arXiv:2306.17836v1 [astro-ph.CO] 30 Jun 2023

Do pulsar timing arrays observe merging primordial black holes?

Paul Frederik Depta,[a] **Kai Schmidt-Hoberg,**[b] **and Carlo Tasillo**[b]

[a]Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, 69117 Heidelberg, Germany
[b]Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany

E-mail: frederik.depta@mpi-hd.mpg.de, kai.schmidt-hoberg@desy.de, carlo.tasillo@desy.de

Abstract. In this letter we evaluate whether the gravitational wave background recently observed by a number of different pulsar timing arrays could be due to merging *primordial* supermassive black hole binaries. We find that for homogeneously distributed primordial black holes this possibility is inconsistent with the strong cosmological and astrophysical constraints on their total abundance. If the distribution exhibits some clustering, however, the merger rate will in general be enhanced, opening the window for a consistent interpretation of the PTA data in terms of merging primordial black holes.

Preprint number: DESY-23-093

Contents

1 Introduction **1**

2 Clustered primordial black holes **2**
2.1 GWs from mergers of primordial black holes 2
2.2 Expected number of binaries 5

3 Production and constraints **6**

4 Fitting to pulsar timing data **7**

5 Results **8**

6 Discussion and conclusions **9**

1 Introduction

The discovery of gravitational waves (GWs) with frequencies $\mathcal{O}(100)\,\mathrm{Hz}$ from binary black hole mergers in 2015 [1] has opened new possibilities for the exploration of our Universe. Depending on the astrophysical or cosmological source, GW signals may exist over a wide range of different frequencies and complementary experimental approaches aim to explore much of the available parameter space [2, 3]. Pulsar timing arrays (PTAs) in particular are sensitive to frequencies in the nHz range. In 2020 NANOGrav [3] announced evidence for a common red process hinting at a stochastic nHz gravitational wave background (GWB), with similar results subsequently published by other PTAs [4–6]. Excitingly, the very recent data releases from NANOGrav [7], EPTA [8], PPTA [9] and CPTA [10] not only strengthen this signal, but also show mounting evidence for the characteristic quadrupolar Hellings-Downs correlation [11]. It is therefore fascinating to speculate about possible origins of this newly discovered GWB. Mergers of supermassive black hole binaries (SMBHB) are expected to contribute to the GWB in this frequency range [3], although the local SMBHB density would need to be $\mathcal{O}(10)$ larger than previously estimated [12–14] in order to match the observed signal amplitude. Given that it is a matter of active debate as to whether this could provide a viable explanation [15–18], it is interesting to consider possible alternative sources. In particular a nHz GWB could also originate from truly cosmological sources [19], such as inflation [20], cosmic strings [21–23], first-order phase transitions [24–27] or the *production* of sub-solar mass primordial black holes (PBHs) [28–31].

In this letter we study the possibility that the observed GW signal could be due to *merging* supermassive PBH binaries. To obtain the correct signal amplitude, a sizeable abundance of PBHs is required, which is subject to strong constraints from various observational probes. It is therefore a non-trivial question as to whether merging PBHs could constitute a viable explanation. Indeed we find that it is *not* possible to consistently explain the NANOGrav data with homogeneously distributed PBHs. Specifically the parameter regions at very large PBH masses which naively allow to fit the PTA data (as done in [32]) no longer result in a stochastic (or even continuous) signal. For clustered PBHs, on the other hand, there is a more efficient formation of binary systems and the merger rate is generally enhanced. In this case we find that it is possible to obtain an overall consistent description of the observed GWB.

This letter is structured as follows: In the next section we review and extend the formalism required to evaluate the GWs from binary mergers for the case of clustered PBHs. In section 3 we discuss the relevant constraints on this scenario before coming to a discussion of how we match the predicted signal to the NANGrav data in section 4. We finally present our results in section 5.

2 Clustered primordial black holes

2.1 GWs from mergers of primordial black holes

A stochastic GWB can be specified in terms of the energy density parameter of GWs per logarithmic frequency interval $h^2\Omega_{\rm GW}(f)$ around a frequency f. In our scenario, this GWB is generated from PBH binary mergers occurring at a rate $R(t_r)$ per comoving volume $V_{\rm c}$ and cosmic time t_r.[1] The frequencies accessible to PTAs can be well below the maximal frequency emitted in a binary merger and might therefore have been radiated long before coalescence. It is hence important to account for this effect. In particular, a given binary merging at time $t_{r,\rm merg}$ has emitted a frequency f_r in the cosmic rest frame, related to a frequency $f = f_r/(1+z)$ today, at time $t_r = t_{r,\rm merg} - \tau_{f_r}$, where

$$\tau_{f_r} = \frac{5\times 2^{1/3}}{256\,\pi^{8/3}}\, f_r^{-8/3}\,(G\,m_{\rm PBH}) \tag{2.1}$$

is the time until coalescence for f_r [33]. As $t_{r,\rm merg}$ is the argument for the *merger* rate, the number of events per comoving volume and cosmic time emitting a frequency f_r in the cosmic rest frame at cosmic time t_r is given by $R(t_r+\tau_{f_r})$. Note that to the best of our knowledge this frequency-dependent rate has not been discussed in the literature concerning GWs generated by PBHs before. The GW energy density parameter is given by (cf. [34])

$$\Omega_{\rm GW}(f) = \frac{f}{\rho_{\rm crit}}\frac{{\rm d}\rho_{\rm GW}}{{\rm d}f} = \frac{f}{\rho_{\rm crit}}\int_0^{t_0}{\rm d}t_r\left(R(t_r+\tau_{f_r})\frac{{\rm d}E_{\rm GW}}{{\rm d}f_r}\right)_{f_r=(1+z)f}\,, \tag{2.2}$$

where t_0 is the current cosmic time, $f_r = (1+z)\,f$ is the frequency that needs to be emitted at the redshift $z = z(t_r)$ corresponding to t_r to detect a frequency f today, and the GW spectrum can be estimated by power-laws for inspiral, merger, and ringdown [35][2]

$$\frac{{\rm d}E_{\rm GW}}{{\rm d}f_r} \simeq \frac{(\pi\,G)^{2/3}\,2^{-1/3}\,m_{\rm PBH}^{5/3}}{3}\begin{cases} f_r^{-1/3} & \text{for } f_r < f_1\,,\\ \frac{f_r^{2/3}}{f_1} & \text{for } f_1 \le f_r < f_2\,,\\ \frac{f_r^2 f_4^4}{f_1 f_2^{4/3}[4(f_r-f_2)^2+f_4^2]^2} & \text{for } f_2 \le f_r < f_3\,,\\ 0 & \text{for } f_3 \le f_r\,. \end{cases} \tag{2.3}$$

Here we assumed BHs with equal mass $m_{\rm PBH}$, $f_i = (a_i\,\eta^2 + b_i\,\eta + c_i)/(2\pi\,G\,m_{\rm PBH})$ with $\eta = 1/4$, and the coefficients from Table I of Ref. [35]

$$\begin{aligned}(a_1,a_2,a_3,a_4) &= (2.9740\times10^{-1}, 5.9411\times10^{-1}, 8.4845\times10^{-1}, 5.0801\times10^{-1})\,,\\ (b_1,b_2,b_3,b_4) &= (4.4810\times10^{-2}, 8.9794\times10^{-2}, 1.2848\times10^{-1}, 7.7515\times10^{-2})\,,\\ (c_1,c_2,c_3,c_4) &= (9.5560\times10^{-2}, 1.9111\times10^{-1}, 2.7299\times10^{-1}, 2.2369\times10^{-2})\,.\end{aligned} \tag{2.4}$$

[1] Cosmic time refers to the time in the corresponding cosmic rest-frame, related to time t measured in the present cosmic rest-frame by ${\rm d}t = (1+z)\,{\rm d}t_r$.

[2] In our calculations we include a cut-off at $f_r = 1/t_r$, since smaller frequencies cannot have completed even one complete orbit.

Note that the GW energy density parameter is in fact a cosmological average of many PBH binaries. We discuss the effect of only having access to a local realisation of the GWB within PTA measurements in Sec. 2.2.

For the merger rate we adapt the calculation from Ref. [36]. We assume a monochromatic initial PBH mass distribution, expecting that going to an extended distribution does not qualitatively change our results. A PBH binary forms in the early Universe when the gravitational attraction between two neighbouring PBHs overcomes the Hubble flow. A third close-by PBH provides angular momentum such that the two PBHs do not simply collide [37–39].

Clearly the merger rate depends on the local PBH density at binary formation. It is therefore interesting to study the effect of clustering, which increases the global comoving number density $n_{\rm PBH}$ by the local density contrast $\delta_{\rm dc}$. The density contrast can be considered constant on the scales relevant for binary formation [36]. The comoving number density $n_{\rm PBH}$ can be expressed in terms of the PBH mass $m_{\rm PBH}$, the fraction of PBH DM $f_{\rm PBH}$ and the DM density $\rho_{\rm DM,0}$ today

$$n_{\rm PBH} = f_{\rm PBH}\,\frac{\rho_{\rm DM,0}}{m_{\rm PBH}}\,. \tag{2.5}$$

The number density $\mathrm{d}n_3$ of the three-body configurations relevant for binary formation, where the comoving distance from a given PBH in the binary to the other PBH in the binary is within x and $x+\mathrm{d}x$ and the comoving distance to the third PBH is within y and $y+\mathrm{d}y$, is [36]

$$\mathrm{d}n_3(x,y) = \frac{n_{\rm PBH}}{2}\,\mathrm{e}^{-N_{\rm PBH}(y)}\,(4\pi\,n_{\rm PBH}\,\delta_{\rm dc})^2\,x^2\,y^2\,\mathrm{d}x\,\mathrm{d}y\,. \tag{2.6}$$

Here, the factor of $1/2$ removes over-counting as either one of the PBHs in the binaries can be chosen and $\mathrm{e}^{-N(y)}$ makes sure that there is no fourth binary within the comoving distance y (assuming Poisson statistics), where

$$N_{\rm PBH}(y) = \frac{4\pi}{3}\,n_{\rm PBH}\,\delta_{\rm dc}\,y^3 \tag{2.7}$$

is the expected number of binaries within a sphere of comoving distance y. The coalescence time τ of this binary is given by [36, 40]

$$\tau(x,y) = \tilde{\tau}\left(\frac{x}{\tilde{x}}\right)^{37}\left(\frac{y}{\tilde{x}}\right)^{-21}\,, \tag{2.8}$$

where

$$\tilde{\tau} = \frac{3}{170}\,\frac{(a_{\rm eq}\tilde{x})^4}{(G\,m_{\rm PBH})^3}\,, \tag{2.9}$$

$$\tilde{x} = \frac{3}{4\pi}\,\frac{2\,m_{\rm PBH}}{a_{\rm eq}^3\,\rho_{\rm eq}}\,, \tag{2.10}$$

$a_{\rm eq}$ and $\rho_{\rm eq}$ are the scale factor and total energy density at matter-radiation equality. The merger rate at time t_r is then given by

$$\begin{aligned} R(t_r) &= \int_0^{\tilde{x}}\mathrm{d}x\int_x^{\infty}\mathrm{d}y\,\frac{\partial^2 n_3}{\partial x\,\partial y}\,\delta(t_r-\tau(x,y)) \\ &= \frac{9\,\tilde{N}_{\rm PBH}^{53/37}}{296\pi\,\delta_{\rm dc}\,\tilde{x}^3\,\tilde{\tau}}\left(\frac{t_r}{\tilde{\tau}}\right)^{-34/37}\left(\Gamma\left[\frac{58}{37},\tilde{N}_{\rm PBH}\left(\frac{t_r}{\tilde{\tau}}\right)^{3/16}\right]-\Gamma\left[\frac{58}{37},\tilde{N}_{\rm PBH}\left(\frac{t_r}{\tilde{\tau}}\right)^{-1/7}\right]\right) \end{aligned} \tag{2.11}$$

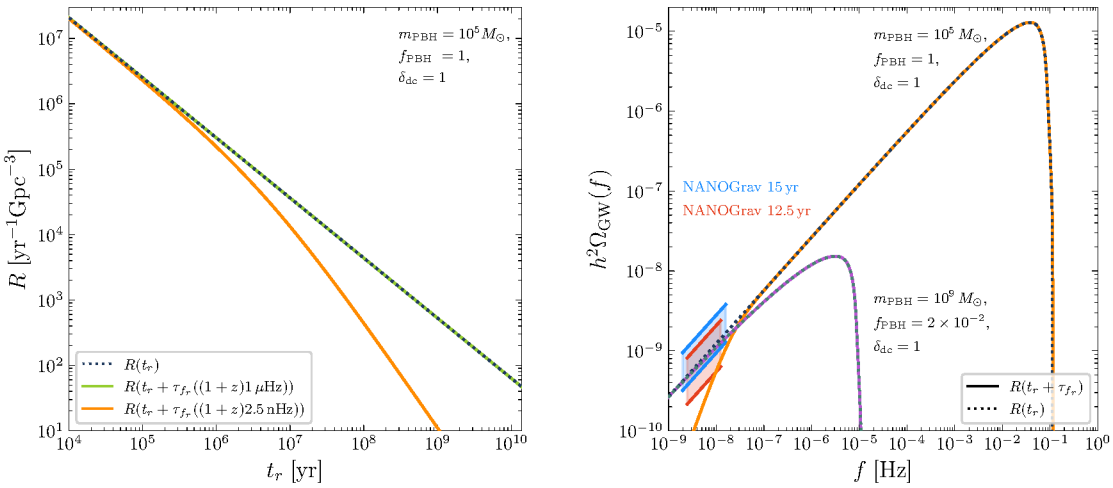


Figure 1. *Left:* Rate for binaries merging at t_r (dotted line) as well as rate for emitting GWs at t_r with frequencies today $f = 1\,\mu$Hz and 2.5 nHz (solid green and orange lines) assuming $m_{\rm PBH} = 10^5\,M_\odot$, $f_{\rm PBH} = 1$, and $\delta_{\rm dc} = 1$. *Right:* GW energy density $h^2\Omega_{\rm GW}(f)$ obtained using the correct rate $R(t_r + \tau_{f_r})$ (solid lines) and instead using the merger rate $R(t_r)$ in Eq. (2.2) (dotted lines) for different PBH masses and abundances. We also show the region, where the NANOGrav 12.5 and 15 yr signals are located assuming a power-law corresponding to the inspiral of binary BHs.

with the incomplete gamma function Γ and $\tilde{N}_{\rm PBH} = N_{\rm PBH}(\tilde{x})$. By substituting $t_r \to t_r + \tau_{f_r}$ we can therefore compute the rate for the emission of this frequency $R(t_r + \tau_{f_r})$. Multiple merger steps, possibly present due to successive hierarchical merging of PBH binaries with increasing mass due to large clustering, can be easily implemented by adding the rates and contributions to the GW energy density parameter for the corresponding steps as detailed in Ref. [41]. We include multiple merger steps when discussing the case of significant clustering, noting that this slightly shifts our results for $\delta_{\rm dc} = 10^3$ to lower PBH abundances compared to only considering a single step.

To illustrate the importance of the time when a given frequency is emitted we show in the left panel of Fig. 1 the rate for binaries merging at t_r (dotted blue line) with the one for GW emission with frequencies $f = 1\,\mu$Hz and 2.5 nHz today (solid green and orange lines), assuming $m_{\rm PBH} = 10^5\,M_\odot$, $f_{\rm PBH} = 1$, and $\delta_{\rm dc} = 1$ in each case. For instance, at $t_r = 10^8$ yr (i.e. $z \approx 30$) one obtains $\tau_{f_r}((1+z)1\,\mu{\rm Hz}) \approx 130$ yr and $\tau_{f_r}((1+z)2.5\,{\rm nHz}) \approx 1.1 \times 10^9$ yr. Hence, the rate for the emission of GWs with the larger frequency (solid green line) is very close to the merger rate (dotted blue line), whereas the rate for the emission of GWs with the smaller frequency (solid orange line) differs significantly, i.e. it takes the value that the dotted blue line attains 1.1×10^9 yr later.

In the right panel of Fig. 1 we show $h^2\Omega_{\rm GW}(f)$ as a function of the GW frequency today according to Eq. (2.2) (solid lines) as well as just inserting the merger rate $R(t_r)$ instead of $R(t_r + \tau_{f_r})$ in the integral (dotted lines). The difference between those calculations is especially important for the low frequencies observed by NANOGrav if the PBH mass is relatively light.

We close the discussion of the GWB signal by mentioning some assumptions that entered in its calculation. These include

- a monochromatic mass distribution for the PBHs [36],

- a circular orbit of the PBH binaries for the GWB spectrum and the time until coalescence for a given frequency,
- neglecting the effect of other PBHs on the binaries [42, 43] as well as other environmental effects e.g. due to accretion, and
- late-time formation of binaries [36].

Apart from the effect of other PBHs on binaries, which can potentially significantly reduce their merger rate [43], but require N-body simulations for an accurate treatment that are beyond the scope of this work, these assumptions are not expected to have a qualitative impact on our results.

2.2 Expected number of binaries

The NANOGrav signal is reported as a stochastic GWB. While there have been searches for the signals of individual binaries in the data of different PTAs [44–46], no compelling evidence for these signals has been found. For sufficiently large PBH masses and small abundances the expected signal will in general no longer resemble a stochastic background, as only very few binaries will contribute to the signal in this case, which would be inconsistent with the observations.

The problem is exacerbated by the question of how well an actual distribution of merging binaries observed by a PTA (corresponding to a local[3] value of GW energy density $\Omega_{\rm GW,loc}$) would reproduce the global average of the GW energy density $h^2\Omega_{\rm GW}(f) = h^2\langle\Omega_{\rm GW,loc}(f)\rangle$, i.e. how well the global mean is reproduced by the binaries that are in our past light cone. As shown in [47], even relatively generic distributions of binary mergers, scaling with the luminosity distance squared, can lead to wide distributions in $\Omega_{\rm GW,loc}$, with width-to-mean ratios scaling like $\Delta\Omega_{\rm GW,loc}/\Omega_{\rm GW} \propto \bar{N}^{-1/3}$ with $\bar{N}$ the expected number of binaries contributing to a certain frequency range, slower than predicted from the central limit theorem $\propto \bar{N}^{-1/2}$.

To calculate the average number of binaries contributing to a certain frequency range, recall that $R(t_r+\tau_{f_r})$ is the number of binaries per comoving volume $V_{\rm c}$ and per time interval in the cosmic rest frame t_r emitting with a frequency f_r. Hence, the number of binaries $\mathrm{d}\bar{N}$ emitting at redshifts between z and $z+\mathrm{d}z$ with a frequency within the logarithmic interval $\mathrm{d}\ln f_r$ around f_r is [48]

$$R(t_r+\tau_{f_r}) = \frac{\mathrm{d}z}{\mathrm{d}t_r}\frac{\mathrm{d}^2\bar{N}}{\mathrm{d}z\,\mathrm{d}\ln f_r}\frac{\mathrm{d}\ln f_r}{\mathrm{d}t_r}\frac{\mathrm{d}t_r}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}V_{\rm c}} = \frac{\mathrm{d}^2\bar{N}}{\mathrm{d}z\,\mathrm{d}\ln f_r}\left(-\frac{\mathrm{d}\ln f_r}{\mathrm{d}\tau_{f_r}}\right)\frac{\mathrm{d}z}{\mathrm{d}V_{\rm c}}\,, \tag{2.12}$$

where we used that the time in the cosmic rest frame a source is emitting within the frequency interval is given by

$$\mathrm{d}t_r\frac{\mathrm{d}\ln f_r}{\mathrm{d}t_r} = -\mathrm{d}t_r\frac{\mathrm{d}\ln f_r}{\mathrm{d}\tau_{f_r}} \tag{2.13}$$

as τ_{f_r} is the time until coalescence for f_r. Since the binaries emitting in the logarithmic frequency interval $\mathrm{d}\ln f_r$ around f_r are detected today in a logarithmic frequency interval $\mathrm{d}\ln f$ around f, where $\mathrm{d}\ln f_r = \mathrm{d}f_r/f_r = \mathrm{d}f/f = \mathrm{d}\ln f$, we have

$$\frac{\mathrm{d}^2\bar{N}}{\mathrm{d}z\,\mathrm{d}\ln f} = \frac{\mathrm{d}^2\bar{N}}{\mathrm{d}z\,\mathrm{d}\ln f_r}\,. \tag{2.14}$$

[3]Local on the scales relevant for a GW signal in PTA observations, i.e. at least within a sphere around earth with a few light-years radius such that correlations between different pulsars can be affected.

With Eq. (2.1) and the definition of the comoving volume in a spatially flat universe

$$V_{\rm c}(z) = \frac{4\pi}{3}[d_{\rm c}(z)]^3 = \frac{4\pi}{3}\left(\int_0^z \frac{{\rm d}z'}{H(z')}\right)^3 , \tag{2.15}$$

where $d_{\rm c}$ is the comoving distance and H is the Hubble rate, we find

$$\frac{{\rm d}^2\bar{N}}{{\rm d}z\,{\rm d}\ln f} = \frac{8}{3}\,\tau_{f_r}\,\frac{4\pi\,[d_{\rm c}(z)]^2}{H(z)}\,R(t_r(z)-\tau_{f_r}) . \tag{2.16}$$

The average number of binaries $\bar{N}(f_-, f_+)$ contributing to GWs of frequencies between f_- and f_+ is therefore given by[4]

$$\bar{N}(f_-, f_+) = \int_{f_-}^{f_+}\frac{{\rm d}f}{f}\int_0^\infty {\rm d}z\,\frac{8}{3}\,\tau_{f_r}\frac{4\pi\,[D_c(z)]^2}{H(z)}\,R(t_r(z)-\tau_{f_r}) . \tag{2.17}$$

As we are interested in the number of binaries contributing to the NANOGrav signal, we take the frequency interval to span over the corresponding lowest five Fourier modes, i.e. $f_- = 1/T_{\rm obs}^{12.5} = 2.46\,{\rm nHz}$ and $f_+ = 5/T_{\rm obs}^{12.5} = 12.3\,{\rm nHz}$. This choice is justified as it is only the first five posteriors of the Bayesian periodogram that go into our analysis, as described in Sec. 4.

The signal prediction for PTAs is crucially affected by the expected number of binaries emitting in the relevant frequency band [47]. Generally, the local GWB $h^2\Omega_{\rm GW,loc}(f)$ that may be observed by a PTA is obtained by adding individual PBH binaries (the total number drawn from a Poisson distribution with mean $\bar{N}(f_-, f_+)$) distributed according to ${\rm d}^2\bar{N}/({\rm d}z\,{\rm d}\ln f)$. In particular, $h^2\Omega_{\rm GW,loc}(f)$ is *not* deterministic given model parameters (mass, abundance, and clustering of PBHs), but instead stochastic in itself. The statistics of $h^2\Omega_{\rm GW,loc}(f)$ can be evaluated using Markov chain Monte Carlo methods or moment generating functions [47]. Due to the considerable additional numerical effort we leave this for future studies, use the global average signal prediction $h^2\Omega_{\rm GW}(f)$, and identify regions in parameter space where we expect a significant deviation from the global average. If $\bar{N}(f_-, f_+) \gg 1$, a lot of different PBH binaries contribute to the GW signal and the local GWB $h^2\Omega_{\rm GW,loc}(f)$ is close to the global average $h^2\Omega_{\rm GW}(f)$. This holds even though uncertainties due to some close-by binaries can be relevant for $\bar{N}(f_-, f_+) \lesssim 100$ [47]. If $\bar{N}(f_-, f_+) \sim \mathcal{O}(\text{few})$, i.e. the GW signal is composed of only a few binaries, the uncertainty in the signal prediction is considerable. Arguably, it would be more appropriate to search for individual GW events instead of a GWB [44–46]. If $\bar{N}(f_-, f_+) \ll 1$, even having a single PBH binary emitting GWs is unlikely and, though it is possible to have an unexpectedly large background as there is a single binary nonetheless, in most realisations $h^2\Omega_{\rm GW,loc}(f) = 0$.

3 Production and constraints

There are a variety of different possible production mechanisms for PBHs, ranging from the collapse of initial Gaussian or non-Gaussian density fluctuations [49], the collapse of domain walls [50], scalar condensates [51], collapse of long-lived string-domain wall networks [52] or first-order phase transitions [53–55]. While significant clustering is not expected for Gaussian primordial fluctuations [56, 57], highly clustered PBH distributions could arise from a variety of different processes, e.g. due to primordial non-Gaussianities [58, 59] or from the collapse

[4] The *actual* number of binaries is then Poisson distributed with mean $\bar{N}(f_-, f_+)$.

of domain walls [50]. In this work we remain agnostic about the origin of the PBHs and about their spatial distribution and concentrate on exploring the phenomenological impact of different assumptions.

Regardless of how PBHs are produced, there are a number of different astrophysical and cosmological constraints on the abundance of heavy PBHs, which we take from [60] and briefly review here. These limits assume a monochromatic mass function for the PBHs as well as a roughly homogeneous spatial distribution (no clustering). While we will briefly comment on the expected impact of sizeable clustering, a full re-evaluation of these limits would require going beyond some of the simple approximations made in the original derivations and is beyond the scope of this letter. Also note that many of the different constraints come with different uncertainties and sometimes also with additional caveats.

The most relevant limits in the mass range of interest come from the heating of stars in the Galactic disk [61], the tidal disruption of galaxies [61], the dynamical friction effect on halo objects [61], requiring successful formation of the observed large scale structure [62], and the non-observation of X-ray binaries [63]. For many of these limits to be valid at least one PBH per relevant cosmic structure (e.g. per galaxy or cluster) is required. For the case of a clustered PBH distribution, we expect that in some structures more than one PBH will reside while in others there will be no PBH at all, given the overall conservation of number density. This will likely weaken a number of these limits and move them to smaller masses for the case of large clustering. A proper evaluation of the PBH spatial distribution and the resulting limits will however require detailed simulations, which will be done elsewhere.

Depending on the production mechanism of the PBHs, strong limits may also arise from the observation of the CMB. In particular if PBHs form due to the tail of Gaussian density fluctuations, Silk damping will generally lead to μ-distortions of the CMB and result in strong constraints over a sizeable mass range [64]. However, this limit crucially depends on the assumed production mechanism and will in general be relaxed or may even be completely evaded if the origin of the PBHs is of a different nature. In fact, as discussed above, for the case of large clustering a different production mechanism will generally be required, calling into question the validity of these limits.

4 Fitting to pulsar timing data

We base our analysis on pulsar observations in the 12.5 yr data set of the NANOGrav collaboration, as the most recent PTA data is not yet publicly available. We plan to update our analysis as soon as the NANOGrav 15 yr data set is available and comment in Sec. 6 on the minor impact of the new data set on our analysis.

To perform fits of the measured NANOGrav signal to a given GW spectrum from PBH mergers we use a two-fold approach: We employ *(i)* the tools `enterprise` and `enterprise-extensions` [65, 66] as well as *(ii)* the much quicker code `cefyll` [67]. The latter relies on the simplifying assumption that the likelihood for a GW signal factorizes into the product of the five lowest Fourier modes' likelihoods. We checked that the best-fit regions in our final scans are consistent with an analysis of the full "production-level" likelihood in `enterprise`, whereas all shown results have been produced with `cefyll`. In both cases we assume spatially uncorrelated signals. This simplification reduces the run-time of the otherwise even more computationally expensive evaluations of the likelihood, while giving an almost identical fit to the data, cf. Fig. 6 in Ref. [7].

For the analysis using `enterprise` we follow the tutorial available at [68], in which 45 pulsars of the NANOGrav 12.5 yr data set are used to derive bounds on the GW spectrum. While we use `PTMCMC` [69] in the `enterprise` analysis, we use the nested sampling algorithm [70] implemented in `UltraNest` [71] to draw samples in the $(m_{\rm PBH}, f_{\rm PBH})$ parameter space, keeping the density contrast $\delta_{\rm dc}$ fixed to either 1 or 10^3, corresponding to the cases of no or relevant PBH clustering, respectively. We fix the prior ranges of the PBH parameters to $f_{\rm PBH} \in [10^{-5}, 1]$ and $m_{\rm PBH} \in [10^3, 10^{12}]\, M_\odot$ and use log priors on both quantities. We find no considerable difference between their posterior distributions using both approaches over most of the the parameter space, justifying the much quicker fit using the Bayesian periodogram.

To obtain constraints on the (clustered) PBH scenario, we perform an additional scan over an extended model parameter space $(m_{\rm PBH}, f_{\rm PBH}, A_{\rm GWB})$ including an additional mock GW signal contribution (with $\gamma = 13/3$ and $A_{\rm GWB} \in [10^{-18}, 10^{-14}]$, cf. Eq. (2) in Ref. [3]) and then marginalise over $A_{\rm GWB}$. The region in the remaining $(m_{\rm PBH}, f_{\rm PBH})$ plane excluded with $2\,\sigma$ corresponds to the parameter space in which the GW signal would be too strong to account for the signal. To obtain our constraints on clustered PBHs we further conservatively remove the parameter space in which less than $\bar{N} = 10$ merger events would contribute to the NANOGrav signal.

We perform calculations of the produced chains using `numpy` [72] and plot them with the help of `matplotlib` [73] and a customized version of `ChainConsumer` [74].

5 Results

In Fig. 2 we show the regions in PBH mass $m_{\rm PBH}$ and DM fraction $f_{\rm PBH}$, where the NANOGrav signal can be explained assuming merging PBHs with no clustering (left) and significant clustering with $\delta_{\rm dc} = 10^3$ (right). In deriving these regions we have used the global average of the GWB signal. Increasing deviations of the signal observable in NANOGrav from the global signal can be expected if the expected number of binaries $\bar{N}$ contributing within the relevant frequency range drops to around or below 10 (dashed dark red line), cf. Sec. 2.2. If $\bar{N} < 1$ (full red line) one in fact would not even expect any GWB signal in most realisations of PBH binary distributions or else a signal which has an amplitude in excess of what is observed when naively extrapolating the best-fit region. In addition a possible signal will no longer be consistent with a stochastic background in this parameter region. These effects were neglected in [32] that falsely concluded that the NANOGrav signal could be explained in a region of parameter space without clustering, where only $\bar{N} \lesssim 1$ PBH binaries are expected to contribute. In this work we only consider the NANOGrav contours to be reliable when $\bar{N} \gtrsim 10$ around and to the left of the dark red line (dashed). In summary we find that the case without clustering cannot explain the NANOGrav signal once taking into account cosmological and astrophysical constraints.

For the case of clustered PBHs, the best-fit NANOGrav region shifts towards smaller $f_{\rm PBH}$, as clustering increases the merger rate and therefore also the GWB signal. Overall we find that in this case there will be an allowed region provided that the PBH formation mechanism does not result in significant μ-distortions. Note that in addition the astrophysical constraints are likely to become weaker for the case of clustering as described in Sec. 3, even though a proper re-evaluation of these constraints is beyond the scope of this work. The potentially allowed region will therefore likely even be larger than indicated in the right panel of Fig. 2, opening up a sizeable region in parameter space where merging PBHs can explain the PTA observations.

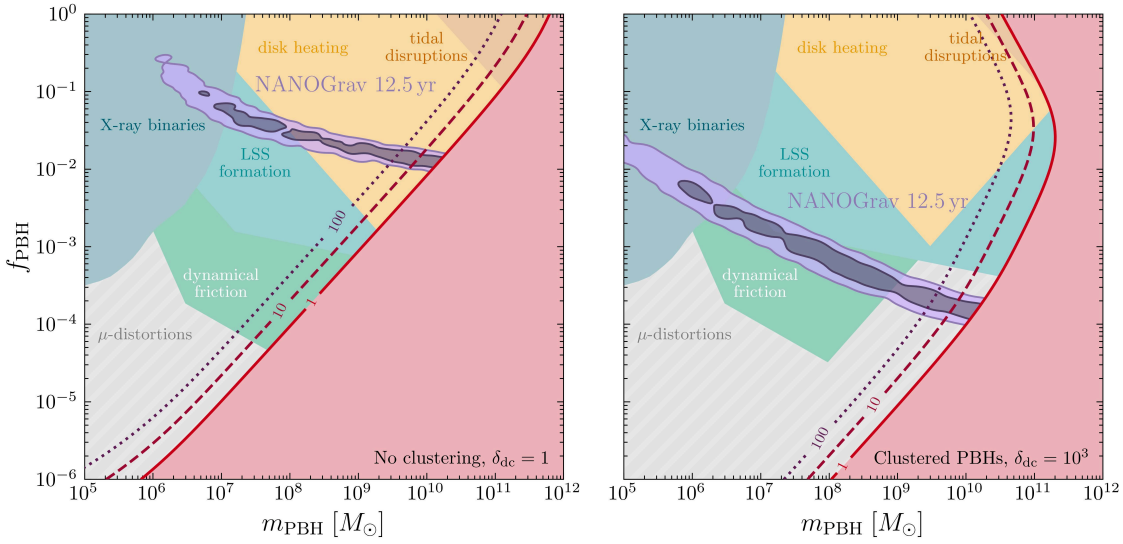


Figure 2. Best-fit region (purple) in the plane of PBH mass $m_{\rm PBH}$ vs. abundance $f_{\rm PBH}$, where the NANOGrav signal can be explained by merging PBH binaries without clustering (left) and with significant clustering (right). We indicate regions, where $\bar{N} < 1$, 10, and 100 PBH binaries are expected to contribute to the signal, noting that for $\bar{N} \lesssim 10$ uncertainties in the signal prediction would become relevant and for $\bar{N} < 1$ no signal is expected for most of the time. We also show the relevant constraints on PBHs as discussed in Sec. 3, which were derived assuming no clustering, and indicate that the constraints from μ-distortions can be very relevant, but depend on the production mechanism.

If the GWB signal generated by merging PBH binaries becomes too large, corresponding to parameters above the purple regions in Fig. 2, the abundance of PBHs can be *constrained* by PTA observations. Here we conservatively require an expected number of at least $\bar{N} = 10$ binaries contributing to the signal. The constraints are shown in Fig. 3 without clustering (purple) and with significant clustering $\delta_{\rm dc} = 10^3$ (red).

6 Discussion and conclusions

In this work we have studied the possibility that the signal observed by various pulsar timing arrays is due to merging primordial supermassive black hole binaries. If the PBHs are "homogeneously" distributed at their formation, i.e. follow a Poisson distribution, significant cosmological and astrophysical constraints exclude the possibility of explaining the PTA signal with merging PBHs. Instead considering a clustered spatial distribution of PBHs increases the binary merger rate and thus enables a consistent explanation of the PTA signals with merging PBH binaries. Crucially, we have checked that also the signal prediction is reliable in the relevant parameter space by computing the expected number of binaries contributing to the gravitational wave signal. Further, we used PTA data to constrain the PBH parameter space when the GWB generated during the mergers would result in stronger signal strengths than the one detected.

Due to publicly available data that can be reanalysed for new models we concentrated on the 12.5 yr NANOGrav data set, but intend to update our results once the new 15 yr data

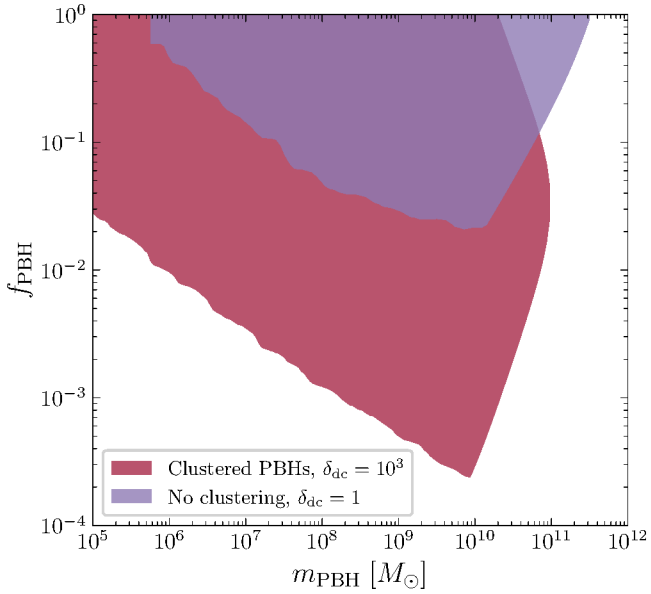


Figure 3. Constraints on PBHs from PTA observations due to a GWB signal strength exceeding observations with no clustering (purple) and significant clustering $\delta_{\rm dc} = 10^3$ (red). We conservatively cut the limits when the expected number of binaries contributing to the signal falls below 10, cf. Sec. 2.2.

set (or the data sets of other PTAs) becomes publicly available. This new data for the first time showed significant evidence for the Hellings-Downs correlation. Still, we do not expect a strong influence on the findings of our letter. The preferred signal amplitudes in the results of the NANOGrav collaboration increased only slightly, which will translate to a small shift of our signal and constraint regions to larger PBH abundances.

This work may serve as a motivation for model builders to construct scenarios that can generate clustered supermassive PBHs without running into cosmological and astrophysical constraints, in particular due to μ-distortions of the cosmic microwave background arising from PBH formation. While the latest PTA data finds no evidence for individual compact binary merger events on top of a stochastic background or anisotropies of the GW spectrum, the situation might change in the future. Such a detection would likely invalidate most other cosmological explanations, but is also a prediction of merging supermassive PBH binaries.

Acknowledgments

We would like to thank Pedro Schwaller for the initial idea which led to this project as well as for extensive discussions. We would further like to thank Andrea Mitridate, Xiao Xue, and Felix Kahlhöfer for helpful discussions on related works. This work is funded by the Deutsche Forschungsgemeinschaft (DFG) through Germany's Excellence Strategy – EXC 2121 "Quantum Universe" — 390833306.

References

[1] LIGO Scientific, Virgo collaboration, *Observation of Gravitational Waves from a Binary*

Black Hole Merger, *Phys. Rev. Lett.* **116** (2016) 061102 [1602.03837].

[2] LISA collaboration, *Laser Interferometer Space Antenna*, 1702.00786.

[3] NANOGrav collaboration, *The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background*, *Astrophys. J. Lett.* **905** (2020) L34 [2009.04496].

[4] B. Goncharov, R.M. Shannon, D.J. Reardon, G. Hobbs, A. Zic, M. Bailes et al., *On the evidence for a common-spectrum process in the search for the nanohertz gravitational-wave background with the parkes pulsar timing array*, 2107.12112v2.

[5] S. Chen, R.N. Caballero, Y.J. Guo, A. Chalumeau, K. Liu, G. Shaifullah et al., *Common-red-signal analysis with 24-yr high-precision timing of the european pulsar timing array: Inferences in the stochastic gravitational-wave background search*, 2110.13184v1.

[6] J. Antoniadis, Z. Arzoumanian, S. Babak, M. Bailes, A.S.B. Nielsen, P.T. Baker et al., *The international pulsar timing array second data release: Search for an isotropic gravitational wave background*, 2201.03980v1.

[7] NANOGrav collaboration, *The NANOGrav 15-year Data Set: Evidence for a Gravitational-Wave Background*, *Astrophys. J. Lett.* **951** (2023) [2306.16213].

[8] J. Antoniadis et al., *The second data release from the European Pulsar Timing Array III. Search for gravitational wave signals*, 2306.16214.

[9] D.J. Reardon et al., *Search for an isotropic gravitational-wave background with the Parkes Pulsar Timing Array*, *Astrophys. J. Lett.* **951** (2023) [2306.16215].

[10] H. Xu et al., *Searching for the nano-Hertz stochastic gravitational wave background with the Chinese Pulsar Timing Array Data Release I*, 2306.16216.

[11] R.w. Hellings and G.s. Downs, *UPPER LIMITS ON THE ISOTROPIC GRAVITATIONAL RADIATION BACKGROUND FROM PULSAR TIMING ANALYSIS*, *Astrophys. J. Lett.* **265** (1983) L39.

[12] J.A. Casey-Clyde, C.M.F. Mingarelli, J.E. Greene, K. Pardo, M. NaÃ±ez and A.D. Goulding, *A quasar-based supermassive black hole binary population model: implications for the gravitational-wave background*, 2107.11390v2.

[13] L.Z. Kelley, L. Blecha and L. Hernquist, *Massive black hole binary mergers in dynamical galactic environments*, 1606.01900v2.

[14] L.Z. Kelley, L. Blecha, L. Hernquist, A. Sesana and S.R. Taylor, *The gravitational wave background from massive black hole binaries in illustris: spectral features and time to detection with pulsar timing arrays*, 1702.02180v2.

[15] H. Middleton, A. Sesana, S. Chen, A. Vecchio, W. Del Pozzo and P.A. Rosado, *Massive black hole binary systems and the NANOGrav 12.5 yr results*, *Monthly Notices of the Royal Astronomical Society: Letters* **502** (2021) L99 [https://academic.oup.com/mnrasl/article-pdf/502/1/L99/36276026/slab008.pdf].

[16] D. Izquierdo-Villalba, A. Sesana, S. Bonoli and M. Colpi, *Massive black hole evolution models confronting the n-Hz amplitude of the stochastic gravitational wave background*, *Monthly Notices of the Royal Astronomical Society* **509** (2021) 3488 [https://academic.oup.com/mnras/article-pdf/509/3/3488/41360112/stab3239.pdf].

[17] M. Curyło and T. Bulik, *Predictions for lisa and pta based on shark galaxy simulations*, 2108.11232v2.

[18] J.J. Somalwar and V. Ravi, *The origin of the nano-hertz stochastic gravitational wave background: the contribution from $z \gtrsim 1$ supermassive black-hole binaries*, 2306.00898v1.

[19] NANOGrav collaboration, *The NANOGrav 15-year Data Set: Search for Signals from New Physics*, *Astrophys. J. Lett.* **951** (2023) [2306.16219].

[20] S. Vagnozzi, *Implications of the NANOGrav results for inflation*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) L11 [2009.13432].

[21] S. Blasi, V. Brdar and K. Schmitz, *Has nanograv found first evidence for cosmic strings?*, 2009.06607v3.

[22] J. Ellis and M. Lewicki, *Cosmic string interpretation of nanograv pulsar timing data*, 2009.06555v3.

[23] W. Buchmuller, V. Domcke and K. Schmitz, *From NANOGrav to LIGO with metastable cosmic strings*, *Phys. Lett. B* **811** (2020) 135914 [2009.10649].

[24] Y. Nakai, M. Suzuki, F. Takahashi and M. Yamada, *Gravitational waves and dark radiation from dark phase transition: Connecting nanograv pulsar timing data and hubble tension*, 2009.09754v2.

[25] W. Ratzinger and P. Schwaller, *Whispers from the dark side: Confronting light new physics with NANOGrav data*, *SciPost Phys.* **10** (2021) 047 [2009.11875].

[26] NANOGrav collaboration, *Searching for Gravitational Waves from Cosmological Phase Transitions with the NANOGrav 12.5-Year Dataset*, *Phys. Rev. Lett.* **127** (2021) 251302 [2104.13930].

[27] T. Bringmann, P.F. Depta, T. Konstandin, K. Schmidt-Hoberg and C. Tasillo, *Does NANOGrav observe a dark sector phase transition?*, 2306.09411.

[28] V. De Luca, G. Franciolini and A. Riotto, *NANOGrav Data Hints at Primordial Black Holes as Dark Matter*, *Phys. Rev. Lett.* **126** (2021) 041303 [2009.08268].

[29] V. Vaskonen and H. Veermäe, *Did NANOGrav see a signal from primordial black hole formation?*, *Phys. Rev. Lett.* **126** (2021) 051303 [2009.07832].

[30] K. Kohri and T. Terada, *Solar-Mass Primordial Black Holes Explain NANOGrav Hint of Gravitational Waves*, *Phys. Lett. B* **813** (2021) 136040 [2009.11853].

[31] A. Ashoorioon, K. Rezazadeh and A. Rostami, *NANOGrav signal from the end of inflation and the LIGO mass and heavier primordial black holes*, *Phys. Lett. B* **835** (2022) 137542 [2202.01131].

[32] V. Atal, A. Sanglas and N. Triantafyllou, *NANOGrav signal as mergers of Stupendously Large Primordial Black Holes*, *JCAP* **06** (2021) 022 [2012.14721].

[33] M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments*, Oxford University Press (2007), 10.1093/acprof:oso/9780198570745.001.0001.

[34] E.S. Phinney, *A Practical theorem on gravitational wave backgrounds*, astro-ph/0108028.

[35] P. Ajith et al., *A Template bank for gravitational waveforms from coalescing binary black holes. I. Non-spinning binaries*, *Phys. Rev. D* **77** (2008) 104017 [0710.2335].

[36] M. Raidal, V. Vaskonen and H. Veermäe, *Gravitational Waves from Primordial Black Hole Mergers*, *JCAP* **09** (2017) 037 [1707.01480].

[37] T. Nakamura, M. Sasaki, T. Tanaka and K.S. Thorne, *Gravitational waves from coalescing black hole MACHO binaries*, *Astrophys. J. Lett.* **487** (1997) L139 [astro-ph/9708060].

[38] K. Ioka, T. Chiba, T. Tanaka and T. Nakamura, *Black hole binary formation in the expanding universe: Three body problem approximation*, *Phys. Rev. D* **58** (1998) 063003 [astro-ph/9807018].

[39] M. Sasaki, T. Suyama, T. Tanaka and S. Yokoyama, *Primordial Black Hole Scenario for the Gravitational-Wave Event GW150914*, *Phys. Rev. Lett.* **117** (2016) 061101 [1603.08338].

[40] P.C. Peters, *Gravitational Radiation and the Motion of Two Point Masses*, *Phys. Rev.* **136** (1964) B1224.

[41] T. Bringmann, P.F. Depta, V. Domcke and K. Schmidt-Hoberg, *Towards closing the window of primordial black holes as dark matter: The case of large clustering*, *Phys. Rev. D* **99** (2019) 063532 [1808.05910].

[42] Y. Ali-Haïmoud, E.D. Kovetz and M. Kamionkowski, *Merger rate of primordial black-hole binaries*, *Phys. Rev. D* **96** (2017) 123523 [1709.06576].

[43] M. Raidal, C. Spethmann, V. Vaskonen and H. Veermäe, *Formation and Evolution of Primordial Black Hole Binaries in the Early Universe*, *JCAP* **02** (2019) 018 [1812.01930].

[44] IPTA collaboration, *Searching for continuous Gravitational Waves in the second data release of the International Pulsar Timing Array*, *Mon. Not. Roy. Astron. Soc.* **521** (2023) 5077 [2303.10767].

[45] J. Antoniadis et al., *The second data release from the European Pulsar Timing Array IV. Search for continuous gravitational wave signals*, 2306.16226.

[46] NANOGrav collaboration, *The NANOGrav 15-year Data Set: Bayesian Limits on Gravitational Waves from Individual Supermassive Black Hole Binaries*, 2306.16222.

[47] J. Ellis, M. Fairbairn, G. Hütsi, M. Raidal, J. Urrutia, V. Vaskonen et al., *Prospects for Future Binary Black Hole GW Studies in Light of PTA Measurements*, 2301.13854.

[48] A. Sesana, A. Vecchio and C.N. Colacino, *The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays*, *Mon. Not. Roy. Astron. Soc.* **390** (2008) 192 [0804.4476].

[49] B.J. Carr and S.W. Hawking, *Black holes in the early Universe*, *Mon. Not. Roy. Astron. Soc.* **168** (1974) 399.

[50] K.M. Belotsky, V.I. Dokuchaev, Y.N. Eroshenko, E.A. Esipova, M.Y. Khlopov, L.A. Khromykh et al., *Clusters of primordial black holes*, *Eur. Phys. J. C* **79** (2019) 246 [1807.06590].

[51] E. Cotner and A. Kusenko, *Primordial black holes from scalar field evolution in the early universe*, *Phys. Rev. D* **96** (2017) 103002 [1706.09003].

[52] F. Ferrer, E. Masso, G. Panico, O. Pujolas and F. Rompineve, *Primordial Black Holes from the QCD axion*, *Phys. Rev. Lett.* **122** (2019) 101301 [1807.01707].

[53] M. Lewicki, P. Toczek and V. Vaskonen, *Primordial black holes from strong first-order phase transitions*, 2305.04924v1.

[54] Y. Gouttenoire and T. Volansky, *Primordial black holes from supercooled phase transitions*, 2305.04942v1.

[55] M.J. Baker, M. Breitbach, J. Kopp and L. Mittnacht, *Primordial black holes from first-order cosmological phase transitions*, 2105.07481v2.

[56] Y. Ali-Haïmoud, *Correlation Function of High-Threshold Regions and Application to the Initial Small-Scale Clustering of Primordial Black Holes*, *Phys. Rev. Lett.* **121** (2018) 081304 [1805.05912].

[57] V. Desjacques and A. Riotto, *Spatial clustering of primordial black holes*, *Phys. Rev. D* **98** (2018) 123533 [1806.10414].

[58] S. Young and C.T. Byrnes, *Long-short wavelength mode coupling tightens primordial black hole constraints*, *Phys. Rev. D* **91** (2015) 083521 [1411.4620].

[59] T. Matsubara, T. Terada, K. Kohri and S. Yokoyama, *Clustering of primordial black holes formed in a matter-dominated epoch*, *Phys. Rev. D* **100** (2019) 123544 [1909.04053].

[60] B. Carr, K. Kohri, Y. Sendouda and J. Yokoyama, *Constraints on primordial black holes*, *Rept. Prog. Phys.* **84** (2021) 116902 [2002.12778].

[61] B.J. Carr and M. Sakellariadou, *Dynamical constraints on dark compact objects*, *Astrophys. J.* **516** (1999) 195.

[62] B. Carr and J. Silk, *Primordial Black Holes as Generators of Cosmic Structures*, *Mon. Not. Roy. Astron. Soc.* **478** (2018) 3756 [1801.00672].

[63] Y. Inoue and A. Kusenko, *New X-ray bound on density of primordial black holes*, *JCAP* **10** (2017) 034 [1705.00791].

[64] B.J. Carr and J.E. Lidsey, *Primordial black holes and generalized constraints on chaotic inflation*, *Phys. Rev. D* **48** (1993) 543.

[65] J.A. Ellis, M. Vallisneri, S.R. Taylor and P.T. Baker, "Enterprise: Enhanced numerical toolbox enabling a robust pulsar inference suite." Zenodo, Sept., 2020. 10.5281/zenodo.4059815.

[66] S.R. Taylor, P.T. Baker, J.S. Hazboun, J. Simon and S.J. Vigeland, *enterprise_ extensions*, 2021.

[67] W.G. Lamb, S.R. Taylor and R. van Haasteren, *The need for speed: Rapid refitting techniques for bayesian spectral characterization of the gravitational wave background using ptas*, 2303.15442v2.

[68] S. Taylor, S. Vigeland, J. Simon, B. Becsy and A. Johnson, *Pta gwb analysis*, 2021.

[69] J. Ellis and R. van Haasteren, *Ptmcmcsampler: Official release*, 2017. 10.5281/zenodo.1037579.

[70] J. Skilling, *Nested Sampling*, *AIP Conference Proceedings* **735** (2004) 395 [https://pubs.aip.org/aip/acp/article-pdf/735/1/395/11702789/395_1_online.pdf].

[71] J. Buchner, *UltraNest - a robust, general purpose Bayesian inference engine*, *The Journal of Open Source Software* **6** (2021) 3001 [2101.09604].

[72] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau et al., *Array programming with NumPy*, *Nature* **585** (2020) 357.

[73] J.D. Hunter, *Matplotlib: A 2d graphics environment*, *Computing in Science & Engineering* **9** (2007) 90.

[74] S.R. Hinton, *ChainConsumer*, *The Journal of Open Source Software* **1** (2016) 00045.