Hierarchical Bayesian method for detecting continuous gravitational waves from an ensemble of pulsars

M. Pitkin[*] and C. Messenger[†]

SUPA, School of Physics & Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom

X. Fan[‡]

Department of Physics & Astronomy, Hubei University of Education, 430205 Wuhan, China
(Dated: September 7, 2018)

When looking for gravitational wave signals from known pulsars, targets have been treated using independent searches. Here we use a hierarchical Bayesian framework to combine observations from individual sources for two purposes: to produce a detection statistic for the whole ensemble of sources within a search, and to estimate the hyperparameters of the underlying distribution of pulsar ellipticities. Both purposes require us to assume some functional form of the ellipticity distribution, and as a proof of principle we take two toy distributions. One is an exponential distribution, defined by its mean, and the other is a half-Gaussian distribution defined by its width. We show that by incorporating a common parameterized prior ellipticity distribution we can be more efficient at detecting gravitational waves from the whole ensemble of sources than trying to combine observations with a simpler non-hierarchical method. This may allow us to detect gravitational waves from the ensemble before there is confident detection of any single source. We also apply this method using data for 92 pulsars from LIGO's sixth science run. No evidence for a signal was seen, but 90% upper limits of 3.9×10^{-8} and 4.7×10^{-8} were set on the mean of an assumed exponential ellipticity distribution and the width of an assumed half-Gaussian ellipticity distribution, respectively.

PACS numbers: 02.50.Ng, 04.80.Nn, 95.55.Ym, 95.85.Sz, 97.60.Gb

I. INTRODUCTION

So far, Advanced LIGO [1] and Advanced Virgo [2] have detected gravitational waves from the time varying mass quadrupole caused by the orbital motion of two compact objects: binary systems consisting of two black holes [3, 4] or two neutron stars [5]. However, individual neutron stars with deformations can also have a time varying mass quadrupole that is not detectable as a transient signal with a chirplike nature, but is long-lived and quasimonochromatic. From such sources, the expected gravitational wave signal strength at Earth is given by[6]

$$h_0 \approx 4.2 \times 10^{-24} \frac{f_{\rm kHz}^2}{D_{\rm kpc}} \left(\frac{I_{zz}}{10^{38}\,{\rm kg\,m^2}} \right) \left(\frac{\varepsilon}{10^{-6}} \right), \quad (1)$$

where $f_{\rm kHz}$ is the star's rotation frequency in kHz, $D_{\rm kpc}$ is its distance in kpc, and the combination of the fiducial ellipticity [7], $\varepsilon = \frac{|I_{xx} - I_{yy}|}{I_{zz}}$, and principal moment of inertia, I_{zz}, are related to the mass quadrupole moment (based on the definition of [8]) by

$$Q_{22} = \sqrt{\frac{15}{8\pi}} I_{zz} \varepsilon. \quad (2)$$

Galactic neutron stars that are observed as pulsars are an enticing target for such signals. The times of arrival of electromagnetic pulses from pulsars can, in many cases, be used to fit their phase evolution to great accuracy. This leads to high precision measurements of the spin frequency and frequency derivative, often sub-milliarcsecond position reconstruction, and for sources in multiple systems, estimates of their orbital parameters. Under the assumption that any pulsar's mass quadrupole is formed by its having a triaxial moment of inertia ellipsoid (or, more colloquially, a "hill," "mountain," or "bump"), and that the rotation is phase locked with the electromagnetic emission, then the gravitational wave signal will be at exactly twice the observed frequency. Therefore, the precise phase evolution of any associated gravitational wave signal can be used as a filter for the signal in gravitational wave data. The only unknowns in the signal model are therefore the gravitational wave amplitude at Earth, as given in Eq. (1), and the orientation of the pulsar given by the inclination angle, ι, polarization angle, ψ, and relative phase at the epoch of the electromagnetic timing model, ϕ_0. As in, e.g., [6, 9–13], following the method of [14], Bayesian inference can be used to produce a joint posterior probability density function on these four parameters, and the orientation parameters can be marginalized over to give a posterior on only h_0. If the distance to a pulsar is known then this can be used to make inferences on, or set a limit on, the mass quadrupole moment, Q_{22}, or alternatively, the pulsars fiducial ellipticity assuming a canonical moment of inertia of $I_{zz} = 10^{38}\,{\rm kg\,m^2}$ [7].[1]

In previous searches for known pulsars in gravitational

---

* matthew.pitkin@glasgow.ac.uk
† christopher.messenger@glasgow.ac.uk
‡ fanxilong@outlook.com

---

[1] The canonical moment of inertia is roughly what you get from

wave data from the LIGO, Virgo and GEO600 detectors [6, 9–13] no signal has been seen, but upper limits, at a 95% credible level, were set on h_0 for all included pulsars. In the latest results [13], using data from the first observing run (O1) of the advanced LIGO detectors [15], this produced limits on h_0 for 200 pulsars, which in turn gave limits on Q_{22} and ε using the best-fit measured distances given in the ATNF Pulsar Catalogue [16]. The lowest limits on Q_{22} and ε found for any pulsar were $9.7 \times 10^{29}\,\mathrm{kg\,m^2}$ and 1.3×10^{-8} respectively, for the relatively nearby ($\sim 0.2\,\mathrm{kpc}$) pulsar PSR J0636+5129.

Different neutron star equations of state can allow different sizes of deformations to be sustained in the star; stiffer equations of state allow larger deformations than softer ones. A good review of the *maximum* allowed ellipticities is given in [7], which shows that in very extreme cases they could be in the range of 10^{-3}, but more realistically, for crustal deformations, they could reach a few $\times 10^{-6}$. Internal magnetic fields can also induce an ellipticity in a neutron star (e.g., [17]), with field strengths of order $10^{15}\,\mathrm{G}$ required to give rise to ellipticities of $\sim 10^{-6}$. For known pulsars, assuming their rate of loss of rotational energy is all due to gravitational wave emission, one can infer the ellipticity required to give the observed spin-down (called the "spin-down limit"). For the population of millisecond pulsars, such maximum possible ellipticities are generally well above the spin-down limit, which give values in the range $\sim 10^{-9} - 10^{-8}$. So, it is fair to assume that this population of pulsars is not strained to their maximum allowed values. The recent work in [18] provides some tentative evidence that millisecond pulsars may actually have a *minimum* ellipticity of $\sim 10^{-9}$.

Searches for gravitational waves from pulsars have always treated each pulsar individually. However, as we will describe further in this paper, it is possible to combine observations from many pulsars to try and detect the ensemble of all pulsars used in a search. The idea is that there may be several sources that are *individually* below some allowed threshold for detection, but when combined the ensemble rises above a detection threshold. This was proposed in [19], in which values of the $\mathcal{F}$-statistic [20], commonly used in continuous gravitational wave searches, for individual sources are summed. The method of [19] depends rather sensitively on the relative strengths of the brightest few sources, and is unlikely to be able to detect an ensemble of similar strength but individually undetectable, sources, i.e., it does not "win" by much over single source searches. This idea was extended in [21] to weight the individual $\mathcal{F}$-statistic values based on the expected detectability of each source, i.e., favoring close-by sources, which was found to always provide a more sensitive detection statistic than [19].

In this paper we also address the task of combining an

ensemble of sources, but for two purposes: to try and estimate the functional form of the underlying distribution of pulsar ellipticities, *and* to use the combined sources as an ensemble detection method. The former of these could allow us to make some physical inferences about the equation of state, or internal magnetic field strength, of the population of neutron stars. The latter, as discussed above, could allow us to detect the ensemble before an individual source is confidently detected.

II.   METHOD

Our aim is to combine data from all pulsars within a gravitational wave signal search to try and estimate the underlying distribution of their fiducial ellipticities (or alternatively, Q_{22}). We can do this within the context of hierarchical Bayesian inference, whereby we use the data to infer parameters of a probability distribution that represents the prior on the ellipticities (in this context these parameters are often called *hyper*parameters). This is essentially the same underlying method as described in [22], where in that context the main use is for detecting a stochastic background of compact binaries coalescences that are individually unresolvable.

In previous searches, each pulsar has been treated individually, and a posterior distribution on the gravitational wave amplitude, h_0, has been estimated, via

$$p(h_0|\mathbf{x}_i, I) \propto \int^{\boldsymbol{\theta}_i} p(\mathbf{x}_i|h_0, \boldsymbol{\theta}_i, I) p(h_0|I) p(\boldsymbol{\theta}_i|I) \mathrm{d}\boldsymbol{\theta}_i, \quad (3)$$

where $\mathbf{x}_i$ is the data for the ith pulsar, $\boldsymbol{\theta}_i = \{\cos\iota, \phi_0, \psi\}_i$ are the pulsar's orientation parameters, $p(\mathbf{x}_i|h_0, \boldsymbol{\theta}_i, I)$ is the likelihood of $\mathbf{x}_i$ given the signal model and particular parameter values, $p(\boldsymbol{\theta}_i|I)$ is the prior probability on the $\boldsymbol{\theta}_i$ parameters (which is flat over the allowed parameter ranges, see Sec. 3.2 of [23]), and $p(h_0|I)$ is the prior on h_0. More recently, in [13], the marginal likelihood (or Bayesian evidence) for each pulsar's data being consistent with the expected signal model has also been calculated, via

$$p(\mathbf{x}_i|I) = \int^{\boldsymbol{\theta}_i} \int^{h_0} p(\mathbf{x}_i|h_0, \boldsymbol{\theta}_i, I) p(h_0|I) p(\boldsymbol{\theta}_i|I) \mathrm{d}\boldsymbol{\theta}_i \mathrm{d}h_0. \quad (4)$$

The prior on h_0 has often been flat between zero and some hard upper cutoff, or having a Fermi-Dirac-like distribution (see, e.g., Sec. 2.3.5 of [24]) with a flat section followed by an exponential-like decay at some predefined value.

Alternatively, we can rearrange Eq. (1), and for each pulsar we can estimate the joint posterior on ε_i and the pulsar's distance, D_i,

$$p(\varepsilon_i, D_i|\mathbf{x}_i, I) = \frac{p(\varepsilon_i|I)}{Z_i} \int^{\boldsymbol{\theta}_i} p(\mathbf{x}_i|\varepsilon_i, D_i, \boldsymbol{\theta}_i, I)$$
$$\times p(D_i|I) p(\boldsymbol{\theta}_i|I) \mathrm{d}\boldsymbol{\theta}_i, \quad (5)$$

---

assuming a uniform density spherical star with a mass of $1.4\,\mathrm{M_\odot}$ and radius of $10\,\mathrm{km}$.

where $Z_i \equiv p(\mathbf{x}_i|I)$ is the evidence of the data (we will use the term *evidence*, rather than *marginal likelihood* in the rest of this paper as it is more compact). To produce a posterior on ε_i we need to set the priors for D_i and ε_i. We can use a Gaussian prior on D_i defined by a mean value μ_{D_i}, taken as the best-fit estimate from the ATNF Pulsar Catalogue [16] (which is generally a dispersion measure-derived distance based on the galactic electron density model of [25]), and a standard deviation, σ_{D_i},[2]

$$p(D_i|\mu_{D_i}, \sigma_{D_i}, I) = \frac{1}{\sqrt{2\pi}\sigma_{D_i}} \exp\left(-\frac{(D_i - \mu_{D_i})^2}{2\sigma_{D_i}^2}\right). \quad (6)$$

For each pulsar the distance is independent, and as the distances are not of interest to us, we can marginalize over them in Eq. (5) as nuisance parameters.[3]

The prior on ε_i can be a function that represents the underlying distribution of pulsar ellipticities. We will say that the function is defined by a set of parameters, which in this context are called the hyperparameters, $\boldsymbol{\Theta}$ that are intrinsic to the pulsar population. We can combine likelihoods for all N pulsars in a search, where for each we marginalize over the parameters $\boldsymbol{\theta}_i$, D_i, and ε_i, leaving a likelihood for $\boldsymbol{\Theta}$,

$$p(\mathbf{X}|\boldsymbol{\Theta}, I) = \prod_i^N \int^{\varepsilon_i} \int^{\boldsymbol{\theta}_i} \int^{D_i} p(\mathbf{x}_i|\varepsilon_i, \boldsymbol{\theta}_i, D_i, I)p(\boldsymbol{\theta}_i|I)p(\varepsilon_i|\boldsymbol{\Theta}, I)p(D_i|\mu_{D_i}, \sigma_{D_i}, I)\mathrm{d}\varepsilon_i\mathrm{d}\boldsymbol{\theta}_i\mathrm{d}D_i. \quad (7)$$

where $\mathbf{X} \equiv \{\mathbf{x}_i\}$ means the combined data from all pulsars.[4] We are interested in estimating the parameters of the underlying ellipticity distribution, $\boldsymbol{\Theta}$, and also calculating the evidence for the data given our particular chosen ellipticity distribution. To do this we further need to define a prior on $\boldsymbol{\Theta}$ and apply Bayes theorem,

$$p(\boldsymbol{\Theta}|\mathbf{X}, I) = \frac{p(\boldsymbol{\Theta}|I)p(\mathbf{X}|\boldsymbol{\Theta}, I)}{p(\mathbf{X}|I)}, \quad (8)$$

where the evidence is

$$p(\mathbf{X}|I) = \int^{\boldsymbol{\Theta}} p(\mathbf{X}|\boldsymbol{\Theta}, I)p(\boldsymbol{\Theta}|I)\mathrm{d}\boldsymbol{\Theta}. \quad (9)$$

The posterior on $\boldsymbol{\Theta}$ allows us to estimate the distribution of ε, while we can use the evidence, $p(\mathbf{X}|I)$, for the hypothesis that the data is consistent with that particular distribution to perform model selection. For example,

we could calculate the ratio of this evidence to one where the hypothesis is that the data for all pulsars consists purely of noise, i.e. the Bayes factor or odds (assuming prior odds of unity), and use this as a detection statistic for the ensemble of sources.

A. Ellipticity distribution priors

We study two different toy models for the underlying ellipticity distribution: an exponential distribution defined only by its mean, μ_ε,

$$p(\varepsilon|\mu_\varepsilon, I) = \frac{1}{\mu_\varepsilon}e^{-\varepsilon/\mu_\varepsilon} \quad (10)$$

such that $\boldsymbol{\Theta} \equiv \mu_\varepsilon$; and, a half-Gaussian distribution peaking at zero, and defined by its width, σ_ε,

$$p(\varepsilon|\sigma_\varepsilon, I) = \frac{2}{\sqrt{2\pi}\sigma_\varepsilon}e^{-\varepsilon^2/2\sigma_\varepsilon^2}, \quad (11)$$

such that $\boldsymbol{\Theta} \equiv \sigma_\varepsilon$. We note that more attention is given to the exponential model in most of our examples.

We also need to set a prior on $\boldsymbol{\Theta}$ for both cases. For both models we chose a prior on the hyperparameter that is uniform in log-space between some lower and upper bounds, e.g., for the exponential distribution we have

$$p(\mu_\varepsilon|I) = [\ln(\mu_{\varepsilon\max}/\mu_{\varepsilon\min})]^{-1}\frac{1}{\mu_\varepsilon}, \quad (12)$$

and equivalently for σ_ε. For the analyses described in Sec. III we use a lower bound of 10^{-10} and an upper bound of 10^{-5}.

In our analyses we assume that all pulsar ellipticities are drawn from these toy distributions. However, in reality they are likely to be too simplistic to describe the true ε distribution. For example, the population of young pulsars and old recycled millisecond pulsars, have undergone different evolutions, with the latter having most

---

[2] Throughout this work we will take the distance uncertainty to be 20% of the μ_{D_i} value, which is roughly what Figure 12 in [26] suggests for dispersion measure-based distance estimates. For some pulsars better distance estimates are available (using, for example, parallax measurements obtained using very long baseline interferometry), while for others the uncertainties can be considerably worse, so in a detailed analysis more reliable uncertainties could be used for individual pulsars.

[3] For pulsars in the same globular cluster the distance measurements and error would essentially be the same, or at least very highly correlated, but we will ignore that fact in this study. Of the pulsars currently listed in the ATNF Pulsar Catalogue [16] with rotation frequencies above 10 Hz, 139 of the 451 pulsars are in globular clusters. The clusters with the largest numbers of pulsars are 47 Tucanae, with 25, and Terzan 5, with 35.

[4] It is worth emphasizing that even though the same raw data is used for all pulsars, the precise signal and the Gaussian component of the colored noise spectrum for each pulsar will be entirely independent. Tracking the signal's precise phase over a long (weeks, months, or years) observation time means that the phase templates for different sources (even if there were billions of sources!) are highly orthogonal.

likely gone through a stage of accretion (see, e.g., the review in [27]) that spins them up and reduces their external dipole magnetic field.[5] This could mean that the distribution of ellipticities for these two populations may be quite different, and therefore a bimodal distribution, or two independent exponential distributions with different means, could be more appropriate. Alternatively, it may be more appropriate to use a quasinonparametric approach, such as the histogram binning in [28], an infinite Gaussian mixture model (see, e.g., [29] for a recent example of this in the context of gravitational waves), or a Gaussian process to model the ellipticity distribution function space. All these approaches have larger $\boldsymbol{\Theta}$ parameter spaces to marginalize over and are therefore computationally more challenging. We leave the exploration of these ideas to future work.

B. Using posterior samples

If we have many pulsars, e.g., the 200 used in [13], and wanted to directly calculate Eqs. (8) and (9) then it would require integrals over 200×4 independent parameters ($\boldsymbol{\theta}_i = \{\cos\iota, \phi_0, \psi\}_i$ and D_i for each pulsar), and a large data set consisting of the data for each pulsar. However, as those parameters and the noise in each pulsar's gravitational wave data are independent, we can calculate the likelihood over ε_i for each pulsar individually, as is already done for current targeted pulsar searches [13]. This simplifies Eq. 7 to be

$$p(\mathbf{X}|\boldsymbol{\Theta}, I) = \prod_i^N \int^{\varepsilon_i} p(\mathbf{x}_i|\varepsilon_i, I)p(\varepsilon_i|\boldsymbol{\Theta}, I)\mathrm{d}\varepsilon_i. \qquad (13)$$

In reality those searches do not output a functional form of the likelihood for each pulsar, but they use nested sampling [30] to compute the evidence that the data contains a signal (coherently combined over multiple detectors), $p(\mathbf{x}_i|I)$, and produce samples drawn from the marginalized *posterior* distribution $p(\varepsilon_i|\mathbf{x}_i, I)$. The posterior samples therefore need to be converted into a functional form of the likelihood for use in Eq. (8).[6]

First, to convert these samples into a smooth functional form we can use kernel density estimation (KDE), with a Gaussian kernel. To remove edge effects for samples that rail against the lower bound of zero we reflect all

_____

[5] Current searches for gravitational waves from pulsars [13] consider all pulsars with rotation frequencies above $10\,\mathrm{Hz}$. Within this subset of ~ 450 pulsars, as given by the ATNF Pulsar Catalogue [16], about 50 of them have high spin-down rates and as such would be considered "young", while the rest are old recycled pulsars.

[6] In reality the software that performs the nested sampling [24] and outputs the posterior samples uses Q_{22} rather than ε, but we can easily convert between the two using Eq. (2) and assuming a principal moment of inertia of $10^{38}\,\mathrm{kg\,m^2}$.

the samples about zero and concatenate these with the original samples before performing the KDE. When evaluating the resulting KDE only at allowed positive values of ε it must then be multiplied by two to get the correct probability density.

Next, rearranging Bayes theorem shows how to turn a posterior for an individual pulsar into a likelihood

$$p(\mathbf{x}_i|\varepsilon_i, I) = \frac{p(\varepsilon_i|\mathbf{x}_i, I)}{p(\varepsilon_i|I)}p(\mathbf{x}_i|I), \qquad (14)$$

where in this case ε_i is the prior on ε used for the individual pulsar. If the prior $p(\varepsilon_i|I)$ is uniform (i.e. a constant) in some range within which the likelihood goes to zero then this is simple to calculate. In all our examples this is the case, with ε_i being defined between zero and an upper range $\varepsilon_{\max}$, such that

$$p(\varepsilon_i|I) = \varepsilon_{\max}^{-1}. \qquad (15)$$

To be explicit, we are undoing an effective uniform prior that was used when calculating the original pulsar posteriors, so that we can then reapply our new ellipticity distribution prior for the whole ensemble. We can therefore substitute

$$p(\mathbf{x}_i|\varepsilon, I) = \varepsilon_{\max}p(\varepsilon_i|\mathbf{x}_i, I)p(\mathbf{x}_i|I), \qquad (16)$$

into Eq. (7). We could also work with samples in h_0 rather than ellipticity, and numerically marginalize over distance errors, which we show in Appendix A, but we will not use h_0 samples in this paper.

We have found that in our analysis, when we have many posteriors that peak at, or close to, zero, and a prior function that increases rapidly as it approaches zero, we cannot use the method as used in, e.g., [28]. That method approximates the integral over the likelihood multiplied by the prior with the expectation value of the prior evaluated at each of the samples. We find that the accuracy of this approach is significantly reduced for small values of prior parameters like μ_ε (see Sec. II A above) when there are no samples with roughly equivalent values.

C. Spin-down limits

We could use the electromagnetically-derived spin-down limits described briefly in Sec. I to infer the hyperparameters of our ellipticity distribution. For each pulsar we have a spin-down limit on ε based on its observed electromagnetic spin-down, independent of the pulsar's distance (combining Eqs. (5) and (6) of [12])

$$\varepsilon^{\mathrm{sd}} = \frac{1.9 \times 10^{-8}}{f_{\mathrm{kHz}}^{5/2}}\left(\frac{|\dot{f}_{\mathrm{rot}}|}{10^{-11}\,\mathrm{Hz\,s^{-1}}}\right)^{1/2}\left(\frac{10^{38}\,\mathrm{kg\,m^2}}{I_{zz}}\right)^{1/2}. \qquad (17)$$

The observed frequency derivatives of pulsars are not necessarily their true values, as they will be contaminated by

proper motion effects (the Shklovskii effect [31]), differential galactic rotation [32], and, for pulsars in globular clusters, local accelerations. Indeed, these effects lead to some pulsars being observed to spin-up. Therefore in any inference using spin-down limits one would have to either correct for these effect, or exclude pulsars for which the effects cannot be estimated.

We can use these spin-down limit derived values to provide a likelihood on ε_i for each pulsar when estimating the underlying distribution's hyperparameters. For an individual pulsar we can (simplistically) say that the spin-down limit gives us a flat likelihood on ε

$$p(d_{\text{EM}i}|\varepsilon_i, I) = \begin{cases} 1/\varepsilon_i^{\text{sd}} & \text{if } 0 \leq \varepsilon_i \leq \varepsilon_i^{\text{sd}}, \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

where $d_{\text{EM}i}$ is the electromagnetically derived information, i.e., from observed pulse time of arrivals, that gives the spin-down limit for a given pulsar.

These likelihoods for each pulsar can be combined into a joint likelihood, and form a likelihood on the distribution's hyperparameters, Θ,

$$p(\{\mathbf{d}_{\text{EM}}\}|\Theta, I) = \prod_i^N \left(\int p(d_{\text{EM}i}|\varepsilon_i, I)p(\varepsilon_i|\Theta, I)\mathrm{d}\varepsilon_i \right). \quad (19)$$

If we assume the underlying distribution is exponential (so $\Theta \equiv \mu_\varepsilon$) we have

$$p(\{\mathbf{d}_{\text{EM}}\}|\mu_\varepsilon, I) = \prod_i^N \left(\frac{1}{\varepsilon_i^{\text{sd}}} \right) \int_0^{\varepsilon_i^{\text{sd}}} \frac{1}{\mu_\varepsilon} e^{-\varepsilon_i/\mu_\varepsilon} \mathrm{d}\varepsilon_i, \quad (20)$$

where we can see that the integral is just the cumulative distribution function for an exponential distribution, which is given by $\text{CDF}(\varepsilon_i^{\text{sd}}, \mu_\varepsilon) = 1 - e^{-\varepsilon_i^{\text{sd}}/\mu_\varepsilon}$. So, we have

$$p(\{\mathbf{d}_{\text{EM}}\}|\mu_\varepsilon, I) \propto \prod_i^N 1 - e^{-\varepsilon_i^{\text{sd}}/\mu_\varepsilon}. \quad (21)$$

A similar result can be found for the half-Gaussian distribution using its CDF.

Using the priors given in Sec. II A we can calculate the posteriors $p(\mu_\varepsilon|\{\mathbf{d}_{\text{EM}}\}, I)$ and $p(\sigma_\varepsilon|\{\mathbf{d}_{\text{EM}}\}, I)$ for the exponential and half-Gaussian distributions, respectively. To do this we use all pulsars in the ATNF Pulsar Catalogue [16], excluding those in globular clusters and with observed spin-ups, estimate their intrinsic period derivatives (by calculating the combined Shklovskii and galactic rotation effects using best-fit distances from the catalogue), and calculate their spin-down limits using Eq. (17). For this simple analysis we ignore uncertainties in each pulsar's moment of inertia. The posterior probability distributions on μ_ε and σ_ε for the exponential and half-Gaussian are shown in Fig. 1. We find 90% credible upper bounds on these two hyperparameters of $\mu_\varepsilon^{90\%} \leq 3.0 \times 10^{-10}$ and $\sigma_\varepsilon^{90\%} \leq 4.1 \times 10^{-10}$.
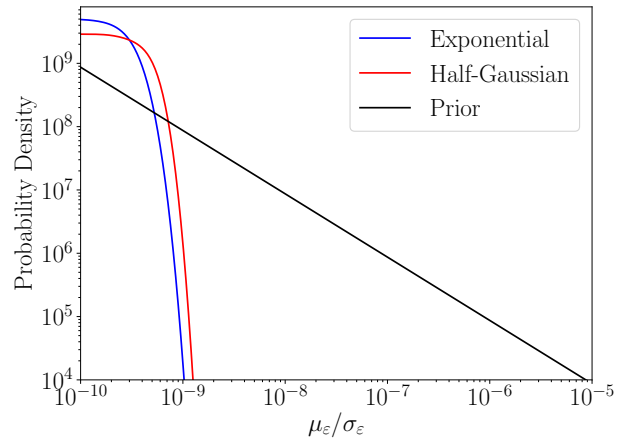

FIG. 1. The posterior probability distributions on the ellipticity distribution prior hyperparameters μ_ε and σ_ε for the exponential and half-Gaussian distributions when using electromagnetically-derived spin-down limits.

In the rest of this paper we will purely assume the use of gravitational wave data alone for making inferences about the ellipticity distribution. However, an interesting extension to this work could be combining the gravitational wave data for pulsars, where that places a more stringent limit than the spin-down limit, with spin-down limits for the pulsars where it does not.

D. Bayes factor

We saw in Eq. (9) that we can calculate the evidence that the data is consistent with a particular parameterized ellipticity distribution (i.e., the exponential or half-Gaussian distributions). However, to make this number informative we need to be able to compare it to another model. We can form the odds between this evidence and one given by a different hypothesis, such as the data from all pulsars being entirely consistent with Gaussian noise, e.g.

$$\mathcal{O}_{\text{n}}^\varepsilon = \frac{p(\mathbf{X}|H_\varepsilon, I)}{p(\mathbf{X}|H_n, I)} \frac{p(H_\varepsilon|I)}{p(H_n|I)}, \quad (22)$$

where H_ε explicitly denotes the hypothesis of our data containing signals drawn from the particular underlying ellipticity distribution, and H_n denotes the noise-only hypothesis. The first term on the right-hand side, the ratio of evidences, is often called the Bayes factor, while the second term is the prior odds of the two hypotheses, which we will generally take as unity unless otherwise stated.

1. Noise hypothesis

The simplest way to form the evidence for the noise hypothesis is just as the product of the null likelihoods for all pulsars, where the null likelihood is the likelihood with the signal model explicitly set of zero (see, e.g., Sec. 2.2.3 of [24] for the null likelihoods used in the analysis presented here). For a particular pulsar i and detector j the null likelihood for the hypothesis that the data is consistent with noise alone, $H_{\rm n}$, is given by

$$p(\mathbf{x}_i^j|H_{\rm n}^j, I) \equiv p(\mathbf{x}_i^j|\varepsilon_i = 0, I), \qquad (23)$$

i.e., it is just the likelihood [as seen in, e.g., Eq. (5)] evaluated with the signal model set to zero. Therefore, with $\mathbf{x}_i \equiv \{\mathbf{x}_i^j\}$, the multidetector ($N_{\rm det}$) noise hypothesis for a single pulsar is

$$p(\mathbf{x}_i|H_{\rm n}, I) = \prod_j^{N_{\rm det}} p(\mathbf{x}_i^j|H_{\rm n}^j, I), \qquad (24)$$

and evidence for the noise hypothesis for all pulsars is

$$p(\mathbf{X}|H_{\rm n}, I) = \prod_i p(\mathbf{x}_i|H_{\rm n}, I). \qquad (25)$$

In reality such a noise hypothesis is not robust against instrumental spectral disturbances and lines in the data (see, e.g., [33]), which can contaminate the frequencies near, or at, those of an expected astrophysical pulsar signal. If data from more than one detector is available, a more robust noise hypothesis is to allow it to incorporate incoherent signals *or* noise between detectors. When using multiple detectors, the individual pulsar signal hypothesis enforces coherence between them, so we can more explicitly call the hypothesis the *coherent* hypothesis, $H_{\rm coh}$, which is what the evidence term $p(\mathbf{x}_i|I) \equiv p(\{\mathbf{x}_i^j\}|H_{\rm coh}, I)$ in Eqs. (5), (14) and (16) truly represents.

The evidence for the new noise hypothesis for each pulsar, which we will call the *incoherent* hypothesis, is described in more detail in Sec. 2.6 of [24], or the Appendix of [13]. In all the analyses presented here we use two detectors (the LIGO Hanford and Livingston detectors, which we will denote as H1 and L1 respectively from here onwards), so for each pulsar the evidence for the incoherent hypothesis, $H_{\rm in}$ is

$$p(\mathbf{x}_i|H_{\rm in}) = \prod_{j\in\{\rm H1,L1\}} \left[p(\mathbf{x}_i^j|H_{\rm s}^j, I) + p(\mathbf{x}_i^j|H_{\rm n}^j, I)\right], \quad (26)$$

where $H_{\rm s}^j$ is the evidence for the hypothesis that the data is consistent with a signal for a single detector. This is generalizable to any number of detectors. Again, we can take the product of this to give the incoherent hypothesis for all N pulsars

$$p(\mathbf{X}|H_{\rm in}) = \prod_i^N p(\mathbf{x}_i|H_{\rm in}). \qquad (27)$$

This implies that in Eq. (22) we would substitute $H_{\rm n} \equiv H_{\rm in}$, which is the case for all results presented here.

2. Nonhierarchical statistic

Without using any of the machinery in defining an underlying ellipticity distribution, and hyperparameters, described previously in this section, we can form a more naïve odds. This assumes all the pulsars have mass quadrupoles (not ellipticities) drawn from the same flat prior distribution with fixed bounds (between zero and a large value at which all likelihoods have approached zero). To differentiate it from the above odds formed using an ellipticity distribution with unknown hyperparameters, we call this our *nonhierarchical* statistic.

As described above, for each pulsar we have the evidence for the coherent signal model $p(\{\mathbf{x}_i^j\}|H_{\rm coh}, I)$ and the null likelihood $p(\mathbf{x}_i|H_{\rm n}, I)$. Following a similar route to Eqs. (49–52) of [34] we can form the probability (*not* the evidence here as we are explicitly including priors on the hypotheses) for a compound hypothesis of any combination of individual signal and noise hypotheses

$$p(H_{\rm com}|\mathbf{X}, I) \propto \prod_i^N \Big[p(\mathbf{x}_i|H_{{\rm coh}\,i}, I)p(H_{{\rm coh}\,i}|I)+$$
$$p(\mathbf{x}_i|H_{{\rm in}\,i}, I)p(H_{{\rm in}\,i}|I)\Big]-$$
$$\prod_i^N p(\mathbf{x}_i|H_{{\rm in}\,i}, I)p(H_{{\rm in}\,i}|I), \qquad (28)$$

where $p(H_{{\rm coh}\,i}|I)$ and $p(H_{{\rm in}\,i}|I)$ are the priors for each hypothesis, which here we explicitly state for each pulsar i. The second product term on the right-hand side of Eq. (28) is also the probability for the all-pulsar noise hypothesis, so we can produce an odds between the two probabilities (noting that they would both have the same proportionality coefficient and it would thus cancel out) as

$$\mathcal{O} \equiv \frac{p(H_{\rm com}|\mathbf{X}, I)}{p(H_{\rm in}|\mathbf{X}, I)} = \frac{p(\mathbf{X}|H_{\rm com}, I)}{p(\mathbf{X}|H_{\rm in}, I)}\frac{p(H_{\rm com}|I)}{p(H_{\rm in}|I)}$$
$$= \left(\prod_i^N \Big[p(\mathbf{x}_i|H_{{\rm coh}\,i}, I)p(H_{{\rm coh}\,i}|I)+\right.$$
$$\left.p(\mathbf{x}_i|H_{{\rm in}\,i}, I)p(H_{{\rm in}\,i}|I)\Big]\right)\times$$
$$\left(\prod_i^N p(\mathbf{x}_i|H_{{\rm in}\,i}, I)p(H_{{\rm in}\,i}|I)\right)^{-1} - 1. \qquad (29)$$

If we state that we have equal *a priori* probability for every combination of the coherent signal and incoherent

hypotheses, e.g.,

$$\prod_{i=1}^{N} p(H_{\text{in}\,i}|I) = \prod_{i=2}^{N} p(H_{\text{in}\,i}|I) \prod_{j=1}^{1} p(H_{\text{coh}\,j}|I)$$

$$= \prod_{i=3}^{N} p(H_{\text{in}\,i}|I) \prod_{j=1}^{3} p(H_{\text{coh}\,j}|I)$$

$$= \dots, \tag{30}$$

then Eq. (29) becomes

$$\mathcal{O} = \frac{\prod_{i}^{N} [p(\mathbf{x}_i|H_{\text{coh}\,i}, I) + p(\mathbf{x}_i|H_{\text{in}\,i}, I)]}{\prod_{i}^{N} p(\mathbf{x}_i|H_{\text{in}\,i}, I)} - 1, \tag{31}$$

where implicitly $p(H_{\text{com}}|I) = (2^N - 1)p(H_{\text{in}}|I)$. Therefore, to assign equal prior odds for the compound hypothesis that data for any pulsar contains a signal one must use

$$\mathcal{O}_{\text{n}}^{\text{NH}} = \frac{\mathcal{O}}{2^N - 1}. \tag{32}$$

We will use this as a comparison to Eq. (22) in the subsequent analyses.

3. $\mathcal{B}$-statisticlike comparison

It is interesting to compare the above statistics to that proposed in [19]. Due to how we have constructed our data sets, it is not simple to calculate the $\mathcal{F}$-statistic (which is the log-likelihood ratio maximized over the parameter space) for our simulated data for each pulsar. However, the natural logarithm of our signal versus noise evidence ratios for each pulsar are essentially the $\mathcal{B}$-statistic of [35] (although we additionally marginalize over pulsar distance) including the line robust incoherent noise denominator [33] in our Bayes factors. Therefore, as a comparison using an ensemble of pulsars, we can simply sum these $\mathcal{B}$-statistics. In our notation, where we are not implicitly in log-space, we would have a product rather than a sum of

$$\mathcal{O}_{\text{n}}^{\mathcal{B}} = \prod_{i}^{N} \frac{p(\mathbf{x}_i|H_{\text{coh}\,i}, I)}{p(\mathbf{x}_i|H_{\text{in}\,i}, I)}. \tag{33}$$

As was shown in [35] the $\mathcal{B}$-statistic is a slightly more efficient discriminator between signal and noise than the $\mathcal{F}$-statistic, so we would expect this to produce a comparable, but slightly more efficient statistic than that in [19].

On a final note, all the above likelihood, prior, posterior, and odds calculations are in practice computed entirely using the natural logarithm of the values to avoid numerical underflow and overflow.

III. ANALYSIS

Here we will discuss analyses that have been performed on simulated data to assess the performance of the three odds given in Eqs. (22), (31) and (33) for detecting an ensemble of gravitational wave signals from pulsars. In the former case this assumes that all pulsars have ellipticities drawn from some underlying distribution defined by unknown hyperparameters, while the latter two assume no unknown hyperparameters for the distribution. In the former case we also assess how well we can recover the hyperparameters defining the ellipticity distribution.

A. Simulations

To make these assessments we have produced a series of simulated data sets to account for different ellipticity distributions, hyperparameter values, and for different realizations of noise. In all cases we take the 200 pulsars searched for in the analysis of LIGO Observing Run 1 (O1) [13] as our sample of sources, with their sky positions and best-fit distances obtained from the ATNF Pulsar Catalogue [16]. We also create data from two detectors, the LIGO Hanford (H1) and Livingston (L1) observatories, assuming that they are operating at their advanced design sensitivities [36, 37] over one year with a 100% duty cycle. For each pulsar a complex time series is simulated at a sample rate of once per 1800 s, to replicate the data that would be produced in a real targeted pulsar search (e.g., [13]) following the application of a heterodyne procedure [14] to remove the rapidly varying signal phase evolution.

Firstly, we will discuss generating sources from the exponential ellipticity distribution defined in Eq. (10). There is a single hyperparameter that defines the distribution, μ_ε, the mean of the distribution. We take 15 values of μ_ε spaced uniformly in log-space between 5×10^{-10} and 5×10^{-8}. For an exponential distribution defined by a particular μ_ε value, we randomly draw ε values for each pulsar, which (assuming $I_{zz} = 10^{38}\,\text{kg}\,\text{m}^2$) we convert into that pulsar's equivalent h_0 via Eq. (1). We also randomly generate values of $\cos\iota$, ψ and ϕ_0 for each pulsar, drawn uniformly from the ranges $[-1, 1]$, $[0, \pi/2]$ and $[0, \pi]$, respectively. Using these values, the time series' for the two detectors are generated via the signal model defined in Equation 13 of [14] with additive Gaussian noise with zero mean and standard deviation derived from the Advanced LIGO (aLIGO) design curve sensitivity [36] at the appropriate gravitational wave frequency (twice the pulsar's rotation frequency). For each pulsar we also calculate the signal-to-noise ratio it would have via Eq. (2) of [38] for the two-detector fully coherent analysis.

For μ_ε values below $\sim 9.6\times10^{-10}$ and above 4.5×10^{-9} we regenerate the ensemble of sources 10 times using different random seeds, while for values between that range we regenerate the ensemble 100 times to provide a better statistical sample. So, to summarize, assuming an underlying exponential distribution of ellipticities, we have $(9 \times 10) + (6 \times 100) = 690$ realizations of data for two detectors containing signals from an ensemble of 200 pulsars.

We perform exactly the same steps for the half-Gaussian distribution defined in Eq. (11), where we take 15 values of σ_ε spaced uniformly in log-space over the same range as above.[7] In this case we only create 10 ensembles of pulsars for each value of σ_ε. So, we have $15 \times 10 = 150$ realizations of data for two detectors containing signals from an ensemble of 200 pulsars.

It is very useful to have "background" data sets with which to compare odds values from ensembles containing signals. So, for this purpose we have also generated 400 realizations of our data sets of 200 pulsars, with each pulsar's data containing simulated Gaussian noise derived from the aLIGO design sensitivity as above, but containing no signal.

B. Processing the data

For each pulsar in each ensemble realization we run the parameter estimation and evidence evaluation code [24] used for real known pulsar searches such as [13]. For each pulsar the code is run individually for the two detectors, and with data from both detectors combined coherently. The code uses a nested sampling algorithm [30, 39, 40] to evaluate the model evidence given by the integral in Eq. (5), and also outputs the null likelihood in each case. For all pulsars, the code set up was identical and 512 "live," or "active," points were used to initialize the nested sampling algorithm. For the signal variables Q_{22}, $\cos\iota$, ϕ_0, and ψ we defined uniform priors in the ranges $[0, 10^{37}]\,\mathrm{kg\,m^2}$, $[-1, 1]$, $[0, \pi]\,\mathrm{rad}$, and $[0, \pi/2]\,\mathrm{rad}$, respectively. The distance was also included as a variable and assigned a Gaussian prior with a mean given by each pulsar's best fit distance, and a standard deviation of 20% of that value, and a hard cutoff at zero. Note that we did not sample in ε, but the Q_{22} samples were easily converted into ε samples via the relation in Eq. (2).

Nested sampling outputs a chain of samples, where each sample is a vector containing particular values of each of the variable parameters $\{\varepsilon_j, \cos\iota_j, \phi_{0_j}, \psi_j, D_j\}$. These are then resampled to provide draws from the posterior distribution using the method described in [40]. As described in Sec. II B, for Eq. (16), which must be inserted into Eq. (13), we cannot use posterior samples, but instead need a functional form of the posterior. We use Gaussian kernel density estimation (KDE), in particular the method implemented in the SCIKIT-LEARN Python package [41], to convert the ε samples into a function that can be evaluated. In practice, when samples are cut-off at a hard boundary, such as not being allowed to be negative for the ε value, it can lead to boundary artifacts in the KDE. So, to avoid such artifacts we produce a copy of the samples, reflect them about zero, and concatenate

them to the original samples before performing the KDE. KDEs are also dependent on the bandwidth chosen for the Gaussian kernels. We use the Scott rule-of-thumb [42] to estimate the bandwidth of the kernel, but only using the original samples rather than the concatenated version to avoid the kernel being too broad. The use of a finite number of posterior samples, and their conversion into a KDE, means that there will be some associated uncertainties that will propagate through the analysis, which are discussed briefly in Appendix B.

For each ensemble of pulsars the outputs from the above processing for the coherent two-detector analyses are inserted into Eq. (13) to form the likelihood for a particular ellipticity distribution model. This is then used in Eq. (9), along with the hyperparameter priors in Sec. II A, to evaluate the evidence for that distribution (in practice trapezoidal integration is used over the hyperparameter range). This then gives the evidence for the numerator of Eq. (22). We also obtain the posterior distribution on the hyperparameter via Eq. (8). For both the simulated distributions (the exponential and half-Gaussian) we calculate Eq. (9) twice, once with the actual distribution used for the simulations and once assuming the alternative distribution, i.e., for the ensembles containing sources with ellipticities drawn from an exponential distribution we calculate the evidence that the distribution was exponential *and* the evidence that the distribution was half-Gaussian. This enables us to do model comparison between the two distributions.

Likewise the outputs from the individual detector analyses can be combined via Eqs. (26) and (27) to form the denominator of Eq. (22). This means that for each ensemble of pulsars drawn from a given distribution we have two values of $\mathcal{O}_\mathrm{n}^\varepsilon$. The first is $\mathcal{O}_\mathrm{n}^{\varepsilon_\mathrm{exp}}$ for the assumption that the distribution is exponential, and the second is $\mathcal{O}_\mathrm{n}^{\varepsilon_\mathrm{hg}}$ for the assumption that it is half-Gaussian.

Similarly, we use all the individual pulsar signal and noise evidences to produce values for $\mathcal{O}_\mathrm{n}^\mathrm{NH}$ via Eqs. (31) and (32), and values for $\mathcal{O}_\mathrm{n}^\mathcal{B}$ via Eq. (33), for each ensemble.

IV. RESULTS

A. Odds values

For each ensemble we have calculated the value of the odds for a given ε distribution: $\mathcal{O}_\mathrm{n}^{\varepsilon_\mathrm{exp}}$ and $\mathcal{O}_\mathrm{n}^{\varepsilon_\mathrm{hg}}$. We have also calculated the non-hierarchical odds assuming no unknown hyperparameters $\mathcal{O}_\mathrm{n}^\mathrm{ind}$, and $\mathcal{O}_\mathrm{n}^\mathcal{B}$. This has been done for the 400 background ensembles in which the data for each pulsar purely contains Gaussian noise. For reasons we discuss in Appendix B 2 the true scaling of odds values may not be reliable. However, as we are able to calculate the odds for a background distribution in an identical way to those containing signals, we can remove any scaling dependent effects by looking at ratios with respect to the background. So, we produce a detection

---

[7] The mean of a half-Gaussian distribution, which may be more directly comparable to the exponential distribution mean, is given by $(\sqrt{2/\pi})\sigma_\varepsilon$.

statistic from the ratio between the observed odds for each ensemble containing signals and the *mean* odds from the background ensembles

$$\mathcal{D} = \frac{\mathcal{O}}{\langle \mathcal{O}_{\mathrm{background}} \rangle}, \qquad (34)$$

where we might have, e.g.,

$$\mathcal{D}^{\varepsilon_{\exp}} = \frac{\mathcal{O}_{\mathrm{n}}^{\varepsilon_{\exp}}}{\langle \mathcal{O}_{\mathrm{n}}^{\varepsilon_{\exp}}{}_{\mathrm{background}} \rangle}, \qquad (35)$$

or equivalent values of $\mathcal{D}^{\mathrm{NH}}$ for the nonhierarchical odds $\mathcal{O}_{\mathrm{n}}^{\mathrm{NH}}$, and $\mathcal{D}^{\mathcal{B}}$ for the $\mathcal{B}$-statistic based odds from Eq. (33). The base-10 logarithm of these three ratios is what is shown in Fig. 2, and subsequent figures.

In the upper panel of Fig. 2 we see that the statistic $\mathcal{D}^{\varepsilon_{\exp}}$ appears to become completely disjoint from the background distribution at values of $\mu_{\varepsilon} \approx 5 \times 10^{-9}$, while the nonhierarchical statistic $\mathcal{D}^{\mathrm{NH}}$ becomes disjoint a little higher at $\mu_{\varepsilon} \approx 7 \times 10^{-9}$. We see that the $\mathcal{B}$-statistic-based distributions, $\mathcal{D}^{\mathcal{B}}$, have generally larger values, but have a much broader background distribution than the other two statistics (we see how this translates into efficiency in Sec. IV A 1). We see a rather similar situation in the lower panel of Fig. 2 for $\mathcal{D}^{\varepsilon_{\mathrm{hg}}}$ with the distributions becoming completely disjoint from the background at $\sigma_{\varepsilon} \approx 5 \times 10^{-9}$, or $(\sqrt{2/\pi})\sigma_{\varepsilon} \approx 6 \times 10^{-9}$.

In Fig. 3 we show a zoomed in version of Fig. 2 for both the exponential distribution (left) and half-Gaussian (right), with the results for $\mathcal{D}^{\varepsilon_{\exp}}/\mathcal{D}^{\varepsilon_{\mathrm{hg}}}$ and $\mathcal{D}^{\mathrm{NH}}$ split into separate panels (top and middle respectively). The background distribution is shown in two ways: horizontal solid lines show the upper and lower extent of the 400 background realizations; shaded bands show the bounds from 1σ to 5σ on the distribution. For comparison the distributions of $\mathcal{D}^{\mathcal{B}}$ and its background, also with bounds from 1σ to 5σ, are shown faintly. For $\mathcal{D}^{\varepsilon_{\exp}}/\mathcal{D}^{\varepsilon_{\mathrm{hg}}}$ the background distribution is roughly symmetrical and Gaussian, while for $\mathcal{D}^{\mathrm{NH}}$ the distribution is not symmetric and is more similar to a Gamma or χ^2-distribution. However, to reflect the true distribution as best we can we form a KDE from them, and use that to estimate the σ-bounds by finding the intervals that bound the probabilities defining the $1 - 5\sigma$ bounds of a Gaussian distribution.

The bottom panels of Fig. 3 show binned distributions of signal-to-noise ratios—the numbers within the boxes show the number of pulsars in that bin for the ensemble, whilst the number *above* the boxes shows the maximum signal-to-noise ratio of all pulsars within that ensemble. For each adjacent pair the left set of bins show the signal-to-noise ratio distribution for the ensemble with the *smallest* $\mathcal{O}_{\mathrm{n}}^{\varepsilon_{\exp}}/\mathcal{O}_{\mathrm{n}}^{\varepsilon_{\mathrm{hg}}}$ (which will be those that gave rise to the lower extent of the distributions shown above); and the right set of bins show the signal-to-noise ratio distribution for the ensemble with the *largest* $\mathcal{O}_{\mathrm{n}}^{\varepsilon_{\exp}}/\mathcal{O}_{\mathrm{n}}^{\varepsilon_{\mathrm{hg}}}$ (which will be those that gave rise to the upper extent of the distributions shown above). If we look at the rightmost value of μ_{ε} in the left panel of Fig. 3 we see that the

lowest value of $\mathcal{O}_{\mathrm{n}}^{\varepsilon_{\exp}}$ came from an ensemble in which all pulsars had signal-to-noise ratios of ≤ 5.4. For an individual pulsar analysis, such a signal-to-noise ratio could easily not be enough to confidently assign it as detected (the probable noise outlier in Fig. 2 of [13] had a similar signal-to-noise ratio, as did the rejected outliers in [43]), but this gave rise to a $\mathcal{D}^{\varepsilon_{\exp}}$ value well outside the background distribution as seen in the top panel. However, from the middle panel it is interesting to note that this ensemble is just on the edge of the background distribution for the nonhierarchical statistic $\mathcal{D}^{\mathrm{NH}}$. In the bottom panels on both sides we see what would be expected from the two different distributions; there is a longer tail giving rise to larger outlier signal-to-noise ratios for the exponential case than for the half-Gaussian case.

We will see what that means in terms of detection efficiency for the exponential distribution in Sec. IV A 1 and Fig. 4.

1. Detection efficiency

Using our distributions of statistics from the background analysis we can define a threshold in $\mathcal{D}$ at which to claim detection with a given false alarm probability (FAP). Here we do this in two ways, noting that we only do this for the exponential distribution as we have enough simulations to get reasonable statistics unlike for the half-Gaussian. The simplest threshold is one based on the maximum background value of $\mathcal{D}$, which, given our 400 background realizations, will represent a FAP of $1/400 = 0.0025$. The other way that we use is via extrapolation from the KDEs of the background distributions out to values of $\mathcal{D}$ that yield equivalent cumulative probabilities to that of a Gaussian distribution at 5σ. To compute efficiencies we find the number of ensembles for each μ_{ε} value that are above the two FAP $\mathcal{D}$ thresholds for $\mathcal{D}^{\varepsilon_{\exp}}$, $\mathcal{D}^{\mathrm{NH}}$, and $\mathcal{D}^{\mathcal{B}}$, compared to the total number of simulations at that value. These efficiency curves are shown in Fig. 4, which also shows (as shaded regions) the 90% credible regions for each curve, and sigmoid fits to the efficiencies as dashed lines.

We see that the false alarm probability based on the number of background realizations $\mathcal{D}^{\varepsilon_{\exp}}$ clearly outperforms both the $\mathcal{D}^{\mathrm{NH}}$ and $\mathcal{D}^{\mathcal{B}}$ statistics. For the false alarm probability extrapolated from the background out to 5σ the efficiencies of the two statistics, $\mathcal{D}^{\varepsilon_{\exp}}$ and $\mathcal{D}^{\mathrm{NH}}$, are far more comparable. It is worth noting that the efficiency curves for $\mathcal{D}^{\mathrm{NH}}$ for both FAPs are very similar. This is because the actual threshold value in both cases is very similar due to there being one large outlier that dominates the first FAP, but which only contributes a small amount to the extrapolated threshold (which will also depend on the KDE kernel width used for the background). Therefore, this KDE based threshold extrapolated to 5σ for $\mathcal{D}^{\mathrm{NH}}$ is probably unreliable and if more background realizations were performed larger outliers may be found. So, we expect the performance of the two
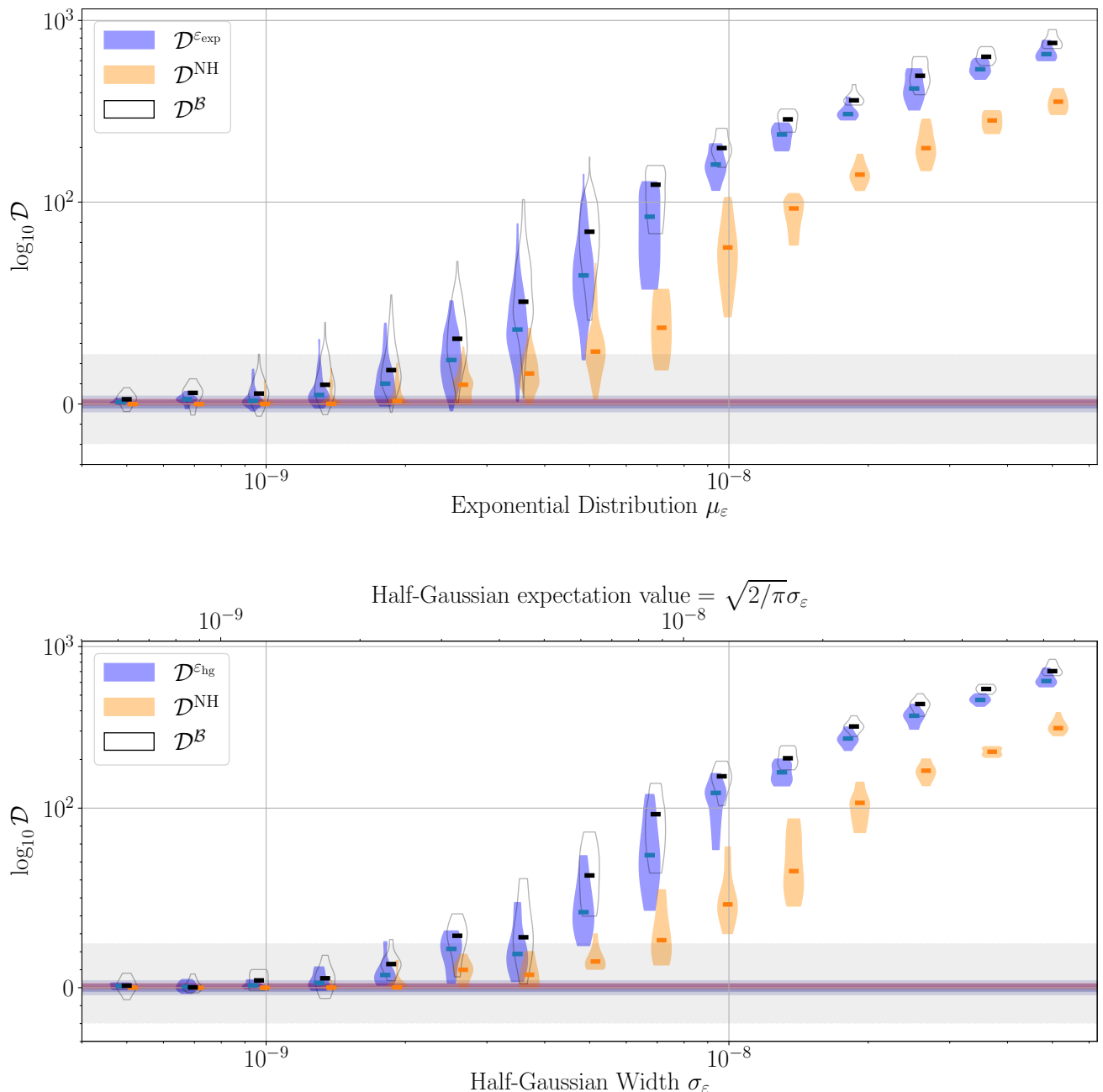
FIG. 2. The distributions of odds values for the ensembles of pulsars with ellipticities drawn from, in the upper panel, exponential distributions with a range of mean values, μ_ε, and in the lower panel, half-Gaussian distributions with a range of widths, σ_ε (the mean of the half-Gaussian distributions, $(\sqrt{2/\pi})\sigma_\varepsilon$, is displayed on the top axis). The distributions for the statistic assuming there is a common distribution $\mathcal{D}^{\varepsilon_{\mathrm{exp/hg}}}$, using the nonhierarchical method $\mathcal{D}^{\mathrm{NH}}$, and using the $\mathcal{B}$-statistic-based method $\mathcal{D}^{\mathcal{B}}$, are shown. The shaded bands show the 5σ extent of the background distributions in all cases (for clarity these are shown in greater detail in Fig. 3). Note that the y-axis is linear below 10^2 and logarithmic above it.

statistics, as observed in the left panel of Fig. 4 to be more reliable, and better represent the true gain by including the common ellipticity distribution. Both $\mathcal{D}^{\varepsilon_{\mathrm{exp}}}$ and $\mathcal{D}^{\mathrm{NH}}$ considerably outperform $\mathcal{D}^{\mathcal{B}}$ for the smallest distributions, i.e., when the contribution from the single strongest source is not overwhelming.
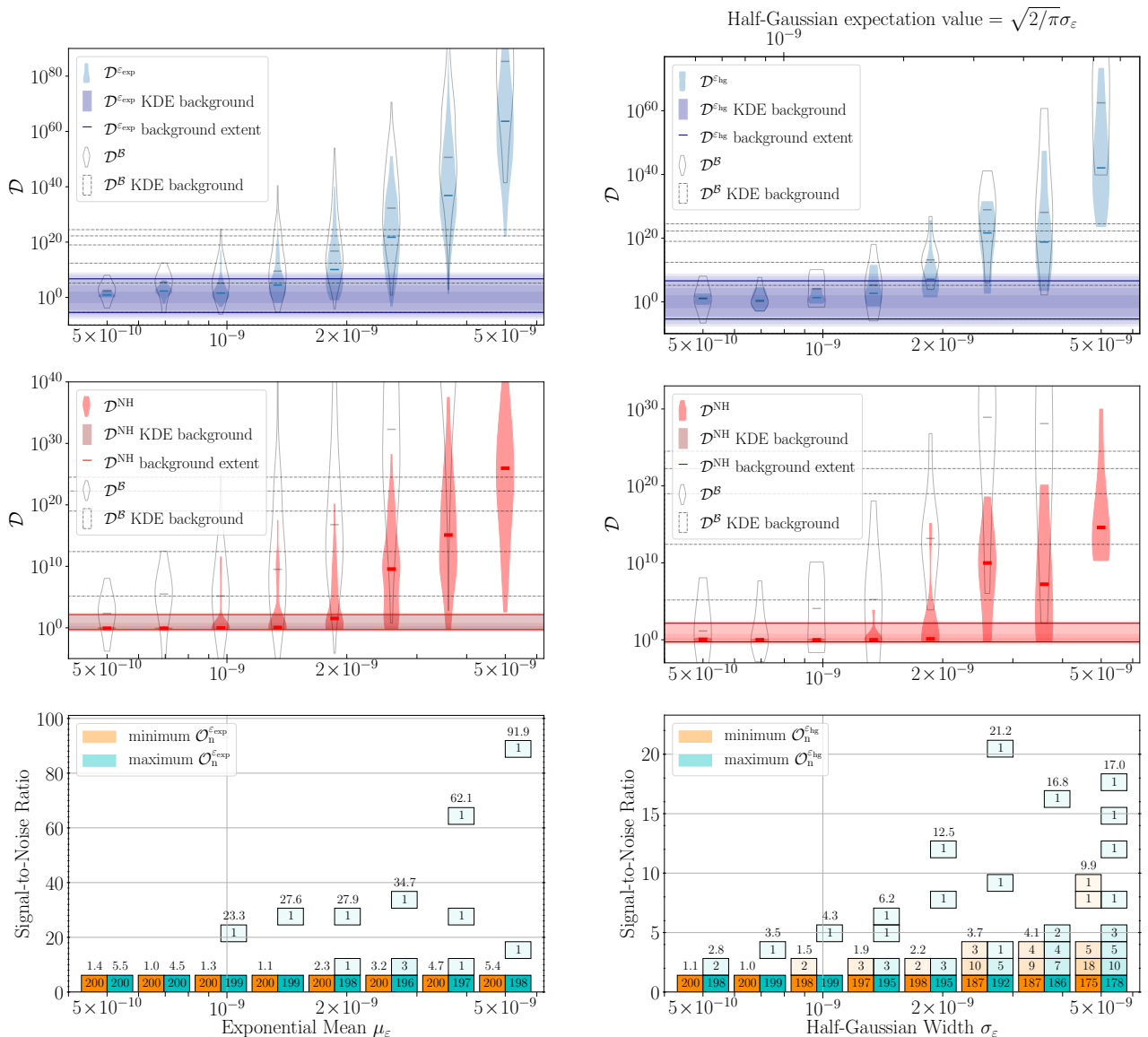
FIG. 3. The left panel shows results for the exponential distribution simulations and the right panel shows results for the half-Gaussian distribution simulations. On each side the top panel shows the distribution of detection statistics $\mathcal{D}^{\varepsilon\exp}$ (left)/$\mathcal{D}^{\varepsilon\mathrm{hg}}$ (right) for ensembles with distributions defined by eight different means/widths, μ_ε (left)/σ_ε (right), with the equivalent $\mathcal{D}^{\mathcal{B}}$ distribution shown faintly for comparison. The middle panel shows the distributions of $\mathcal{D}_{\mathrm{NH}}$ (again with $\mathcal{D}^{\mathcal{B}}$ shown for comparison). In both these panels the horizontal shaded regions show the distribution for the background (noise-only) ensembles (see text), with the solid horizontal lines showing the maximum and minimum extent of the distribution. The background distributions for the $\mathcal{D}^{\mathcal{B}}$ values are shown as the faint dashed lines. The bottom panel shows the distribution of signal-to-noise ratios for the ensemble with the minimum (left box of each pair) and maximum (right box of each pair) $\mathcal{D}^{\varepsilon\exp}$ (left)/$\mathcal{D}^{\varepsilon\mathrm{hg}}$ (right) value, with numbers inside the boxes showing the histogram count, and numbers above the boxes showing the maximum signal-to-noise ratio for that ensemble.

B. Parameter estimation

One of the purposes of this paper has been to develop a method to estimate the hyperparameters defining a common ellipticity distribution from which the ellipticities of pulsars are drawn. As described in Sec. II A we have

taken two simple distributions with which to test this: an exponential distribution and a half-Gaussian.

For each of the ensembles of pulsars described in Sec. III A we have estimated the posterior for the distribution means, μ_ε, for the exponential distribution simulations, and for the distribution widths, σ_ε, for the half-Gaussian distribution simulations. These can be seen as
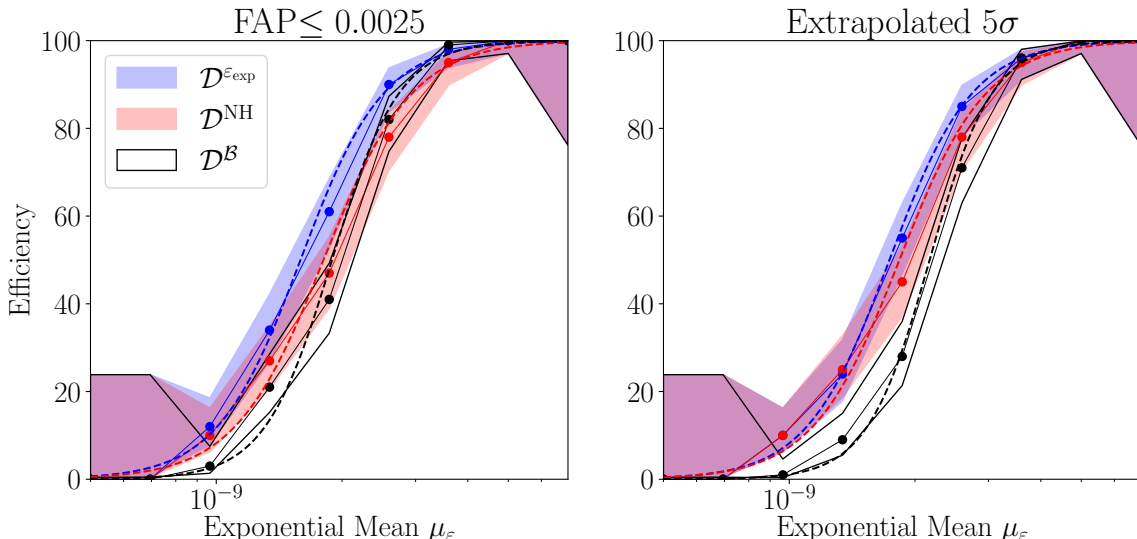
FIG. 4. Efficiency curves for the $\mathcal{D}^{\varepsilon_{\exp}}$, $\mathcal{D}^{\mathrm{NH}}$ and $\mathcal{D}^{\mathcal{B}}$ detection statistics for two different false alarm probabilities: a false alarm probability of 0.0025 based on the number of background realizations, and an equivalent 5σ false alarm probability extrapolated from KDEs of the background distributions. The dashed curves show sigmoid fits to the measured efficiencies.

a function of the simulated values used to construct the ensembles in Fig. 5 (top panel for the exponential distribution and bottom panel for the half-Gaussian). The plot shows the 90% credible interval on the posterior for μ_ε (top)/σ_ε (bottom) for the ensemble with the largest (dark error bars) and smallest (light error bars) odds, $\mathcal{O}_\mathrm{n}^{\varepsilon_{\exp}/\mathrm{hg}}$, for each simulated value. We see that the true distribution parameters are recovered accurately and, in all bar one case for the exponential distribution simulations, when no signal is detected (based on the FAP of 0.0025) the posteriors include the lower prior boundary on μ_ε (and thus in these case the maximum bound represents an upper limit).

C. Model comparison

We can compare the evidences for the two different models of the ellipticity distribution. As discussed earlier we have specified two toys distributions (the exponential and half-Gaussian distributions) and have calculated the evidence for both for all our ensembles of pulsars. We can take the ratio of these evidences and see which distribution is preferred in each case. Figure 6 shows this ratio for the ensembles created from sources with ellipticities drawn from the exponential distribution (top) and the half-Gaussian distribution (bottom). In the top panel, while there is a definite trend towards strongly favoring the correct distribution there are still ensembles with the largest exponential mean value that favor the half-Gaussian model. At the smallest values of μ_ε the half-Gaussian is always favored, probably due to it having a more sharply falling off tail and therefore smaller

prior volume. We find that the cases where the true exponential distribution is most highly favored are when there are a few outliers with large signal-to-noise ratios compared to the bulk of the distributions, which would be allowed by the longer tail exponential distribution, but not by the half-Gaussian. In the bottom panel of Fig. 6 we find that the true half-Gaussian distribution is generally favoured in the majority of cases, although we have fewer simulations with which to truly probe the tails of the distribution.

V. S6 RESULTS

Rather than purely working with simulated data with have also run on real data from the LIGO detectors. Here we present the results of this analysis using data from LIGO's sixth science run (S6) [44, 45].[8] The run had data from the two LIGO detectors at Hanford and Livingston operating in enhanced configuration between 8 July 2009 and 20 October 2010. We use 92 of the pulsars included in the analysis of [12].[9] For each pulsar

----

[8] The full LIGO S6 data set is publicly available through the LIGO Open Science Center https://losc.ligo.org/S6/ [46].

[9] For this analysis there was a selection cut meaning not all the pulsars in [12] were used. To enable the relatively quick production of background realizations of the data we made use of the more efficient spectral interpolation algorithm [47], rather than the standard heterodyne method [14], to process the S6 data. This method makes use of Fourier transforms of 1800 s chunks of data, which had been created for the analyses in [48, 49] and included a low frequency cut-off at 40 Hz, so pulsars with gravi-
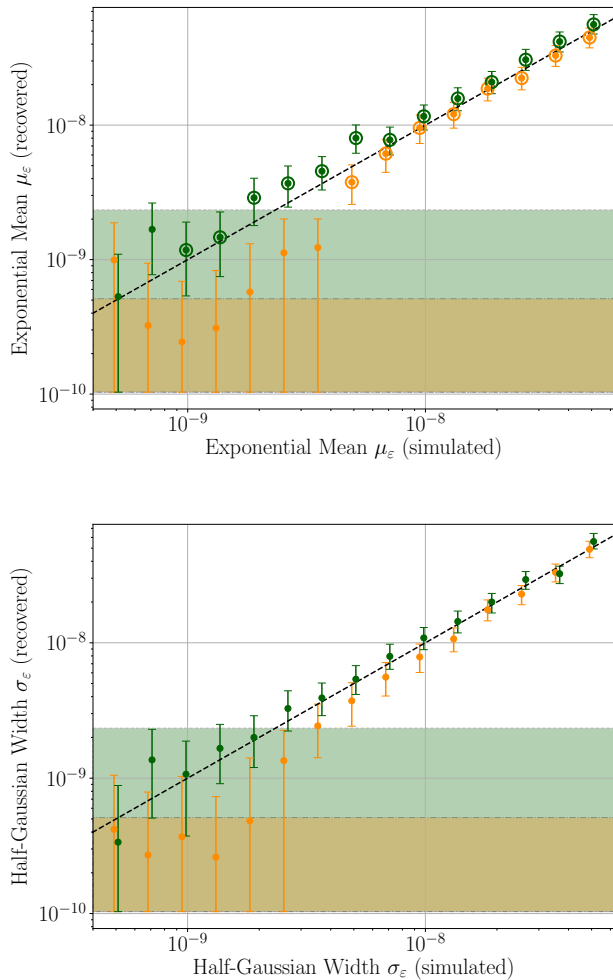
FIG. 5. Credible intervals (90%) about the median value from the recovered posterior probability distributions on the exponential ellipticity distribution hyperparameter μ_ε, across the range of simulated μ_ε values (top), and the half-Gaussian ellipticity distribution hyperparameter σ_ε, across a range of simulated σ_ε values (bottom). The dark green error bars represent the credible interval for the ensemble with the largest odds, while the orange error bars are for the ensemble with the lowest odds. These, therefore, are indicative of the general range of possible results. In the top plot error bars with open circles around the median values are those that would be considered "detected" above the $\mathcal{D}^{\varepsilon \exp}$ false alarm probability (see Sec. IV A 1) of 0.0025. The shaded bands show the intervals with the largest and smallest upper bound for ensembles consisting purely of noise.

FIG. 6. The distributions of ratio $\mathcal{O}_n^{\varepsilon \exp}/\mathcal{O}_n^{\varepsilon \mathrm{hg}}$ (top) for sources with ellipticities drawn from exponential distributions with given means μ_ε, and $\mathcal{O}_n^{\varepsilon \mathrm{hg}}/\mathcal{O}_n^{\varepsilon \exp}$ (bottom) for sources with ellipticities drawn from half-Gaussian distributions with given widths σ_ε.

and for each individual detector, and for both detectors combined, the evidence and posterior samples (marginalized over orientation and pulsar distance) as required in Eq. (16) were produced. As with the simulated data sets described in Sec. III A, using these and the noise evidence values we have calculated $\mathcal{O}_n^{\varepsilon \exp}$ and $\mathcal{O}_n^{\varepsilon \mathrm{hg}}$ for the ensemble of pulsars. As before, we require a background distribution of these values to compare our "foreground" to. To create a background we required data that shares the same noise characteristics as the foreground, but in which an astrophysical signal from a given pulsar would not be present. To achieve this we use the method described briefly in Section V.A. of [43] (also see [50]), in which for each pulsar we reprocess (using the spectral interpolation method from [47]) the data with the pulsar's sky location changed to a different randomly selected sky location. We performed this background generation 100

___

tational wave frequencies below this cut-off were excluded. The spectral interpolation method also has limitations for high spin-down pulsars or those in tight binaries, so many where excluded on these criteria.

times to find the distribution of values.

The foreground and background distribution of $\mathcal{D}^{\varepsilon_{\exp}}$, $\mathcal{D}^{\varepsilon_{\mathrm{hg}}}$, and $\mathcal{D}^{\mathrm{NH}}$ are shown in Fig. 7. We see that the foreground for both ellipticity distributions are well within the background distribution, so we see no evidence for gravitational wave emission from the ensemble of pulsars.

However, we can set upper limits on the ellipticity distribution hyperparameters. Figure 8 shows the posteriors for μ_ε and σ_ε for the exponential and half-Gaussian distributions, respectively. From these we find 90% credible upper limits of $\mu_\varepsilon^{90\%} \leq 3.9 \times 10^{-8}$ and $\sigma_\varepsilon^{90\%} \leq 4.7 \times 10^{-8}$. These are about two orders of magnitude less constraining than the purely spin-down limit based limits discussed in Sec. II C, although they are the first such limits to be set based purely on gravitational wave observations.

VI. CONCLUSIONS

In this work we have described a Bayesian hierarchical method for combining gravitational wave observations from an ensemble of known pulsars for two purposes: to create a detection statistic for identifying a signal from the ensemble, and to estimate the parameters of the distribution of pulsars' fiducial ellipticities.[10] For two toy ellipticity distributions, an exponential and a half-Gaussian, we have used simulations to find that incorporating this distribution as a common prior on the ellipticity of stars, with an unknown hyperparameter, can produce a more efficient detection statistic than combining the data for the ensemble of pulsars in a nonhierarchical way. We also show that it is more efficient than a statistic derived in a similar way to that in [19]. We find that the detection of the ensemble could even be seen in cases where individual sources may not be individually detectable with high confidence. However, we should note that the efficiency may not be improved if the true distribution does not well match our assumed prior form.

For ensembles for which gravitational wave emission would be considered detected we have shown in Fig. 5 that we can correctly constrain the hyperparameters of the simulated ellipticity distribution. If no signal is seen we can also set upper limits on these. However, as shown in Fig. 6, we have also found that it is difficult to distinguish between our two toy distributions as they are broadly similar.

---

[10] Here we have worked with fiducial ellipticities as they are a convenient and relatable quantity (i.e., they express the relative deformation of the star). However, the analysis actually estimates the mass quadrupole moment of the stars and converts that into the ellipticity given the canonical moment of inertia of $10^{38}\,\mathrm{kg\,m^2}$. So, one could convert back to the moment of inertia independent mass quadrupole if required. We should note that simulated signals were drawn from the fiducial ellipticity ε parameter, and we converted to mass quadrupoles using the canonical moment of inertia, so they do not incorporate a realistic equation-of-state dependent spread of moments of inertia.

We have performed the analyses using real data for 92 pulsars from the LIGO S6 science run, with the assumption of the same two ellipticity distributions: an exponential and a half-Gaussian. We saw no evidence of a signal from the ensemble, but set upper limits on the two distributions hyperparameters of $\mu_\varepsilon^{90\%} \leq 3.9 \times 10^{-8}$ and $\sigma_\varepsilon^{90\%} \leq 4.7 \times 10^{-8}$. These upper limits are ~ 2 orders of magnitude less constraining than those that can be produced using the electromagnetically derived pulsar spin-down limits. However, they are the first such limits to be produced purely from gravitational wave observations.

We note that the exponential and half-Gaussian distributions used are rather simple. They were chosen as simple toy models that were easy to use due to being defined by a single hyperparameter. However, they are not necessarily physically realistic distributions. Observationally, we know that there are different populations of pulsars, like the old recycled millisecond pulsars and the young pulsars. The fact that the former are most likely to have undergone an accretion phase, which could alter the structure of their crust and magnetic field strength compared to non-recycled pulsars, meaning they could well have a different distribution. So, it could be that the two populations should be treated independently, or a more complex distribution that allows separation of the two distributions should be used (for a simple case it could be a bimodal Gaussian).

It should also be noted that in this work we assume a 20% uncertainty of the distance to all pulsars, but in reality there are range of distance uncertainties from a few percent, or hundreds of percent. In a more thorough analysis the actual measurement uncertainties for each pulsar should be included, although we do not imagine it would lead to a particularly significant change in the results.

As discussed in Sec. II C the electromagnetic-observation-derived spin-down limits could be incorporated more fully into the ellipticity distribution analysis. The simplest way to do this would be use the spin-down limits as a priors on the ellipticity (or mass quadrupole) for each pulsar. In this way, for pulsars for which the gravitational wave data alone is not particularly informative the spin-down limit-based prior would dominate.

Finally, it is worth highlighting that this type of analysis would only constrain the underlying distribution of known pulsars, but not necessarily the entire neutron star population. There could, for example, be a different distribution for accreting stars, or stars that are purely gravitars (i.e. neutron stars that are purely spinning down due to gravitational wave emission.)
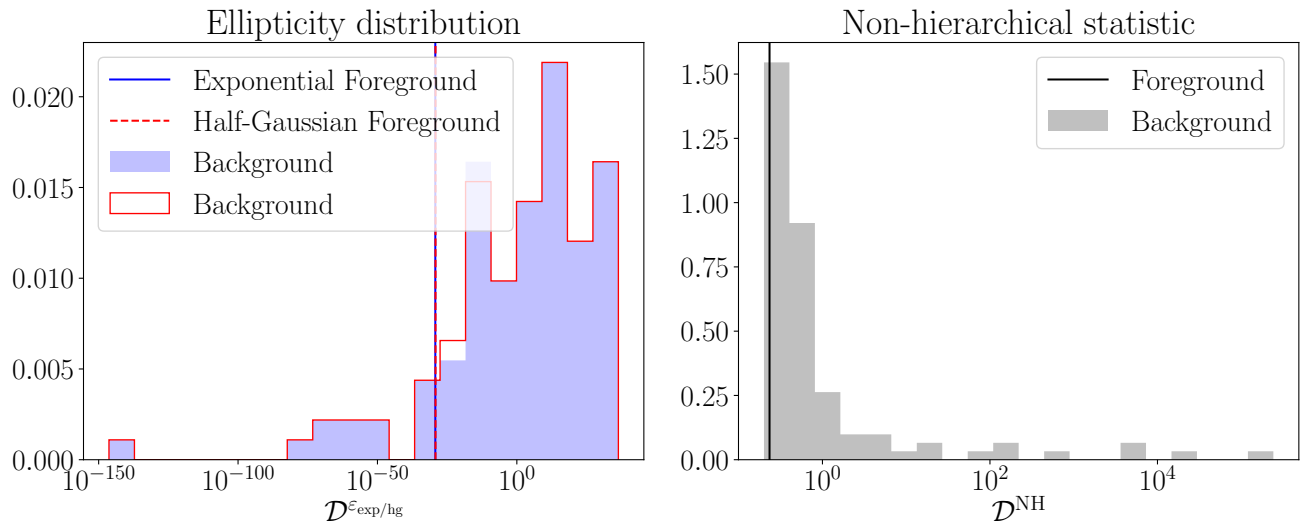
FIG. 7. The left panel shows the distributions of $\mathcal{D}^{\varepsilon_{\mathrm{exp}}}$ and $\mathcal{D}^{\varepsilon_{\mathrm{hg}}}$ for the foreground and background distributions from LIGO S6 data. The right panel shows $\mathcal{D}^{\mathrm{NH}}$ for the same data.
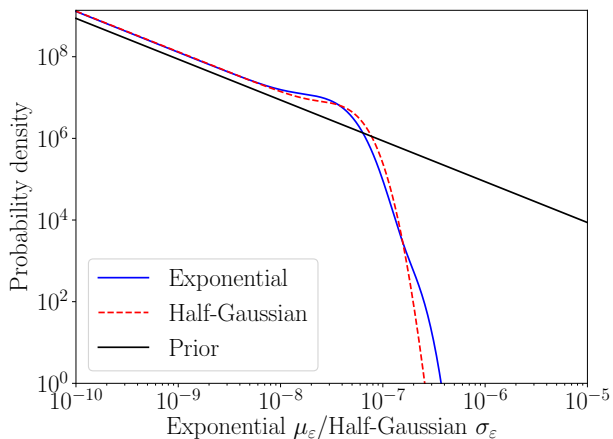

FIG. 8. The posterior probability distributions for the hyperparameters μ_ε and σ_ε defining the exponential and half-Gaussian distributions respectively.

Appendix A: h_0 posteriors

If an analysis produced posterior samples on the observed gravitational wave amplitude, h_0, rather than on ε or Q_{22} (as has been the case for previous known pulsar searches), the analysis described in this paper could still be performed, although the marginalizations over the each pulsar's distance would have to be explicitly computed. Assuming that a kernel density estimate has been used to turn sets of h_0 posterior samples into probability densities, then for each pulsar the likelihood on

ε_i (assuming a uniform prior on h_0 that is only nonzero between zero and some maximum $h_{0\max}$) is

$$p(\mathbf{x}_i|\varepsilon_i, I) = \int^{D_i} p(\mathbf{x}_i|h_0 = \frac{\kappa_i \varepsilon_i}{D_i}, D_i, I) \frac{p(\mathbf{x}_i|I)}{p(h_{0i}|I)} dD_i$$

$$= \int^{D_i} p(h_{0i} = \frac{\kappa_i \varepsilon_i}{D_i}|\mathbf{x}_i, D, I) h_{0\max} \times$$
$$p(\mathbf{x}_i|I) p(D_i|I) dD_i, \qquad (A1)$$

where

$$\kappa_i = \frac{16\pi^2 G f_{\text{rot}_i}^2}{c^4} I_{zz}. \qquad (A2)$$

Appendix B: Evidence evaluation

Here we describe the various features/issues that we have found regarding evaluating the evidences. These relate to the systematic biases on evidences from nested sampling (see, e.g., [24]), and statistical uncertainties on the final evidence due to a combination of uncertainties from the KDE posterior estimates and statistical uncertainties on the individual pulsar evidences.

1. Statistical uncertainties on evidence values

To produce our final evidence values given by Eq. (9) we rely on the output of a code that uses a stochastic sampler to perform the required integrals. We also rely on a finite number of posterior samples from each pulsar's estimate of ε to form a kernel density estimate of the true posterior. Both of these mean that even on identical initial data (with identical noise realizations) there will be some stochastic variation in the results.

To estimate these variations we have performed the processing described in Sec. III B 10 times on the same ensemble set, which in this case is one containing no pulsar signals. A different random seed is used for the sampler in each case, otherwise the results really should be identical. To estimate the variations in the final evidence caused by the finite number of samples used for the KDE of the ε posteriors, we use Eq. (16) but with $p(\mathbf{x}_i|I) = 1$ for each pulsar, so the only variation is from differences between $p(\varepsilon_i|\mathbf{x}_i, I)$ in each analysis. We find that there is a standard deviation on the base-10 logarithm of the final evidence from Eq. (9) of ~ 0.6, which comes from the variation in the estimates of the individual pulsar ε posteriors. Each KDE is estimated using ~ 1000 posterior samples, so we can check if this is roughly the variation you might expect. We can draw 1000 samples from 200 half-Gaussian distributions with known standard deviations, for each distribution produce a KDE and evaluate it at a range of points, sum the logarithms of these KDEs, and then numerically integrate it over the range of points. Doing this multiple times, but with the same set of standard deviations for each half-Gaussian, we find

the variation in the final integral is of the same order as that which we see for our analysis using simulated pulsar data. This uncertainty can be reduced by increasing the number of posterior samples used for the KDE estimate. To get more samples in our actual analysis we would need to use a greater number of nested sampling live points, which increases the run time. However, in future real analysis it may be worth doing this to cut down the uncertainties.

To see the uncertainty in the final evidence cause by both the stochastic variation in the individual pulsar evidences *and* the KDE, we repeat the above with the actual estimated $p(\mathbf{x}_i|I)$ for each pulsar. We find a standard deviation on in the base-10 logarithm of the final evidence from Eq. (9) of ~ 1.3. This suggests that the stochastic variations in the individual pulsar evidences and KDEs contribute roughly equally to the overall uncertainty. Again, this could be reduced by using a larger number of live points for the nested sampling algorithm. However, this variation is smaller than the distribution of background values, so will not be of great significance.

2. Systematic uncertainties on evidence values

When creating the odds for signals being drawn from a particular distribution compared to the noise distribution, we find that there can be a severe bias on the odds value. This comes from individually small systematic biases on the signal evidence values calculated for individual pulsars by the nested sampling algorithm (see, for example, the study on biases in nested sampling in [57]). For example, over the 200 pulsars used, if the average natural logarithm of the signal evidence was 0.05 smaller than the (unknowable) truth, then when the likelihoods for individual pulsars are combined there will be a $0.05 \times 200 = 10$ offset compared to what the value should be. In tests performed on identical realizations of a population of 200 pulsars, all containing only noise, where the only difference in the analysis is the number of live points used for the nested sampling, we find, for example, a mean difference in the pulsar's signal natural logarithm evidence values of -0.08 between using 512 live points compared to 2048. While this difference is fairly insignificant for a single pulsar, it adds up to a ~ 16 offset between the final odds values. We do not have a handle on whether using more live points produces evidences that are systematically closer to the truth (although Fig. 11 of [24] suggests that systematic errors are present even for large numbers of live points in similar cases when the true evidence can be calculated), as we do not have access to the truth for this specific case, it means that we cannot use the actual value of the odds as a trustworthy indicator of the true odds. Therefore, to gauge whether our model comparison favours one model over the other we have to resort to comparing it to a *background* distributions of odds, and see at what point our simulations diverge from the background.

We believe that these individually small systematic errors in evidences may be important for other detection statistics based on hierarchical analysis that combine evidences from multiple independent analyses, such as that proposed in [22]. This deserves more study (like that in [57]) to see how if similar features are observed in nested sampling algorithms that differ from that used here [24, 40].

[1] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso, *et al.*, "Advanced LIGO," Classical Quantum Gravity **32**, 074001 (2015), arXiv:1411.4547.

[2] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin, *et al.*, "Advanced Virgo: a second-generation interferometric gravitational wave detector," Classical Quantum Gravity **32**, 024001 (2015), arXiv:1408.3978.

[3] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger," Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837.

[4] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, *et al.*, "Binary Black Hole Mergers in the First Advanced LIGO Observing Run," Phys. Rev. X **6**, 041015 (2016), arXiv:1606.04856.

[5] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, *et al.*, "GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral," Phys. Rev. Lett. **119**, 161101 (2017), arXiv:1710.05832.

[6] B. P. Abbott, R. Abbott, F. Acernese, R. Adhikari, P. Ajith, B. Allen, G. Allen, M. Alshourbagy, R. S. Amin, S. B. Anderson, *et al.*, "Searches for Gravitational Waves from Known Pulsars with Science Run 5 LIGO Data," Astrophys. J. **713**, 671–685 (2010), arXiv:0909.3583.

[7] N. K. Johnson-McDaniel and B. J. Owen, "Maximum elastic deformations of relativistic stars," Phys. Rev. D **88**, 044004 (2013), arXiv:1208.5227.

[8] G. Ushomirsky, C. Cutler, and L. Bildsten, "Deformations of accreting neutron star crusts and gravitational wave emission," Mon. Not. R. Astron. Soc. **319**, 902–932 (2000), astro-ph/0001136.

[9] B. Abbott, R. Abbott, R. Adhikari, A. Ageev, B. Allen, R. Amin, S. B. Anderson, W. G. Anderson, M. Araya, H. Armandula, *et al.*, "Setting upper limits on the strength of periodic gravitational waves from PSR J1939+2134 using the first science data from the GEO 600 and LIGO detectors," Phys. Rev. D **69**, 082004 (2004), gr-qc/0308050.

[10] B. Abbott, R. Abbott, R. Adhikari, A. Ageev, B. Allen, R. Amin, S. B. Anderson, W. G. Anderson, M. Araya, H. Armandula, *et al.*, "Limits on Gravitational-Wave Emission from Selected Pulsars Using LIGO Data," Phys. Rev. Lett. **94**, 181103 (2005), gr-qc/0410007.

[11] B. Abbott, R. Abbott, R. Adhikari, J. Agresti, P. Ajith, B. Allen, R. Amin, S. B. Anderson, W. G. Anderson, M. Arain, *et al.*, "Upper limits on gravitational wave emission from 78 radio pulsars," Phys. Rev. D **76**, 042001 (2007), gr-qc/0702039.

[12] J. Aasi, J. Abadie, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, T. Accadia, F. Acernese, C. Adams, T. Adams, *et al.*, "Gravitational Waves from Known Pulsars: Results from the Initial Detector Era," Astrophys. J. **785**, 119 (2014), arXiv:1309.4027.

[13] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, *et al.*, "First Search for Gravitational Waves from Known Pulsars with Advanced LIGO," Astrophys. J. **839**, 12 (2017), arXiv:1701.07709.

[14] R. J. Dupuis and G. Woan, "Bayesian estimation of pulsar parameters from gravitational wave data," Phys. Rev. D **72**, 102002 (2005), gr-qc/0508096.

[15] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, *et al.*, "GW150914: The Advanced LIGO Detectors in the Era of First Discoveries," Phys. Rev. Lett. **116**, 131103 (2016), arXiv:1602.03838.

[16] R. N. Manchester, G. B. Hobbs, A. Teoh, and M. Hobbs, "The Australia Telescope National Facility Pulsar Catalogue," Astronomical Journal **129**, 1993–2006 (2005), `http://www.atnf.csiro.au/people/pulsar/psrcat/`, astro-ph/0412641.

[17] C. Cutler, "Gravitational waves from neutron stars with large toroidal B fields," Phys. Rev. D **66**, 084025 (2002).

[18] G. Woan, M. D. Pitkin, B. Haskell, D. I. Jones, and P. D. Lasky, "Evidence for a Minimum Ellipticity in Millisecond Pulsars," Astrophys. J. Lett. **863**, L40 (2018), arXiv:1806.02822.

[19] C. Cutler and B. F. Schutz, "Generalized F-statistic: Multiple detectors and multiple gravitational wave pulsars," Phys. Rev. D **72**, 063006 (2005), gr-qc/0504011.

[20] P. Jaranowski, A. Królak, and B. F. Schutz, "Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection," Phys. Rev. D **58**, 063001 (1998), gr-qc/9804014.

[21] X. Fan, Y. Chen, and C. Messenger, "Method to detect gravitational waves from an ensemble of known pulsars," Phys. Rev. D **94**, 084029 (2016), arXiv:1607.06735.

[22] Rory Smith and Eric Thrane, "Optimal Search for an Astrophysical Gravitational-Wave Background," Phys. Rev. X **8** (2018), 10.1103/PhysRevX.8.021019.

[23] M. Pitkin, C. Gill, D. I. Jones, G. Woan, and G. S. Davies, "First results and future prospects for dual-harmonic searches for gravitational waves from spinning neutron stars," Mon. Not. R. Astron. Soc. **453**, 4399–4420 (2015), arXiv:1508.00416.

[24] M. Pitkin, M. Isi, J. Veitch, and G. Woan, "A nested sampling code for targeted searches for continuous gravitational waves from pulsars," arXiv:1705.08978.

[25] J. M. Yao, R. N. Manchester, and N. Wang, "A New Electron-density Model for Estimation of Pulsar and FRB Distances," Astrophys. J. **835**, 29 (2017), arXiv:1610.09448.

[26] J. M. Cordes and T. J. W. Lazio, "NE2001.I. A New

Model for the Galactic Distribution of Free Electrons and its Fluctuations," astro-ph/0207156.

[27] D. R. Lorimer, "Binary and Millisecond Pulsars," Living Rev. Relativity **11** (2008), 10.12942/lrr-2008-8, arXiv:0811.0762.

[28] D. W. Hogg, A. D. Myers, and J. Bovy, "Inferring the Eccentricity Distribution," Astrophys. J. **725**, 2166–2175 (2010), arXiv:1008.4146.

[29] W. Del Pozzo, C. P. L. Berry, A. Ghosh, T. S. F. Haines, L. P. Singer, and A. Vecchio, "Dirichlet process Gaussian-mixture model: An application to localizing coalescing binary neutron stars with gravitational-wave observations," Mon. Not. R. Astron. Soc. **479**, 601–614 (2018), arXiv:1801.08009.

[30] J. Skilling, "Nested sampling for general bayesian computation," Bayesian Anal. **1**, 833–860 (2006).

[31] I. S. Shklovskii, "Possible Causes of the Secular Increase in Pulsar Periods." Sov. Astron. **13**, 562 (1970).

[32] T. Damour and J. H. Taylor, "On the orbital period change of the binary pulsar PSR 1913 + 16," Astrophys. J. **366**, 501–511 (1991).

[33] D. Keitel, R. Prix, M. A. Papa, P. Leaci, and M. Siddiqi, "Search for continuous gravitational waves: Improving robustness versus instrumental artifacts," Phys. Rev. D **89**, 064023 (2014), arXiv:1311.5738.

[34] M. Isi, M. Pitkin, and A. J. Weinstein, "Probing dynamical gravity with the polarization of continuous gravitational waves," Phys. Rev. D **96**, 042001 (2017), arXiv:1703.07530.

[35] R. Prix and B. Krishnan, "Targeted search for continuous gravitational waves: Bayesian versus maximum-likelihood statistics," Classical Quantum Gravity **26**, 204013 (2009), arXiv:0907.2569.

[36] LIGO Scientific Collaboration, *Advanced LIGO anticipated sensitivity curves*, Tech. Rep. LIGO-T0900288-v3 (2010) https://dcc.ligo.org/LIGO-T0900288/public.

[37] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, *et al.*, "Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo," Living Rev. Relativity **19**, 1 (2016).

[38] M. Pitkin, "Prospects of observing continuous gravitational waves from known pulsars," Mon. Not. R. Astron. Soc. **415**, 1849–1863 (2011), arXiv:1103.5867.

[39] J. Veitch and A. Vecchio, "Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network," Phys. Rev. D **81**, 062003 (2010), arXiv:0911.3820.

[40] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, *et al.*, "Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LAL-Inference software library," Phys. Rev. D **91**, 042003 (2015), arXiv:1409.7215.

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, *et al.*, "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res. **12**, 2825 (2011), arXiv:1201.0490.

[42] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley Series in Probability and Statistics (Wiley, New York, 1992).

[43] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, *et al.*, "First narrow-band search for continuous gravitational waves from known pulsars in advanced detector data," Phys. Rev. D **96**, 122006 (2017), arXiv:1710.02327.

[44] B. P. Abbott, R. Abbott, R. Adhikari, P. Ajith, B. Allen, G. Allen, R. S. Amin, S. B. Anderson, W. G. Anderson, M. A. Arain, *et al.*, "LIGO: the Laser Interferometer Gravitational-Wave Observatory," Rep. Prog. Phys. **72**, 076901 (2009), arXiv:0711.3041.

[45] The LIGO Scientific Collaboration and The Virgo Collaboration, "Sensitivity Achieved by the LIGO and Virgo Gravitational Wave Detectors during LIGO's Sixth and Virgo's Second and Third Science Runs," arXiv:1203.2674.

[46] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, "The LIGO Open Science Center," J. Phys. Conf. Ser. **610**, 012021 (2015), arXiv:1410.4839.

[47] G. S. Davies, M. Pitkin, and G. Woan, "A targeted spectral interpolation algorithm for the detection of continuous gravitational waves," Classical Quantum Gravity **34**, 015010 (2017), arXiv:1603.00412.

[48] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, *et al.*, "Searches for Continuous Gravitational Waves from Nine Young Supernova Remnants," Astrophys. J. **813**, 39 (2015), arXiv:1412.5942.

[49] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, *et al.*, "Results of the deepest all-sky survey for continuous gravitational waves on LIGO S6 data running on the Einstein@Home volunteer distributed computing project," Phys. Rev. D **94**, 102002 (2016), arXiv:1606.09619.

[50] M. Isi, S. Mastrogiovanni, M. Pitkin, and O. J. Piccinni, (to be published).

[51] LIGO Scientific Collaboration, "LALSuite," https://wiki.ligo.org/DASWG/LALSuite (2017).

[52] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D.S. Seljebotn, and K. Smith, "Cython: The best of both worlds," Comput. Sci. Eng. **13**, 31 –39 (2011), http://cython.org.

[53] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, *et al.*, "Jupyter notebooks—a publishing format for reproducible computational workflows," in *Positioning and Power in Academic Publishing: Players, Agents and Agendas: Proceedings of the 20th International Conference on Electronic Publishing* (IOS Press, Amsterdam, 2016) p. 87, http://jupyter.org/.

[54] J. D. Hunter, "Matplotlib: A 2d graphics environment," Computing In Science & Engineering **9**, 90–95 (2007).

[55] M. Droettboom, T. A. Caswell, J. Hunter, *et al.*, "matplotlib/matplotlib: v2.0.0," (2017), 10.5281/zenodo.248351.

[56] M. Pitkin, "psrqpy: a python interface for querying the ATNF pulsar catalogue," JOSS **3**, 538 (2018), arXiv:1806.07809.

[57] J. Buchner, "A statistical test for Nested Sampling algorithms," J. Stat. Comput. **26**, 383 (2016), arXiv:1407.5459.