

Mitigating cosmic variance in the Hellings-Downs curve: a Cosmic Microwave Background analogy

Cyril Pitrou^{1,*} and Giulia Cusin^{1,2,†}

¹*Institut d'Astrophysique de Paris, UMR-7095 du CNRS et de Sorbonne Université, Paris, France*

²*Département de Physique Théorique and Center for Astroparticle Physics,
Université de Genève, Quai E. Ansermet 24, CH-1211 Genève 4, Switzerland*

(Dated: April 25, 2025)

The Hellings-Downs (HD) correlation, which characterizes the signature of a stochastic gravitational wave background measured via Pulsar Timing Arrays (PTA), is derived using a harmonic formalism. This approach closely follows the theoretical framework traditionally employed to compute correlations of temperature fluctuations in the Cosmic Microwave Background (CMB). This parallel enables a direct comparison between the correlations observed in PTA and those in CMB experiments. After providing analytic estimates of the transmission functions due to the model-fitting procedure, we show that the covariance matrix in frequency space becomes very nondiagonal. We then build formally the quadratic estimator for the HD correlation in multipolar space, for both a perfect experiment, and for a realistic pulsar noise model. For a perfect experiment, we show that the signal to noise ratio grows with the observation time and the number of frequency bins, in turn determined by the cadence of observation. For an imperfect experiment, the behaviour is similar, with an effective multipole-dependent number of frequency, obtained after weighting with noise. We predict that with ~ 200 pulsars monitored for 25 years, multipoles of the HD correlation up to $\ell = 4$ can be measured. Our findings clarify that what has been referred to as *cosmic variance* in some of the previous literature is not an intrinsic limitation for PTA measurements. Instead, when an optimal estimator is used, this variance can be mitigated by accumulating more observation time or improving the cadence of pulsar monitoring. Therefore, unlike CMB angular correlations, where cosmic variance represents an irreducible constraint, it can be reduced in PTA measurements and will continue to diminish in future experiments. Finally, we show that if the primordial power spectrum of tensor fluctuations was very blue with $n_T > 4$, the CMB angular correlation due to these tensor modes would also exhibit a HD correlation. While our results are derived for the case of an isotropic bath of gravitons, we also discuss the case in which the graviton distribution function is anisotropic.

CONTENTS

I. Introduction	2	D. Multipole space and frequency estimator	10
II. GW background: general concepts and definitions	3	VI. Fitting the residuals	11
A. GW mode expansion	3	A. Discrete Fourier series	11
B. Statistics of the GW background	4	B. Templates fitting	11
C. GW background spectrum	4	VII. Combining different estimators	13
III. Pulsar redshift	5	A. Nonweighted time average	14
A. General expression	5	B. Noise contributions	15
B. Observer term	5	1. Noise due to Poissonian pulsar statistics	16
C. Pulsar term	6	2. Noise of individual pulsars	16
IV. Pulsar redshift correlation	7	C. Optimal weighting	17
A. Direction and time basis	7	1. Perfect experiment	17
B. Direction and frequency basis	7	2. Imperfect experiment	18
C. Multipolar and time basis	7	VIII. Background anisotropy estimator	20
D. Multipolar and frequency basis	8	A. Anisotropy in the time residual map: estimator and variance	20
E. Adding the pulsar term	8	B. Angular power spectrum of energy density: estimator and variance	21
F. Timing residuals	8	IX. The Hellings-Downs curve in the CMB	22
V. Quadratic estimators	8	A. GW in cosmology	22
A. Directional space and time estimator	9	B. Redshift multipole transfer function	23
B. Directional space and frequency estimator	10	C. Correlation function	24
C. Multipole space and time estimator	10	X. Discussion and Conclusions	25
		A. Rotations and spherical harmonics	25
		B. Quadratic estimator for perfect experiments	26

* pitrou@iap.fr

† cusin@iap.fr

C. Quadratic estimator for imperfect experiments	27
D. Pulsar shot noise	28
E. Hellings-Downs correlation in angular space	29
References	30

I. INTRODUCTION

Pulsar Timing Arrays (PTAs) were used to provide the first evidence of a stochastic gravitational wave (GW) background in the nHz band [1–4], with evidence of spatial correlation among different pulsar redshifts following the Hellings-Downs (HD) function. Various detection methods are reviewed in [5], and the physical interpretation of the HD correlation is discussed in [6, 7]. Methods for mapping the background have also been developed [8–11] (see also [12, 13] for a discussion of polarisation with Stokes parameters), and it has recently been proposed that a loud source may account for the evidence of anisotropy reported in [14]. Due to the all-sky nature of the signal, a harmonic formulation of the HD analysis has been introduced in [15–23], which forms the basis for our harmonics analysis. The most recent analysis of the HD signature by NANOGrav [24] is expressed in multipole space, with a clear detection of the quadrupole and only marginal evidence for the octupole. The significance of the background detection is expected to improve in the future, as the signal-to-noise ratio (SNR) of PTA observables increases with the observation time [19, 25, 26]. The spectral function of the GW background from binaries is known to be approximately a power law [27], and its exponent can also be constrained. However, to eliminate sources of contamination, time residuals for each pulsar must be fit with a linear drift and a spin-down, while the modulation due to Earth’s motion must be accounted for, which alters the low-frequency part of the spectrum [28–30].

To evaluate the error and construct estimators for the pulsar correlation function, it is crucial to understand the covariance of the quadratic estimators. Recent efforts have focused on obtaining the covariance in the most general scenarios [31], with particular attention to source discreteness [7, 31–35] and the anisotropic nature of the underlying large-scale structure [32, 36, 37]. It has recently been argued [7, 31–35] that, since PTAs operate in a regime where time averaging over observation periods limits the sampling of many realizations of the Universe, any attempt to measure the HD correlation is affected by cosmic variance. This cosmic variance is often considered an irreducible limitation on our ability to constrain the HD correlation with any PTA experiment, even in the absence of instrumental noise [7, 31, 34, 38]. This is in contradiction with the intuition proposed a few years ago in [16] that since PTAs are sensitive to a range of frequencies, the effect of cosmic variance should be reduced by properly combining different frequency bins. Indeed, estimates of the SNR scaling based on simulated PTA data suggest that sensitivity continues to improve with increasing observation time [19],

with no indication of reaching a fundamental limitation. This intuition was further developed theoretically in the recent article [39] where the authors build an optimal estimator for HD correlation as function of pulsar separation, exploiting the frequency information. However, confusion persists in the literature, with cosmic variance still being referred to as a fundamental limitation in very recent studies.

In this article, we present a pedagogical derivation of how the cosmic variance issue is mitigated by optimally combining different frequency bins. Unlike [39], when building the optimal estimator, we work in multipole space rather than in real space. The advantage of working in multipole space is that, in the strong signal limit, the covariance is fully diagonal, and a quadratic optimal estimator can be built, with a more transparent structure than the one derived in [39] in real space. We also consider the effect of adding instrumental noise and we present in detail how the optimal estimator is modified when accounting for the fact that in a real experiment one has to fit time residuals with a deterministic template, to subtract the effect of a linear drift and a quadratic term. We show that the final claim about the SNR is not altered by this filtering procedure.

Since PTAs measure timing residuals, which are closely related to pulsar redshifts, the HD correlation in PTAs has an analogous counterpart in CMB observations, when considering the correlation between redshift perturbations (or temperature anisotropies) induced by GWs in the early Universe. We identify the CMB analogue of the HD correlation, which, although conceptually similar, exhibits a more complex structure compared to the PTA HD correlation. This comparison allows us to evaluate the impact of cosmic variance in both types of experiments. We find that while, for the CMB, cosmic variance is an irreducible source of uncertainty on our ability to reconstruct the angular power spectrum, it does not pose a fundamental limitation for PTA experiments, provided that the correct estimator is used. When using an optimal estimator, we find indeed that the SNR grows with the observation time and with the cadence of pulsar monitoring, hence there is no fundamental limitation in our ability to measure the HD curve. For the CMB, the redshift is indirectly inferred from the spectrum or the energy of the photons received, which is already averaged over the duration of the experiment. Additionally, the intrinsic timescale of variations in the underlying signal is of the order of millions of years. As a result, the situation is akin to performing a single measurement of the CMB at a specific moment in time. The duration of the experiment mainly serves to reduce the experimental error in this single measurement, while it plays no role in a perfect experiment. Exploiting this parallel, we also show that, if the primordial power spectrum of tensor fluctuations was very blue with $n_T > 4$, the CMB angular correlation due to tensor modes would also exhibit a HD correlation.

In our derivation of an estimator for the HD correlation, we first work in real space of time and directions. The dependence on directions can be reformulated in spherical harmonic components, and the time dependence can be expressed in frequency space. Hence we will consider four different but equivalent formulations of the correlation function and the

corresponding four (equivalent) estimators. Multipole space turns out to be a particularly suitable representation for the HD correlation as the angular power spectrum of pulsar redshifts is diagonal. The advantage of this approach is that the covariance of any estimator is diagonal in both frequency and multipole space, taking the standard form that is familiar from CMB studies (see also [15] for CMB mapping methods applied to PTA mapping).

Our article is structured as follows: in section II, we present general concepts and definitions and in section III we discuss pulsar redshift in multipole space, followed by the computation of the two-point correlation in both real and multipole space in section IV. Section V introduces different representations (in time and direction space, in frequency and direction space, and in multipole space) of a quadratic estimator for the pulsar redshift correlation function. In section VI we discuss how to consistently go from the ideal case of a continuous frequency spectrum, to discrete frequency space, and we compute how the pulsar two-point function gets modified when accounting for the finite observation time and filtering out spurious contributions with a proper time modeling. In section VII, we show that cosmic variance arises when different time (or frequency) measurements are naively combined, and we demonstrate that an optimal estimator that properly weights low frequencies mitigates the cosmic variance problem. We build formally the quadratic estimator for the HD correlation in multipolar space, for both a perfect experiment, and for a realistic pulsar noise model. For a perfect experiment, we show that the signal to noise ratio increases with the observation time and the cadence of measurements. For an imperfect experiment, the signal to noise ratio depends on an effective multipole-dependent number of frequency, which increases with the observation time, but which also depends on the frequency dependent level of noise. We predict that with ~ 200 pulsars monitored for 25 years, multipoles of the HD correlation up to $\ell = 4$ can be measured. Our findings show that variance is not an intrinsic limitation, as it can be mitigated by accumulating more observation time or improving the cadence of pulsar monitoring. Finally in section IX we present the terms of the analogy with the CMB and we discuss our findings in section X. Our results are derived for the case of an isotropic bath of gravitons. In section VIII we discuss the case in which the graviton distribution function is anisotropic: we build an optimal estimator for anisotropies and we compute its variance, recovering results of [18]. The fact that a given time-residual map has a finite angular resolution fixed by the number of pulsars in a network reflects on our ability to constrain the angular power spectrum of the background energy density, even for a perfect experiment. Technical details and mathematical derivations are presented in a series of appendices.

II. GW BACKGROUND: GENERAL CONCEPTS AND DEFINITIONS

In this section we detail how gravitational waves are expanded on tensor valued plane waves, or harmonic functions,

and we review the statistical properties of the modes of this expansion. Note that we are eventually interested in writing the redshift of a pulsar in terms of GW from the pulsar to the observer. Since the pulsars monitored are galactic objects, we can assume spacetime to be flat (neglecting cosmological expansion). This implies that GWs satisfy $\square h_{ij} = 0$ where the d'Alembertian is the flat space one. On the other hand, when working out the CMB analogy, we will have to extend the treatment of this section to a cosmological background, accounting for the Universe expansion.

A. GW mode expansion

Tensor fluctuations, describing GW, can be expanded on a suitable basis of tensor valued harmonic functions. In flat space, and for a wave propagating along the z axis, the harmonic functions are given by¹

$$Q_{ij}^{(\pm 2)}(\mathbf{k} = k\mathbf{e}_z, \mathbf{x}) = e_{ij}^{(\pm 2)}(e_z) e^{i\mathbf{k} \cdot \mathbf{x}}, \quad (1)$$

where the helicity tensor basis is

$$e_{ij}^{(\pm 2)}(e_z) \equiv (e_i^x \pm i e_i^y)(e_j^x \pm i e_j^y). \quad (2)$$

For a wave propagating in a general direction, we simply rotate the previous construction. If the mode direction is $\hat{\mathbf{k}} = (\theta_k, \phi_k)$, then we use the rotation $R_{\hat{\mathbf{k}}} = R(\phi_k, \theta_k, 0)$, with the Euler angles defined in (A2), so as to build another harmonic function

$$Q_{ij}^{(\pm 2)}(\mathbf{k}, \mathbf{x}) = R_{\hat{\mathbf{k}}} [Q_{ij}^{(\pm 2)}(k\mathbf{e}_z, \mathbf{x})]. \quad (3)$$

Hence a general harmonic function can be expressed in terms of a suitable helicity basis as

$$Q_{ij}^{(\pm 2)}(\mathbf{k}, \mathbf{x}) \equiv e_{ij}^{(\pm 2)}(\hat{\mathbf{k}}) e^{i\mathbf{k} \cdot \mathbf{x}}, \quad (4a)$$

$$e_{ij}^{(\pm 2)}(\hat{\mathbf{k}}) \equiv (R_{\hat{\mathbf{k}}})_i^k (R_{\hat{\mathbf{k}}})_j^l e_{kl}^{(\pm 2)}(e_z). \quad (4b)$$

The essential properties of this construction are

$$e_{ij}^{(\pm 2)}(-\hat{\mathbf{k}}) = e_{ij}^{(\mp 2)}(\hat{\mathbf{k}}), \quad (5a)$$

$$Q_{ij}^{(\pm 2)}(-\mathbf{k}, \mathbf{x}) = Q_{ij}^{(\pm 2)*}(\mathbf{k}, \mathbf{x}). \quad (5b)$$

With this basis, there are several, but equivalent, ways to decompose a GW background as a superposition of plane waves coming from different directions, depending on how we treat the time dependence. The first method consists in using progressive waves. The GW reads in the local Minkowski metric with time coordinate η

$$h_{ij}(\eta, \mathbf{x}) = \text{Re} \left(\int \frac{d^3 \mathbf{k}}{(2\pi)^3} \sum_{m=\pm 2} h^{(m)}(\mathbf{k}) e^{ik\eta} Q_{ij}^{(m)}(\mathbf{k}, \mathbf{x}) \right), \quad (6)$$

¹ The normalization is different from the one used in [40] or [41] in the context of CMB fluctuations (tensorial contributions), but it is consistent with the usual notation for the treatment of a stochastic background of GW.

where we used the definition (4a) and $h^{(m)}(\mathbf{k})$ is the coefficient in front of a plane wave propagating in the direction $-\hat{\mathbf{k}}$, that is a wave which we see coming from the direction $\hat{\mathbf{k}}$.² For future use we introduce the following modes

$$h_{\cos}^{(m)}(\mathbf{k}, \eta) \equiv \frac{1}{2} \left(h^{(m)}(\mathbf{k}) e^{ik\eta} + h^{(m)*}(-\mathbf{k}) e^{-ik\eta} \right), \quad (7)$$

such that

$$h_{\cos}^{(m)*}(\mathbf{k}, \eta) = h_{\cos}^{(m)}(-\mathbf{k}, \eta), \quad (8)$$

and using (5b) the GW expansion reads

$$h_{ij} = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \sum_{m=\pm 2} h_{\cos}^{(m)}(\mathbf{k}, \eta) Q_{ij}^{(m)}(\mathbf{k}, \mathbf{x}), \quad (9)$$

without need to take the real part as in (6). Finally, we introduce the combination

$$H^{(m)}(\mathbf{k}, \eta) \equiv \frac{1}{2} \left(h^{(m)}(\mathbf{k}) e^{ik\eta} + h^{(-m)*}(\mathbf{k}) e^{-ik\eta} \right), \quad (10)$$

which satisfies

$$H^{(m)*}(\mathbf{k}, \eta) = H^{(-m)}(\mathbf{k}, \eta). \quad (11)$$

B. Statistics of the GW background

For an unpolarized and stationary background, the two-point function of the wave modes satisfy the following relations

$$\begin{aligned} \langle h^{(m)}(\mathbf{k}) h^{(m')*}(\mathbf{k}') \rangle &= (2\pi)^3 \delta_{mm'} \delta(\mathbf{k} - \mathbf{k}') 2\mathcal{A}(\mathbf{k}), \\ \langle h^{(m)}(\mathbf{k}) h^{(m')*}(\mathbf{k}') \rangle &= 0, \end{aligned} \quad (12)$$

which are independent correlation rules. The average $\langle \dots \rangle$ denotes an ensemble average over the many possible realization of the GW background.

The 2-point function of the modes $h_{\cos}^{(m)}$ can be derived from their definition (7)

$$\begin{aligned} \langle h_{\cos}^{(m)}(\mathbf{k}, \eta) h_{\cos}^{(m')*}(\mathbf{k}', \eta') \rangle &= \delta_{mm'} (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}') \\ &\quad \times \mathcal{A}(\mathbf{k}) \cos[k(\eta - \eta')], \quad (13) \\ \langle h_{\cos}^{(m)}(\mathbf{k}, \eta) h_{\cos}^{(m')*}(\mathbf{k}', \eta') \rangle &= \delta_{mm'} (2\pi)^3 \delta(\mathbf{k} + \mathbf{k}') \\ &\quad \times \mathcal{A}(\mathbf{k}) \cos[k(\eta - \eta')]. \end{aligned}$$

where the latter is not independent from the former because of the relation (8). The fact that the time dependence of the two-point correlation is only a function of $\eta - \eta'$ is a consequence of the assumed stationarity. Similarly, using the definition (10) we get for $H^{(m)}$

$$\begin{aligned} \langle H^{(m)}(\mathbf{k}, \eta) H^{(m')*}(\mathbf{k}', \eta') \rangle &= \delta_{mm'} (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}') \\ &\quad \times \mathcal{A}(\mathbf{k}) \cos[k(\eta - \eta')]. \end{aligned} \quad (14)$$

Again the result for $\langle H^{(m)} H^{(m')} \rangle$ is not independent and can be found from property (11).

C. GW background spectrum

The GW spectrum $\mathcal{A}(\mathbf{k})$, introduced in section II B, can be interpreted as a distribution function of a gas of gravitons up to some factors and powers of \mathbf{k} . In the context of PTA observations, this interpretation implicitly assumes that the function $\mathcal{A}(\mathbf{k})$ has been promoted to a local function $\mathcal{A}(\mathbf{k}, \mathbf{x} = 0)$ at the observer position, in turn identified with the pulsar network position. In other words, we consider that the set of pulsars is very localized and that we have a distribution function describing the gas of gravitons at the pulsar network position. This approach is valid only if the sources are very far compared to the distance between pulsars (local detector assumption). Note that when defining a distribution function from quantum field theory we do exactly the same thing: we define an occupation number average which is in full generality global, and then we localize it (e.g. with a Wigner transform), see [42, 43].

Let us parametrize the angular dependence of the distribution function as

$$\mathcal{A}(\mathbf{k}) = \mathcal{A}(k) \hat{\mathcal{A}}(\hat{\mathbf{k}}), \quad \hat{\mathcal{A}}(\hat{\mathbf{k}}) = \sum_{LM} \hat{\mathcal{A}}_{LM} Y_L^M(\hat{\mathbf{k}}), \quad (15)$$

where $k = |\mathbf{k}|$ and $\hat{\mathcal{A}}_{L-M} = (-1)^M \hat{\mathcal{A}}_{LM}^*$ since the GW spectrum is real valued. This decomposition assumes the existence of a factorization between a spectral shape $\mathcal{A}(k)$, and an angular-dependent function $\hat{\mathcal{A}}(\hat{\mathbf{k}})$. Unless stated otherwise we shall assume an isotropic function, that is $\hat{\mathcal{A}}(\hat{\mathbf{k}}) = 1$.

The isotropic part $\mathcal{A}(k)$ is related to the background energy density, which can be quantified in units of the critical energy density with

$$\Omega_{\text{GW}} = \int_0^\infty \Omega_{\text{GW}}(f) d \ln f, \quad \Omega_{\text{GW}}(f) = \frac{8\pi^2}{3H_0^2} f^3 \hat{\mathcal{P}}(f), \quad (16)$$

where $k = 2\pi f$, and the (two-sided, meaning that needs to be integrated on both positive and negative frequencies) spectral function is

$$\hat{\mathcal{P}}(f) \equiv 4\pi f^2 \mathcal{A}(2\pi f). \quad (17)$$

The spectral function in turn is directly related to the variance of the norm of the GW fluctuations since with (9) and (13) we find

$$\langle h_{ij}(\eta_0) h^{ij}(\eta_0) \rangle = 4h^2, \quad (18)$$

where following the notation of [22] we introduced

$$h^2 \equiv 2 \int_0^\infty \frac{4\pi k^2 dk}{(2\pi)^3} \mathcal{A}(k) = \int_{-\infty}^\infty \hat{\mathcal{P}}(f) df, \quad (19)$$

with the convention $\hat{\mathcal{P}}(-f) = \hat{\mathcal{P}}(f)$. The factor 4 in (18) comes from the normalization of the polarization basis (2), and the factor 2 in (19) is the contribution of the two polarizations. The quantity $2h^2$ is interpreted as the variance due to one polarization, and we therefore define the characteristic strain in a (one-sided) log-frequency band as $h_c^2(f) = 4f \hat{\mathcal{P}}(f)$ so that $2h^2 = \int_0^\infty h_c^2(f) d \ln f$.

² This decomposition is close to but different from the (7.1) of Allen [22]: we do not use negative frequencies and we introduce both left and right polarizations. The normalization we use is also different.

It is common in the literature to parametrize the spectral density as [44]

$$\hat{\mathcal{P}}(f) = \hat{\mathcal{P}}_* \left(\frac{f}{f_*} \right)^\gamma, \quad (20)$$

with a spectral index $\gamma = -7/3$, such that $\Omega_{\text{GW}}(f) \propto f^{2/3}$, for a background composed by the superposition of GW emitted by a population of compact binaries in the inspiralling phase (see e.g. chapter 23 of [44]).

III. PULSAR REDSHIFT

We detail how gravitational waves induce a redshift z of the pulsar signal, distinguishing an observer and a pulsar contribution, and discussing their multipolar decomposition.

A. General expression

For a pulsar in a given direction \mathbf{n} and at distance $\chi \equiv |\mathbf{x}| = \eta_0 - \eta$, the redshift to the pulsar is defined as (see e.g. chapter 23 of [44])

$$z(\chi, \mathbf{n}) = \frac{1}{2} \int_0^\chi d\chi' n^i n^j \partial_\eta h_{ij}(\eta_0 - \chi', \chi' \mathbf{n}), \quad (21)$$

where to simplify the notation the dependence on η_0 is omitted here on the left-hand side of equations. In the following, we split the redshift into an observer term and a pulsar term as

$$z(\chi, \mathbf{n}) = z^o(\mathbf{n}) - z^p(\chi, \mathbf{n}), \quad (22)$$

with

$$z^o(\mathbf{n}) = \frac{1}{2} \int_0^\infty d\chi' n^i n^j \partial_\eta h_{ij}(\eta_0 - \chi', \chi' \mathbf{n}), \quad (23a)$$

$$z^p(\chi, \mathbf{n}) = \frac{1}{2} \int_\chi^\infty d\chi' n^i n^j \partial_\eta h_{ij}(\eta_0 - \chi', \chi' \mathbf{n}). \quad (23b)$$

The first contribution corresponds to the redshift one would measure for an infinitely distant pulsar, the second term is the same but computed from the pulsar position (as if the observer was located on the pulsar).

We now seek to expand the observer and pulsar terms as the product of a wave term and a geometrical factor expanded in spherical harmonics, in the form

$$z^{o/p}(\mathbf{n}) = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \sum_{m=\pm 2} \sum_{\ell} z_{\ell m}^{o/p}(\mathbf{k}, \eta_0) R_{\hat{\mathbf{k}}} [Y_\ell^m](\mathbf{n}), \quad (24)$$

where from now on we always consider $\ell \geq 2$, unless specified otherwise. To that purpose, we use the GW plane wave expansion (6), hence we need an explicit expression for the tensor $Q_{ij}^{(m)}(\mathbf{k}, \mathbf{x})$ contracted with $n^i n^j$. For a GW propagating along the e_z axis one has (just like for tensor modes in the CMB context [40, 41])

$$Q_{ij}^{(m)}(ke_z, \mathbf{x}) n^i n^j = - \sum_{\ell \geq 2} c_\ell \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \frac{j_\ell(k\chi)}{(k\chi)^2} Y_\ell^m(\mathbf{n}), \quad (25)$$

and

$$c_\ell \equiv i^\ell \sqrt{4\pi(2\ell+1)}. \quad (26)$$

If the plane wave is in a general direction as defined in (3) we only need to rotate the result using that

$$\begin{aligned} R_i^k R_j^l Q_{kl}^{(m)}(ke_z, \mathbf{x}) n^i n^j &= Q_{kl}^{(m)}(ke_z, \mathbf{x}) R_i^k R_j^l n^i n^j \\ &= Q_{ij}^{(m)}(ke_z, \mathbf{x}) (R^{-1} \cdot \mathbf{n})^i (R^{-1} \cdot \mathbf{n})^j, \end{aligned} \quad (27)$$

and one finally gets

$$Q_{ij}^{(m)}(\mathbf{k}, \mathbf{x}) n^i n^j = - \sum_{\ell \geq 2} c_\ell \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \frac{j_\ell(k\chi)}{(k\chi)^2} R_{\hat{\mathbf{k}}} [Y_\ell^m(\mathbf{n})], \quad (28)$$

where, using (A3), a rotated spherical harmonics is

$$R_{\hat{\mathbf{k}}} [Y_\ell^m(\mathbf{n})] \equiv \sum_M D_{Mm}^\ell(R_{\hat{\mathbf{k}}}) Y_\ell^M(\mathbf{n}), \quad (29)$$

where D_{Mm}^ℓ are Wigner rotation coefficients defined in appendix A.

B. Observer term

We now focus on the observer term. From eqs. (6), (23a) and (28) we first get the expansion

$$z^o(\mathbf{n}) = \text{Re} \left(\int \frac{d^3 \mathbf{k}}{(2\pi)^3} \sum_{m=\pm 2} h^{(m)}(\mathbf{k}) e^{ik\eta_0} F_0^{(m)}(\hat{\mathbf{k}}, \mathbf{n}) \right), \quad (30)$$

where $F_0^{(m)}$ is the antenna pattern function in the helicity basis, explicitly given by

$$F_0^{(m)}(\hat{\mathbf{k}}, \mathbf{n}) = \sum_{\ell \geq 2} F_\ell^o R_{\hat{\mathbf{k}}} [Y_\ell^m(\mathbf{n})], \quad (31)$$

with multipolar coefficients

$$F_\ell^o \equiv -\frac{i}{2} \sqrt{4\pi(2\ell+1)} \int_0^\infty i^\ell \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \frac{j_\ell(x)}{x^2} e^{-ix} dx. \quad (32)$$

Using the relation [21]³

$$\int_0^\infty \frac{j_\ell(x)}{x^2} e^{\mp ix} dx = 2(\pm i)^{1-\ell} \frac{(\ell-2)!}{(\ell+2)!}, \quad (33)$$

one gets the compact expression

$$F_\ell^o \equiv \sqrt{4\pi(2\ell+1)} \sqrt{\frac{(\ell-2)!}{(\ell+2)!}}, \quad (34)$$

³ It is a special case of Eq. 6.621.1 in [45] using $j_\ell(x) = \sqrt{\pi/(2x)} J_{\ell+1/2}(x)$, $\alpha = \pm i$, $\beta = 1$, $\nu = \ell + 1/2$ and $\mu = -5/2$ along with ${}_2F_1(\ell/2 - 1/2, \ell/2; \ell + 3/2; 1) = 2^{2+\ell} \Gamma(\ell + 3/2) / \Gamma(\ell + 3) / \sqrt{\pi}$.

where we recall that $\ell \geq 2$. Given the relation (A7) between Wigner rotation coefficients and spin-weighted spherical harmonics, the expansion (31) with (29) is equivalent to (2.10) in [22].

Let us examine the properties of the antenna pattern function (31). It is first clear that

$$F_o^{(m)}(R \cdot \hat{\mathbf{k}}, R \cdot \mathbf{n}) = F_o^{(m)}(\hat{\mathbf{k}}, \mathbf{n}), \quad (35)$$

which implies that changing the direction of a GW wave or the direction of observation in the opposite direction have the same effect. From the conjugation properties (A10) and (A8) we deduce

$$(R_{\hat{\mathbf{k}}}[Y_\ell^m])^* = (-1)^m R_{\hat{\mathbf{k}}}[Y_\ell^{-m}], \quad (36)$$

hence using that for antenna patterns $m = \pm 2$, we deduce

$$F_o^{(m)*}(\hat{\mathbf{k}}, \mathbf{n}) = F_o^{(-m)}(\hat{\mathbf{k}}, \mathbf{n}). \quad (37)$$

With this property, the real part in (30) can be removed if we use the $H^{(m)}$ combination (10), as

$$z^o(\mathbf{n}) = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \sum_{m=\pm 2} H^{(m)}(\mathbf{k}, \eta_0) F_o^{(m)}(\hat{\mathbf{k}}, \mathbf{n}), \quad (38)$$

hence we deduce that the multipole expansion (24) is satisfied with

$$z_{\ell m}^o(\mathbf{k}, \eta_0) = H^{(m)}(\mathbf{k}, \eta_0) F_\ell^o. \quad (39)$$

However, we also get from (A9)

$$R_{-\hat{\mathbf{k}}}[Y_\ell^m(\mathbf{n})] = (-1)^\ell R_{\hat{\mathbf{k}}}[Y_\ell^{-m}(\mathbf{n})] = R_{\hat{\mathbf{k}}}[Y_\ell^{-m}(-\mathbf{n})], \quad (40)$$

hence we deduce

$$F_o^{(m)}(-\hat{\mathbf{k}}, -\mathbf{n}) = F_o^{(-m)}(\hat{\mathbf{k}}, \mathbf{n}) = F_o^{(m)*}(\hat{\mathbf{k}}, \mathbf{n}). \quad (41)$$

Therefore, the expansion (24) is equally satisfied with

$$z_{\ell m}^o(\mathbf{k}, \eta_0) = \frac{1}{2} \left[h^{(m)}(\mathbf{k}) e^{ik\eta_0} + (-1)^\ell h^{(m)*}(-\mathbf{k}) e^{-ik\eta_0} \right] F_\ell^o. \quad (42)$$

C. Pulsar term

We now consider the pulsar term. We get

$$z^p(\mathbf{n}) = \text{Re} \left(\int \frac{d^3 \mathbf{k}}{(2\pi)^3} \sum_{m=\pm 2} h^{(m)}(\mathbf{k}) e^{ik\eta_0} F_p^{(m)}(k\chi, \hat{\mathbf{k}}, \mathbf{n}) \right), \quad (43)$$

where the antenna pattern function is now defined as

$$F_p^{(m)}(k\chi, \hat{\mathbf{k}}, \mathbf{n}) = \sum_{\ell \geq 2} F_\ell^p(k\chi) R_{\hat{\mathbf{k}}}[Y_\ell^m(\mathbf{n})], \quad (44)$$

with

$$F_\ell^p(k\chi) = -\frac{i}{2} \sqrt{4\pi(2\ell+1)} \int_{k\chi}^\infty i^\ell \sqrt{\frac{(\ell+2)!}{(\ell-2)!}} \frac{j_\ell(x)}{x^2} e^{-ix} dx, \quad (45)$$

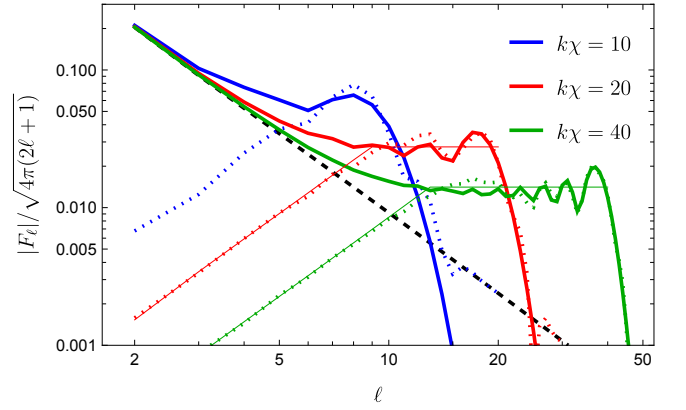


FIG. 1. Multipolar coefficients of the antenna pattern function: observer term (F_ℓ^o , black dashed line), pulsar term (F_ℓ^p , dotted lines) and total term (F_ℓ , continuous lines). The thin solid straight lines correspond to the approximation (50). Different colors correspond to different values of $k\chi$ used for evaluating the pulsar term.

and it explicitly depends on the distance χ and on the frequency k (no compact expression of the type of (34) exists in this context even though the integral can be expressed in a quite involved manner with hypergeometrical functions).

One can derive an approximation for the pulsar angular pattern modes F_ℓ^p . Recalling that for large arguments, spherical Bessel functions behave as

$$j_\ell(x) \sim \frac{1}{x} \sin\left(x - \ell \frac{\pi}{2}\right), \quad (46)$$

and using that

$$\int_X^\infty \frac{\cos^2(x)}{x^3} dx \simeq \int_X^\infty \frac{\sin^2(x)}{x^3} dx \simeq \frac{1}{4X^2}, \quad (47)$$

we can evaluate approximately (45) for low ℓ (meaning $\ell \ll \ell_{\max} \equiv k\chi$)

$$\frac{-F_\ell^{p,\text{low}}}{\sqrt{4\pi(2\ell+1)}} \simeq \frac{1}{8\ell_{\max}^2} \sqrt{\frac{(\ell+2)!}{(\ell-2)!}}. \quad (48)$$

For larger ℓ , one can approximate the norm of the left-hand side by a constant. Choosing

$$\frac{-F_\ell^{p,\text{high}}(k\chi)}{\sqrt{4\pi(2\ell+1)}} \simeq \frac{1}{\sqrt{3}\ell_{\max}}, \quad (49)$$

allows to recover the expected form of the correlation function at zero separation, as it will become clear in section IV E. A reasonable description consists in choosing

$$-F_\ell^p = \min\left(-F_\ell^{p,\text{low}}, -F_\ell^{p,\text{high}}\right), \quad (50)$$

for $2 \leq \ell \leq \ell_{\max} = k\chi$. In Fig. 1 we plot F_ℓ for the observer term and for the pulsar contribution for various $k\chi$.

IV. PULSAR REDSHIFT CORRELATION

We now want to compute the expectation of the pulsar redshift correlation $z^o(\mathbf{n}_1, \eta_1)z^o(\mathbf{n}_2, \eta_2)$ under the statistics of the graviton gas introduced in section II B. We will from now on denote this average as $\langle \dots \rangle_{\mathcal{A}}$ to avoid confusion with other averaging procedures that will be introduced later in the text. We will focus on the observer term only, and add the contribution of the pulsar term to the pulsar redshift in a second step. The dependence on directions can be reformulated in spherical harmonic components, and the time dependence can be expressed in frequency space. Hence we will consider four different but equivalent formulations of the correlation function.

A. Direction and time basis

The distributional average of the pulsar redshift correlation is

$$\langle z^o(\mathbf{n}_1, \eta_1)z^o(\mathbf{n}_2, \eta_2) \rangle_{\mathcal{A}} = \int \frac{d^3 \mathbf{k}_1}{(2\pi)^3} \frac{d^3 \mathbf{k}_2}{(2\pi)^3} \sum_{m_1 m_2} \quad (51)$$

$$\langle H^{(m_1)}(\mathbf{k}_1, \eta_1) H^{(m_2)*}(\mathbf{k}_2, \eta_2) \rangle_{\mathcal{A}} F_0^{(m_1)}(\hat{\mathbf{k}}_1, \mathbf{n}_1) F_0^{(m_2)*}(\hat{\mathbf{k}}_2, \mathbf{n}_2).$$

With the statistics (14), this selects $p_1 = p_2$, and assuming an isotropic distribution $\mathcal{A}(k)$ for the graviton gas, this reduces to a single integral on momenta

$$\langle z^o(\mathbf{n}_1, \eta_1)z^o(\mathbf{n}_2, \eta_2) \rangle_{\mathcal{A}} = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \quad (52)$$

$$\sum_{m=\pm 2} \mathcal{A}(k) \cos[k(\eta_1 - \eta_2)] F_0^{(m)}(\hat{\mathbf{k}}, \mathbf{n}_1) F_0^{(m)*}(\hat{\mathbf{k}}, \mathbf{n}_2).$$

Expressing the antenna pattern functions with (31) and using the definition (29) we need to take the angular average over the mode direction $\hat{\mathbf{k}}$ since $d^3 \mathbf{k} = k^2 dk d^2 \hat{\mathbf{k}}$. From (A7) and the orthogonality relation (A13), this is

$$\int d^2 \hat{\mathbf{k}} D_{M_1 m}^{\ell_1}(R_{\hat{\mathbf{k}}}) D_{M_2 m}^{\ell_2*}(R_{\hat{\mathbf{k}}}) = \delta_{\ell_2}^{\ell_1} \delta_{M_2}^{M_1} \frac{4\pi}{2\ell_1 + 1}. \quad (53)$$

Finally, using the addition theorem (A14) we get

$$\langle z^o(\mathbf{n}_1, \eta_1)z^o(\mathbf{n}_2, \eta_2) \rangle_{\mathcal{A}} = \mathcal{P}(\eta_1 - \eta_2) \mu(\beta_{12}), \quad (54)$$

where the unequal time power is

$$\mathcal{P}(\eta_1 - \eta_2) \equiv \int \frac{4\pi k^2 dk}{(2\pi)^3} 2\mathcal{A}(k) \cos[k(\eta_1 - \eta_2)], \quad (55)$$

while the function μ , expressed in terms of the angle β_{12} between \mathbf{n}_1 and \mathbf{n}_2 ($\cos \beta_{12} \equiv \mathbf{n}_1 \cdot \mathbf{n}_2$), is the HD curve (without pulsar term)

$$\mu(\beta_{12}) = \sum_{\ell \geq 2} \mu_{\ell}(\beta_{12}), \quad \mu_{\ell}(\beta_{12}) \equiv \frac{|F_{\ell}^o|^2}{4\pi} P_{\ell}(\cos \beta_{12}). \quad (56)$$

With the results of appendix E, this takes the famous re-summed form

$$\mu(\beta_{12}) = \frac{1}{3} - \frac{1}{6}x + x \ln x, \quad x \equiv \frac{1 - \cos \beta_{12}}{2}. \quad (57)$$

For equal times, we recover

$$\langle z^o(\mathbf{n}_1, \eta)z^o(\mathbf{n}_2, \eta) \rangle_{\mathcal{A}} = h^2 \mu(\beta_{12}), \quad (58)$$

where h^2 is defined in (19). The shape (57) of the HD correlation is plotted in Fig. 10.

B. Direction and frequency basis

Defining the Fourier transform of the pulsar redshift as

$$z(\mathbf{n}, \eta) = \int_{-\infty}^{\infty} \hat{z}(\mathbf{n}, f) e^{i2\pi f \eta} df \quad (59)$$

where

$$\hat{z}^*(\mathbf{n}, f) = \hat{z}(\mathbf{n}, -f), \quad (60)$$

and using

$$\mathcal{P}(\eta_1 - \eta_2) = \int_{-\infty}^{\infty} \hat{\mathcal{P}}(f) e^{i2\pi f(\eta_1 - \eta_2)} df, \quad (61)$$

where the spectral function is defined in (17), the correlation in frequency basis becomes

$$\langle \hat{z}^o(\mathbf{n}_1, f) \hat{z}^{o,*}(\mathbf{n}_2, f') \rangle_{\mathcal{A}} = \delta(f - f') \hat{\mathcal{P}}(f) \mu(\beta_{12}). \quad (62)$$

C. Multipolar and time basis

The spherical harmonics expansion of the pulsar redshift is defined as

$$z^o(\mathbf{n}, \eta) = \sum_{\ell m} z_{\ell m}^o(\eta) Y_{\ell}^m(\mathbf{n}). \quad (63)$$

Since z is real valued, $(-1)^m z_{\ell -m}^o = z_{\ell m}^{o*}$. We want to find the two-point function of multipolar coefficients (the angular power spectrum). To this aim, one can repeat the steps of the computation leading to (54) until the point in which the summation on spherical harmonics is used. Then using the orthogonality of spherical harmonics (A13), the statistics of the $z_{\ell m}^o$ is inferred immediately and one gets

$$\langle z_{\ell m}^o(\eta) z_{\ell' m'}^{o*}(\eta') \rangle_{\mathcal{A}} = \delta_{\ell \ell'} \delta_{m m'} C_{\ell}^o(\eta, \eta'), \quad (64)$$

with

$$C_{\ell}^o(\eta, \eta') = \mathcal{P}(\eta - \eta') C_{\ell}^{\text{HD}}, \quad (65)$$

and the HD multipoles are

$$C_{\ell}^{\text{HD}} \equiv \frac{|F_{\ell}^o|^2}{2\ell + 1} = 4\pi \frac{(\ell - 2)!}{(\ell + 2)!}. \quad (66)$$

For equal times we get simply $C_{\ell}^o(\eta) \equiv C_{\ell}^o(\eta, \eta)$ which is

$$C_{\ell}^o(\eta) = h^2 C_{\ell}^{\text{HD}}, \quad (67)$$

and does not depend on η .

D. Multipolar and frequency basis

Finally, we consider the Fourier transform of (63). The statistics of the multipoles $\hat{z}_{\ell m}^0(f)$, which are the Fourier transform of $z_{\ell m}^0(\eta)$, is given by

$$\langle \hat{z}_{\ell m}^0(f) \hat{z}_{\ell' m'}^{0*}(f') \rangle_{\mathcal{A}} = \delta_{\ell\ell'} \delta_{mm'} \delta(f-f') C_{\ell}^0(f), \quad (68)$$

where

$$C_{\ell}^0(f) = \hat{\mathcal{P}}(f) C_{\ell}^{\text{HD}}, \quad (69)$$

is the Fourier transform of (65). **From the reality of z the multipoles in Fourier space $\hat{z}_{\ell m}^{0*}(f) = (-1)^m \hat{z}_{\ell -m}^0(-f)$.** In the multipolar and frequency basis, the correlation is fully diagonal. For the frequency part, this is due to the assumed stationarity of the underlying GW background. For the multipolar components, this is due to the fact that we assumed an isotropic function $\mathcal{A}(k)$.

Let us briefly discuss the effects of an anisotropic spectrum: if $\hat{\mathcal{A}}(\hat{\mathbf{k}}) \neq 1$, the correlation in equation (68) acquires off-diagonal contributions. This can be demonstrated by following the steps outlined in section IV A up to equation (53). However, this latter expression cannot be directly applied due to the angular dependence of $\hat{\mathcal{A}}(\hat{\mathbf{k}})$. Therefore, one must rely on the properties of spherical harmonics, as detailed in appendix A. The Wigner coefficients are written in terms of spin-weighted spherical harmonics, as shown in equation (A7), and the integral involving three spin-weighted spherical harmonics is given by equation (A16). Ultimately, we obtain

$$\begin{aligned} \langle \hat{z}_{\ell_1 m_1}^0(f_1) \hat{z}_{\ell_2 m_2}^0(f_2) \rangle_{\mathcal{A}} &= \delta_{\ell_1 + \ell_2 + L}^{\text{even}} \delta(f_1 + f_2) \hat{\mathcal{P}}(f_1) F_{\ell_1}^0 F_{\ell_2}^0 \\ &\times \sum_{LM} \frac{\hat{\mathcal{A}}_{LM}^*}{\sqrt{4\pi}} \sqrt{2L+1} \begin{pmatrix} L & \ell_1 & \ell_2 \\ M & m_1 & m_2 \end{pmatrix} \begin{pmatrix} L & \ell_1 & \ell_2 \\ 0 & -2 & 2 \end{pmatrix}, \quad (70) \end{aligned}$$

where $\delta_{\ell_1 + \ell_2 + L}^{\text{even}} = 1$ if $\ell_1 + \ell_2 + L$ is even and vanishes if it is odd. The isotropic result (68) is easily recovered when using $\hat{\mathcal{A}}_{LM} = \delta_L^0 \delta_M^0 \sqrt{4\pi}$ and (A19). The structure of the correlations (70) is similar to the one induced in the CMB by an anisotropic power spectrum [46, 47]. Note that the anisotropic part of $\hat{\mathcal{A}}(\hat{\mathbf{k}})$ is responsible for the off-diagonal contributions, but it also modifies the diagonal correlations.

E. Adding the pulsar term

Let us now look at the contribution of the pulsar term to the correlation defined in (57). We define the total correlation function including both observer and pulsar terms as

$$\mu^{\text{tot}}(k\chi, \beta_{12}) = \sum_{\ell \geq 2} \frac{|F_{\ell}^0 - F_{\ell}^{\text{p}}(k\chi)|^2}{4\pi} P_{\ell}(\cos \beta_{12}), \quad (71)$$

and we plot it in Fig. 2, using for the mode functions F_{ℓ} eqs. (34) and (45). Interference between pulsar and observer terms can be neglected and the pulsar contribution converges to the usual $(1/3)\delta_{n_1 n_2}$ variance in the limit $k\chi \gg 1$. Indeed,

using for the pulsar term the parametrization in (50), we observe that for $k\chi \rightarrow \infty$

$$\sum_{\ell=2}^{\infty} \frac{|F_{\ell}^{\text{p}}(k\chi)|^2}{4\pi} \simeq \sum_{\ell=2}^{\ell_{\text{max}}} \frac{|F_{\ell}^{\text{p,high}}(k\chi)|^2}{4\pi} \rightarrow \frac{1}{3}. \quad (72)$$

In other words, at zero separation the pulsar contribution to the correlation function tends to 1/3 for $k\chi \rightarrow \infty$ (this is not exact for finite $k\chi$). Explicitly, for $k\chi \gg 1$

$$\mu^{\text{tot}}(\beta_{12}) \simeq \mu(\beta_{12}) + \frac{1}{3} \delta^c(\beta_{12}), \quad (73)$$

where $\delta^c(\beta_{12})$ is a kind of generalized Kronecker function δ , on the set of real numbers, i.e. a function which is 1 when $\beta_{12} = 0$, and vanishes otherwise. In other words, we have $\mu^{\text{tot}}(0) \rightarrow 2/3$ as both the observer and pulsar parts converge to 1/3, but $\mu^{\text{tot}}(\beta_{12}) = \mu(\beta_{12})$ for $\beta_{12} > 0$. The total correlation is plotted in the left panel of Fig. 2, and the pulsar term alone is plotted in its right panel.

Notice that in a realistic PTA situation the condition $k\chi \gg 1$ is largely satisfied. Indeed even for a distance χ as small as 156 pc (distance to the closest pulsar) and k of order of the minimum value in the PTA frequency range, $k = 2\pi f \simeq 2\pi \times 3 \times 10^{-9}$ Hz, $k\chi$ is of order 300, which becomes about 1900 for a millisecond pulsar at a more typical distance of 1 kpc.

From now on in our analysis we discard the contribution of the pulsar term, since its support is zero (and in observations often only pairs of different pulsars are used to get rid of noise in individual pulsars).

F. Timing residuals

In PTA experiments we observe timing residuals which are related to the pulsar redshift evolution by

$$r(\eta) = \int_{\eta_0}^{\eta} z(\eta') d\eta'. \quad (74)$$

This implies that timing residuals are easily obtained from the redshift expressions by integration. As detailed below in section VIB, we can ignore the constant term, related to the time of reference when expressing residuals, as it is removed when fitting the residuals [28, 30]. Therefore, the expressions for timing residuals are obtained from the redshift ones just by the replacements

$$z \rightarrow r, \quad \hat{\mathcal{P}}(f) \rightarrow \frac{\hat{\mathcal{P}}(f)}{(2\pi f)^2}, \quad \mathcal{A}(k) \rightarrow \frac{\mathcal{A}(k)}{k^2}, \quad (75)$$

meaning that in practice for residuals one has to use $\gamma = -13/3$ in (20) when considering a background from massive black hole binaries.

V. QUADRATIC ESTIMATORS

We now introduce a quadratic estimator for the pulsar 2-point function. Similar to how we defined four types of

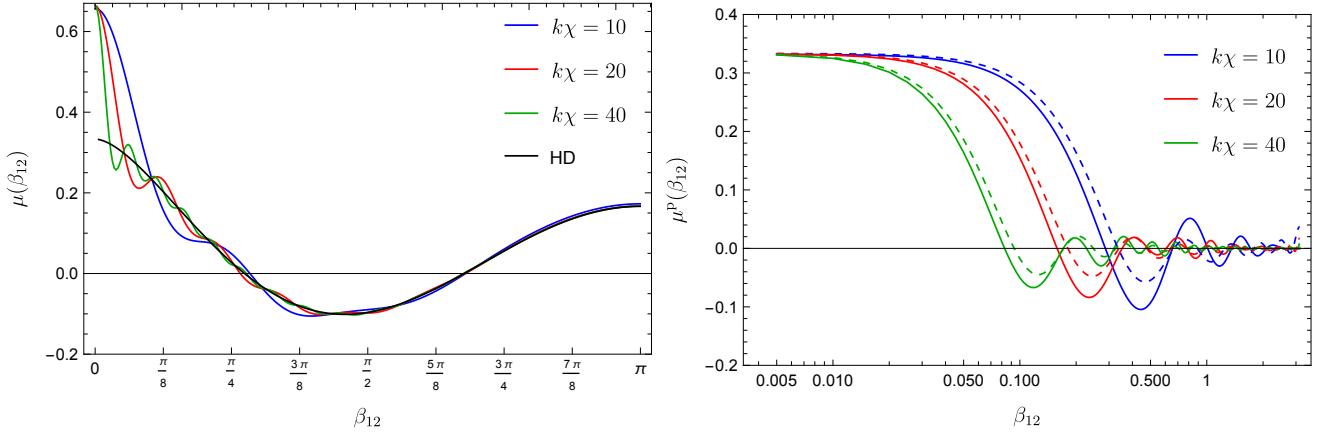


FIG. 2. *Left*: Total correlation function, including the pulsar term contribution. Different lines correspond to different values of $k\chi$ while the black line is the HD curve. We observe that as we increase $k\chi$ the pulsar term rapidly converges to give a contribution equal to $1/3$ at zero separation, and vanishing contributions for finite separation. *Right*: Pulsar term contribution to the correlation function. The continuous line is the exact result, and the dashed line is the approximation (50).

quadratic averages based on different combinations of direction or multipoles, and time or frequency domain, we will have four corresponding flavors of quadratic estimators. In section VIII, we will also comment on how, in the presence of background anisotropies, an estimator can be built to extract multipoles of the spectral function.

A. Directional space and time estimator

Observationally, in order to build the correlation function, one measures pulsar redshifts and take a spatial average over all pulsar pairs separated by a given angle: a pulsar average. We define the following estimator

$$\Gamma^e(\beta_{12}; \eta_1, \eta_2) = \langle z(\mathbf{n}_1, \eta_1) z(\mathbf{n}_2, \eta_2) \rangle_R, \quad (76)$$

where β_{12} is the angle between \mathbf{n}_1 and \mathbf{n}_2 . The average over pairs is defined as

$$\langle z(\mathbf{n}_1, \eta_1) z(\mathbf{n}_2, \eta_2) \rangle_R \equiv \int \frac{d^3 R}{8\pi^2} z(R \cdot \mathbf{n}_1, \eta_1) z(R \cdot \mathbf{n}_2, \eta_2), \quad (77)$$

where the differential rotation element $d^3 R$ is defined in appendix A. Using (31) along with orthogonality of Wigner rotation matrices (A11), we get

$$\langle F^{(m_1)}(\hat{\mathbf{k}}, \mathbf{n}_1) F^{(m_2)*}(\hat{\mathbf{k}}, \mathbf{n}_2) \rangle_R = \delta_{m_1 m_2} \sum_{\ell} \frac{|F_{\ell}|^2}{4\pi} P_{\ell}(\mathbf{n}_1 \cdot \mathbf{n}_2), \quad (78)$$

hence using (38), the estimator is expressed as

$$\Gamma^e(\beta_{12}; \eta_1, \eta_2) = \int \frac{d^3 \mathbf{k}_1}{(2\pi)^3} \frac{d^3 \mathbf{k}_2}{(2\pi)^3} \times \sum_{m=\pm 2} H^{(m)}(\mathbf{k}_1, \eta_1) H^{(m)*}(\mathbf{k}_2, \eta_2) \sum_{\ell} \frac{|F_{\ell}|^2}{4\pi} P_{\ell}(\cos \beta_{12}). \quad (79)$$

We want to compute the mean and variance of this estimator, averaging over graviton realizations (distributional average). With the statistics (14), the mean is

$$\langle \Gamma^e(\beta_{12}; \eta_1, \eta_2) \rangle_{\mathcal{A}} = \mathcal{P}(\eta_1 - \eta_2) \mu(\beta_{12}), \quad (80)$$

hence (76) is a good estimator which averages to what we are looking for. The assumption of statistical isotropy implies that the mean of the estimator depends only on the difference of directions, while stationarity implies it depends only on the difference of times. In both cases the reciprocal space is much better adapted as we shall detail below.

To simplify the notation, to compute the covariance we consider equal-time correlators and we define

$$\Gamma^e(\beta_{12}; \eta) \equiv \Gamma^e(\beta_{12}; \eta, \eta). \quad (81)$$

The covariance is given by

$$\text{Cov}(\beta_{12}, \eta; \beta'_{12}, \eta') \equiv \langle \hat{\Gamma}(\beta_{12}, \eta) \hat{\Gamma}(\beta'_{12}, \eta') \rangle_{\mathcal{A}} - \langle \hat{\Gamma}(\beta_{12}, \eta) \rangle_{\mathcal{A}} \langle \hat{\Gamma}(\beta'_{12}, \eta') \rangle_{\mathcal{A}}. \quad (82)$$

Using the definition of the estimator (76) and replacing (38), we see that the distributional average acts on products of 4 modes $H^{(m)}(\mathbf{k})$: Wick's theorem gives three pair of terms. The first is canceled by the product of averages, hence we are left with two contributions, the first pairing 13 and 24, and the second pairing 14 and 23, which are equal. We have to perform the integral on the relative direction between \mathbf{k}_1 and \mathbf{k}_2 for which we use

$$\langle F^{(m_1)}(\hat{\mathbf{k}}_1, \mathbf{n}_1) F^{(m_2)*}(\hat{\mathbf{k}}_2, \mathbf{n}_2) \rangle_R = \sum_{\ell} \frac{|F_{\ell}|^2}{4\pi} D_{m_2 m_1}^{\ell}(R_{\hat{\mathbf{k}}_2}^{-1} \cdot R_{\hat{\mathbf{k}}_1}) P_{\ell}(\mathbf{n}_1 \cdot \mathbf{n}_2), \quad (83)$$

which is consistent with (78) since $D_{m_2 m_1}^{\ell}(\mathbb{I}) = \delta_{m_1 m_2}$. Even-

tually, we get

$$\begin{aligned} \text{Cov}(\beta_{12}, \eta; \beta'_{12}, \eta') &= \left(\frac{\mathcal{P}(\eta - \eta')}{2} \right)^2 \sum_{\ell} \frac{|F_{\ell}|^4}{(4\pi)^2} P_{\ell}(\cos \beta_{12}) P_{\ell}(\cos \beta'_{12}) \\ &\quad \times \sum_{m_1=\pm 2} \sum_{m_2=\pm 2} \int_{-1}^1 d\cos \theta |d_{m_1 m_2}^{\ell}(\theta)|^2. \end{aligned} \quad (84)$$

Using (A5) and (A6), the integrals are

$$\int_{-1}^1 d\cos \theta |d_{22}^{\ell}(\theta)|^2 = \int_{-1}^1 d\cos \theta |d_{2-2}^{\ell}(\theta)|^2 = \frac{2}{2\ell+1}, \quad (85)$$

and the final result is

$$\begin{aligned} \text{Cov}(\beta_{12}, \eta; \beta'_{12}, \eta') &= 2\mathcal{P}^2(\eta - \eta') \sum_{\ell} \frac{|F_{\ell}|^4}{(4\pi)^2 (2\ell+1)} P_{\ell}(\cos \beta_{12}) P_{\ell}(\cos \beta'_{12}), \end{aligned} \quad (86)$$

in agreement with (8.4) of [22].

B. Directional space and frequency estimator

One can introduce the Fourier transform of the estimator (76) to work in frequency space. The mean and the covariance are

$$\langle \Gamma^e(\beta_{12}; f_1, f_2) \rangle_{\mathcal{A}} = \delta(f_1 - f_2) \hat{\mathcal{P}}(f_1) \mu(\beta_{12}) \quad (87)$$

and

$$\begin{aligned} \text{Cov}(\beta_{12}, f; \beta'_{12}, f') &= \delta(0) \delta(f - f') \\ \hat{\mathcal{P}}^2(f) \sum_{\ell} \frac{2|F_{\ell}|^4}{(4\pi)^2 (2\ell+1)} P_{\ell}(\cos \beta_{12}) P_{\ell}(\cos \beta'_{12}), \end{aligned} \quad (88)$$

and we observe that everything is diagonal in frequency space, as a consequence of stationarity. In section VI A, we will show that the function $\delta(0)$ has to be interpreted as $\delta(0) = T$, where T is the observation time.

C. Multipole space and time estimator

Now we can build an estimator for the angular power spectrum

$$C_{\ell}^e(\eta, \eta') \equiv \sum_{m=-\ell}^{\ell} \frac{z_{\ell m}(\eta) z_{\ell m}^*(\eta')}{2\ell+1}. \quad (89)$$

We observe that averaging over all m for a given ℓ , is the multipole space equivalent of averaging over all pulsar pairs with a given separation in direction space. The relation to the estimator (76) can be easily found using the expansion (63) and

$$\langle Y_{\ell}^m(\mathbf{n}) Y_{\ell'}^{m'}(\mathbf{n}') \rangle_R = \delta_{\ell\ell'} \delta_{mm'} \frac{1}{4\pi} P_{\ell}(\mathbf{n} \cdot \mathbf{n}'). \quad (90)$$

One finds indeed the relations

$$\Gamma^e(\beta_{12}; \eta_1, \eta_2) = \sum_{\ell} \frac{2\ell+1}{4\pi} C_{\ell}^e(\eta_1, \eta_2) P_{\ell}(\cos \beta_{12}), \quad (91)$$

hence the estimator C_{ℓ}^e is the counterpart of the estimator $\Gamma^e(\beta_{12})$ in multipole space.

From the Gaussian statistics (64), the average of the estimator (89) gives

$$\langle C_{\ell}^e(\eta, \eta') \rangle_{\mathcal{A}} = \mathcal{P}(\eta - \eta') C_{\ell}^{\text{HD}}. \quad (92)$$

with definition (66). Also here, for simplicity we just focus on equal time for the covariance, introducing

$$C_{\ell}^e(\eta) = C_{\ell}^e(\eta, \eta), \quad (93)$$

whose variance, found from Wick theorem and the statistics (64), is given by

$$\begin{aligned} \text{Cov}_{\ell\ell'}(\eta, \eta') &\equiv \langle C_{\ell}^e(\eta) C_{\ell'}^e(\eta') \rangle_{\mathcal{A}} - \langle C_{\ell}^e(\eta) \rangle_{\mathcal{A}} \langle C_{\ell'}^e(\eta') \rangle_{\mathcal{A}} \\ &= \delta_{\ell\ell'}^e \frac{2}{2\ell+1} [C_{\ell}(\eta, \eta')]^2. \end{aligned} \quad (94)$$

This is the usual expression of cosmic variance on an angular power spectrum estimator, where the factor $2\ell+1$ at the denominator corresponds to the independent values of m at a given ℓ . As expected, the covariance of the estimator (89) is related to the one of (76) (eq. (86)) via

$$\begin{aligned} \text{Cov}(\beta_{12}, \eta; \beta'_{12}, \eta') &= \sum_{\ell} \frac{(2\ell+1)^2}{(4\pi)^2} \text{Cov}_{\ell\ell}(\eta, \eta') P_{\ell}(\cos \beta_{12}) P_{\ell}(\cos \beta'_{12}), \end{aligned} \quad (95)$$

as it can be checked directly. At equal time and for equal separations it is exactly Eq. (7) of [48] which was obtained in the CMB context.

D. Multipole space and frequency estimator

Finally, we introduce the Fourier transform of the estimator (89)

$$C_{\ell}^e(f, f') \equiv \sum_{m=-\ell}^{\ell} \frac{\hat{z}_{\ell m}(f) \hat{z}_{\ell m}^*(f')}{2\ell+1}. \quad (96)$$

Its average is given by

$$\langle C_{\ell}^e(f, f') \rangle_{\mathcal{A}} = \delta(f - f') \hat{\mathcal{P}}(f) C_{\ell}^{\text{HD}}. \quad (97)$$

We compute the covariance only at equal frequency defining

$$C_{\ell}^e(f) = C_{\ell}^e(f, f), \quad (98)$$

and we get

$$\begin{aligned} \text{Cov}_{\ell\ell'}(f, f') &= \langle C_{\ell}^e(f) C_{\ell'}^e(f') \rangle_{\mathcal{A}} - \langle C_{\ell}^e(f) \rangle_{\mathcal{A}} \langle C_{\ell'}^e(f') \rangle_{\mathcal{A}} \\ &= \delta_{\ell\ell'}^e \frac{2}{2\ell+1} \delta(0) \delta(f - f') [C_{\ell}(f)]^2. \end{aligned} \quad (99)$$

We observe that the covariance is now fully diagonal. As a result, the multipole and frequency representation of the quadratic estimator is the most convenient, since the inverse covariance is straightforward to compute due to its diagonal form.

VI. FITTING THE RESIDUALS

In this section, we discuss how to consistently go from the ideal case of a continuous frequency spectrum, to discrete frequency space, and we compute how the pulsar two-point function gets modified when accounting for the finite observation time and filtering out spurious contributions with a proper time modelling.

A. Discrete Fourier series

In practice, we do not have access to the full spectrum of frequency as the experiment is run only for a finite duration. For an observation time T , we can only access the discrete Fourier series coefficients for the frequency

$$f_p = p/T \quad \text{with} \quad 1 \leq |p| \leq T f_{\max}, \quad (100)$$

where

$$f_{\max}^{-1} \equiv 2\Delta\eta, \quad (101)$$

is set from the cadence of observations $\Delta\eta$ and the Nyquist criterion. The discrete Fourier series coefficients are defined as

$$\tilde{z}(f_p) = \int_{-T/2}^{T/2} z(t) e^{-2\pi i f_p t} dt. \quad (102)$$

Hence the relation between the Fourier discrete series coefficients and the Fourier transform is

$$\tilde{z}(f_p) = T \int df \hat{z}(f) \text{sinc}(\pi(f - f_p)T). \quad (103)$$

It is standard to use the approximation

$$\text{sinc}(\pi x) \simeq \delta(x), \quad (104)$$

such that

$$\tilde{z}(f_p) \simeq \hat{z}(f_p). \quad (105)$$

We chose not to normalize (102) with the observation time $1/T$ to get this approximate equality between the Fourier transform and the Fourier series coefficients.

The relations traditionally used when going from discrete to continuous and backward are

$$\frac{1}{T} \sum_n \leftrightarrow \int_{\text{obs}} df \equiv \int_{-f_{\max}}^{-1/T} df + \int_{1/T}^{f_{\max}} df, \quad (106a)$$

$$T \delta_{pq} \leftrightarrow \delta(f_p - f_q). \quad (106b)$$

By using these results, the correlations (62) and (68) for discrete frequencies f_p, f_q , are approximately replaced by

$$\langle \tilde{z}(\mathbf{n}_1, f_p) \tilde{z}(\mathbf{n}_2, f_q) \rangle_{\mathcal{A}} \simeq \mu(\beta_{12}) T \delta_{pq} \hat{\mathcal{P}}(f_p), \quad (107a)$$

$$\langle \tilde{z}_{\ell_1 m_1}(f_p) \tilde{z}_{\ell_2 m_2}(\mathbf{n}_2, f_q) \rangle_{\mathcal{A}} \simeq \delta_{\ell_1}^{\ell_2} \delta_{m_1}^{m_2} C_{\ell_1}^{\text{HD}} T \delta_{pq} \hat{\mathcal{P}}(f_p) \quad (107b)$$

Finally, given the correspondence (106b), it follows that the terms $\delta(0)$ appearing in the expressions for the covariance found in the previous section for a continuous frequency spectrum ((88) and (99)) have to be interpreted as $\delta(0) = T$.

B. Templates fitting

The approach in section VI A to go from a continuous frequency spectrum to a discrete one and back, is too simplistic in two aspects. First for a very red spectrum the approximation (104) in (103) fails badly. And second we also need to take into account the fact that the observed timing residual is fitted with a deterministic template, and this has the effect of removing the low frequencies. We now deal with both complications.

As shown in [30], errors in the estimate of timing model parameters lead to deterministic features in the timing residuals. For example, an error in the pulse period leads to timing residuals that grow linearly with time, while an error in the period derivative leads to residuals that grow quadratically with time. Thus, we can improve our estimates of the timing model parameters by fitting for a constant, linear and quadratic term (in some cases also a cubic spin-down model is fitted for), and subtract these terms from the observed residuals. When working with redshift, related to time residual via the time integration (74), one fits with a constant and a linear term (eventually with a quadratic term as well if a cubic spin-down effect is considered in the timing residuals). In the following part of this section, for consistency with the rest of the paper, we will work with pulsar redshifts. However, the treatment of time residuals is straightforward, recalling the relation (74), and when showing results we will consider both these observables.

The redshift of a given pulsar in real space is expressed as a function of the pulsar direction \mathbf{n} and in multipole space as a series of multipoles (ℓ, m) . The treatment we will present in the following can be applied to both real and multipole space, hence, to keep the discussion general, we will simply denote redshift as $z(t)$, omitting direction/multipole dependence. Pulsar redshift is related to its Fourier transform through

$$z(t) = \int \hat{z}(f) e^{i2\pi f t} df, \quad (108)$$

with autocorrelation

$$\langle \hat{z}(f) \hat{z}^*(f') \rangle_{\mathcal{A}} = \delta(f - f') S(f). \quad (109)$$

Concretely, for autocorrelations in direction space $S(f) = \hat{\mathcal{P}}(f) \mu(0)$, and for multipole space autocorrelations $S(f) = \hat{\mathcal{P}}(f) C_{\ell}^{\text{HD}}$. Let us define the scalar product (this corresponds to the ordinary least square described in [30])

$$\{f(t), g(t)\} = \frac{1}{T} \int_{-T/2}^{T/2} f^*(t) g(t) dt. \quad (110)$$

We notice that a set of functions which are orthogonal for this scalar product are the Legendre polynomials P_n since

$$\{P_n(2t/T), P_{n'}(2t/T)\} = \delta_{nn'} \frac{1}{2n+1}. \quad (111)$$

Using

$$\{P_n(2t/T), e^{i2\pi f t}\} = i^n j_n(\pi f T), \quad (112)$$

the pulsar redshift is expanded on Legendre polynomials as

$$z(t) = \sum_n z_n P_n(2t/T), \quad (113)$$

with coefficients

$$z_n \equiv \frac{\{P_n(2t/T), z(t)\}}{\{P_n(2t/T), P_n(2t/T)\}} = i^n (2n+1) \int j_n(\pi f T) \hat{z}(f) df. \quad (114)$$

Since we want to subtract spurious effects by fitting a constant, a linear drift and a quadratic shape (corresponding to a pulsar spin-down), the signal after fitting is

$$z^{\text{pos}}(t) = z(t) - \sum_{n=0}^2 z_n P_n(2t/T). \quad (115)$$

The total initial power (before subtraction) is

$$\begin{aligned} \langle \{z(t), z(t)\} \rangle_{\mathcal{A}} &= \sum_{n=0}^{\infty} \frac{\langle z_n z_n^* \rangle_{\mathcal{A}}}{2n+1} \\ &= \sum_{n=0}^{\infty} \int (2n+1) j_n^2(\pi f T) S(f) df, \end{aligned} \quad (116)$$

which using $\sum_n (2n+1) j_n^2(x) = 1$, can be resummed to give

$$\langle \{z(t), z(t)\} \rangle_{\mathcal{A}} = \int S(f) df, \quad (117)$$

as expected. For the residual signal after fitting polynomials, we get simply

$$\begin{aligned} \langle \{z^{\text{pos}}(t), z^{\text{pos}}(t)\} \rangle_{\mathcal{A}} &= \langle \{z(t), z(t)\} \rangle_{\mathcal{A}} - \sum_{n=0}^2 \frac{\langle z_n z_n^* \rangle_{\mathcal{A}}}{2n+1} \\ &= \sum_{n=3}^{\infty} \int (2n+1) j_n^2(\pi f T) S(f) df, \end{aligned} \quad (118)$$

or equivalently

$$\langle \{z^{\text{pos}}(t), z^{\text{pos}}(t)\} \rangle_{\mathcal{A}} = \int S(f) \mathcal{T}(f) df, \quad (119)$$

where we introduced the transmission, or damping, function

$$\mathcal{T}(f) = 1 - \sum_{n=0}^2 (2n+1) j_n^2(\pi f T). \quad (120)$$

It is easy to check that $\mathcal{T}(f) \propto f^6$ when $fT \ll 1$ as found in [30].

If we also want to fit an annual modulation with frequency f_{earth} , then we only need to fit $e^{\pm i 2\pi f_{\text{earth}} t}$ along with the previous polynomials. If $f_{\text{earth}} \gg 1/T$ then we have approximately $\{P_n(2t/T), e^{\pm i 2\pi f_{\text{earth}} t}\} \simeq 0$ for $n = 0, 1, 2$, hence we can proceed in two steps and fit the annual modulation after the low order polynomials. This amounts to projecting in the space orthogonal to P_0, P_1, P_2 first and then in the space orthogonal to $e^{\pm i 2\pi f_{\text{earth}} t}$. After fitting we then get the residual

$$z^{\text{pos,earth}} = z^{\text{pos}} - \sum_{s=\pm 1} \{e^{i 2\pi s f_{\text{earth}} t'}, z^{\text{pos}}(t')\} e^{i 2\pi s f_{\text{earth}} t}, \quad (121)$$

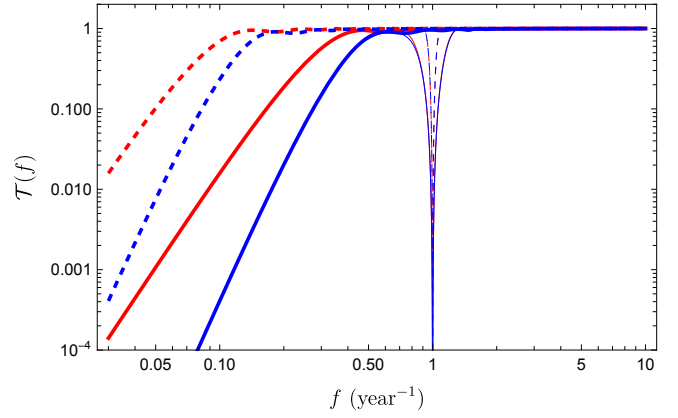


FIG. 3. Transmission function \mathcal{T} (thick lines) and $\mathcal{T}\mathcal{T}^{\text{earth}}$ (thin lines) for $T = 3$ years (continuous lines) and $T = 10$ years (dashed lines). In the red curves, only P_0, P_1 were used in the fitting, whereas for the blue curves P_2 was also used. Hence the first case corresponds to the transmission function for redshifts and the second to the transmission function of time residuals. The slope at low frequency in the former case is $\propto f^4$ and it is $\propto f^6$ in the latter case. The dip at the Earth frequency has a typical width $\Delta f = 1/T$ hence gets narrower for longer observations. This figure is to be compared with Fig. 2 of [30], where a similar results is obtained with different filtering methods.

where the projections are

$$\{e^{\pm i 2\pi f_{\text{earth}} t'}, z(t')\} = \int \hat{z}(f) \text{sinc}[\pi(f \mp f_{\text{earth}})T] df. \quad (122)$$

This translates into a damping of the initial power as

$$\langle \{z^{\text{pos,earth}}(t), z^{\text{pos,earth}}(t)\} \rangle_{\mathcal{A}} \simeq \int S(f) \mathcal{T}(f) \mathcal{T}^{\text{earth}}(f) df, \quad (123)$$

where a product of two transmission functions appears, the transmission function defined in (120) and a new damping term

$$\mathcal{T}^{\text{earth}}(f) = 1 - \text{sinc}^2[\pi(f - f_{\text{earth}})T] - \text{sinc}^2[\pi(f + f_{\text{earth}})T]. \quad (124)$$

This transmission function is plotted in figure 3, together with the transmission function \mathcal{T} defined in (120). Hereafter, we ignore the Earth frequency filtering and consider that we only fit the lowest order polynomials which are essential for the regularity when $f \rightarrow 0$. Finally, if a given pulsar noise is not stationary, one might weight the functional scalar product (110) by replacing $dt \rightarrow W(t)dt$, and one must revise the previous construction by using the polynomials which are orthogonal with this new measure [30]. For simplicity we ignore this aspect.

Up to now we assumed a continuous frequency spectrum. However, to make our model realistic, we need to take into account that the frequency spectrum is discrete with $f_p = p/T$, and the redshift in Fourier space has the form (102). The Fourier modes after filtering can be found from the real space ones as a projection

$$\tilde{z}^{\text{pos}}(f) = T \{e^{i 2\pi f t}, z^{\text{pos}}(t)\}. \quad (125)$$

Hence, for $f_p = p/T$, using the decomposition (115), with coefficients (114) and the projections (112), we get

$$\tilde{z}^{\text{pos}}(f_p) = \int df \hat{z}(f) \delta^T(f, f_p), \quad (126)$$

where the pseudo Dirac is

$$\begin{aligned} \delta^T(f, f_p) &= T \text{sinc}[\pi(f - f_p)T] \\ &- T \sum_{n=0}^2 (2n+1) j_n(\pi f T) j_n(\pi f_p T). \end{aligned} \quad (127)$$

Since $\delta^T(f, f_p) = \delta^T(-f, -f_p)$ the reality property (60) implies $\tilde{z}^{\text{pos},*}(f_p) = \tilde{z}^{\text{pos}}(-f_p)$.

Restoring the directional dependence in (126), let us compute the two-point function in direction space. From (126) and (62), we find

$$\langle \tilde{z}^{\text{pos}}(\mathbf{n}_1, f_1) \tilde{z}^{\text{pos},*}(\mathbf{n}_2, f_2) \rangle_{\mathcal{A}} = \mu(\beta_{12}) \mathcal{P}^{\text{pos}}(f_1, f_2), \quad (128)$$

where

$$\mathcal{P}^{\text{pos}}(f_1, f_2) \equiv \int \hat{\mathcal{P}}(f) \delta^T(f, f_1) \delta^T(f, f_2) df. \quad (129)$$

This matrix can be defined for general values of f_1, f_2 but it only makes sense for the discrete frequency values and we use the compact notation $\mathcal{P}_{pq}^{\text{pos}} \equiv \mathcal{P}^{\text{pos}}(f_p, f_q)$. If we correlate multipoles, from (68), we now get

$$\langle \tilde{z}_{\ell_1 m_1}^{\text{pos}}(f_p) \tilde{z}_{\ell_2 m_2}^{\text{pos},*}(f_q) \rangle_{\mathcal{A}} = \delta_{\ell_2}^{\ell_1} \delta_{m_2}^{m_1} C_{\ell}^{\text{HD}} \mathcal{P}_{pq}^{\text{pos}}. \quad (130)$$

The diagonal part of this correlation matrix $\mathcal{P}_{pp}^{\text{pos}}$ is plotted in Fig. 4. We observe that there is a huge leakage from the low frequencies to the high frequencies since at high frequencies $\mathcal{P}^{\text{pos}}(f_p, f_p)$ differs substantially from $\mathcal{T}(f) \hat{\mathcal{P}}(f_p)$ that is what one obtains assuming the spectrum of frequencies is continuous (see eq. (62)) and applying a transmission function that accounts for the effect of having filtered out linear drift and constant term. The effect becomes more apparent the more the spectrum is red, i.e. $\gamma \ll -1$. Let us try to develop an intuition of why this is the case, by inspecting the definition of the correlation $\mathcal{P}^{\text{pos}}(f_p, f_p)$ (129). We consider first only the term $\text{sinc}[\pi(f - f_p)T]$ in the definition of the pseudo-Dirac function $\delta^T(f_1, f_2)$ (see eq. (127)). With the approximation $\text{sinc}(\pi(f - f_p)T) \sim fT/p$ valid at small f , and using the parametrization (20) of the spectral function, the integral (129) defining \mathcal{P}^{pos} is dominated by a term of the type $\sim T^2 \int df f^{\gamma+2}/(pq)$. For timing residuals this strongly diverges in $f \rightarrow 0$ since $\gamma = -13/3$. However with the extra terms of (127) coming from the fitting, $\delta^T(f - f_p) \sim (fT)^3/p$ and the integral is regular. However, it is still the low frequency part of the integral (129), that is the range in which $|f|T$ is of order unity, which dominates, even for large values of f_p and f_q . It follows that

$$\mathcal{P}_{pq}^{\text{pos}} \propto \frac{T^2}{pq}. \quad (131)$$

This dominant contribution has very strong off-diagonal correlations.⁴ However, in a realistic data analysis, one will maximize the likelihood with respect to the injected parameter, e.g. as in Fig. 2a of [49], instead of measuring the diagonal values of $\mathcal{P}_{pq}^{\text{pos}}$. Because of the off-diagonal structure, it is expected that the eigenvalues of the correlation matrix are the relevant quantities to estimate the effect of power removal at low frequencies, and we find that they differ substantially from the diagonal value as the eigenmodes mix substantially the various functions f_p . This is illustrated in Fig 5: while for large frequencies the eigenvalues of $\mathcal{P}_{pq}^{\text{pos}}$ tend to follow the values of $\hat{\mathcal{P}}(f_p)$, the first eigenvalues are lower since power has been removed at low frequency via the fitting process.

To summarize, when working with the postfitting residuals $\tilde{z}_{\ell m}^{\text{pos}}(f_p)$ or $\tilde{z}^{\text{pos}}(\mathbf{n}, f_p)$, the correspondence (106b) between Fourier transform (evaluated on the discrete set of frequencies) and Fourier series coefficients is modified to

$$\mathcal{P}^{\text{pos}}(f_p, f_q) = \mathcal{P}^{\text{pos}}(-f_p, -f_q) \leftrightarrow \delta(f_p - f_q) \hat{\mathcal{P}}(f_p). \quad (132)$$

In other words, once postfitting residuals are used, when going from continuous to discrete frequency space the correlation function acquires off-diagonal terms.

VII. COMBINING DIFFERENT ESTIMATORS

In section V, we considered quadratic estimators of pulsar redshifts for a given pair of times or frequencies. Exploiting statistical isotropy, directions were combined with a rotational average (pulsar average), which in multipole space corresponds to a summation over m . We now need to combine the various times or frequencies to construct estimators that account for the entire dataset, making use of the underlying stationarity of the GW background. There are two key features of the signal we wish to measure.

- First, we aim to verify that the multipole correlations follow the shape of the C_{ℓ}^{HD} given in equation (66), which involves measuring these multipoles individually [1]. We thus use the parametrization

$$C_{\ell}^{\text{HD}} \rightarrow c_{\ell} C_{\ell}^{\text{HD}} \quad (133)$$

and wish to constrain the c_{ℓ} . In the CMB context, this corresponds to revealing the Sachs-Wolfe plateau [50] dependence $C_{\ell}^{\text{CMB}} \propto 1/[\ell(\ell+1)]$ for a scale-invariant primordial scalar spectrum (see e.g. Fig. 2.4 of [51]).

- Second, we seek to constrain the global amplitude $\hat{\mathcal{P}}_*$ and, possibly, the spectral index γ , by gathering all available information. In the CMB context it is equivalent of constraining the global amplitude A_s^2 of scalar

⁴ Note that we assumed that the biggest contribution to the integral comes from $|f|T$ of order unity, but part of the integral comes from the usual Dirac type contribution of $\text{sinc}[\pi(f - f_p)T]$ and $\text{sinc}[\pi(f - f_q)T]$ which is around $f = f_p$ and $f = f_q$. However, the redder the spectrum is, the less important these contributions are compared to the low f part.

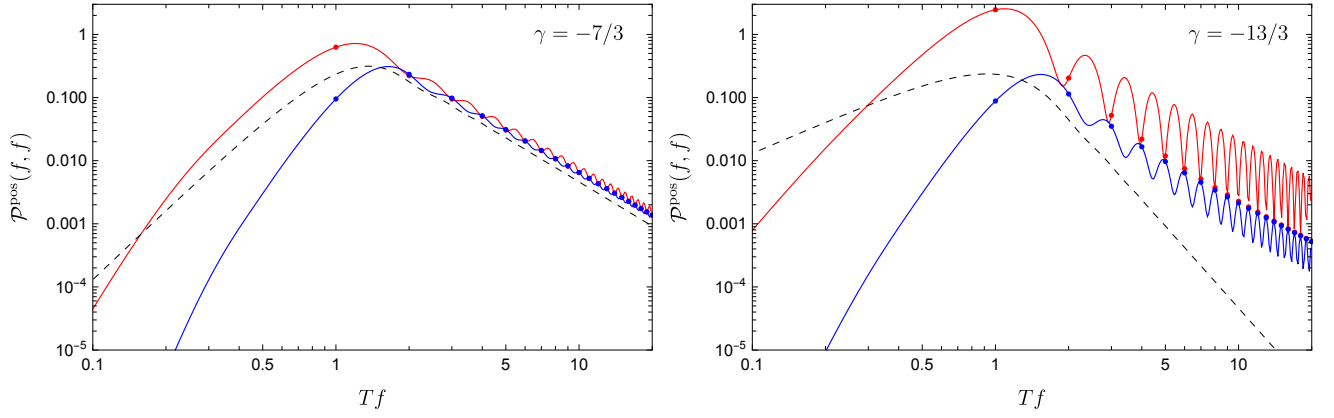


FIG. 4. Diagonal part of the correlation matrix defined in (130) and (129). In the left panel we consider a spectral index $\gamma = -7/3$ (corresponding to redshift measurements) whereas the right panel uses $\gamma = -13/3$ (time residual measurement). In red we subtract from redshift/time residuals only P_0 and P_1 whereas in blue we fit P_0, P_1, P_2 . The continuous line is $\mathcal{P}^{\text{pos}}(f, f)$ and the dots are the same function evaluated for finite frequency values, i.e. $f = f_p$ with $p = 1, 2, \dots$ which are the only values of frequency accessible in a realistic observation. The dashed line is the simple estimate via $\mathcal{T}(f)\hat{\mathcal{P}}(f)$ which is the initial spectrum multiplied by the transmission function (120). We see that for timing residuals (right panel) a lot of power from the low f spills to larger frequencies, i.e. $\mathcal{P}^{\text{pos}}(f_p, f_p)$ differs substantially from $\mathcal{T}(f)\hat{\mathcal{P}}(f_p)$ for high frequencies. The spectrum of redshift and timing residuals is so red that even a tiny spillover of low frequency over large ones completely modifies the resulting spectrum.

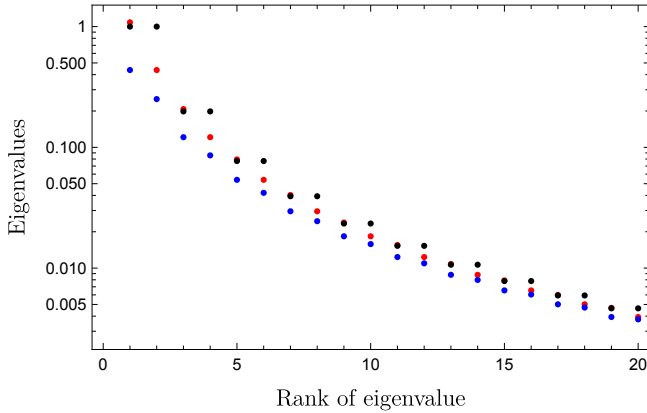


FIG. 5. Eigenvalues of $T^{-2}\mathcal{P}^{\text{pos}}(f_p, f_q)$ in red (when only P_0 and P_1 are fitted) and in blue (when P_2 is also fitted). The black dots corresponds to the values $T^{-1}\hat{\mathcal{P}}(1/T) = T^{-1}\hat{\mathcal{P}}(-1/T)$, $T^{-1}\hat{\mathcal{P}}(2/T) = T^{-1}\hat{\mathcal{P}}(-2/T), \dots$

fluctuations and the primordial spectrum spectral index n_s [50, 52].

PTA work in the regime where the average over the observation time does not allow to sample many realisations of the Universe, contrary to the case of Earth-based and space-based instruments. For a given observation time T , the frequency resolution of a given instrument is $\Delta f \simeq 1/T$. Assuming T of the order of three year, one has $\Delta f \simeq 10^{-8}$ Hz, which is about two orders of magnitude below the maximum frequency set by the cadence of pulsar observations. Note that for Earth- and space-based detectors the frequency resolution is similar to the PTA one, being set by the observation time, however the frequency we have access to is order of magnitude higher

(mHz for space-based and Hz for ground-based instruments). Recently, some studies have claimed that cosmic variance imposes an irreducible limitation on our ability to reconstruct the Hellings-Downs correlation [7, 22, 34]. In this section, we demonstrate in multipole space that this limitation arises only when different time measurements are combined in a naive manner (see [39] for an analogue result derived in configuration space). However, in section VII C we detail that when appropriate weighting is applied, the impact of the so-called cosmic variance is mitigated.

A. Nonweighted time average

Let us introduce the time-averaged version of the equal-time estimator (89) (observationally taking such an average helps in reducing instrumental noise disturbance)

$$C_\ell^{Te} = \frac{1}{T} \int_{-T/2}^{T/2} d\eta C_\ell^e(\eta). \quad (134)$$

From (92) and using $\mathcal{P}(0) = h^2$ where h^2 is defined in (19), its distributional average is given by

$$\langle C_\ell^{Te} \rangle_{\mathcal{A}} = h^2 C_\ell^{\text{HD}}. \quad (135)$$

From (94), the variance of the estimator reads

$$\begin{aligned} & \langle C_\ell^{Te} C_{\ell'}^{Te} \rangle_{\mathcal{A}} - \langle C_\ell^{Te} \rangle_{\mathcal{A}} \langle C_{\ell'}^{Te} \rangle_{\mathcal{A}} \\ &= \delta_{\ell\ell'} \frac{2}{2\ell+1} \int_{-T/2}^{T/2} \frac{d\eta}{T} \int_{-T/2}^{T/2} \frac{d\eta'}{T} [C_\ell(\eta, \eta')]^2. \end{aligned} \quad (136)$$

To simplify our notation, and following [22], we introduce

$$\begin{aligned} h^4 &\equiv \int_{-T/2}^{T/2} \frac{d\eta}{T} \int_{-T/2}^{T/2} \frac{d\eta'}{T} \varphi^2(\eta - \eta') \\ &= 4 \int \frac{4\pi k_1^2 dk_1}{(2\pi)^3} \frac{4\pi k_2^2 dk_2}{(2\pi)^3} \mathcal{A}(k_1) \mathcal{A}(k_2) f_T(k_1, k_2), \end{aligned} \quad (137)$$

where

$$\begin{aligned} f_T(k_1, k_2) &\equiv \int_{-T/2}^{T/2} \frac{d\eta d\eta'}{T^2} \cos[k_1(\eta - \eta')] \cos[k_2(\eta - \eta')] \\ &= \frac{1}{2} [\text{sinc}^2((k_1 + k_2)T/2) + \text{sinc}^2((k_1 - k_2)T/2)], \end{aligned}$$

as in (7.14) of [22]. Then the variance can be written as

$$\langle C_\ell^{Te} C_{\ell'}^{Te} \rangle_{\mathcal{A}} - \langle C_\ell^{Te} \rangle_{\mathcal{A}} \langle C_{\ell'}^{Te} \rangle_{\mathcal{A}} = \frac{2\delta_{\ell\ell'}}{2\ell+1} \frac{h^4}{(h^2)^2} (\langle C_\ell^{Te} \rangle_{\mathcal{A}})^2. \quad (138)$$

The relative size variance to mean is set by the ratio $h^4/(h^2)^2$. Indeed for an ideal experiment, one gets for a given ℓ

$$\text{SNR}_\ell^2 = \frac{2\ell+1}{2} N_f^{\text{eff}}, \quad N_f^{\text{eff}} \equiv \left(\frac{h^4}{h^2} \right). \quad (139)$$

Approximating $\text{sinc}(\pi x) \simeq \text{sinc}^2(\pi x) \simeq \delta(x)$ we have roughly

$$h^4 \simeq \frac{1}{T} \int_{\text{obs}} [\hat{\mathcal{P}}(f)]^2 df, \quad h^2 = \int_{\text{obs}} \hat{\mathcal{P}}(f) df. \quad (140)$$

If the frequency integrals in (140) are not sensitive to the low frequency boundary $1/T$, that is if the spectrum is not red, then it is clear $\sqrt{N_f^{\text{eff}}} = h^2/h^2 \propto \sqrt{T}$ and the SNR improves with time. However for a red spectrum (expected in PTA experiment, see eq. (20)) with $\gamma < -1$ there is always more power at low frequency, hence $h^4 \propto T^{-2-2\gamma}$ and $h^2 \propto T^{-1-\gamma}$, such that

$$N_f^{\text{eff}} \simeq -\frac{(2+4\gamma)}{(1+\gamma)^2}, \quad (141)$$

implying that the \sqrt{T} scaling is lost. For the time residual spectrum with $\gamma = -13/3$ as expected from an astrophysical population, N_f^{eff} is of order (slightly larger than) unity and one gets $\text{SNR}_\ell^2 \simeq (2\ell+1)/2$. This fact is at the origin of the cosmic variance problem discussed in [7, 22, 34].

In order to understand the issue more clearly, let us insert the Fourier transform (59) in the time average estimator (134), to cast it as

$$C_\ell^{Te} = \int df \int df' \text{sinc}[\pi(f-f')T] C_\ell^e(f, f'). \quad (142)$$

Using approximation (104), this is approximately

$$C_\ell^{Te} \simeq \frac{1}{T} \int C_\ell^e(f) df = \int \sum_m \frac{\hat{z}_{\ell m}(f) \hat{z}_{\ell m}^*(f)}{(2\ell+1)T} df, \quad (143)$$

where it is understood that the integral over frequencies is \int_{obs} defined in (106a). The integral (143) is not an optimal way to

combine the various frequencies because all frequencies are weighted in the same way. As a consequence, the SNR, which is given by (139), is not optimal either. Combining frequencies as in (143) effectively amounts to consider as contributing to the signal only the lowest frequency $1/T$ and a few larger ones, and as explained in appendix B this is the worst choice if one aims to have a high SNR estimator. The effective number of frequencies used with this estimator is precisely N_f^{eff} .

If we instead aim to constrain the global amplitude $\hat{\mathcal{P}}_*$, in the infinite pulsar number we will combine the information from all ℓ , and even though the frequencies are not optimally combined, this brings and infinite number of angular modes as PTA are sensitive to all ℓ in perfect experiments, hence the cosmic variance in that case vanishes (see e.g. section VII.A of [31]).

We conclude this section with a remark. As detailed in section VI, the integral on frequencies in (143) is in fact a sum on the discrete functions $f_p = p/T$, and the $\hat{z}_{\ell m}(f)$ are to be replaced by the $\hat{z}_{\ell m}^{\text{pos}}(f_p)$ obtained as in (115) from $\hat{z}_{\ell m}(f)$ by fitting low order polynomials and extracting the Fourier series components. Since the power in low frequencies is reduced by this procedure, one might wonder whether this alleviates the variance problem, giving rise to a higher SNR compared to the case in which the SNR size is set by N_f^{eff} given by (141). The estimator (134) for a discrete frequency spectrum and after having filtered low-order polynomial, becomes

$$\begin{aligned} C_\ell^{Te} &\equiv \frac{1}{T} \int_{-T/2}^{T/2} \sum_{m=-\ell}^{\ell} \frac{\hat{z}_{\ell m}^{\text{pos}}(\eta) \hat{z}_{\ell m}^{\text{pos}*}(\eta)}{2\ell+1} d\eta \\ &= \frac{1}{T^2} \sum_{|p| \geq 1} \sum_{m=-\ell}^{\ell} \frac{\hat{z}_{\ell m}^{\text{pos}}(f_p) \hat{z}_{\ell m}^{\text{pos}*}(f_p)}{2\ell+1}, \end{aligned} \quad (144)$$

where the Parseval relation has been used to obtain the second line. Repeating the computation of the signal-to-noise ratio and using (130) we find that we must replace in (139)

$$N_f^{\text{eff}} \rightarrow \frac{(\sum_{|p| \geq 1} \mathcal{P}_{pp}^{\text{pos}})^2}{\sum_{|p|, |q| \geq 1} |\mathcal{P}_{pq}^{\text{pos}}|^2}. \quad (145)$$

A comparison between these two quantities is presented in Fig. 6, which shows that the correct estimation with Fourier series coefficients (145) leads to a slightly better SNR ($N_f^{\text{eff}} \simeq 1.9$ for $\gamma = -13/3$), but overall the approximation (141) is not far off. For very negative γ we checked that it converges toward the minimum value which is unity.

After reviewing briefly in the next section the various noise contributions that affect the measurements of the angular power spectrum of the pulsar redshift two-point function, we explain in section VII C how cosmic variance can be mitigated by properly weighting the different frequencies so that the effective number of frequencies reaches the maximum.

B. Noise contributions

The first noise contribution arises from the fact that, in our theoretical derivation of the correlation function, we assumed

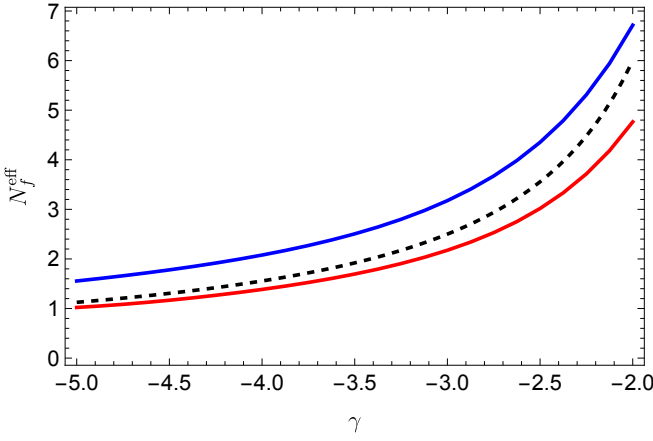


FIG. 6. The effective number of frequency modes N_f^{eff} as a function of the spectral index γ . The dashed line is the approximation (141). The colored lines are the estimation from the Fourier series coefficients correlations (145). The red line corresponds to the residual after fitting only P_0, P_1 whereas the blue line corresponds to also fitting P_2 from the raw signal.

an infinite number of pulsar pairs. However, in a realistic experiment, the number of pairs is finite, which introduces a Poisson noise, or shot noise, contribution to the angular power spectrum. We emphasize that this shot noise is due to the discrete distribution of pulsars, not the discreteness of the emitting sources. This effect has been discussed in [22, 34], where the authors examine the redshift correlation function in real space.⁵ The second noise contribution comes from the noise of individual pulsars, that add up and give a flat in ℓ -space (shot-noise-like) offset to the angular power spectrum of the pulsar redshift correlation function, usually modeled as a sum of white and red noises [53].

We find it more convenient to work in multipole space, where the structure of the estimator is diagonal, as discussed in Section V.

1. Noise due to Poissonian pulsar statistics

Let us estimate the effect of extracting the $z_{\ell m}$ coefficients from a finite number of pixels, which may not necessarily be optimally placed. If we have a finite number N of pulsars we get the naive estimation of the multipoles (either functions of time or frequency) as

$$z_{\ell m}^N = \frac{4\pi}{N} \sum_{i=1}^N z(\mathbf{n}_i) Y_{\ell}^{m*}(\mathbf{n}_i). \quad (146)$$

⁵ Although the authors in these references do not explicitly mention pulsar shot noise, they introduce the concept of the total variance of the HD correlation, which accounts for the fact that a realistic experiment involves a finite number of pulsars.

If we perform a double average on realization of the GW background, and on pulsar positions, we get

$$\langle z_{\ell m}^N z_{\ell' m'}^{N*} \rangle_{\mathcal{A}, \mathbf{P}} = \delta_{\ell}^{\ell'} \delta_m^{m'} (C_{\ell}^{\text{GW}} + N_{\ell}^{\text{P}}), \quad (147)$$

where we have split the total angular power spectrum on the right hand side into a contribution that is always present and a contribution going to zero in the limit $N \rightarrow \infty$, that we identify with a shot noise contribution. It is given by

$$N_{\ell}^{\text{P}} = \frac{1}{N} \left(-C_{\ell}^{\text{GW}} + \sum_L (2L+1) C_L^{\text{GW}} \right). \quad (148)$$

The computation is quite technical and can be found in appendix D. Note that the second term in the parenthesis is $4\pi\mu^{\text{tot}}(\beta_{12}=0)$, that is it is directly proportional to the total variance at zero separation and this contributes equally for all multipoles N_{ℓ} . We observe that this noise contribution (148) scales as $1/N$ with nearly no ℓ dependence as the second term is the dominant one: this is precisely the reason why we decided to call it shot noise, even if it is not ℓ -independent. To summarize, even in a perfect experiment we will have shot noise from the finite number of pulsars used to evaluate the multipoles.

We observe that the shot noise of the estimator (146) can be reduced by introducing proper weights in the sum over pulsars (using a quadrature scheme). Concretely, the estimator can be replaced by

$$z_{\ell m}^N = 4\pi \sum_{i=1}^N w_i z(\mathbf{n}_i) Y_{\ell}^{m*}(\mathbf{n}_i), \quad (149)$$

where the w_i are appropriate weights (the naive estimate (146) corresponds to $w_i = 1/N$). A simple choice consists in extending the trapezoidal Riemann sum to a sphere. For a set of points randomly placed this is equivalent to weighting each point by the fraction of the area encompassed. The fractional area w_i around each pulsar direction can be obtained by generating many random directions and evaluating the fraction which are the closest to each individual pulsar. Another choice consists in finding the best quadrature coefficients, see e.g. [54, 55], such that (146) is exact when $z(\mathbf{n})$ can be expressed as a finite sum of spherical harmonics.

The pulsar shot noise with the naive method (146) and with the quadrature scheme (149) (using area weights) is represented in Fig. 7 and we check that with the quadrature the pulsar shot noise decreases as $1/N^2$.⁶

2. Noise of individual pulsars

Finally, we need to account for the noise in individual pulsars, either coming from intrinsic irregularity or experimental

⁶ This does not mean that we can beat the optimal estimator (165) whose SNR^2 scales as $1/N$, because when estimating the total variance we must add the contributions of all multipoles (all ℓ values) and the prefactor in front of the $1/N^2$ scaling increases with ℓ , as shown in Fig. 7.

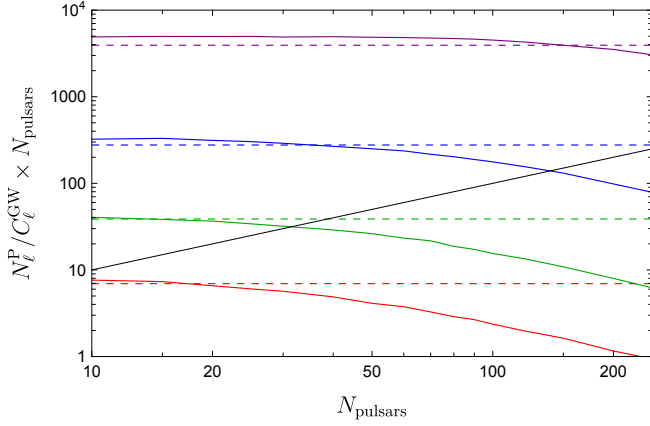


FIG. 7. Ratio of the shot noise contribution and the correlation function (rescaled by the pulsar number), as a function of the pulsar number. The dashed lines are the pulsar shot noise (148) of the naive estimate (146) whereas the continuous lines are obtained with the quadrature weighting (149) where the weights are the area of each pulsar direction. The curves are from bottom to top for $\ell = 2, 3, 5, 10$ respectively. The thin black line corresponds to $C_\ell^{\text{GW}} = N_\ell^{\text{P}}$ where the pulsar noise is as large as the HD correlation, hence points located above this line are dominated by pulsar shot noise. We check that $N_\ell^{\text{P}} \propto 1/N$ with the naive estimate whereas it scales as $1/N^2$ when weighting by the area. However we see that the prefactor of this latter scaling increases with ℓ , i.e. the weighting suppression is more efficient at small multipoles.

errors. Each pulsar signal is supplemented with an independent noise n_i , and if stationary, the statistics in Fourier space is of the form

$$\langle \hat{n}_i(f) \hat{n}_j^*(f') \rangle_n = (4\pi)^{-1} \delta_{ij} \delta(f - f') n_i(f). \quad (150)$$

where the average on the left-hand side has to be understood as an average of different noise realisations. Different noise contributions will add up to give a shot noise contribution to the angular power spectrum, flat in multipole space. Indeed, replacing $\hat{z}(\mathbf{n}_i, f) \rightarrow \hat{z}(\mathbf{n}_i, f) + \hat{n}_i(f)$ in (146), and repeating the analysis which led to (147), the average over the GW ensemble, the pulsar directions and the pulsar noises becomes (accounting for weights as in eq. (149))

$$\langle \hat{z}_{\ell m}^N(f) \hat{z}_{\ell' m'}^{N*}(f') \rangle_{\mathcal{A}, \mathbf{P}, n} = \delta_\ell^{\ell'} \delta_m^{m'} \delta(f - f') [C_\ell^{\text{GW}}(f) + N_\ell^{\text{P}}(f) + n(f)], \quad (151)$$

where

$$n(f) \equiv \sum_{i=1}^N w_i^2 n_i(f), \quad (152)$$

and where from (68) $C_\ell^{\text{GW}}(f) = \hat{\mathcal{P}}(f) C_\ell^{\text{HD}}$. It is understood that the pulsar shot noise now takes the same form as in (148) if the naive estimate (146) is used, but with $C_\ell(f) \rightarrow C_\ell(f) + n(f)$. For a family of pulsars whose noises are of the same order of magnitude, increasing their number reduces the noise since the weights are of order $1/N$ hence $n(f) \propto 1/N$.

In the case where the noise (152) is larger than the pulsar shot noise, we could decide to use weights that minimize this

contribution instead of weights that minimize the pulsar shot noise. Under the constraint $\sum_i w_i = 1$, this leads to the choice

$$w_i = \frac{n_i^{-1}}{\sum_{j=1}^N n_j^{-1}}, \quad (153)$$

such that

$$n^{-1}(f) = \sum_{i=1}^N n_i^{-1}(f). \quad (154)$$

However, with this choice while $n(f)$ is minimized, the second term of the pulsar shot noise (148) is worsened by a multiplicative factor

$$\frac{N (\sum_{i=1}^N n_i^{-2})}{(\sum_{i=1}^N n_i^{-1})^2}. \quad (155)$$

This factor is always larger than unity unless all the noises are equal.⁷ In that case, this corresponds to directly using the naive estimate (146). Ideally, a good set of weights would take into account both the area of each pixel and its intrinsic noise to minimize the total noise.

C. Optimal weighting

In this section, we explain how cosmic variance can be mitigated by properly weighting the angular power spectrum of pulsars at different frequencies, to build an optimal estimator. We work in multipole space as this makes derivations simpler. We first consider the case of a perfect experiment, for which an optimal quadratic estimator for the correlation amplitude can be built using the procedure detailed in appendix B. When adding instrumental noise, optimal estimators are formally more involved and use the Fisher information matrix. We will show that in the limit of strong signal, the SNR scales as the product of the observation time and cadence of observation, hence there is no fundamental limitation that prevents one from improving it in future observations.

1. Perfect experiment

Let us first discuss how an optimal estimator can be built in a perfect experiment, that is one in which there is no noise (no intrinsic pulsar noise nor pulsar shot noise). The $\hat{z}_{\ell m}^N(f) = \hat{z}_{\ell m}(f)$ can be viewed as independent measurements where it is understood that we must consider the discrete set of frequencies (one should use the $\hat{z}_{\ell m}^{\text{pos}}(f_p)$). Let us use a generalized index $a = (\ell, m, f)$ to make the notation compact. If we know the shape of the correlation of these measurements,

⁷ This follows from Cauchy-Schwarz inequality applied to the N dimensional vectors \mathbf{A} and \mathbf{B} with components $A_i = 1/n_i$ and $B_i = 1$. It saturates when \mathbf{A} is aligned with \mathbf{B} hence when all noises are equal.

and want to measure its global amplitude, as detailed in appendix B, we need to form an optimal estimator as

$$Q \propto \sum_{ab} z_b^* C_{ba}^{-1} z_a. \quad (156)$$

This estimator, applicable to perfect experiments, is formally simpler than the estimator for imperfect experiments reviewed in appendix C. As a result, it offers a simpler explanation of how the optimal combination of frequencies consistently mitigates cosmic noise.

When detecting the HD correlation, we are first interested in measuring the correlation for each ℓ separately [24], that is we aim at constraining the c_ℓ defined in (133), hence we fix ℓ and perform the previous sums on $a, b = (m, f)$ only. Using that the correlation matrix (151) is diagonal (and ignoring the noise), the optimal estimator (with the continuous frequency approximation) is

$$Q_\ell = \sum_{m=-\ell}^{\ell} \int_{\text{obs}} \frac{df}{N_f} \frac{\hat{z}_{\ell m}(f) \hat{z}_{\ell m}^*(f)}{(2\ell+1) \hat{\mathcal{P}}(f) C_\ell^{\text{HD}}} = \int_{\text{obs}} \frac{df}{N_f} \frac{C_\ell^e(f)}{\hat{\mathcal{P}}(f) C_\ell^{\text{HD}}} \quad (157)$$

where $\bar{\mathcal{P}}(f) \equiv \hat{\mathcal{P}}(f)/\hat{\mathcal{P}}_\star$ is the normalized spectral function and $N_f = 2Tf_{\text{max}}$ the number of frequency modes. The average and variance of this optimal estimator are

$$\langle Q_\ell \rangle_{\mathcal{A}} = \hat{\mathcal{P}}_\star, \quad (158a)$$

$$\langle (Q_\ell)^2 \rangle_{\mathcal{A}} - (\langle Q_\ell \rangle_{\mathcal{A}})^2 = \frac{2}{2\ell+1} (2Tf_{\text{max}})^{-1} \hat{\mathcal{P}}_\star^2, \quad (158b)$$

hence the estimator can be used to measure the c_ℓ . The associated signal-to-noise ratio (SNR) is

$$\text{SNR}_\ell^2 = \frac{2\ell+1}{2} (2Tf_{\text{max}}). \quad (159)$$

In practice, when considering the Fourier series coefficients $\hat{z}_{\ell m}^{\text{pos}}(f_p)$ instead of the Fourier transform $\hat{z}_{\ell m}(f)$, the correlation is not diagonal anymore, as detailed in section VI B. But the argument is the same and one needs only to retain the correct correlation matrix in (156) to combine the discrete set of measured frequencies, that is we must replace

$$\int_{\text{obs}} \frac{df}{N_f} \frac{\hat{z}_{\ell m}(f) \hat{z}_{\ell m}^*(f)}{\hat{\mathcal{P}}(f)} \rightarrow \frac{1}{N_f} \sum_{p,q} \hat{z}_{\ell m}^{\text{pos}}(f_p) \hat{z}_{\ell m}^{\text{pos}*}(f_q) \bar{S}_{pq}^{\text{pos},-1}, \quad (160)$$

where the sums on p, q run on the values $1 \leq |p| \leq Tf_{\text{max}}$ and $1 \leq |q| \leq Tf_{\text{max}}$ and where the normalized covariance matrix $\bar{S}_{pq}^{\text{pos}}$ is defined from (129) with $\hat{\mathcal{P}}(f)$ instead of $\hat{\mathcal{P}}(f)$. In all cases for a perfect experiment we have $2\ell+1$ modes at a given ℓ from the isotropy assumption (all the possible values of m), and we have $2Tf_{\text{max}}$ frequency modes. The product of these quantities gives the number of modes available and the (square of) SNR is 1/2 for each of them.

Now if we aim at constraining the global amplitude $\hat{\mathcal{P}}_\star$, that is the second goal mentioned at the beginning of section VII, several ℓ can also be optimally combined as

$$Q_{\text{opt}} \equiv \sum_{\ell=2}^{\ell_{\text{max}}} (2\ell+1) Q_\ell, \quad (161)$$

where $\Sigma_{\text{max}} \equiv \sum_{\ell=2}^{\ell_{\text{max}}} (2\ell+1) = (\ell_{\text{max}}+3)(\ell_{\text{max}}-1)$ is the number of multipoles available until ℓ_{max} . This estimator is also obtained directly from (156) with $a, b = (\ell, m, f)$. The corresponding SNR with which $\hat{\mathcal{P}}_\star$ can be measured is

$$\text{SNR}^2 = \frac{\Sigma_{\text{max}}}{2} (2Tf_{\text{max}}). \quad (162)$$

Since we can only combine N_P pulsars, it is expected that $\Sigma_{\text{max}} \simeq N_P$.

Let us show that this is indeed the case by building the optimal estimator in real space directly from the pulsar pairs so that we do not have to account for the pulsar shot noise inherently contained in the multipole estimators (146). With the results of appendix B, that is using (156) with $a, b = (\mathbf{n}, f)$ and the covariance (62), an optimal estimator for $\hat{\mathcal{P}}_\star$ is

$$Q_{\text{opt}}^{\text{real}} = \frac{1}{N_P} \sum_{i=1}^{N_P} \sum_{j=1}^{N_P} \int_{\text{obs}} \frac{df}{N_f} \frac{\hat{z}(\mathbf{n}_i, f) \hat{z}^*(\mathbf{n}_j, f)}{\hat{\mathcal{P}}(f)} [\mu(\beta_{ji})]^{-1}. \quad (163)$$

It is clear using (62) that $\langle Q_{\text{opt}}^{\text{real}} \rangle_{\mathcal{A}} = \hat{\mathcal{P}}_\star$ and $\langle (Q_{\text{opt}}^{\text{real}})^2 \rangle_{\mathcal{A}} - (\langle Q_{\text{opt}}^{\text{real}} \rangle_{\mathcal{A}})^2 = (2Tf_{\text{max}})^{-1} \hat{\mathcal{P}}_\star^2 / N_P$. Again, in a rigorous treatment, one should not use the continuous frequency approximation, and one should consider a modification of the type (160) for pulsar directions, that is

$$Q_{\text{opt}}^{\text{real}} = \frac{1}{N_P N_f} \sum_{i,j=1}^{N_P} \sum_{p,q} \hat{z}^{\text{pos}}(\mathbf{n}_i, f_p) \hat{z}^{\text{pos}*}(\mathbf{n}_j, f_q) \bar{S}_{qp}^{\text{pos},-1} [\mu(\beta_{ji})]^{-1}, \quad (164)$$

but this does not alter these results. Hence, when measuring $\hat{\mathcal{P}}_\star$ with this optimal estimator

$$\text{SNR}^2 = \frac{N_P}{2} (2Tf_{\text{max}}). \quad (165)$$

This result is to be compared with (7.12) of [31] in which $N_f = 2Tf_{\text{max}}$ is replaced by $N_f^{\text{eff}} = h^4/\hbar^4 \ll 2Tf_{\text{max}}$. Indeed in this reference while the angular information is optimally combined, frequencies are not. It corresponds to the perfect experiment version of (30) in [39].

2. Imperfect experiment

When considering an imperfect experiment with noise, a similar procedure has to be followed to weight different frequencies in an optimal manner. If the noise is exactly known, then a quadratic estimator which takes noise into account can be built [39]. Since the noise influences the signal, the signal in turn affects the estimation of the noise, meaning that the noise can never be fully separated from the signal. Therefore, to accurately measure the correlations induced by the gravitational wave (GW) background, the estimation should involve maximizing the likelihood over the GW parameters, the template-fitting parameters, and the noise parameters. The final step is to analyze the marginalized distribution. With the compact index notation, the Gaussian likelihood which is maximized is of the form

$$\mathcal{L} = G(0, C_{ab}; z_a), \quad (166)$$

where $G(M_a, C_{ab}; z_a)$ is a multidimensional Gaussian of average M_a and total covariance C_{ab} evaluated in the set of values z_a .

The set of variables which is directly related to what is observed is the set of $\hat{z}(\mathbf{n}_i, f)$, or more precisely the $\tilde{z}^{\text{pos}}(\mathbf{n}_i, f_p)$ (after filtering). As detailed in appendix C, if we can estimate the noise statistics independently, it becomes possible to build quadratic estimators for imperfect experiments, either to measure the c_ℓ of (133) or to constrain the global amplitude $\hat{\rho}_\star$. However estimating the SNR is very difficult analytically as it requires the inversion of the covariance matrix, and this is not feasible analytically for noisy experiments with finite number of pulsars.

Due to the impossibility of assessing analytically the SNR in imperfect experiments with the optimal quadratic estimators built in appendix C, let us work directly on the multipole estimators (146) and let us assume that they follow a Gaussian distribution of the type (166), whose covariance (151) is due to three sources of randomness: the GW signal, the intrinsic noises, and the pulsar directions. Instead of considering the fundamental likelihood of the $\tilde{z}^{\text{pos}}(\mathbf{n}_i, f_p)$ observables, we thus consider a likelihood in the intermediate quantities (146) built on sums of the former observables. In addition, even though the directions of the pulsars are well measured, we will estimate the pulsar noise induced by the finite number of pulsars, that is the pulsar shot noise, by averaging over the realizations of the pulsar directions. This means that we replace the pulsar noise of a given realization by the average pulsar noise on all possible direction realizations. Hence an estimator which maximizes this hybrid likelihood will not be optimal. Nonetheless its associated SNR will shed light on the SNR of the optimal estimator. Indeed, under these three averages (GW signal, intrinsic noises and pulsar directions) the covariance is diagonal in the $a = (\ell, m, f)$ indices, hence its inversion, which is required to estimate the SNR, is straightforward. In fact, as we work with the Fourier series coefficients after fitting, one should use (130) which is not diagonal in frequency space. For simplicity and to make our derivation more transparent, we use the continuous frequency approximation hereafter. The Fisher matrix provides the SNR with which different parameters can be measured. Focusing on the space of the parameters c_ℓ , the corresponding Fisher matrix is diagonal and given by

$$F_{\ell\ell} \equiv \frac{1}{2} \text{Tr} [C^{-1} \cdot \partial_{c_\ell} C \cdot C^{-1} \cdot \partial_{c_\ell} C], \quad (167)$$

where the matrix products and trace are taken in the space $a = (m, f)$ at fixed ℓ . We find (using the continuous frequency approximation for simplicity)

$$F_{\ell\ell} = \frac{2\ell+1}{2} T \int_{\text{obs}} \left[\frac{C_\ell^{\text{GW}}(f)}{c_\ell C_\ell^{\text{GW}}(f) + N_\ell^{\text{P}}(f) + \mathcal{N}(f)} \right]^2 df. \quad (168)$$

Then the SNR for the c_ℓ is

$$\begin{aligned} \text{SNR}_\ell^2 &= c_\ell^2 / F_{\ell\ell}^{-1} |_{c_\ell=1} = c_\ell^2 F_{\ell\ell} |_{c_\ell=1} \\ &= \frac{2\ell+1}{2} N_f^{\text{eff}}(\ell), \end{aligned} \quad (169)$$

where the ℓ -dependent effective number of frequencies⁸ is defined by

$$N_f^{\text{eff}}(\ell) \equiv \int_{\text{obs}} T \left[\frac{C_\ell^{\text{GW}}(f)}{C_\ell^{\text{GW}}(f) + N_\ell^{\text{P}}(f) + \mathcal{N}(f)} \right]^2 df. \quad (170)$$

We deduce immediately that when increasing observation time or the number of pulsars (the latter improvement decreasing both the pulsar shot noise and the total intrinsic noise) the SNR increases, as already predicted with mock PTA data (see e.g. Fig. 4 in [19]). When translated into discrete frequencies with (106a), and retaining only the dominant frequency mode $1/T$ (and its mirror negative frequency $-1/T$), this reduces to the estimation (26) of [16]. When ignoring pulsar shot noise $N_f^{\text{eff}}(\ell) \simeq 2T f_{\text{max}}^{\text{eff}}(\ell)$ where the effective maximum frequency is the frequency for which $C_\ell^{\text{GW}}(f_{\text{max}}^{\text{eff}}(\ell)) = \mathcal{N}(f_{\text{max}}^{\text{eff}}(\ell))$ since for larger frequencies the integrand becomes negligible. If the noise has a spectral index $\gamma_n > \gamma$ defined by $\mathcal{N}(f) \propto (f/f_\star)^{\gamma_n}$ then from the shape (66) we deduce the scaling $f_{\text{max}}^{\text{eff}} \propto \ell^{4/(\gamma-\gamma_n)}$. Therefore the effective number of frequencies depends on ℓ contrary to what is assumed in the equivalence (7) of [39]. This explains why, even when using many pulsars such that the pulsar shot noise becomes subdominant, the larger multipoles remain the most challenging to constrain [24].

We now seek to get a rough estimate of $N_f^{\text{eff}}(\ell)$ with the pulsar noise characteristics of [56]. For any pulsar in a set, the covariance matrix of timing residuals is written as

$$\langle r(\eta_a) r(\eta_b) \rangle = C_{ab}^{\text{GW}} + C_{ab}^{\mathcal{N}}, \quad (171)$$

where we have distinguished a signal and a noise component, with noise covariance

$$C_{ab}^{\mathcal{N}} = \int_{-f_{\text{max}}}^{f_{\text{max}}} df \cos(2\pi f(\eta_a - \eta_b)) \frac{\mathcal{N}(f)}{4\pi}. \quad (172)$$

$\mathcal{N}(f)$ is defined in (150) for a given pulsar and $P_{\mathcal{N}}(f) \equiv \mathcal{N}(f)/(4\pi)$ is the two-sided noise spectral density (for timing residuals it has dimensions of inverse cubic frequency). This quantity is usually parametrized for each pulsar separating a contribution of white noise and a red noise one

$$P_{\mathcal{N}}(f) = \frac{\mathcal{N}(f)}{4\pi} = \sigma^2 \Delta\eta + \frac{1}{2} A_{\text{RN}}^2 \left(\frac{f}{1\text{yr}^{-1}} \right)^{\gamma_{\text{RN}}}, \quad (173)$$

where σ is the standard deviation due to white noise and $\Delta\eta$ is the cadence of observation. We consider the 68 pulsars in tables 5 and 6 of [56], and take into account red noise only for the 23 pulsars for which it has been measured. The cadence of observation is estimated by dividing the time of observation by the number of epochs for each pulsar, which is

⁸ Note that in section VII A, we used the term "effective number of frequencies" N_f^{eff} when an optimal estimator is not used. Here *effective* rather means that the computation accounts for instrumental noise. We always have $N_f^{\text{eff}}(\ell) \leq N_f$ with the total number of available frequencies $N_f = 2T f_{\text{max}}$, and the equality is only reached for perfect experiments with an infinite number of pulsars.

a crude estimation of the cadence since there are a few gaps in the observations. The noises, divided by the transmission factor (120) computed with their respective observing time, are combined with (154), that is using weights which are optimized to reduce the total noise. As a consequence, when estimating the shot noise due to the finite number of pulsars which we also take into account, we must replace $1/N$ by $\sum_{i=1}^N w_i^2$ in (148) and this factor is approximately 0.12 for $f = 1 \text{ yr}^{-1}$, hence larger than $1/N = 1/68$. Finally, the signal spectrum amplitude at $f_* = 1 \text{ yr}^{-1}$ is set to $\mathcal{P}_* = \frac{1}{4}(h_c^{1\text{yr}})^2 \text{ yr}$ with $h_c^{1\text{yr}} = 3 \times 10^{-15}$, taking into account (75) for time residuals.

Eventually, we find that $N_f^{\text{eff}}(\ell = 2) \simeq 4.2$, $N_f^{\text{eff}}(\ell = 3) \simeq 0.43$ and $N_f^{\text{eff}}(\ell = 4) \simeq 0.06$. This corresponds to $\text{SNR}_2 \simeq 3.2$ and $\text{SNR}_3 \simeq 1.2$ (and a combined $\text{SNR} \simeq 3.5$) which is in line with the constraints on the HD quadrupole and octupole in Fig. 2 of [24] and the amplitude detection at $(3.5 - 4)\sigma$ reported in [1]. If we had three times more pulsars (204) with the same observational characteristics we would find $N_f^{\text{eff}}(\ell = 2) \simeq 10$ ($\text{SNR}_2 \simeq 5$) and $N_f^{\text{eff}}(\ell = 3) \simeq 2.2$ ($\text{SNR}_3 \simeq 2.7$). If the 68 pulsars were all observed during 25 years, we would get $N_f^{\text{eff}}(\ell = 2) \simeq 8.6$ ($\text{SNR}_2 \simeq 4.6$) and $N_f^{\text{eff}}(\ell = 3) \simeq 1$ ($\text{SNR}_3 \simeq 1.9$). If we combine both improvements, that is 204 pulsars all monitored during 25 years, we would get $N_f^{\text{eff}}(\ell = 2) \simeq 18$ ($\text{SNR}_2 \simeq 6.8$), $N_f^{\text{eff}}(\ell = 3) \simeq 4.7$ ($\text{SNR}_3 \simeq 4.0$), and $N_f^{\text{eff}}(\ell = 4) \simeq 1$ ($\text{SNR}_4 \simeq 2.1$) and a combined $\text{SNR} \simeq 8.3$ for the amplitude of the background. Since both the observing time and the number of pulsars monitored is set to increase, we can state that we have entered the era of effective number of frequencies much larger than unity. From the above values we can hope that the octupole of the HD correlation should soon be clearly detected, and in the long run one should be able to detect the hexadecapole ($\ell = 4$). The multipole $\ell = 5$ would however require around 450 pulsars monitored during 25 years.

We also recover that in a regime where the sources of noise are negligible (ideal experiment), the SNR scales as in (159). Again, this has to be compared to (139), which is what we obtain when combining naively different frequencies. Taking one residual per month, that is $2Tf_{\text{max}} = 1/12 \text{ yr}^{-1}$, and an observation time of 15 years, one gets (independent of the spectrum) $\text{SNR}_\ell^2 \sim 180(2\ell + 1)/2$, i.e. a factor about 120 larger than what one gets by doing (139) for a spectrum with $\gamma = -13/3$. Notice that both the cadence of observation, which determines f_{max} , and the observation period will improve in the future, further increasing the SNR.

Summarizing: if we correctly combine the contributions of different frequencies (using optimal weights), in the context of a perfect experiment, we improve the SNR over a single frequency by a factor corresponding to the number of frequency modes. In contrast, by performing a brute-force computation as in Eq. (139), where no weights are applied, we only gain a factor which is an effective number of frequency modes $N_f^{\text{eff}} = h^4/\hbar^4$, and it can be very small for red spectra (since we effectively give too much weight to a few low frequencies). For an imperfect experiment (i.e. when instrumental noise is accounted for) the counting of the frequency bins is

weighted by the noise, and given by the multipole-dependent integral (170). Still, it grows with the observation time. Our simple derivation in multipole space makes it clear that by optimally combining different frequency bins, we can construct an estimator for the HD correlation multipoles whose SNR grows with observation time, even for red spectra. Hence, cosmic variance is actually not an irreducible noise component in this context, contrary to what happens for the CMB where we effectively have access to one single measurement for each direction and there is nothing to combine to improve the SNR of the correlation estimator.

VIII. BACKGROUND ANISOTROPY ESTIMATOR

Up to now we have considered an isotropic bath of gravitons, see section II C: in this context, the HD correlation is a signature of an isotropic spectral function $\mathcal{A}(k)$. The anisotropic part $\hat{\mathcal{A}}(\hat{k})$ of the GW background is reflected instead in the off-diagonal correlations in the (ℓ, m) space, see (70). For a finite number of pulsars, the constraints on anisotropy are better placed using a set of adapted functions [57, 58]. However for a large number of pulsars one can estimate (see section VII B 1) $z_{\ell m}(f)$, and estimators of anisotropy can be built out of the two-point function of $z_{\ell_1 m_1}(f_1)z_{\ell_2 m_2}(f_2)$, looking at the off-diagonal structure of the correlation. As an illustration, let us restore the anisotropic part of the GW spectrum in this section only. Our starting point is (70) for the two-point function of pulsar redshift in the presence of anisotropy. Our goal is to build an estimator for the anisotropic coefficient \mathcal{A}_{LM} in (70), evaluate its variance (along the lines of [18]) due to the limited angular resolution of a time-residual map, in the context of a perfect experiment with zero instrumental noise. We will then discuss how this noise propagates into a variance on the angular power spectrum of the background energy density, which is the observable usually used to quantify the deviation from isotropy. We will show that, even for a perfect experiment and in the absence of other noise components, the fact that the time-residual map has a limited angular resolution, limits our ability to measure the angular power spectrum from the background intensity (energy density) map.

A. Anisotropy in the time residual map: estimator and variance

We start by isolating the \mathcal{A}_{LM} modes in (70) by exploiting the orthonormality of Wigner 3j. Hence we introduce the combinations (contractions of the 2-point function with Wigner 3j symbol)

$$\begin{aligned} \mathcal{E}_{\ell_1 \ell_2}^{LM}(f_1, f_2) &\equiv \sqrt{2L+1} \sum_{m_1 m_2} \begin{pmatrix} L & \ell_1 & \ell_2 \\ M & m_1 & m_2 \end{pmatrix} \\ &\times \hat{z}_{\ell_1 m_1}(f_1) \hat{z}_{\ell_2 m_2}(f_2), \end{aligned} \quad (174)$$

which satisfies $\mathcal{E}_{\ell_1\ell_2}^{LM*}(f_1, f_2) = (-1)^M \mathcal{E}_{\ell_1\ell_2}^{L-M}(-f_1, -f_2)$ and obviously

$$\mathcal{E}_{\ell_1\ell_2}^{LM}(f_1, f_2) = \mathcal{E}_{\ell_2\ell_1}^{LM}(f_2, f_1). \quad (175)$$

The average of (174) can be computed using (70) and (A17) and is given by

$$\begin{aligned} \langle \mathcal{E}_{\ell_1\ell_2}^{LM}(f_1, f_2) \rangle_{\mathcal{A}} &= \delta_{\ell_1+\ell_2+L}^{\text{even}} \delta(f_1 + f_2) \hat{\mathcal{P}}(f_1) \\ &\quad \frac{\hat{\mathcal{A}}_{LM}^*}{\sqrt{4\pi}} F_{\ell_1} F_{\ell_2} \begin{pmatrix} L & \ell_1 & \ell_2 \\ 0 & -2 & 2 \end{pmatrix}. \end{aligned} \quad (176)$$

Hereafter we only consider parameters for which $\ell_1 + \ell_2 + L$ is even. In the case where the anisotropy is mild, we can assume the covariance is only due to the isotropic part. In that case, using (68) and (A17), the covariance takes the simple form

$$\begin{aligned} &\langle \mathcal{E}_{\ell_1\ell_2}^{LM}(f_1, f_2) \mathcal{E}_{\ell'_1\ell'_2}^{L'M'*}(f'_1, f'_2) \rangle_{\mathcal{A}} \\ &- \langle \mathcal{E}_{\ell_1\ell_2}^{LM}(f_1, f_2) \rangle_{\mathcal{A}} \langle \mathcal{E}_{\ell'_1\ell'_2}^{L'M'*}(f'_1, f'_2) \rangle_{\mathcal{A}} \\ &\simeq \delta_L^{L'} \delta_M^{M'} \left[\delta_{\ell_1}^{\ell'_1} \delta_{\ell_2}^{\ell'_2} \delta(f_1 - f'_1) C_{\ell_1}(f_1) \delta(f_2 - f'_2) C_{\ell_2}(f_2) \right. \\ &\quad \left. + \delta_{\ell_1}^{\ell'_2} \delta_{\ell_2}^{\ell'_1} \delta(f_1 - f'_2) C_{\ell_1}(f_1) \delta(f_2 - f'_1) C_{\ell_2}(f_2) \right]. \end{aligned} \quad (177)$$

If we want to combine the expressions (174) to form an optimal estimator for the $\hat{\mathcal{A}}_{LM}$, we must normalize them and weight them by their inverse variance [18]. Explicitly, after having discretized the frequency space (with the $f = 0$ mode removed), one builds the following estimator

$$\hat{\mathcal{A}}_{LM,f}^{\ell_1\ell_2,e} \equiv \frac{\sqrt{4\pi}}{TF_{\ell_1}F_{\ell_2}} \frac{1}{\hat{\mathcal{P}}(f)} \begin{pmatrix} L & \ell_1 & \ell_2 \\ 0 & -2 & 2 \end{pmatrix}^{-1} \mathcal{E}_{\ell_1\ell_2}^{LM}(f, -f)^*, \quad (178)$$

where on the left-hand side we have explicitly indicated that one can build an estimator for each given value of ℓ_1 and ℓ_2 . From (176) the mean of this object is trivially given by $\hat{\mathcal{A}}_{LM}$ while, assuming that anisotropies are mild, and using (177), the variance of this object (we define $\text{Var}(X) \equiv \langle XX^* \rangle - \langle X \rangle \langle X^* \rangle$ hence $\sqrt{\text{Var}}$ is the associated error) is given by

$$\text{Var}(\hat{\mathcal{A}}_{LM,f}^{\ell_1\ell_2,e}) = \frac{4\pi}{(2\ell_1+1)(2\ell_2+1)} \begin{pmatrix} L & \ell_1 & \ell_2 \\ 0 & -2 & 2 \end{pmatrix}^{-2}. \quad (179)$$

We recall that all averages here are taken over GW realisations, i.e. we make use of the average $\langle \dots \rangle_{\mathcal{A}}$ defined in section IV. We then combine all the estimators $\hat{\mathcal{A}}_{LM,f}^{\ell_1\ell_2,e}$ with inverse variance weighting to obtain the minimum-variance estimator. However, from (175) we deduce the symmetry property

$$\hat{\mathcal{A}}_{LM,f}^{\ell_1\ell_2,e} = \hat{\mathcal{A}}_{LM,-f}^{\ell_2\ell_1,e}, \quad (180)$$

hence in order to avoid double counting when combining estimators we can either restrict to $f > 0$ and use all ℓ_1, ℓ_2 combinations, or we can use both $f > 0$ and $f < 0$ and restrict to $\ell_1 \leq \ell_2$. We make the former choice which is simpler in

our case;⁹ and obtain (with an implied restriction of the sums below to $\ell_1 + \ell_2 + L$ being even)

$$\hat{\mathcal{A}}_{LM}^e \equiv \frac{\sum_{\ell_1\ell_2} \sum_{f>0} \hat{\mathcal{A}}_{LM,f}^{\ell_1\ell_2,e} \text{Var}^{-1}(\hat{\mathcal{A}}_{LM,f}^{\ell_1\ell_2,e})}{\sum_{\ell_1\ell_2} \sum_{f>0} \text{Var}^{-1}(\hat{\mathcal{A}}_{LM,f}^{\ell_1\ell_2,e})}. \quad (181)$$

The variance of this estimator is by construction

$$\text{Var}^{-1}(\hat{\mathcal{A}}_{LM}^e) = \sum_{\ell_1\ell_2} \sum_{f>0} \text{Var}^{-1}(\hat{\mathcal{A}}_{LM,f}^{\ell_1\ell_2,e}), \quad (182)$$

or explicitly

$$\begin{aligned} &\text{Var}^{-1}(\hat{\mathcal{A}}_{LM}^e) \\ &= \frac{N_f}{2} \frac{1}{(4\pi)} \sum_{\ell_1\ell_2} \delta_{\ell_1+\ell_2+L}^{\text{even}} (2\ell_1+1)(2\ell_2+1) \begin{pmatrix} L & \ell_1 & \ell_2 \\ 0 & -2 & 2 \end{pmatrix}^2. \end{aligned} \quad (183)$$

In this equation we assumed that in the absence of instrumental noise all frequencies equally contribute, and N_f is equal to the number of frequency modes, i.e. $N_f = 2Tf_{\text{max}}$. In the presence of noise, a proper weighting should be introduced. This result agrees with what obtained in [18] (where it is implicitly assumed that $N_f = 1$). Notice that all these sums run over $\ell_1, \ell_2 \leq \ell_{\text{max}}$ with $\ell_{\text{max}} \sim \sqrt{N_p}$, where N_p is the number of pulsars in the network, and the multipole measured is constrained by $0 \leq L \leq 2\ell_{\text{max}}$.

B. Angular power spectrum of energy density: estimator and variance

We conclude this section by relating this anisotropy estimator to the estimator of the energy density angular power spectrum, which is the observable usually used to quantify anisotropies [59]. Let us define $(\Omega_{\text{GW}})_{LM} = 8\pi^2/(3H_0^2) f^5 \mathcal{A}(2\pi f) \hat{\mathcal{A}}_{LM}$ whose estimator is

$$(\Omega_{\text{GW}})_{LM}^e \equiv \frac{8\pi^2}{3H_0^2} f^5 \mathcal{A}(2\pi f) \hat{\mathcal{A}}_{LM}^e. \quad (184)$$

The mean and variance of this estimator can be trivially computed from the ones of $\hat{\mathcal{A}}_{LM}^e$. An estimator for the angular power spectrum of the anisotropies is given by

$$C_L^{\Omega_{\text{GW}},e} \equiv \frac{1}{2L+1} \sum_M (\Omega_{\text{GW}})_{LM}^e (\Omega_{\text{GW}})_{LM}^{e*}, \quad (185)$$

⁹ The latter choice is however more adapted when we consider an equal time correlator, as is implicitly the case in [18]. The associated estimator is $\hat{\mathcal{A}}_{LM}^{\ell_1\ell_2,e} \equiv \sum_f \hat{\mathcal{A}}_{LM,f}^{\ell_1\ell_2,e}$ where the sum is on both positive and negative frequencies. When considering the symmetry property (180) for $\ell_1 = \ell_2$, we deduce that $\text{Var}(\hat{\mathcal{A}}_{LM}^{\ell\ell,e})$ is twice its counterpart with $\ell_1 \neq \ell_2$ due to covariance between positive and negative frequencies, that is $\text{Var}(\hat{\mathcal{A}}_{LM}^{\ell_1\ell_2,e})$ is given by the r.h.s of (179) multiplied by $1 + \delta_{\ell_1\ell_2}$ as in Eq. (26) of [18]. When combining the estimators we must also restrict to $\ell_1 \leq \ell_2$ so as to avoid double counting. Then, when the variance of the optimal estimator is recast as a sum on all ℓ_1, ℓ_2 combinations, the $1/(1 + \delta_{\ell_1\ell_2})$ is exactly what is needed to recover (183) with $N_f = 1$, as is expected when only a single equal time measurement is performed.

with variance given by

$$\text{Var}\left(C_L^{\Omega_{\text{GW},e}}\right) = \frac{2}{(2L+1)^2} \quad (186)$$

$$\times \sum_M \left[\text{Var}^2((\Omega_{\text{GW}})_{LM}^e) + 2\text{Var}((\Omega_{\text{GW}})_{LM}^e)|(\Omega_{\text{GW}})_{LM}|^2 \right].$$

We stress again that the average is an average over GW realizations, and there is a residual stochasticity due to cosmological initial conditions (see also [36] for a detailed discussion of this double average procedure). In other words, the result above holds for any cosmological realization. To get predictions, one has to take an additional average over cosmological initial conditions, and using (16) and (17) the cosmological average of the variance is

$$\langle \text{Var}\left(C_L^{\Omega_{\text{GW},e}}\right) \rangle_{\text{cos}} = \frac{2}{2L+1} \times \quad (187)$$

$$\left[\left(\frac{\bar{\Omega}_{\text{GW}}}{4\pi} \right)^4 \text{Var}^2(\hat{\mathcal{A}}_{LM}^e) + 2 \left(\frac{\bar{\Omega}_{\text{GW}}}{4\pi} \right)^2 \text{Var}(\hat{\mathcal{A}}_{LM}^e) C_L^{\Omega_{\text{GW}}} \right],$$

which can be directly computed from (183). It follows that for a given pulsar number N_p (determining the angular resolution of a time-residual map, as $\ell_{\text{max}} \sim \sqrt{N_p}$), (187) provides us with the precision with which, in a perfect experiment, the angular power spectrum of the background energy density can be computed from a time-residual map.

IX. THE HELLINGS-DOWNS CURVE IN THE CMB

There is a formal analogy between the redshift to a pulsar and the tensor modes contribution to the CMB temperature fluctuations. We want to address the following questions: is there a HD-like correlation in the context of the CMB temperature fluctuations? And what is cosmic variance in this case? These questions have been addressed in [60, 61] in the case of a narrow GW background generated within the horizon. We study them for the GW background generated by amplification of the vacuum fluctuations in the primordial Universe, highlighting the main differences when identifying the HD-like correlation. In this section, η will denote the conformal time associated with a cosmological scale factor $a(\eta)$.

A. GW in cosmology

The background of gravitational waves generated in the early Universe [50] during an inflationary phase, differs from the background considered so far for PTA by three important details.

- First, the modes satisfy a damped wave equation instead of a simple wave equation, because of Hubble friction. Namely the GW, which are tensor cosmological perturbations, satisfy

$$h''_{ij} + 2\mathcal{H}h'_{ij} + k^2 h_{ij} = 0. \quad (188)$$

where a prime indicates a derivative with respect to conformal time η . In a matter dominated era, the conformal Hubble rate scales as $\mathcal{H} = a'/a = 2/\eta$ and the independent solutions of this equations are

$$\frac{j_1(k\eta)}{(k\eta)}, \quad \frac{n_1(k\eta)}{(k\eta)}. \quad (189)$$

When generalizing (7) to a matter dominated cosmological context, we thus need to use the replacement

$$e^{ik\eta} \rightarrow s(k\eta), \quad e^{-ik\eta} \rightarrow s^*(k\eta), \quad (190)$$

with the cosmological solution s defined in terms of Hankel functions

$$s(x) = -\frac{j_1(x) + in_1(x)}{x}. \quad (191)$$

The solutions (189) are only valid for a matter dominated Universe. In radiation domination, $\mathcal{H} = 1/\eta$ and the solutions are simply $j_0(k\eta)$ and $n_0(k\eta)$, and $s(x) = [j_0(k\eta) + in_0(k\eta)]$.

- The solution must be regular when outside the Hubble horizon at early times ($k\eta \ll 1$). Hence since only $j_1(x)/x$ is regular for the matter era solution (resp. $j_0(x)$ for the radiation era solution), the initial conditions for a mode \mathbf{k} and $-\mathbf{k}$ must be related to eliminate $n_1(k\eta)/(k\eta)$ (resp. $n_0(k\eta)$). It follows that the initial conditions are not set with random phases in cosmology. This translates into a modification of the statistics of the modes as we must have

$$h^{(m)\star}(\mathbf{k}) = h^{(m)}(-\mathbf{k}), \quad (192)$$

such that

$$h_{\text{cos}}^{(m)}(\mathbf{k}, \eta) = h^{(m)}(\mathbf{k}) \text{Re}[s(k\eta)]. \quad (193)$$

The link (192) implies that the statistics of the mode is modified, and instead of (12), they must obey

$$\langle h^{(m)}(\mathbf{k}) h^{(m')\star}(\mathbf{k}') \rangle_{\mathcal{A}} = (2\pi)^3 \delta_{mm'} \delta(\mathbf{k} - \mathbf{k}') \mathcal{A}(k),$$

$$\langle h^{(m)}(\mathbf{k}) h^{(m')\star}(\mathbf{k}') \rangle_{\mathcal{A}} = (2\pi)^3 \delta_{mm'} \delta(\mathbf{k} + \mathbf{k}') \mathcal{A}(k). \quad (194)$$

The statistics of $h_{\text{cos}}^{(m)}$ is also altered compared to (13), and it becomes

$$\langle h_{\text{cos}}^{(m)}(\mathbf{k}, \eta) h_{\text{cos}}^{(m')\star}(\mathbf{k}', \eta') \rangle_{\mathcal{A}} = \delta_{mm'} (2\pi)^3 \delta(\mathbf{k} - \mathbf{k}') \quad (195)$$

$$\times \mathcal{A}(k) \text{Re}(s(k\eta)) \text{Re}(s^*(k'\eta')).$$

- The general expression for the redshift (21) and its splitting between an observer and a pulsar term remains valid, if we replace the pulsar distance χ by the last-scattering surface (LSS) distance χ_{LSS} (and the integrals (23) do not extend to infinity but go up to the big-bang distance). However a major difference, detailed hereafter, is that it is generally not correct to neglect the pulsar term by assuming $k\chi_{\text{LSS}} \gg 1$ since in cosmology the spectrum is nearly scale invariant, and large modes which evade this condition, contribute substantially.

The difference in statistics, that is namely the replacement of (12) with (194) when the time evolution is modified according to (190), has an important consequence for the equivalent of the HD correlation in the CMB correlations. Apart from the argument of regularity on super-Hubble scales, the origin of the difference is rooted in the quantum origin of fluctuations. A correlation of the type (12) might be of classical origin, due to independence of a large number of binary sources. But equivalently it could be of quantum origin as we could promote, as in (2) of [61], the tensor perturbations to a quantum field by promoting the mode functions to operators (note that in this discussion a hat refers to an operator and not to a Fourier transform)

$$h^{(m)}(\mathbf{k}) \rightarrow \hat{h}^{(m)}(\mathbf{k}) = \hat{a}_m(\mathbf{k})h^{(m)}(\mathbf{k}), \quad (196)$$

where the creation and annihilation operators satisfy the usual commutation rules

$$[\hat{a}_m(\mathbf{k}), \hat{a}_{m'}^\dagger(\mathbf{k}')] = (2\pi)^3 \delta_{mm'} \delta^3(\mathbf{k} - \mathbf{k}'). \quad (197)$$

The statistics (12) is then interpreted as an average in the quantum vacuum using $\hat{a}_m(\mathbf{k})|0\rangle = 0$, and defining $2\mathcal{A}(\mathbf{k}) = |h^{(2)}(\mathbf{k})|^2 = |h^{(-2)}(\mathbf{k})|^2$ for an unpolarized background. However tensor fluctuations generated during inflation are assumed to be generated in the Bunch-Davies vacuum when asymptotically inside the Hubble radius, and then this leads to frozen solutions when their scale exits the Hubble radius. It is this frozen value which serves as an initial state for the subsequent standard cosmological model, but the statistical properties are then inherited from the inflationary phase. The important consequence is that in inflation this translates for the initial conditions of the standard model into the operators

$$\hat{h}^{(m)}(\mathbf{k}) = \hat{A}_m(\mathbf{k})h^{(m)}(\mathbf{k}), \quad \hat{A}_m(\mathbf{k}) \equiv \frac{\hat{a}_m(\mathbf{k}) + \hat{a}_m^\dagger(-\mathbf{k})}{\sqrt{2}} \quad (198)$$

and the mode functions must satisfy (192). Although of quantum origin, the tensor modes behave classically since $[\hat{A}_m(\mathbf{k}), \hat{A}_{m'}^\dagger(\mathbf{k}')] = 0$, and we find the modified statistics (194). Physically, the tensor modes of inflationary origin are a superposition of counter propagating waves forming a standing wave, hence a mode with wavevector \mathbf{k} and the one with opposite wavevector $-\mathbf{k}$ are not statistically independent. This is very different from a classical GW background from binary sources, where the binary sources in one directions have absolutely no reason to be correlated with the ones in the opposite direction. As we detail hereafter this triggers a small difference for the HD correlation in the CMB.

B. Redshift multipole transfer function

Similarly to the PTA case, we can decompose CMB redshifts in a basis of spherical harmonics [51] as (24)

$$z^{\text{CMB}}(\mathbf{n}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \sum_{m=\pm 2} \sum_{\ell} z_{\ell m}^{\text{CMB}}(\mathbf{k}, \eta_0) R_{\hat{\mathbf{k}}} [Y_{\ell}^m](\mathbf{n}). \quad (199)$$

Observe that what is measured in CMB observations is the temperature: on this context $z^{\text{CMB}}(\mathbf{n})$ stands for $-\delta T^{\text{CMB}}(\mathbf{n})/\bar{T}^{\text{CMB}}$ as the GW affect energies, and thus temperatures, via the integrated Sachs-Wolfe effect. This can be split into the sum of an observer and a pulsar term, where the pulsar term is in fact a LSS term as we must use $\chi = \chi_{\text{LSS}}$ in (21). Note that η cannot go infinitely in the past but is restricted to $0 \leq \eta \leq \eta_0$.

The multipole transfer function is defined in Boltzmann codes such as CLASS [62, 63], as the ratio

$$\mathcal{T}_{\ell m}(\mathbf{k}) \equiv \frac{z_{\ell m}^{\text{CMB}}(\mathbf{k}, \eta_0)}{h^{(m)}(\mathbf{k})}. \quad (200)$$

A key difference with the background measured by PTA, is that we must replace the evolution of the modes with the damped solution according to (190). The general form of the transfer requires the resolution of the Boltzmann hierarchy, usually with a line of sight method. However we can estimate rather easily the observer contribution by approximating the time evolution for large modes (locally, near the observer) by

$$s(k\eta) \sim \frac{e^{ik\eta}}{(k\eta)^2} \simeq \frac{e^{ik\eta}}{(k\eta_0)^2}, \quad (201)$$

that is we keep the oscillatory behaviour but we consider only the final value for the damping which has occurred due to Hubble friction. Then we can just follow the same steps as for a GW background observed with PTA. The final result (42) is then only altered by this damping $1/(k\eta_0)^2$ but can be written in a simpler form due to the relation (192). Eventually we find that the observer contribution can be approximately written as

$$z_{\ell m}^{\text{CMB,o}}(\mathbf{k}, \eta_0) \simeq (-i)^\ell \frac{\cos(k\eta_0 + \ell\frac{\pi}{2})}{(k\eta_0)^2} h^{(m)}(\mathbf{k}) F_\ell^o, \quad (202)$$

with the same F_ℓ^o as in (34). The observer part of the transfer function $\mathcal{T}_{\ell m}^o$ can be read directly from this result (202) recalling its definition (200), and it depends only on k .

The multipole transfer function (200) is plotted in Fig. 8, for different values of k . The modulation $\cos(k\eta_0 + \ell\frac{\pi}{2})$, is the main difference with a classical GW background measured by PTA. Since cosmological GW are standing waves, they have nodes and maxima, and depending on where these are located with respect to the observer, this affects how the integrated SW effect is split in the various multipoles. It is fundamentally due to using the statistics (194) instead of (12) as required by the origin of primordial GW modes. This feature is absent for backgrounds generated within inside the horizon in a later phase, hence the modulation is not seen in the Fig. 1 of [61] even though it corresponds to sharply peaked spectra. Apart from this modulation we see that the rescaled transfer function $\mathcal{T}_{\ell m}/\sqrt{4\pi(2\ell+1)}$ scales like the observer term $F_\ell^o/\sqrt{4\pi(2\ell+1)} \propto \sqrt{(\ell-2)!/(\ell+2)!}$ for small ℓ and large modes and then it gets dominated by a pulsar term (which is a last scattering term in cosmology).

We also notice that this pulsar term, which has roughly a slope $\propto \ell^2$ stops when $\ell \geq k\eta_0$ as expected since the distance to the last scattering surface is approximately η_0 . For small

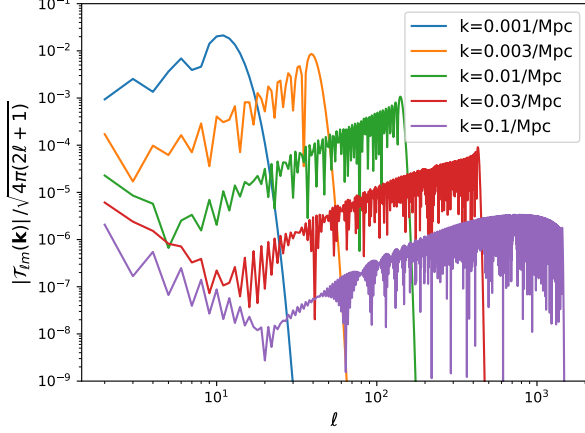


FIG. 8. Transfer function (200) for several modes and $m=2$ obtained from a modified version of the Boltzmann solver CLASS [62, 63]. For large modes ($k > 0.01/\text{Mpc}$ approximately) and low ℓ we see an observer term with its characteristic decay $\propto (\ell-2)!/(\ell+2)!$ with a modulation coming from the factor $\cos(k\eta_0 + \ell\frac{\pi}{2})$, which is then dominated by the pulsar (also known as the LSS term) for large values of ℓ . It is the observer part of the transfer function which is responsible for the appearance of a HD-like correlation in the CMB correlation function. The background cosmology parameters are $\Omega_\Lambda = 0$, $\tau_{\text{reio}} = 0$ and $h_0 = 0.4$.

modes k , we are not in the regime where the wavelength is much shorter than the distance to the last scattering surface (playing the counterpart of pulsars in our analogy) and the observer term is always subdominant compared to the pulsar term.

C. Correlation function

In full analogy with the PTA case, we define a redshift (or CMB temperature fluctuations) correlation function or its equivalent in multipolar space when decomposing $z^{\text{CMB}}(\mathbf{n}) = \sum_{\ell m} z_{\ell m}^{\text{CMB}} Y_{\ell}^m(\mathbf{n})$ which is

$$\langle z_{\ell m}^{\text{CMB}} z_{\ell' m'}^{\text{CMB}*} \rangle_{\mathcal{A}} = \delta_{\ell \ell'} \delta_{m m'} C_{\ell}^{\text{CMB}}. \quad (203)$$

All time dependence has been eliminated compared to the PTA case because an observation of the CMB sky amounts to a single observation given the extremely slow evolution of the CMB (over a scale of millions of years) compared with the observation time. The expression of the correlation in multipole space (the angular power spectrum) is rather simple in terms of the transfer function

$$C_{\ell}^{\text{CMB}} = \int \frac{k^2 dk}{2\pi^2} \mathcal{A}(k) \sum_{m=\pm 2} |\mathcal{T}_{\ell m}(k)|^2. \quad (204)$$

In Fig. 9 we plot the C_{ℓ}^{CMB} for various spectral indices defined in the cosmological context as $\mathcal{A}(k) \propto (k/k_*)^{n_T-3}$. We see that for the scale invariant cosmological spectrum with

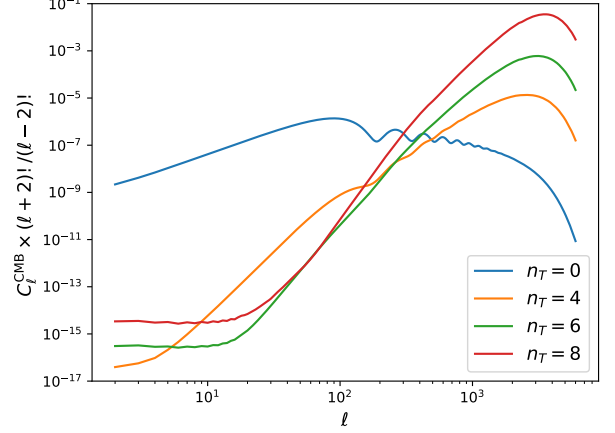


FIG. 9. Multipoles of the CMB fluctuations from tensor modes obtained from CLASS [62, 63]. The background cosmology parameters are $\Omega_\Lambda = 0$, $\tau_{\text{reio}} = 0$, $h_0 = 0.4$, $A_s = 2.1 \times 10^{-9}$ and unit tensor to scalar ratio. We plot $(\ell+2)!/(\ell-2)!C_{\ell}^{\text{CMB}}$ so that a HD correlation appears as a plateau, which is visible for $n_T = 6$ and $n_T = 8$.

$n_T = 0$, we do not see the Hellings-Downs correlation as large angular scales are dominated by small modes which exhibit only the pulsar contribution (see Fig. 8). However if we consider a very blue spectrum, then these small modes are subdominant and even the large scale angular correlations are dominated by the observer Hellings-Downs correlation of the large modes. As we plot $(\ell+2)!/(\ell-2)!C_{\ell}^{\text{CMB}}$, the HD correlation appears as a plateau. The modulation in $\cos(k\eta_0 + \ell\frac{\pi}{2})$ of the transfer is not visible because when integrating on k it is averaged over since the modes which contribute to the HD plateau satisfy $k\eta_0 \gg 1$. In order to have a large scale HD plateau we need at least $n_T > 4$ so that the primordial conditions can compensate the subsequent damping $[1/(k\eta_0)^2]^2$ due to Hubble friction.

Finally, as done in the PTA context, we can build an estimator for the redshift correlations as (we stress that for the sake of the comparison with PTA, we are here assuming unrealistically that there are no scalar fluctuations in the CMB anisotropies) $C_{\ell}^{\text{CMB},e} = \sum_m z_{\ell m}^{\text{CMB}}(\eta_0) z_{\ell m}^{\text{CMB}*}(\eta_0)/(2\ell+1)$. Since in this context there are no different time of observations, nor frequency bins, this is the only estimator one can build. The variance is estimated in the usual way with Wick's theorem and the signal to noise reduces to $\text{SNR}_{\ell}^2 = (2\ell+1)/2$ which is the usual CMB cosmic variance, irreducible, since there is no way to mitigate it, contrary to the PTA case.¹⁰

¹⁰ When we say that nothing can be done to mitigate cosmic variance in the context of the CMB, we mean that we will never achieve the sensitivity required to directly observe the timescale of GWs that influence the CMB. If we could observe over a timescale of one million years, techniques like those used in PTA might allow us to mitigate cosmic variance. In other words, when we refer to a *perfect experiment*, we are implying an experiment that can be conducted within a humanly feasible timescale.

X. DISCUSSION AND CONCLUSIONS

In this article, we critically examined the derivation of the cosmic variance associated with the HD correlation curve and explored its implications for PTA observations. By working in real and multipole space, we showed that an optimal estimator for the angular power spectrum of the HD correlation can be constructed. This estimator's signal-to-noise ratio improves with increased observation time and increasing the cadence of observations, even in the presence of a very red spectrum (see eq. (159)), hence it will keep on improving in the future. The key to this improvement lies in the introduction of an optimal weighting scheme, which reduces the impact of cosmic variance by mitigating the contribution of low-frequency components in the presence of a red spectrum. We also considered the effect of instrumental noise, we accounted for pulsar shot noise, and discussed a weighting procedure to minimize the impact of these noise components. Additionally, we provided a detailed analysis of how the optimal estimator is modified when accounting for the fact that, in a real experiment, time residuals must be fitted with a deterministic template to subtract the effects of a linear drift and a quadratic term. We showed that this filtering procedure does not alter the final conclusion regarding the SNR in the strong signal limit.

In this regard, the term *cosmic variance* is somewhat of an abuse of language in the PTA context, as it suggests that nothing can be done to reduce its impact, even with perfect experiments, and the authors of [19] have advocated to rename it *sample variance*. This stands in contrast to the situation in CMB observations, where even in a perfect experiment, more observation time cannot reduce cosmic variance as the CMB signal is essentially static on any human time scale. In the CMB context, cosmic variance is indeed an irreducible limitation.

To make this point clear, we developed a formal analogy between the redshift to a pulsar and the tensor modes contribution to the redshift to the large scattering surface. We showed that two contributions can be distinguished in the CMB redshift: an observer term which dominates small multipoles and a pulsar term which becomes important at higher multipoles (smaller scales). The transition scale depends on the mode k under consideration (the observer term becoming visible at small ℓ for $k > 0.01/\text{Mpc}$ approximately). We computed the redshift correlation function, and we showed that its multipolar decomposition (i.e. the angular power spectrum) exhibits a HD-like behaviour at small ℓ for a sufficiently blue spectrum with $n_T > 4$, while for redder spectra small modes, which exhibit only the pulsar contribution, dominate the spectrum hence the HD-behaviour is lost. Since the tensor modes have a subdominant effect compared to the scalar modes, and since in standard inflationary models the tensor spectrum is nearly scale-invariant, the HD correlation in the CMB will realistically never be observed. We recovered that any estimator for the angular power spectrum is cosmic-variance limited in this context since no time weighting is possible.

We conclude with two remarks. First, we stress that in PTA data analysis techniques, estimators for the correlation function are constructed using optimal weightings, as discussed,

for example, in [44]. However, since several recent theoretical papers have referred to cosmic variance as a fundamental limitation for PTA observations, we believe it is important to avoid further confusion by providing a pedagogical derivation of the weighting procedure used to optimally combine frequency bins. To this end, we also draw a parallel with the CMB case, where such a weighting procedure is not possible, even though a HD correlation is present for very blue primordial spectra. While working in multipole space offers a direct parallel with CMB observables, it is also the simplest representation for HD analysis. Indeed, we have shown that in multipole space, the covariance matrix is diagonal, significantly simplifying computations. This suggests that the multipole space framework, rather than real space, is the most effective approach for studying HD correlations in PTA observations.

Acknowledgements

We thank Camille Bonvin, Nastassia Grimm, Guillaume Faye, Marc Kamionkowski and Martin Pijenburg for discussions. G.C. acknowledges support from CNRS and from the Swiss National Science Foundation (Ambizione grant, "Gravitational wave propagation in the clustered universe").

Appendix A: Rotations and spherical harmonics

We gather all relations related to spherical harmonics and rotations which are used in this article. We use the conventions of [64] to define spherical harmonics and the Wigner rotation coefficients, which are defined by

$$Y_\ell^m = \langle \mathbf{n} | \ell m \rangle, \quad D_{m'm}^\ell = \langle \ell m' | R | \ell m \rangle \quad (\text{A1})$$

where an active rotation is parameterized with Euler angles as

$$R(\phi, \theta, \gamma) = R_z(\phi) \cdot R_y(\theta) \cdot R_z(\gamma). \quad (\text{A2})$$

The active rotation of spherical harmonics is defined by the inverse passive transformation of the direction since

$$R[Y_\ell^m](\mathbf{n}) \equiv \langle \mathbf{n} | R | \ell m \rangle = Y_\ell^m(R^{-1} \cdot \mathbf{n}) = D_{Mm}^\ell(R) Y_\ell^M(\mathbf{n}). \quad (\text{A3})$$

The Wigner coefficients are related to the small Wigner coefficients $d_{mm'}^\ell$ associated with a R_y rotation only by

$$D_{mm'}^\ell(\phi, \theta, \gamma) = e^{-im\phi} d_{mm'}^\ell(\theta) e^{-im'\gamma}. \quad (\text{A4})$$

In particular when $m = \pm 2$ and $m' \pm 2$ and noting $\cos \theta = \beta$, the small Wigner coefficients are expressed in terms of Jacobi polynomials as

$$d_{m'm}^\ell(\theta) = \left(\frac{1+\beta}{2} \right)^{\frac{|m+m'|}{2}} \left(\frac{1-\beta}{2} \right)^{\frac{|m-m'|}{2}} P_{\ell-2}^{(|m-m'|, |m+m'|)}(\beta), \quad (\text{A5})$$

which satisfy from their orthonormality relation the particular relations

$$\begin{aligned} \frac{2}{2\ell+1} &= \int_{-1}^1 dx \left(\frac{1+x}{2} \right)^2 \left(P_{\ell-2}^{(0,4)}(x) \right)^2 \\ &= \int_{-1}^1 dx \left(\frac{1-x}{2} \right)^2 \left(P_{\ell-2}^{(4,0)}(x) \right)^2. \end{aligned} \quad (\text{A6})$$

The Wigner coefficients are also related to the spin-weighted spherical harmonics through

$$D_{ms}^\ell(\phi, \theta, \gamma) = {}_sY_\ell^{-m}(\theta, \phi)(-1)^m \sqrt{\frac{4\pi}{2\ell+1}} e^{-is\gamma}. \quad (\text{A7})$$

The most useful properties are

$$D_{m'm}^\ell(R^{-1}) = D_{mm'}^{\ell*}(R) = (-1)^{m-m'} D_{-m-m'}^\ell(R), \quad (\text{A8})$$

$$D_{Mm}^\ell(\varphi + \pi, \pi - \theta, 0) = (-1)^\ell D_{M-m}^\ell(\varphi, \theta, 0), \quad (\text{A9})$$

from which we deduce in particular

$${}_sY_\ell^{m*} = (-1)^{s+m} {}_{-s}Y_\ell^{-m}. \quad (\text{A10})$$

The orthogonality property of Wigner coefficients is expressed as

$$\int \frac{d^3R}{8\pi^2} D_{ms}^\ell(R) D_{m's'}^{\ell'*}(R) = \frac{\delta_{\ell\ell'} \delta_{mm'} \delta_{ss'}}{2\ell+1}, \quad (\text{A11})$$

where an integral on all rotations is defined as

$$\int d^3R \equiv \int_0^{2\pi} d\gamma \int_0^\pi \sin\theta d\theta \int_0^{2\pi} d\phi = 8\pi^2. \quad (\text{A12})$$

In particular this leads to the orthonormality of spin-weighted spherical harmonics

$$\int d^2n {}_sY_\ell^m(\mathbf{n}) {}_sY_{\ell'}^{m'*}(\mathbf{n}) = \delta_{\ell\ell'} \delta_{mm'}, \quad (\text{A13})$$

whose counterpart for $s = 0$ are the addition theorems

$$\sum_m Y_\ell^m(\mathbf{n}) Y_\ell^{m*}(\mathbf{n}') = \frac{2\ell+1}{4\pi} P_\ell(\mathbf{n} \cdot \mathbf{n}'), \quad (\text{A14})$$

$$\sum_\ell \frac{(2\ell+1)}{4\pi} P_\ell(\mathbf{n} \cdot \mathbf{n}') = \delta_D^2(\mathbf{n}, \mathbf{n}'), \quad (\text{A15})$$

where δ_D^2 is the Dirac delta on the unit sphere.

The integral of three spin-weighted spherical harmonics is

$$\begin{aligned} \int {}_{s_1}Y_{\ell_1}^{m_1} {}_{s_2}Y_{\ell_2}^{m_2} {}_{s_3}Y_{\ell_3}^{m_3} d^2\mathbf{n} &= \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \\ &\times \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ -s_1 & -s_2 & -s_3 \end{pmatrix}. \end{aligned} \quad (\text{A16})$$

The Wigner 3j symbols which enter the previous expression satisfy

$$\sum_{m_1 m_2} \begin{pmatrix} \ell_1 & \ell_2 & L \\ m_1 & m_2 & M \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & L' \\ m_1 & m_2 & M' \end{pmatrix} = \frac{\delta_{L'L}^M \delta_{M'M}^L}{2L+1}, \quad (\text{A17})$$

when the triangular inequality is satisfied, and

$$\sum_{LM} (2L+1) \begin{pmatrix} L & \ell_1 & \ell_2 \\ M & m_1 & m_2 \end{pmatrix} \begin{pmatrix} L & \ell_1 & \ell_2 \\ M & m'_1 & m'_2 \end{pmatrix} = \delta_{m'_1 m_1}^{m'_2 m_2}. \quad (\text{A18})$$

Compatibility of (A16) with (A13) implies

$$\begin{pmatrix} 0 & \ell & \ell \\ 0 & m & -m' \end{pmatrix} = \delta_{mm'} \frac{(-1)^{\ell+m'}}{\sqrt{2\ell+1}}. \quad (\text{A19})$$

Appendix B: Quadratic estimator for perfect experiments

Let us assume we have N real-valued Gaussian random variable x^i with correlation $\langle x_i x_j \rangle = AC_{ij}$. The x^i correspond to residual or redshift measurements at the i/T frequency for instance. We want to evaluate A which is unknown, assuming that the overall shape of the correlations, C_{ij} , is known theoretically.

Let us consider a quadratic estimator (we use implicit Einstein summation)

$$Q = x_j W_{ji} x_i, \quad (\text{B1})$$

where W_{ij} is symmetric. Its average and variance are

$$\langle Q \rangle = AW_{ij} C_{ji} = A \text{Tr}(C \cdot W), \quad (\text{B2a})$$

$$\begin{aligned} \langle Q^2 \rangle - (\langle Q \rangle)^2 &= 2A^2 C_{ji} W_{ij} C_{ji'} W_{i'i'} \\ &= 2A^2 \text{Tr}(C \cdot W \cdot C \cdot W). \end{aligned} \quad (\text{B2b})$$

With a positive definite symmetric matrix \mathcal{Q} one can always define a scalar product in the space of symmetric matrices by

$$\{M_{ij}; N_{ij}\}_{\mathcal{Q}} \equiv \mathcal{Q}_{ji} M_{ij} \mathcal{Q}_{ji'} N_{i'i'} = \text{Tr}(\mathcal{Q} \cdot M \cdot \mathcal{Q} \cdot N), \quad (\text{B3})$$

which is positive definite since the correlation matrix is also positive definite. Then we can rewrite

$$\langle Q \rangle = A \{W_{ij}; C_{ij}^{-1}\}_C, \quad (\text{B4a})$$

$$\langle Q^2 \rangle - (\langle Q \rangle)^2 = 2A^2 \{W_{ij}; W_{ij}\}_C. \quad (\text{B4b})$$

The SNR ratio is

$$\text{SNR} = \frac{\langle Q \rangle}{\sqrt{\langle Q^2 \rangle - (\langle Q \rangle)^2}}, \quad (\text{B5})$$

which can be written as $\text{SNR} = \sqrt{N^{\text{eff}}/2}$ with

$$N^{\text{eff}} \equiv \frac{\{W_{ij}; C_{ij}^{-1}\}_C^2}{\{W_{ij}; W_{ij}\}_C}. \quad (\text{B6})$$

We find from Cauchy-Schwarz inequality

$$\text{SNR} \leq \sqrt{\frac{\{C_{ij}^{-1}; C_{ij}^{-1}\}_C}{2}} = \sqrt{\frac{N}{2}}, \quad (\text{B7})$$

and this inequality saturates for

$$W_{ij}^{\text{opt}} = C_{ij}^{-1}. \quad (\text{B8})$$

To summarize, the optimal weight in (B1) is the inverse of the correlation shape for which the effective degrees of freedom used is $N^{\text{eff}} = N$, which is the maximum available. Note that since $\{W_{ij}; C_{ij}^{-1}\}_C$ can be made as small as desired (it can be made to vanish by projection of W_{ij} in the space orthogonal to C_{ij}^{-1}) the SNR can be extremely small, by taking a suboptimal choice for W .

Clearly the minimum requirement is that W_{ij} is positive definite. But even in that case the SNR can be quite damaged

without optimization, as in full generality N^{eff} is only constrained by $N^{\text{eff}} \geq 1$. To show this we can diagonalize the correlation in an orthogonal basis, hence there exists an orthogonal matrix P and diagonal matrix D with positive entries such that $C = P \cdot D \cdot P^{-1}$ and the square root is defined by $\sqrt{C} = P \cdot \sqrt{D} \cdot P^{-1}$. Then $C \cdot W = \sqrt{C} \cdot (\sqrt{C} \cdot W \cdot \sqrt{C}) \sqrt{C}^{-1}$ shows that $C \cdot W$ is similar to the symmetric matrix $\sqrt{C} \cdot W \cdot \sqrt{C}$ which can also be diagonalized in an orthonormal basis by using the matrix R , with diagonal matrix d such that $\sqrt{C} \cdot W \cdot \sqrt{C} = R \cdot d \cdot R^{-1}$. The diagonal values d_i are positive if W is positive definite. Indeed for any vector X , $X^T \cdot \sqrt{C} \cdot W \cdot \sqrt{C} \cdot X = Y^T \cdot W \cdot Y$ with $Y = \sqrt{C} \cdot X$. Eventually one obtains $C \cdot W = (\sqrt{C} \cdot R) \cdot d \cdot (\sqrt{C} \cdot R)^{-1}$ and

$$N^{\text{eff}} = \frac{(\text{Tr}(C \cdot W))^2}{\text{Tr}(C \cdot W \cdot C \cdot W)} = \frac{(\sum_i d_i)^2}{\sum_i d_i^2}. \quad (\text{B9})$$

Multiplying all d_i by the same factor does not alter this ratio. However if we have two different eigenvalues d_I and d_J with $d_I > d_J$ then for $\varepsilon > 0$, the shifts $d_I \rightarrow d_I + \varepsilon$ and $d_J \rightarrow d_J - \varepsilon$ reduce the denominator. Therefore for a given fixed $\sum_i d_i$ the minimum of N^{eff} is reached when one eigenvalue is maximal and all other vanish and in that case $N^{\text{eff}} = 1$. For instance we could choose $W_{ij} = \delta_i^1 \delta_j^1$ (plus negligible positive values for all other diagonal entries to make sure it is positive definite), such that $\text{SNR} \simeq 1/\sqrt{2}$. It amounts to throwing all points except the first one. This is essentially what is achieved when considering the naive time-average with a very red spectrum for which N^{eff} is only slightly larger than unity.

Finally, redshift Fourier components are complex valued, but satisfy the condition (60). Let us build an estimator for complex-valued random variables such that $x_i^* = x_{-i}$ and $\langle x_i x_j^* \rangle = AC_{ij}$. The hermitian correlation satisfies $C_{-j-i} = C_{ji}^* = C_{ij}$. The general quadratic estimator of A is of the form

$$Q = x_j^* W_{ji} x_i, \quad (\text{B10})$$

with $W_{-j-i} = W_{ji}^* = W_{ij}$. Repeating the previous analysis, now (B3) becomes a hermitian product on hermitian matrices which is again positive definite because C is. The average and variance of the estimator are then similar to (B4), hence we find that the optimal estimator is (B8) also in this context.

Appendix C: Quadratic estimator for imperfect experiments

We aim at finding an estimator which maximizes the likelihood (166) where the set of variables are the $\tilde{z}^{\text{pos}}(\mathbf{n}_i, f_p)$. We use the multi-indices of the type $a = (\mathbf{n}_i, f_p)$, $b = (\mathbf{n}_j, f_q)$, or in compact form $a = (i, p)$, $b = (j, q)$, $c = (k, r)$, $d = (l, s)$, such that in (166) $z_a = \tilde{z}^{\text{pos}}(\mathbf{n}_i, f_p)$. In that case the correlation in (166), which we separate into signal and noise, is

$$C_{ab} \equiv \langle z_a z_b^* \rangle_{\mathcal{A}} = C_{ab}^S + C_{ab}^N \\ \equiv \mu(\beta_{ij}) \mathcal{P}^{\text{pos}}(f_p, f_q) + \delta_{ij} (4\pi)^{-1} \mathcal{N}_i^{\text{pos}}(f_p, f_q), \quad (\text{C1})$$

where $\mathcal{N}_i^{\text{pos}}(f_p, f_q)$ is defined from $\mathcal{N}_i(f)$ in the same way as the signal part was defined using (129). The correlation of the

signal is further expanded as

$$C_{ab}^S = \sum_{\ell} c_{\ell} C_{ab}^{S,\ell}, \quad C_{ab}^{S,\ell} \equiv \mu_{\ell}(\beta_{ij}) \mathcal{P}^{\text{pos}}(f_p, f_q), \quad (\text{C2})$$

where we used the multipolar component of the HD curve $\mu_{\ell}(\beta)$ defined in (56).

Following [65], a simple quadratic estimator can be devised to update the values of the c_{ℓ} of (133) in an iterative procedure which aims at maximizing the likelihood. Let us first define a hermitian inner product in the space of hermitian matrices using the (inverse) total covariance matrix in (B3). With the current notation, this is

$$\{M_{ab}; N_{cd}\}_{C^{-1}} \equiv \text{Tr}(M \cdot C^{-1} \cdot N \cdot C^{-1}) = M_{ab} C_{ca}^{-1} C_{bd}^{-1} N_{dc}, \quad (\text{C3})$$

with implicit summation in the last expression. The Fisher matrix for the c_{ℓ} parameters is

$$F_{\ell\ell'} \equiv \frac{1}{2} \text{Tr} [C^{-1} \cdot \partial_{c_{\ell}} C \cdot C^{-1} \cdot \partial_{c_{\ell'}} C] = \frac{1}{2} \{C_{ab}^{S,\ell}; C_{cd}^{S,\ell'}\}_{C^{-1}}. \quad (\text{C4})$$

The quadratic estimator which updates c_{ℓ} , requires to inverse the Fisher matrix, see e.g. (2.18) of [65], and is

$$c_{\ell}^e = F_{\ell\ell'}^{-1} X_{\ell'}, \quad X_{\ell} \equiv \frac{1}{2} \{C_{ab}^{S,\ell}; z_c z_d^* - C_{cd}^N\}_{C^{-1}}. \quad (\text{C5})$$

From (C1) and the expansion (C2) it is immediate to obtain $\langle X_{\ell} \rangle_{\mathcal{A}} = F_{\ell\ell'}^{-1} c_{\ell'}$, hence the estimator (C5) is unbiased, that is if the guess was correct, there is nothing to update. This can also be seen by rewriting $c_{\ell}^e - c_{\ell} = \frac{1}{2} F_{\ell\ell'}^{-1} \{C_{ab}^{S,\ell}; z_c z_d^* - C_{cd}\}_{C^{-1}}$ which vanishes if the initial guess is already unbiased, as can be seen using (C1). This estimator is the counterpart in multipole space of the estimator in angular separation space built in [39]. While submitting the present article, a similar approach based on the Fisher matrix inversion was developed in [66], where the sums are only performed on pairs of pulsars such that there is no need for an offset in the definition of X_{ℓ} . The estimator is obtained by approximating that the quadratic observables $z_a z_b$ follow a Gaussian distribution with covariance $C_{ab;cd} = C_{ac} C_{bd} + C_{ad} C_{bc}$ and thus it uses a maximization of the associated χ^2 . On the other hand, the method [65] that we follow is directly based on the maximization of the Gaussian distributions for the z_a with correlation C_{ab} . Contrary to the former approach, in the latter one the estimator must be computed iteratively until convergence is reached so as to take into account that the Fisher matrix itself depends on the estimated parameters.

The SNR can be evaluated with the Fisher matrix. Indeed, since $\langle c_{\ell}^e c_{\ell'}^e \rangle_{\mathcal{A}} = F_{\ell\ell'}^{-1}$, the SNR for c_{ℓ} is $\text{SNR}_{\ell}^2 = c_{\ell}^2 / F_{\ell\ell}^{-1}$. Sadly it is in general impossible to find analytic forms for the Fisher matrix and thus for the SNR, since when there is noise (C1) does not factorize into a frequency part and an angular part. Hence let us first lower our ambitions and consider the case where there is no instrumental noise but shot noise due to the finite pulsar number is present, i.e. the imperfection comes from the finite number of pulsars. We assume that there are $N_f = 2T f_{\text{max}}$ frequencies which are measured perfectly. The inverse of the correlation factorizes as

$$C_{ab}^{-1} = [\mu(\beta_{ij})]^{-1} \mathcal{P}_{pq}^{\text{pos},-1}, \quad (\text{C6})$$

where it is understood that $[\mu(\beta_{ij})]^{-1}$ means the inverse of the matrix $\mu(\beta_{ij})$. Hence the Fisher matrix reduces to

$$F_{\ell\ell'} = \frac{N_f}{2} \sum_{ijkl} \mu_\ell(\beta_{ij}) [\mu(\beta_{ik})]^{-1} [\mu(\beta_{jl})]^{-1} \mu_{\ell'}(\beta_{kl}), \quad (\text{C7})$$

where the frequency-dependent part has disappeared into a trace of the identity of the frequency space, which yields a factor N_f . Again we must accept that there is no general simplification for a generic given set of pulsar directions. We would like to average this expression over the possible realizations of the pulsar directions to assess the noise introduced by the randomness of a finite number of pulsars, but there is to our knowledge no simple analytic method. At best, we could use the approximation $[\mu(\beta_{ij})]^{-1} \simeq (4\pi/N_P)^2 \mu^{-1}(\beta_{ij})$, where the latter is the inverse in the limit of infinite pulsar number, obtained by inversion in the $\ell \geq 2$ multipolar subspace (see section IV.C of [31])

$$\mu^{-1}(\beta_{ij}) = \sum_{\ell \geq 2} \frac{1}{C_\ell^{\text{HD}}} \frac{(2\ell+1)}{4\pi} P_\ell(\cos \beta_{ij}). \quad (\text{C8})$$

In that case the Fisher matrix simplifies drastically, as sums over pulsar directions are replaced by angular integrations, and with (A13) and (A14) we recover

$$F_{\ell\ell'} = \delta_{\ell\ell'} \frac{N_f}{2} (2\ell+1). \quad (\text{C9})$$

Having neglected noise, and assumed an infinite number of pulsars, our imperfect experiment is now perfect hence we are not surprised to recover (159). The difference between (C7) and (C9) is precisely due to the finite number of pulsars. However we know no simple method to assess the magnitude of this difference, hence in section VII C 2 we take an alternative approach based on the estimators (146) and their variance under the realizations of the pulsar directions to estimate this pulsar shot noise.

Finally, if we want to constrain the global amplitude $\hat{\rho}_\star$ with an imperfect experiment, the Fisher matrix is just a number

$$F = \frac{1}{2} \hat{\rho}_\star^{-2} \{C_{ab}^S, C_{cd}^S\}_{C^{-1}} \quad (\text{C10})$$

and the estimator is simply

$$\hat{\rho}_\star^e \equiv F^{-1} X, \quad X \equiv \frac{1}{2} \hat{\rho}_\star^{-1} \{C_{ab}^S, z_c z_d^* - C_{cd}^N\}_{C^{-1}}. \quad (\text{C11})$$

Up to the constant shift due to noise, this is exactly the estimator (7.2) of [31]. For a perfect experiment $C_{ab}^S = C_{ab}$, $F = 1/2 \hat{\rho}_\star^{-2} N_f N_P$, and this reduces to $\hat{\rho}_\star^e = \hat{\rho}_\star / (N_P N_f) \sum_{ab} z_a z_b^* C_{ba}^{-1}$, hence to the estimator (164) when using (C6).

Appendix D: Pulsar shot noise

Let us start from the estimator (146). We want to compute its covariance when averaging over both the statistics of $z(\mathbf{n}_i)$

which is a statistics of the $z_{\ell m}$, and over the randomness of the position of the pulsars (indicated with $\langle \rangle_P$ in that case). The underlying redshift function is decomposed as

$$z(\mathbf{n}) = \sum_{L \geq 2} \sum_{M=-L}^L z_{LM} Y_L^M(\mathbf{n}), \quad (\text{D1})$$

and the multipole ensemble statistics is of the form

$$\langle z_{LM} z_{L'M'}^* \rangle_{\mathcal{A}} = \delta_L^{L'} \delta_M^{M'} C_\ell. \quad (\text{D2})$$

Now we shall also need

$$\langle Y_\ell^m(\mathbf{n}_i) \rangle_P = \int \frac{d^2 \mathbf{n}}{4\pi} Y_\ell^m(\mathbf{n}) = \delta_0^\ell \delta_0^m \frac{1}{\sqrt{4\pi}}, \quad (\text{D3})$$

$$\langle Y_\ell^m(\mathbf{n}_i) Y_{\ell'}^{m'*}(\mathbf{n}_i) \rangle_P = \int \frac{d^2 \mathbf{n}}{4\pi} Y_\ell^m(\mathbf{n}) Y_{\ell'}^{m'*}(\mathbf{n}) = \frac{\delta_\ell^\ell \delta_{m'}^m}{4\pi}, \quad (\text{D4})$$

whereas for two distinct pulsars

$$\langle Y_\ell^m(\mathbf{n}_i) Y_{\ell'}^{m'*}(\mathbf{n}_j) \rangle_P = \frac{\delta_0^\ell \delta_0^m \delta_0^{\ell'} \delta_0^{m'}}{4\pi}. \quad (\text{D5})$$

Let us evaluate now the statistics of the estimator (146) under the distribution of pulsar positions. The one-point function is simply

$$\langle z_{\ell m}^N \rangle_P = z_{\ell m}. \quad (\text{D6})$$

In order to evaluate the covariance, we need to distinguish when we have the same pulsar or not hence this brings the two contributions

$$\begin{aligned} \langle z_{\ell m}^N z_{\ell' m'}^{N*} \rangle_P &= \frac{N-1}{N} z_{\ell m} z_{\ell' m'}^* + \frac{4\pi}{N} \sum_{LL'MM'} z_{LM} z_{L'M'}^* \int d^2 \mathbf{n} Y_\ell^{m*} Y_{L'}^{M'*} Y_{\ell'}^{m'} Y_L^M. \end{aligned} \quad (\text{D7})$$

The first term comes from the $N(N-1)$ terms for which $\mathbf{n}_i \neq \mathbf{n}_j$ and the second term from the N other terms.

Then we take the GW ensemble statistics and get

$$\langle \langle z_{\ell m}^N \rangle_P \rangle_{\mathcal{A}} = \langle z_{\ell m} \rangle_{\mathcal{A}} = 0. \quad (\text{D8})$$

For the covariance we have a product of four spherical harmonics. We can pair them using the addition of spherical harmonics which is inferred from (A17), (A13) and (A10)

$$\begin{aligned} Y_\ell^m Y_{\ell'}^{m'} &= \sum_{LM} (-1)^M \sqrt{\frac{(2\ell+1)(2\ell'+1)(2L+1)}{4\pi}} \\ &\quad \begin{pmatrix} \ell & \ell' & L \\ m & m' & -M \end{pmatrix} \begin{pmatrix} \ell & \ell' & L \\ 0 & 0 & 0 \end{pmatrix} Y_L^M. \end{aligned} \quad (\text{D9})$$

We pair Y_ℓ^{m*} with $Y_{L'}^{m'*}$ to give $Y_{L'}^{m'*}$, and we pair $Y_{\ell'}^{m'}$ with Y_L^M to give Y_L^M . Using the GW ensemble average brings $\delta_L^{L'} \delta_M^{M'}$ and using the orthogonality property of spherical harmonics

brings $\delta_{\mathcal{L}'}^{\mathcal{L}} \delta_{m'}^m$. Putting all this together we obtain

$$\begin{aligned} \langle \langle z_{\ell m}^N z_{\ell' m'}^{N*} \rangle_{\mathcal{P}} \rangle_{\mathcal{A}} &= \frac{N-1}{N} \delta_{\ell'}^{\ell} \delta_{m'}^m C_{\ell} \\ &+ \frac{1}{N} \sum_{LM} \sum_{\mathcal{L}m} C_L (2L+1) (2\mathcal{L}+1) \sqrt{(2\ell+1)(2\ell'+1)} \\ &\begin{pmatrix} L & \ell' & \mathcal{L} \\ M & m' & -m \end{pmatrix} \begin{pmatrix} L & \ell' & \mathcal{L} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} L & \ell & \mathcal{L} \\ M & m & -m \end{pmatrix} \begin{pmatrix} L & \ell & \mathcal{L} \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (\text{D10})$$

With the orthonormality (A17), we get

$$\begin{aligned} \langle \langle z_{\ell m}^N z_{\ell' m'}^{N*} \rangle_{\mathcal{P}} \rangle_{\mathcal{A}} &= \frac{N-1}{N} \delta_{\ell'}^{\ell} \delta_{m'}^m C_{\ell} \\ &+ \frac{1}{N} \sum_{\mathcal{L}} C_L (2L+1) (2\mathcal{L}+1) \delta_{\ell'}^{\ell} \delta_{m'}^m \begin{pmatrix} L & \ell & \mathcal{L} \\ 0 & 0 & 0 \end{pmatrix}^2. \end{aligned} \quad (\text{D11})$$

Using finally

$$\sum_L (2L+1) \begin{pmatrix} \ell_1 & \ell_2 & L \\ 0 & 0 & 0 \end{pmatrix}^2 = 1, \quad (\text{D12})$$

which is a particular case of (A18), this reduces to

$$\langle \langle z_{\ell m}^N z_{\ell' m'}^{N*} \rangle_{\mathcal{P}} \rangle_{\mathcal{A}} = \delta_{\ell'}^{\ell} \delta_{m'}^m (C_{\ell} + N_{\ell}^{\mathcal{P}}), \quad (\text{D13})$$

where the pulsar shot noise is

$$N_{\ell}^{\mathcal{P}} = -\frac{1}{N} C_{\ell} + \frac{1}{N} \sum_L (2L+1) C_L. \quad (\text{D14})$$

Finally if we repeat the analysis of this section starting from (149), that is with general weights $w_i \neq 1/N$, then the factors $1/N$ in the previous expression are replaced by $\sum_{i=1}^N w_i^2 \geq 1/N$.

Appendix E: Hellings-Downs correlation in angular space

We summarize how (57) is obtained with a method adapted from Sec III.E of [15]. Let us start by decomposing $\ln x$, where $x = (1 - \cos \beta)/2 = \sin^2(\beta/2)$, on Legendre polynomials, that is we seek $\ln x = \sum_{\ell} a_{\ell} P_{\ell}(\beta)$ and the coefficients are given from the orthogonality of Legendre polynomials as

$$\begin{aligned} a_{\ell} &= \frac{2\ell+1}{2} \int_{-1}^1 \ln x P_{\ell}(\cos \beta) d\cos \beta \\ &= (2\ell+1) \int_0^1 \ln x P_{\ell}(1-2x) dx. \end{aligned} \quad (\text{E1})$$

Using

$$P_{\ell}(1-2x) = \frac{1}{\ell!} \frac{d^{\ell}}{dx^{\ell}} [x^{\ell} (1-x)^{\ell}], \quad (\text{E2})$$

and

$$\frac{d^n}{dx^n} \ln x = \frac{(-1)^{n-1} (n-1)!}{x^n}, \quad (\text{E3})$$

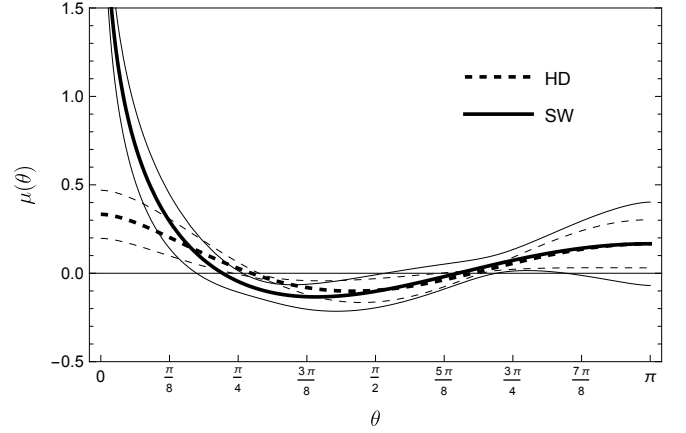


FIG. 10. Hellings-Downs correlation (57) in thick dashed line and Sachs-Wolfe plateau correlation from scalar perturbations (E6) in thick continuous line, the latter being normalized to the same value as the former when $\theta = \pi$. The thin lines correspond to the one standard deviation contour due to cosmic variance for $N_f^{\text{eff}} = 1$, obtained from (95). The contours will of course shrink when the value of N_f^{eff} is increased by improving observation time and pulsar monitoring cadence.

the coefficients are easily evaluated with successive integration by parts and for $\ell \geq 1$ we get

$$a_{\ell} = -\frac{2\ell+1}{\ell(\ell+1)}, \quad (\text{E4})$$

whereas $a_0 = -1$. Therefore we find

$$-\ln x = -\ln \sin^2(\theta/2) = 1 + \sum_{\ell \geq 1} \frac{2\ell+1}{\ell(\ell+1)} P_{\ell}(\cos \theta). \quad (\text{E5})$$

The Sachs-Wolfe plateau in the CMB temperature fluctuations for a scale-invariant spectrum of scalar perturbations behaves as $C_{\ell}^{\text{SW}} \propto 1/[\ell(\ell+1)]$. Hence once translated with (E5) as a correlation for angular separation where the dipole is conventionally removed, the Sachs-Wolfe plateau correlation function is

$$\begin{aligned} \mu^{\text{SW}}(\theta) &\equiv \sum_{\ell \geq 2} \frac{(2\ell+1)}{4\pi} C_{\ell}^{\text{SW}} P_{\ell}(\theta) \\ &\propto -1 - \ln \sin^2(\theta/2) - \frac{3}{2} \cos \theta. \end{aligned} \quad (\text{E6})$$

This captures the shape of the CMB correlations in angular separation space (see e.g. Fig. 2 of [48]), and it diverges for $\theta \rightarrow 0$ contrary to the Hellings-Downs curve, see Fig. 10.

Finally, we need only to use

$$\cos \theta P_{\ell}(\cos \theta) = \frac{\ell+1}{2\ell+1} P_{\ell+1}(\cos \theta) + \frac{\ell}{2\ell+1} P_{\ell-1}(\cos \theta),$$

in combination with (E5), and rearranging the summations so as to sum only on the P_{ℓ} , we obtain

$$x \ln x = \sum_{\ell \geq 2} (2\ell+1) \frac{(\ell-2)!}{(\ell+2)!} P_{\ell}(\cos \theta) - \frac{1}{3} + \frac{x}{6}, \quad (\text{E7})$$

which can be used to find the Hellings-Downs correlation function (57).

-
- [1] G. Agazie *et al.* (NANOGrav), *Astrophys. J. Lett.* **951**, L8 (2023), arXiv:2306.16213 [astro-ph.HE].
- [2] D. J. Reardon *et al.*, *Astrophys. J. Lett.* **951**, L6 (2023), arXiv:2306.16215 [astro-ph.HE].
- [3] J. Antoniadis *et al.* (EPTA), *Astron. Astrophys.* **678**, A48 (2023), arXiv:2306.16224 [astro-ph.HE].
- [4] H. Xu *et al.*, *Res. Astron. Astrophys.* **23**, 075024 (2023), arXiv:2306.16216 [astro-ph.HE].
- [5] J. D. Romano and N. J. Cornish, (2016), arXiv:1608.06889 [gr-qc].
- [6] F. A. Jenet and J. D. Romano, *Am. J. Phys.* **83**, 635 (2015), arXiv:1412.1142 [gr-qc].
- [7] J. D. Romano and B. Allen, (2023), arXiv:2308.05847 [gr-qc].
- [8] M. Anholm, S. Ballmer, J. D. E. Creighton, L. R. Price, and X. Siemens, *Phys. Rev. D* **79**, 084030 (2009), arXiv:0809.0701 [gr-qc].
- [9] C. M. F. Mingarelli, T. Sidery, I. Mandel, and A. Vecchio, *Phys. Rev. D* **88**, 062005 (2013), arXiv:1306.5394 [astro-ph.HE].
- [10] P. F. Depta, V. Domcke, G. Franciolini, and M. Pieroni, (2024), arXiv:2407.14460 [astro-ph.CO].
- [11] F. Semenzato, J. A. Casey-Clyde, C. M. F. Mingarelli, A. Racanelli, N. Bellomo, N. Bartolo, and D. Bertacca, (2024), arXiv:2411.00532 [astro-ph.CO].
- [12] R. C. Bernardo, G.-C. Liu, and K.-W. Ng, *JCAP* **04**, 034 (2024), arXiv:2312.03383 [gr-qc].
- [13] G. Cusin, R. Durrer, and P. G. Ferreira, *Phys. Rev. D* **99**, 023534 (2019), arXiv:1807.10620 [astro-ph.CO].
- [14] K. Grunthal *et al.*, (2024), 10.1093/mnras/stae2573, arXiv:2412.01214 [astro-ph.HE].
- [15] J. Gair, J. D. Romano, S. Taylor, and C. M. F. Mingarelli, *Phys. Rev. D* **90**, 082001 (2014), arXiv:1406.4664 [gr-qc].
- [16] E. Roebber and G. Holder, *Astrophys. J.* **835**, 21 (2017), arXiv:1609.06758 [astro-ph.CO].
- [17] W. Qin, K. K. Boddy, M. Kamionkowski, and L. Dai, *Phys. Rev. D* **99**, 063002 (2019), arXiv:1810.02369 [astro-ph.CO].
- [18] S. C. Hotinli, M. Kamionkowski, and A. H. Jaffe, *Open J. Astrophys.* **2**, 8 (2019), arXiv:1904.05348 [astro-ph.CO].
- [19] J. Nay, K. K. Boddy, T. L. Smith, and C. M. F. Mingarelli, *Phys. Rev. D* **110**, 044062 (2024), arXiv:2306.06168 [gr-qc].
- [20] R. C. Bernardo and K.-W. Ng, *JCAP* **11**, 046 (2022), arXiv:2209.14834 [gr-qc].
- [21] R. C. Bernardo and K.-W. Ng, *Phys. Rev. D* **107**, 044007 (2023), arXiv:2208.12538 [gr-qc].
- [22] B. Allen, (2024), arXiv:2404.05677 [gr-qc].
- [23] R. C. Bernardo and K.-W. Ng, (2024), arXiv:2409.07955 [astro-ph.CO].
- [24] G. Agazie *et al.*, (2024), arXiv:2411.13472 [astro-ph.HE].
- [25] X. Siemens, J. Ellis, F. Jenet, and J. D. Romano, *Class. Quant. Grav.* **30**, 224015 (2013), arXiv:1305.3196 [astro-ph.IM].
- [26] N. Pol, S. R. Taylor, and J. D. Romano, *Astrophys. J.* **940**, 173 (2022), arXiv:2206.09936 [astro-ph.HE].
- [27] A. Sesana, A. Vecchio, and C. N. Colacino, *MNRAS* **390**, 192 (2008), arXiv:0804.4476.
- [28] R. Blandford and S. A. Teukolsky, *Astrophys. J.* **205**, 580 (1976).
- [29] R. van Haasteren, Y. Levin, P. McDonald, and T. Lu, *Mon. Not. Roy. Astron. Soc.* **395**, 1005 (2009), arXiv:0809.0791 [astro-ph].
- [30] J. S. Hazboun, J. D. Romano, and T. L. Smith, *Phys. Rev. D* **100**, 104028 (2019), arXiv:1907.04341 [gr-qc].
- [31] B. Allen and J. D. Romano, *Phys. Rev. D* **108**, 043026 (2023), arXiv:2208.07230 [gr-qc].
- [32] B. Allen, D. Agarwal, J. D. Romano, and S. Valtolina, *Phys. Rev. D* **110**, 123507 (2024), arXiv:2406.16031 [gr-qc].
- [33] B. Allen, *Frascati Phys. Ser.* **74**, 65 (2022).
- [34] B. Allen, *Phys. Rev. D* **107**, 043018 (2023), arXiv:2205.05637 [gr-qc].
- [35] B. Allen and S. Valtolina, (2024), arXiv:2401.14329 [gr-qc].
- [36] N. Grimm, M. Pijenburg, G. Cusin, and C. Bonvin, (2024), arXiv:2411.08744 [gr-qc].
- [37] N. Grimm, M. Pijenburg, G. Cusin, and C. Bonvin, (2024), arXiv:2404.05670 [astro-ph.CO].
- [38] D. Agarwal and J. D. Romano, *Phys. Rev. D* **110**, 043044 (2024), arXiv:2404.08574 [gr-qc].
- [39] B. Allen and J. D. Romano, (2024), arXiv:2407.10968 [gr-qc].
- [40] W. Hu and M. J. White, *Phys. Rev. D* **56**, 596 (1997), arXiv:astro-ph/9702170.
- [41] C. Pitrou and T. S. Pereira, *Phys. Rev. D* **100**, 123535 (2019), arXiv:1909.13687 [gr-qc].
- [42] C. Fidler and C. Pitrou, *JCAP* **06**, 013 (2017), arXiv:1701.08844 [cond-mat.stat-mech].
- [43] C. Pitrou, *Astropart. Phys.* **125**, 102494 (2021), arXiv:1902.09456 [astro-ph.CO].
- [44] M. Maggiore, *Gravitational Waves. Vol. 2: Astrophysics and Cosmology* (Oxford University Press, 2018).
- [45] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, seventh ed. (Elsevier/Academic Press, Amsterdam, 2007).
- [46] L. Ackerman, S. M. Carroll, and M. B. Wise, *Phys. Rev. D* **75**, 083502 (2007), [Erratum: *Phys. Rev. D* **80**, 069901 (2009)], arXiv:astro-ph/0701357.
- [47] R. Kothari, S. Ghosh, P. K. Rath, G. Kashyap, and P. Jain, *Mon. Not. Roy. Astron. Soc.* **460**, 1577 (2016), arXiv:1503.08997 [astro-ph.CO].
- [48] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, *Adv. Astron.* **2010**, 847541 (2010), arXiv:1004.5602 [astro-ph.CO].
- [49] L. Lentati, P. Alexander, M. P. Hobson, S. Taylor, J. Gair, S. T. Balan, and R. van Haasteren, *Phys. Rev. D* **87**, 104021 (2013), arXiv:1210.3578 [astro-ph.IM].
- [50] P. Peter and J.-P. Uzan, *Primordial cosmology*, Oxford graduate texts (Oxford Univ. Press, Oxford, 2009).
- [51] R. Durrer, *The Cosmic Microwave Background* (Cambridge University Press, 2008).
- [52] N. Aghanim *et al.* (Planck), *Astron. Astrophys.* **641**, A6 (2020), [Erratum: *Astron. Astrophys.* **652**, C4 (2021)], arXiv:1807.06209 [astro-ph.CO].
- [53] S. J. Chamberlin, J. D. E. Creighton, X. Siemens, P. Demorest, J. Ellis, L. R. Price, and J. D. Romano, *Phys. Rev. D* **91**, 044048 (2015), arXiv:1410.8256 [astro-ph.IM].
- [54] J. Keiner and D. Potts, *Math. Comput.* **77**, 397 (2008).
- [55] K. P. Drake and G. B. Wright, *J. Comput. Phys.* **416**, 109544 (2020), arXiv:1904.10514 [math.NA].
- [56] G. Agazie *et al.* (NANOGrav), *Astrophys. J. Lett.* **951**, L9 (2023), arXiv:2306.16217 [astro-ph.HE].
- [57] Y. Ali-Haïmoud, T. L. Smith, and C. M. F. Mingarelli, *Phys. Rev. D* **102**, 122005 (2020), arXiv:2006.14570 [gr-qc].

- [58] Y. Ali-Haïmoud, T. L. Smith, and C. M. F. Mingarelli, *Phys. Rev. D* **103**, 042009 (2021), [arXiv:2010.13958 \[gr-qc\]](#).
- [59] G. Cusin, C. Pitrou, and J.-P. Uzan, *Phys. Rev. D* **96**, 103019 (2017), [arXiv:1704.06184 \[astro-ph.CO\]](#).
- [60] T. Namikawa, S. Saga, D. Yamauchi, and A. Taruya, *Phys. Rev. D* **100**, 021303 (2019), [arXiv:1904.02115 \[astro-ph.CO\]](#).
- [61] K.-W. Ng, *Phys. Rev. D* **106**, 043505 (2022), [arXiv:2106.12843 \[astro-ph.CO\]](#).
- [62] J. Lesgourgues, *arXiv e-prints*, [arXiv:1104.2932](#) (2011), [arXiv:1104.2932 \[astro-ph.IM\]](#).
- [63] D. Blas, J. Lesgourgues, and T. Tram, *JCAP* **2011**, 034 (2011), [arXiv:1104.2933 \[astro-ph.CO\]](#).
- [64] A. Challinor and A. Lewis, *Phys. Rev. D* **71**, 103010 (2005), [arXiv:astro-ph/0502425](#).
- [65] J. R. Bond, A. H. Jaffe, and L. Knox, *Phys. Rev. D* **57**, 2117 (1998), [arXiv:astro-ph/9708203](#).
- [66] B. Allen and J. D. Romano, (2024), [arXiv:2412.14852 \[gr-qc\]](#).