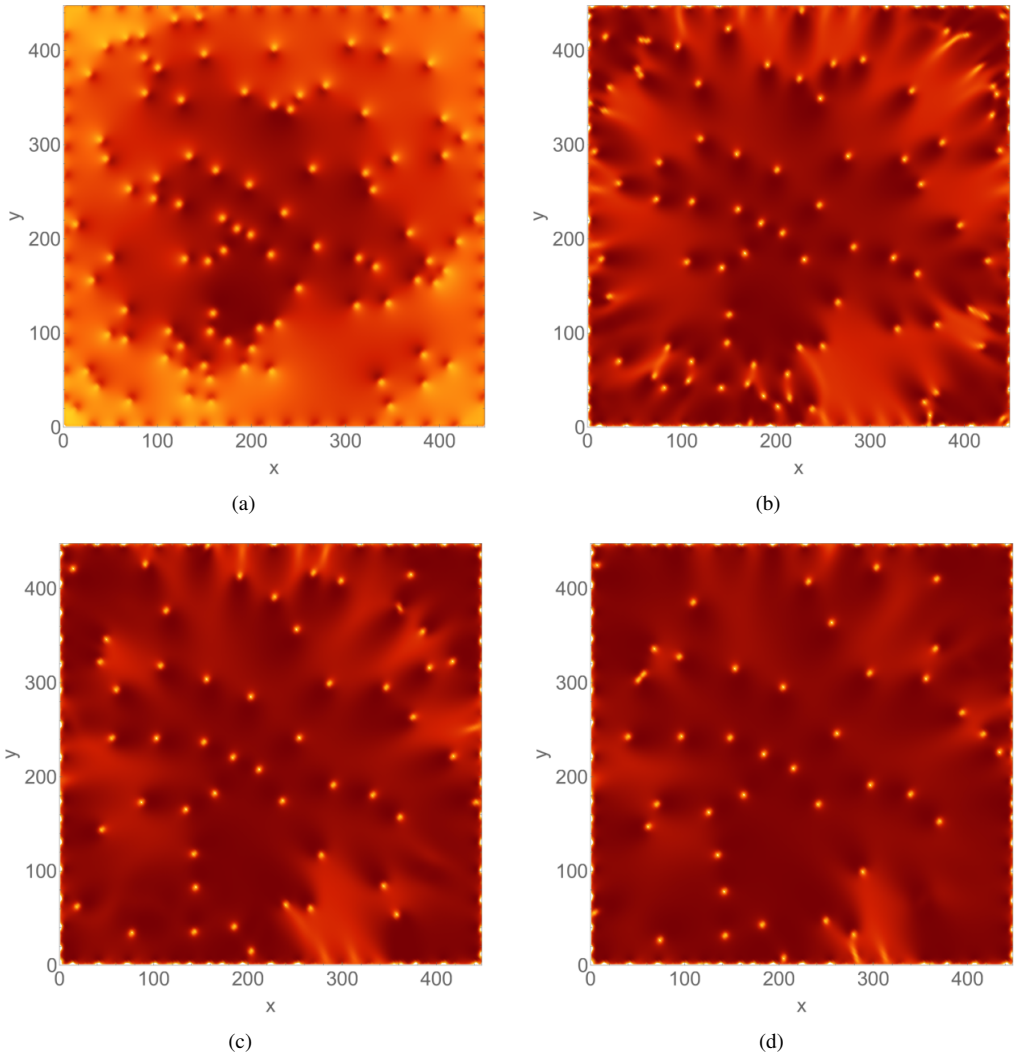

Figure 4. Energy density snapshots of the time evolution of an initial arrangement of 90 vortices at random positions with minimal initial vortex-vortex separation of 15 lattice units, using the improved fixed boundary condition discussed in Section 3. The panels 4(a),4(b),4(c) and 4(d) show the state of the system at initial time, and at $t = 1000$, $t = 2000$ and $t = 3000$ respectively. In order to prevent a violent separation of the vortices, a high damping factor has been used in this evaluation. The lighter color shading of the initial state in panel 4(a) comes from the fact that the initial condition has a lot of excess energy, which is then dissipated throughout the time-evolution.

References

[1] M.G. Alford, K. Rajagopal, F. Wilczek, Phys. Lett. **B422**, 247 (1998), hep-ph/9711395

[2] M.G. Alford, S.K. Mallavarapu, T. Vachaspati, A. Windisch, Phys. Rev. **C93**, 045801 (2016), 1601.04656

[3] A.P. Balachandran, S. Digal, T. Matsuura, Phys. Rev. **D73**, 074009 (2006), hep-ph/0509276

[4] E. Nakano, M. Nitta, T. Matsuura, Phys. Lett. **B672**, 61 (2009), 0708.4092

[5] K. Iida, G. Baym, Phys. Rev. **D63**, 074018 (2001), [Erratum: Phys. Rev.D66,059903(2002)], hep-ph/0011229

[6] I. Giannakis, H.c. Ren, Phys. Rev. **D65**, 054017 (2002), hep-ph/0108256

[7] M. Eto, M. Nitta, N. Yamamoto, Phys. Rev. Lett. **104**, 161601 (2010), 0912.1352

Gravitational Waves, Gamma Ray Bursts, and Black Stars*

Tanmay Vachaspati
Physics Department, Arizona State University,
Tempe, AZ 85287, USA.
E-mail: tvachasp@asu.edu
(Dated: November 14, 2016)

Stars that are collapsing toward forming a black hole but appear frozen near their Schwarzschild horizon are termed "black stars". The collision of two black stars leads to gravitational radiation during the merging phase followed by a delayed gamma ray burst during coalescence. The recent observation of gravitational waves by LIGO, followed by a possible gamma ray counterpart by Fermi, suggests that the source may have been a merger of two black stars with profound implications for quantum gravity and the nature of black holes.

From an asymptotic observer's viewpoint, a collapsing body is forever suspended just above its Schwarzschild radius. The strong gravitational redshift near the surface of the collapsing body causes the body to appear black. Such objects are known as "frozen stars" or "black stars". Black *holes* are the infinite time limit of black stars and traditionally black stars are viewed as indistinguishable from black holes (*e.g.* Chapter 33, [1]). However, there are good reasons to maintain the distinction between black stars and black holes. First, *quantum* analyses of gravitational collapse show that a black star evaporates in a finite time [2] and so it is impossible to take the infinite time limit. Second, theories of quantum gravity often predict that black holes have structure such as a string theory fuzzball [3] or a firewall [4]. Third, observations of the collision of two black objects can tell us if the objects are black holes or black stars and hence the distinction between these objects is experimentally meaningful [5]. The resolution of these issues has taken on a new urgency after the recent exciting observations by LIGO [6] and Fermi [7] in which gravitational wave emission from two coalescing black hole-like objects appears to have been followed by a gamma ray burst.

The essential idea behind equating black stars to black holes is that a collapsing star very quickly fades from an observer's view, and there is no way to send in probes (*e.g.* light rays) at late times so as to see the surface of the black star. This idea is captured in the spacetime diagram shown in Fig. 1 where certain rays can hit the surface of the star but later rays arrive at the surface of the star after it has crossed into its own event horizon. So the surface of the black star can only be probed by rays that arrive sufficiently early. This is the usual interpretation, also described in [1], and if the object is probed at sufficiently late times, there is no way to send in probes to distinguish between a black star and a black hole.

The picture changes in a *quantum* analysis, since then a collapsing body produces a time-dependent metric that leads to quantum radiation and causes the body to slowly evaporate even as it collapses [2]. This radiation, very similar to Hawking radiation from a black hole [8], does not require the event horizon of a black hole and holds for any collapsing body, even when it is outside its own Schwarzschild radius. Gravitational collapse leads to a black star that is continually collapsing and concurrently evaporating into quantum radiation. Then the collapsing object spacetime is shown in Fig. 2. Now every null ray that hits the collapsing object and reflects off of it will reach future null infinity $\mathscr{I}^+$. This happens if the null ray collides with the collapsing object at any stage of the collapse. The only caveat is that the interaction of the null ray and the black star will lead to a time delay in the escape of the reflected ray but the amount of time delay will depend on when the reflection actually occurs.

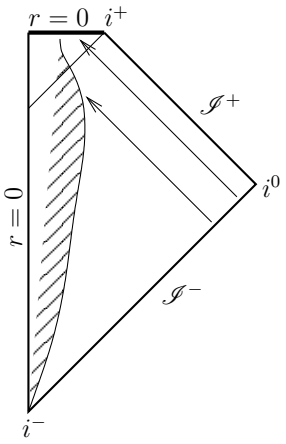

FIG. 1: Spacetime of a classical collapsing object (shaded region) that appears as a black star to an observer located near i^+. Two null rays from $\mathscr{I}^-$ are shown, one that collides with the black star and the other that does not.

In contrast to a black star, a "black hole" is a classical *vacuum* solution of Einstein's equations and there is no matter distribution anywhere in spacetime except perhaps at the central singularity. Hence the collision of black holes will only lead to gravitational radiation because the spacetime is devoid of matter. On the other hand, black star collisions will lead to gravitational *and*

*Essay written for the Gravity Research Foundation 2016 Awards for Essays on Gravitation.

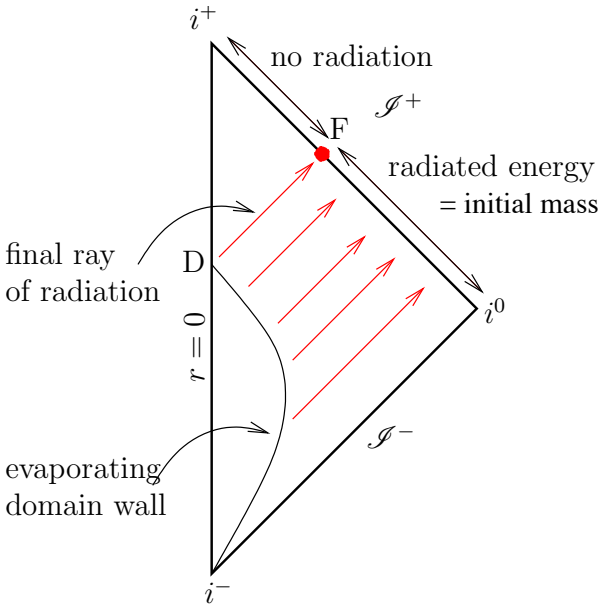

FIG. 2: Spacetime of a quantum collapsing object (labeled a "domain wall" in the figure) [2]. Now there is no event horizon and singularity, and all rays that hit the collapsing object reach $\mathscr{I}^+$, even if they are delayed due to their interactions with the gravitational field.

electromagnetic radiation. Collisions of black stars can be a source of gamma ray bursts (*e.g.* [9, 10]), and such bursts will be preceded by gravitational wave emission whose characteristics are similar to those of black hole mergers. Thus, when two black stars collide, they can be distinguished from the collision of black holes by the presence of electromagnetic radiation.

In a realistic astrophysical setting the collision of two black holes might be accompanied by the collision of other accompanying matter, such as an accretion disk, around the black holes. Stellar mass black holes would have devoured surrounding matter and are therefore considered relatively clean environments, though new astrophysical scenarios might include such matter [11]. Even if there is surrounding matter that collides and produces gamma rays, this electromagnetic radiation and the gravitational radiation would be produced at the same time, with no specific time delay between them. On the other hand, the black star scenario clearly predicts a time delay between the gravitational wave emission and the gamma ray burst because first the metric outside the black stars coalesces and only then the material of the black stars coalesces.

We now estimate the energy radiated from the collision of two black stars each of mass M. The radius of each of the black stars is $R = 2GM$ and so the density is

$$\rho = \frac{1}{8G^3M^2} \sim 10^{14} \left(\frac{30M_\odot}{M}\right)^2 \text{gms/cm}^3 \qquad (1)$$

where $M_\odot \approx 2\times 10^{33}$ gms is the solar mass. This density is high but still below the QCD density. Hence a $30M_\odot$ black star is mostly composed of ordinary protons and neutrons.

The collision of two black stars of mass M involves the collision of their constituent protons and neutrons which will lead to radiation of photons and other light particles. Since the gravitational binding between all constituents in this system is very strong, we treat the collision as being totally inelastic. Then the initial kinetic energy gets converted to radiation resulting in the release of total energy

$$E \sim Mv^2\delta \approx 4\times 10^{49} \left(\frac{M}{30M_\odot}\right) \left(\frac{v}{0.5c}\right)^2 \left(\frac{\delta}{10^{-6}}\right) \text{ ergs} \qquad (2)$$

where δ is the gravitational redshift of the energy as it escapes the collision region. The collision velocity will typically be an $O(1)$ fraction of the speed of light, and the gravitational redshift factor $\delta \ll 1$ will depend on how close the black star is to being a black hole.

We assume that the energy is released in a light crossing time $\sim R/c = 2GM/c \sim 3\times 10^{-4}$sec (for a $\sim 30M_\odot$ black star), which is the only relevant length scale in the problem. This time interval will be dilated by δ^{-1} and the emitted power in photons will be

$$P \sim 2\times 10^{46} \left(\frac{v}{0.5c}\right)^2 \left(\frac{\delta}{10^{-6}}\right)^2 \text{ ergs/sec} \qquad (3)$$

Note that the power is independent of the mass M. We can estimate the frequency of the photons by once again treating the collision as being totally inelastic. Then every proton in the black star gets stopped on collision and the emitted photon energy is simply the initial kinetic energy of the proton

$$E_\gamma \sim m_p v^2\delta \approx 0.25 \left(\frac{v}{0.5c}\right)^2 \left(\frac{\delta}{10^{-6}}\right) \text{ keV} \qquad (4)$$

Even though we cannot precisely estimate the event rate of black star collisions, we do know that the rate is lower for lower initial velocity since then the stars will take a long time to collide *i.e.* δ will be smaller. Then the emitted energy will redshift by a greater amount and there will be a greater time delay between the gravitational and electromagnetic emissions, making the gamma ray burst very faint and also temporally uncorrelated with the gravitational wave event. On the other hand, we expect that the number of black stars falls off with higher velocity. These two arguments suggest that there should be a velocity at which black star collisions peak. In terms of gamma ray bursts, it implies that the gamma ray bursts should have a typical photon energy. Further, the total power emitted should scale with this photon energy as seen by dividing Eq. (3) by (4),

$$\frac{P}{E_\gamma} \approx 10^{56} \left(\frac{\delta}{10^{-6}}\right) \text{ sec}^{-1} \qquad (5)$$

This formula does not depend on the mass of the colliding black stars and neither on their velocities, and hence is an invariant of the model.

If an observed gamma ray burst is indeed due to colliding black stars, the burst should be preceded by gravitational wave radiation from the coalescing spacetimes of the black stars. The gravitational wave emission should be very similar to that calculated numerically for black hole collisions [12–14], and the final gravitational wave emission due to coalescence should be accompanied by the gamma ray burst when the material of the black stars coalesce. The waveforms of the emitted electromagnetic radiation will depend on the normal modes of the two black star system. Indeed, characteristics of the gravitational radiation preceding the gamma ray burst, together with the gamma ray burst, may allow us to infer the parameters of the colliding black stars and the initial conditions.

LIGO [6] has recently detected the gravitational wave signature from the merger of two black holes, each with mass $\approx 30M_{\odot}$. This stunning announcement has been followed by a cautious but equally stunning claim by the Fermi collaboration [7] that they may have seen a gamma ray burst counterpart of the LIGO event. The energy and emission frequency of the gamma ray burst are broadly consistent with those estimated for black star collisions. If future gravitational wave events are followed by delayed gamma ray burst events, it would be strong support for the black star picture and would provide deep insight into gravitational collapse, black holes, and quantum gravity.

Acknowledgments

This work was supported by the U.S. Department of Energy, Office of High Energy Physics, under Award No. de-sc0013605 at ASU.

[1] C. W. Misner, K. S. Thorne and J. A. Wheeler, San Francisco 1973, 1279p
[2] T. Vachaspati, D. Stojkovic and L. M. Krauss, Phys. Rev. D **76**, 024005 (2007) doi:10.1103/PhysRevD.76.024005 [gr-qc/0609024].
[3] S. D. Mathur, Fortsch. Phys. **53**, 793 (2005) doi:10.1002/prop.200410203 [hep-th/0502050].
[4] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, JHEP **1302**, 062 (2013) doi:10.1007/JHEP02(2013)062 [arXiv:1207.3123 [hep-th]].
[5] T. Vachaspati, arXiv:0706.1203 [astro-ph].
[6] B. P. Abbott *et al.* [LIGO Scientific and Virgo Collaborations], Phys. Rev. Lett. **116**, no. 6, 061102 (2016) doi:10.1103/PhysRevLett.116.061102 [arXiv:1602.03837 [gr-qc]].
[7] V. Connaughton *et al.*, arXiv:1602.03920 [astro-ph.HE].
[8] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975) [Erratum-ibid. **46**, 206 (1976)].
[9] D. Q. Lamb, Phys. Rep. **333**, 505 (2000)
[10] T. Piran, Rev. Mod. Phys. **76**, 1143 (2004) [arXiv:astro-ph/0405503].
[11] A. Loeb, Astrophys. J. **819**, no. 2, L21 (2016) doi:10.3847/2041-8205/819/2/L21 [arXiv:1602.04735 [astro-ph.HE]].
[12] F. Pretorius, Phys. Rev. Lett. **95**, 121101 (2005) doi:10.1103/PhysRevLett.95.121101 [gr-qc/0507014].
[13] M. Campanelli, C. O. Lousto, P. Marronetti and Y. Zlochower, Phys. Rev. Lett. **96**, 111101 (2006) doi:10.1103/PhysRevLett.96.111101 [gr-qc/0511048].
[14] J. G. Baker, J. Centrella, D. I. Choi, M. Koppitz and J. van Meter, Phys. Rev. D **73**, 104002 (2006) [arXiv:gr-qc/0602026].

Prepared for submission to JCAP

Probing stochastic inter-galactic magnetic fields using blazar-induced gamma ray halo morphology

Francis Duplessis[*] **and Tanmay Vachaspati**[*†]

[*]Physics Department, Arizona State University, Tempe, AZ 85287, USA.
[†]Maryland Center for Fundamental Physics, University of Maryland, College Park, MD 20742, USA.

E-mail: fdupless@asu.edu, tvachasp@asu.edu

Abstract. Inter-galactic magnetic fields can imprint their structure on the morphology of blazar-induced gamma ray halos. We show that the halo morphology arises through the interplay of the source's jet and a two-dimensional surface dictated by the magnetic field. Through extensive numerical simulations, we generate mock halos created by stochastic magnetic fields with and without helicity, and study the dependence of the halo features on the properties of the magnetic field. We propose a sharper version of the Q-statistics and demonstrate its sensitivity to the magnetic field strength, the coherence scale, and the handedness of the helicity. We also identify and explain a new feature of the Q-statistics that can further enhance its power.

Contents

1 Introduction

Multiple analyses of observed gamma rays [1–7] provide growing evidence for the existence of inter-galactic magnetic fields (for reviews see [8, 9]). The existence of such magnetic fields poses new questions for cosmology and probably also for particle physics [10, 11]. In addition, a primordial magnetic field can play an important role during structure formation in the universe and could help us understand the ubiquity of magnetic fields in astrophysical bodies.

A critical challenge at this stage is to sharpen observational techniques so that we can better observe and measure inter-galactic magnetic fields. Of the various probes of inter-galactic magnetic fields, blazar-induced gamma ray cascades hold certain key advantages. The gamma ray cascades originate in the voids in the large-scale structure and are mostly immune to complications of a noisy environment. The cascade develops in a relatively small spatial volume and hence is a local probe of the magnetic field in the voids. This is distinct from other methods, such as the Faraday rotation of the cosmic microwave background polarization, that probe an integrated measure of the magnetic field. Gamma ray cascades are also highly sensitive probes and can trace very weak cosmological magnetic fields.

In this paper we study the effect of stochastic inter-galactic magnetic fields on blazar induced gamma ray halos and some results overlap with those of Refs. [12–16]. The cascade process is complicated and all analyses use some simplifying assumptions. For example, the analysis in Ref. [13] only considered non-stochastic magnetic field configurations. Other simplification schemes, such as the "large spherical observer" method employed in Ref. [14], transport arrival directions of gamma rays for distant observers to a single Earth-bound observer. This technique is certainly useful to study spectral properties of the cascade, but there is a danger that it loses or shuffles the spatial information of gamma ray arrival directions that is crucial for morphological studies. Our focus is on the effect of stochastic magnetic fields that are statistically isotropic and with or without helicity. So we carefully

analyse the spatial information of the gamma rays that is useful for deducing properties of the magnetic field but, for the present, we only include an approximate description of the cascade development.

An important helpful concept that we develop in this paper is that of the "PP surface" (see Sec. 3). This spatial surface holds the key to halo morphology and many of the features that we see in our simulations can be understood in terms of the shape of the PP surface and its intersection with the blazar jet.

We have applied a refined version of the Q-statistics first proposed in Ref. [17] to study the morphology of halos. Our results show that this statistic can successfully extract the helicity of the magnetic field. Our simulations also reveal that the plot of $Q(R)$, where R is a variable that will be explained below, has an additional bump. We are able to show that this bump is a genuine feature of the Q-statistic and explain it in terms of properties of the PP surface in Sec. 7. Thus this extra feature of $Q(R)$ may become an observational tool in future.

We give some background information in Sec. 2, discuss our simulation techniques in Sec. 3, discuss features of the halo in Sec. 4, introduce the Q-statistic in Sec. 5 and apply it to stochastic fields in Sec. 6. As mentioned above, we discuss the bump feature in $Q(R)$ in Sec. 7. We summarize our conclusions in Sec. 8. Our stochastic magnetic field generation scheme is described in Appendix A.

2 Blazar Halos from an Intergalactic Magnetic Field

TeV photons from blazars induce electromagnetic cascades through pair production with the extragalactic background light, $\gamma_{\rm EBL}\gamma_{\rm TeV} \to e^+e^-$. In the presence of a magnetic field, the charged leptons follow spiral paths as they propagate and lose energy due to inverse Compton scattering with the cosmic microwave background (CMB) photons. The up-scattered CMB photons have gamma ray energies and produce extended halos around the direction of the blazar. In this section we briefly discuss the formation of the halo under simplifying assumptions.

Consider a blazar located at the origin of our coordinate system described by the unit basis vectors $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$. We choose $\hat{\mathbf{z}}$ so that Earth is located at $\mathbf{r}_{\rm E} = -d_s\hat{\mathbf{z}}$ where d_s is the comoving distance to the source,

$$d_s = \frac{1}{a_0 H_0}\int_0^{z_s} \frac{1}{\sqrt{\Omega_m(1+z)^3+\Omega_\Lambda}}dz \simeq \frac{z_s}{0.22}\text{Gpc}. \tag{2.1}$$

To perform the integral, we have used $\Omega_\Lambda \approx 0.69$, $\Omega_m \approx 0.31$, $H_0 \approx 0.67h$ as found in Ref. [18] and we have also assumed that $z_s \ll 1$ and used natural units so that $c = 1$. For all the simulations in this paper we will choose $d_s = 1$ Gpc.

The blazar will typically emit photons in a collimated jet which we approximate to be a conical region with half-opening angle $\theta_{\rm jet} \approx 5^\circ$. The energy $E_{\gamma 0}$ of these photons must lie above some threshold of about a TeV if they are to produce an electron-positron pair from interaction with the Extragalactic Background Light (EBL). Due to the opacity of the EBL, the TeV photons will travel a mean free path (MFP) determined by the pair production cross section $\sigma_{\gamma\gamma}$ and the number density of the EBL photons $n_{\rm EBL}$,

$$D_{\gamma 0} = \langle \sigma_{\gamma\gamma} n_{\rm EBL}\rangle^{-1} \simeq (80\text{Mpc})\frac{\kappa}{(1+z_{\gamma\gamma})^2}\Big(\frac{10\text{TeV}}{E_{\gamma 0}}\Big), \tag{2.2}$$

We have assumed that $n_{\rm EBL} \propto (1+z_{\gamma\gamma})^{-2}$ to approximate the MFP in the final equality [19]. Following [13] we will set $\kappa = 1$ as this dimensionless constant is estimated to lie in the range of $0.3 < \kappa < 3$. The comoving distance from the source to the pair production event is given by $D^c_{\gamma 0} = (1+z_{\gamma\gamma})D_{\gamma 0}$.

The redshift of the produced lepton pairs will depend on the relative position of the leptons to the source. Since $D^c_{\gamma 0} \ll d_s$, we make the approximation $z_{\gamma\gamma} \approx z_s$ and we can write,

$$D^c_{\gamma 0} \simeq (80{\rm Mpc})\frac{\kappa}{(1+z_s)}\Big(\frac{10{\rm TeV}}{E_{\gamma 0}}\Big). \tag{2.3}$$

The energy of each of the produced leptons will be $E_e \approx E_{\gamma 0}/2$. These leptons are expected to travel a distance D_e before losing most of their energy through inverse Compton (IC) cooling which occurs by upscattering CMB photons. The cooling distance is

$$D_e = \frac{3m_e^2}{4\sigma_T U_{CMB} E_e} \simeq (31\ {\rm kpc})\Big(\frac{5\ {\rm TeV}}{E_e}\Big)\Big(\frac{1.22}{1+z_{\gamma\gamma}}\Big)^4 \tag{2.4}$$

where $\sigma_T = 6.65 \times 10^{-25}{\rm cm}^2$ is the Thomson scattering cross section and $U_{CMB}(z_{\gamma\gamma}) \simeq (0.26\ {\rm eV/cm}^3)(1+z_{\gamma\gamma})^4$ is the CMB energy density. Note that we can assume the whole cascade development happens around redshift $z_{\gamma\gamma}$ as $D_e \ll D_{\gamma 0}$. At that redshift, the average energy of a CMB photon is

$$E_{CMB} \simeq (6\times 10^{-4}\ {\rm eV})(1+z_{\gamma\gamma}), \tag{2.5}$$

which implies that, from energy conservation, the upscattered photons will have energy

$$E_\gamma = \frac{4}{3}E_{CMB}\frac{E_e^2}{m_e^2} \simeq (77\ {\rm GeV})\Big(\frac{E_{\gamma 0}}{10\ {\rm TeV}}\Big)^2. \tag{2.6}$$

As the lepton propagates, it upscatters $\approx ({\rm 10TeV})/({\rm 10GeV}) \sim 10^3$ photons, and produces a gamma ray cascade in the 1-100 GeV range if the initial gamma ray had an energy of a few TeV. Clearly not every photon upscattered by the leptons will reach Earth. Those that do must come from a set of events that satisfy a set of three constraints given in Ref. [13] that we now describe.

After pair production, the lepton's initial velocity will be almost parallel to the momentum of the parent photon with a negligible deviation of order the inverse Lorentz boost factor $m_e/E_e \sim 10^{-6}$. Their subsequent trajectory will be determined by the magnetic field $\mathbf{B}(\mathbf{x})$ through the Lorentz force. If the magnetic field is incoherent on length scales smaller than the cooling distance $D_e \sim 30$ kpc, the lepton trajectories will be diffusive and this situation is much harder to analyze. So we focus on magnetic fields that are coherent on scales that are much larger than D_e. Then the lepton trajectories are bent in an effectively constant magnetic field and follow a helical trajectory with gyroradius

$$R_L = R_{L0}|\mathbf{v}_\perp|, \quad \text{with } R_{L0} = \frac{E_e}{e|\mathbf{B}|}, \tag{2.7}$$

which depends on the lepton's perpendicular velocity to $\mathbf{B}$, *i.e.* $\mathbf{v}_\perp = \mathbf{v} - (\mathbf{v}\cdot\hat{\mathbf{B}})\hat{\mathbf{B}}$.

The quantity $2\pi R_{L0}$ is useful as it denotes the distance the lepton must travel in order to perform a full revolution. The value of R_{L0} is a function of the redshift as it depends on $|\mathbf{B}|$.

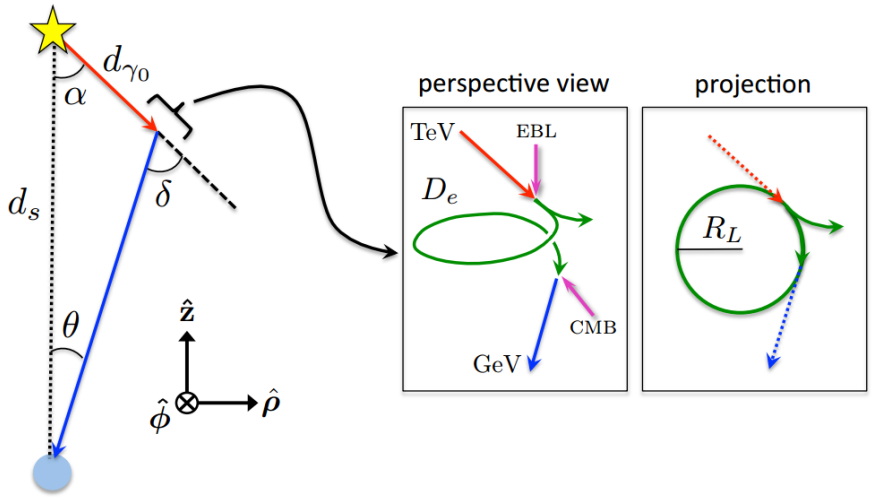

Figure 1. A TeV photon emitted from a blazar travels a comoving distance of $d_{\gamma 0}$ before scattering off an EBL photon and pair producing leptons. The lepton trajectories are bent due to the local magnetic field over a very short distance compared to the distance to the source, d_s, and are shown in the insets. (The insets show huge bending whereas we have only considered magnetic fields that give small bending.) Inverse-Compton scattering of a lepton and CMB photons results in a cascade of GeV energy gamma rays arriving at Earth from the direction of the pair production. [Sketch taken from [13].]

For magnetic fields frozen in the plasma, the field strength redshifts as $|\mathbf{B}| = B_0(1+z_{\gamma\gamma})^2 \approx B_0(1+z_s)^2$ where B_0 is the magnetic field magnitude today. With

$$R_{L0} \simeq 3.5 \text{ Mpc } \left(\frac{E_e}{5 \text{ TeV}}\right)\left(\frac{B_0}{10^{-15}G}\right)^{-1}\left(\frac{1+z_s}{1.22}\right)^{-2}, \tag{2.8}$$

we can evaluate the ratio

$$\frac{D_e}{2\pi R_{L0}} \simeq 0.0106 \left(\frac{E_\gamma}{10 \text{ GeV}}\right)^{-1}\left(\frac{B_0}{10^{-15}G}\right)\left(\frac{1+z_s}{1.22}\right)^{-2} \tag{2.9}$$

which determines the angular deflection of the leptons.

As depicted in Fig. 1, let us introduce the following angles: θ is the arrival angle of the GeV photon with respect to the source location, δ represents the angle between the upscattered photon and the TeV photon, α is the angle subtended by the TeV photon's momentum, $\mathbf{r}_E$ the vector from the source to Earth, and finally ϕ is the azimuthal angle in which the whole (planar) scattering process take place. We also introduce the polar vectors $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\phi}}$, in the $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ plane. It is important to emphasize that the whole process occurs in a plane to a good approximation because $D_e \ll D^c_{\gamma 0}, d_s$ and the length of the lepton trajectory can be ignored. Then there are only 3 points that are relevant (the source, the pair production point, and the observer) and they always lie in a plane.

Applying the sine formula to the triangle in Fig. 1 we get our first constraint

$$d_s \sin(\theta) = d_{\gamma 0} \sin(\delta). \tag{2.10}$$

where $d_{\gamma 0}$ is the distance traveled by the TeV photon and is a random variable drawn from a distribution that depends on the MFP $D^c_{\gamma 0}$ in Eq. (2.3). This is discussed in detail in Sec. 3.

The bending angle δ is related to the distance traveled by the lepton through the local magnetic field which we write as $\mathbf{B} = B\hat{\mathbf{n}}_{||}$. We also decompose the lepton's initial velocity at time $t_i = 0$, $\mathbf{v}(t_i = 0) = v_{||}\hat{\mathbf{n}}_{||} + v_\perp \hat{\mathbf{n}}_\perp$, where $\hat{\mathbf{n}}_\perp \cdot \hat{\mathbf{n}}_{||} = 0$. At some later time t the velocity is

$$\mathbf{v}(t) = v_{||}\hat{\mathbf{n}}_{||} + v_\perp \cos(\omega t)\hat{\mathbf{n}}_\perp \pm v_\perp \sin(\omega t)(\hat{\mathbf{n}}_\perp \times \hat{\mathbf{n}}_{||}), \tag{2.11}$$

Here we introduced the angular frequency of the orbital motion $\omega = v_\perp/R_L = 1/R_{L0}$ and the $+$ $(-)$ sign refers to the positron (electron) trajectory. A CMB photon upscattered at time t_{IC} will be directed along the lepton's trajectory and so the deflection angle of Figure 1 can be expressed as $\cos(\delta) = \hat{\mathbf{v}}(0) \cdot \hat{\mathbf{v}}(t_{IC})$. Using Eq. (2.11) we can derive the second constraint,

$$1 - \cos(\delta) = \Big(1 - (\hat{\mathbf{v}}(0) \cdot \hat{\mathbf{B}})^2\Big)\Big(1 - \cos(t_{IC}/R_{L0})\Big). \tag{2.12}$$

The time of inverse Compton scattering t_{IC} is a stochastic variable. Given its value and the magnetic field direction, the constraints determine the bending angle, δ.

A single propagating lepton will be able to upscatter CMB photons towards Earth only at certain times when the lepton's momentum is directed towards Earth. Photons upscattered at other times will not reach Earth and we can safely ignore them. The number of photons upscattered by a lepton is very large ($\sim 10^3$), with mean deviation angles between the photons $\sim 10^{-3} \times 0.01$ (see Eq. (2.9)). This angle is large enough that we only expect ~ 1 of the cascade photons from any lepton to reach Earth. This allows us to adopt the strategy that we first select a value of t_{IC} from an exponential probability distribution as described in Sec. 3 and then solve the constraint equations to find all TeV gamma rays from the blazar that upscatter CMB photons that reach Earth. For different values of t_{IC}, different TeV gamma rays from the blazar will lead to observed photons. In this way, we will be able to track the photons that arrive on Earth and not waste computational effort on those that go elsewhere.

The third and final constraint is that the cascade gamma ray lies in the plane specified by $\hat{\phi}$. This requires that the Lorentz force in the azimuthal $\hat{\phi}$ direction vanishes between the time of pair production and IC scattering. Namely the ϕ component of the impulse must vanish,

$$J_\phi = \hat{\boldsymbol{\phi}} \cdot \mathbf{J} = \hat{\boldsymbol{\phi}} \cdot \Big(\pm e \int_0^{t_{IC}} dt\ \mathbf{v}(t) \times \mathbf{B}\Big) = 0. \tag{2.13}$$

The impulse can be simplified by pulling out the assumed constant magnetic field of the integral and defining

$$\mathbf{v}_{\rm avg} = \frac{1}{t_{IC}} \int_0^{t_{IC}} dt\ \mathbf{v}(t). \tag{2.14}$$

The geometrical setup of Fig. 1 forces $\mathbf{v}_{\rm avg}$ to bisect the angle δ and therefore its unit vector can be written as

$$\hat{\mathbf{v}}_{\rm avg} = \sin\big(\delta/2 - \theta\big)\hat{\boldsymbol{\rho}} - \cos\big(\delta/2 - \theta\big)\hat{\mathbf{z}}. \tag{2.15}$$

Decomposing $\mathbf{B}$ as

$$\mathbf{B} = b_\rho \hat{\boldsymbol{\rho}} + b_\phi \hat{\boldsymbol{\phi}} + b_z \hat{\mathbf{z}} \tag{2.16}$$

allows us to write

$$\hat{\boldsymbol{\phi}} \cdot \hat{\mathbf{v}}_{\rm avg} \times \hat{\mathbf{B}} = -b_\rho \cos\big(\delta/2 - \theta\big) - b_z \sin\big(\delta/2 - \theta\big) = 0. \tag{2.17}$$

To summarize this section, Eqs. (2.10), (2.12) and (2.17) are the constraints that need to be satisfied by the variables (θ, δ, ϕ) given a magnetic field realization and initial velocity

of the TeV gamma ray (both of which depend on (θ, δ, ϕ)), the source-observer distance (d_s), the distance to pair production ($d_{\gamma 0}$), and the photon upscattering time (t_{IC}).

3 Halo Simulations

For events that satisfy the constraints in Eqs. (2.10), (2.12) and (2.17), an observer on Earth will receive flux at a polar angle of θ from the line of sight (LoS) to the blazar and at an azimuthal angle ϕ. Solving these constraints requires the use of numerical methods when considering general $\mathbf{B}(\mathbf{x})$ and when including the stochasticity in the propagating distances (PDs) of the initial gamma ray and pair produced leptons.

Therefore to simulate one observed photon, we supply the distance $d_{\gamma 0}$ traversed by some TeV gamma ray of energy $E_{\gamma 0}$ emitted from the source before it pair produces leptons, one of which in turn travels a distance ct_{IC} before emitting a photon of energy E_γ. Once these 4 values $d_{\gamma 0}$, $E_{\gamma 0}$, ct_{IC} and E_γ, are set and an ambient magnetic field is given, one can numerically solve the constraint equations for θ, δ and ϕ. The process is repeated until N (which we chose to be 1000 or 5000 per simulation) observed photons are simulated. This will create the halo that one would observe if the source was emitting isotropically. For a source with a specific jet orientation we only retain the events whose initial TeV photons lie within the jet. These small number of events give us the observed halo that will be shown in our plots.

Let us go through the details regarding the generation of $d_{\gamma 0}$, $E_{\gamma 0}$, ct_{IC} and E_γ. We must supply some energy distribution for gamma rays emitted by blazars; for this we assume a power law spectrum [20, 21] and follow Ref. [15] by choosing a spectral index of $\Gamma \simeq 2.5$ that is characteristic of the TeV sources. This yields a spectrum given by,

$$\frac{dN_{\gamma 0}}{dE_{\gamma 0}} \sim \left(\frac{E_{\gamma 0}}{\text{TeV}}\right)^{-2.5}. \tag{3.1}$$

We shall also impose a 10 TeV cutoff on the emitted photon energy. The distance $d_{\gamma 0}$ traversed by a TeV gamma ray before turning into a pair of leptons is drawn from the exponential distribution,

$$P[d_{\gamma 0}] = \frac{1}{D^c_{\gamma 0}(E_{\gamma 0})} e^{-d_{\gamma 0}/D^c_{\gamma 0}(E_{\gamma 0})}. \tag{3.2}$$

The resulting leptons will have energies E_e given by Eq. (2.6) and they will upscatter numerous CMB photons along their trajectories. The mean free path between each scattering is given by $l_{\text{MFP}} = (n_{CMB}\sigma_T)^{-1}$, where n_{CMB} is the number density of CMB photons. The lepton loses energy with each scattering and subsequent scatterings lead to lower energy cascade gamma rays. Hence we run Monte Carlo simulations to determine the distributions $P(ct_{IC}, E_\gamma | E_e^{(\text{ini})})$, giving us the probability that a lepton with initial energy $E_e^{(\text{ini})}$ upscatters a CMB photon to energy E_γ after traveling a distance $d_e = ct_{IC}$. Examples of these distributions are shown in Fig. 2. Note that d_e will generally be much smaller than the cooling distance D_e. Only events that lead to observed photons of energy between $E_{\text{min}} = 5$ GeV and $E_{\text{max}} = 50$ GeV will be retained as these are in energies of observational interest for the statistical analysis done in Sec. 5.

We will solve the constraint equations in a variety of magnetic field backgrounds, starting with simple analytic configurations for illustration purposes, and then move on to the more

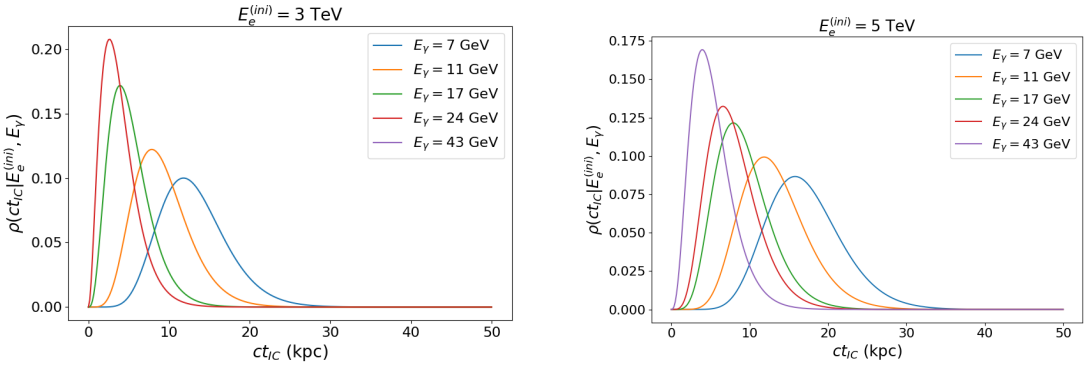

Figure 2. Examples of the probability distribution for the distance traveled by a lepton with initial energy $E_e = E_{\gamma_0}/2$ before it upscatters a CMB photon to E_γ. These distributions are for leptons evolving in the CMB light at a redshift of $z \approx 0.24$, corresponding to a source located at $d_s \approx 1$ Gpc from Earth. Note that the final distribution for $E_\gamma = 43$ GeV does not appear on the left as the lepton does not possess enough energy to upscatter the CMB photons to these energies.

realistic case of stochastic, isotropic magnetic fields. Our procedure to generate stochastic, isotropic magnetic fields is described in Appendix A.

As a warm up, and to compare our method with the results of Ref. [13], we consider a source that radiates TeV photons isotropically in two different magnetic field backgrounds. The first background,

$$\mathbf{B} = B_0\Big(\cos(\beta)\hat{\mathbf{y}} - \sin(\beta)\hat{\mathbf{z}}\Big), \tag{3.3}$$

with $\beta = \pi/4$, is a uniform magnetic field pointing at an angle $\pi/4$ from the line of sight. The second background is a maximally helical field

$$\mathbf{B} = B_0\Big(\sin(2\pi z/\lambda)\hat{\mathbf{x}} + \cos(2\pi z/\lambda)\hat{\mathbf{y}}\Big). \tag{3.4}$$

Here λ is the coherence length of the helical field. We will take $d_s = 1$ Gpc, $B_0 = 10^{-14}$ G and $\lambda = 500$ Mpc as the prototypical values and eventually vary them one at a time to see their effect on the halo morphology.

Next we solve the constraint equations and determine the arrival directions θ, ϕ for several different energies E_γ, for an isotropically emitting source. The points located further away from the source direction usually corresponds to lower energy photons. This is expected since leptons that travel long distances (and hence allow for a large bending angle) will have already lost a lot of energy and upscatter less energetic photons. This behavior can be seen from the distribution shown in Fig 2.

We show halos for the simple field configurations of Eqs.(3.3) and (3.4) in Fig. 3. Looking closely at Fig. 3, the drawn points are triangular; upright triangles are gamma rays that originate from electrons and inverted triangles are those that originate from positrons. The distinction is made clearer in Fig. 4 where red (black) points originate from positron (electron) processes. If the source was taken to be a jet, gamma rays predominantly from one of the two leptons will be observed.

The constraint equations are quite complicated to solve but there is a helpful visualization. First consider the third constraint equation, Eq. (2.17), and note that b_ρ and b_ϕ are also

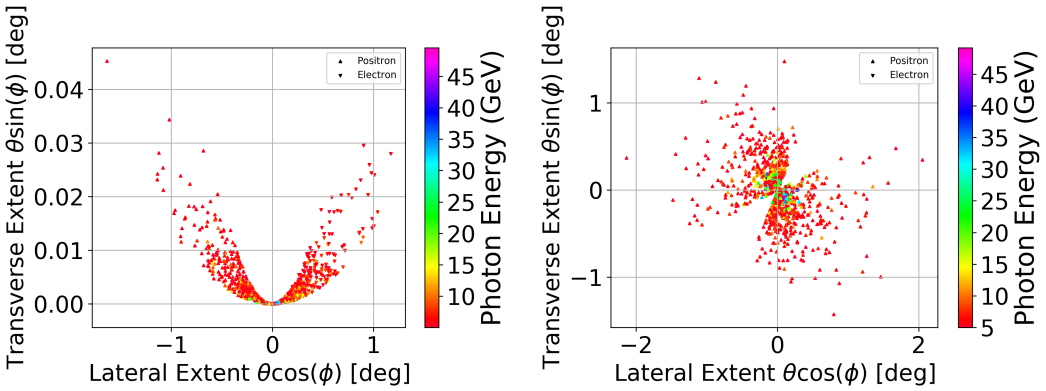

Figure 3. Example of halos from blazars in a uniform (left) and maxially helical (right) inter-galactic magnetic field as given in Eqs. (3.3) and (3.4). The colors denote the energy of the observed gamma ray.

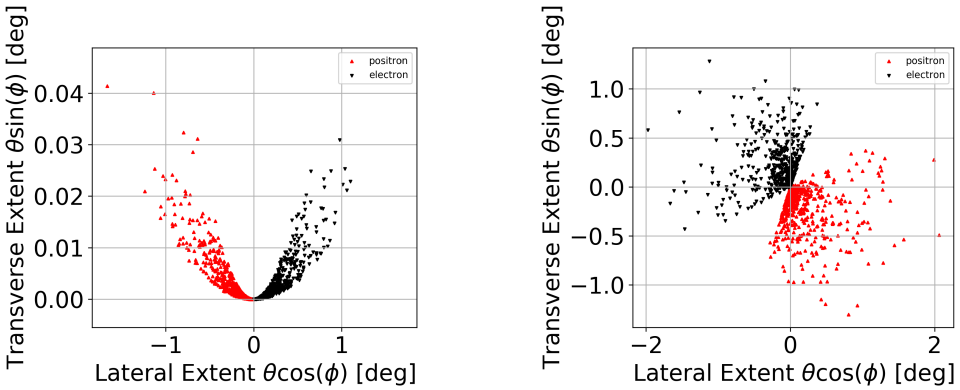

Figure 4. Same as in Fig. 3 but now red (black) points originate from inverse Compton scattering due to positrons (electrons). If the source was taken to emit along a jet, most of the observed gamma rays would originate from either positron or electron processes but not both.

functions of θ, δ and ϕ. So Eq. (2.17) provides one functional relation between these variables that only depends on the magnetic field background. Hence the magnetic field defines a two-dimensional surface in space. We will call this the "Pair Production surface" or the "PP surface" since only lepton pair production at this surface can send GeV gamma rays to the observer. In Fig. 5 we show the PP surface for the magnetic fields of Eqs. (3.3) and (3.4). On these plots we also show the pair production locations, "PP locations", that resulted in the halos of Fig. 3. Note that a gamma ray from the source will propagate a certain distance, $d_{\gamma 0}$ and then pair produce. So the pair production points also lie on a sphere of radius $d_{\gamma 0}$. This is partly enforced by the law of sines in Eq. (2.10), which gives a relation between δ and θ. The intersection of this sphere and the PP surface define a one-dimensional curve in space; CMB photons that are inverse Compton scattered along the one-dimensional curve can propagate to Earth. However, not all points on this one-dimensional curve will satisfy the final constraint. Namely, Eq. (2.12), picks out a limited set of points on the one-dimensional curve and these give the trajectories of the gamma rays that are observed.

The PP surface can be found analytically for simple cases. For instance, the constraint

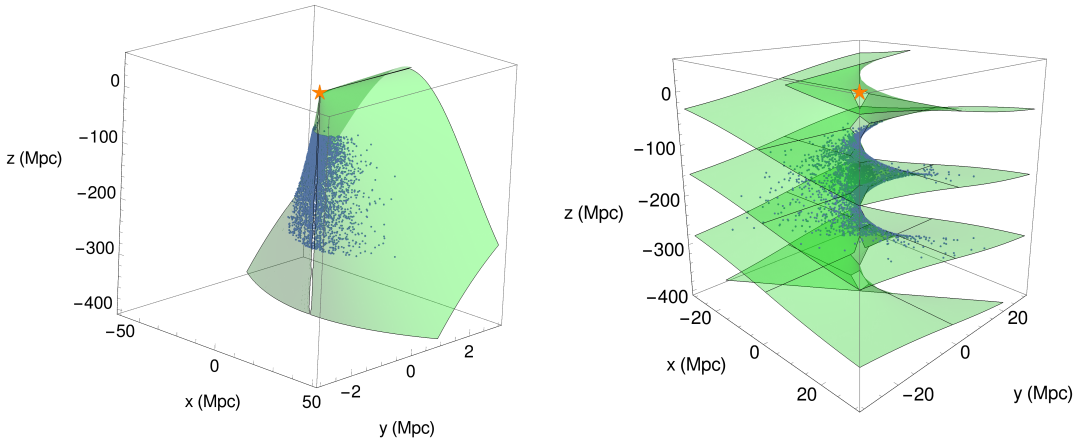

Figure 5. The PP surface for the uniform magnetic field of Eq. (3.3) (left) and the maximally helical magnetic field of Eq. (3.4) (right). The source is located at the orange star; the observer is at $z = -1$ Gpc. The blue points are the events that give rise to the halos shown in Figs. 3.

in Eq. (2.17) with the helical magnetic field from Eq. (3.4), which has $b_z = 0$, reduces to,

$$b_\rho \cos(\delta/2 - \theta) = 0, \tag{3.5}$$

with

$$b_\rho = \boldsymbol{B} \cdot \hat{\boldsymbol{\rho}} = \sin(2\pi z/\lambda + \phi). \tag{3.6}$$

As $\cos(\delta/2 - \theta) = 0$ has only one solution at $\delta = \pi,\ \theta = 0$ in the physical range $\theta \in [0, \pi/2]$, $\delta \in [0, \pi]$, the surface is mainly determined by $b_\rho = 0$ which translates to,

$$\phi = -\frac{2\pi z}{\lambda}. \tag{3.7}$$

This equation describes a spiral structure as seen in Figure 5.

Until now, we have been assuming that the source emits photons isotropically. Below, we will also consider the case when the source emits photon in a collimated jet. In that case, there is a fourth constraint restricting the relevant part of the PP surface to where it intersects the jet, and it is quite possible that there is no solution. We ignore such cases as they are observationally irrelevant. In following figures we will show PP locations, even if they do not lie within the jet. Only those PP locations that lie within the jet will lead to observed gamma rays. For instance, Fig. 6 presents an example in which the source has a jet with half-opening angle $\theta_{\rm jet} = 5^\circ$ and the magnetic field is given by Eq. (3.4). The jet direction is chosen so that the Earth lies within the cone of the jet and the blazar can be seen directly. The left plots in both figures shows that the jet picks out a small region of the PP surface, and the right plots show the resulting halo.

4 Parameter Dependence of Halo

In this section we discuss the structure of the halo as the parameters B_0, λ and the sign of the helicity of the maximally helical magnetic field in Eq. (3.4) are varied. The concept of the PP surface will be a useful tool for this discussion as it allows us to clearly see how the magnetic field dictates the halo's shape.

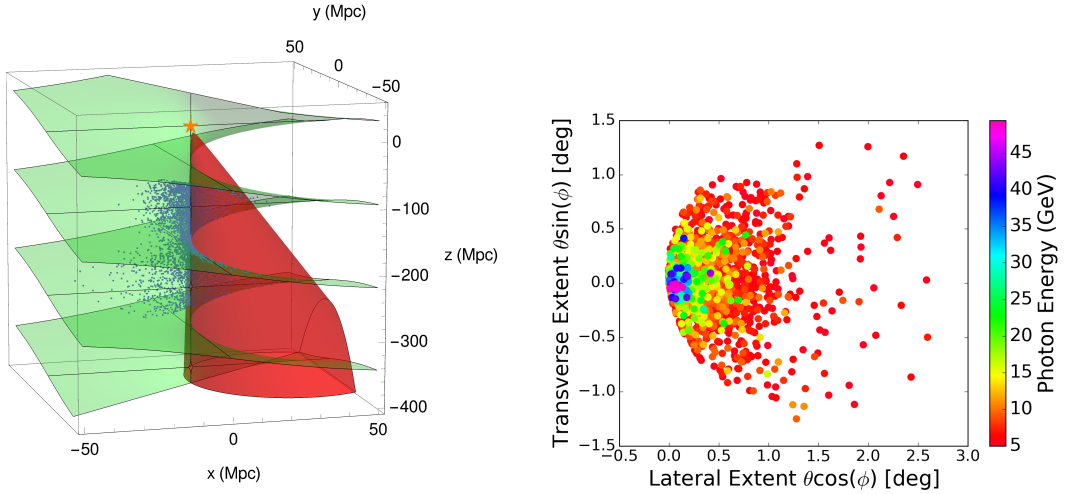

Figure 6. Example of how a blazar with a jet will only shine and activate a small region of the PP surface (left) and the resultant halo (right). The magnetic field is given in Eq. (3.4) with $B_0 = 10^{-14}$G, $\lambda = 250$ Mpc.

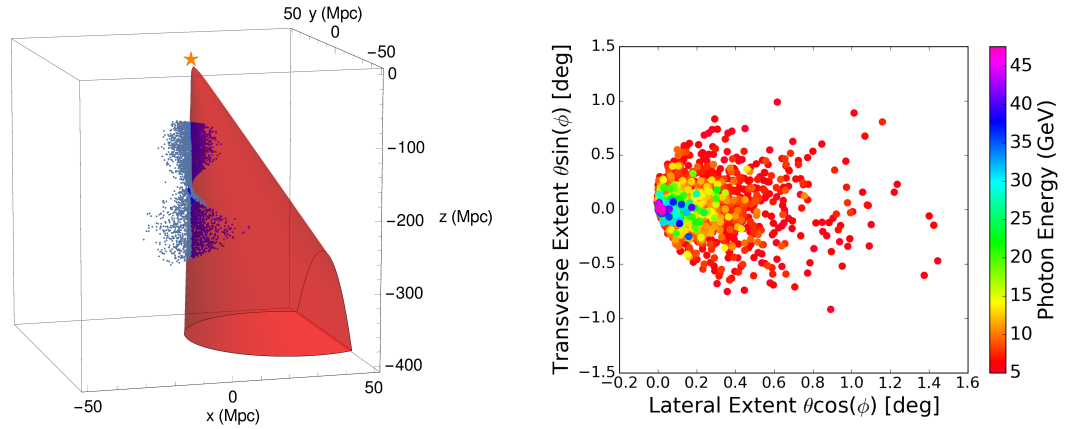

Figure 7. Monte Carlo simulation with stochastic PDs using the magnetic field of Eq. (3.4) and with the same setup as in Fig. 6 but with B_0 reduced to $5 \times 10^{-15}G$. Compared to the right panel of Fig. 6, we see that the high energy gamma rays (blue and green points) are more clustered and so the halo size is smaller at fixed energy.

The magnetic field strength directly affects the amount of bending of the lepton trajectories since the gyroradius $R_L \propto 1/B_0$. Therefore a weak magnetic field will require that the initial TeV gamma ray is already propagating nearly towards Earth. Thus reducing B_0 will shrink the size of the halo at any given gamma ray energy, although lower energy gamma rays may now enter the field of view. This can be seen in Fig. 7 which was created using $B_0 = 5 \times 10^{-15}$ G and $\lambda = 250$ Mpc. The plot looks almost identical to Fig. 6, which was created using $B_0 = 10^{-14}$ G, except that the extent of the halo in the x and y directions, for photons of the same energy, has shrunk by a factor of ~ 2.

If one does not track the photon's energy, the effect of a change in B_0 is not easily

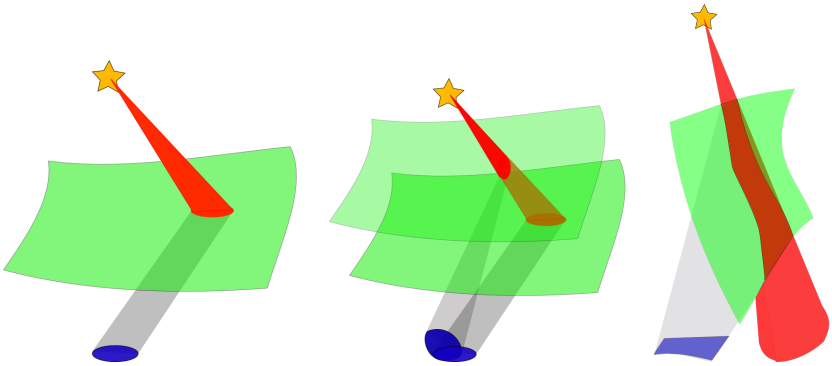

Figure 8. Sketch of a blazar jet shown in red intersects the green PP surface which delimits the shape of the halo (shown in blue) as seen by some observer. The halo photons must be distributed in the blue region. A situation similar to the one depicted in the middle and third sketch can be seen from the simulations in Fig. 6 and Fig. 9 below. The differences between the many possible shapes arise due the characteristics of the intersection between the blazar's jet and the PP surface.

seen through the morphology of the halos as their shapes and sizes are determined by the intersection of the jet and the PP surface. We show a few examples of this interplay in the sketch of Fig. 8. Understanding this could allow us to learn valuable information about the inter-galactic magnetic field in the region probed by the PP locations by observing the halo's shape. In a real situation, we cannot observe the full halo shape as it will be contaminated by background photons coming from other sources. However, we can still extract certain useful halo information since the background is expected to be stochastically isotropic and certainly not parity odd.

Another important thing to note is that we have assumed the jet and the power spectrum to be fixed on the timescale necessary for the creation of the halos. Namely, the path length of two events (i.e. the sum of the magnitude of the two vectors shown in Fig. 1) will differ as a function of their bending angle. Hence this can introduce a significant time delay between the subsequent observations of two initial TeV photons emitted from the source at the same time. If this time delay is large, as would occur for events whose PP locations are Megaparsecs apart, we would expect the source dynamics to alter its power spectrum and jet direction in that timeframe – making our fixed jet assumption false. A quick change in jet direction would translate in observed events arising from potentially very different regions of the PP surface. Fortunately, this should not affect our final results once we average over many realizations as this already stacks random jet oritentations together. A large power spectrum variability could introduce more drastic effects but we will neglect this complication in this initial exploration.

Let us quickly comment on the dependence of the morphology on the coherence length of $\mathbf{B}$. As can be seen from Eq. (2.3), the MFP $D^c_{\gamma 0}$ of the TeV photons are of the order of $10 - 100$ Mpc. Any magnetic field with coherence length much larger than $D^c_{\gamma 0}$ will appear constant in space. On the other hand, for $\lambda_c \ll D^c_{\gamma 0}$, the halo will be produced from a rapidly varying part of the PP surface and will be more scattered.

The helicity of the magnetic field in Eq. (3.4) can be flipped by changing $x \to -x$. A flip in the helicity simply leads to a parity inversion of the PP surface and the halo spiral also changes handedness. However, to get more statistics, we will investigate both helicities using independent simulations in Sec. 6.

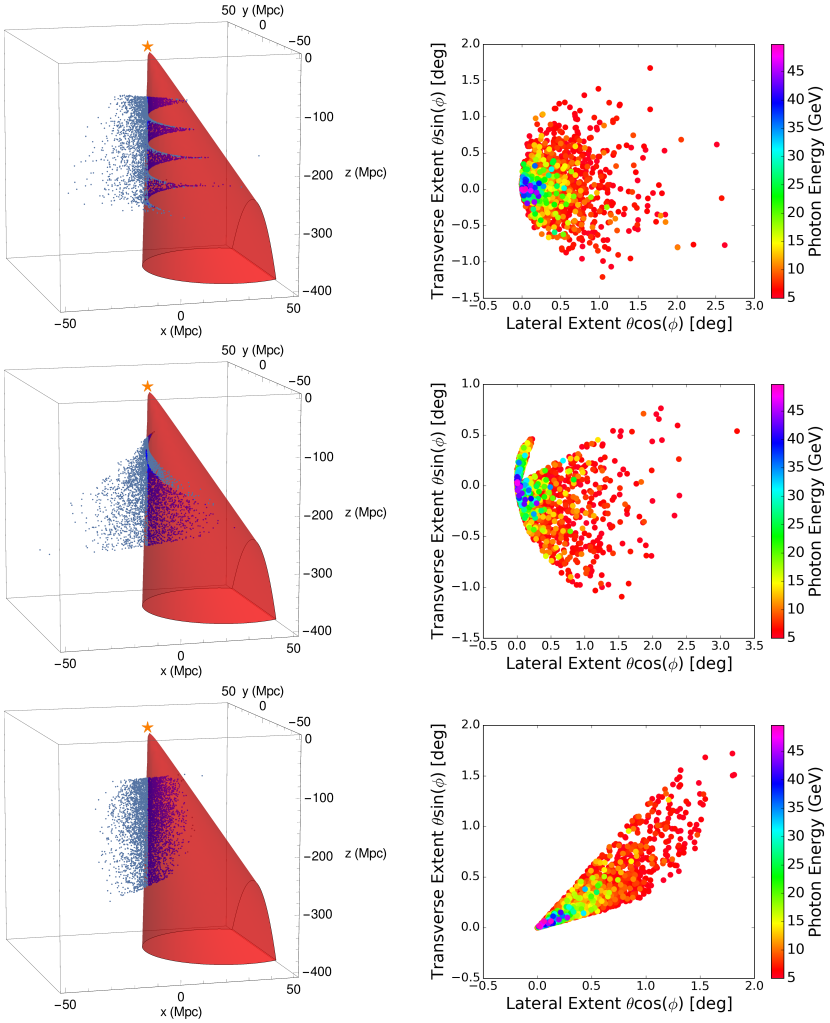

Figure 9. The PP locations on the PP surface and jet (left) and corresponding halo (right) for $\lambda = 100, 500, 2000$ Mpc for the magnetic field in Eq. (3.4) with $B_0 = 10^{-14}$ G. The direction to the source is at $\theta = 0$.

5 The Q statistic

One of the main goals of this work is to determine if the helicity of the inter-galactic magnetic field can be deduced from the shape of the blazar halos. As we have seen, under certain conditions, a helical inter-galactic magnetic field can produce a clear spiral-like structure in a gamma ray halo. Hence it is important to develop a statistical technique that is sensitive to this structure. A statistic, called Q, was developed in Ref. [17], and was applied to the *diffuse* gamma ray background observed by the Fermi telescope in Refs. [4, 5]. A non-zero value of Q was observed with high confidence in comparison to Monte Carlo simulations that assume no inter-galactic magnetic field.

One can see from the halo plots above, *e.g.* Fig. 3, that the arrival direction of high energy photons tend to lie closer to the blazar line-of-sight than those for lower energy photons. Hence different locations of the PP surface are sampled by photons of different energies and the observed gamma rays can carry an imprint of any curvature or twist of the

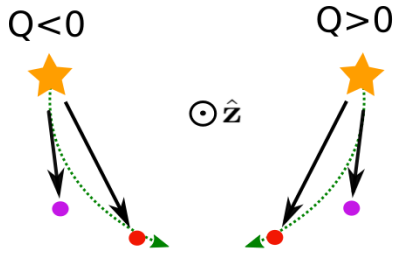

Figure 10. Illustration of the Q-statistic. The plot represents two blazar halos in the observational plane, each halo with just two photons, one at high energy (purple) and the other at low energy (red). The image of the blazars is denoted by the stars and the line of sight to the blazar (along $\hat{\mathbf{z}}$) points out of the plane of the page as denoted by the arrow tip. The left sketch shows a situation where $Q < 0$ since $\mathbf{n}_{\rm red} \times \mathbf{n}_{\rm purple} \cdot \mathbf{n}_{\rm blazar} < 0$; similarly the sketch of the blazar on the right shows a $Q > 0$ situation.

PP surface. More precisely, the work of Ref. [17] showed that a left (right) handed helical magnetic field will create left (right) handed spiral patterns in the observed photons.

Below we briefly review the idea behind a slightly modified version of the Q statistic proposed in Ref. [17]. We will apply the statistics on regions surrounding an observed blazar whose angular position will be denoted by the unit vector $\mathbf{n^{(3)}} = \hat{\mathbf{z}}$. We consider a disk of radius R centered on the location of the source and consider the set of photons within this disk. These photons are binned according to their energies into non-overlapping bins ΔE_1, ΔE_2. We use N_2 to denote the number of photons in bin ΔE_2 within the disk of radius R. We then perform the sum,

$$Q(\Delta E_1, \Delta E_2, R) = -\mathbf{n}^{(3)} \cdot \left(\frac{1}{N_2} \sum_{j=1}^{N_2} \mathbf{n}_j^{(2)} \times \left[\frac{\sum_{i=1}^{N_1} \mathbf{n}_i^{(1)} \Theta(\mathbf{m}_i^{(1)} \cdot \mathbf{m}_j^{(2)})}{\sum_{i=1}^{N_1} \Theta(\mathbf{m}_i^{(1)} \cdot \mathbf{m}_j^{(2)}) + \epsilon} \right] \right) \tag{5.1}$$

where $\mathbf{n}_i^{(a)} \equiv \mathbf{n}_i(\Delta E_a)$ is the unit vector denoting the arrival direction of photon i in bin a; $\mathbf{m}_i^{(a)}$ is the unit vector obtained by projecting $\mathbf{n}_i^{(a)}$ on to the xy-plane: $\mathbf{m}_i^{(a)} = \mathbf{n}_i^{(a)} - (\mathbf{n}_i^{(a)} \cdot \hat{\mathbf{z}})\hat{\mathbf{z}}$. We've also introduced the infinitesimal quantity ϵ to keep the denominator from vanishing. The original Q statistic in Ref. [17] was defined without the Heaviside function (Θ) in Eq. (5.1). We illustrate the $Q-$statistic in Fig. 10.

In the presence of background gamma rays in addition to the blazar gamma rays, we expect $Q(R)$ to start near 0 at $R = 0$, grow to a peak value near $R = R_{\rm halo}$, where $R_{\rm halo}$ is the angular radius of the halo, and finally come back down towards 0 at large R where the signal becomes background dominated. However in mock maps with no background, the value of Q should asymptotically flatten out to its maximal value attained at $R_{\rm halo}$, and its value will be negative (positive) for right (left) handed magnetic fields. We can see this behavior in Fig. 11 where we simulate the halo without any stochasticity and with the magnetic field of Eq. (3.4).

To showcase the Q-statistic in this paper we will separate the gamma rays in three energy bins:

$$\Delta E_1 = (5, 20), \ \Delta E_2 = (20, 35), \ \Delta E_3 = (35, 50), \tag{5.2}$$

all numbers in GeV. This choice was the real reason we only simulated photons between 5 to 50 GeV.

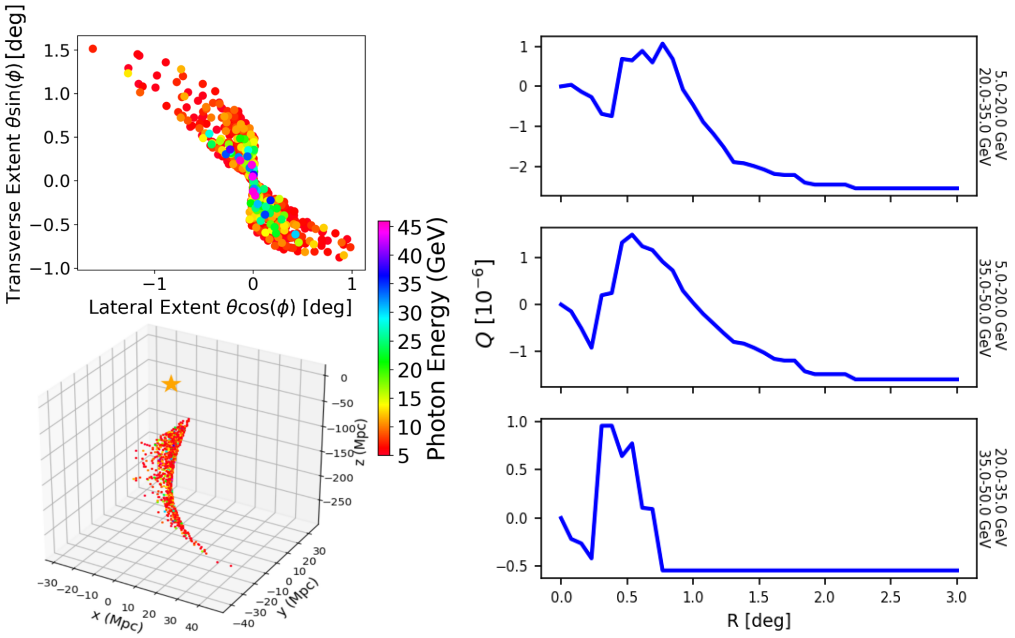

Figure 11. (Top Left) A set of simulated observed photons from the halo formed by a blazar's jet with half-opening angle of 5°. The surrounding random magnetic field was created with parameters described in Eq. (6.2) and has the form given by Eq. (6.1). (Bottom Left) The PP locations of the photons responsible for the halo. (Right) The result of applying the Q-statistic to the observed photons on the right.

6 The Q statistic Applied to Stochastic Magnetic Fields

The result of Fig. 11 is noisy and can be misleading as we are dealing with random magnetic fields. Indeed, these fields can sometimes create halos whose Q-statistics suggest the wrong helicity. It is therefore important to average over many realizations of the magnetic field and the jet orientation. Each realization will simulate a blazar with a jet of half-opening angle $\theta_{\rm jet} = 5°$ and having Earth in its LoS. The jet is also constrained to generate a halo with at least 3 events in order for the statistics to be applied; this condition is easily satisfied if Earth is in the jet's LoS. Jets pointing further away from the LoS might still yield observable photons but we would not be able to identify these blazars and so we don't simulate those cases.

We will consider magnetic fields of the form,

$$\mathbf{B}(\mathbf{x}) = \frac{1}{2N^2+2} \sum_{\mathbf{k}\in K} \mathbf{b}(\mathbf{k}, f_{\rm H}, B_{\rm rms}) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{6.1}$$

with the set K consisting of $2N^2+2$ vectors which have magnitude $k_{\rm mag}$ and whose directions are approximatively uniformly spread over the unit sphere. Half of the Fourier coefficients $\mathbf{b}(\mathbf{k}, f_{\rm H}, B_{\rm rms})$ are drawn from their respective distribution as outlined in Appendix A, while the other half are set by the requirement $\mathbf{b}(\mathbf{k}, f_{\rm H}, B_{\rm rms}) = \mathbf{b}^*(-\mathbf{k}, f_{\rm H}, B_{\rm rms})$, necessary for obtaining a real value for the magnetic field. The value of $-1 \leq f_{\rm H} \leq 1$ controls the

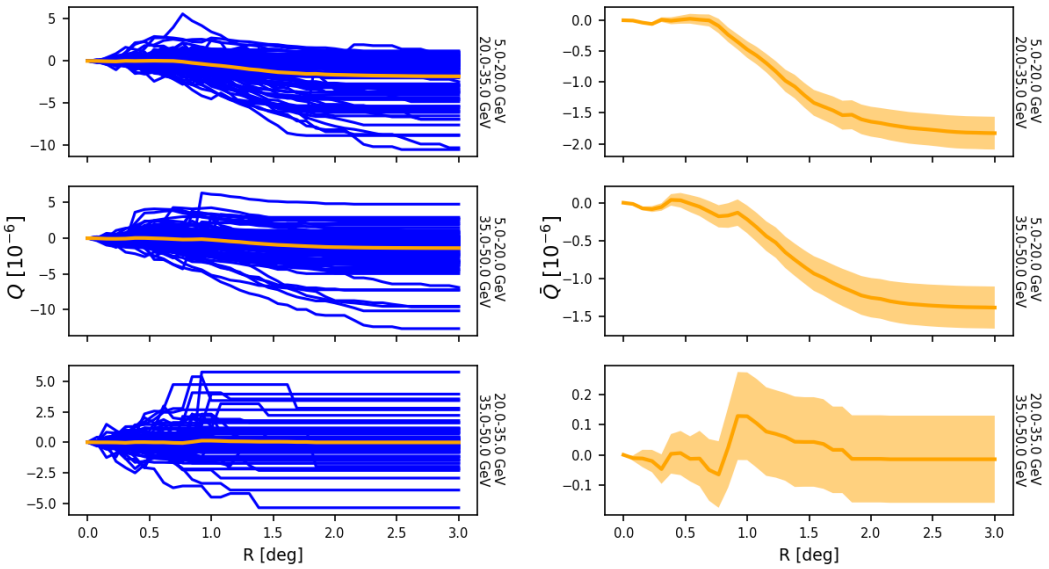

Figure 12. (Left) $Q(R)$ versus R for 100 Monte Carlo runs when the stochastic magnetic fields are generated using the parameters shown in Eq. (6.2) with $f_H = +1$. The mean $\overline{Q}(R)$ is shown by the orange curve. (Right) A zoomed-in view of $\overline{Q}(R)$. The width of the error band is given by the standard error *i.e.* standard deviation of the 100 Monte Carlo $Q(R)$ values divided by the square root of the sample size (100).

handedness of the field, namely $f_{\rm H} = 1$ (-1) corresponds to a maximally right-handed (left-handed) helical field. Finally $B_{\rm rms}$ determines the root mean square of $\mathbf{B}(\mathbf{x})$.

In Fig. 11 we compute the Q-statistics for 100 realizations of halos created with a random magnetic field created using the parameters

$$\begin{aligned} B_{\rm rms} &= 1 \times 10^{-14}\,{\rm G},\ k_{\rm mag} = 0.01/{\rm Mpc}, \\ f_{\rm H} &= +1,\ 2N^2 + 2 = 27, \end{aligned} \tag{6.2}$$

where $2N^2 + 2$ is the number of directions of the $\mathbf{k}$ vector in Eq. (6.1). In Fig. 12 we plot the average of Q over all the realizations, denoted $\overline{Q}(R)$, for the same runs as in Fig. 11 and for the three gamma ray energy bin combinations. By doing so, we have in mind of averaging the Q-statistics obtained from small regions around multiple observed blazars. The plot also shows the standard error in $\overline{Q}(R)$ which is given by the standard deviation of $Q(R)$ divided by the square root of the number of realizations in the Monte Carlo simulations. The standard error follows from the central limit theorem and is the error in using the sample mean to estimate the population mean. However, it assumes that the samples, Monte Carlo simulations in our case, from which values of $Q(R)$ are drawn are independent and identically distributed. This is certainly true in our setup but may not be true for actual observations in which the same photons might contribute to the $Q(R)$ calculated for blazars that are close to each other. In addition, there will be variation in the distance to observed blazars and other source characteristics. We plan to take some of these factors into account in a follow-up analysis.

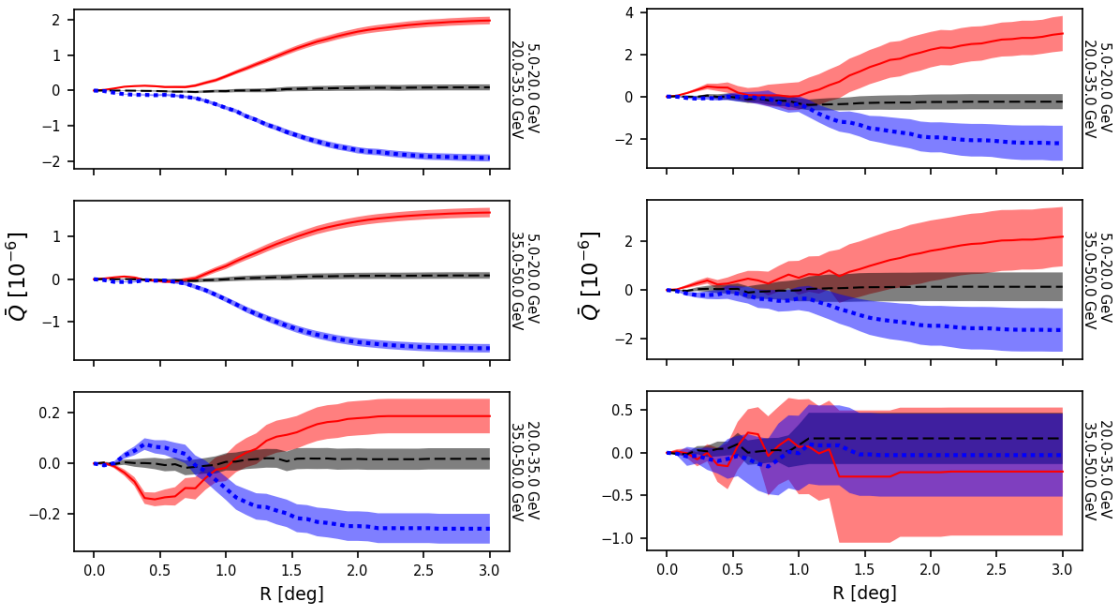

Figure 13. $\overline{Q}(R)$ versus R when averaging for 1000 simulations (left) and 20 simulations (right). The realizations had parameters $f_{\rm H} = -1$ (red, solid), $f_{\rm H} = 0$ (black, dashed) and $f_{\rm H} = +1$ (blue, dotted) for the three energy bin combinations: $(\Delta E_1, \Delta E_2)$ (top row), $(\Delta E_1, \Delta E_3)$ (middle row) and $(\Delta E_2, \Delta E_3)$ (bottom row) where the bins are defined in Eq. (5.2). The other parameters of the stochastic magnetic fields are given in Eq. (6.2) and the bands denote standard error of $\overline{Q}$.

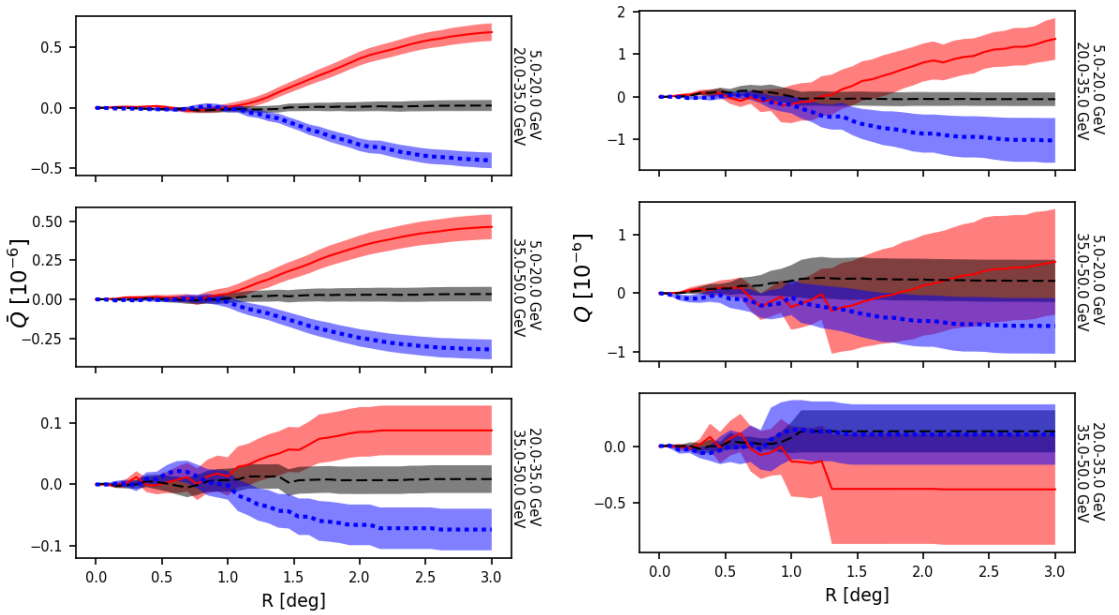

Figure 14. Analysis of the same data as used for Fig. 13 but computed with the original definition of Q. Namely, the Θ term in Eq. (5.1) is simply replaced by the value 1. It is clear both from the magnitude and the size of the error bars that the modified Q is a sharper statistics.

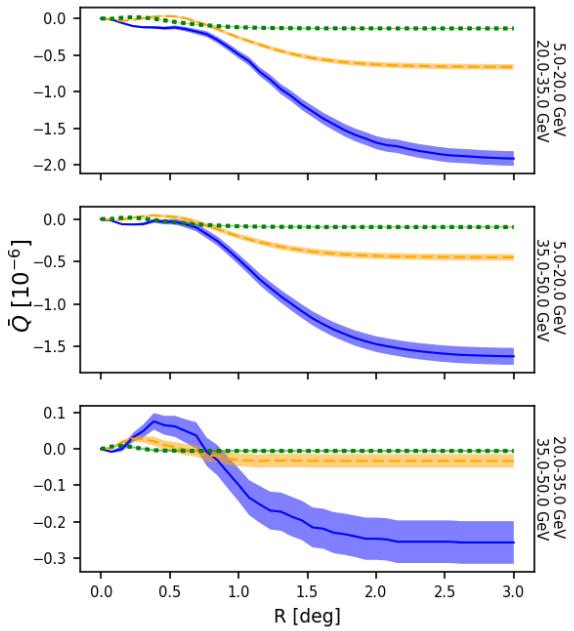

Figure 15. $\overline{Q}$ versus R for 1000 simulations with $B_{\rm rms} = 10^{-14}$ G (blue,solid), 5×10^{-15} G (orange, dashed), and $2 \times 5 \times 10^{-15}$ G (green, dotted) and $f_H = 1$, $k = 0.01$/Mpc.

Pushing the statistics further, we are clearly able to distinguish between many different properties of the magnetic field as the number of realization increases. For instance, we have plotted $\overline{Q}(R)$ versus R for 1000 Monte Carlo simulations for $f_H = 0, \pm 1$ (Fig. 13) and the plots show a clear correlation between $\overline{Q}(R)$ and the helicity of the magnetic field. We can also notice distinct oscillations that occur at small R. It is also reassuring that the $f_H = +1$ plot is the mirror image of the $f_H = -1$ curve, just as we would expect due to parity reflection. Within the clean context studied here, we only need $\sim 10 - 20$ halos before we can detect the sign of the helicity through the sign of the Q's at large R. However this number depends heavily on the properties of the magnetic field, source variability and background. Hence the determination of the exact amount of data required to make such detection will require a careful analysis of these parameters and therefore is relegated to future work.

In Fig. 14, we show that the Θ factor in our definition of Q in Eq. (5.1) improves the resolution. Without the Θ factor, the Q-statistic is determined by the cross product of the *average* arrival direction of the photons (in two energy bins) within a radius R of the source. Because of the electron-positron symmetry, photons tend to arrive on either side of the source (see Fig. 4), and the average arrival direction tends to be near the origin. Introducing the Θ factor ensures that for every high energy photon selected, we only average the low energy photons that arrive on the same side with respect to the source. Then there is a larger contribution to the value of Q. Essentially the Θ term limits the sum to gamma rays within the electron (or the positron) branch of the halo (see Fig. 4).

Next we examine the dependence of $\overline{Q}(R)$ on magnetic field parameters. In Fig. 15 we plot $\overline{Q}(R)$ for several different magnetic field strengths and for fixed helicity $f_H = 1$. We see that increasing the magnetic field strength leads to an increasing amplitude of $\overline{Q}$ for *all* energy combinations. The increase is due to the magnitude of $\mathbf{n_1} \times \mathbf{n_2}$ which becomes larger as the bending allows $\mathbf{n_1}$, $\mathbf{n_2}$ to point further apart.

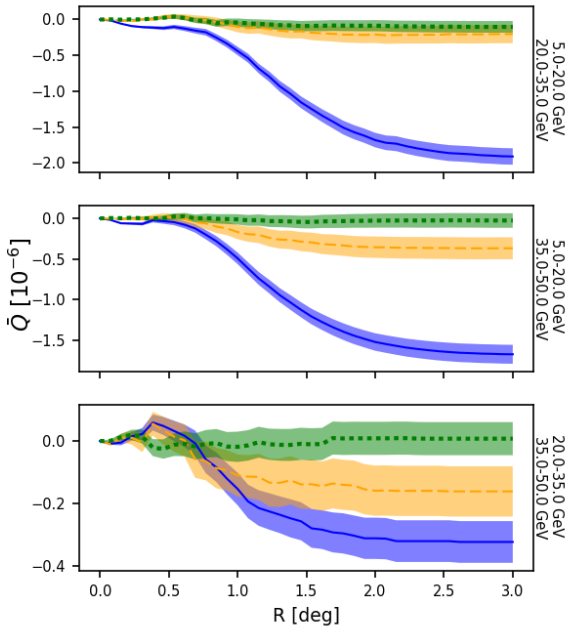

Figure 16. $\overline{Q}$ versus R for 1000 simulations with $k = 0.01$/Mpc (solid, blue), $k = 0.05$/Mpc (orange, dashed), $k = 0.1$/Mpc (green, dotted) with $B_{\rm rms} = 1 \times 10^{-14}$ G and $f_H = +1$.

The effect of changing the magnetic field coherence scale is shown in Fig. 16 where the magnetic field strength and other parameters are fixed and only $k_{\rm mag}$ is varied. The magnetic field with larger coherence length gives a larger signal, but there is a turning point as extremely large coherence scale fields will behave like uniform fields. The signal for smaller correlation length is washed out but the suppression depends on the energy combination. This is to be expected from the analysis of Ref. [17] since $\overline{Q}$ with a certain energy combination is sensitive to the magnetic helicity power spectrum at a definite coherence scale that is determined by the combination of energies. To probe magnetic fields on small length scales, it is necessary to consider gamma rays whose energies are close together [17]. Thus the energy bins also have to be smaller and this means that the statistics is poorer.

7 The features at small R

The Q-statistic is essentially a measure of the differential rotation found in the arrival direction of photons of different energies. Q will be negative (positive) if the rotation is right (left)-handed as is depicted in Fig. 10. The results of our Monte Carlo simulations, for example in Fig. 13, show that $\overline{Q}$ has some oscillation at small R which is made abundantly clear for the $\overline{Q}$ using the bin with highest energies. Here we provide an explanation of this small R feature. What the Q-statistic allow us to probe is the shape of the PP surface. As already mentioned, it is clear from Fig. 11 that high energy photons are found close to the LoS and the low energy ones are further out. However there is another important piece of information, namely the z coordinate of their PP locations.

Remember that to give a contribution to Q, one requires a high energy photon (γ_{HE}) and a low energy photon (γ_{LE}) as is shown in Fig. 10. If most γ_{LE} photons have PP locations higher up on the PP surface than the γ_{HE}'s, the Q-statistic measures the twist of the PP

surface as one traverses it from bottom to the top. This is in contrast to whenever the low energy photons originate from PP locations close to Earth when compared to those of γ_{HE}. As the twist is parity odd, these two cases contribute to $\overline{Q}$ with opposite signs and is the reason for these oscillations.

We can understand this effect explicitly with a little more thought. When R is very small, we expect to see events with small bending angles. These mainly occur when the lepton upscatters a photon towards Earth early after it was pair produced. The small R observed γ_{HE} (γ_{LE}) photons must therefore have originated from TeV leptons with high (low) energies. Because TeV gamma ray have a MFP that decreases with energy, we then expect the γ_{LE} to be produced at PP locations closer to Earth than the γ_{HE} ones, therefore Q initially measures the twist from top to bottom. Assuming that the helicity of the magnetic field is right-handed, the value of Q will become increasingly negative as R departs from 0.

However as R increases further, two new things occur. First, the observed photons entering the field of view come from events who experiences more bending. This means that the lepton had to travel further and in the process lost more energy, therefore the maximal energy of the new photons will be lower than those observed at small R and won't contribute to the highest energy bin. Hence the average z of the PP location of the high energy photons is determined by the events at small R. Second, most of the volume entering the field of view will be located closer to the source due to projection effects; the observed volume is shown by the green cone of Fig. 17. We can then statistically expect the newly observed photons to have PP locations located near $z = 0$. We are now in the reverse situation: a large majority of photons γ_{LE} have PP locations that are higher than the ones for the γ_{HE}. These give a contribution to Q with the opposite sign, pulling the its value toward 0 and sometime even changes its overall sign (compare the top and bottom panel of Fig. 13). Finally when R becomes large, the low energy events with PP locations at small z and far from the LoS start contributing to Q and dominate. This process, depicted in Fig. 17, is responsible for the features before the value of Q flattens out to its asymptotic value.

8 Conclusions

We have studied the effect of stochastic inter-galactic magnetic fields on the morphology of gamma ray halos. The dependence of the morphology on the magnetic field strength, the coherence length, and the helicity were investigated. Most importantly, we have provided an understanding of the structure of the halo in geometrical terms, as arising due to the "PP surface" as determined by the magnetic field. In simple cases, the PP surface can be found analytically (for example, Eq. (3.7)).

To analyze the halo morphology, we have proposed a sharper version of the Q-statistic in Eq. (5.1) and applied it to simulated halos. Our key finding is that Q is a powerful diagnostic of the magnetic helicity (Fig. 13), field strength (Fig. 15) and coherence scale (Fig. 16). Based on the analytical work of Ref. [17], we expect the sensitivity of Q to the coherence scale to depend crucially on the energies of the gamma rays that are used. It would be interesting to quantitatively examine how the sensitivity of Q to the coherence scale can be improved with a choice of energy bins.

In addition, our Monte Carlo simulations have revealed a bump in $\overline{Q}(R)$ at small values of R (see Fig. 13). We have understood and explained this feature in terms of the PP surface in Sec. 7. This new feature may become useful in the analysis of real data in the future.

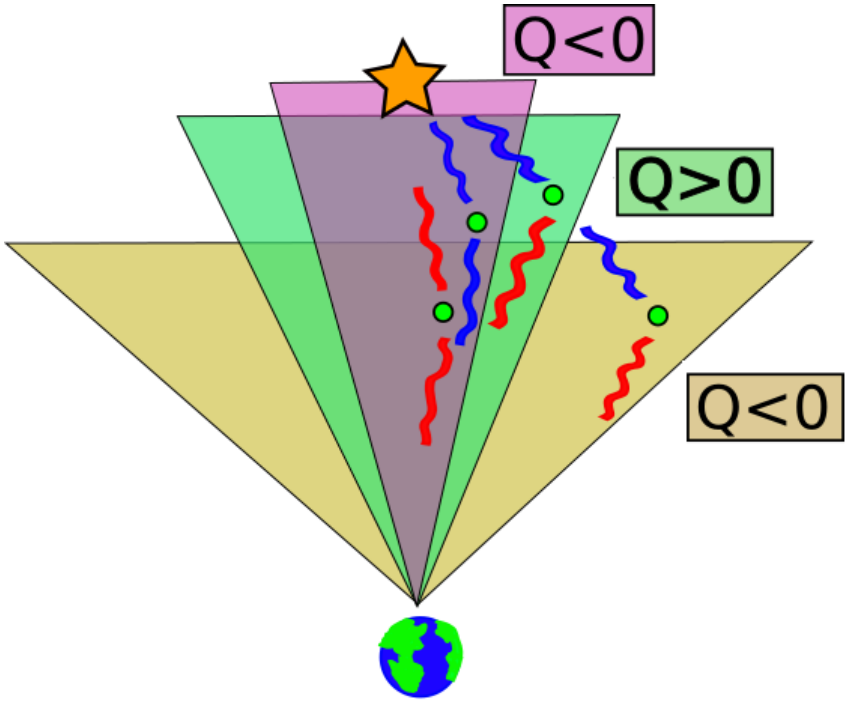

Figure 17. The sign of the Q-statistic when applied to a halo produced by a magnetic field with $f_H = 1$. Here the squiggly lines represents photons of high (blue) and low (red) energies. The top photons are the initial TeV gamma rays, the green points are the PP locations and the bottom photons are the upscattered GeV photons that are eventually observed. At small R, the events with small bending allows the Q-statistic to measure the twist of the PP surface from top to bottom (purple shaded region, $Q < 0$). When R gets larger and PP surface region near the blazar enters the field of view, the new events entering the field of view contribute to Q with the opposite sign as they are located further up the PP surface than the high energy events which only occurs close to the LoS (green shaded region, $Q > 0$). Finally as R gets large and the whole halo is exposed, the overwhelming low energy events at low z and far from the LoS dominate the signal and drive Q toward its asymptotic value (brown shaded region, $Q < 0$).

Our present study is limited in a few ways that we plan to overcome in future work. First, we have not included any background gamma rays. These will introduce noise in the evaluation of $\overline{Q}$ and the error bars will increase. We have also limited ourselves to stochastic isotropic magnetic fields but with only one $|\mathbf{k}|-$mode. This is useful at this stage as it allows us to diagnose the effects of changing the coherence scale. In future, we plan to include a spectrum of the magnetic field as motivated by current observations [11]. In future we also plan to incorporate the full development of the electromagnetic cascade into our numerical code, perhaps along the lines of Ref. [14] or [16]. Once we have understood individual blazar halos, we will apply our techniques to the diffuse gamma ray background which is expected to contain halos due to unseen blazars as well as those due to identified blazars.

Acknowledgement

We thank Kohei Kamada and Andrew Long for comments. TV is supported by the U.S. Department of Energy, Office of High Energy Physics, under Award No. DE-SC0013605 at

ASUs.

A Generation of Isotropic Random Magnetic Fields

To generate helical magnetic fields, we first decompose the magnetic field $\mathbf{B}(\mathbf{x})$ in circularly polarized modes with basis vectors $\mathbf{K}^{\pm}(\mathbf{k})$ that are divergence-free eigenfunctions of the Laplace operator

$$\mathbf{K}^{\pm}(\mathbf{k}) = \mathbf{e}^{\pm}(\mathbf{k})\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}} \equiv \frac{\mathbf{e_1}(\mathbf{k}) \pm i\mathbf{e_2}(\mathbf{k})}{\sqrt{2}}\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}}. \tag{A.1}$$

The triad of unit vectors, $\{\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}\}$, is constructed as

$$\mathbf{e}_1 \equiv \frac{\mathbf{n}_0 \times \hat{\mathbf{k}}}{|\mathbf{n}_0 \times \hat{\mathbf{k}}|},\ \mathbf{e}_2 \equiv \frac{\hat{\mathbf{k}} \times \mathbf{e}_1}{|\hat{\mathbf{k}} \times \mathbf{e}_1|},\ \mathbf{e}_3 = \frac{\mathbf{k}}{k} \equiv \hat{\mathbf{k}} \tag{A.2}$$

where $\mathbf{n}_0$ is any chosen unit vector such that $\mathbf{n}_0 \neq \hat{\mathbf{k}}$.

With these definitions, the $\mathbf{e}$'s form a right-handed orthonormal system and we have,

$$\nabla \cdot \mathbf{K}^{\pm} = 0,\ \nabla \times \mathbf{K}^{\pm} = \pm k\mathbf{K}^{\pm},\ \mathbf{K}^{\pm *}(\mathbf{k}) = -\mathbf{K}^{\pm}(-\mathbf{k}) \tag{A.3}$$

Hence any magnetic field can be decomposed as,

$$\begin{aligned}\mathbf{B}(\mathbf{x}) &= \int \frac{d^3k}{(2\pi)^3}\mathbf{b}(\mathbf{k})\mathrm{e}^{i\mathbf{k}\cdot\mathbf{x}} \\ &= \int \frac{d^3k}{(2\pi)^3}\left[b^+\mathbf{K}^+ + b^-\mathbf{K}^-\right],\end{aligned} \tag{A.4}$$

with the condition that

$$b^{\pm *}(\mathbf{k}) = -b^{\pm}(-\mathbf{k}) \tag{A.5}$$

to ensure that $\mathbf{B}(\mathbf{x})$ is real. The divergence-free condition, $\nabla \cdot \mathbf{B} = 0$, is automatically satisfied in this procedure.

We are interested in generating random magnetic fields with given energy ($E_B(k)$) and helical ($H_B(k)$) power spectra. The relations between the modes $b^{\pm}(\mathbf{k})$ and the power spectra are given by

$$\frac{1}{8\pi}\langle|\mathbf{B}(\mathbf{x})|^2\rangle = \int \frac{k^2dk}{16\pi^3}\left[|b^+|^2 + |b^-|^2\right] \equiv \int E_B(k)d\ln(k) \tag{A.6}$$

and

$$\langle\mathbf{A}(\mathbf{x}) \cdot \mathbf{B}(\mathbf{x})\rangle = \int \frac{kdk}{2\pi^2}\left[|b^+|^2 - |b^-|^2\right] \equiv \int H_B(k)d\ln(k). \tag{A.7}$$

The ratio of E_B and H_B will be written in terms of a function $f_H(k)$ as [8],

$$H_B(k) = f_H(k)\frac{8\pi}{k}E_B(k), \tag{A.8}$$

and the "realizability condition" leads to the restriction

$$-1 \leq f_H(k) \leq 1 \tag{A.9}$$

The field is non-helical if $f_H = 0$, maximally right-handed if $f_H = +1$, and maximally left-handed if $f_H = -1$.

Eqs. (A.6), (A.7) and (A.8) allow us to write,

$$|b^{\pm}|^2 = \left(\frac{2\pi}{k}\right)^3 [1 \pm f_H(k)] E_B(k). \tag{A.10}$$

Hence the modes $|b^{\pm}(\mathbf{k})|$ are drawn from a normal distribution with mean $\mu^{\pm} = 0$ and standard deviation $\sigma^{\pm} = (1 \pm f_H)(2\pi/k)^3 E_B(k)$. We then include a uniformly drawn phase angle $\theta^{\pm}(\mathbf{k}) \in [0, 2\pi)$ which yields,

$$b^{\pm}(\mathbf{k}) = |b^{\pm}(\mathbf{k})| e^{i\theta^{\pm}(\mathbf{k})} \tag{A.11}$$

In this paper we focus on stochastic magnetic fields that are *isotropic* but have power on a single length scale $\lambda_c = 2\pi/k_{\rm mag}$ and that have either $f_H(k) = 0$ or $f_H(k) = \pm 1$. This corresponds to a delta function distribution for $E_B(k)$ and vanishing or maximal helicity of either sign. To ensure that the magnetic fields are stochastically isotropic, we choose $N^2 + 1$ vectors $\mathbf{k}_n$ $(n = 1, ..., N^2 + 1)$ that discretize half of the two-sphere of directions in $\mathbf{k}-$space,

$$\mathbf{k}_n = k_{\rm mag}(\sin\theta_i \cos\phi_j, \sin\theta_i \sin\phi_j, \cos\theta_i), \tag{A.12}$$

with

$$\theta_i = \cos^{-1}\left(\frac{2i-1}{N} - 1\right), \ \phi_j = 2\pi\frac{(j-1)}{N}, \tag{A.13}$$

for $i, j = 1, \ldots, N$, and

$$k_{N^2+1} = k_{\rm mag}(0, 0, 1) \tag{A.14}$$

Once we have $\mathbf{k}$ and $b^{\pm}(\mathbf{k})$ as described above, we compute

$$\mathbf{b}(\mathbf{k}) = b^{+}\mathbf{K}^{+} + b^{-}\mathbf{K}^{-} \tag{A.15}$$

for every $\mathbf{k} = \mathbf{k}_n$. We also find $\mathbf{b}(-\mathbf{k})$ using the reality condition

$$\mathbf{b}(-\mathbf{k}) = \mathbf{b}^{*}(\mathbf{k}). \tag{A.16}$$

Finally we obtain the random magnetic field,

$$\mathbf{B}(\mathbf{x}) = \frac{\mathbf{1}}{(\mathbf{2N^2 + 2})} \sum_{\mathbf{k}\in\mathbf{K}} \mathbf{b}(\mathbf{k}) \mathbf{e}^{\mathbf{ik\cdot x}} \tag{A.17}$$

where K is the set of vectors $\{\mathbf{k}_n, -\mathbf{k}_n\}$ for $n = 1, \ldots, N^2 + 1$.

References

[1] A. Neronov and I. Vovk, Science **328**, 73 (2010) doi:10.1126/science.1184192 [arXiv:1006.3504 [astro-ph.HE]].

[2] S. Ando and A. Kusenko, Astrophys. J. **722**, L39 (2010) doi:10.1088/2041-8205/722/1/L39 [arXiv:1005.1924 [astro-ph.HE]].

[3] W. Essey, S. Ando and A. Kusenko, Astropart. Phys. **35**, 135 (2011) doi:10.1016/j.astropartphys.2011.06.010 [arXiv:1012.5313 [astro-ph.HE]].

[4] H. Tashiro, W. Chen, F. Ferrer and T. Vachaspati, Mon. Not. Roy. Astron. Soc. **445**, no. 1, L41 (2014) doi:10.1093/mnrasl/slu134 [arXiv:1310.4826 [astro-ph.CO]].

[5] W. Chen, B. D. Chowdhury, F. Ferrer, H. Tashiro and T. Vachaspati, Mon. Not. Roy. Astron. Soc. **450**, no. 4, 3371 (2015) doi:10.1093/mnras/stv308 [arXiv:1412.3171 [astro-ph.CO]].

[6] W. Chen, J. H. Buckley and F. Ferrer, Phys. Rev. Lett. **115**, 211103 (2015) doi:10.1103/PhysRevLett.115.211103 [arXiv:1410.7717 [astro-ph.HE]].

[7] J. D. Finke, L. C. Reyes, M. Georganopoulos, K. Reynolds, M. Ajello, S. J. Fegan and K. McCann, Astrophys. J. **814**, no. 1, 20 (2015) doi:10.1088/0004-637X/814/1/20 [arXiv:1510.02485 [astro-ph.HE]].

[8] A. Brandenburg and K. Subramanian, Phys. Rept. **417**, 1 (2005) doi:10.1016/j.physrep.2005.06.005 [astro-ph/0405052].

[9] R. Durrer and A. Neronov, Astron. Astrophys. Rev. **21**, 62 (2013) doi:10.1007/s00159-013-0062-7 [arXiv:1303.7121 [astro-ph.CO]].

[10] J. M. Wagstaff and R. Banerjee, JCAP **1601**, 002 (2016) doi:10.1088/1475-7516/2016/01/002 [arXiv:1409.4223 [astro-ph.CO]].

[11] T. Vachaspati, arXiv:1606.06186 [astro-ph.CO].

[12] A. Elyiv, A. Neronov and D. V. Semikoz, Phys. Rev. D **80**, 023010 (2009) doi:10.1103/PhysRevD.80.023010 [arXiv:0903.3649 [astro-ph.CO]].

[13] A. J. Long and T. Vachaspati, JCAP **1509**, no. 09, 065 (2015) doi:10.1088/1475-7516/2015/09/065 [arXiv:1505.07846 [astro-ph.CO]].

[14] R. Alves Batista, A. Saveliev, G. Sigl and T. Vachaspati, Phys. Rev. D **94**, no. 8, 083005 (2016) doi:10.1103/PhysRevD.94.083005 [arXiv:1607.00320 [astro-ph.HE]].

[15] A. E. Broderick, P. Tiede, M. Shalaby, C. Pfrommer, E. Puchwein, P. Chang and A. Lamberts, arXiv:1609.00387 [astro-ph.HE].

[16] T. Fitoussi, R. Belmont, J. Malzac, A. Marcowith, J. Cohen-Tanugi and P. Jean, doi:10.1093/mnras/stw3365 arXiv:1701.00654 [astro-ph.HE].

[17] H. Tashiro and T. Vachaspati, Phys. Rev. D **87**, no. 12, 123527 (2013) doi:10.1103/PhysRevD.87.123527 [arXiv:1305.0181 [astro-ph.CO]].

[18] P. A. R. Ade *et al.* [Planck Collaboration], Astron. Astrophys. **594**, A13 (2016) doi:10.1051/0004-6361/201525830 [arXiv:1502.01589 [astro-ph.CO]].

[19] A. Neronov and D. Semikoz, Phys.Rev. D80 (2009) 123012, arXiv:0910.1920.

[20] M. Ackermann *et al.* [Fermi-LAT Collaboration], Astrophys. J. Suppl. **222**, no. 1, 5 (2016) doi:10.3847/0067-0049/222/1/5 [arXiv:1508.04449 [astro-ph.HE]].

[21] M. Ackermann *et al.* [Fermi-LAT Collaboration], Phys. Rev. Lett. **116**, no. 15, 151105 (2016) doi:10.1103/PhysRevLett.116.151105 [arXiv:1511.00693 [astro-ph.CO]].