PURPOSE-LED PUBLISHING™

PAPER · OPEN ACCESS

A sensitivity curve approach to tuning a pulsar timing array in the detection era

To cite this article: Jeremy G Baier *et al* 2025 *Class. Quantum Grav.* **42** 075008

View the article online for updates and enhancements.

A sensitivity curve approach to tuning a pulsar timing array in the detection era

Jeremy G Baier[1,*] , **Jeffrey S Hazboun**[1]
and Joseph D Romano[2]

[1] Oregon State University, 1500 SW Jefferson Ave, Corvallis, OR 97331, United States of America
[2] Department of Physics and Astronomy, University of Texas Rio Grande Valley, One West University Boulevard, Brownsville, TX 78520, United States of America

E-mail: baierj@oregonstate.edu

Abstract

As pulsar timing arrays (PTAs) transition into the detection era of the stochastic gravitational wave background (GWB), it is important for PTA collaborations to review and possibly revise their observing campaigns. The detection of a 'single source' would be a boon for gravitational astrophysics, as such a source would emit gravitational waves for millions of years in the PTA frequency band. Here we present generic methods for studying the effects of various observational strategies, taking advantage of detector sensitivity curves, i.e. noise-averaged, frequency-domain detection statistics. The statistical basis for these methods is presented along with myriad examples of how to tune a detector towards single, deterministic signals or a stochastic background. We demonstrate that trading observations of the worst pulsars for high cadence campaigns on the best pulsars increases sensitivity to single sources at high frequencies while hedging losses in GWB and single source sensitivity at low frequencies. We also find that sky-targeted observing campaigns yield minimal sensitivity improvements compared with other PTA tuning options. Lastly, we show the importance of the uncorrelated half of the GWB, i.e. the pulsar-term,

* Author to whom any correspondence should be addressed.

as an increasingly prominent sources of noise and show the impact of this emerging noise source on various PTA configurations.

1. Introduction

Pulsar timing arrays (PTAs) around the world recently published [1–4] compelling evidence for Hellings–Downs correlations [5] for a stochastic gravitational wave background (GWB), ushering PTAs into their detection era. This exciting new era brings the expectation of more types of gravitational waves (GWs) to be detected in the near future, including but not limited to resolvable binary supermassive black holes [6] and GW bursts [7] as well as other GW features such as anisotropies in the background [8–10] and GW memory [11]. With a tentative background in hand, it is necessary to incorporate this new information—which may soon be regarded as a foreground with respect to other sources—and reassess sensitivity to those same sources. This work investigates PTA sensitivity to single sources of GWs as individual binaries emerge above the background. Specifically, we answer the question of how a PTA can optimize its sensitivity to these single sources in the presence of a GWB.

1.1. PTAs and pulsar observing campaigns

PTAs operate by regularly timing a set of galactic millisecond pulsars (MSPs) [12]. Deviations between the anticipated arrival time and measured arrival time of pulses of light can be correlated across these pulsars to uncover GW signatures in the data [5]. A number of collaborations [13–15] and an international consortium of those collaborations [16] have been monitoring more than 100 MSPs since the early 2000s in an effort to detect these GWs. Their early efforts mostly established persistent timing campaigns of the lowest residual RMS error pulsars over many years, along with stable telescope equipment that minimized systematic noise.

Once these observing campaigns were well established, the community was able to consider optimizing observations of the galactic population of MSPs in order to detect a particular GW signal. A variety of factors play a role in this type of optimization including which pulsars to observe, how frequently to observe them, and which electromagnetic frequencies to observe them with [17, 18]. Limitations on telescope time and funding constrain PTA collaborations in their optimizations of these pulsar observing campaigns. The two most likely types of sources of GWs sought in the nanohertz band are a stochastic GWB and continuous waves (CWs) from single sources [19]. Siemens *et al* [20] derives scaling laws for PTAs and finds that for a GWB the dominant factor in signal-to-noise (S/N) growth is the number of pulsars observed as this yields the most pairs of pulsars for Hellings–Downs correlations. On the other hand, various groups [21–23] found that an optimal campaign for CWs involves allocating more observing time into one's 'best' pulsars (in this case the ones with the smallest residual RMS errors) – whether that is longer observations or more frequent observations. In a similar vein, [24] finds that the majority of a CW signal comes from just a subset of pulsars.

Most of these optimizations are analytic or semi-analytic in nature, approximating what the most important aspects of the PTA likelihood will be over time. Many other studies, like [24], carry out full simulations of PTAs with injected GW signals to understand the sensitivity of PTAs to various sources [6, 19, 25–27] and their ability to estimate GW parameters accurately [23, 28, 29]. Here we use frequency-domain detector characterization and sensitivity curves

built from the noise-averaged expectation value of detection statistics to study not just the time-to-detection of various sources, but also myriad tunable parameters of the observational campaigns that may change the time to detection. These methods mix the best of both worlds, speeding up calculations by using only the expectation value of many noise realizations of a PTA configuration, but also using the frequentist statistics at the heart of many PTA GW searches. All the techniques mentioned above are valuable in their own right; possessing pros and cons for particular applications. Tuning a PTA is the work of sensitivity curves.

1.2. PTA detector characterization

Instrument sensitivity is commonly employed to quantify and explore a detector's capacity to differentiate signals of interest from sources of noise. PTAs can be well characterized as detectors by quantifying noise and timing model effects in the pulse arrival times [30–32].

The software package `hasasia` [33] was developed to robustly and efficiently characterize PTAs as detectors. In this paper we introduce additional functionality to `hasasia` in order to expand upon techniques that leverage detector characterization as an efficacious tool in both forecasting and optimizing sensitivity to single sources of GWs.

The outline of the paper is as follows. In section 2, we compare existing detection statistics for continuous GWs emitted by supermassive binary black holes (SMBBHs), elaborating on a polarization, inclination, and phase-marginalized statistic. In section 3, we describe our model for a PTA similar in characteristics to the international PTA (IPTA) [16]. In section 4, we explore the detection capabilities of a simulated PTA by comparing its sensitivity to simulated SMBBH populations. In section 5, we demonstrate techniques to optimize pulsar observing campaigns for single-source detection and explore how the GWB and a transdimensional simultaneous fit of the GWB and CWs affect single-source sensitivity. Lastly, we conclude in section 6 with remarks on how these techniques could be used with a reduced PTA dataset to optimize pulsar timing campaigns for single-source detections and multi-messenger detections. Appendix A contains a derivation of a marginalized $\mathcal{F}$ statistic that is used throughout the work.

2. Detection statistics

In order to forecast the ability of PTAs to detect various GW sources, we use sensitivity curves [30–33] for pulsars based on their noise properties and timing model transmission function to determine the spectra for individual pulsars. These are then combined using two GW frequentist statistics, written in the Fourier domain, to characterize the PTA as a GW detector. Specifically, if one writes down a signal-to-noise ratio, often derived from a maximized likelihood ratio, the denominator, i.e. the 'noise' as a function of frequency, is taken to be the sensitivity. One such statistic, known as the optimal cross-correlation statistic (OS) [34, 35], is used for assessing sensitivity to an isotropic GWB. The square of the expected signal-to-noise ratio of the optimal statistic is

$$\rho_{\text{OS}}^2 = 2T_{\text{obs}} \int_0^{f_{\text{Nyq}}} \mathrm{d}f > \frac{S_h^2(f)}{S_{\text{eff}}^2(f)} \tag{1}$$

where S_h is the strain power spectral density of the GWB, T_{obs} is the time span of the full PTA dataset and $S_{\text{eff}}(f)$ is the *effective* strain-noise power spectral density for the whole PTA

$$S_{\text{eff}}(f) = \left(\sum_I \sum_{J>I} \frac{T_{IJ}}{T_{\text{obs}}} \frac{\chi_{IJ}^2}{S_I(f) S_J(f)} \right)^{-1/2}. \tag{2}$$

Note that $S_{\mathrm{eff}}(f)$ can be interpreted as the sensitivity of the PTA to a stochastic GWB. The above expansion includes contributions from the Hellings–Downs factors χ_{IJ}, the overlapping timespan between two pulsar's datasets, T_{IJ}, and the individual pulsar strain-noise power spectral densities $S_I(f)$. Note that $S_{\mathrm{eff}}(f)$ has dimensions of strain^2 Hz-1. See [32] for more details.

The other statistic we use is a matched filter statistic for resolvable binary supermassive black holes [32, 36]. Binary SMBHs are the leading candidate for the source of the GWB seen by PTA collaborations [37], and is the main focus of this work. As we will demonstrate, tuning a PTA requires not only making it as sensitive as possible to these individual sources, but also maintaining sensitivity to the cross correlations in the GWB. Pulsar observing campaigns can be fine-tuned to maximize sensitivity to the GWB, individual sources, or optimized for both.

There are a few frequentist statistics in the literature developed for single sources in PTAs. See for example [32, 36, 38–40]. In appendix A we review the relationships between these and the matched filter signal-to-noise ratio used for sensitivity curves in this work. The signal-to-noise ratio for a circular, non-evolving binary black hole source can be written as

$$\rho_{\mathrm{MF}}^2\left(h_0, \iota, \psi, \hat{k}\right) = 2T_{\mathrm{obs}} \int_0^{f_{\mathrm{Nyq}}} \mathrm{d}f > \frac{S_h\left(f, h_0\right)}{S_{\mathrm{eff}}\left(f, \hat{k}, \iota, \psi\right)}, \tag{3}$$

where

$$S_{\mathrm{eff}}\left(f, \hat{k}, \iota, \psi\right) \equiv \left(\frac{4}{5} \sum_I \frac{T_I}{T_{\mathrm{obs}}} \frac{1}{S_I\left(f, \hat{k}, \iota, \psi\right)}\right)^{-1}. \tag{4}$$

The individual pulsar strain-noise power spectral densities are functions of GW frequency f, source direction $-\hat{k}$, inclination angle ι, and polarization angle ψ. This sensitivity curve formalism and implementation is robust and has been corroborated with Bayesian methods in [32] and [41].

Of particular interest for these frequentist statistics is the ability to quote false alarm probabilities given a null distribution and detection probabilities (DPs) given a distribution of the noise with a signal in it. The (null) distribution of the optimal statistic follows a generalized χ^2 distribution [42] and can be calculated for a given dataset given a set of noise parameters. This work focuses on single-source detections with the assumption that we are far enough in the future to assume a significant detection of the GWB.

In appendix A we introduce the $\mathcal{F}_e$ statistic precisely because it has a well-defined false alarm probability/DP and can be related to ρ_{MF}^2 in equation (3). For our purposes, the main result of appendix A is demonstrated in figure A1 where we motivate the use of an ι, ψ and ϕ_0-averaged DP, $\langle \mathrm{DP}_{\mathcal{F}_e}(\rho(\iota, \psi)) \rangle_{\iota, \psi, \phi_0}$, based on the fact that 2 times the $\mathcal{F}_e$ statistic follows a noncentral χ^2 distribution [38, 39]. Therefore, we can calculate the DP given a particular value of $\rho_{\mathrm{MF}}^2(h_0, \iota, \psi, \hat{k})$. As we will see in section 4, this allows us to use our sensitivity curves to efficiently carry out the analyses like those in [6, 19, 43].

3. PTA sensitivity

3.1. Simulating an IPTA-like detector

To demonstrate our methods, we simulate a detector which is similar in its number of pulsars and time span to the current IPTA [1] dataset and similar in its noise characteristics to the NANOGrav 15 year data set [44].
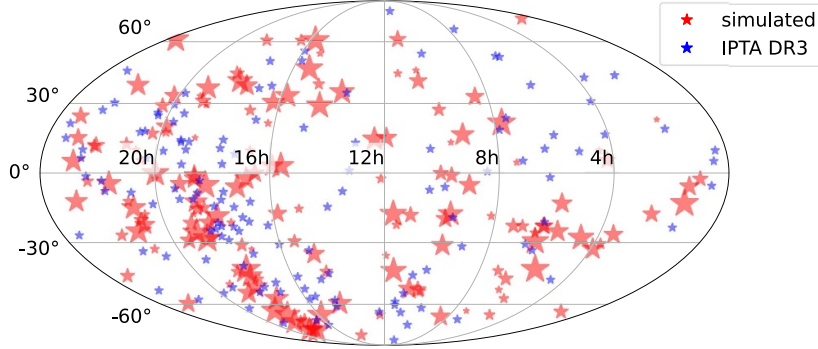
Figure 1. Pulsar sky locations. IPTA DR3 pulsars are marked by a blue star and pulsars in our simulated dataset are marked by a red star. The red stars are sized according to their individual pulsar strain-noise power spectral densities, $S_I(f)$ (see equation (2)), where a larger star corresponds to a larger value of $S_I(f)$.

Using the approximate sky locations of the pulsars in [1], we create empirical distributions of pulsar ecliptic longitude and latitude from which we initially draw 115 pulsars, which are shown in figure 1.

Similarly, we construct an empirical distribution of the logarithm of pulsar white noise reported in [44], using the 'whitened' values in cases where red noise is found to be significant. Next, we create a 2–dimensional empirical distribution of the amplitudes and spectral indices of the power-law model of achromatic red noise used in [41],

$$\varphi_i \sim \frac{A^2}{12\pi^2 T_{\text{span}}} \left(\frac{f_i}{f_{\text{yr}}}\right)^{-\gamma} \text{yr}^3. \tag{5}$$

We construct this distribution from the subset of pulsars whose intrinsic red noise remained significant in a common red noise search. (See the bolded values in table 2 of [41].) This allows us to later inject a common red noise process without double counting red noise in the pulsars. Since only $\sim 18\%$ of the pulsars in the NANOGrav 15 year dataset maintain significant levels of red noise after removing the common red noise signal, we only inject intrinsic red noise in a proportional number of pulsars in our simulated dataset. Lastly, we construct an empirical distribution of pulsar time spans from [44] to model historical PTA growth.

From the constructed empirical distributions for pulsar sky coordinates, white noise, red noise, and time span, we simulate a PTA of 115 pulsars with a total time span of 16 years and an observing cadence of 20 observations per year for all pulsars. We simulate these data with the `hasasia` package and proceed to inject into all pulsars a power-law GWB, which we construct from the median of the joint posterior from [1]: $\log_{10} A_{\text{GWB}} = -14.29$ and $\gamma_{\text{GWB}} = 3.44$. From here, we adjust the overall sensitivity of our detector to mirror that of contemporary PTAs by adjusting the overall power in the white noise of the detector, given by,

$$P_{\text{WN}} = 2\sigma^2/c, \tag{6}$$

where c is the cadence of observations and σ is the uncertainty on pulse arrival time. By shifting our white noise distribution for all the pulsars, we adjust the overall sensitivity until the injected GWB has a S/N of 7 at 16 years to match preliminary projects for IPTA Data Release 3 (DR3) [1]. In other words, we are calibrating our PTA to have a sensitivity at 16 years, which is comparable to the projections of DR3.
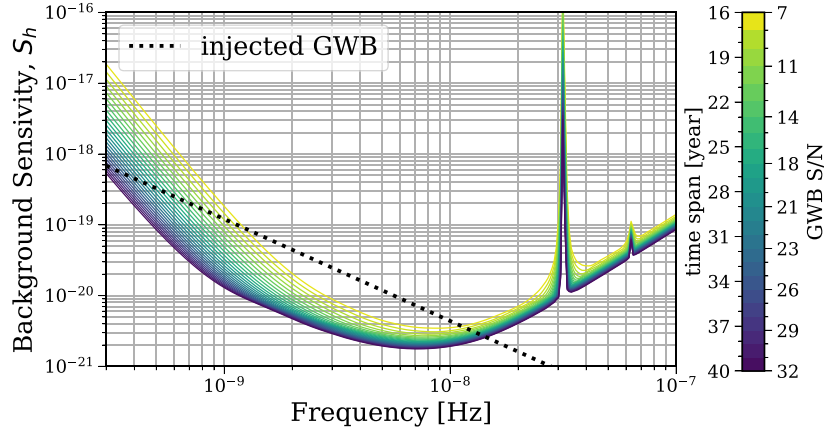
Figure 2. Standard GWB sensitivity. The background sensitivity curves of the simulated PTA are plotted for every year between 16 and 40 years. The color bar encodes both the year and the GWB S/N at that year. The injected GWB is shown as a dashed black line.

From here, we simulate our detector out to a total time span of 40 years. We assume that the white noise in our detector does not change over the course of the simulations (i.e. the radio telescopes or other noise mitigation techniques do not improve), and we add 2 pulsars per year whose sky locations and noise properties are drawn from the same empirical distributions used above. As is standard, we only include pulsars in our dataset or a time slice of the data set if the pulsar has at least 3 years of data. Generally, PTAs do this so that a timing model may be accurately fit. So at 40 years, we have a detector with 156 pulsars and a GWB S/N of $\sim$32. The time evolution of the GWB sensitivity is shown in figure 2, and the time evolution of the sky-averaged single-source characteristic strain sensitivity is shown in figure 3. Note that the main purpose of this PTA is to be a high fidelity test bed in order to demonstrate the various techniques developed for helping to inform observing era strategies for pulsar timing. A large scale effort is underway to forecast the sensitivity of PTAs using next generation facilities.

4. Single SMBBH detection prospects from a GWB

One way we can assess the single source detection capacity of our simulated detector is by computing the detection statistics from [19] (see section 4.2 for a brief review) over a simulated population of SMBBHs. By looking at many realizations of these populations, we can make statistical statements about how many binaries we detect, when we detect them, and at which frequencies we detect them.

4.1. SMBBH population simulations

In this study, we use 1000 realizations of a SMBBH population simulated with `holodeck` [45], a software package designed for astrophysical inference of black hole population models using PTA data. See [37, 43] for a description. As demonstrated in [6, 43] populations of this type vary widely between choices of the black hole-population parameter space and also from realization to realization. Therefore, we use ten random samples from the same phenomenological
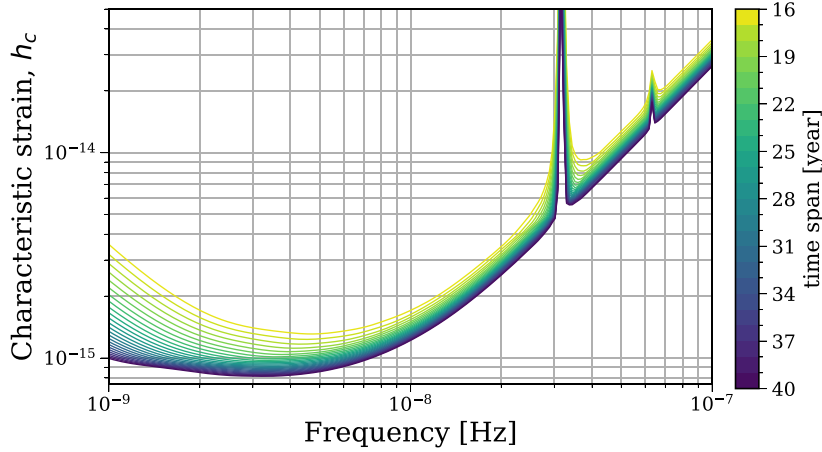
Figure 3. Standard single source sensitivity. The single-source sensitivity curves of the simulated PTA are plotted for every year between 16 and 40 years. The color bar encodes the dataset timespan for which the sensitivity curve is calculated.

parameter space in [37], but for each sample of the astrophysical black hole-population parameters, we generate 100 population realizations which we then adjust by 'tilting' the median background to match the 3P+ spectral index, $\gamma_{\mathrm{GWB}} = 3.44$. In order to minimize distractions in this methods paper from the huge diversity in these binary populations, which are **not** conditioned on any PTA data, we choose a set of realizations that gave us a 'middle of the road' total DP (TDP) (see section 4.2 below) median of roughly $\sim$50% at 30 years. This population is merely a control from which to measure *changes* in the detection statistics throughout the various optimizations we discuss in this work. Each realization contains binaries in 40 frequency bins evenly spaced from $\sim$2 nHz to $\sim$80 nHz. We consider detection statistics on the loudest (i.e. the largest strain amplitude) SMBBH in each frequency bin. Figure 4 shows the characteristic strain (h_c) of the loudest binary in each frequency bin for all 1000 realizations considered.

Recall that in the course of simulating our detector, we injected a power-law GWB into all the pulsars which is manifested in the detector sensitivity as a GWB self-noise. For this study, we elect to not recompute detector sensitivity for specific GWB realizations for the following reasons. As noted above, the SMBBH populations are not well constrained yet by GWB measurements so for this study, we have modified the realizations to ensure that all the population realizations are broadly consistent with the injected background, which means that detector sensitivity does not change very dramatically with different realizations. By leaving the GWB self-noise fixed, we do not need to recompute detector sensitivity for each realization of the background, providing substantial computational savings. We leave an in-depth investigation of the effects of power-law deviations on detector sensitivity as future work after more computational savings are implemented into `hasasia`. We do, however, explore the effects of several substantially different backgrounds on detector sensitivity and single-source detection in section 5.5.

We focus on how one can harness sensitivity curves as an efficient alternative to costly simulations, allowing us to easily adjust the signal(s) and noise in our detector to optimize the myriad knobs for single-source detections. In future work, we will use `holodeck` to generate many SMBBH populations which have been conditioned on the free spectral posteriors from
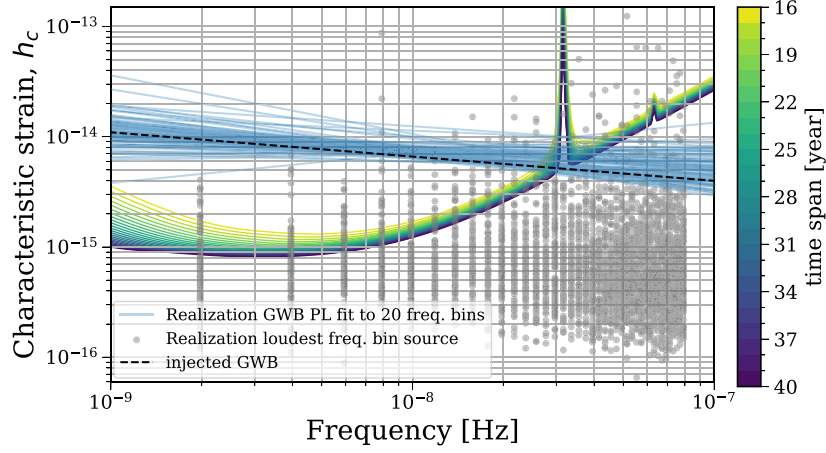
Figure 4. Simulated binary population sourcing the GWB. The loudest strain in each frequency bin for 100 of the 1000 total SMBBH population realizations compared to a time-evolving detector sensitivity. As the data set time span increases, the binaries grow in S/N, some eventually reaching a detection threshold.

[2]. This more in-depth treatment should yield more robust time-to-detection predictions. In this work, we present the methods developed for such studies.

4.2. Application of detection statistics to population realizations

There are a number of ways one could apply the detection statistics from [19] for these SMBBH population realizations using the matched-filter (squared) S/N ρ_{MF}^2 described in section 2. We consider two useful approaches: a per–frequency bin DP as well as a TDP. The per–frequency bin S/N calculates the S/N described in section 2 and appendix A for the loudest source in each frequency bin across all 1000 population realizations. A per–frequency DP, DP_i, is calculated using the probability density function of the $\mathcal{F}_e$ statistic given a value for the S/N (ρ), see for example equation (32) from [19] and the discussion in appendix A surrounding equation (A.18). The TDP is then given by (equation (33) in [19]),

$$\mathrm{TDP} = 1 - \prod_i \left[1 - \mathrm{DP}_i\right], \tag{7}$$

where i indexes frequency bins. The TDP can be interpreted as the probability of detecting at least one binary in *any* frequency bin. For a '3σ' detection, we use the false alarm rate to be 0.00135, which yields $\bar{\mathcal{F}}_e \sim 8.89$, requiring a S/N of ~ 3.7 under this detection statistic's null distribution which is a non-central χ^2 distribution with 4 degrees of freedom and a non-centrality parameter of ρ^2. For the entirety of this work, we use this detection threshold of $\bar{\mathcal{F}}_e \sim 8.89$. We apply these detection statistics to the simulated population realizations and the simulated PTA described in section 3. Firstly, we compute the TDP for each realization of the Universe every year from 16 to 40 years in figure 5. Again, this probability can be thought of as the probability of making a detection of *at least* one single source in *any* of the frequency bins. We note that this is asymptotic to $\mathrm{TDP}|_{\mathrm{DP}_i} = \mathrm{FAP} \sim 0.05$ at low probabilities since for a given non-zero false alarm probability, there is a non-zero chance of falsely claiming a detection
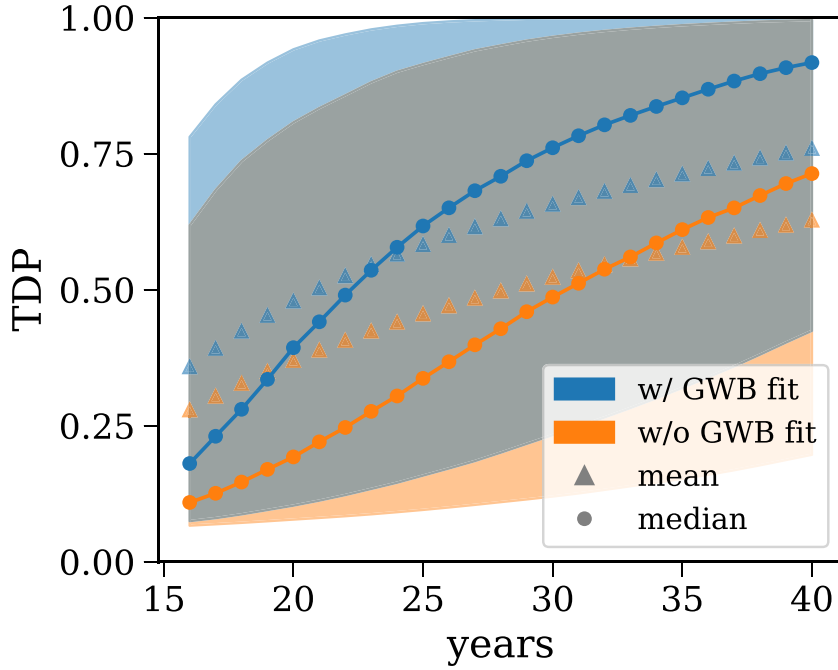
Figure 5. Total detection probability of the standard observing campaign. Both the mean (triangles) and the median (dots) TDP across realizations are shown at each year from years 16 to 40 for the simulated detector. They are calculated both with (blue) and without (orange) a GWB Earth-term simultaneous fit. (See section 5.4 for details on a simultaneous fit.) The shaded regions enclose the middle 68% of the TDP distribution.

in each frequency bin. Figure 5 shows the change in the TDP over the dataset timespan, by tracking the evolution of the mean, median, and the central 68% region.

Figure 6 shows how the per–frequency DP_i changes in time by plotting the percent change in the median DP_i from a dataset timespan of 16 years to 20, 30, and 40 years. Here and throughout the work we define the percent change in a quantity as

$$100 \times \left(\frac{\text{New} - \text{Original}}{\text{Original}} \right). \tag{8}$$

Figure 6 reveals how low frequency binaries tend to move to higher DPs more quickly than high frequency binaries for our simulated populations and PTA, which is expected since the sensitivity at low frequencies is primarily a function of the timespan of the data set.

We return to this methodology of comparing sensitivity to simulated populations in the next section, where we optimize our pulsar observing campaigns to single-source detections.

5. Tuning the detector

A number of substantial changes can be made to pulsar observing campaigns, i.e. observing some pulsars more or less frequently, or with different instruments, or not observing them at all, with the goal of optimizing a PTA's dataset (and thus sensitivity) to a specific source. By using sensitivity curves and frequency-domain detection statistics, we can explore how these changes in a campaign will affect sensitivity without having to re-simulate entire datasets. This
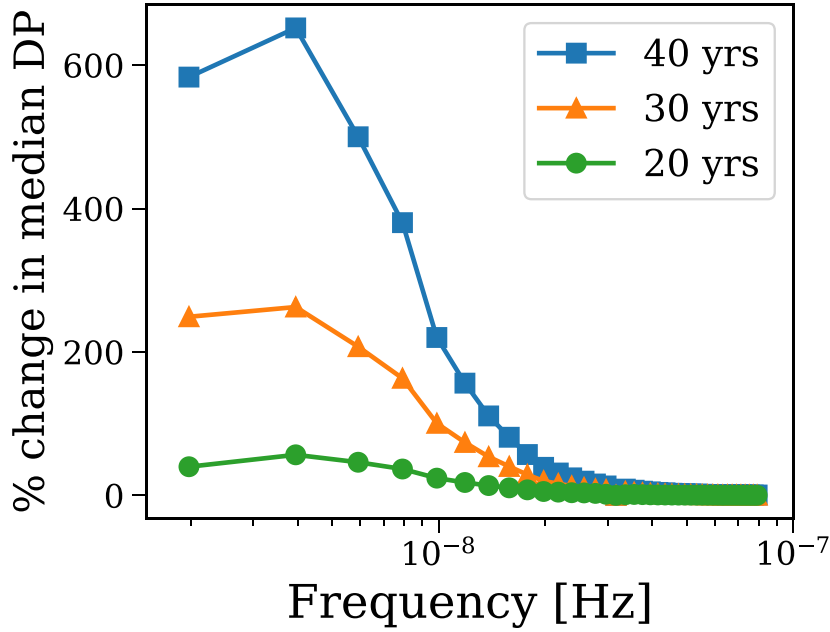
Figure 6. The percent change in median detection probability. The %ΔDP is shown from year 16 to years 20, 30, and 40 is plotted for the 40 frequency bins considered in this study.

efficient approach, paired with the detection statistics described in sections 2 and 4, enables us to robustly quantify our sensitivity to the GWB and the detectability of single sources under a variety of campaigns and noise assumptions. In this section, we highlight our ability to answer the question of how changes to pulsar observing campaigns impact single-source sensitivity. For the rest of this section, we compare the standard, status quo, campaign (see figure 3) to 'alternate' campaigns at 30TDPyears assuming that these changes were initiated at 20TDPyears, putting the alternate campaigns in effect for a total of 10TDPyears, which is 1/3 of the considered data set time span. We settle on an optimal campaign for single-source detections under realistic observing constraints and proceed to use the techniques from section 4 to assess how detection prospects improve under this optimal campaign. We proceed to explore other topics beyond simply single source sensitivity, such as a simultaneous fit of the GWB alongside single sources, the dependence of sensitivity on the GWB spectrum itself, multi-messenger detection prospects, and sensitivity to new-physics models.

5.1. High cadence campaigns

To start, we simulate a hypothetical observing campaign in which we add high cadence campaigns on our 'most sensitive' N pulsars. These high cadence campaigns consist of timing these pulsars four times as frequently from year 20 to year 30. For the rest of the work, we define the 'most sensitive' pulsars as those which have the smallest residual root-mean-square (RMS) error. More frequent observations of these pulsars increase the sensitivity at high frequencies, where white noise dominates in PTAs. High cadence campaigns on ten pulsars constitutes a 30% increase in telescope time beyond the assumed yearly increases in telescope time to add two pulsars per year. As expected, figure 7 shows large sensitivity gains at high frequencies but
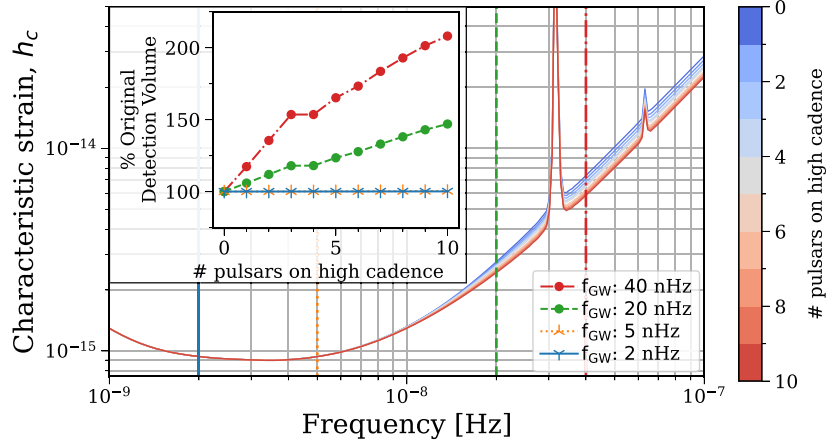
Figure 7. Adding observations. Sky-marginalized single-source sensitivity curves are plotted for a variable number of high cadence campaigns, which are encoded in the color bar. The vertical lines denote frequencies at which we compute the change in detector volume with respect to the standard campaign. The change in detector volume as a function of the number of pulsars on high cadence campaigns is shown in the inset plot.

leaves the sensitivity relatively unaffected at low frequencies, where sensitivity is primarily a function of the timespan of the dataset. The GWB S/N increases $<5\%$ with all 12 pulsars on high cadence campaigns. Another astrophysically useful way to quantify the sensitivity in our detector is by computing its detection volume (DV). We define the DV at a given frequency as

$$\mathrm{DV}(\mathrm{f}) = \frac{1}{3} \iint \mathrm{d}r\, \mathrm{d}\Omega\, r^2, \tag{9}$$

where Ω is the solid angle on the sky in steradians and r is the maximum luminosity distance at which the detector is capable of making a 3σ detection to a binary with a chirp mass of $10^9\mathrm{M}_\odot$. Note that our choice in chirp mass and significance threshold are both inconsequential as we quote the percent original DV, which we define as

$$\%\,\mathrm{Original\,DV} = 100 \times (\mathrm{New\,DV}/\mathrm{Original\,DV}). \tag{10}$$

The inset plot in figure 7 shows that the detector volume remains essentially the same at low frequency and nearly doubles at high frequency, which follows our gains in sensitivity. The hypothetical observing campaign examined in this section provides a useful baseline for our more realistic fixed observing-time campaigns.

5.2. Comparison of fixed observing-time campaigns

It is more realistic to consider a scenario where a PTA has a fixed amount of observing time and can only optimize its sensitivity to single sources by reallocating that time amongst pulsars and observations. For this study, we assume a fixed integration time for each observation (i.e. each observation is the same duration) but vary the cadence of observations, which we define as the number of observations of a pulsar per year.
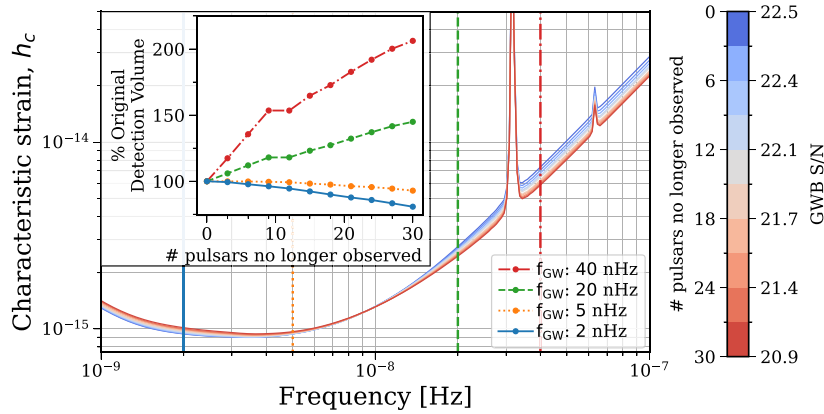
Figure 8. Dropping pulsars. Sky-marginalized single-source sensitivity curves are plotted for a variable number of pulsars dropped at 20 years to free telescope time for high cadence campaigns on the most sensitive pulsars. The color bar encodes the number of pulsars dropped (in sets of 3 pulsars) as well as the GWB S/N. The inset plot tracks the change in detector volume at 4 different reference frequencies as a function of changes to the observing campaign.

First, we consider a campaign which stops observing the least sensitive $3N$ pulsars at 20 years and instead observes the most sensitive N pulsars with a high cadence campaign of quadrupled observations, ensuring that telescope time remains fixed. We test various metrics for 'least sensitive' pulsar, including largest residual RMS error, but we ultimately use the pulsar's contribution to the GWB. Accordingly, we drop pulsars (in sets of 3) which contribute the least to the GWB as calculated by leaving out pulsars one-by-one and recomputing the GWB S/N. Again, we define the most sensitive pulsars as those with the smallest residual RMS error. The single-source sensitivity curves in figure 8 illustrate how discontinuing observations on the least sensitive pulsars and quadrupling the cadence of the most sensitive pulsars in a fixed-observing timing scenario yields sensitivity gains at high frequencies and sensitivity losses at low frequencies. This sensitivity loss is an important conclusion of this work. The sensitivity of PTAs to single sources at low frequencies is critically dependent on the ability of the detector to mitigate the effects of the GWB, which we explore further in section 5.5. The change in detector volume is nearly identical to the previous campaign (figure 7) at high frequencies, but we note a decrease of up to 10% in detector volume at low frequencies due to the discontinuation of observations on pulsars which were contributing to the GWB mitigation.

We now consider a campaign which puts the least sensitive $6N$–pulsars on half-cadence while keeping the most sensitive N–pulsars on high cadence campaigns, which again keeps the total telescope time fixed. This strategy allows for the continued contribution of the least sensitive pulsars at low frequencies but still frees up telescope time for high cadence campaigns on the most sensitive pulsars. Figure 9 reveals how this campaign, compared to the previous campaign (figure 8), results in a higher GWB S/N at 30 years, less sensitivity losses at low frequencies, and roughly the same sensitivity gains at higher frequencies. Accordingly, the detector volume losses for low frequencies are reduced from 10% to $<5\%$ while the increases in detector volume are maintained. In this way, we are able to preserve the gains of the previous campaign and minimize the losses at low frequencies, all while still keeping telescope
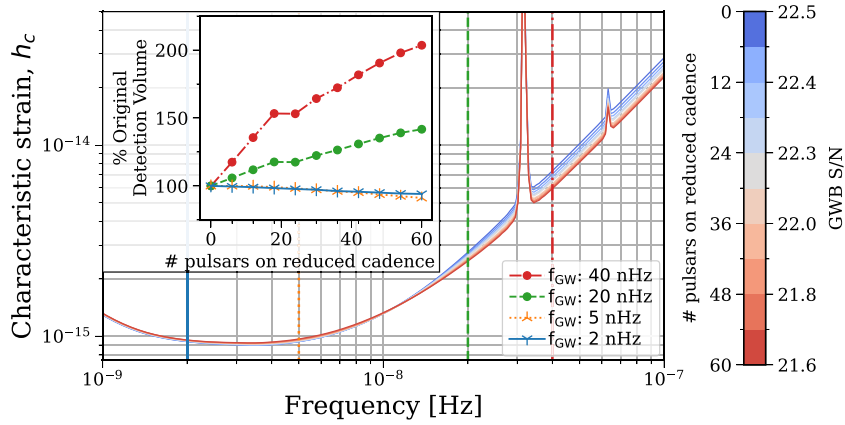
Figure 9. Shifting observations. Sky-marginalized single-source sensitivity curves are plotted for a variable number of pulsars whose cadence gets halved at 20 years to free telescope time for high cadence campaigns on the most sensitive pulsars. The color bar encodes the number of pulsars on low cadence (in sets of 6 pulsars) as well as the GWB S/N. The inset plot tracks the change in detector volume at 4 different frequencies as a function of changes to the observing campaign.

time fixed; for this reason, we refer to this campaign as the 'optimized' campaign for the rest of the work. Figure 10 visualizes what effect this type of optimal campaign has on detecting individual binaries by plotting the loudest strain in each frequency bin for a sample of 100 of the total 1000 realizations. We proceed with this optimized campaign for single-source sensitivity and employ the functionality demonstrated in section 4 to further probe these sensitivity gains. Similar to before, we calculate the per-frequency DP_i and the TDP, but now we plot the change in DP_i and the change in TDP between the standard campaign and the optimal campaign. Figure 11(a) illustrates how sensitivity gains at high frequencies correspond to increases in the DPs of individual binaries at those frequencies while we see a decrease in DP at lower frequencies. We see the overall effect that the optimal campaign has on detectability in figure 11(b), which traces TDP as a function of the number of pulsars on low cadence. Most of the population realizations exhibit sizable increases in TDP with more pulsars being moved to the optimal campaign, however some fraction of the realizations—those corresponding to a population-loudest binary at low frequencies—exhibit a small decrease in TDP at 30 years with this campaign due the slight losses in sensitivity at those frequencies.

5.3. Sky-targeted high cadence campaigns

A 'directional tuning' of PTAs has been previously explored in several works [21, 23]. The prevailing idea is that observing pulsars more frequently in the approximate (but not exact) direction of a particular source yields higher single-source S/N in that direction on the sky. Using sensitivity curves, we can rapidly test the directional tunability of a PTA by creating a sensitivity sky map, which encodes directional detector sensitivity in discrete, equal-area pixels across the sky. We can add observations or adjust noise for particular pulsars before recomputing these sky maps and calculating the S/N to hypothetical test sources. We consider four test sources of equal strain but differing sky locations to investigate how tunable our simulated detector is to different regions of the sky. We place the nearest ten pulsars (which exist in the detector at year 20) to source of interest on high (×4) cadence campaigns for
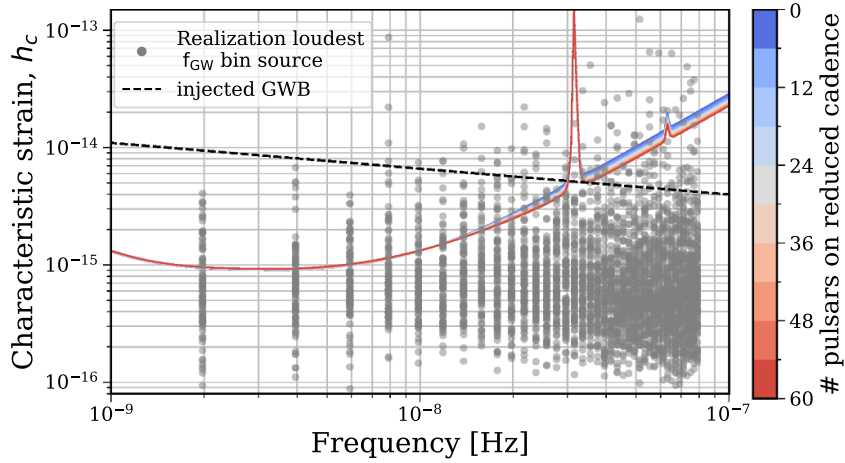
Figure 10. The loudest strain in each frequency bin across 100 population realizations plotted next to the sky-marginalized, single-source strain sensitivity of the detector reveals which binaries might be detectable.
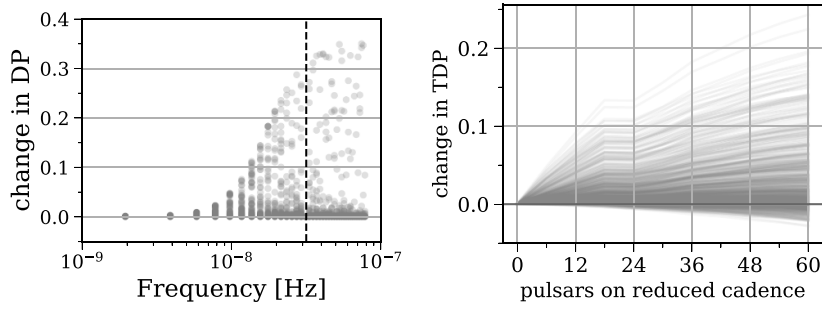

Figure 11. Changes in detection statistics of the optimal vs. the standard campaign. (a) The change in detection probability between the standard campaign and the optimal campaign with 60 (10) pulsars on reduced (high) cadence is plotted for each frequency bin across realizations. The green line partitions the PTA frequency band into a low and high GW frequency sub bands, the latter of which we might expect to see more multi-messenger candidates. This is further explored in section 5.7. (b) The change in total detection probability is graphed as a function of change in observing campaign for the optimal campaign.

the next 10 years while we put the weakest 60 pulsars (in terms of GWB S/N contribution excluding those closest to the source) on half cadence campaigns for the same duration. The percent change in source S/N is laid out in four different panels in figure 12 corresponding to hypothetical candidates in four distinct sky regions, corresponding to differing pulsar density and hence sky sensitivity. We note that these fixed observing time targeted campaigns do not increase our S/N very dramatically in the source direction and decrease it by varying amounts in the other directions, which is in agreement with what was found for the 'Targeted' search in [23]. We do note that this directional tuning is most effective in the direction with medium but not lowest pulsar density (bottom right) since fewer observations were being taken in that direction to begin with compared to the higher density pulsar regions (top and bottom left) but
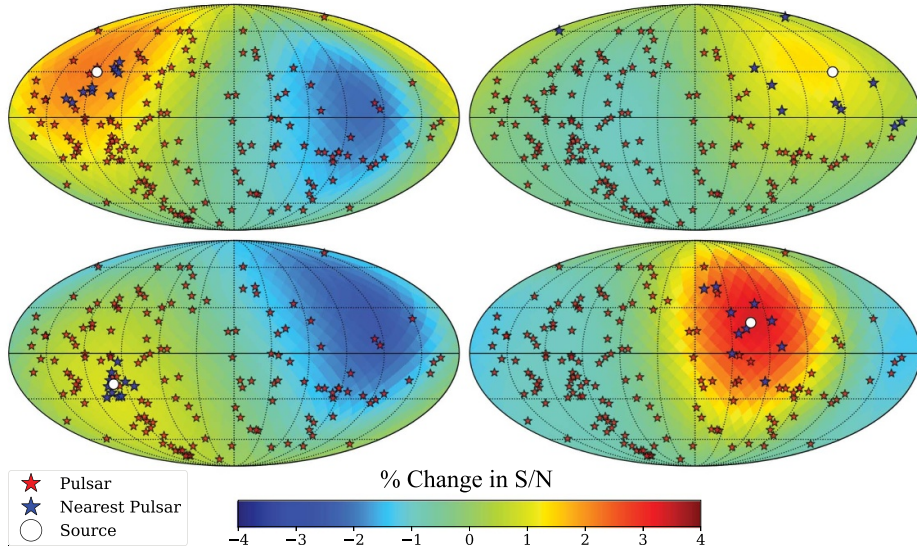
Figure 12. Sky-targeted campaign. In this alternate campaign, we quadruple the observing cadence of the nearest ten pulsars (marked by a blue star) to each source (marked by a white circle) while halving the cadence of the weakest 60 pulsars (by GWB S/N contribution) such that telescope time is held fixed. The percent changes in S/N are plotted as sky maps for the source frequency of 8 nHz. The largest percent increase in S/N is <4%. Note that we increase the cadence of the nearest 10 pulsars to the source at 20 years so pulsars added to the array past 20 years are on standard cadence and remain marked in red. We omit the axis labels for legibility, but the sky orientation remains the same as in figure 1.

the nearest pulsars are closer to the source compared to the lowest pulsar density region (top right).

Since telescope time is fixed in this framework, one can think of this as reallocating observations from across the sky, weighted by pulsar density, to a particular region on the sky. So it makes sense that the lower pulsar density regions are affected more since more observations are actually moving to that sky region. This detector in particular is simulated to be IPTA-like in sky-coverage; regional PTAs typically exhibit less homogenous sky-coverage. Therefore, one could expect more directional tunability in those cases. One caveat of our method, however, is that we cannot precisely say how these targeted, high cadence campaigns impact the parameter estimation of the signal. One recent study, [46] (which generally agrees with previous studies [47, 48]), concludes that S/N is the dominant factor in source sky localization followed by source proximity to pulsars and source chirp mass. While we observe increased S/N to the source and can thereby surmise better sky localization, we cannot full quantify the impact of these campaigns on parameter estimation; moreover, these effects also depend upon GW frequency and source chirp mass in a way that is not explored in this work, but efforts are already underway to incorporate these effects into `hasasia` for future studies.

5.4. Single-source sensitivity gains of a simultaneous GWB fit

The presence of the GWB will necessitate searches for CWs along with a correlated red noise process in order to model out the power in the Earth-term of the GWB [49, 50]. Improvements in single-source sensitivity with a simultaneous fit depend upon how well the GWB is fit in the
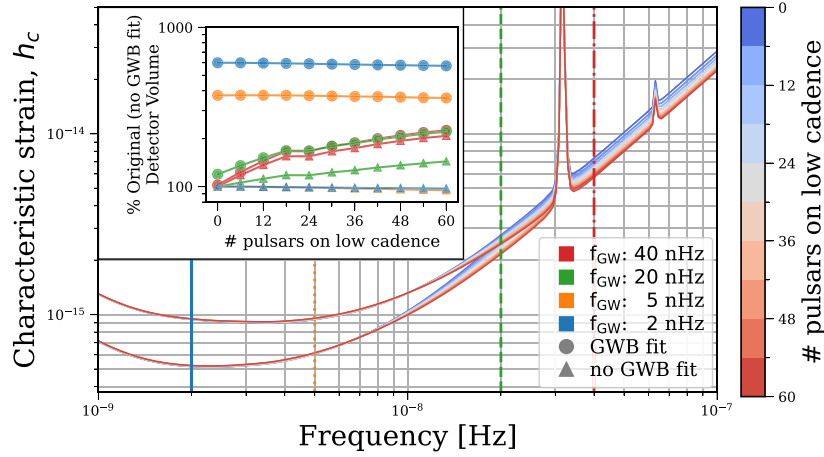
Figure 13. The GWB fit. Under the same optimal campaign, we compare single-source sensitivity with a perfect GWB Earth-term fit to no fit at all. The inset plot compares changes in the detector volume with respect to the standard campaign without a GWB Earth-term fit. This is plotted as a function of changes in observing campaign for the cases of with (dots) and without (triangles) a GWB Earth-term fit at the 4 reference frequencies marked.

search [51]. We bound the sensitivity gains in this simultaneous fit by considering the limiting cases: a perfect GWB Earth-term simultaneous fit, i.e. only the half of the GWB power in the so-called pulsar term is unmitigated, and no simultaneous fit at all. While our actual sensitivity to single-sources lies somewhere between these two extremes, the potentially drastic difference in sensitivity between these extremes is evident in figure 13 nonetheless. Our optimal campaign is even more optimal in the case of a simultaneous fit since there is less GWB foreground in the data, which ultimately results in higher sensitivity gains at high frequencies and lower sensitivity losses at low frequencies compared to not performing a simultaneous fit. Notably, the optimal campaign yields sensitivity gains which extend to lower frequencies since the PTA crosses over from being white noise dominated to red noise dominated at lower frequencies with less GWB to contend with. Unsurprisingly, fitting out the power in the GWB Earth-term dramatically improves single-source sensitivity at low frequencies. Recall that the GWB fit is dependent on the number of pulsar pairs and their time-spans, so there is a trade-off between GWB fit and high cadence campaigns in single-source sensitivity.

This hints at the fact that single-source sensitivity as well as the sensitivity gains (and losses) in the optimal campaign and otherwise are dependent on the spectrum of the GWB itself. We explore the relationship between single-source sensitivity and the spectrum of the GWB in the next section. Again, this type of analysis proves trivial in our sensitivity curve framework since all we have to do is inject a differing signal in our noise covariance before recomputing sensitivity.

5.5. Dependence of single-source sensitivity on GWB parameters

The spectral characterization of the tentative GWB observed in PTAs is currently unconstrained and noise-model dependent [2]. Advances in noise modeling techniques as well as more data are expected to improve this in the near future [52–57], but for now, we demonstrate
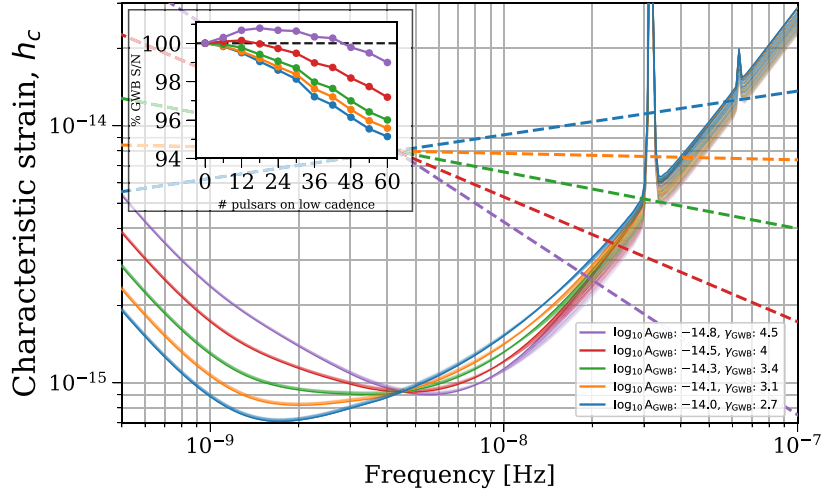
Figure 14. Spectral index of the GWB and single-source sensitivity. We simulate single-source sensitivity curves for our detector at 30 years for various power-law GWB parameter values which are consistent with contemporary PTA observations from [1]. Single source sensitivity curves for the optimal campaign (see section 5.2) are plotted above and the scaling of each GWB's S/N under the optimal campaign is displayed in the inset plot.

in figure 14 how different power-law GWBs affect single-source sensitivity and the efficacy of the optimal observing campaign. We select parameter values of the GWB which are consistent with the joint posterior in [1] and roughly follow the covariance of the amplitude and spectral index. At high frequencies, the GWB has little impact on single-source sensitivity, and thus the sensitivity curves all align here. Whereas at the lowest frequencies, single-source sensitivity is largely dictated by the power-spectrum of the GWB. We note that the optimal campaign is serendipitously even more optimal for GWBs with steeper spectra than shallower spectra due to the differing frequencies at which the PTA crosses over from being GWB-limited (and red noise-limited) to white noise-limited. In fact, as the inset plot demonstrates for the steepest spectrum ($\gamma_{GWB} = 4.5$), the optimal campaign *increases* S/N to the GWB itself at 30 years for a few pulsars on low cadence. This effect is a result of high cadence campaigns uncovering more GWB from underneath the white noise, thereby boosting the S/N.

We can also explore the most probable frequency bin of the first single source detection as a function of the GWB. Using these sensitivity curves and the population simulations (which we adjust for each different GWB), we note the frequency bin of the first binary in each population realization to reach a 3σ detection. Despite appreciable impacts on single-source sensitivity as seen in figures 14 and 15 reveals that differing GWBs only slightly impact the most probable frequency bin of the first detection, which is typically in the second or fourth frequency bin ($\sim 4 - 8$ nHz). We note that a shallowed GWB corresponds to a more likely chance to detect the first binary at higher frequencies.
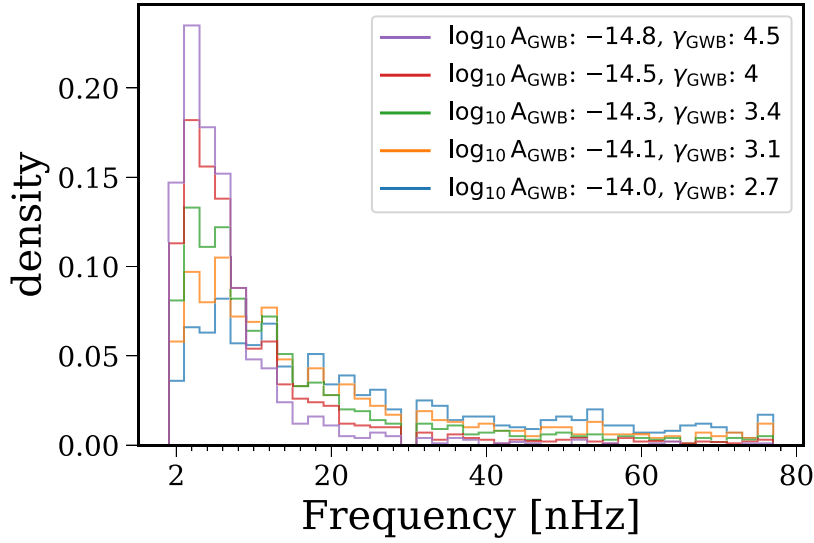
Figure 15. Spectral index of the GWB and first single-source detections. A histogram of the frequency bin of the first detected binary across all 1000 realizations of the SMBBH population. The color encodes injected GWB parameters. Note that the median of first detection shifts to higher frequencies for shallower spectral indices, but the most probable frequency bin remains between the second and fourth frequency bins across different GWBs.

5.6. Next-generation radio telescope campaigns

Current and future improvements to radio telescopes hold the promise of significantly reducing white noise in PTA data sets. Here we consider high cadence campaigns conducted by a 'Next-Generation' radio telescope *without* reducing telescope time. In an IPTA-wide setting, various PTAs have the potential to tune their observations in concert with one another to produce an optimized IPTA data set. For instance, PTAs could use the most sensitive, Next-Generation, radio telescopes such as FAST [58] or the SKA [59] to perform targeted, high cadence campaigns on the most sensitive pulsars. These targeted campaigns are all the more potent as the Next-Generation telescopes achieve observational precision, which is unrivaled by other radio telescopes. To model this scenario, in addition to the standard campaign, we place the most sensitive N pulsars on an observing campaign of 52 observations per year and also adjust the observational error (RMS) by a factor of $1/10$ to model increased timing precision. Figure 16 shows the sensitivity curves for such a campaign for up to 24 pulsars on high cadence. The sensitivity gains resemble those of section 5.1, except much more extreme. Figure 17 highlights the dramatic increases in detector volume that PTAs could reap from such coordinated campaigns and sensitive telescopes. We note that at high frequencies, the detector volume is more than tripled as compared to the standard campaign. Lastly, we recompute the change in per frequency DP and TPD in figure 18 to further demonstrate the benefits of having such an extreme campaign. Since we do not relegate any pulsars to a reduced cadence, we only observe increases in DP across all frequencies and this change is even higher than some frequencies as compared with optimal campaign in section 5.2. This is a testament to the power of high cadence campaigns and the potential of future radio telescopes to bolster the field of nanohertz GW astronomy.
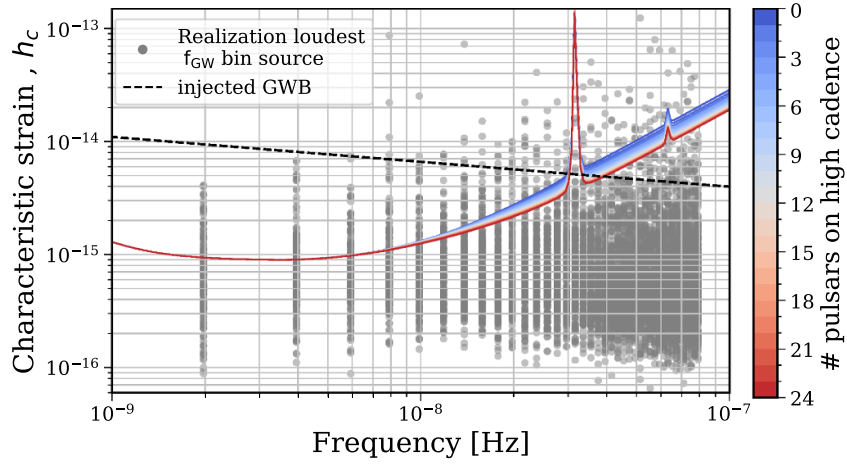
Figure 16. Next-generation telescope campaign sensitivity. With the *N*–most sensitive pulsars on a weekly observing cadence with a smaller observational error, we compare single-source strain sensitivity with an optimistic GWB Earth-term fit to the loudest strains in each frequency of 100 SMBBH population realizations.
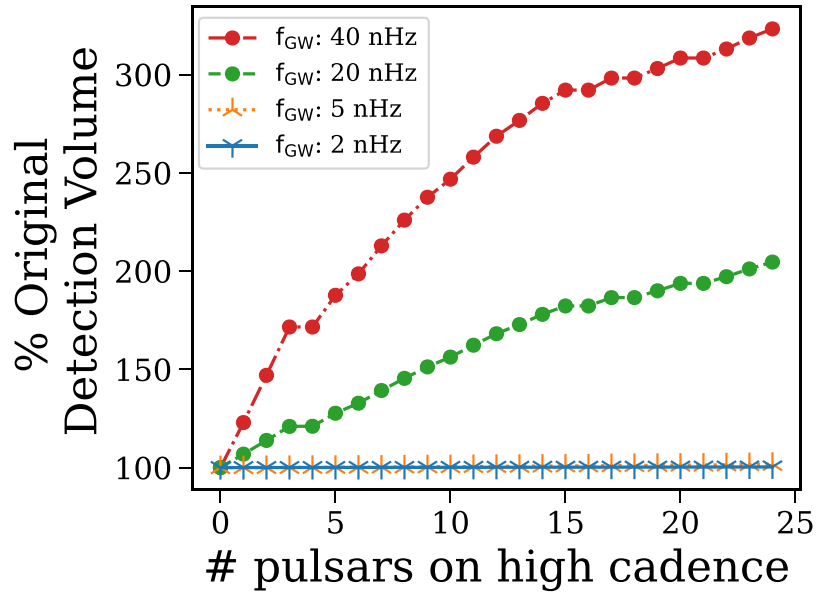

Figure 17. Next-generation telescope campaign detector volume. The change in detector volume is plotted as a function of the number of pulsars on high cadence campaigns with a next-generation telescope.
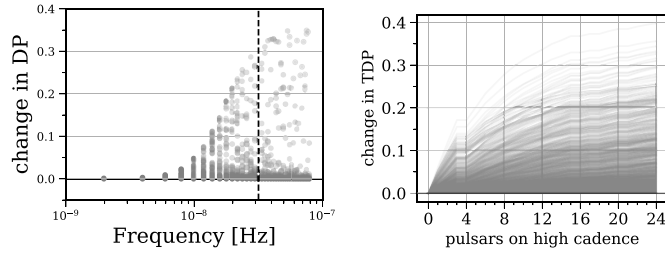
Figure 18. Next-generation telescope campaign detection statistics. (a) The change in DP for each frequency is shown for the case of 24 pulsars on high cadence campaigns with a next generation telescope. The black line demarcates 1/yr. (b) We calculate the change in TDP as a function of number of pulsars on a high cadence campaign.

5.7. Tuning towards a multi-messenger detection

Multi-messenger astrophysics (MMA) combines multiple heralds of information to corroborate and enrich the scientific understanding from each source (e.g. [60]). The legacy survey of space and time (LSST), which will be carried out by the Vera Rubin Observatory, is expected to give us of order 10^9 SMBBH candidates from periodic variability in active galactic nuclei (AGN) light curves [61, 62] due to the unprecedented sensitivity and cadence. However, since the survey will last only for 10 years, many of these candidates will not extend into the lowest, most sensitive PTA frequency bins. For this reason, it would be beneficial to tune our detector to higher frequencies in the hopes of achieving a multi-messenger detection with the LSST. With the planned time span of 10 years for the LSST survey, we will not be able to flag any AGN light curves as being potentially periodic below $\sim$10 nHz, which corresponds to the 5th frequency bin in our population simulations.

Referring to figure 11 we see that if we focus on frequencies > 10 nHz the per frequency DP_i all increase for the optimal campaign. This increase in sensitivity at higher frequencies[3] would most likely produce a detection of anisotropy sooner, thus revealing the source of the GWB in the nHz band to be SMBBHs rather than cosmological.

Another parameter one might consider with respect to an MMA detection with LSST is the interplay between LSST and PTA sensitivity in terms of binary orientation. For instance, periodicity in AGN is more likely to be discernible from noise for near edge-on ($\iota \sim 90°$) binaries due to Doppler boosting and GW self-lensing effects [64–66], whereas PTAs are more sensitive to face-on binaries ($\iota \sim 0$) as can be seen in the sky response functions for GWs (see appendix A). Figure 19 compares detection prospects between the optimal PTA binary orientation and edge-on candidate orientations for detection. A 'partial-marginalization' over only the polarization angle and phase of the S/N is used to do this calculation and is available in `hasasia`.

5.8. Sensitivity beyond black hole binaries

Binary black holes are not the only possible source for the emerging stochastic background in PTAs. NANOGrav explored a variety of new-physics models which could produce GWs in the PTA band [67]. Taking the maximum posterior values reported in [67, 68]

---

[3] Note that anisotropy is predicted to be detected first at higher frequencies due to the larger scatter of binary amplitudes there [8, 63].
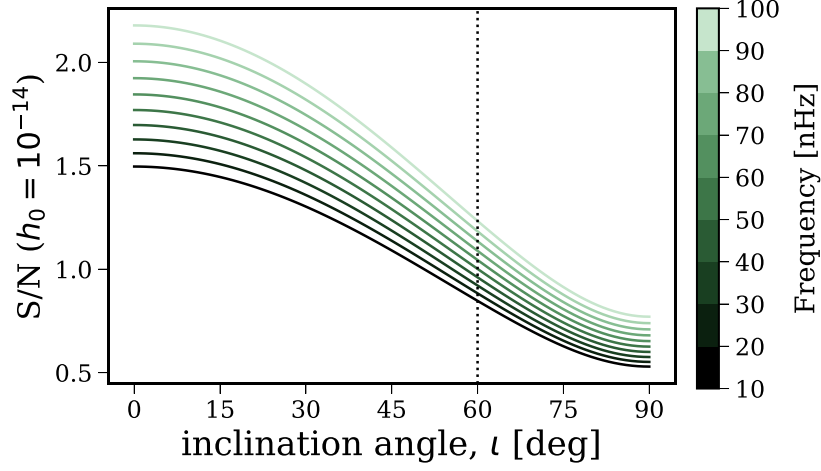
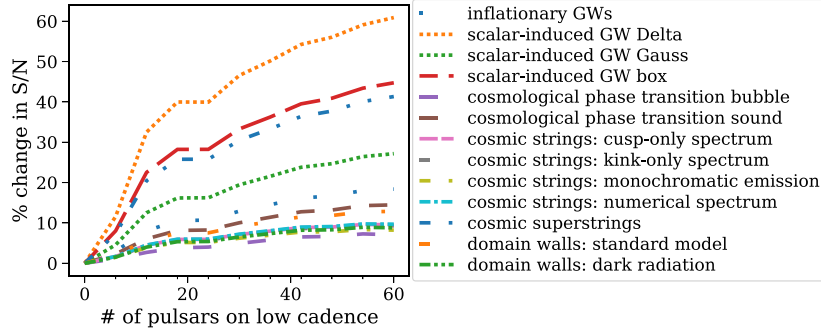**Figure 19.** Dependence of S/N on inclination angle, ι.

Figure 20. New physics sensitivity. The increase in S/N of the optimal campaign considered in section 5.2 over the standard campaign using the maximum posterior values of several of the new-physics models in [67, 68].

(see appendix A, table 4), we calculate the change in various S/N values in figure 20 between the standard and optimal campaigns. We find that the optimal observing campaign again yields higher S/N at higher frequencies. In general, the high frequency regime of many of these new-physics signals would more readily allow PTAs to constrain their parameter spaces as they tend to have a shallower effective spectral index than the SMBBH-sourced GWB.

5.9. Forecasting parameter estimation uncertainties

Currently the `hasasia` framework cannot forecast how the uncertainties of GW parameters will change over time. However, the S/N can be related to the amplitude and spectral index of the GWB [28] and the sky localization of single sources [46]. Pol *et al* [28] found the following empirical scaling relations for the GWB parameter uncertainties as a function of the Hellings-Downs correlated S/N, ρ_{HD}, $\Delta A_{\mathrm{GWB}}/A_{\mathrm{GWB}} \approx 44\% \times (\rho_{\mathrm{HD}}/3)^{-\frac{17}{25}}$ and $\Delta\alpha/\alpha \approx 40\% \times (\rho_{\mathrm{HD}}/3)^{-\frac{17}{20}}$.

In Petrov *et al* [46] they find that the sky localization area in square degrees can be related to the S/N for single sources, ρ_{ss}, as $\sim 1/\rho_{\mathrm{ss}}^2$.

In recent years Fisher matrix analyses have been used to predict the parameter uncertainties for PTAs of the GWB [29] and single sources [23]. These calculations are done in the frequency domain for GWBs [29] pointing to the ability to incorporate these types of calculations directly into `hasasia` and PTA sensitivity investigations in the future.

6. Discussion and outlook

As PTAs enter the detection era, with significance of the GWB anticipated to reach a detection in the coming years, it is natural to focus next on the resolution of anisotropies in the GWB and the detection of individual SMBBHs. This study compares detection statistics for individual sources then demonstrates a powerful method to improve PTA campaigns, using sensitivity curves, which can accurately account for GWB self-noise. Specifically, we find that for our simulated IPTA-like detector,

(i) Adding high cadence campaigns reduces the power in the white noise, yielding higher sensitivities at $f_{\mathrm{GW}} > 10^{-8}$ Hz.

(ii) For fixed observing resources, a high cadence (low cadence) campaign on the most (least) sensitive pulsars facilitates high frequency sensitivity gains while yielding minimal sensitivity losses at lower frequencies.

(iii) For an IPTA-like detector under fixed observing time constraints, a sky-targeted campaign only minimally boosts sensitivity in that direction. Increasing the sensitivity of the best pulsars, across the sky, gives better gains.

(iv) Simultaneously fitting for a GWB and single-sources provides a large sensitivity boost at low frequencies which depends on the accuracy of the fit.

(v) The parameters of the GWB impacts at what frequencies we expect to see single sources, and impacts where PTAs cross over from being red noise dominated to white noise dominated.

(vi) Next generation radio telescopes will drastically reduce the white noise in PTAs thus opening up the possibility for more high frequency candidates.

(vii) The sensitivity gains at higher frequencies of the optimal campaign synergizes well with the frequencies possible for an MMA detection with LSST and the expectation that anisotropies will be seen at higher frequencies first.

(viii) The optimal observational campaign for PTAs improves sensitivity to various new-physics models at high frequencies, where they could much more easily be disentangled from a SMBBH GWB.

The use of sensitivity curves skirts the computational cost of simulating many realizations of datasets, and instead allows for an efficient recalculation of the noise in the detector for myriad imaginable scenarios. We demonstrate this method to be versatile in the scenarios it can consider and to be computationally cheap with most simulations being run on a laptop. However, this methodology currently lacks the ability to compute or forecast the parameter estimation of both the GWB and single sources. Future work includes application of this method on a real, reduced PTA dataset with more finely conditioned `holodeck` SMBBH population simulations as well as an exploration of the interplay between GWB and single-source parameter estimation.

As PTAs approach the detection of GWs from individual supermassive black hole binaries, the successful PTA must account for every iota of sensitivity like a diligent bookkeeper. While

additional observational resources are the ideal path to sensitivity gains, realistic constraints leave PTAs with a resource optimization problem. This problem must be addressed sooner rather than later as any changes to a campaign will take time to integrate into the sensitivity.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/jeremy-baier/pta_tuning.

Acknowledgments

Appendix A. The polarization, inclination and phase averaged $\mathcal{F}_e$ statistic

The $\mathcal{F}$ statistic is a maximum-likelihood detection statistic for continuous gravitational-wave searches, originally developed in the context of ground-based detectors by Jaranowski, Królak and Schutz [72, 73]. It was later extended to space-based searches by [74], and adapted to PTAs as the $\mathcal{F}_e$ statistic[4] by [38].

What follows is a summary of the $\mathcal{F}_e$ statistic with enough information to relate the traditional $\mathcal{F}$ statistic to the statistic used in [32] and originally incorporated into `hasasia`, which is averaged over the gravitational-wave polarization angle ψ, inclination angle ι, and phase ϕ_0. Readers interested in more details should refer to the references mentioned above.

As shown in [73, 74], the $\mathcal{F}_e$ statistic can be written as

$$2\mathcal{F}_e = x_a \left[\mathcal{M}^{-1} \right]^{ab} x_b \tag{A.1}$$

where

$$x_a \equiv (x|h_a) , \qquad \mathcal{M}_{ab} \equiv (h_a|h_b) , \tag{A.2}$$

---

[4] Note that `hasasia` uses the $\mathcal{F}_e$ statistic instead of the incoherent $\mathcal{F}_p$ statistic [39] in order to study the sky position-dependence of single-source sensitivity in the PTA band.

are defined with respect to the inner product

$$(x|y) = 4\mathrm{Re} \left\{ \int_0^\infty \mathrm{d}f > \sum_{I,J} \tilde{x}_I(f) \left[S^{-1}(f) \right]_{IJ} \tilde{y}_J^*(f) \right\}. \tag{A.3}$$

Here $[S^{-1}(f)]_{IJ}$ is the inverse matrix to the matrix of one-sided power spectral densities $S_{IJ}(f)$ defined via

$$\langle \tilde{n}_I(f) \tilde{n}_J^*(f') \rangle = \frac{1}{2} S_{IJ}(f) \, \delta(f - f') , \tag{A.4}$$

and $S_I(f) \equiv S_{II}(f)$. In addition, $\tilde{x}_I(f)$ is the Fourier transform of the time-domain data $x_I(t)$ for pulsar I:

$$x_I(t) = n_I(t) + \mathcal{A}^a h_{aI}(t) , \tag{A.5}$$

where $\mathcal{A}^a$ are amplitude parameters which depend on $\{h_0, \cos\iota, \psi, \phi_0\}$ via

$$\begin{aligned}
\mathcal{A}^1 &= +A_+ \cos 2\psi \cos\phi_0 - A_\times \sin 2\psi \sin\phi_0 , \\
\mathcal{A}^2 &= +A_+ \sin 2\psi \cos\phi_0 + A_\times \cos 2\psi \sin\phi_0 , \\
\mathcal{A}^3 &= -A_+ \cos 2\psi \sin\phi_0 - A_\times \sin 2\psi \cos\phi_0 , \\
\mathcal{A}^4 &= -A_+ \sin 2\psi \sin\phi_0 + A_\times \cos 2\psi \cos\phi_0 ,
\end{aligned} \tag{A.6}$$

with

$$A_+ \equiv h_0 \left(1 + \cos^2\iota \right), \qquad A_\times \equiv 2h_0 \cos\iota , \tag{A.7}$$

and $h_{aI}(t)$ are basis functions defined by

$$\begin{aligned}
h_{1I}(t) &= +\frac{1}{2\pi f_0} \sin(2\pi f_0 t) \, F_I^+(\hat{k}) , \\
h_{2I}(t) &= +\frac{1}{2\pi f_0} \sin(2\pi f_0 t) \, F_I^\times(\hat{k}) , \\
h_{3I}(t) &= -\frac{1}{2\pi f_0} \cos(2\pi f_0 t) \, F_I^+(\hat{k}) , \\
h_{4I}(t) &= -\frac{1}{2\pi f_0} \cos(2\pi f_0 t) \, F_I^\times(\hat{k}) .
\end{aligned} \tag{A.8}$$

Here

$$F_I^+(\hat{k}) \equiv \frac{1}{2} \frac{\hat{p}_I^i \hat{p}_I^j}{1 + \hat{p}_I \cdot \hat{k}} e_{ij}^+(\hat{k}), \quad F_I^\times(\hat{k}) \equiv \frac{1}{2} \frac{\hat{p}_I^i \hat{p}_I^j}{1 + \hat{p}_I \cdot \hat{k}} e_{ij}^\times(\hat{k}), \tag{A.9}$$

are the $+, \times$ antenna pattern functions for pulsar I, $\hat{p}_I^i$ is the unit vector from the Earth to pulsar I, and $e_{ij}^{+,\times}(\hat{k})$ are the $+$ and $\times$ polarization tensors.

It follows that the expectation value of the $\mathcal{F}_e$ statistic with respect to different noise realizations is

$$\langle 2\mathcal{F}_e \rangle_{\mathrm{noise}} = 4 + \rho_{\mathrm{MF}}^2(\mathcal{A}) , \tag{A.10}$$

where

$$\rho_{\mathrm{MF}}^2(\mathcal{A}) \equiv (s|s) \tag{A.11}$$

is the (expected) squared matched-filter signal-to-noise ratio, evaluated for the true value of the signal parameters. Explicitly,

$$
\rho_{\mathrm{MF}}^2(\mathcal{A}) = \frac{T}{(2\pi f_0)^2} \left[\sum_I \frac{\left(F_I^+(\hat{k})\right)^2}{S_I(f)} \left(A_+^2 \cos^2 2\psi + A_\times^2 \sin^2 2\psi\right) \right.
$$
$$
+ \sum_I \frac{\left(F_I^\times(\hat{k})\right)^2}{S_I(f)} \left(A_+^2 \sin^2 2\psi + A_\times^2 \cos^2 2\psi\right)
$$
$$
\left. + 2 \sum_I \frac{F_I^+(\hat{k}) F_I^\times(\hat{k})}{S_I(f)} \left(A_+^2 - A_\times^2\right) \cos 2\psi \sin 2\psi \right] . \tag{A.12}
$$

The above result for $\langle 2\mathcal{F}_e\rangle_{\mathrm{noise}}$ allows us to conclude that $2\mathcal{F}_e$ is described by a non-central χ^2 distribution with 4 DOFs and non-centrality parameter $\rho_{\mathrm{MF}}^2(\mathcal{A})$ [39]

$$
\mathcal{P}_{2\mathcal{F}_e} = \frac{(2\mathcal{F}_e)^{1/2}}{\rho} I_1\left(\rho\sqrt{2\mathcal{F}_e}\right) e^{-\left(\mathcal{F}_e + \frac{1}{2}\rho^2\right)} , \tag{A.13}
$$

where $I_1(x)$ is the modified Bessel function of the first kind of order 1.

　　If instead we average $2\mathcal{F}_e$ over the GW polarization angle ψ, inclination angle ι, and phase ϕ_0, we obtain

$$
\langle 2\mathcal{F}_e\rangle_\Omega = 2\mathcal{F}_0 + \langle \rho^2(h_0)\rangle_\Omega, \tag{A.14}
$$

where $\mathcal{F}_0$ is the equivalent of $\mathcal{F}_e$ but for the null hypothesis (i.e. the absence of a GW signal), and $\langle \rho^2(h_0)\rangle_\Omega$ is the squared signal-to-noise ratio defined in [32] and used in `hasasia`, obtained by averaging over angles and phase, $\langle\ \rangle_\Omega \equiv \langle\ \rangle_{\cos\iota,\psi,\phi_0}$. Explicitly,

$$
\begin{aligned}
\langle \rho^2(h_0)\rangle_\Omega &\equiv \langle \rho_{\mathrm{MF}}^2(\mathcal{A})\rangle_{\cos\iota,\psi,\phi_0} \\
&= \frac{1}{2}\int_{-1}^1 \mathrm{d}(\cos\iota) \> \frac{2}{\pi}\int_{-\pi/4}^{\pi/4} \mathrm{d}\psi \> \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\phi_0 \> \rho_{\mathrm{MF}}^2(\mathcal{A}) \\
&= \left(\frac{1}{2}\right)\left(\frac{4}{5}\right) T \frac{h_0^2}{(2\pi f_0)^2} \sum_I \frac{\left(F_I^+(\hat{k})\right)^2 + \left(F_I^\times(\hat{k})\right)^2}{S_I(f)},
\end{aligned} \tag{A.15}
$$

where the factor of $1/2$ in the last line comes from averaging over the polarization angle ψ, and the factor of $4/5$ comes from averaging over $\cos\iota$.

　　If we further average (A.14) over noise realizations, we obtain

$$
\langle\langle 2\mathcal{F}_e\rangle_\Omega\rangle_{\mathrm{noise}} = 4 + \langle \rho^2(h_0)\rangle_\Omega = \langle\langle 2\mathcal{F}_e\rangle_{\mathrm{noise}}\rangle_\Omega, \tag{A.16}
$$

which shows that noise averaging and angle averaging commute. With angle averaging the distribution of the $\langle 2\mathcal{F}_e\rangle_\Omega$ statistic is an *ordinary* χ^2 distribution with four DOFs, but shifted to the right by $\langle \rho^2(h_0)\rangle_\Omega$

$$
\mathcal{P}_{\langle 2\mathcal{F}_e\rangle_\Omega} = \left(\mathcal{F}_e - \frac{1}{2}\rho^2\right) \exp^{-\left(\mathcal{F}_e - \frac{1}{2}\rho^2\right)} . \tag{A.17}
$$

In the absence of a signal, both of these statistics are identical and are described by an ordinary χ^2 distribution with four DOFs. This is the distribution used for the false alarm probability versus threshold relation.
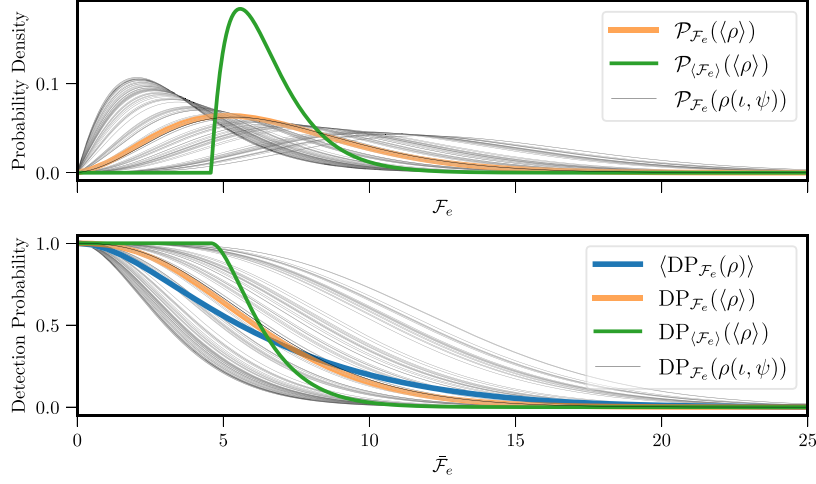
Figure A1. Comparison of the probability density functions and detection probabilities for various versions of the $\mathcal{F}_e$ statistic. In the legends above all $\langle \cdot \rangle$ refer to averaging over $\cos \iota$, ψ and ϕ. These curves are constructed for a signal that has an angle-averaged $\langle \rho \rangle_\Omega \approx 3$. Note particularly in the lower figure that $\langle \mathrm{DP}_{\mathcal{F}_e}(\rho) \rangle \neq \mathrm{DP}_{\mathcal{F}_e}(\langle \rho \rangle)$.

With these two probability distributions in hand we next demonstrate the differences in the detection probabilities (DPs), which depend on *what stage* one averages over $\cos \iota$, ψ and ϕ_0. Following [19], the DP gives the probability that a given source will be detected given a threshold for the detection statistic, $\bar{\mathcal{F}}_e$, and is calculated using the probability density function of the $\mathcal{F}_e$ statistic,

$$\mathrm{DP} = \int_{\bar{\mathcal{F}}_e}^{\infty} \mathcal{P}\left(\rho, \mathcal{F}_e\right) \mathrm{d}\mathcal{F}_e . \tag{A.18}$$

Either equation (A.13) or (A.17) can be used for $\mathcal{P}$ in equation (A.18). The probability density functions and DP functions are shown figure A1. The angle-averaged S/N , $\langle \rho \rangle_\Omega$, was derived in [32] and was the first single source detection statistic incorporated into `hasasia` and used to compare the single-source strain threshold to the Bayesian strain upper limits from [75]. For the purposes of forecasting, formalisms that predict the detection probability of single sources, like the one in Rosado *et al* [19], are preferred because they can be easily related to a false alarm probability and a TDP.

The angle-averaged signal to noise, $\langle \rho^2(h_0) \rangle_\Omega$, is very closely related to the $\mathcal{B}$-statistic, first developed for ground-based detectors by Prix and Krishnan [73] and derived for PTAs by Taylor *et al* [40]. However, in those works all four of the $\mathcal{A}$ parameters are marginalized over to form a Bayes factor. As noted in those works, the $\mathcal{A}$ priors are unphysical and lead to slight errors in recovery and injection studies. Here we instead marginalize over the physical priors of only the angle parameters in $\bar{\mathcal{A}} \equiv \{h_0, \cos \iota, \psi, \phi_0\}$. This suits the objective of our work here to build sensitivity curves for single-sources in the PTA band where we would like the freedom to change h_0. While the angle-marginalized signal-to-ratio was used in previous studies, `hasasia` now has the ability to look at sensitivity curves that depend on $\cos \iota$, ψ and ϕ_0. In order to more closely mimic a set of simulations, where sources are chosen to have specific values of ι and ψ and DPs are calculated and, only then, averaged, we use the angle-averaged DP, $\langle \mathrm{DP}_{\mathcal{F}_e}(\rho(\iota, \psi)) \rangle_\Omega$. We are able to calculate the integral over these angles using

Python-based numerical integration functionality, retaining the calculational speed necessary for these campaign-tuning investigations.

ORCID iDs

Jeremy G Baier ⓘ https://orcid.org/0000-0002-4972-1525
Jeffrey S Hazboun ⓘ https://orcid.org/0000-0003-2742-3321
Joseph D Romano ⓘ https://orcid.org/0000-0003-4915-3246

References

[1] Agazie G *et al* (The International Pulsar Timing Array Collaboration) 2023 arXiv:2309.00693
[2] Agazie G *et al* (Nanograv Collaboration) 2023 *Astrophys. J. Lett.* **951** L8
[3] Antoniadis J *et al* 2023 arXiv:2306.16214
[4] Reardon D J *et al* 2023 *Astrophys. J. Lett.* **951** L6
[5] Hellings R W and Downs G S 1983 *Astrophys. J. Lett.* **265** L39–L42
[6] Kelley L Z, Blecha L, Hernquist L, Sesana A and Taylor S R 2018 *Mon. Not. R. Astron. Soc.* **477** 964–76
[7] Finn L S and Lommen A N 2010 *Astrophys. J.* **718** 1400–15
[8] Mingarelli C M F, Sidery T, Mandel I and Vecchio A 2013 *Phys. Rev.* D **88** 062005
[9] Taylor S R and Gair J R 2013 *Phys. Rev.* D **88** 084001
[10] Taylor S R, van Haasteren R and Sesana A 2020 *Phys. Rev.* D **102** 084039
[11] Madison D R, Cordes J M and Chatterjee S 2014 *Astrophys. J.* **788** 141
[12] Foster R S and Backer D C 1990 *Astrophys. J.* **361** 300
[13] McLaughlin M A 2013 *Class. Quantum Grav.* **30** 224008
[14] Manchester R N *et al* 2013 *Publ. Astron. Soc. Aust.* **30** e017
[15] Kramer M and Champion D J 2013 *Class. Quantum Grav.* **30** 224009
[16] Manchester R N *et al* IPTA 2013 *Class. Quantum Grav.* **30** 224010
[17] Lam M T *et al* 2016 *Astrophys. J.* **819** 155
[18] Lam M T, McLaughlin M A, Cordes J M, Chatterjee S and Lazio T J W 2018 *Astrophys. J.* **861** 12
[19] Rosado P A, Sesana A and Gair J 2015 *Mon. Not. R. Astron. Soc.* **451** 2417–33
[20] Siemens X, Ellis J, Jenet F and Romano J D 2013 *Class. Quantum Grav.* **30** 224015
[21] Christy B, Anella R, Lommen A, Finn L S, Camuccio R and Handzo E 2014 *Astrophys. J.* **794** 163
[22] Lam M T 2018 *Astrophys. J.* **868** 33
[23] Liu T, Cohen T, McGrath C, Demorest P B and Vigeland S J 2023 *Astrophys. J.* **945** 78
[24] Speri L, Porayko N K, Falxa M, Chen S, Gair J R, Sesana A and Taylor S R 2023 *Mon. Not. R. Astron. Soc.* **518** 1802–17
[25] Taylor S R, Vallisneri M, Ellis J A, Mingarelli C M F, Lazio T J W and van Haasteren R 2016 *Astrophys. J. Lett.* **819** L6
[26] Mingarelli C M F, Lazio T J W, Sesana A, Greene J E, Ellis J A, Ma C-P, Croft S, Burke-Spolaor S and Taylor S R 2017 *Nat. Astron.* **1** 886–92
[27] Kelley L Z, Haiman Z, Sesana A and Hernquist L 2019 *Mon. Not. R. Astron. Soc.* **485** 1579–94
[28] Pol N S *et al* (The NANOGrav Collaboration) 2021 *Astrophys. J. Lett.* **911** L34
[29] Babak S, Falxa M, Franciolini G and Pieroni M 2024 arXiv:2404.02864
[30] Thrane E and Romano J D 2013 *Phys. Rev.* D **88** 124032
[31] Moore C J, Taylor S R and Gair J R 2015 *Class. Quantum Grav.* **32** 055004
[32] Hazboun J S, Romano J D and Smith T L 2019 *Phys. Rev.* D **100** 104028
[33] Hazboun J S, Romano J D and Smith T L 2019 *J. Open Soft. Sci.* **4** 1775
[34] Chamberlin S J, Creighton J D,E, Siemens X, Demorest P, Ellis J, Price L R and Romano J D 2015 *Phys. Rev.* D **91** 044048
[35] Anholm M, Ballmer S, Creighton J D E, Price L R and Siemens X 2009 *Phys. Rev.* D **79** 084030
[36] Ellis J A, Jenet F A and McLaughlin M A 2012 *Astrophys. J.* **753** 96
[37] Agazie G 2023 arXiv:2306.16220
[38] Babak S and Sesana A 2012 *Phys. Rev.* D **85** 044034
[39] Ellis J A, Siemens X and Creighton J D E 2012 *Astrophys. J.* **756** 175
[40] Taylor S, Ellis J and Gair J 2014 *Phys. Rev.* D **90** 104028

[41] Agazie G *et al* (The Nanograv Collaboration) 2023 *Astrophys. J. Lett.* **951** L10
[42] Hazboun J S, Meyers P M, Romano J D, Siemens X and Archibald A M 2023 *Phys. Rev.* D **108** 104050
[43] Gardiner E C, Kelley L Z, Lemke A-M and Mitridate A 2024 *Astrophys. J.* **965** 164
[44] Agazie G *et al* (The Nanograv Collaboration) 2023 *Astrophys. J. Lett.* **951** L9
[45] Kelley L Z *et al* 2023 holodeck *Zenodo* (available at: https://doi.org/10.5281/zenodo.10966412)
[46] Petrov P, Taylor S R, Charisi M and Ma C P 2024 Identifying host galaxies of supermassive black hole binaries found by ptas (arXiv:2406.04409)
[47] Sesana A and Vecchio A 2010 *Phys. Rev.* D **81** 104008
[48] Goldstein J M, Sesana A, Holgado A M and Veitch J 2019 *Mon. Not. R. Astron. Soc.* **485** 248–59
[49] Bécsy B and Cornish N J 2020 *Class. Quantum Grav.* **37** 135011
[50] Agazie G *et al* (The NANOGrav Collaboration) 2023 *Astrophys. J. Lett.* **951** L50
[51] Ferranti I, Shaifullah G, Chalumeau A and Sesana A 2024 arXiv:2407.21105
[52] Simon J, Hazboun J, Larsen B and Baier J *Astrophys. J.* in preparation
[53] The NANOGrav Collaboration *Astrophys. J.* in preparation
[54] Miles M T, Shannon R M, Reardon D J and Bailes M (MPTA collaborators and MeerTime Builders 2024 The MeerKAT Pulsar Timing Array: the 4.5-yr data release and the noise and stochastic signals of the millisecond pulsar population) *Mon. Not. R. Astron. Soc.* **536** 1467–88
[55] Antoniadis J *et al* 2023 arXiv:2306.16225
[56] Srivastava A *et al* 2023 *Phys. Rev.* D **108** 023008
[57] Reardon D J *et al* 2023 *Astrophys. J. Lett.* **951** 7
[58] Jiang P *et al* 2019 *Sci. China Phys. Mech. Astron.* **62** 959502
[59] Braun R, Bourke T, Green J A, Keane E and Wagg J 2015 Advancing Astrophysics with the Square Kilometre Array *Advancing Astrophysics With the Square Kilometre Array (AASKA14)* p 174
[60] Abbott B P *et al* (LIGO scientific and virgo and fermi-GBM and INTEGRAL Collaborations) 2017 *Astrophys. J.* **848** L13
[61] Ivezić Ž *et al* LSST: from science drivers to reference design and anticipated data products 2019 *Astrophys. J.* **873** 111
[62] Witt C A, Charisi M, Taylor S R and Burke-Spolaor S 2022 *Astrophys. J.* **936** 89
[63] Pol N, Taylor S R and Romano J D 2022 *Astrophys. J.* **940** 173
[64] D'Orazio D J, Haiman Z and Schiminovich D 2015 *Nature* **525** 351–3
[65] D'Orazio D J and Di Stefano R 2018 *Mon. Not. R. Astron. Soc.* **474** 2975–86
[66] Kelley L Z, D'Orazio D J and Di Stefano R 2021 *Mon. Not. R. Astron. Soc.* **508** 2524–36
[67] Afzal A *et al* (The NANOGrav Collaboration) 2023 *Astrophys. J. Lett.* **951** L11
[68] Afzal A *et al* (The NANOGrav Collaboration) 2024 *Astrophys. J. Lett.* **971** L27
[69] Hunter J D 2007 *Comput. Sci. Eng.* **9** 90–95
[70] Harris C R *et al* 2020 *Nature* **585** 357–62
[71] Virtanen P *et al* 2020 (SciPy 10 Contributors) *Nat. Methods* **17** 261–72
[72] Jaranowski P, Królak A and Schutz B F 1998 *Phys. Rev.* D **58** 063001
[73] Prix R and Krishnan B 2009 *Class. Quantum Grav.* **26** 204013
[74] Cutler C and Schutz B F 2005 *Phys. Rev.* D **72** 063006
[75] Aggarwal K *et al* (The NANOGrav Collaboration) 2019 *Astrophys. J.* **880** 116