

# The Hellings and Downs correlation of an arbitrary set of pulsars

Bruce Allen\*

*Max Planck Institute for Gravitational Physics (Albert Einstein Institute),  
Leibniz Universität Hannover, Callinstrasse 38, D-30167, Hannover, Germany*

Joseph D. Romano†

*Department of Physics and Astronomy, Texas Tech University, Lubbock, TX 79409-1051, USA*

(Dated: August 16, 2022)

Pulsar timing arrays (PTAs) detect gravitational waves (GWs) via the correlations that the waves induce in the arrival times of pulses from different pulsars. The mean correlation  $\mu_u(\gamma)$  as a function of the angle  $\gamma$  between the directions to two pulsars was calculated by Hellings and Downs in 1983. The variance  $\sigma_{\text{tot}}^2(\gamma)$  in this correlation was recently calculated for a single pulsar pair at angle  $\gamma$ . Averaging over many such pairs, uniformly distributed on the sky, reduces this to an intrinsic cosmic variance  $\sigma_{\text{cos}}^2(\gamma)$ . We extend that analysis to an arbitrary finite set of pulsars, distributed at specific sky locations, for which the pulsar pairs are grouped into finite-width bins in  $\gamma$ . Given (measurements or calculations of) the correlations for any set of pulsars, we find the best way to estimate the mean in each bin. The optimal estimator of the correlation takes into account correlations among all of the pulsars that contribute to that angular bin. We also compute the variance in the binned estimate. For narrow bins, as the number of pulsar pairs grows, the variance drops to the cosmic variance. For wider bins, by sacrificing angular resolution in  $\gamma$ , the variance can even be reduced below the cosmic variance. Our calculations assume that the GW signals are described by a Gaussian ensemble, which provides a good description of the confusion noise produced by expected PTA sources. We illustrate our methods with plots of the GW variance for the sets of pulsars currently monitored by several PTA collaborations. The methods can also be applied to future PTAs, where the improved telescopes will provide larger pulsar populations and higher-precision timing.

## I. INTRODUCTION

A pulsar timing array (PTA) is a galactic-scale gravitational-wave (GW) detector that searches for low-frequency (nanohertz) GWs by precisely monitoring the arrival times of pulses from an array of galactic pulsars [1]. Gravitational waves (e.g., from inspiraling supermassive black-hole binaries in the centers of merging galaxies) influence the pulse arrival times in a way that is correlated between different pulsars. The mean correlation between a pair of pulsars depends upon the angular separation  $\gamma$  between the lines of sight to each member of the pair, as seen from Earth.

The average angular correlation was calculated by Hellings and Downs [2] for a unit amplitude, isotropic and unpolarized GW background. It has the simple analytic form

$$\mu_u(\gamma) = \frac{1}{4} + \frac{1}{12} \cos \gamma + \frac{1}{2} (1 - \cos \gamma) \log \left( \frac{1 - \cos \gamma}{2} \right), \quad (1.1)$$

where  $\gamma$  is the angular separation between a pair of pulsars, and “u” means “unpolarized”. Observation of a correlation proportional to this Hellings and Downs curve is the “smoking gun” signature that a PTA has detected GWs [3].

Several groups are searching for such correlations. However, there is currently little evidence for the Hellings and Downs angular dependence, which makes it difficult to claim that a GW background is responsible. Nonetheless, these groups do report strong statistical evidence for fluctuations in the *individual* pulsar arrival times that share the same “red” spectrum [4–7] that a GW background is expected to produce [8].

Could the lack of evidence for the Hellings and Downs angular dependence be a statistical fluctuation? To answer this, it is important to understand what variations away from the Hellings and Downs predicted mean might be expected. The size of such fluctuations is quantified by the variance of the Hellings and Downs correlation.

---

\* bruce.allen@aei.mpg.de

† joseph.d.romano@ttu.edu

### A. Total variance, pulsar variance, and cosmic variance

Recent work [9] calculates this variance for several GW source models, neglecting all sources of noise. The most important model contains  $N$  unpolarized point sources, uniformly distributed (statistically) in space, radiating GWs at the same frequency but with independent random phases and sky positions. For large  $N$ , this creates “confusion noise”, giving rise to a stationary and Gaussian stochastic process.

The (total) variance  $\sigma_{\text{tot}}^2$  computed in [9] is a sum of “pulsar variance” and “cosmic variance”  $\sigma_{\text{cos}}^2$ . Pulsar variance arises because different pairs of pulsars separated by the same angle  $\gamma$  have correlations that differ from the average, in a way that depends (unpredictably) upon their sky positions. Reference [9] computes the total variance (there denoted  $\sigma^2$  *without* a subscript) for a *single randomly selected pulsar pair* separated by angle  $\gamma$ . In contrast, the cosmic variance is the variance of the correlation *after* the correlation has been averaged over all possible locations/orientations of the pair. This corresponds to employing an *infinite number of pulsar pairs* separated by angle  $\gamma$ , uniformly distributed about the sky, and is called *pulsar averaging*.

In the confusion-noise model of [9], cosmic variance arises because, even after pulsar averaging, the correlation depends upon the relative phases of the GW sources. Each realization of the universe exhibits different pulsar-averaged correlations. The cosmic variance is the amount by which the pulsar-averaged correlation curve is expected to differ from the Hellings and Downs prediction in an ideal world containing an infinite number of noise-free pulsars. Unlike the pulsar variance, it cannot be reduced: the cosmic variance is a fundamental limit to the precision with which the Hellings and Downs predicted mean might be observed at a particular angle  $\gamma$ .

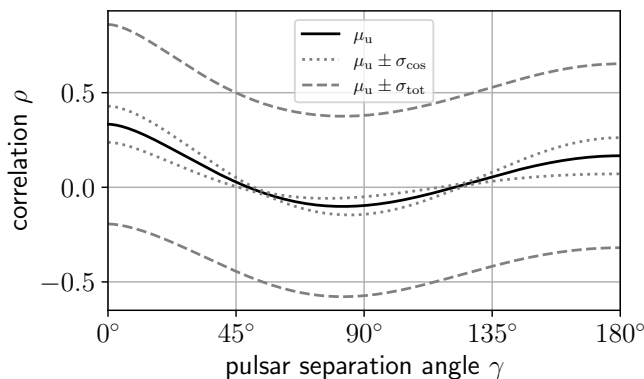


FIG. 1. The mean Hellings and Downs correlation  $\mu_u$  as a function of the angular separation  $\gamma$  between a pair of pulsars. Also plotted are  $\mu_u \pm \sigma_{\text{tot}}$  and  $\mu_u \pm \sigma_{\text{cos}}$ , where  $\sigma_{\text{tot}}^2$  is the total variance for a single pulsar pair and  $\sigma_{\text{cos}}^2$  is the cosmic variance. (For this plot, we have set  $\hbar^4/h^4 = 1/2$  and  $h^2 = 1$ ; see text for details of the GW source model.) Note: although the correct name is “standard deviation”, we sometimes call  $\sigma$  the “variance”.

Figure 1 illustrates the total and cosmic variance, comparing them to the mean  $\mu_u$  of the Hellings and Downs correlation. The total variance at angle  $\gamma$  is the uncertainty associated with the determination of the Hellings and Downs correlation when a single (randomly selected) pulsar pair at angle  $\gamma$  is used to estimate that correlation. The cosmic variance at angle  $\gamma$  is the uncertainty that remains when large numbers of uniformly distributed pulsar pairs at angle  $\gamma$  are used to do the estimation. *In this paper, we study the transition between these two limits, when a finite set of pulsars at specific sky locations are used to estimate the correlation.* This reflects observational reality, because PTA pulsars are nonuniformly distributed on the sky, and the pairs formed from them have no separation angles in common.

The plots in Fig. 1 assume that the gravitational waves arise from the incoherent sum of many weak sources, giving rise, via the central-limit theorem, to a Gaussian ensemble. For such sources, the scaling relation between the (squared) mean and the variance is described in App. B; these plots take  $\hbar^4/h^4 = 1/2$  and  $h^2 = 1$ . This corresponds to the large source-number limit of the “narrowband” discrete confusion noise model of [9] where the source frequency is assumed to be commensurate with the inverse observation time.

### B. Variance of the Hellings and Downs correlation for an arbitrary set of pulsars

In this paper, we assume that the reader is familiar with Ref. [9], and extend that analysis in three ways:

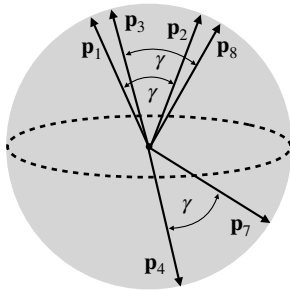


FIG. 2. Three pulsar pairs with angular separation  $\gamma$ . The naive correlation estimator  $\rho_{\text{naive}} = (\rho_{12} + \rho_{38} + \rho_{47})/3$  weights all pulsar pairs equally, but is suboptimal. The optimal estimator  $\rho_{\text{opt}} = 0.3\rho_{12} + 0.3\rho_{38} + 0.4\rho_{47}$  gives more weight to the correlation  $\rho_{47}$ . This is because the 12 and 38 pairs are close on the sky, so  $\rho_{12}$  and  $\rho_{38}$  give nearly redundant estimates. (Note: if there are a large number of pairs, uniformly distributed about the sky, then rotational symmetry implies that all have the same weight.)

First, we place a finite number of pulsar pairs, all separated by the *same* angle  $\gamma$ , at arbitrary sky positions. This allows a more precise estimate of the Hellings and Downs correlation  $\rho$  at angle  $\gamma$  than is possible with only one pulsar pair. In this situation, the only sensible way we have found to define the Hellings and Downs correlation is as a weighted average of the measured pulsar pair correlations  $\rho_{ab}$ , where  $a$  and  $b$  label the pulsars. We solve an optimization problem to find the best choice of weights, which depend upon the relative sky locations of the pulsar pairs as shown in Fig. 2. In the limit of low pulsar noise, these weights tell an observer (in a universe described by the Gaussian ensemble) the optimal way to combine the measured pulsar-pair correlations.

We have normalized this optimal estimator to agree with the single-pulsar-pair correlation, because it averages the individual pulsar-pair correlation measurements, rather than summing them. Not surprisingly, when there is more than a single pulsar pair, the variance in the Hellings and Downs correlation is reduced below the total variance given in [9], which (as explained above) is computed for a single pulsar pair.

But what the reader may not expect is that as the number of pairs grows, this variance does not decrease to zero. Instead, as the number of pairs grows, the variance approaches the cosmic variance, as (later) illustrated in Fig. 4. We examine in detail how this limit is approached, and how much the pulsar variance is reduced for a particular set of pulsars on the sky.

Our approach to deriving the cosmic variance is different than that given in [9], providing a useful alternative interpretation. We construct the weights by requiring that the resulting estimator of the Hellings and Downs correlation is (1) unbiased and (2) has minimum variance. That minimum variance *defines* the total variance for the given set of pulsar pairs, and approaches the cosmic variance in the limit of many pulsar pairs. It does not approach zero in this limit, because the measurements between different pulsar pairs are correlated with one another. Once there are sufficient pulsar pairs at a given angular separation, including more pairs at that angle does not provide additional information about the GW-induced correlations.

Our second extension to [9] is for the (realistic) case where none of the pulsar pairs have exactly the same angular separation  $\gamma$ . This extends the multi-pulsar-pair treatment just described, which assumes that all of the pulsar pairs have exactly the same angular separation  $\gamma$ . In practice, with real pulsars in a real analysis, this is never the case. One must split the range of  $\gamma$  into a discrete set of nonzero width angular bins, and carry out a “binned” analysis. This binned analysis follows a similar approach to the previous case, which (starting from here) we refer to as the “unbinned” case. We (1) find the optimal choice of weights for combining the correlation measurement within a given bin, and then (2) compute the variance of that estimator.

To illustrate this, Fig. 3 shows the variance that would be obtained by using all of the pulsars currently employed by the three major PTAs. If the Gaussian ensemble is a good description of the GWs in our Universe, then this shows the expected deviations away from the Hellings and Downs mean, if the correlation measurements are free of experimental or intrinsic pulsar noise. Thus, it is an upper limit on the ability to which that set of 88 pulsars could be expected to recover the Hellings and Downs curve. Similar plots for the three individual PTAs (but employing only “their” pulsars) are given in Sec. VI B.

Our third extension to [9] is to also consider the effects of noise. While most of our analysis assumes noise-free measurements, we also explain how to include noise contributions to the pulsar-pair covariance matrix. These are fundamentally different than the GW-induced timing correlations. For the GW-induced correlations, as the number of pulsar pairs increases, the variance in the Hellings and Downs correlation approaches the (nonzero) cosmic variance. In contrast, if the different pulsars have independent noise, then the contribution of that noise to the Hellings and

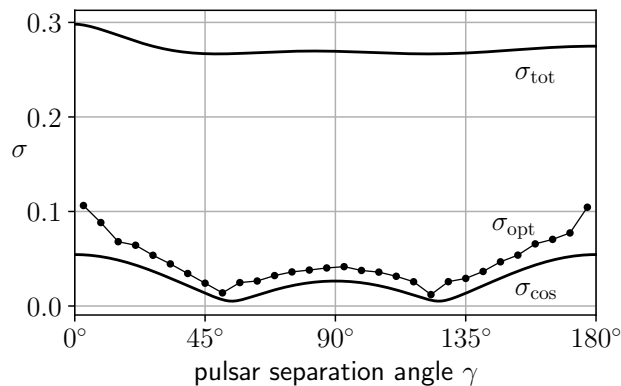


FIG. 3. The dots (lines are to guide the eye) show the predicted GW-induced variance  $\sigma_{\text{opt}}^2$  for the current set of 88 IPTA pulsars (assumed free of noise) after optimally combining timing-residual correlations in  $30 \times 6^\circ$  angular separation bins. This optimal averaging reduces the variance far below the single pulsar pair (total) variance  $\sigma_{\text{tot}}^2$ , bringing it close to the cosmic variance  $\sigma_{\text{cos}}^2$ . The dots indicate the scale of the expected fluctuations away from the Hellings and Downs curve, for a Gaussian ensemble GW background with the binary inspiral spectrum described in App. B. This plot has the same scale and assumptions as the IPTA plot in Fig. 7, i.e.,  $\alpha = 1$  for timing residuals and  $\hat{h}^2 \approx 0.4$ . See Secs. VIB and F for more details.

Downs correlation *vanishes* as the number of pulsar pairs increases. Thus, it is reasonable to expect that as PTAs add additional pulsars and improve their measurements, the limit of the cosmic variance can be achieved in practice.

### C. Outline

Here, we give a brief outline of the paper, and introduce the most important notation. Note that our methods and results apply both to redshifts (set  $\alpha = 0$ ) and to timing residuals (set  $\alpha = 1$ ). See App. A for details.

We begin in Sec. II by defining the time-averaged correlation  $\rho_{ab}$  between the redshifts of two pulsars  $a$  and  $b$ , where the subscript denotes the pulsar pair  $ab$ . For the Gaussian ensemble, in Sec. II A we calculate the first moment  $\langle \rho_{ab} \rangle$  of the correlation. For two distinct pulsars ( $a \neq b$ ) this is proportional to the Hellings and Downs curve  $\mu_{\text{u}}(\gamma_{ab})$  given in (1.1), where  $\gamma_{ab}$  denotes the angle between the directions to the two pulsars. To gracefully accommodate the case  $a = b$ , for which the pulsar term doubles the correlation, we also define the closely related quantity  $\mu_{ab}$ . In Sec. II B, we use this to compute the second moment  $\langle \rho_{ab} \rho_{cd} \rangle$ . From the two moments, we obtain the covariance matrix  $C_{ab,cd} \equiv \langle \rho_{ab} \rho_{cd} \rangle - \langle \rho_{ab} \rangle \langle \rho_{cd} \rangle$ , where  $cd$  denotes a second pulsar pair constructed from pulsars  $c$  and  $d$ .

Note that in this paper, angle brackets  $\langle Q \rangle$  denote the average of a function or functional  $Q$ . This average is always normalized so that  $\langle 1 \rangle = 1$ . If not explicitly stated otherwise, then this is an average over the Gaussian ensemble of GW realizations. However, in some places we (state that we) use it differently. For example, in (7.1) it denotes an average over a Gaussian ensemble of noise sources *and* GW sources. In (4.8) it denotes an average over pulsar pairs at fixed separation angles (indicated by the subscript  $cd, \gamma_k$  on the rightmost angle bracket).

The covariance matrix is the most important quantity in our analysis, and used throughout. In Sec. II C, we derive an identity satisfied by  $C_{ab,cd}$ , which is used later to obtain the cosmic variance limit. To further simplify the notation, we sometimes drop the indices entirely and write it as a (square, real, positive-definite symmetric) matrix  $C$ .

We then turn our attention to defining the Hellings and Downs correlation for a collection of pulsars. In Sec. III, we define it as the optimal estimator of the correlation. This is a weighted sum of pulsar-pair correlations in a (zero- or nonzero-width) angular bin. In Sec. III A, we derive the weights  $w_{ab}$  needed to form that optimal estimator by requiring that the estimator be unbiased and minimize the variance. This derivation is general, and applies both to the unbinned case (where the pulsar pairs all have the same angular separation) or the binned case (where they do not). In Sec. III B, we specialize to the unbinned case, and in Sec. III C, we show how the variance simplifies if the vector  $\mathbb{1}$  containing all ones is an eigenvector of the covariance matrix  $C$ .

Using this framework, we now study the smooth decrease of the variance as the number of pulsar pairs is increased from a single pair to many pairs. (In [9], the variances for these two extremes are called the ‘total variance’ and the ‘cosmic variance’ respectively.) We begin in Sec. IV with the unbinned case. We examine the limit as the number of pulsar pairs becomes large, showing that the variance of the optimal estimator converges to the cosmic variance computed in [9].

The limit is obtained by showing that for large numbers of pulsar pairs uniformly distributed on the sky,  $\mathbb{1}$  becomes

an eigenvector of the unbinned covariance matrix  $C$ . Its eigenvalue is simply related to the average value of the entries of  $C$ . For finding this average value, particularly in the binned case, it is helpful to write  $C$  in “block form” with blocks corresponding to the different angular bins. In Sec. IV and the related Appendices D and E, we use single indices (i.e.,  $j$  and  $k$ ) to denote these angular correlation bins. Then,  $C_{jk}$  denotes a sub-matrix (we use the term “block”) of the covariance matrix, whose rows/columns correspond to the pairs in the  $j$ ’th/ $k$ ’th angular bins.

In Sec. V, we examine the unbinned case numerically, showing that as pulsar pairs are added, the optimal estimator decreases smoothly from the single-pair variance to the cosmic variance.

In Sec. VI, we move on to the more realistic “binned” case. Since the bins have nonzero width in  $\gamma$ , some angular resolution is lost. But we show in Sec. VIA that for some ranges of angular separation  $\gamma$ , this allows us to reduce the variance *below* the cosmic variance. Then, in Sec. VIB, we apply the binned formalism to the sets of pulsars currently used by the three major PTAs. With plots, we show how much these pulsar sets could reduce pulsar variance, illustrating (for the ideal noiseless case) how closely they approach the cosmic variance.

In Sec. VII, we examine the effects that noise and errors might have on the optimal estimator. Previously we neglected all sources of noise. In Sec. VIIA, we extend the calculation of the covariance matrix to include noise contributions, and we show in Sec. VIB that if the noise is uncorrelated between pulsars, then for large numbers of pulsar pairs, its effect vanishes. The argument is presented in detail in App. C.

Our methodology could be applied to any source of GWs. However, for much of the paper, we assume that the GW background arises from a Gaussian ensemble, which implies a special form for the covariance matrix. In Sec. VIII, we explain the consequence: our predictions for the cosmic variance create a consistency test. If the Hellings and Downs correlation (estimated according to the recipe we present) matches the Hellings and Downs prediction much more (or much less) closely than the variance we compute, then it is very unlikely that our Universe has a GW background which is described by the Gaussian ensemble. This shows that the Hellings and Downs variance is an observable quantity, since it could (in principle or practice) be used to falsify this standard model of the GW background.

This is followed by a short conclusion in Sec. IX and a number of technical appendices, App. A – App. F. At the start of the Appendix, we give an outline of their contents.

## II. STATISTICS OF PULSAR-PAIR REDSHIFT CORRELATIONS

Consider a pulsar timing array built from a collection of  $M$  distinct pulsars. We label the pulsars with indices  $a$ ,  $b$ ,  $c$ , and  $d$ , which take values in the range  $1, 2, \dots, M$ . Associated with each pulsar we have pulse redshift measurements  $Z_a(t)$  which are a function of time. Given a pair of pulsars  $a$  and  $b$ , their time-averaged redshift correlation is

$$\rho_{ab} \equiv \overline{Z_a Z_b}, \quad (2.1)$$

where the overline denotes a time average between  $-T/2$  and  $T/2$ , where  $T$  is the total observation period. For simplicity we take  $T$  to be the same for all pulsars. Note that our calculations, formalism, results and plots also apply to pulsar timing residual correlations (as well as to redshifts) as explained in App. A.

Our main objects of analysis are pulsar pairs, labeled by  $a$  and  $b$ ; we denote the angles between their sky directions by  $\gamma_{ab}$ . The total number of distinct pulsar pairs is  $N_{\text{pairs,tot}} \equiv M(M-1)/2$ . We can index these distinct pairs by  $ab$  with  $a < b$ , so  $a = 1, 2, \dots, M-1$  and  $b = a+1, a+2, \dots, M$ . However (unless stated otherwise) our formulae also allow for the case  $a = b$ .

Many calculations only include pulsar pairs that lie within a bin (a small angular interval) about  $\gamma$ . These pairs are not described by any simple indexing system, but should rather be thought of as a collection of discrete ordered pairs of integers,  $ab, cd, \dots, ef$ . For example, Fig. 2 shows three pulsar pairs  $\{1, 2\}$ ,  $\{3, 8\}$ , and  $\{4, 7\}$  that lie in the  $\gamma \in [30^\circ, 31^\circ]$  angular interval. So, when labeling or indexing pulsar pairs with angular separation  $\gamma$ , we will use  $n_{\text{pairs}}$  (which in general depends upon  $\gamma$ , and in the example has value  $n_{\text{pairs}} = 3$ ) to denote the number of pairs. Then, we treat  $ab$  as a *single* index for those pairs, running over the range  $ab = 1, 2, \dots, n_{\text{pairs}}$ .

We now compute the first moment  $\langle \rho_{ab} \rangle$  of the correlation  $\rho_{ab}$  and its covariance matrix, which is defined by the combination of first and second moments

$$C_{ab,cd} \equiv \langle \rho_{ab} \rho_{cd} \rangle - \langle \rho_{ab} \rangle \langle \rho_{cd} \rangle. \quad (2.2)$$

The angle brackets denote the average over a Gaussian ensemble. As shown in [9], this ensemble corresponds to an infinite collection of weak, unpolarized, time-stationary GW sources, distributed uniformly in space. In any narrow frequency band, there are an infinite number of sources radiating at indistinguishably close GW frequencies, but with different (random) phases. This generates “confusion noise”, giving rise via the central-limit theorem to a stationary and Gaussian stochastic process.

### A. First moment

For the unpolarized isotropic stationary Gaussian ensemble, App. C from [9] gives the first moment of the pulsar- $a$  pulsar- $b$  redshift correlation as

$$\langle \rho_{ab} \rangle = h^2 \mu_{ab}, \quad (2.3)$$

where

$$\mu_{ab} \equiv \mu_{\text{u}}(\gamma_{ab}) + \delta_{ab} \mu_{\text{u}}(0). \quad (2.4)$$

Here,  $\delta_{ab}$  is the Kronecker delta, which vanishes if pulsars  $a$  and  $b$  are distinct, and is unity if they are identical. The function  $\mu_{\text{u}}(\gamma_{ab})$  is the Hellings and Downs curve given by (1.1). Its fundamental underlying definition is [2]

$$\mu_{\text{u}}(\gamma_{ab}) \equiv \frac{1}{4\pi} \int d\Omega \sum_A F_a^A(\Omega) F_b^A(\Omega). \quad (2.5)$$

This is an average of the products of antenna pattern functions  $F_a^A(\Omega)$  and  $F_b^A(\Omega)$  over source directions, where  $\Omega$  denotes a unit vector on the two-sphere and  $A = +, \times$  labels two orthogonal polarization states of the GWs. Definitions of the antenna pattern functions and a full derivation of (1.1) starting from (2.5) are given in App. D of [9].

The function  $\mu_{\text{u}}(\gamma_{ab})$  is the sky-averaged correlation for *distinct* pulsars  $a$  and  $b$  separated by angle  $\gamma_{ab}$ , taking only Earth terms into account. The delta function which appears in (2.4) for time-stationary GW signals is discussed in detail in App. C.2 of [9]. It arises because the expected correlation for a pair of *identical* pulsars  $a = b$  is twice the expected correlation for a pair of *distinct* pulsars  $a$  and  $b$  which lie along the same line of sight, but are at different distances [10] from Earth. If pulsars  $a$  and  $b$  are identical and the GW sources are (statistically) stationary in time, then the pulsar-pulsar term in the correlation has the same expectation value as the Earth-Earth term, which produces a factor of two.

The overall scale of the correlation  $\langle \rho_{ab} \rangle$  is determined by the constant

$$h^2 \equiv 4\pi \int_{-\infty}^{\infty} df H(f) (2\pi f)^{-2\alpha}. \quad (2.6)$$

Set  $\alpha = 0$  for redshift correlations, or set  $\alpha = 1$  for timing residual correlations, as explained in App. A. For redshifts,  $h$  is a dimensionless strain, whereas for timing residuals,  $h$  has units of time  $\times$  (dimensionless) strain.

The real two-sided squared-strain spectral density  $H(f)$  is related to the one-sided GW energy-density spectrum  $\Omega_{\text{gw}}(f)$  or to the characteristic strain spectrum  $h_c^2(f)$  via

$$H(f) = \frac{3H_0^2}{32\pi^3} \frac{1}{|f|^3} \Omega_{\text{gw}}(|f|) = \frac{1}{16\pi} \frac{1}{|f|} h_c^2(|f|), \quad (2.7)$$

with  $H(f) \geq 0$  and  $H(f) = H(-f)$ , as shown in [9, 11].

Note that the derivation of (2.3) assumes that the spectral function  $H(f)$  has a coherence time / length which is much less than the typical pulsar-pulsar and Earth-pulsar distances, so that the Earth-pulsar terms which appear in the correlation can be neglected. This should be the case for the expected PTA sources.

### B. Second moment and covariance matrix

For a Gaussian ensemble, the second moment of the pulsar-pair redshift correlation can be found following the methods in App. C.3 of [9]. In that reference, the ensemble average  $\langle \rho_{12} \rho_{12} \rangle$  is evaluated, where 1 and 2 refer to distinct pulsars. This evaluates to three terms that arise from Isserlis' theorem [12]. (This theorem expresses the four-point function for zero-mean Gaussian variables as a sum of three products of two-point functions.) A calculation identical to that in [9], but using four arbitrary pulsars  $a, b, c$ , and  $d$  (which might or might not be the same) gives three terms:

$$\begin{aligned} \langle \rho_{ab} \rho_{cd} \rangle &= (h^2 \mu_{ab}) (h^2 \mu_{cd}) + h^4 (\mu_{ac} \mu_{bd} + \mu_{ad} \mu_{bc}) \\ &= \langle \rho_{ab} \rangle \langle \rho_{cd} \rangle + h^4 (\mu_{ac} \mu_{bd} + \mu_{ad} \mu_{bc}). \end{aligned} \quad (2.8)$$

The second line of (2.8) is obtained from the first line by using (2.3).

The factor which appears on the rhs of (2.8) sets the scale of the covariance, and is

$$\hbar^4 = (4\pi)^2 \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \operatorname{sinc}^2(\pi(f-f')T) H(f)H(f')(4\pi^2 f f')^{-2\alpha}. \quad (2.9)$$

The constant  $\alpha$  is explained in App. A. For redshift correlations,  $\alpha = 0$  and  $\hbar$  is a dimensionless strain, whereas for timing residual correlations,  $\alpha = 1$  and  $\hbar$  has units of time  $\times$  (dimensionless) strain. Note that since  $H(f) \geq 0$  and  $0 \leq \operatorname{sinc}^2\phi \leq 1$ , it follows immediately from (2.6) and (2.9) that for  $T > 0$  one has  $\hbar^4 < \hbar^4$ .

To compare the variance (proportional to  $\hbar^4$ ) to the squared mean (proportional to  $\hbar^4$ ) we must have a value for their ratio, which depends upon the spectrum of the GW sources. We evaluate this ratio for several models in App. B, and it is also discussed in App. C of [9]. For example, for a Gaussian ensemble of binary inspiral sources, the ratio  $\hbar^2/h^2$  is shown in Fig. 8.

The covariance matrix is defined by (2.2). For the Gaussian ensemble, making use of (2.8) we obtain

$$C_{ab,cd} = \hbar^4 (\mu_{ac}\mu_{bd} + \mu_{ad}\mu_{bc}). \quad (2.10)$$

As a sanity check, consider the special case computed in App. C.3 of Ref. [9], for which the pulsars are  $a = 1, b = 2, c = 1, d = 2$ . For this case, (2.10) simplifies to give the total (pulsar plus cosmic) variance of  $\rho_{ab}$

$$C_{ab,ab} = \sigma_{\text{tot}}^2 = \hbar^4 (\mu_u^2(\gamma_{ab}) + 4\mu_u^2(0)). \quad (2.11)$$

This assumes that the pulsars are distinct, so that  $a \neq b$ .

In Secs. IV and V, we will be interested in the case where the angle  $\gamma$  between the  $ab$  pulsar pair is the same as the angle  $\gamma$  between the  $cd$  pulsar pair. However, the expression (2.10) for the covariance matrix is completely general. The pulsars  $a, b, c$  and  $d$  can be freely chosen: they may have arbitrary angular separations, and any or all of them may be identical or distinct.

### C. Useful identities

For later use, we provide two useful identities involving the covariance matrix  $C_{ab,cd}$ . The first follows by using (2.4) to expand the rhs of (2.10) for  $C_{ab,cd}$ , giving

$$\begin{aligned} \mu_{ac}\mu_{bd} + \mu_{ad}\mu_{bc} &= \mu_u(\gamma_{ac})\mu_u(\gamma_{bd}) + \mu_u(\gamma_{ad})\mu_u(\gamma_{bc}) + \mu_u^2(0) [\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}] \\ &+ \mu_u(0) [\mu(\gamma_{bd})\delta_{ac} + \mu_u(\gamma_{ac})\delta_{bd} + \mu_u(\gamma_{bc})\delta_{ad} + \mu_u(\gamma_{ad})\delta_{bc}]. \end{aligned} \quad (2.12)$$

The second identity provides an alternative form for the first two terms on the rhs of (2.12). Making use of (2.5), the products of Hellings and Downs curves may be written in terms of antenna pattern functions as

$$\begin{aligned} &\mu_u(\gamma_{ac})\mu_u(\gamma_{bd}) + \mu_u(\gamma_{ad})\mu_u(\gamma_{bc}) \\ &= \sum_A \sum_{A'} \int \frac{d\Omega}{4\pi} \int \frac{d\Omega'}{4\pi} \left( F_a^A(\Omega)F_c^A(\Omega)F_b^{A'}(\Omega')F_d^{A'}(\Omega') + F_a^A(\Omega)F_d^A(\Omega)F_b^{A'}(\Omega')F_c^{A'}(\Omega') \right). \end{aligned} \quad (2.13)$$

We will use these identities in Sec. IV to calculate the average value of a row of the correlation matrix, when there are many pulsar pairs distributed uniformly on the sky. We use this to prove that in the limit of many such pulsar pairs, the variance of the optimal estimator converges to the cosmic variance first calculated in [9].

## III. OPTIMAL ESTIMATION OF THE HELLINGS AND DOWNS CORRELATION

Suppose that we have  $M$  pulsars at arbitrary sky positions, creating  $M(M-1)/2$  distinct pulsar pairs. We wish to combine the measurements of their redshift correlations  $\rho_{ab}$  with weights  $w_{ab}$  to get the best possible estimate of the Hellings and Downs correlation at separation angle  $\gamma$ . Here, by ‘‘best possible estimate’’ we mean the assignment of weights to the pulsar pairs measurements that *minimizes* the variance of an *unbiased* estimator for (our choice of) some statistical ensemble of universes. We show below how these two conditions determine the form of the weights.

The main complication, as illustrated in Fig. 2, is that the optimal choice of weights depends upon the relative sky locations of *all* the pulsars in the array. This is because the correlation between two pulsar pairs  $ab$  and  $cd$  depends on the angular separation between the pairs  $ac, bd, ad$  and  $bc$ , as indicated in (2.8) and (2.10). The solution to the

optimization problem must account for the nonzero correlations between redshift correlation measurements from all of these different pairings.

A second complication arises because, in practice, no pulsar pair will have a separation angle exactly equal to  $\gamma$ . We need to split the range of angular separations into a discrete set of nonzero-width bins, and combine the (subset of) measured correlations for the pulsar pairs that fall into each bin. How to optimally combine the data for this general case is described in Sec. III A.

We also consider the hypothetical ‘‘unbinned’’ case, corresponding to multiple pulsar pairs with exactly the same separation angle. This is a special case of the binned analysis, corresponding to the limit of infinitesimally narrow bins. The method for optimally combining such data is given in Sec. III B.

### A. Derivation of the optimal binned estimator and its variance

Here, we derive the optimal way to combine pulsar-pair correlation measurements  $\rho_{ab}$ . We consider the general case where the separation angles  $\gamma_{ab}$  are distributed within a nonzero-width bin centered on  $\gamma$ . To simplify the equations, we use vector/matrix notation, where the dimension of the vectors and (square) matrices is the number  $n_{\text{pairs}}$  of pulsar pairs in the angular correlation bin. For example,  $\rho \equiv \rho_{ab}$  denotes the column vector of redshift correlation measurements, and  $C \equiv C_{ab,cd}$  denotes the real, positive-definite symmetric matrix whose elements are the covariances between  $\rho_{ab}$  and  $\rho_{cd}$ .

The optimal estimator is a linear combination

$$\rho_{\text{opt}} = \rho^\top w = \sum_{ab=1}^{n_{\text{pairs}}} \rho_{ab} w_{ab}, \quad (3.1)$$

where  $\rho$  is a column vector of the individual pulsar-pair correlation measurements  $\rho_{ab}$ , and  $w$  is a column vector of corresponding dimensionless weights  $w_{ab}$ . Here,  $\rho^\top$  denotes the transpose of the column vector  $\rho$ . (The second equality gives an explicit expression for  $\rho_{\text{opt}}$  in terms of the vector components. For subsequent formulae, we only give the more compact vector/matrix equations.)

We start by calculating the variance of the optimal estimator:

$$\sigma_{\text{opt}}^2 \equiv \langle \rho_{\text{opt}}^2 \rangle - \langle \rho_{\text{opt}} \rangle^2, \quad (3.2)$$

where angle brackets denote the expectation value for the Gaussian ensemble used in e.g., (2.3) and (2.8). Substituting (3.1) into (3.2), we obtain

$$\sigma_{\text{opt}}^2 = w^\top \left( \langle \rho \rho^\top \rangle - \langle \rho \rangle \langle \rho^\top \rangle \right) w = w^\top C w = (C w, C w), \quad (3.3)$$

where the second equality follows from the definition (2.2) of the covariance matrix  $C \equiv C_{ab,cd}$ . For the third equality we introduce an inner product of column vectors  $A$  and  $B$  via

$$(A, B) = A^\top C^{-1} B, \quad (3.4)$$

and use the property that  $C$  is a symmetric matrix, so  $C^\top = C$ .

To completely determine the weights, we need a normalization constraint. Allowing for a nonzero-width angular correlation bin, the expected value of the estimator follows immediately from (2.3) and (3.1), as

$$\langle \rho_{\text{opt}} \rangle = h^2 \mu^\top w = h^2 (\mu, C w), \quad (3.5)$$

where  $\mu$  denotes the column vector with components  $\mu_{ab} = \mu_u(\gamma_{ab})$ , assuming  $a \neq b$ . Now, we must make an explicit and somewhat arbitrary choice: what value should we pick for the expected value of the Hellings and Downs correlation estimator  $\langle \rho_{\text{opt}} \rangle$  for this particular angular bin? This choice effectively defines what it means for the optimal estimator to be unbiased.

Using the subscript ‘‘norm’’ to denote this somewhat arbitrary normalization constraint, the mean value of our estimator should correspond to the expected Hellings and Downs correlation

$$\langle \rho_{\text{opt}} \rangle \equiv \rho_{\text{norm}} = h^2 \mu_u(\gamma_{\text{norm}}) \equiv h^2 \mu_{\text{norm}} \quad (3.6)$$

for *some* choice of angle  $\gamma_{\text{norm}}$  within our bin, corresponding to a specific value of  $\mu_{\text{norm}} \equiv \mu_u(\gamma_{\text{norm}})$ . But what choice is most sensible? Three reasonable options are:



- a)  $\mu_{\text{norm}} = \mu_{\text{u}}(\gamma_{\text{norm}})$ , where  $\gamma_{\text{norm}}$  is the central angle of the bin. In practice, such bin centers are typically set in advance of any analysis, as evenly spaced “round numbers”.
- b)  $\mu_{\text{norm}} = \mu^{\top} \mathbb{1} / n_{\text{pairs}}$ , which is a uniform average of the expected Hellings and Downs correlation values for the different pulsar pairs in that bin. Here,  $\mathbb{1}$  denotes a column vector of dimension  $n_{\text{pairs}}$  containing ones:

$$\mathbb{1} \equiv (1, 1, \dots, 1)^{\top}. \quad (3.7)$$

- c)  $\mu_{\text{norm}} = \mu_{\text{u}}(\gamma_{\text{norm}})$ , where  $\gamma_{\text{norm}} = \gamma^{\top} \mathbb{1} / n_{\text{pairs}}$  is the mean angular separation of the pulsar pairs that lie within the bin. Here,  $\gamma$  denotes a column vector of the values  $\gamma_{ab}$ .

We will see that this arbitrary choice only enters the mean and the (square root of the) variance via an overall scale factor. So, while there may be other justifiable choices which are not listed above, we will see that the choice has no effect on the most important quantity: the fractional uncertainty  $\sigma_{\text{opt}} / \rho_{\text{opt}}$ .

For any of these arbitrary normalization choices, we can now identify the weights that define the optimal (minimum variance) estimator. Since the expected value  $\langle \rho_{\text{opt}} \rangle = \rho_{\text{norm}} = h^2 \mu_{\text{norm}}$  is independent of the weights  $w$ , we can divide  $\sigma_{\text{opt}}^2$  by the square of this quantity before minimization over  $w$ . Hence, we select  $w$  to minimize

$$\frac{\sigma_{\text{opt}}^2}{\rho_{\text{norm}}^2} = \frac{\sigma_{\text{opt}}^2}{\langle \rho_{\text{opt}} \rangle^2} = \frac{(Cw, Cw)}{h^4 (\mu, Cw)^2}, \quad (3.8)$$

where we have used (3.3) and (3.5).

The inner product defined by (3.4) is positive definite and obeys the Schwarz inequality. Hence, the denominator of (3.8) is maximum when the vectors  $\mu$  and  $Cw$  are parallel to (and hence proportional to) one another. This implies that  $Cw = q\mu$  for some  $q$ , or equivalently  $w = qC^{-1}\mu$ . Substituting this into the normalization condition (3.5) then lets us solve for  $q$  and completely determines the optimal weights:

$$h^2(\mu, q\mu) = h^2 \mu_{\text{norm}} \quad \Longrightarrow \quad q = \frac{\mu_{\text{norm}}}{(\mu, \mu)} \quad \Longrightarrow \quad w = \frac{\mu_{\text{norm}}}{\mu^{\top} C^{-1} \mu} C^{-1} \mu. \quad (3.9)$$

In the final equality we have used the definition of the inner product (3.4).

From the optimal weights given in (3.9), the binned optimal estimator and its variance are then given by (3.1) and (3.3) as

$$\rho_{\text{opt}} = \mu_{\text{norm}} \frac{\rho^{\top} C^{-1} \mu}{\mu^{\top} C^{-1} \mu}, \quad \text{and} \quad (3.10)$$

$$\sigma_{\text{opt}}^2 = \frac{\mu_{\text{norm}}^2}{\mu^{\top} C^{-1} \mu}. \quad (3.11)$$

As discussed after (3.7), the ratio  $\rho_{\text{opt}} / \sigma_{\text{opt}}$  is independent of the arbitrary choice  $\mu_{\text{norm}}$ .

While this paper is specifically focused on GW backgrounds that are described by a Gaussian ensemble, the results of this section apply to any GW ensemble. This is because any ensemble has a covariance matrix, from which one can define the (best estimate of the) pulsar-pair correlation (3.10) and the total variance (3.11). So, this way of defining the Hellings and Downs correlation for some set of pulsar pairs, and computing its variance, is perfectly general.

## B. Simplifications for unbinned correlations

The above formulae simplify for the case of unbinned correlation measurements  $\rho_{ab}$ , where the pulsar pairs  $ab$  all have exactly the same angular separation  $\gamma = \gamma_{ab}$ . For this case, the correlation measurements  $\rho_{ab}$  all have the same expected value,  $\langle \rho_{ab} \rangle = h^2 \mu_{ab} = h^2 \mu_{\text{u}}(\gamma)$ , which in vector notation reads

$$\langle \rho \rangle = h^2 \mu = h^2 \mu_{\text{u}}(\gamma) \mathbb{1}. \quad (3.12)$$

This implies that the three “different” reasonable choices for the normalization condition described in Sec. III A all correspond to the same condition:

$$\mu_{\text{norm}} = \mu_{\text{u}}(\gamma) \quad \Leftrightarrow \quad \mathbb{1}^{\top} w = 1. \quad (3.13)$$

For this case, the optimal weights (3.9), optimal estimator (3.10), and optimal variance (3.11) are

$$w = \frac{C^{-1} \mathbb{1}}{\mathbb{1}^\top C^{-1} \mathbb{1}}, \quad (3.14)$$

$$\rho_{\text{opt}} = \frac{\rho^\top C^{-1} \mathbb{1}}{\mathbb{1}^\top C^{-1} \mathbb{1}}, \quad \text{and} \quad (3.15)$$

$$\sigma_{\text{opt}}^2 = \frac{1}{\mathbb{1}^\top C^{-1} \mathbb{1}}. \quad (3.16)$$

The denominator of (3.16) for  $\sigma_{\text{opt}}^2$  is the inverse covariance matrix summed over all of its elements. This is called the ‘‘grand sum’’ of  $C^{-1}$ . Note that these are standard expressions for the optimal combination of correlated measurements which have the same expectation value, and may be found in, e.g., [8, 13, 14].

An important special case is that of a single pulsar pair  $ab$ . For this, the optimal estimator of the correlation is  $\rho_{ab}$ . The optimal variance for this single pair is  $\sigma_{\text{opt}} = C_{ab,ab}$ , which is given in (2.11). As expected, this is precisely the (one-pulsar-pair) total variance  $\sigma^2(\gamma)$  derived in App. C of Ref. [9] for the Gaussian ensemble.

### C. Simplifications for unbinned correlations when $\mathbb{1}$ is an eigenvector of $C$

We continue to restrict attention to the case where the pulsar pairs  $ab$  all have the same angular separation  $\gamma = \gamma_{ab}$ . We obtain a further simplification of (3.14)–(3.16) if the sum of every row of the covariance matrix has the same value  $\lambda$  as the sum of any other row. This means that  $\sum_{cd} C_{ab,cd} = \lambda$  is independent of the row index  $ab$ . Equivalently, the condition may be written in matrix notation as

$$C \mathbb{1} = \lambda \mathbb{1}, \quad (3.17)$$

meaning that  $\mathbb{1}$  is an eigenvector of  $C$  with eigenvalue  $\lambda$ . For this case, the variance of the optimal estimator, (3.16), can be expressed in terms of the average value of the entries of  $C$ .

To see this, note that since  $C$  is real, symmetric, and positive definite, the eigenvalue  $\lambda$  must be real and positive. Now multiply (3.17) from the left by  $\lambda^{-1} C^{-1}$  to obtain

$$C^{-1} \mathbb{1} = \lambda^{-1} \mathbb{1}. \quad (3.18)$$

This means that  $\mathbb{1}$  is also an eigenvector of  $C^{-1}$  with a real positive eigenvalue  $\lambda^{-1}$ . So we can take the dot product of (3.18) on the left with  $\mathbb{1}^\top$ , leading to

$$\mathbb{1}^\top C^{-1} \mathbb{1} = \mathbb{1}^\top \lambda^{-1} \mathbb{1} = n_{\text{pairs}}/\lambda, \quad (3.19)$$

where the number of pairs has appeared because it is the dimension of the vector  $\mathbb{1}$ .

So, with these assumptions (3.16) simplifies to

$$\sigma_{\text{opt}}^2 = \lambda/n_{\text{pairs}} \equiv s, \quad (3.20)$$

where  $s$  is the average value of the entries of  $C$ . In addition, the properly normalized optimal weights of (3.14) are then given by

$$w = \mathbb{1}/n_{\text{pairs}}. \quad (3.21)$$

In this case, the optimal weights are all equal and (3.15) becomes

$$\rho_{\text{opt}} = \rho^\top \mathbb{1}/n_{\text{pairs}}. \quad (3.22)$$

This is a uniform average of the correlation measurements  $\rho_{ab}$ .

To summarize: if the sum of each row of the covariance matrix gives the same number  $\lambda$ , then the variance of the optimal estimator is simply the average value,  $s \equiv \lambda/n_{\text{pairs}}$ , of the entries of  $C$ , and the optimal weights are all equal.

We will show in Sec. IV that, in the limit of an infinite number of pulsar pairs distributed uniformly on the sky, each row of the covariance matrix  $C$  for the pulsar-pair redshift correlations sums to the same value. Thus, the variance of the optimal estimator may be found from the average value of the entries of  $C$ , and related to the cosmic variance. This follows from symmetry, because with an infinite number of uniformly distributed pairs, any pair at angle  $\gamma$  is equivalent to any other pair at the same angle: the considerations of Fig. 2 do not apply.

### D. Simple example: an infinite set of measurements with nonzero variance

We normally expect that as more and more noise-free measurements are included in some estimate, the uncertainty decreases to zero. However, as we have explained, the Hellings and Downs correlation does not behave this way. In the limit of many measurements, the uncertainty decreases to a nonzero value, which is the cosmic variance.

Here, we provide a simple example to illustrate how correlations between the different measurements (off-diagonal terms in the covariance matrix) are responsible for this behavior. Consider the following  $n_{\text{pairs}} \times n_{\text{pairs}}$  covariance matrix, where the terms proportional to  $a$  arise from correlations between the different measurements:

$$C = \begin{pmatrix} 1 & a & \dots & a \\ a & 1 & \dots & a \\ \vdots & \vdots & \ddots & \vdots \\ a & a & \dots & 1 \end{pmatrix}. \quad (3.23)$$

The parameter  $a$  must lie in the range  $0 \leq a < 1$  in order for  $C$  to be a positive-definite matrix (i.e.,  $v^\top C v \geq 0$ , with equality holding if and only if  $v = 0$ ).

We find the optimal estimator and its variance using the results of Sec. III C. It is easy to verify that  $\mathbb{1}$  is an eigenvector of  $C$  with eigenvalue

$$\lambda = 1 + (n_{\text{pairs}} - 1)a. \quad (3.24)$$

Thus, the optimal estimator is

$$\rho_{\text{opt}} = \rho^\top \mathbb{1} / n_{\text{pairs}}, \quad (3.25)$$

and its variance is

$$\sigma_{\text{opt}}^2 = \frac{\lambda}{n_{\text{pairs}}} = \frac{1 + (n_{\text{pairs}} - 1)a}{n_{\text{pairs}}}. \quad (3.26)$$

Now, one might think that since  $\rho_{\text{opt}}$  is a uniform average of the correlation measurements  $\rho_{ab}$ , then the variance of the optimal estimator should scale like  $1/n_{\text{pairs}}$ , and hence vanish in the limit  $n_{\text{pairs}} \rightarrow \infty$ . But it follows immediately from (3.26) that in the limit of large numbers of measurements the variance approaches

$$\lim_{n_{\text{pairs}} \rightarrow \infty} \sigma_{\text{opt}}^2 = a, \quad (3.27)$$

which is finite.

In the next section, we will see this same behavior. It is the nonzero off-diagonal elements of the covariance that are responsible for the cosmic variance, not unequal weighting of the correlation measurements.

## IV. RELATING THE VARIANCE OF THE OPTIMAL ESTIMATOR TO THE COSMIC VARIANCE

In this section, we show that if there are large numbers of pulsar pairs uniformly distributed around the sky, then the variance (in a narrow bin) of the optimal estimator at angle  $\gamma$  approaches the cosmic variance.

We begin by randomly placing a large number of pulsars uniformly on the sky. For pedagogic clarity we use one million pulsars, so that the total number of pairs is  $N_{\text{pairs,tot}} \approx 5 \times 10^{11}$ . Order the pairs by  $\cos \gamma_{ab} \in [-1, 1]$ , and divide them into (say) 1000 angular bins of width  $d(\cos \gamma) = 0.002$ . On average, each bin will contain  $n_{\text{pairs}} \approx 5 \times 10^8$  pairs, with values of  $\cos \gamma$  differing by at most  $\pm 0.002$ . Label the bins by  $j = 1, 2, \dots, 1000$ , and let  $\gamma_j$  denote the average value of  $\gamma$  in bin  $j$ .

The covariance matrix  $C$  now takes the form shown in (D1), with  $\kappa = 1000$ . It consists of one million blocks. Each block has approximately  $5 \times 10^8$  rows and columns, and hence approximately  $25 \times 10^{16}$  entries.

Pick any one of these blocks, which are identified by “row/column” indices  $j, k$  in the range  $j, k = 1, 2, \dots, 1000$  (see (D1)). In Sec. IV A, we will show that, within a block, the sum of each row is the same (in the limit of large numbers of pairs) as the sum of any other row. We will also show that the sum of any column is the same (in the limit of large numbers of pairs) as the sum of any other column. Finally, we will show that the average value of an entry in any one of these million blocks is given by

$$s_{jk} = h^4 \int_0^\pi d\beta \sin \beta \left( \mu_{++}(\gamma_j, \beta) \mu_{++}(\gamma_k, \beta) + \mu_{\times\times}(\gamma_j, \beta) \mu_{\times\times}(\gamma_k, \beta) \right). \quad (4.1)$$

Here,  $\gamma_j$  and  $\gamma_k$  are the average values of  $\gamma$  for the corresponding row and column, and  $\mu_{AA}(\gamma, \beta)$  are the two-point functions defined in App. G of [9].

From (4.1) and the results of Sec. III B, the desired result follows immediately. The variance of the optimal estimator in the unbinned analysis corresponds to picking  $\gamma_j = \gamma_k$  to have the same value. For the blocks along the diagonal, this is true to a good approximation. So, setting  $j = k$  and  $\gamma_j = \gamma_k \equiv \gamma$  in (4.1), we find that the average value of an entry on one of these diagonal blocks is

$$\begin{aligned} s_{jj} &= 2\hbar^4 \frac{1}{2} \int_0^\pi d\beta \sin \beta \left( \mu_{++}^2(\gamma_j, \beta) + \mu_{\times\times}^2(\gamma_j, \beta) \right) \\ &= 2\hbar^4 \widetilde{\mu}^2(\gamma). \end{aligned} \quad (4.2)$$

The (cosmic variance) function  $\widetilde{\mu}^2(\gamma)$  is computed in App. G of [9]. It is

$$\begin{aligned} \widetilde{\mu}^2(\gamma) &= -\frac{5}{48} - \frac{1}{12}\pi^2 + \left(\frac{49}{432} - \frac{\pi^2}{36}\right) \cos^2 \gamma + \frac{1}{6}(\cos^2 \gamma + 3) \left[ \text{Li}_2\left(\frac{1 - \cos \gamma}{2}\right) + \text{Li}_2\left(\frac{1 + \cos \gamma}{2}\right) \right] \\ &\quad + \frac{1}{12}((\cos \gamma - 1)(\cos \gamma + 3)) \log\left(\frac{1 - \cos \gamma}{2}\right) + \frac{1}{12}((\cos \gamma + 1)(\cos \gamma - 3)) \log\left(\frac{1 + \cos \gamma}{2}\right), \end{aligned} \quad (4.3)$$

where  $\text{Li}_2$  denotes the dilogarithm function.

The variance of the optimal estimator of the Hellings and Downs correlation is given by (3.20) as the average value  $s$  of the entries of the covariance matrix, given by (4.2). Hence, the optimal estimator has variance

$$\sigma_{\text{opt}}^2(\gamma) = s = 2\hbar^4 \widetilde{\mu}^2(\gamma) = \sigma_{\text{cos}}^2(\gamma), \quad (4.4)$$

where  $\sigma_{\text{cos}}^2(\gamma)$  is exactly the cosmic variance found by different arguments in App. C of [9]. This provides a simple and alternative way to derive the cosmic variance.

### A. Pulsar averaging the covariance matrix

We now show that the sum of any row of a  $C_{jk}$  block of the full covariance matrix  $C$  has the same value as the sum of any other row of that block, showing that  $\mathbb{1}$  is an eigenvector of the block. We also show that the average value of the entries in that block is given by (4.1). Since we restrict ourselves to just that single block, in this section we drop the indices  $j, k$  from this average value.

Start with (2.10) for the covariance matrix, and use the identities (2.12) and (2.13) to obtain

$$\begin{aligned} C_{ab,cd} &= \hbar^4 (\mu_{ac}\mu_{bd} + \mu_{ad}\mu_{bc}) \\ &= \hbar^4 \sum_A \sum_{A'} \int \frac{d\Omega}{4\pi} \int \frac{d\Omega'}{4\pi} \left( F_a^A(\Omega) F_c^{A'}(\Omega) F_b^{A'}(\Omega') F_d^{A'}(\Omega') + F_a^A(\Omega) F_d^A(\Omega) F_b^{A'}(\Omega') F_c^{A'}(\Omega') \right) \\ &\quad + \hbar^4 \mu_{\text{u}}(0) \left[ \mu_{\text{u}}(0) (\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}) + \mu_{\text{u}}(\gamma_{bd})\delta_{ac} + \mu_{\text{u}}(\gamma_{ac})\delta_{bd} + \mu_{\text{u}}(\gamma_{bc})\delta_{ad} + \mu_{\text{u}}(\gamma_{ad})\delta_{bc} \right]. \end{aligned} \quad (4.5)$$

This can be written more compactly using ‘‘symmetrization notation’’ [15] for indices:

$$\begin{aligned} C_{ab,cd} &= 2\hbar^4 \left[ \sum_A \sum_{A'} \int \frac{d\Omega}{4\pi} \int \frac{d\Omega'}{4\pi} F_a^A(\Omega) F_{(c}^A(\Omega) F_{d)}^{A'}(\Omega') F_b^{A'}(\Omega') \right. \\ &\quad \left. + \mu_{\text{u}}^2(0) \delta_{a(c} \delta_{d)b} + \mu_{\text{u}}(0) \left( \mu_{\text{u}}(\gamma_{a(c} \delta_{d)b} + \mu_{\text{u}}(\gamma_{b(c} \delta_{d)a} \right) \right] \end{aligned} \quad (4.6)$$

Here, the index symmetrization is defined by

$$Q_{(cd)} = \frac{1}{2} (Q_{cd} + Q_{dc}). \quad (4.7)$$

As we stressed earlier, this equation holds for arbitrary choices of the four pulsars  $a, b, c$ , and  $d$ .

We now compute the average value of the entries along the  $ab$  row of one of the million blocks of (4.6). To do this we fix  $ab$  to one of the rows of the block, and average along the columns  $cd$  within the block. The pairs labeled by  $cd$

all have (about) the same angular separation  $\gamma_k$ , and since the pulsars were uniformly distributed about the sky, so are these  $cd$  pairs. Thus, the average is

$$s_{ab} = \langle C_{ab,cd} \rangle_{cd,\gamma_k}, \quad (4.8)$$

where  $\langle \rangle_{cd,\gamma_k}$  denotes the average over all pulsar pairs  $cd$  separated by angle  $\gamma_k$ . In [9] this is called a ‘‘pulsar average’’ and is explicitly defined in App. A of that reference as a normalized integral over three variables.

There are three terms that contribute to the average (4.8). The first explicitly involves the antenna pattern functions, and the second and third terms involve either a single Kronecker delta function or a product of two Kronecker deltas. We show in App. C that the latter two terms correspond to sets of measure zero, and make negligible contributions in the limit of a large number of pulsar pairs.

Thus, for large numbers of pulsar pairs, the only term that contributes is

$$s_{ab} = 2\hbar^4 \sum_A \sum_{A'} \int \frac{d\Omega}{4\pi} \int \frac{d\Omega'}{4\pi} F_a^A(\Omega) F_b^{A'}(\Omega') \langle F_c^A(\Omega) F_d^{A'}(\Omega') \rangle_{cd,\gamma_k}. \quad (4.9)$$

Pulsar averages of this type are discussed in App. G of [9], where they are called two-point functions, and the specific average that appears in (4.9) is evaluated. The uniform sky average over all pulsar pairs  $cd$  separated by angle  $\gamma_k$  yields

$$\langle F_c^A(\Omega) F_d^{A'}(\Omega') \rangle_{cd,\gamma_k} = \mu_{AA'}(\gamma_k, \beta), \quad (4.10)$$

where the functions  $\mu_{AA'}(\gamma, \beta)$  are given explicitly in App. G of [9]; they vanish if  $A \neq A'$ . Here,  $\beta$  is the angle between  $\Omega$  and  $\Omega'$ , so  $\cos \beta = \Omega \cdot \Omega'$ . Thus, substituting (4.10) into (4.9) and summing over polarizations  $A'$ , we find

$$s_{ab} = 2\hbar^4 \sum_A \int \frac{d\Omega}{4\pi} \int \frac{d\Omega'}{4\pi} F_a^A(\Omega) F_b^A(\Omega') \mu_{AA}(\gamma_k, \cos^{-1} \Omega \cdot \Omega'). \quad (4.11)$$

It is now easy to see that the value of  $s_{ab}$  is the same for all rows of the block.

We obtained  $s_{ab}$  by averaging over  $cd$  pairs which are uniformly distributed over the sphere. After this averaging, by rotational symmetry,  $s_{ab}$  can only depend upon the angle between the directions to pulsars  $a$  and  $b$ . But since all the  $ab$  pairs in the block have an angle (close to)  $\gamma_j$ , it follows that  $s_{ab}$  is the same for all rows, since all rows have (about) the same  $\gamma_{ab} = \gamma_j$ . Since the value of  $s_{ab}$  does not depend upon the row  $ab$ , we can write  $s_{ab} = s$ .

To compute the average value of any row, simply use (4.10) to average (4.11) over  $ab$  pairs separated by angle  $\gamma_j$ . Since each row has the same value, this averaging does not change that value. Using (4.10) with pulsar labels  $cd$  replaced by  $ab$  and  $\gamma_k$  replaced by  $\gamma_j$ , we obtain

$$\begin{aligned} s &= 2\hbar^4 \sum_A \int \frac{d\Omega}{4\pi} \int \frac{d\Omega'}{4\pi} \mu_{AA}(\gamma_j, \cos^{-1} \Omega \cdot \Omega') \mu_{AA}(\gamma_k, \cos^{-1} \Omega \cdot \Omega') \\ &= 2\hbar^4 \frac{1}{2} \int_0^\pi d\beta \sin \beta \left( \mu_{++}(\gamma_j, \beta) \mu_{++}(\gamma_k, \beta) + \mu_{\times\times}(\gamma_j, \beta) \mu_{\times\times}(\gamma_k, \beta) \right), \end{aligned} \quad (4.12)$$

which establishes (4.1). Note that to obtain the final line of (4.12), we have used

$$\int \frac{d\Omega}{4\pi} \int \frac{d\Omega'}{4\pi} Q(\Omega \cdot \Omega') = \frac{1}{2} \int_0^\pi d\beta \sin \beta Q(\cos \beta) \quad (4.13)$$

for the spherical average of any function  $Q(x)$  of a single variable. The normalization is easily checked by setting  $Q(x) = 1$ .

## V. DECREASING THE PULSAR VARIANCE TO APPROACH THE COSMIC VARIANCE

We have shown how to construct the optimal estimator for the Hellings and Downs correlation at angle  $\gamma$ , by combining correlations from pairs of pulsars separated by that angle, and have calculated the variance of that estimator. In this section, we use numerical studies to understand how that variance decreases as the number of pulsar pairs increases.

There are two limiting cases for which we can give analytic expressions for the variance. If there is only a single pulsar pair at each angular separation, then the variance is the total variance, given by  $\sigma_{\text{tot}}^2(\gamma)$  in (2.11). If there are

a large number of uniformly distributed pulsar pairs, then the variance is the cosmic variance  $\sigma_{\text{cos}}^2(\gamma)$ , given by (4.4). Here, we study the transition between these two limits, when the number of pulsar pairs is finite.

In Fig. 4, we show the transition from  $\sigma_{\text{tot}}^2(\gamma)$  to  $\sigma_{\text{cos}}^2(\gamma)$  as we average together more and more pulsar pairs having the same angular separation  $\gamma$ . For these plots, we simulated 10 different realizations of  $n_{\text{pairs}} = 1, 2, 5, 10, 25, 50, 100$ , and 200 pulsar pairs, all separated by the same angle  $\gamma$  and distributed uniformly on the sky. We then calculated  $\sigma_{\text{opt}}^2(\gamma)$  for each value of  $n_{\text{pairs}}$  and each value of  $\gamma$ . (To reduce the fluctuations associated with the random placement of the pulsar pairs on the sky, we averaged together the results of the 10 different realizations to get the final set of transition plots.)

Also plotted in Fig. 4 is the minimum number of pulsar pairs needed for the optimal variance  $\sigma_{\text{opt}}^2(\gamma)$  to reach a value that lies within  $1/e \approx 0.37$  of the cosmic variance  $\sigma_{\text{cos}}^2(\gamma)$ . In terms of the standard deviation  $\sigma_{\text{opt}}(\gamma)$ , this corresponds to reaching a value of approximately  $1.17\sigma_{\text{cos}}(\gamma)$  for each value of  $\gamma$ . As shown in the figure, in the vicinity of the minima of  $\sigma_{\text{cos}}(\gamma)$ , which are around  $54^\circ$  and  $126^\circ$ , a minimum of approximately 6000 pulsar pairs is required.

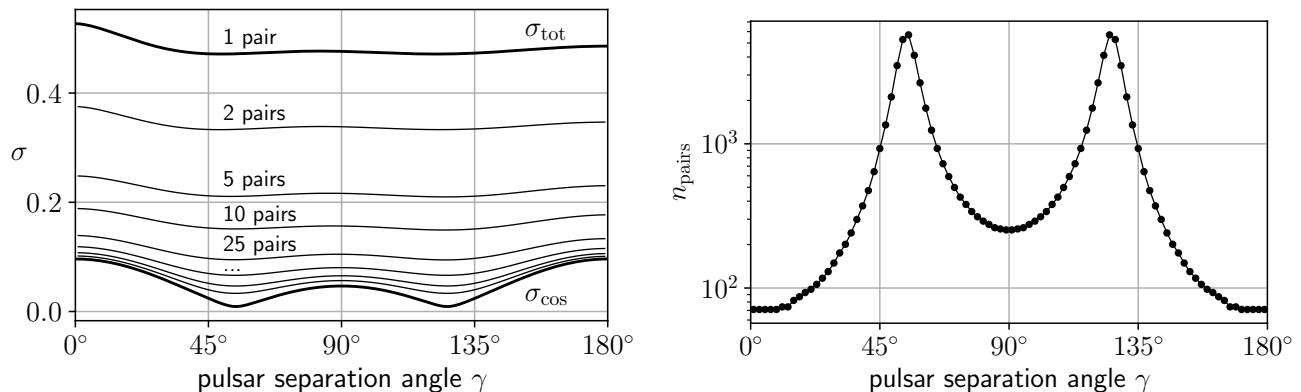


FIG. 4. Left panel: The variance of the optimal estimator as the number of pulsar pairs at each angular separation  $\gamma$  is increased. Shown in the plot, from top to bottom, is  $\sigma_{\text{opt}}(\gamma)$  calculated for 1, 2, 5, 10, 25, 50, 100, and 200 pairs. (For this plot we have set  $\hbar^4 = 1/2$ .) Right panel: The minimum number of pulsar pairs needed for  $\sigma_{\text{opt}}^2(\gamma)$  to reach a value that is  $1/e \approx 37\%$  larger than the cosmic variance. For this number of pulsar pairs,  $\sigma_{\text{opt}}^2(\gamma) \approx (1 + 1/e)\sigma_{\text{cos}}^2(\gamma)$ .

## VI. BINNED PULSAR AVERAGING

In Secs. IV and V, we assumed that we had correlation measurements  $\rho_{ab}$  from multiple pulsar pairs  $ab$  that shared the *exact same separation angle*  $\gamma$ . By optimally combining the  $n_{\text{pairs}}$  measured correlations to best estimate the Hellings and Downs mean correlation using (3.15), we were able to reduce the pulsar variance, bringing the total variance (3.16) of the optimal estimator closer to the cosmic variance. This was proven analytically in Sec. IV in the limit where  $n_{\text{pairs}} \rightarrow \infty$ , and demonstrated numerically in Fig. 4 for increasingly larger values of  $n_{\text{pairs}}$ .

However, as mentioned in the introduction and Sec. III, a real PTA has *no* pulsar pairs that are separated by the exact same angle, so in practice these methods cannot be directly applied. In a real PTA,  $M$  pulsars are distributed nonuniformly on the sky, and there are  $M(M-1)/2$  distinct pairs, each of which has a *distinct* separation angle. To compare their correlations with the Hellings and Downs prediction, these must be *binned* into sets that have similar (but not identical) angular separations. For example, a PTA composed of  $M = 40$  pulsars will have 780 distinct angular separations. These could be distributed across 30 angular separation bins in  $\gamma$ , on average containing 26 pairs each.

### A. Beating the cosmic variance

In Sec. IIIA, we derive the optimal estimator (3.10) for binned correlation measurements. One interesting consequence of doing the binned estimation is that for some ranges of angular separation  $\gamma$ , the reduction in angular resolution allows one to “beat” the cosmic variance for a sufficiently large number of pulsar pairs in the bin. This is illustrated numerically in Fig. 5 for the case of a simulation involving 400 pulsars distributed uniformly over the

sky and 6-degree angular separation bins. An analytic demonstration of this result, for a simple “two-component” angular separation bin is given in App. E.

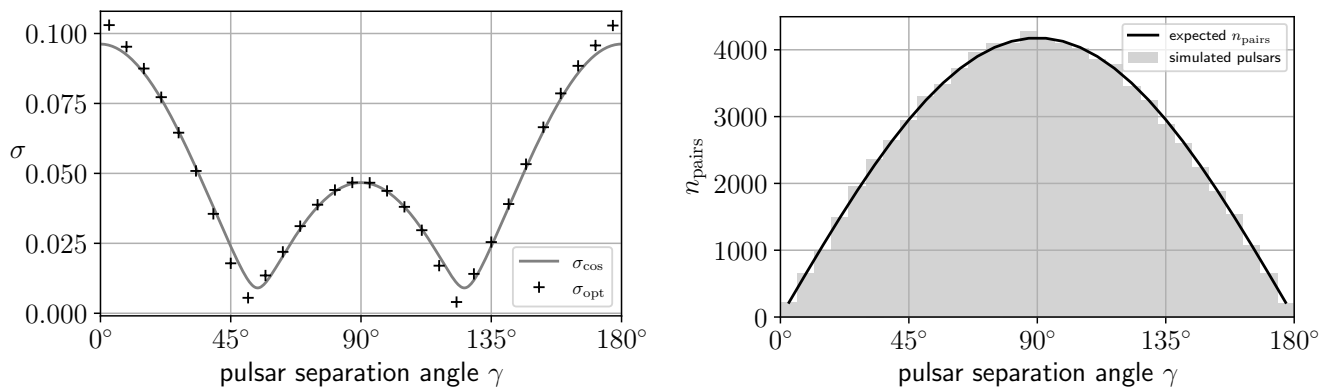


FIG. 5. Left panel: Comparison of the variance of the optimal binned estimator and the cosmic variance for a simulation consisting of 400 pulsars distributed uniformly over the sky and 6-degree angular separation bins. The variance of the optimal binned estimator dips below that of the cosmic variance for sufficiently large numbers of pulsar pairs per bin. This is possible because angular resolution has been sacrificed, see App. E. (For this plot we have set  $h^4 = 1/2$ .) Right panel: The expected number of such pairs in each angular separation bin for the case described in the left-hand panel. The expected number of such pairs is proportional to  $\sin \gamma$ .

## B. Examples

Here, we apply the results of Sec. III A to make plots showing the total variance  $\sigma_{\text{tot}}^2(\gamma)$ , cosmic variance  $\sigma_{\text{cos}}^2(\gamma)$ , and variance  $\sigma_{\text{opt}}^2(\gamma)$  of the optimal binned estimator for a finite number of pulsars distributed nonuniformly over the sky, and for finite-sized angular separation bins. We do so using the current sky locations of the pulsars monitored by the three major pulsar timing array collaborations: European Pulsar Timing Array (EPTA), North American Nanohertz Observatory for Gravitational Waves (NANOGrav), and Parkes Pulsar Timing Array (PPTA). We also construct an International Pulsar Timing Array (IPTA) by forming the union of the pulsars monitored by the individual PTAs. See Table I in App. F for the names and angular coordinates of the pulsars.

The individual PTA collaborations are currently monitoring 42, 66, and 26 pulsars respectively, with a total of 88 distinct pulsars for the IPTA. A skymap of the pulsars is shown in Fig. 6, which is a Mollweide projection in equatorial coordinates. Note that the pulsars are highly concentrated in the direction of the galactic center, which has equatorial coordinates  $(\text{ra}, \text{dec}) = (17\text{h}45\text{m}, -29^\circ)$ , and is indicated by a black dot in Fig. 6. For reference, the center of the sky maps is  $(\text{ra}, \text{dec}) = (12\text{h}, 0^\circ)$ .

Plots showing the expected Hellings and Downs correlation  $\pm$  the uncertainties associated with the total variance, cosmic variance, and variance of the optimal binned estimator are in Fig. 7.

For these plots, we have chosen 30 angular separation bins, equally spaced between 0 and 180 degrees, corresponding to 6 degree bin widths. We have also chosen to model the relative amplitude of the expected correlation and their uncertainties, setting  $h^4 = 0.16h^4$  and  $h^4 = 1$ , which is appropriate for a Gaussian ensemble of timing-residual measurements as described in App. B. Finally, for the optimal binning, we have chosen to normalize the weights according to condition (b) in Sec. III A,  $\mu_{\text{norm}} \equiv \mu^\top \mathbb{1}/n_{\text{pairs}}$ , which is the uniform average of the expected Hellings and Downs correlation across the bin.

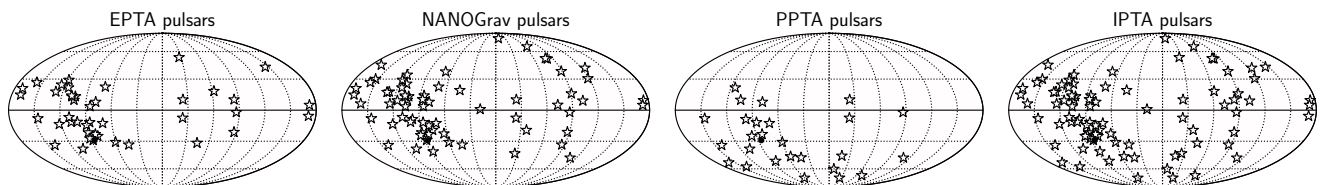


FIG. 6. Sky locations of the pulsars used by the EPTA, NANOGrav, PPTA, and IPTA collaborations. The black dot indicates the direction to the Galactic center. Table I in App. F lists the pulsar names and sky locations.

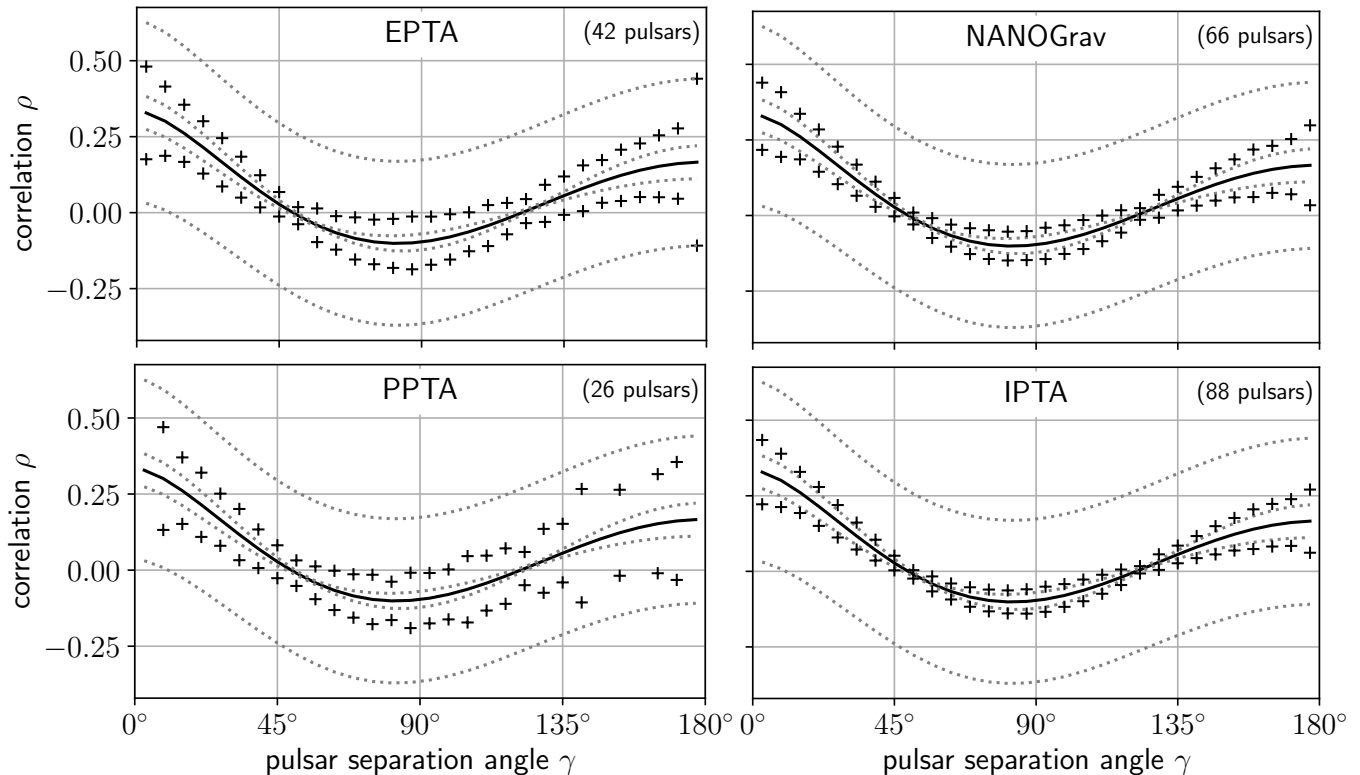


FIG. 7. The mean Hellings and Downs timing-residual correlations (solid black curve) and the  $\pm\sigma_{\text{opt}}$  uncertainty of the optimal binned estimator with  $30 \times 6^\circ$  angular separation bins (“+”-symbols). This illustrates the expected match for data which has no noise or errors. We take the pulsars currently used by the EPTA, NANOGrav, and PPTA collaborations, and merge these for the IPTA. (See Fig. 3 for a zoom of the IPTA variances.) We assume the binary inspiral GW background described in App. B, which has  $\dot{h}^2/h^2 \approx 0.4$  (timing residual correlations,  $\alpha = 1$ ) and normalize the mean with  $h^2 = 1$ . Two other curves are shown for comparison. The dotted lines furthest from the solid black curve are the single pulsar pair variance:  $\pm\sigma_{\text{tot}}$ . The dotted lines closest to the solid black curve are the cosmic variance corresponding to an infinite number of pulsar pairs:  $\pm\sigma_{\text{cos}}$ . The PPTA plot has no “+”-symbols at  $3^\circ$ ,  $147^\circ$ ,  $159^\circ$ , and  $177^\circ$  since  $n_{\text{pairs}} = 0$  at those bins.

These plots do not include the contribution from measurement noise or intrinsic pulsar noise to the total and optimal estimator variances. As such, these plots show the fundamental limits on the recovery of the Hellings and Downs correlation, based on using optimal binned estimators for the current nonuniform distribution of pulsars and a specific choice of binning. Interested readers can produce similar plots for different arrays of pulsars, different choices of binning, etc. following this general procedure.

## VII. INCLUDING NOISE IN THE OPTIMAL ESTIMATOR CALCULATION

Here, we consider what happens when the pulsar-pair correlation measurements have noise contributions (e.g., from measurement noise, intrinsic pulsar noise, etc.). The optimal combination formalism presented in Secs. III and IV can also be applied to this case. The expressions for the mean redshift correlation  $\langle\rho_{ab}\rangle$  and for the covariance matrix  $C_{ab,cd}$  acquire additional terms which incorporate the effect of noise. These expressions are a slight generalization of those found in [8], which considered an optimal cross-correlation estimator of the amplitude of the GW background contribution for a very simple GW signal+noise model.

### A. Combined signal + noise covariance matrix

In what follows, we will assume that the noise in each pulsar’s redshift measurement is described by a spectral function  $N_a(f)$  (analogous to  $H(f)$  for the GW background) where  $a$  labels the pulsar. We will also assume that the noise associated with redshift measurements from different pulsars  $a$  and  $b$  are uncorrelated with one another and with



the GW background. These assumptions imply that if  $\tilde{h}_a(f)$  and  $\tilde{n}_a(f)$  denote the Fourier domain representation of the GW and noise contributions to the redshift measurement  $Z_a$  in pulsar  $a$ , then

$$\langle \tilde{n}_a^*(f) \tilde{n}_b(f') \rangle = N_a(f) \delta_{ab} \delta(f - f'), \quad \langle \tilde{h}_a^*(f) \tilde{h}_b(f') \rangle = 4\pi H(f) \mu_{ab} \delta(f - f'), \quad \langle \tilde{n}_a^*(f) \tilde{h}_b(f') \rangle = 0, \quad (7.1)$$

where angle brackets now denote an average over a Gaussian ensemble of different noise realizations as well as an average over an ensemble of different GW source realizations. Since the measured redshift data is a sum of the GW signal and noise contributions, we have

$$\tilde{Z}_a(f) = \tilde{h}_a(f) + \tilde{n}_a(f), \quad (7.2)$$

and using (7.1),

$$\langle \tilde{Z}_a^*(f) \tilde{Z}_b(f') \rangle = (4\pi H(f) \mu_{ab} + N_a(f) \delta_{ab}) \delta(f - f'), \quad (7.3)$$

where  $\tilde{Z}_a(f)$  is the Fourier domain representation of the time-domain measurements  $Z_a = Z_a(t)$ .

From these assumptions, it follows that the pulsar-pair redshift measurements have expectation value

$$\langle \rho_{ab} \rangle = h^2 \mu_{ab} + n_a^2 \delta_{ab}, \quad n_a^2 \equiv \int_{-\infty}^{\infty} df N_a(f) (2\pi f)^{-2\alpha}, \quad (7.4)$$

where  $\alpha = 0$  or  $1$  for pulsar redshift or timing residual correlations respectively, as explained in App. A. Note that for distinct pulsar pairs  $a \neq b$ , we have  $\langle \rho_{ab} \rangle = h^2 \mu_u(\gamma_{ab})$ , which agrees with the noise-free case. This means that if one restricts attention to a single pair, then only the autocorrelation measurements  $\rho_{aa}$  are affected by the noise. If one looks at correlations between different pairs, then the noise enters in a more complicated way.

For the covariance matrix  $C_{ab,cd}$ , it is straightforward to do a calculation similar to that in Sec. II B. Assuming that both the pulsar noise and the GW background are described by a central-limit-theorem Gaussian ensemble, we use Isserlis' theorem to convert the expectation values of products of four redshift measurements in the frequency domain to products involving two redshift measurements. The final result (which allows for the possibility that  $a = c$ ,  $b = d$ , etc.) is

$$C_{ab,cd} = h^4 (\mu_{ac} \mu_{bd} + \mu_{ad} \mu_{bc}) + (\delta_{ac} \delta_{bd} + \delta_{ad} \delta_{bc}) N_{ab}^2 + (\delta_{ac} M_a \mu_{bd} + \delta_{bd} M_b \mu_{ac} + \delta_{ad} M_a \mu_{bc} + \delta_{bc} M_b \mu_{ad}), \quad (7.5)$$

where

$$\begin{aligned} N_{ab}^2 &\equiv \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \operatorname{sinc}^2(\pi(f - f')T) N_a(f) N_b(f') (4\pi^2 f f')^{-2\alpha} \text{ and} \\ M_a &\equiv 4\pi \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \operatorname{sinc}^2(\pi(f - f')T) N_a(f) H(f') (4\pi^2 f f')^{-2\alpha}. \end{aligned} \quad (7.6)$$

With the assumptions we have made, this is the form of the covariance matrix including both the GW signal contributions and the pulsar-noise contributions.

## B. Properties of the noise contribution to the covariance matrix

Does the inclusion of pulsar noise in the covariance matrix affect our previous analyses? For a finite number of pulsars and sufficiently loud noise, the answer is yes. Current searches for GW backgrounds using pulsar timing array data *do* include the noise contributions in their analyses (both Bayesian and frequentist) as described briefly in Sec. IX. But in the limit of a sufficiently large number of pulsar pairs distributed uniformly on the sky, the noise contributions average to zero.

To see this, we can repeat the calculation of Sec. IV A including the additional noise terms. As before, the relevant entries of the covariance matrix are those where the rows and columns correspond to pulsar pairs  $ab$  all separated by the same angle  $\gamma$ . We showed in Sec. IV A that the  $\mathbb{1}$  vector is an eigenvector of the noise-free component of the covariance matrix in the limit of an infinite number of pulsar pairs distributed uniformly across the sky. So is the  $\mathbb{1}$  vector also an eigenvector of the noise terms in (7.5)? Examining those terms by first symmetrizing over the indices  $c$  and  $d$ , and then expanding  $\mu_{ac}$ , etc. in terms of the corresponding Hellings and Downs functions  $\mu_u(\gamma_{ac})$  and Kronecker deltas  $\delta_{ac}$ , etc. using (2.4), we obtain

$$\begin{aligned} N_{ab,cd} &\equiv (\delta_{ac} \delta_{bd} + \delta_{ad} \delta_{bc}) N_{ab}^2 + (\delta_{ac} M_a \mu_{bd} + \delta_{bd} M_b \mu_{ac} + \delta_{ad} M_a \mu_{bc} + \delta_{bc} M_b \mu_{ad}) \\ &= 2 [N_{ab}^2 + (M_a + M_b) \mu_u(0)] \delta_{a(c} \delta_{d)b} + 2 M_a \delta_{a(c} \mu_u(\gamma_{d)b}) + 2 M_b \delta_{b(c} \mu_u(\gamma_{d)a}). \end{aligned} \quad (7.7)$$

But since all of the above terms involve either a single Kronecker delta or a product of two Kronecker deltas, they give zero in the limit of an infinite number of pulsar pairs as shown explicitly in App. C. Thus, the  $\mathbf{1}$  vector *is* an eigenvector of the noise terms, but it has eigenvalue 0. So the noise terms do not contribute to the limiting behavior of the full covariance matrix, and

$$\sigma_{\text{opt}}^2(\gamma) = \sigma_{\text{opt,signal}}^2(\gamma) + \sigma_{\text{opt,noise}}^2(\gamma) = \sigma_{\text{cos}}^2(\gamma), \quad (7.8)$$

just as we found for the noise-free case.

### VIII. HOW TO APPLY THE OPTIMAL ESTIMATOR

It is helpful to think about how the methods described in this paper would be used in practice if we are handed a set of measurements  $\rho_{ab}$  from an angular bin at angle  $\gamma$ . From these measured correlations, we want to construct an optimal estimator of the correlation at angle  $\gamma$ , and proceed as described in this paper. First, we compute the weights  $w_{ab}$  from (3.9) or (3.14). For this, we need the (inverse of the) covariance matrix  $C_{ab,cd}$  appropriate for that angular separation bin. *But the covariance matrix cannot be obtained or estimated from the data that we have – a theoretical model is needed.*

For the Gaussian ensemble, the theoretical model is (2.10). The coefficient  $h^4$  can be accurately estimated via (2.11) from the pulsar-pair self-correlations (also using other angular bins, if desired). Then, since we know the locations of all of the pulsars in our data set, the rhs of (2.10) can be computed using (2.4) and (1.1). From that, we can proceed to form the weights and the optimal estimator.

The important point to note is that in following this procedure *we have assumed that the Gaussian ensemble model is correct*. Logically, there are only two possibilities: (a) the data/Universe is well-described by the Gaussian ensemble, or (b) the data/Universe is *not* well-described by the Gaussian ensemble. In case (a), the optimal estimator that we form, and the variance that we attribute to it, will both be meaningful and consistent. However, in case (b), neither the weights that we have derived nor the variance of the optimal estimator that we compute, have any justification.

Consider a specific example: an ensemble of universes containing GW sources which radiate at distinct nonoverlapping frequencies. Such an ensemble is constructed in Sec. IIIb of Ref. [9], where it is demonstrated that the cosmic variance is zero, so the pulsar-averaged correlations curves always follow the  $\mu_{\text{u}}(\gamma)$  shape exactly. But if we followed the procedure described above, we would obtain the same estimated variance for the optimal estimator as in the Gaussian confusion-noise case, where the pulsar-averaged correlations curves do not follow  $\mu_{\text{u}}(\gamma)$  precisely.

The conclusion is that the construction provided here is mainly useful as a consistency test, given that we have a finite set of observations that are not uniform on the sky. It can be used to determine the number of pulsar pairs to place in each angular bin, and also used to weight the correlations of those pairs. Finally, it can be used to estimate (for the Gaussian ensemble) the variance that would be expected for that set of pulsar pairs. If the correlation curve that is obtained has fluctuations of about  $\pm 1\sigma_{\text{opt}}$  around the  $h^2\mu_{\text{u}}(\gamma)$  curve, then we conclude that the Universe is consistent with the Gaussian ensemble predictions. If the fluctuations are significantly smaller or significantly larger, then the opposite conclusion is reached.

In this sense, the results of this paper are useful because they predict the total Gaussian ensemble variance (pulsar plus cosmic) for the case of a finite number of pulsars at specific sky locations, whereas the variance  $\sigma_{\text{cos}}^2$  computed in Ref. [9] is for the limit of an infinite number of uniformly distributed pulsars.

### IX. CONCLUSION

The standard definition of the Hellings and Downs correlation is for a single pair of pulsars. We have extended and generalized that, defining the Hellings and Downs correlation for a set of pulsar pairs whose angular separations  $\gamma$  lie in some narrow range.

We have shown the optimal way to weight the correlations of the individual pulsar pairs to recover the Hellings and Downs mean as closely as possible. These weights are normalized to produce an average that agrees with the single-pair correlation, enabling direct comparison. As pulsar pairs are added at some angle, the variance of the correlation decreases. This variance starts at the total variance  $\sigma_{\text{tot}}^2$  of [9] for a single pulsar pair. It decreases smoothly as pulsar pairs are added, eventually approaching the cosmic variance  $\sigma_{\text{cos}}^2$  of that same reference. The plots in Fig. 4 illustrate this transition.

In future pulsar timing array searches, using improved telescopes and larger pulsar populations, *we expect that the variance of the Hellings and Downs correlation will reveal exactly this cosmic variance*. In practice, as a function of pulsar sky-separation angle  $\gamma$ , the optimal estimator of the Hellings and Downs correlation will not agree exactly in shape with the classic Hellings and Downs curve  $\mu_{\text{u}}(\gamma)$ . The cosmic variance is an estimator of the difference.

Current searches face considerable challenges since they contend with many sources of noise (e.g., measurement noise, intrinsic pulsar noise, etc). The noise is typically included in pulsar timing analyses via a noise covariance matrix in the construction of a likelihood function for a Bayesian analysis, or as part of a filter function for frequentist detection statistics/estimators of the GW background, similar to what we describe in Sec. VII. If this noise is uncorrelated between pulsars, then we have shown that, as more and more pulsars are added to the PTA, the pulsar noise eventually becomes insignificant, and the cosmic variance should emerge.

Since the data is noise-dominated (possibly until recently, see below) current implementations of the optimal cross-correlation statistic [16–18] ignore the (GW-induced) off-diagonal elements of the covariance matrix when forming the weights that are used to combine the measured correlations. If the GW background is small compared to the noise, then these off-diagonal elements may safely be ignored. However, if the GW signal is comparable to (or larger than) the noise, then it is critical to include these off-diagonal elements. As we have shown here, these GW-induced off-diagonal covariances reflect the correlations between measurements in a given angular separation bin. If they are ignored, then the Hellings and Downs correlation is misestimated, and we find that the errors are typically underestimated.

If we could safely ignore these covariances, then the variance of the optimal estimator would go to zero in the limit of an infinite number of pulsar pairs distributed uniformly on the sky. But as we showed in Sec. IV, this is not the case. The limiting expression for the variance of the optimal estimator is not zero; instead it is the cosmic variance.

This topic is of particular relevance because pulsar timing array searches are now entering a regime where the GW background may be dominating the correlation in the lowest frequency bins [4–7]. Failing to include the off-diagonal elements of the covariance matrix when constructing the optimal estimator for the Hellings and Downs correlation or the amplitude of the GW background will lead to incorrect statistical assessments of the data (e.g., confidence intervals that do not have the proper statistical coverage). As such, effort are currently underway [19] to include the full form of the covariance matrix, cf. (7.5), to fully implement the optimal pulsar-pair estimator and to improve our recovery of the GW background.

## ACKNOWLEDGMENTS

The authors thank EPTA, NANOGrav, and PPTA for providing lists of their pulsar names and sky positions. JDR acknowledges support from NSF Physics Frontiers Center Award PFC-2020265 and start-up funds from Texas Tech University.

## OUTLINE OF APPENDICES

The appendices contain calculations and arguments which are useful for the main body of the paper. Here, we outline their contents.

Appendix A shows how our results for the Hellings and Downs correlation of pulsar redshifts can be easily modified to treat pulsar timing residuals.

The relationship between the mean and the variance for the Gaussian ensemble depends upon certain integrals of the power spectrum. We discuss this dependence in App. B and calculate the relationship for a simple model of the GW background.

In App. C, we examine some Kronecker delta terms which appear in the covariance matrix. We show that if there are many pulsar pairs uniformly distributed on the sky, then the delta terms constitute a set of measure zero and do not affect the variance.

In App. D, we consider the case where there are large numbers of pulsar pairs at a small (integer) number  $\kappa$  of distinct separation angles. We show that the quantity relevant for the variance of the Hellings and Downs correlation may be calculated from  $\kappa(\kappa - 1)/2$  quantities, obtained from the two-point function given in [9] via (4.1). We employ this in App. E to show how (with a binned analysis employing large number of measurements at two different angles) it is possible to “beat” the cosmic variance.

Finally, in App. F, we list the sky positions of the 88 pulsars currently monitored by the various PTA collaborations. These are used in Sec. VI B for the plots shown in Fig. 7.

### Appendix A: Mean and variance of timing residual correlations (versus redshift correlations)

Much of the literature and experimental/observational work for PTAs is based on pulsar timing residuals rather than redshifts. Pulsar timing residuals may be obtained by integrating the pulse redshifts with respect to time, which inserts a factor of  $1/2\pi if$  into frequency domain formulae. This makes it straightforward to modify formulae for the

first moment, the second moment, and the variance and covariance of pulsar timing redshifts so that they can also be applied to pulsar timing residuals.

Following this procedure, formulae identical to (2.3) and (2.10) are obtained for the first moment and covariance matrix of timing residual correlations, by replacing  $H(f)$  with  $H(f)/4\pi^2 f^2$  in the definitions of  $h^2$  and  $\hat{h}^4$  given in (2.6) and (2.9). We implement this by introducing a constant  $\alpha$  in those formulae. Set  $\alpha = 0$  to describe pulsar redshift correlations, or set  $\alpha = 1$  to describe pulsar timing residual correlations.

### Appendix B: Relative scaling between the correlation mean and variance

The GW Gaussian ensemble is completely specified by the real spectral function  $H(f) \geq 0$  defined by (2.6) and (2.7). For example,  $H(f)$  sets the overall scale  $h^2$  of the expected Hellings and Downs correlation via (2.3). The spectral function  $H(f)$  also determines the scale  $\hat{h}^4$  of the covariance matrix  $C_{ab,cd}$  via (2.9) and (2.10).

This implies that  $\hat{h}^4$  sets the scale of the total variance and the cosmic variance, since those are determined entirely by  $C_{ab,cd}$ . It follows immediately that the ratio of the total variance (for a given number of pulsar pairs and binning procedure) to (say) the cosmic variance is identical for *any* Gaussian ensemble: the quantities being compared only depend upon  $\hat{h}^4$ . Thus, plots in this paper that only contain variances are *universal*: they apply to any Gaussian ensemble.

In contrast, the ratio of the variance to the (squared) mean is model-dependent: it depends upon the ratio  $\hat{h}^4/h^4 < 1$  and hence upon  $H(f)$ . To compare the (square of the) mean and the variance, we must specify the spectrum  $H(f)$  which defines the Gaussian ensemble. Hence, plots in this paper that contain variances and means (for example, Fig. 7) are *model-dependent*. To make them, we must assume some form of the spectrum  $H(f)$ .

One simple source model is constructed in Sec. 3 of Ref. [9]. This “confusion-noise model” consists of a large set of GW sources, uniformly distributed in space, *all radiating GWs at exactly the same frequency, but with different random phases*. The corresponding spectral function  $H(f)$  is the sum of a delta function centered at that frequency, plus another equal one at the corresponding negative frequency. In App. C of [9] it is shown that for this model  $\hat{h}^4/h^4 \in [1/2, 1)$ ; if the frequency is an integer multiple of the inverse observation time, then  $\hat{h}^4/h^4 = 1/2$ . Note that the same values for this ratio are obtained for pulsar redshift correlations ( $\alpha = 0$ ) and for pulsar timing residual correlations ( $\alpha = 1$ ). This source model is easy to construct and understand, so we use this value  $\hat{h}^4/h^4 = 1/2$  in Fig. 1. However, the model is unrealistic because in our Universe the sources are distributed over frequency.

Here, we construct a more realistic spectral model, which we then use for Fig. 7. This more realistic model assumes that the GW background is produced by large numbers of compact binary systems, whose energy loss is dominated by GW emission in the Newtonian limit (orbital velocities  $\ll$  speed of light). In this case,  $H(f) \propto |f|^{-7/3}$ . Physically, this spectrum extends to very low frequencies (corresponding to periods longer than thousands of years).

Although it continues to much lower frequencies, for the purpose of computing PTA correlations, the spectrum is effectively “cut off” below a characteristic frequency  $f_0 > 0$ . Hence, we take

$$H(f) = \begin{cases} q|f|^{-7/3} & \text{for } |f| > f_0, \text{ and} \\ 0 & \text{for } |f| \leq f_0, \end{cases} \quad (\text{B1})$$

where  $q$  is an irrelevant constant. The cutoff frequency  $f_0$  is of order  $1/T$ , where  $T$  is the total observation time. It arises because PTA data analysis in effect “removes” low-frequency components during the fitting process that subtracts (quadratic and higher-order terms in) the pulsar’s intrinsic spin-down to obtain timing residuals [20–22]. For further details, see the “transmission functions” illustrated in Fig. 2 of Ref. [22].

The ratio of the variance to the squared mean is proportional to  $\hat{h}^4/h^4$ . For our simple spectral model, this ratio only depends [23] upon the low-frequency cutoff  $f_0$ . This can be seen by substituting (B1) into (2.6) and (2.9) to obtain

$$\frac{\hat{h}^4}{h^4} = \frac{\int_{\pi f_0 T}^{\infty} dx \int_{\pi f_0 T}^{\infty} dy \left[ \frac{1}{2} \text{sinc}^2(x-y) + \frac{1}{2} \text{sinc}^2(x+y) \right] (xy)^{-7/3-2\alpha}}{\left( \int_{\pi f_0 T}^{\infty} dx x^{-7/3-2\alpha} \right)^2}. \quad (\text{B2})$$

Here, we have changed to dimensionless variables  $x = \pi T f$  and  $y = \pi T f'$ , and used  $H(f) = H(-f)$  to write the integrals in one-sided form. The constant  $\alpha$  that appears in (B2) is explained in App. A. Set  $\alpha = 0$  to describe the ratios of pulsar redshift correlations, or set  $\alpha = 1$  to describe the ratios of pulsar timing residual correlations.

The behavior of  $\hat{h}^4/h^4 < 1$  as given in (B2) has interesting limits. As  $f_0 T \rightarrow 0$ , the integrands and integrals are dominated by values of  $x$  and  $y$  which are small enough that both sinc functions approach unity. The numerator

of (B2) then approaches the denominator, and  $\hbar^4/\hbar^4 \rightarrow 1$ . For values of  $f_0T$  which are somewhat larger, the first sinc function approaches unity, whereas the second sinc function falls off quickly, and  $\hbar^4/\hbar^4 \rightarrow 1/2$ . For the cases of interest, where  $f_0T$  is of order unity, the first sinc function is also less than one (although it contributes more than the second one).

The ratio  $\hbar^2/h^2$  (the square root of (B2)) is evaluated numerically and shown in Fig. 8 as a function of the effective lower cutoff frequency  $f_0T$ . A reasonable choice is  $f_0 = 1/T$ , which is justified by Fig. 2 of Ref. [22]. Taking  $f_0T = 1$  in Fig. 8 implies  $\hbar^2/h^2 \approx 0.4$  for timing residuals and  $\hbar^2/h^2 \approx 0.26$  for redshifts. We use the first of these values for the plots in Fig. 7. For the other plot in the paper (Fig. 1) that shows a mean value, we use  $\hbar^4/h^4 = 1/2$ , corresponding to the single-frequency confusion-noise model described above and in [9].

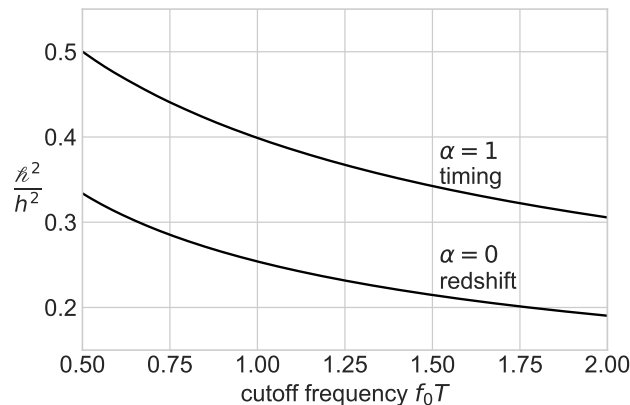


FIG. 8. The quantity  $\hbar^2/h^2$  determines the ratio of the standard deviation  $\sigma$  in the Hellings and Downs correlation to its mean  $\langle \rho \rangle$ . The ratio depends upon the spectrum  $H(f)$  of the Gaussian ensemble. Here, this corresponds to a set of binary inspiral sources (see text). The different values of the cutoff frequency  $f_0$  are expressed in units of the inverse observation time  $1/T$ . The  $\alpha = 0$  curve is for redshift correlations, and the  $\alpha = 1$  curve is for timing-residual correlations (see App. A).

### Appendix C: Proof that terms involving Kronecker deltas do not contribute to the cosmic variance

Here, we show that terms in the covariance matrix  $C_{ab,cd}$  involving one or two Kronecker delta terms, such as

$$2\hbar^4 \left[ \mu_u^2(0) \delta_{a(c)\delta_{d)b} + \mu_u(0) \left( \delta_{a(c)\mu_u(\gamma_{d)b)} + \delta_{b(c)\mu_u(\gamma_{d)a)} \right) \right] \quad \text{or} \quad (C1)$$

$$2 \left[ N_{ab}^2 + (M_a + M_b)\mu_u(0) \right] \delta_{a(c)\delta_{d)b} + 2M_a \delta_{a(c)\mu_u(\gamma_{d)b)} + 2M_b \delta_{b(c)\mu_u(\gamma_{d)a)}$$

from (4.6) or (7.7), do not contribute to the average value of the covariance matrix components over a row,  $\langle C_{ab,cd} \rangle_{cd,\gamma}$  defined by (4.8), where  $\gamma$  denotes the angular separation of the  $ab$  and  $cd$  pairs. Basically, these terms constitute a set of measure zero in the limit where the number of pulsar pairs goes to infinity.

We begin by considering the terms proportional to  $\delta_{a(c)\delta_{d)b}$ . The factors  $\mu_u^2(0)$ ,  $N_{ab}^2$ , or  $(M_a + M_b)$  multiplying these terms are all finite, and hence won't affect the following analysis. Such terms are easy to evaluate, since there is only one value of  $c$  and  $d$  that will match the fixed values of  $a$  and  $b$ :

$$2\langle \delta_{a(c)\delta_{d)b} \rangle_{cd,\gamma} = 1/n_{\text{pairs}}. \quad (C2)$$

This is manifestly independent of the fixed pulsar pair  $ab$ , and vanishes in the limit of an infinite number of pulsar pairs distributed uniformly on the sky: it is a set of measure zero.

The terms proportional to a single Kronecker delta are also sets of measure zero. This may be understood by looking at just one of four subterms

$$s_{ab} \equiv \langle \delta_{ac}\mu_u(\gamma_{db}) \rangle_{cd,\gamma}. \quad (C3)$$

We can again ignore the factors  $\mu_u(0)$ ,  $M_a$ , and  $M_b$  since they are finite multiplicative constants. This average also vanishes in the limit of an infinite number of pulsar pairs distributed uniformly on the sky. To see this, note that (C3)

can be bounded above by replacing  $\mu_u(\gamma_{db})$  by its maximum value  $\mu_u(0)$  and bounded below by replacing  $\mu_u(\gamma_{db})$  with its minimum value. Looking first at the upper bound, one must have

$$s_{ab} \leq \mu_u(0) \langle \delta_{ac} \rangle_{cd, \gamma}. \quad (\text{C4})$$

Since  $\delta_{ac}$  is independent of  $d$ , the averaging over  $d$  (within a cone of angle  $\gamma$  about  $c$ ) has no effect; that is, the  $cd$  averaging reduces to an average just over  $c$ . Now, imagine explicitly carrying out this average as a sum of  $N$  terms divided by  $N$ . Cover the surface of the sphere with  $N$  equal-area uniformly distributed patches, which are the discrete values of  $c$ . In the sum over  $c$ , the Kronecker delta in (C4) only gets a contribution from a single term (the term for which the  $c$  patch contains  $a$ ). Hence, we obtain

$$s_{ab} \leq \mu_u(0)/N, \quad (\text{C5})$$

so  $s_{ab}$  is bounded above by  $\mu_u(0)/N$ . A similar argument shows that  $s_{ab}$  is bounded below by a constant/ $N$ . Since both bounds converge to zero for large  $N$ , we conclude that  $s_{ab}$  vanishes as  $N \rightarrow \infty$ .

#### Appendix D: Computations with matrices of a certain form

We consider real, symmetric, positive-definite matrices built up from smaller matrices, of the form

$$C = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1\kappa} \\ C_{21} & C_{22} & \cdots & C_{2\kappa} \\ \vdots & \vdots & \ddots & \vdots \\ C_{\kappa 1} & C_{\kappa 2} & \cdots & C_{\kappa \kappa} \end{pmatrix}. \quad (\text{D1})$$

Here, each of the  $C_{jk}$  is a (possibly nonsquare) matrix, and  $\kappa$  is a positive integer. We assume that each of these  $\kappa^2$  matrices has the property that *the sum of any row of  $C_{jk}$  has the same value as the sum of any other row of  $C_{jk}$* . (Since  $C$  is symmetric, this implies that the sum of any column of  $C_{jk}$  equals the sum of any other column of  $C_{jk}$ .)

We introduce the symbol  $d_k$  (for  $k = 1, \dots, \kappa$ ) to denote the number of columns in the matrix  $C_{jk}$ . Note that this is independent of  $j$ . Note also that since  $C$  is symmetric, the number of rows of the matrix  $C_{jk}$  is given by  $d_j$ . Hence, the matrix  $C_{jk}$  has dimension ( $d_j$  rows)  $\times$  ( $d_k$  columns). The sum

$$N = \dim(C) = \sum_{j=1}^{\kappa} d_j \quad (\text{D2})$$

is the dimension of the matrix  $C$ .

Finally, we define the quantity  $s_{jk}$  to be the *average value of the entries of the matrix  $C_{jk}$* . Note that since the sum of any row of  $C_{jk}$  gives the same number, independent of the row, that sum must be given by  $d_j s_{jk}$ .

Equation (D1) is the form taken by the covariance matrix of the Hellings and Downs correlation, when there are  $\kappa$  distinct values of the pulsar-pair separation angle  $\gamma_j$ , for  $j = 1, \dots, \kappa$ , and there are  $d_j$  distinct pulsar pairs at each angular separation. We have shown in Sec. IV that when there are large numbers of pulsar pairs at each angle, uniformly distributed on the sky, then the average value of any row of the  $C_{jk}$  matrix is row-independent.

In this appendix, we obtain a simple formula for the ‘‘sum of interest’’, which is the quantity  $\mu^\top C^{-1} \mu$ . Here,  $\mu$  is an  $N$ -dimensional column vector containing values of the Hellings and Downs curve  $\mu_u(\gamma_j)$  corresponding to the angular separation of the pairs.

We begin by constructing a set of  $\kappa$  column vectors  $e_1, e_2, \dots, e_\kappa$  of dimension  $N$ . Note that these do not form a basis for  $C$ , since that would require  $N$  vectors, and here we have far fewer. However, we will see that the  $e_j$  form a basis for the subspace of interest, needed to evaluate the sum of interest.

The first of these vectors has  $d_1$  equal nonzero elements followed by  $N - d_1$  zeros. The second of these vectors has  $d_1$  zeros followed by  $d_2$  equal nonzero elements, followed by  $N - d_1 - d_2$  zeros, and so on. Thus, we have

$$\begin{aligned} e_1 &= d_1^{-1/2} (1, \dots, 1, 0, \dots, 0, \dots, 0, \dots, 0)^\top, \\ e_2 &= d_2^{-1/2} (0, \dots, 0, 1, \dots, 1, \dots, 0, \dots, 0)^\top, \\ &\dots \\ e_\kappa &= d_\kappa^{-1/2} (0, \dots, 0, 0, \dots, 0, \dots, 1, \dots, 1)^\top. \end{aligned} \quad (\text{D3})$$

The ( $N$ -dimensional) dot product of any two of these vectors follows immediately from (D3), and is

$$e_j^\top e_k = \delta_{jk}, \quad (\text{D4})$$

where  $\delta_{jk}$  is the Kronecker delta. Hence, the dot product of distinct vectors vanishes.

Because the sum of any row of  $C_{jk}$  yields the same number, the action of  $C$  on any one of the vectors  $e_k$  produces a linear sum of the full set: the vectors  $e_1, e_2, \dots, e_\kappa$  form an invariant subspace under the action of  $C$ . Multiplying  $e_k$  given in (D3) on the left by  $C$  given by (D1) yields

$$C e_k = \sum_{j=1}^{\kappa} S_{jk} e_j, \quad (\text{D5})$$

where we have defined the quantities

$$S_{jk} \equiv (d_j d_k)^{1/2} s_{jk} \quad \text{for } j, k = 1, \dots, \kappa. \quad (\text{D6})$$

The square root factors in (D6) arise from counting the contributions arising from each row of  $C e_k$ .

To evaluate the sum of interest, we exploit the fact that  $s_{jk}$  and  $S_{jk}$  are symmetric  $\kappa \times \kappa$  matrices. This allows us to transform the  $N \times N$  dimensional problem to a smaller  $\kappa \times \kappa$  dimensional problem. Let  $S^{-1}$  denote the matrix inverse of  $S$ , which satisfies (and is also defined by)

$$\sum_{\ell=1}^{\kappa} S_{j\ell} (S^{-1})_{\ell k} = \delta_{jk}. \quad (\text{D7})$$

From (D6) and (D7) it immediately follows that

$$(S^{-1})_{jk} = (d_j d_k)^{-1/2} (s^{-1})_{jk} \quad \text{for } j, k = 1, \dots, \kappa, \quad (\text{D8})$$

where  $s^{-1}$  denotes the matrix inverse of  $s$ .

With the matrix inverse  $S^{-1}$  we define a set of  $N$ -dimensional column vectors

$$v_j = \sum_{k=1}^{\kappa} (S^{-1})_{kj} e_k \quad \text{for } j = 1, \dots, \kappa. \quad (\text{D9})$$

We now show that  $C^{-1} e_j = v_j$ . Consider the action of  $C$  on  $v_j$ . Multiplying (D9) on the left by  $C$  and using (D5) we obtain

$$\begin{aligned} C v_j &= \sum_{k=1}^{\kappa} (S^{-1})_{kj} \sum_{\ell=1}^{\kappa} S_{\ell k} e_\ell \\ &= \sum_{\ell=1}^{\kappa} \left( \sum_{k=1}^{\kappa} S_{\ell k} (S^{-1})_{kj} \right) e_\ell \\ &= e_j, \end{aligned} \quad (\text{D10})$$

where on the second line we use the definition of the matrix inverse (D7) to replace the quantity in curved brackets with the Kronecker delta. Multiplying (D10) on the left by  $C^{-1}$ , it follows immediately that

$$C^{-1} e_j = v_j. \quad (\text{D11})$$

Taking the inner product of (D11) from the left with  $e_k^\top$  and using (D9) and (D4) gives us the inner product

$$e_k^\top C^{-1} e_j = (S^{-1})_{kj}, \quad (\text{D12})$$

from which we can now easily evaluate the quantity of interest.

We write the  $N$ -dimensional column vector  $\mu$  as a linear combination of the  $e_j$ :

$$\mu = \sum_{j=1}^{\kappa} \mu_j \sqrt{d_j} e_j, \quad (\text{D13})$$

where  $\mu_j \equiv \mu_u(\gamma_j)$  is the value of the Hellings and Downs curve at angle  $\gamma_j$ . The quantity of interest is then evaluated as

$$\begin{aligned}
\mu^\top C^{-1} \mu &= \sum_{j=1}^{\kappa} \sum_{k=1}^{\kappa} \mu_j \mu_k \sqrt{d_j} \sqrt{d_k} e_j^\top C^{-1} e_k \\
&= \sum_{j=1}^{\kappa} \sum_{k=1}^{\kappa} \mu_j \mu_k (d_j d_k)^{1/2} (S^{-1})_{jk} \\
&= \sum_{j=1}^{\kappa} \sum_{k=1}^{\kappa} \mu_j \mu_k (s^{-1})_{jk} \\
&= \hat{\mu}^\top s^{-1} \hat{\mu},
\end{aligned} \tag{D14}$$

where on the second line we use (D12) to evaluate the matrix elements of  $C^{-1}$  and on the third line we use (D8) to replace the matrix elements of  $S^{-1}$  by those of  $s^{-1}$ . The  $\kappa$ -dimensional column vector  $\hat{\mu}$  which appears in the final line has components

$$\hat{\mu} = (\mu_1, \mu_2, \dots, \mu_\kappa)^\top. \tag{D15}$$

Thus, the  $N \times N$  dimensional computation of  $\mu^\top C^{-1} \mu$  has been reduced to a  $\kappa \times \kappa$  dimensional problem, where  $\kappa$  is the number of distinct angular correlation values.

An important feature of this result is that when there are enough pulsar pairs at a particular angular separation, then the variance is not reduced by adding more pulsar pairs at that angular separation. Here, “enough pairs” means that the average value of the covariance matrix rows at that angle are approximately constant. This is because (D14) depends only upon the matrix  $s_{jk}$ . That matrix is determined by the the average value of each block of the covariance matrix. Once there are enough pulsar pairs at a given angle, adding more pulsar pairs at that angle does not change the average value of the block.

### Appendix E: “Beating” the cosmic variance (a simple example)

Here, with a simple example, we show that the variance of the binned estimator with two distinct angles can be *smaller* than the cosmic variance for a point estimator at a single angle.

We take an angular separation bin containing a large number of pulsar pairs separated by two distinct angles,  $\gamma_1$  and  $\gamma_2$ , and uniformly distributed around the sky. Let  $\gamma_1 \approx 54^\circ$  lie at the leftmost minimum of the cosmic variance  $\sigma_{\cos}^2(\gamma)$ , as shown in the left panel of Fig. 4. We will demonstrate that the joint (binned) estimator has a *smaller* variance than this. This is possible because the binned estimator uses information from correlation measurements made at *both* angular separations  $\gamma_1$  and  $\gamma_2$ . By sacrificing some angular resolution, the variance of the binned estimator drops below the minimum cosmic variance. Such behavior is visible in Fig. 5.

We assume that there are  $n_{\text{pairs}}$  pulsar pairs at each of the two angles, so that the covariance matrix  $C$  is  $2n_{\text{pairs}} \times 2n_{\text{pairs}}$  in size. As described in App. D, and shown in (D14), in the limit as  $n_{\text{pairs}} \rightarrow \infty$ , the variance of the optimal estimator can be calculated from the  $2 \times 2$  matrix

$$s = s_{jk} = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma_{22}^2 \end{pmatrix}. \tag{E1}$$

The four values that appear in  $s$  are the average values of the entries in each of the four corresponding  $n_{\text{pairs}} \times n_{\text{pairs}}$  blocks  $C'_{jk}$ , where  $j, k = 1, 2$ .

An explicit expression for  $\sigma_{jk}^2$  was derived in (4.1)

$$\sigma_{jk}^2 = \hbar^4 \int_0^\pi d\beta \sin \beta \left( \mu_{++}(\gamma_j, \beta) \mu_{++}(\gamma_k, \beta) + \mu_{\times \times}(\gamma_j, \beta) \mu_{\times \times}(\gamma_k, \beta) \right). \tag{E2}$$

As shown in (4.4), for large  $n_{\text{pairs}}$ , the variances of the individual (single-angle) optimal estimators approach the cosmic variances, so  $\sigma_{11}^2 \equiv \sigma_1^2 = \sigma_{\cos}^2(\gamma_1)$  and  $\sigma_{22}^2 \equiv \sigma_2^2 = \sigma_{\cos}^2(\gamma_2)$ .

Since (E2) has the form of a positive definite inner product, the Schwarz inequality ensures that  $\sigma_{12}^4 \leq \sigma_1^2 \sigma_2^2$ . Thus, we can parameterize the off-diagonal elements of  $s$  as

$$\sigma_{12}^2 = r \sigma_1 \sigma_2, \tag{E3}$$



where the Schwarz inequality ensures that the real number  $r \in [-1, 1]$ . Note that  $\sigma_{12}^2$  may have either sign.

The variance  $\sigma_{\text{opt}}^2$  of the optimal binned estimator is expressed in terms of  $\mu^T C^{-1} \mu$  by (3.11). This is evaluated in (D14), in terms of the matrix  $s$  of (E1). In that way, we obtain

$$\begin{aligned} \frac{\mu_{\text{norm}}^2}{\sigma_{\text{opt}}^2} &= \mu^T C^{-1} \mu = \hat{\mu}^T s^{-1} \hat{\mu} \\ &= \frac{1}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^4} (\mu_1^2 \sigma_2^2 + \mu_2^2 \sigma_1^2 - 2\mu_1 \mu_2 \sigma_{12}^2) \\ &= \frac{1}{1 - r^2} \left( \frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2} - 2r \frac{\mu_1 \mu_2}{\sigma_1 \sigma_2} \right). \end{aligned} \quad (\text{E4})$$

To obtain the second line, we explicitly invert the  $2 \times 2$  matrix in (E1), and set  $\hat{\mu} = (\mu_1, \mu_2)^T$  where  $\mu_1 \equiv \mu_u(\gamma_1)$  and  $\mu_2 \equiv \mu_u(\gamma_2)$ . For the final line, we use (E3) to eliminate  $\sigma_{12}^2$ .

In signal processing and data analysis, the quantities

$$S_{\text{opt}} \equiv h^2 \mu_{\text{norm}} / \sigma_{\text{opt}}, \quad S_1 \equiv h^2 \mu_1 / \sigma_1, \quad S_2 \equiv h^2 \mu_2 / \sigma_2, \quad (\text{E5})$$

which appear in (E4) are called expected signal-to-noise ratios (SNRs). We included the factors of  $h^2$  on the rhs of (E5) so that the numerators have the interpretation of correlation  $\rho$ . Larger (absolute) SNR values indicate that the mean can be determined with higher fractional precision. Because the mean of the Hellings and Downs correlation varies with angle  $\gamma$ , it is these SNRs (rather than the variance) that reflect the additional precision obtained by combining data from multiple angles. We will see that a loss in precision is not possible (the converse Santa Claus Principle in action [24]).

A simple geometric argument shows that combining data from the two angular separation bins always increases the SNR. First, multiply (E4) by  $h^4$  and use (E5) to obtain

$$S_{\text{opt}}^2 = \frac{1}{1 - r^2} (S_1^2 + S_2^2 - 2r S_1 S_2). \quad (\text{E6})$$

Then introduce new variables  $x$  and  $y$  so that

$$\frac{S_1}{S_{\text{opt}}} = \frac{x - y}{\sqrt{2}}, \quad \frac{S_2}{S_{\text{opt}}} = \frac{x + y}{\sqrt{2}}. \quad (\text{E7})$$

Equation (E6) satisfied by the three SNRs may now be written

$$\frac{x^2}{1 + r} + \frac{y^2}{1 - r} = 1, \quad (\text{E8})$$

which is the equation of an ellipse with major radius  $\sqrt{1 + |r|}$  and minor radius  $\sqrt{1 - |r|}$ , rotated by 45 degrees with respect to axes defined by  $S_1$  and  $S_2$ . This is shown in Fig. 9. Note that  $0 \leq \text{minor radius} \leq 1 \leq \text{major radius} \leq \sqrt{2}$ . This means that the ellipse always lies within the square, and hence  $S_{\text{opt}}^2 \geq S_1^2$  and  $S_{\text{opt}}^2 \geq S_2^2$ . Thus, the optimal binned estimator always has a SNR that is greater than or equal to the SNRs for the individual (narrow-bin) estimators corresponding to the angular separations  $\gamma_1$  and  $\gamma_2$ .

Some insight can be gained from two special cases: (i) If the correlation measurements at the two discrete angles are uncorrelated with one another (i.e.,  $r = 0$ ), then the ellipse is simply the unit circle and the signal-to-noise ratios add in quadrature:

$$S_{\text{opt}}^2 = S_1^2 + S_2^2. \quad (\text{E9})$$

(ii) For perfectly correlated or anti-correlated correlation measurements  $r = \pm 1$ , and  $S_2 = \pm S_1$ . The ellipses degenerate into lines as shown in Fig. 9, and one has

$$S_{\text{opt}}^2 = S_1^2 = S_2^2. \quad (\text{E10})$$

In this second case, the SNR of the optimal binned estimator has the same value as that for the (narrow-bin) estimator for either angle separately, since no new information is obtained by considering the second angle.

Finally, to show that  $\sigma_{\text{opt}}^2$  can be less than the global minimum of the cosmic variance, we take  $\gamma_1 \approx 54^\circ$  at a minima of the cosmic variance curve, and  $\gamma_2 \approx 49^\circ$  at a zero of the expected Hellings and Downs correlation, so  $\mu_2 = 0$ . Pick  $\mu_{\text{norm}}$  to lie anywhere between  $\mu_2 = 0$  and  $\mu_1$ . We have proved that  $S_{\text{opt}}^2 \geq S_1^2$ , which implies from (E5) that

$$\sigma_{\text{opt}}^2 \leq \frac{\mu_{\text{norm}}^2}{\mu_1^2} \sigma_1^2 < \sigma_1^2. \quad (\text{E11})$$

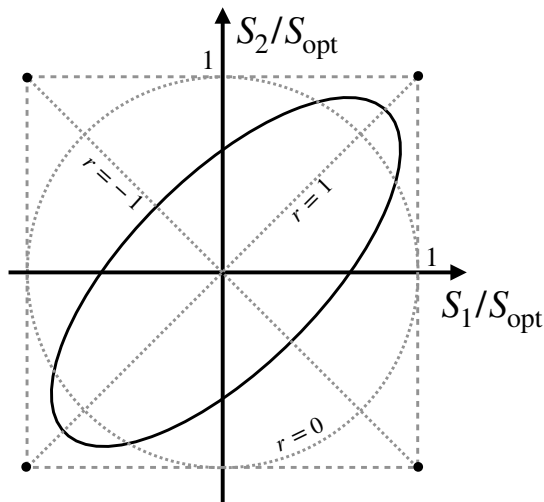


FIG. 9. Equation (E6) constrains the ratios of SNRs,  $S_1/S_{\text{opt}}$  and  $S_2/S_{\text{opt}}$ , to lie on an ellipse. The shape of the ellipse is determined by  $r \in [-1, 1]$ , and is also shown for  $r = 0$  (dotted unit circle) and for  $r = \pm 1$  (degenerate dotted line segments at  $\pm 45^\circ$ ). The ellipse *always* lies within the square, proving geometrically that  $S_{\text{opt}}^2 \geq S_1^2$  and  $S_{\text{opt}}^2 \geq S_2^2$ .

Since  $\sigma_1^2$  is at the minimum of the cosmic variance,  $\sigma_{\text{opt}}^2$  lies below that. As mentioned at the start of this appendix, this behavior is apparent in Fig. 5. (For producing the plot in Fig. 5, we choose the normalization condition on the weights to be  $\mu_{\text{norm}} \equiv \mu^\top \mathbb{1}/n_{\text{pairs}} = (\mu_1 + \mu_2)/2 = \mu_1/2$ , leading to  $\sigma_{\text{opt}}^2 \leq \sigma_1^2/4$  or  $\sigma_{\text{opt}} \leq \sigma_{\text{cos}}/2$ .)

#### Appendix F: Sky locations of pulsars currently monitored by pulsar timing array collaborations

Table I gives the sky locations of the 88 pulsars currently monitored by the European Pulsar Timing Array (EPTA) collaboration, the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) collaboration, and the Parkes Pulsar Timing Array (PPTA) collaboration. We used these to create Figs. 3, 6 and 7; see Sec. VI B) for details. The sky locations are in equatorial coordinates, with right ascension (ra) and declination (dec) given in degrees. They are related to spherical polar coordinates  $\theta$  and  $\phi$  via  $\text{ra} = \phi$  and  $\text{dec} = \pi/2 - \theta$ .

- 
- [1] R. S. Foster and D. C. Backer, “Constructing a pulsar timing array,” *ApJ* **361**, 300–308 (1990).
  - [2] R. W. Hellings and G. S. Downs, “Upper limits on the isotropic gravitational radiation background from pulsar timing analysis.” *ApJ* **265**, L39–L42 (1983).
  - [3] S. R. Taylor, L. Lentati, S. Babak, P. Brem, J. R. Gair, A. Sesana, and A. Vecchio, “All correlations must die: Assessing the significance of a stochastic gravitational-wave background in pulsar timing arrays,” *Phys. Rev. D* **95**, 042002 (2017).
  - [4] Zaven Arzoumanian, Paul T. Baker, Harsha Blumer, Bence Bécsey, Adam Brazier, and others (Nanograv Collaboration), “The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background,” *ApJ* **905**, L34 (2020).
  - [5] Boris Goncharov *et al.*, “On the Evidence for a Common-spectrum Process in the Search for the Nanohertz Gravitational-wave Background with the Parkes Pulsar Timing Array,” *Astrophys. J. Lett.* **917**, L19 (2021).
  - [6] S. Chen *et al.*, “Common-red-signal analysis with 24-yr high-precision timing of the European Pulsar Timing Array: inferences in the stochastic gravitational-wave background search,” *Mon. Not. Roy. Astron. Soc.* **508**, 4970–4993 (2021).
  - [7] J. Antoniadis *et al.*, “The International Pulsar Timing Array second data release: Search for an isotropic gravitational wave background,” *Mon. Not. Roy. Astron. Soc.* **510**, 4873–4887 (2022).
  - [8] Joseph D. Romano, Jeffrey S. Hazboun, Xavier Siemens, and Anne M. Archibald, “Common-spectrum process versus cross-correlation for gravitational-wave searches using pulsar timing arrays,” *Phys. Rev. D* **103**, 063027 (2021).
  - [9] Bruce Allen, “Variance of the Hellings-Downs Correlation,” arXiv e-prints, arXiv:2205.05637 (2022).
  - [10] Here, “different distances” means differing by more than a few characteristic GW wavelengths, so tens of years.
  - [11] Bruce Allen and Joseph D. Romano, “Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities,” *Phys. Rev. D* **59**, 102001 (1999).
  - [12] L. Isserlis, “On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables,” *Biometrika* **12**, 134–139 (1918).

TABLE I. The 88 pulsars currently employed by the European Pulsar Timing Array, the North American Nanohertz Observatory for Gravitational Waves, and the Parkes Pulsar Timing Array collaborations. The final column identifies which PTAs employ this pulsar. The sky location is given by right ascension and declination in degrees. The International Pulsar Timing Array (IPTA) pulsars are the full list.

Pulsar name	ra (deg)	dec (deg)	PTA	Pulsar name	ra (deg)	dec (deg)	PTA
J0023+0923	5.8	9.4	N	J1721-2457	260.3	-25.0	E
J0030+0451	7.6	4.9	EN	J1730-2304	262.6	-23.1	ENP
J0034-0534	8.6	-5.6	E	J1732-5049	263.2	-50.8	P
J0218+4232	34.5	42.5	E	J1738+0333	264.7	3.6	EN
J0340+4130	55.1	41.5	N	J1741+1351	265.4	13.9	N
J0406+3039	61.6	30.7	N	J1744-1134	266.1	-11.6	ENP
J0437-4715	69.3	-47.3	NP	J1745+1017	266.4	10.3	N
J0509+0856	77.3	8.9	N	J1747-4036	267.0	-40.6	N
J0557+1551	89.4	15.8	N	J1751-2857	267.9	-29.0	EN
J0605+3757	91.3	38.0	N	J1801-1417	270.5	-14.3	E
J0610-2100	92.6	-21.0	EN	J1802-2124	270.5	-21.4	EN
J0613-0200	93.4	-2.0	ENP	J1804-2717	271.1	-27.3	E
J0614-3329	93.5	-33.5	N	J1811-2405	272.8	-24.1	N
J0621+1002	95.3	10.0	E	J1824-2452A	276.1	-24.9	P
J0636+5128	99.0	51.5	N	J1832-0836	278.1	-8.6	NP
J0645+5158	101.5	52.0	N	J1843-1113	280.9	-11.2	EN
J0709+0458	107.3	5.0	N	J1853+1303	283.5	13.1	EN
J0711-6830	108.0	-68.5	P	J1857+0943	284.4	9.7	ENP
J0740+6620	115.2	66.3	N	J1903+0327	285.8	3.5	N
J0751+1807	117.8	18.1	E	J1909-3744	287.4	-37.7	ENP
J0900-3144	135.2	-31.7	E	J1910+1256	287.5	12.9	EN
J0931-1902	142.8	-19.0	N	J1911+1347	288.0	13.8	EN
J1012+5307	153.1	53.1	E	J1911-1114	288.0	-11.2	E
J1012-4235	153.1	-42.6	N	J1918-0642	289.7	-6.7	EN
J1017-7156	154.5	-71.9	P	J1923+2515	290.8	25.3	N
J1022+1001	155.7	10.0	ENP	J1939+2134	294.9	21.6	ENP
J1024-0719	156.2	-7.3	ENP	J1944+0907	296.0	9.1	N
J1045-4509	161.5	-45.2	P	J1946+3417	296.6	34.3	N
J1125+7819	171.5	78.3	N	J1955+2908	298.9	29.1	EN
J1125-6014	171.5	-60.2	P	J2010-1323	302.7	-13.4	EN
J1312+0051	198.2	0.9	N	J2017+0603	304.3	6.1	N
J1446-4701	221.6	-47.0	P	J2019+2425	304.9	24.4	E
J1453+1902	223.4	19.0	N	J2033+1734	308.4	17.6	EN
J1455-3330	223.9	-33.5	EN	J2043+1711	310.8	17.2	N
J1545-4550	236.5	-45.8	P	J2124-3358	321.2	-34.0	ENP
J1600-3053	240.2	-30.9	ENP	J2129-5721	322.3	-57.4	P
J1603-7202	240.9	-72.0	P	J2145-0750	326.5	-7.8	ENP
J1614-2230	243.7	-22.5	N	J2214+3000	333.7	30.0	N
J1630+3734	247.7	37.6	N	J2229+2643	337.5	26.7	EN
J1640+2224	250.1	22.4	EN	J2234+0611	338.6	6.2	N
J1643-1224	250.9	-12.4	ENP	J2241-5236	340.4	-52.6	P
J1705-1903	256.4	-19.1	N	J2302+4442	345.7	44.7	N
J1713+0747	258.5	7.8	ENP	J2317+1439	349.3	14.7	EN
J1719-1438	259.8	-14.6	N	J2322+2057	350.6	21.0	EN

- [13] Michael Schmelling, “Averaging correlated data,” *Physica Scripta* **51**, 676–679 (1995).
- [14] Albert Lazzarini, Sukanta Bose, Peter Fritschel, Martin McHugh, Tania Regimbau, Kaice Reilly, Joseph D. Romano, John T. Whelan, Stan Whitcomb, and Bernard F. Whiting, “Optimal combination of signals from colocated gravitational wave interferometers for use in searches for a stochastic background,” *Phys. Rev. D* **70**, 062001 (2004).
- [15] Robert M. Wald, *General Relativity* (Chicago University Press, Chicago, 1984).
- [16] Melissa Anholm, Stefan Ballmer, Jolien D. E. Creighton, Larry R. Price, and Xavier Siemens, “Optimal strategies for gravitational wave stochastic background searches in pulsar timing data,” *Phys. Rev. D* **79**, 084030 (2009).
- [17] Sydney J. Chamberlin, Jolien D. E. Creighton, Xavier Siemens, Paul Demorest, Justin Ellis, Larry R. Price, and Joseph D. Romano, “Time-domain implementation of the optimal cross-correlation statistic for stochastic gravitational-wave background searches in pulsar timing data,” *Phys. Rev. D* **91**, 044048 (2015).
- [18] Sarah J. Vigeland, Kristina Islo, Stephen R. Taylor, and Justin A. Ellis, “Noise-marginalized optimal statistic: A robust hybrid frequentist-bayesian statistic for the stochastic gravitational-wave background in pulsar timing arrays,” *Phys. Rev. D* **98**, 044003 (2018).
- [19] Patrick M. Meyers and Michele Vallisneri, Private communication, May (2022).

- [20] R. Blandford, R. Narayan, and R. W. Romani, “Arrival-time analysis for a millisecond pulsar.” *Journal of Astrophysics and Astronomy* **5**, 369–388 (1984).
- [21] Ronald W. Hellings, “Pulsar timing and gravitational waves.” in *Relativistic Gravitational Experiments in Space: proceedings of NASA Workshop, Annapolis, Maryland June 28-30, 1988.*, NASA Conference Publication, Vol. 3046 (1989) pp. 93–97.
- [22] Jeffrey S. Hazboun, Joseph D. Romano, and Tristan L. Smith, “Realistic sensitivity curves for pulsar timing arrays,” *Phys. Rev. D* **100**, 104028 (2019).
- [23] Since PTA pulsar observations have a maximum cadence of about once per week, the upper limits of the integrals in (B2) should not extend to infinity, so the ratio  $h^4/f^4$  also depends upon this upper cutoff. However, the integrand is dominated by the low-frequency contributions, so this dependence is weak: extending the upper cutoff to infinity has no significant effect.
- [24] The Santa Claus Principle holds that “there is no Santa Claus”, meaning that one does not get free gifts. Conversely, since here we are giving something up (angular resolution) we must gain something in exchange (higher SNR).