Inflationary theory and pulsar timing investigations of primordial black holes and gravitational waves

Nicholas Orlofsky, Aaron Pierce, and James D. Wells

Inflationary theory and pulsar timing investigations of primordial black holes and gravitational waves

Nicholas Orlofsky[a], Aaron Pierce[a], James D. Wells[a,b]

[a]*Michigan Center for Theoretical Physics (MCTP)*
Department of Physics, University of Michigan,
Ann Arbor, MI 48109 USA

[b]*Deutsches Elektronen-Synchrotron (DESY),*
Notkestraße, D-22607, Hamburg, Germany

The gravitational waves measured at LIGO are presumed here to come from merging primordial black holes. We ask how these primordial black holes could arise through inflationary models while not conflicting with current experiments. Among the approaches that work, we investigate the opportunity for corroboration through experimental probes of gravitational waves at pulsar timing arrays. We provide examples of theories that are already ruled out, theories that will soon be probed, and theories that will not be tested in the foreseeable future. The models that are most strongly constrained are those with a relatively broad primordial power spectrum.

I.   INTRODUCTION

The LIGO detection of gravitational waves (GWs) resulting from the merger of black holes with masses ~ 10 to $30 M_\odot$ (where $M_\odot$ is a solar mass) [1, 2] was an important validation of general relativity. In addition, the merger rate observed by LIGO may be consistent with primordial black holes (PBHs) (for PBH reviews, see, *e.g.*, [3, 4]) making up all [5] or a fraction $\sim 10^{-3}$ [6, 7] of the observed dark matter (DM) density. In this study, we investigate the consequences for theory and experiment by assuming the PBH merger explanation.

Primordial black holes within this mass range are subject to several constraints, especially if the fractional relic abundance $f = \Omega_{\rm PBH}/\Omega_{\rm DM} \gtrsim 0.1$. Microlensing measurements [8–10] constrain black hole masses $M_{\rm PBH} \lesssim 30 M_\odot$, with the strongest constraints coming at $M_{\rm PBH} \lesssim M_\odot$. Constraints from the ultra-faint dwarf Eridanus II [11] give complementary bounds for $M_{\rm PBH} \gtrsim 5$ to $30 M_\odot$, depending on assumptions about the density and velocity dispersion of DM. These constraints when taken together may allow a monochromatic PBH mass spectrum at $M_{\rm PBH} = 30 M_\odot$ or allow an extended distribution with $f = \Omega_{\rm PBH}/\Omega_{\rm DM} \lesssim 0.1$ over a wide range of masses around this. Bounds from these experiments on extended mass spectra are discussed in [4, 12]. Bounds may also be placed on PBHs with mass $M_{\rm PBH} \sim 1$ to $1000 M_\odot$ moving relative to pulsar lines of sight [13]. In addition to these, there are strong bounds from WMAP and FIRAS [14] and Planck [15] that may limit $f \lesssim 10^{-2}$ to 10^{-4} around these PBH masses, though these bounds have been disputed [5]. Such a low abundance, nevertheless, may be consistent with the observed rate at LIGO if the higher merger rate estimates such as those given in Refs. [6, 7] obtain.

Let us comment on estimations of PBH merger rates found in the literature. In calculating the merger rate, Ref. [5] estimates binary formation inside DM halos today, while Refs. [6, 7] presume that binaries formed primordially. In the primordial case, they estimate the probability for pairs of PBHs to gravitate enough to decouple from Hubble expansion (see also earlier work by [16, 17]). The estimated rate for primordially-produced binary mergers is much higher than the rate of mergers from binaries formed in halos, so the mergers of primordially produced binaries may be expected to dominate.

However, there are several assumptions and quantities that can only be calculated numerically that lead to uncertainty in the rate of mergers from primordially-formed binaries. In isolation, a primordial PBH pair would merge essentially immediately after they decouple from the Hubble flow, with a time scale set by the gravitational free-

fall time. However, other PBHs may perturb the binary system by creating tidal forces that lead the pair to form eccentric orbits. Once in this eccentric orbit, the time scale to merge via the emission of gravitational radiation is significantly prolonged. To estimate this new infall time, the semi-major and -minor axes for these binaries must be computed, a point on which Refs. [7, 17] disagree numerically (although Ref. [17] uses a more realistic Poisson probability distribution compared to the flat distribution in [6, 7]).

Notably, no uncertainties are given for these calculations. To estimate the theoretical uncertainties, we vary the parameters of the model within an order of magnitude.[1] The uncertainty of the merger rate translates into a range of PBH DM fractions f consistent with the LIGO rate from $f \sim 10^{-3}$ to 10^{-1}. The theoretical uncertainties that may be included in this estimate include those discussed in [17], such as the treatment of angular dependence, three-body collisions, other fluctuations beyond the three bodies considered, initial conditions, and radiation drag. Additionally, all these calculations assume an idealized monochromatic mass spectrum of PBHs. Finally, there is a further question of whether these binaries, which initially have major axes $\gtrsim$ the size of our solar system, would survive the process of halo formation. These questions require further detailed numerical study beyond the scope of this work. An interesting observation is that this collection of binary merger assumptions appears inconsistent with PBH DM fraction $f = 1$. For our purposes, we simply note the present theoretical and experimental uncertainties may allow a wide range of PBH abundance. For concreteness, we focus on the case where PBHs make up all of the DM, which might obtain if the above mechanism were ineffective (and in which case the mechanism of [5] might dominate), but we will see that reducing the PBH abundance by even several orders of magnitude will have only a modest impact on the primary experimental probe we investigate in this paper and now introduce.

If these black holes were produced primordially, they would have resulted from the collapse of large density perturbations. While these would be sourced as scalar perturbations, because of their size they can lead to non-trivial tensor perturbations at second order in cosmological perturbation theory [18–22]. A key observation is PBHs in the mass range detected by LIGO generate tensor perturbations that may be detected as gravitational waves at pulsar timing arrays (PTAs) [23].

To make a clear distinction from the GWs detected at LIGO, we call the GWs probed at PTAs "secondary gravitational waves" (SGWs). SGWs that are correlated with the LIGO GW signal may or may not be detectable at PTAs depending on the details of the formation mechanism of the PBHs. For example, it has been observed that for scalar perturbations that are highly peaked—behaving essentially like Dirac-δ functions—PTA probes of SGWs already exclude the formation of PBHs with masses in the range $10^{-2}M_\odot \lesssim M_{\mathrm{PBH}} \lesssim 10M_\odot$ [24–26]. However, constraints have not been applied to explicit models for forming PBHs with non-idealized scalar power spectra except recently in [27, 28].

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

---

[1] Specifically, we vary α and β as defined in [17] between 0.1 and 10 and use a Poisson probability distribution.

The remainder of this paper is organized as follows. In Section II we review the calculation of PBH and SGW spectra from a primordial scalar spectrum. Section III gives a demonstration of the SGW spectrum for an idealized δ function primordial spectrum. Section IV discusses explicit modes for PBH formation and places bounds on their resulting SGW spectra. In Section V we discussion assumptions and uncertainties in our calculations and their effects on the bounds. We conclude in Section VI.

II. PBH AND GW SPECTRA FROM A PRIMORDIAL SCALAR SPECTRUM

We consider a primordial curvature perturbation spectrum $\mathcal{P}_{\mathcal{R}}(k)$ whose form is determined by early universe dynamics. Our approach is to ensure that PBHs are formed in the right mass range and with the right abundance to explain the LIGO GW signal. The spectrum then has the potential to give rise to a background of stochastic SGWs, whose rate and strength depend on details of the inflationary theory. These details will be described in Section IV. Here we review the formalism for determining PBH formation and SGWs from a generic perturbation spectrum.

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

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

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

Assuming these perturbations are Gaussian, the energy fraction of PBHs with mass in the interval $(M, M+dM)$ at their formation time t_f can be calculated using Press-Schechter formalism [30, 31],

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

where [32]

$$\sigma_{R_M}^2 = \int \frac{dk}{k}W(kR_M)^2 \mathcal{P}_{\delta}(k) = \int \frac{dk}{k}W(kR_M)^2 \left(\frac{4}{9}\right)(kR_M)^4 \mathcal{P}_{\Phi}(k) \tag{3}$$

is the variance for the Gaussian probability distribution for primordial density perturbations on length scale R_M. Here, $W(x) = e^{-x^2/2}$ is a Gaussian window function. We take the threshold for PBH formation $\delta_c = 1/3$ [33] in (2) which will result in a conservative estimate of the SGW abundance, though higher values of e.g. $\delta_c \simeq 0.45$ have also been indicated [34], and non-sphericity effects [34–37] can make δ_c higher still. See the Discussion section for the effect of these different choices. Finally, the relic abundance of PBH today is given by

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

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

----

[2] More properly, $\mathcal{P}_{\delta}(k) = \frac{16}{3}\left(\frac{k}{aH}\right)^2 j_1^2\left(\frac{k}{\sqrt{3}aH}\right)\mathcal{P}_{\mathcal{R}}(k)$ [29], where j_1 is a spherical Bessel function, though the difference mainly appears on sub-horizon scales that are suppressed in (3) by the window function $W(kR)$.

The second-order tensor power spectrum can be calculated from the scalar power spectrum using [25, 26, 38–41],[3]

$$\mathcal{P}_h(k,\eta) = \int_0^\infty d\tilde{k} \int_{-1}^1 d\mu \; \mathcal{P}_\Phi(|\mathbf{k}-\tilde{\mathbf{k}}|)\mathcal{P}_\Phi(\tilde{k})\mathcal{F}(k,\tilde{k},\mu,\eta), \tag{5}$$

where

$$\mathcal{F}(k,\tilde{k},\mu,\eta) = \frac{(1-\mu^2)^2}{a^2(\eta)} \frac{k^3\tilde{k}^3}{|\mathbf{k}-\tilde{\mathbf{k}}|^3} \int_0^\eta d\eta_1 \; a(\eta_1)g_k(\eta,\eta_1)f(\mathbf{k},\tilde{\mathbf{k}},\eta_1) \int_0^\eta d\eta_2 \; a(\eta_2)g_k(\eta,\eta_2) \left[f(\mathbf{k},\tilde{\mathbf{k}},\eta_2) + f(\mathbf{k},\mathbf{k}-\tilde{\mathbf{k}},\eta_2) \right], \tag{6}$$

with

$$f(\mathbf{k},\tilde{\mathbf{k}},\eta) = 12\Phi(\tilde{k}\eta)\Phi(|\mathbf{k}-\tilde{\mathbf{k}}|\eta) + 8\eta\Phi(\tilde{k}\eta)\Phi'(|\mathbf{k}-\tilde{\mathbf{k}}|\eta) + 4\eta^2\Phi'(\tilde{k}\eta)\Phi'(|\mathbf{k}-\tilde{\mathbf{k}}|\eta). \tag{7}$$

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

$$\Phi(\mathbf{k},\eta) = \frac{A(\mathbf{k})}{(\sqrt{w}k\eta)^2} \left(\frac{\sin(\sqrt{w}k\eta)}{\sqrt{w}k\eta} - \cos(\sqrt{w}k\eta) \right), \tag{8}$$

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

$$\langle \Phi(\mathbf{k})\Phi(\mathbf{k}') \rangle = \frac{2\pi^2}{k^3}\delta^3(\mathbf{k}+\mathbf{k}')\mathcal{P}_\Phi(k). \tag{9}$$

The Green's function in (6) is

$$g_k(\eta,\eta') = \frac{\sin(k(\eta-\eta'))}{k}. \tag{10}$$

Finally, the relic abundance of gravitational waves can be calculated in terms of Eqn. (5) as [40, 42],[4]

$$\frac{d}{d\ln k}\Omega_{\mathrm{SGW}}(k,\eta) = \frac{1}{12}\left(\frac{k}{a(\eta)H(\eta)} \right)^2 \mathcal{P}_h(k,\eta). \tag{11}$$

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

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

where z is the redshift.

Finally, let us review how experimental searches connect with gravitational wave abundance. Gravitational wave experiments typically quote results in terms of the characteristic strain h_c, which is related to an abundance of stochastic gravitational waves by [42],

$$\frac{d}{d\ln k}\Omega_{\mathrm{GW}}(k=2\pi f,\eta_0) = \frac{2\pi}{3H_0^2}f^2h_c^2(f). \tag{13}$$

The Ω_{SGW} abundance computed in Eq. (12) can be directly translated to characteristic strain constraints using this formula.

———

[3] See [38, 41] for variable changes to make this computationally simpler, where we have resolved some inconsistencies and ambiguities.
[4] Refs. [25, 38, 39, 41] give a range of differing values for this expression with which we do not agree.
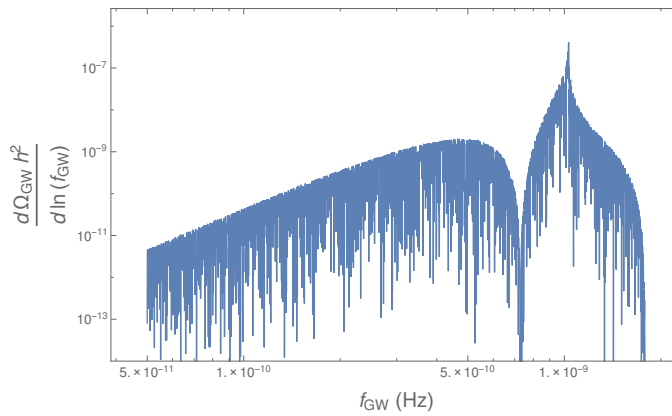
FIG. 1. Gravitational wave abundance for an idealized delta function scalar spectrum peaked at the scale k_f corresponding to PBH mass $30M_\odot$ according to Eqn. (1) and normalized so that $\Omega_{\rm PBH} = \Omega_{\rm DM}$

III.   IDEALIZED DELTA FUNCTION SPECTRUM

Perhaps the simplest model for PBH production is to assume a sharp narrow spike in k space in the scalar power spectrum. For a narrow enough spike, this can be approximated by a δ-function [25]. While a δ function is not physical, it is useful to consider as a mathematical construct.

The SGW spectrum for a δ-function scalar spectrum is plotted in Fig. 1. For reference, the δ-function is chosen to be peaked at a scale k_f corresponding to a horizon PBH mass of $30M_\odot$ (see Eqn. (1)), and its amplitude is chosen so that PBHs that form from this spectrum make up all of the observed DM abundance.

A constructive interference between g_k and $f(\mathbf{k}, \tilde{\mathbf{k}}, \eta)$ in (6) leads to a resonance at $k = 2\sqrt{w}/k_f$ where the amplitude continues to grow at late times, and there is a zero at $k = \sqrt{2w}/k_f$. The spectrum extends up to $k = 2k_f$ where the incoming scalar modes are aligned.

Of course, physical spectra will be extended. Let us briefly discuss the changes to the PBH and SGW spectra as we go to more extended primordial scalar spectra. Regarding the PBH spectrum, note that PBH formation is exponentially sensitive to σ_{R_M} and the integration in (3) samples a somewhat narrow window in k of $\mathcal{P}_\Phi(k)$. Thus, PBHs predominately form near where $\mathcal{P}_\Phi(k)$ peaks. Nevertheless, the integration in (3) may lead the PBH mass spectrum to peak at a smaller or larger scale than $\mathcal{P}_\Phi(k)$ depending on the detailed shape of $\mathcal{P}_\Phi(k)$. Critical collapse effects [34, 43–45] will lead to further corrections to the peak mass—see comments in the Discussion section. In regards to SGWs, the SGW abundance spectra will be smoothed out by the integral in (5) for extended $\mathcal{P}_\Phi(k)$, so features like the resonance and destructive interference in the δ-function spectrum will not be present for extended spectra. Additionally, the SGW abundance is only quadratically sensitive to $\mathcal{P}_\Phi(k)$ and depends on integration over a larger range of k in (5). This leads to an enhanced SGW abundance over a larger range in frequency for extended scalar spectra relative to a narrower spectrum.

IV.   MODELS OF PRIMORDIAL SCALAR SPECTRA

There are many models of inflationary dynamics that can induce primordial power spectra giving rise to PBH formation. Here, we review several classes of models capable of producing PBHs with large enough abundance at masses relevant to LIGO. We do not attempt a full accounting of all models that have been proposed for PBH production. Rather, we survey several models that predict different primordial spectra, allowing us to draw some
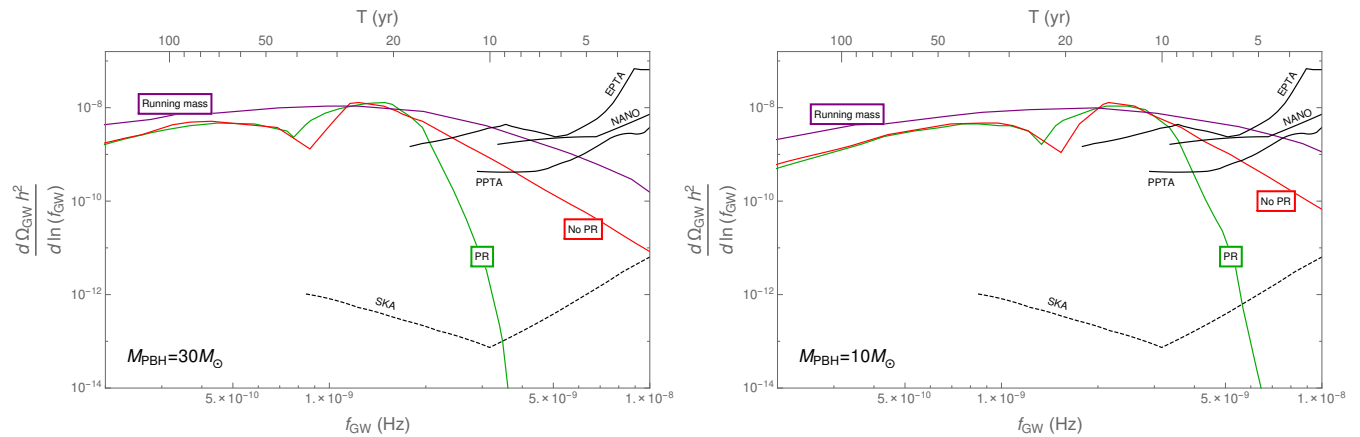
FIG. 2. Gravitational wave abundance (envelopes) as a function of frequency assuming $\Omega_{\mathrm{PBH}} = \Omega_{\mathrm{DM}}$. The PBH abundance spectrum is peaked at 30 (left) or 10 (right) $M_\odot$. We display SGW for a top-hat spectrum with width set by expectations from parametric resonance (green "PR" curve), a red-tilted scalar spectrum with spectral index $n_s = -1$ supplemented by a cutoff at a minimum frequency (red "No PR" curve), and the spectrum from the running mass model (purple curve). Black solid lines are current spectrum-independent bounds from EPTA (upper) [51], NANOGrav (middle) [53], and PPTA (lower) [52]. The black dashed line is a projection for bounds from SKA [54]. The top axis indicates the approximate observing time T to be sensitive to a given minimum frequency $f_{\mathrm{min}} \sim 1/T$.

general conclusions. Other models of PBH production not considered here can be found in Refs. [27, 46–50]. For each model, we calculate the resulting SGW spectrum. Results and present bounds from the European Pulsar Timing Array (EPTA) [51], Parkes Pulsar Timing Array (PPTA) [52], and NANOGrav [53] experiments and projections for SKA [54] are displayed in conjunction with these spectra in Fig. 2. The plots show envelopes of SGW spectra that correspond to a peak in the PBH mass spectrum at masses $M_{\mathrm{PBH}} = 30$ and $10 M_\odot$. We show results for $\Omega_{\mathrm{PBH}} = \Omega_{\mathrm{DM}}$, but the SGW amplitude is only log-dependent on the PBH relic abundance. Thus, results are only slightly changed for other choices. For example, for a spectra peaking at $M_{\mathrm{PBH}} = 30 M_\odot$, reducing the PBH abundance from $f = 1$ (all the DM) to $f = 10^{-4}$ reduces the SGW relic abundance by a factor of 0.4, which would have a minor effect on SGW detectability, whereas the LIGO rate may be accommodated in either case depending on the details of the binary formation and merger rate calculation as discussed in the Introduction.

A. Double inflation with parametric resonance

A close approximation to a highly-peaked δ-function scalar spectrum can be realized with a period of parametric resonance after the end of inflation. Oscillations of the inflaton can lead to specific modes being exponentially enhanced [55]. Since parametric resonance occurs after inflation has ended, the resonantly excited modes would not have a large enough length scale to produce PBHs at masses relevant to LIGO. A solution is to have two periods of inflation, with parametric resonance occurring between the two periods. The second inflation stretches the resonantly amplified modes to the relevant length scale. A model of this type is given in Refs. [56–58], which construct a period of hybrid inflation [59, 60] followed by new inflation [61, 62]. Such a model can be engineered to give a peak at any scale—by varying the length of the second inflation—and with any amplitude—depending on the relationship between the efficiency of the resonance and the decay width of the inflaton causing the resonance.

Because parametric resonance produces a sharply peaked scalar perturbation spectrum, it is the model that most closely mimics the δ-function spectrum of the previous subsection. However, even in this case, the δ-function approximation is not quite applicable; parametric resonance at a scale k results in a resonant band of width $\sim k$ [63].

Such a spectrum can be approximated more closely by a top-hat [25], and the resulting SGW spectrum is shown in Fig. 2 by a green curve.[5] This more realistic spectrum does not have as pronounced a peak as the δ-function case. Furthermore, for a fixed peak mass of PBH production, the scalar power spectrum peak must be shifted to larger k due to the integration in (3), and therefore the SGW frequencies in this more realistic spectrum are larger than those of the δ-function spectrum.

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

Summary: Double inflation with parametric resonance has enough flexibility to allow a narrow range of PBHs needed to produce the LIGO GW signal. SGW spectra for masses $M_{\mathrm{PBH}} \lesssim 30 M_{\odot}$ are constrained before accounting for critical collapse effects. An increase in PTA data collection time will enable detection of the SGWs for spectra peaked at masses greater than this, potentially detecting or excluding this mechanism as an explanation for LIGO GWs.

B. Double inflation without parametric resonance

Double inflation can produce a somewhat peaked spectrum even without parametric resonance. The earlier period of inflation gives rise to perturbations observed in the CMB and large scale structure. This early inflation is thus bounded to have small perturbations $\delta\rho/\rho \simeq 10^{-5}$. However, a later period of inflation can have much larger perturbations. This can be constructed with a model similar to that of the previous subsection; hybrid inflation followed by new inflation [64, 65]. The detectability of this model using PTAs was recently discussed in Ref. [28], with which our conclusions are in agreement. To produce a large peak in the power spectrum, the model is designed so that the scale k_* at which the second inflation starts roughly coincides with the scale relevant to PBH formation. The power spectrum of the second inflation starts out very large but has a red spectral tilt ($n_s < 1$). Thus, the spectrum is $\mathcal{P}_{\mathcal{R}} = \mathcal{P}_{\mathcal{R}}(k_*)(k/k_*)^{n_s-1}$ for $k > k_*$ and $\mathcal{P}_{\mathcal{R}} \ll \mathcal{P}_{\mathcal{R}}(k_*)$ for $k < k_*$. In this model, the slow roll parameters satisfy $|\eta| \gg |\epsilon|$, so using the relationship $n_s - 1 = 2\eta - 6\epsilon$ and the slow roll condition $|\eta| < 1$ implies $-1 \lesssim n_s \lesssim 3$ [61].

For this model, given a peak PBH mass M_{peak}, k_* is smaller than the formation scale $k_f(M_{\mathrm{peak}})$ given by (1). This results from the integration in (3), which is maximized when $R_M^{-1} > k_*$. Thus, the SGW spectrum will be peaked at smaller frequencies in this model compared to the case of a δ-function giving the same peak PBH mass. This effect is more pronounced for larger n_s corresponding to wider spectra, whereas for the minimum $n_s = -1$, $k_* \simeq k_f(M_{\mathrm{peak}})$. On the other hand, the SGW abundance away from the peak scale k_* is enhanced relative to narrower spectra.

Fig. 2 shows the resulting SGW spectra as red curves assuming $n_s = -1$, the minimum consistent with slow roll. This choice gives the narrowest possible peak in the perturbation spectrum. Larger values of n_s would be more strongly

---

[5] In the notation of [25], the resonant band has amplitude $\mathcal{P}_{\Phi} \simeq A^2/(2\Delta)$ on the domain $|\ln(k/k_p)| < \Delta$. For parametric resonance peaked at k_p with width k_p, this implies $\Delta \simeq \sinh^{-1}(1/2) \simeq 0.48$. PBHs are predominantly produced at the smallest masses within this window corresponding to the scale $k = k_p e^{\Delta}$.

constrained because they represent a more extended perturbation spectrum with larger resulting SGW abundance at high frequencies, which more than makes up for the necessary reduction in k_* to ensure the peak PBH mass M_{peak} remains constant as n_s increases. Already, the EPTA and PPTA experiments exclude this model of PBH formation up to masses peaked at greater than $30M_\odot$ due to the contributions to SGWs at high frequency.[6] However, it is possible that critical collapse effects may still narrowly allow this model.

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

C. Running mass model

The running mass model [66–68] supposes just one period of inflation, but with significant running of the spectral index. This approach can achieve a large perturbation amplitude at scales much smaller than those relevant to the CMB (here denoted $k_0 = 0.05 \text{ Mpc}^{-1}$). The amplitude at any point can be parametrized by,

$$\mathcal{P}_\mathcal{R}(k) = \mathcal{P}_\mathcal{R}(k_0) \left(\frac{k}{k_0} \right)^{n(k)-1}, \tag{14}$$

where,

$$n(k) = n_s(k_0) + \frac{1}{2!}\alpha_s(k_0)\ln\left(\frac{k}{k_0}\right) + \frac{1}{3!}\beta_s(k_0)\ln^2\left(\frac{k}{k_0}\right) + \frac{1}{4!}\gamma_s(k_0)\ln^3\left(\frac{k}{k_0}\right) + ... \tag{15}$$

Here, α_s, β_s, and γ_s are the runnings of n_s, giving sequentially higher derivatives of n_s with respect to $\log k$. When they are all allowed to vary, the parameters are bounded at $k = k_0$ as $\mathcal{P}_\mathcal{R} = (2.142\pm.049)\times 10^{-9}$ [69], $n_s = 0.9586\pm0.0056$, $\alpha_s = 0.009 \pm 0.010$, and $\beta_s = 0.025 \pm 0.013$ [70]. To obtain a peaked spectrum, we will generally want one or both of α_s and β_s to be positive, while γ_s (or a higher order term) must be negative. The coefficients are chosen so that the peaking occurs for scales relevant for PBH formation, well outside of the region probed by the CMB.

For an explicit model of this type, see the above references. We will simply set at $k = k_0$ the values $\mathcal{P}_\mathcal{R} = 2.142 \times 10^{-9}$, $n_s = 0.96$, and $\alpha_s = 0.009$ near their preferred experimental values. We then vary β_s and γ_s to achieve a PBH fractional abundance vs. mass spectrum that peaks at a given mass (here, as before, either 30 or $10M_\odot$) and gives the desired total relic abundance equal to the DM abundance.[7] Of note, this requires a tuning of several significant digits in the running parameters to obtain the correct peak position and amplitude in the primordial spectrum. Additionally, for the masses of interest, β_s is several standard deviations away from its measured value.

Another shortcoming is that generically $|\beta_s|$ should be suppressed by a factor of $(n_s - 1)$ in relation to $|\alpha_s|$ if the third derivative of the inflaton potential is not much larger than the lower derivatives. This is the case, *e.g.*, for the explicit model in [68]. However, to produce a large enough peak in the curvature spectrum for the PBH masses of interest, $|\beta_s| \gtrsim |\alpha_s|$ is required. Constructing an explicit model to circumvent this generic suppression poses a challenge.

As in the previous subsection, PBH production occurs predominantly on scales where $\mathcal{P}_\Phi$ is largest, while SGWs are produced at all frequencies. The result is plotted in purple in Fig. 2. While in principle stronger bounds could

---

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

[7] Specifically, $\beta_s = 0.0903091$ (0.0813319) and $\gamma_s = -0.0166717$ (-0.0145293) for peak $M_{\text{PBH}} = 30M_\odot$ $(10M_\odot)$.

be placed by combining PTA data over many frequencies, even the sensitivity of the bin-by-bin exclusions surpasses the SGW abundance for this model. Thus, present PTA experiments exclude this model for PBH production as an explanation for the LIGO events.

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

D. Axion-curvaton model

Unlike the previous models discussed, the axion-curvaton model [71, 72] (see also [73]) supposes that primordial fluctuations on small scales are sourced after inflation from a separate curvaton field [74, 75] that need not induce a second period of inflation. The model consists of a complex field $\Phi = (\varphi/\sqrt{2})e^{i\sigma/f_\sigma}$, where σ is the curvaton. Once φ reaches the minimum of its potential at $\varphi = f_\sigma$ and begins oscillating about it, σ becomes well-defined. After this point, corresponding to comoving scale k_*, the curvaton can induce a blue scalar spectrum ($n_s > 1$) so that $\mathcal{P}_\mathcal{R} = \mathcal{P}_\mathcal{R}(k_*)(k/k_*)^{n_s-1}$ for $k < k_*$ and $\mathcal{P} \ll \mathcal{P}_\mathcal{R}(k_*)$ for $k > k_*$. Here, a blue spectrum with $n_s \sim 2$ to 4 can be obtained. The power spectrum from the curvaton is constrained on large scales to be less than that observed to be coming from inflation—$\mathcal{P}_\mathcal{R}(k) \lesssim 2 \times 10^{-9}$ for $k \lesssim \mathrm{Mpc}^{-1}$.

Because in this model the universe is assumed to always be radiation dominated, the curvaton, which redshifts as matter, decays before it dominates the universe's energy density. So, the perturbations sourced by the curvaton will grow as $(\rho_\sigma/\rho_r)^2 \propto a^2$ until the curvaton decays. Thus, PBHs—whose formation depends exponentially on the perturbation amplitude—will preferentially form at the time of the curvaton decay. Since the scale at which the curvaton decays is unrelated to the scale k_* at which φ reaches its minimum, PBHs are expected to form long after the scale at which their primordial overdensities reenter the horizon.[8] But during the intervening time, the (radiation) energy within the comoving volume containing the primordial overdensity will redshift. The resulting PBH is thus potentially much smaller in mass than if it had formed immediately after horizon reentry. To produce a black hole of equal mass, k_* must in turn be far smaller than if the PBH collapsed immediately after horizon reentry. This drives the expected SGW signal to smaller frequencies than can be probed at pulsar timing arrays. Thus, barring some coincidence between the time it takes φ to reach its minimum and the decay time of the curvaton, PTA experiments will not be sensitive to this PBH production model. Indeed, these time scales are expected to be far apart, with the Hubble parameter at which φ oscillations start corresponding to $H \simeq 2\pi f_\sigma$ and the Hubble parameter when φ decays corresponding to $H \simeq \Gamma_\sigma \simeq m_\sigma^3/f_\sigma^2 \simeq \Lambda^6/f_\sigma^5 \ll f_\sigma$, where $\Lambda \ll f_\sigma$ is the explicit $U(1)$ breaking scale where σ receives a non-perturbative mass.

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

---

[8] For example, Ref. [72] considers benchmark points of $M_{\mathrm{min}}/M_{\mathrm{BH}} = 10^{-3}$ and 10^{-8}.

V. DISCUSSION

It is worth noting that we have neglected effects of critical collapse [34, 43–45], wherein detailed numerical work has shown that the mass of PBHs formed following horizon reentry may differ from the horizon mass depending on the size of the overdensity in δ. Most importantly for our discussion, critical collapse effects shift the peak in the mass spectrum to slightly lower masses. This shift in the peak of the mass spectrum can be at most of order a few for the models considered here [45]. To compensate, $\mathcal{P}_\Phi(k)$ must shift to smaller k (corresponding to larger mass, see (1)) once critical collapse effects are included in order to keep the PBH abundance spectrum peaked at the same mass. This change would require a slightly longer collection time, with the collection time depending on the mass as $T \propto \sqrt{M}$. Of the models we have discussed, the one whose detection prospects are most sensitive to this effect is double inflation with parametric resonance. In addition, critical collapse may also affect the PBH abundance by an order one factor, which will not significantly impact the SGW abundance.

We have also neglected effects of non-sphericity [34–37], which tend to raise the threshold δ_c on the matter perturbation spectrum for PBHs to form. For an increase from δ_c to δ_c', the amplitude of the scalar spectrum $\mathcal{P}_\Phi$ must increase by a factor of $\sim (\delta_c'/\delta_c)^2$, and the corresponding GW abundance increases by a factor $\sim (\delta_c'/\delta_c)^4$. This may increase the SGW abundances in Fig. 2 by as much as a factor of ~ 9. This is not enough to change the qualitative picture of which models are probed by PTA experiments, though it may partially compensate for the effects of critical collapse on the needed observation time.

Another effect that can change the necessary threshold on δ_c is a soft equation of state during the period relevant to PBH formation [76–78]. If new physics exists such that the equation of state is more matter-like ($w < 1/3$), a smaller δ_c is necessary to induce collapse. While for a fixed abundance of PBHs, this can reduce the amplitude of SGWs, it would not change the qualitative picture.

It is worth commenting that the merging of supermassive black holes (SMBH) will also create a stochastic background of gravitational waves. These obey a power-law spectrum with $\Omega_{\rm GW} \propto f^{2/3}$ [79–81]. This differs markedly from the spectra considered in this paper, but disentangling the SGW considered here from the GW from the SMBH will present an additional challenge. This challenge may be acute depending on the amplitude of the SMBH GW.

We now revisit our assumption that PBHs constitute all of the dark matter. Recent work [12] has called into question whether an extended spectrum of PBHs making up all of the DM is allowed by MACHO and faint dwarf cooling constraints. This, perhaps along with bounds from WMAP [14] and Planck [15], may indicate that PBHs are allowed to be at most only a fraction of the DM. Nevertheless, even a small abundance of PBHs may still be consistent with LIGO observations [6, 7]. But we reiterate that a reduction in the the PBH abundance by even several orders of magnitude will have only a small impact on the SGW spectra considered here because the PBH abundance is exponentially sensitive to the primordial power spectrum amplitude, whereas the SGW abundance is only power-law sensitive to it. As discussed in the introduction, the observed LIGO merger rate may be accommodated by PBH DM fractions much smaller than unity. Since our primary study target is explaining LIGO GWs, not the full abundance of DM, our results hold.

VI. CONCLUSION

If LIGO GWs are to be explained by merging PBHs, they will need to be produced in the early universe via a peaked primordial curvature spectrum. A by-product of this will be the production of SGWs. We have shown that the detectability of SGWs at current and future experiments will depend sensitively on the physics that gives rise to PBHs.

[34] I. Musco, J. C. Miller, and L. Rezzolla, Class. Quant. Grav. **22**, 1405 (2005), arXiv:gr-qc/0412063 [gr-qc]; I. Musco, J. C. Miller, and A. G. Polnarev, Class. Quant. Grav. **26**, 235001 (2009), arXiv:0811.1452 [gr-qc]; I. Musco and J. C. Miller, Class. Quant. Grav. **30**, 145009 (2013), arXiv:1201.2379 [gr-qc].

[35] R. K. Sheth, H. J. Mo, and G. Tormen, Mon. Not. Roy. Astron. Soc. **323**, 1 (2001), arXiv:astro-ph/9907024 [astro-ph].

[36] F. Kuhnel and M. Sandstad, Phys. Rev. **D94**, 063514 (2016), arXiv:1602.04815 [astro-ph.CO].

[37] J. R. Bond and S. T. Myers, Astrophys. J. Suppl. **103**, 1 (1996).

[38] K. N. Ananda, C. Clarkson, and D. Wands, Phys. Rev. **D75**, 123518 (2007), arXiv:gr-qc/0612013 [gr-qc].

[39] D. Baumann, P. J. Steinhardt, K. Takahashi, and K. Ichiki, Phys. Rev. **D76**, 084019 (2007), arXiv:hep-th/0703290 [hep-th].

[40] E. Bugaev and P. Klimai, Phys. Rev. **D81**, 023517 (2010), arXiv:0908.0664 [astro-ph.CO].

[41] L. Alabidi, K. Kohri, M. Sasaki, and Y. Sendouda, JCAP **1209**, 017 (2012), arXiv:1203.4663 [astro-ph.CO].

[42] M. Maggiore, Phys. Rept. **331**, 283 (2000), arXiv:gr-qc/9909001 [gr-qc]; *Gravitational Waves. Vol. 1: Theory and Experiments*, Oxford Master Series in Physics (Oxford University Press, 2007).

[43] C. Gundlach, Living Rev. Rel. **2**, 4 (1999), arXiv:gr-qc/0001046 [gr-qc]; Phys. Rept. **376**, 339 (2003), arXiv:gr-qc/0210101 [gr-qc].

[44] J. C. Niemeyer and K. Jedamzik, Phys. Rev. Lett. **80**, 5481 (1998), arXiv:astro-ph/9709072 [astro-ph]; Phys. Rev. **D59**, 124013 (1999), arXiv:astro-ph/9901292 [astro-ph].

[45] F. Kühnel, C. Rampf, and M. Sandstad, Eur. Phys. J. **C76**, 93 (2016), arXiv:1512.00488 [astro-ph.CO].

[46] A. Dolgov and J. Silk, Phys. Rev. **D47**, 4244 (1993).

[47] J. Garcia-Bellido, A. D. Linde, and D. Wands, Phys. Rev. **D54**, 6040 (1996), arXiv:astro-ph/9605094 [astro-ph].

[48] S. Clesse and J. García-Bellido, Phys. Rev. **D92**, 023524 (2015), arXiv:1501.07565 [astro-ph.CO].

[49] S.-L. Cheng, W. Lee, and K.-W. Ng, (2016), arXiv:1606.00206 [astro-ph.CO].

[50] S. Blinnikov, A. Dolgov, N. K. Porayko, and K. Postnov, JCAP **1611**, 036 (2016), arXiv:1611.00541 [astro-ph.HE].

[51] L. Lentati *et al.*, Mon. Not. Roy. Astron. Soc. **453**, 2576 (2015), arXiv:1504.03692 [astro-ph.CO].

[52] R. M. Shannon *et al.*, Science **349**, 1522 (2015), arXiv:1509.07320 [astro-ph.CO].

[53] Z. Arzoumanian *et al.* (NANOGrav), Astrophys. J. **821**, 13 (2016), arXiv:1508.03024 [astro-ph.GA].

[54] G. Janssen *et al.*, *Proceedings, Advancing Astrophysics with the Square Kilometre Array (AASKA14): Giardini Naxos, Italy, June 9-13, 2014*, PoS **AASKA14**, 037 (2015), arXiv:1501.00127 [astro-ph.IM].

[55] L. Kofman, A. D. Linde, and A. A. Starobinsky, Phys. Rev. Lett. **73**, 3195 (1994), arXiv:hep-th/9405187 [hep-th].

[56] M. Kawasaki, T. Takayama, M. Yamaguchi, and J. Yokoyama, Phys. Rev. **D74**, 043525 (2006), arXiv:hep-ph/0605271 [hep-ph].

[57] T. Kawaguchi, M. Kawasaki, T. Takayama, M. Yamaguchi, and J. Yokoyama, Mon. Not. Roy. Astron. Soc. **388**, 1426 (2008), arXiv:0711.3886 [astro-ph].

[58] P. H. Frampton, M. Kawasaki, F. Takahashi, and T. T. Yanagida, JCAP **1004**, 023 (2010), arXiv:1001.2308 [hep-ph].

[59] E. J. Copeland, A. R. Liddle, D. H. Lyth, E. D. Stewart, and D. Wands, Phys. Rev. **D49**, 6410 (1994), arXiv:astro-ph/9401011 [astro-ph].

[60] A. D. Linde and A. Riotto, Phys. Rev. **D56**, R1841 (1997), arXiv:hep-ph/9703209 [hep-ph].

[61] K. I. Izawa and T. Yanagida, Phys. Lett. **B393**, 331 (1997), arXiv:hep-ph/9608359 [hep-ph].

[62] K. I. Izawa, M. Kawasaki, and T. Yanagida, Phys. Lett. **B411**, 249 (1997), arXiv:hep-ph/9707201 [hep-ph].

[63] L. Kofman, A. D. Linde, and A. A. Starobinsky, Phys. Rev. **D56**, 3258 (1997), arXiv:hep-ph/9704452 [hep-ph].

[64] M. Kawasaki, N. Sugiyama, and T. Yanagida, Phys. Rev. **D57**, 6050 (1998), arXiv:hep-ph/9710259 [hep-ph].

[65] M. Kawasaki and T. Yanagida, Phys. Rev. **D59**, 043512 (1999), arXiv:hep-ph/9807544 [hep-ph].

[66] E. D. Stewart, Phys. Lett. **B391**, 34 (1997), arXiv:hep-ph/9606241 [hep-ph]; Phys. Rev. **D56**, 2019 (1997), arXiv:hep-ph/9703232 [hep-ph].

[67] S. M. Leach, I. J. Grivell, and A. R. Liddle, Phys. Rev. **D62**, 043516 (2000), arXiv:astro-ph/0004296 [astro-ph].

[68] M. Drees and E. Erfani, JCAP **1104**, 005 (2011), arXiv:1102.2340 [hep-ph]; JCAP **1201**, 035 (2012), arXiv:1110.6052 [astro-ph.CO].

[69] P. A. R. Ade *et al.* (Planck), (2015), arXiv:1502.01589 [astro-ph.CO].

[70] P. A. R. Ade *et al.* (Planck), (2015), arXiv:1502.02114 [astro-ph.CO].

[71] S. Kasuya and M. Kawasaki, Phys. Rev. **D80**, 023516 (2009), arXiv:0904.3800 [astro-ph.CO].

[72] M. Kawasaki, N. Kitajima, and T. T. Yanagida, Phys. Rev. **D87**, 063519 (2013), arXiv:1207.2550 [hep-ph].

[73] K. Kohri, C.-M. Lin, and T. Matsuda, Phys. Rev. **D87**, 103527 (2013), arXiv:1211.2371 [hep-ph].

[74] D. H. Lyth and D. Wands, Phys. Lett. **B524**, 5 (2002), arXiv:hep-ph/0110002 [hep-ph].

[75] D. H. Lyth, C. Ungarelli, and D. Wands, Phys. Rev. **D67**, 023503 (2003), arXiv:astro-ph/0208055 [astro-ph].

[76] M. Yu. Khlopov and A. G. Polnarev, Phys. Lett. **B97**, 383 (1980).

[77] K. Jedamzik, Phys. Rev. **D55**, 5871 (1997), arXiv:astro-ph/9605152 [astro-ph]; *Sources and detection of dark matter in the universe. Proceedings, 3rd International Symposium, and Workshop on Primordial Black Holes and Hawking Radiation, Marina del Rey, USA, February 17-20, 1998*, Phys. Rept. **307**, 155 (1998), arXiv:astro-ph/9805147 [astro-ph].

[78] K. Jedamzik and J. C. Niemeyer, Phys. Rev. **D59**, 124014 (1999), arXiv:astro-ph/9901293 [astro-ph].

[79] E. S. Phinney, (2001), arXiv:astro-ph/0108028 [astro-ph].

[80] A. H. Jaffe and D. C. Backer, Astrophys. J. **583**, 616 (2003), arXiv:astro-ph/0210148 [astro-ph].

[81] J. S. B. Wyithe and A. Loeb, Astrophys. J. **590**, 691 (2003), arXiv:astro-ph/0211556 [astro-ph].

[82] S. Clesse and J. García-Bellido, Phys. Dark Univ. **10**, 002 (2016), arXiv:1603.05234 [astro-ph.CO]; (2016), arXiv:1610.08479 [astro-ph.CO].

[83] I. Cholis, E. D. Kovetz, Y. Ali-Haïmoud, S. Bird, M. Kamionkowski, J. B. Muñoz, and A. Raccanelli, Phys. Rev. **D94**, 084013 (2016), arXiv:1606.07437 [astro-ph.HE].