

Pre-Merger Localization of Gravitational-Wave Standard Sirens With LISA: Harmonic Mode Decomposition

Bence Kocsis,^{1,2} Zoltán Haiman,³ Kristen Menou,³ and Zsolt Frei¹

¹*Institute of Physics, Eötvös University, Pázmány P. s. 1/A, 1117 Budapest, Hungary*

²*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138*

³*Department of Astronomy, Columbia University, 550 West 120th Street, New York, NY 10027*

The continuous improvement in localization errors (sky position and distance) in real time as LISA observes the gradual inspiral of a supermassive black hole (SMBH) binary can be of great help in identifying any prompt electromagnetic counterpart associated with the merger. We develop a new method, based on a Fourier decomposition of the time-dependent, LISA-modulated gravitational-wave signal, to study this intricate problem. The method is faster than standard Monte Carlo simulations by orders of magnitude. By surveying the parameter space of potential LISA sources, we find that counterparts to SMBH binary mergers with total mass $M \sim 10^5 - 10^7 M_\odot$ and redshifts $z \lesssim 3$ can be localized to within the field of view of astronomical instruments ($\sim \text{deg}^2$) typically hours to weeks prior to coalescence. This will allow a triggered search for variable electromagnetic counterparts as the merger proceeds, as well as monitoring of the most energetic coalescence phase. A rich set of astrophysical and cosmological applications would emerge from the identification of electromagnetic counterparts to these gravitational-wave standard sirens.

PACS numbers:

I. INTRODUCTION

One of the key objectives of the planned, low-frequency gravitational-wave (GW) detector LISA (Laser Interferometric Space Antenna) is the detection of supermassive black hole (SMBH) binary mergers at cosmological distances. The observation of these chirping GW sources would deepen our understanding of (i) general relativity, e.g. by offering unique tests of spacetime physics in the vicinity of SMBHs [1, 2, 3, 4, 5], (ii) cosmology, by providing additional constraints on the luminosity distance–redshift relation [6, 7, 8], (iii) large-scale structure, by indirectly constraining hierarchical structure formation scenarios [9, 10, 11, 13], and (iv) black hole astrophysics, e.g. by allowing accurate determinations of Eddington ratios, and other attributes of black hole accretion, in systems with SMBH mass and spin known independently, from the GW measurements [14, 15, 16].

From a purely astronomical point of view, one of the most attractive features of the LISA mission design is the possibility to constrain the 3-dimensional location (i.e. sky position and distance) of GW inspiral sources to within a small enough volume that the identification of potential electromagnetic (EM) counterparts to SMBH merger events can be contemplated seriously. Indeed, the accuracy of such LISA localizations at merger are encouraging, with an error volume $\delta\Omega \times \delta z = 0.3 \text{ deg}^2 \times 0.1$ for SMBH masses $m_1 = m_2 = 10^6 M_\odot$ at $z = 1$, for instance [17]. In Ref. [15], we have shown that this accuracy may be sufficient to allow a unique identification of the bright quasar activity that may be associated with any such SMBH merger.

Another possibility, examined here in detail, is to monitor the sky for EM counterparts in real time, as the SMBH inspiral proceeds. This is arguably one of the most efficient ways to identify reliably (prompt) EM counterparts to SMBH merger events, since the exact nature of such counterparts is a priori unknown. Using the GW inspiral signal accumulated up to some look–back time, t_f , preceding the final coalescence, one

already has a partial knowledge of where the source of GWs is located on the sky. Since the sky position is deduced primarily from the detector’s motion around the Sun, one anticipates that angular positioning uncertainties will not change too dramatically during the last few days before merger, so that a targeted EM observation of the final stages of inspiral may be a feasible task. Here, we present an in-depth study of the potential for such pre-merger localizations with LISA, while we discuss various astrophysical concepts and observational strategies for EM counterpart identifications in a companion work [18].

The main purpose of the present analysis is thus to determine the accuracy of SMBH inspiral localizations with LISA, as a function of look–back time, t_f , prior to merger. The LISA detector is not uniformly sensitive to sources with different sky positions and angular momentum orientations. Results will thus generally depend on the fiducial values of these angles. Our first objective is to calculate the time-dependence of distributions of localization errors, for randomly oriented sources, over a large range of values for the SMBH masses and source redshift. A second objective of our analysis is to estimate source parameter dependencies for these distributions of localization errors, i.e. how the 3-dimensional (sky position and distance) localization error distributions depend on the fiducial sky position of GW sources. This is useful to understand which regions of the sky may be best suited for the identification of EM counterparts to SMBH merger events. To the best of our knowledge, this angle dependence has not been explored in detail before, not even in terms of final errors at ISCO (i.e. at $t_f = t_{\text{isco}}$, when using the complete inspiral data-stream, up to the innermost stable circular orbit, or ISCO).

Parameter estimation uncertainties for LISA inspirals have been considered previously, under a variety of approximations [5, 7, 8, 13, 17, 19, 20, 21, 22]. These studies differ in the levels of approximation adopted for the GW waveform, using various orders of the post-Newtonian expansion. The LISA signal output for these waveforms are obtained through a lin-

arXiv:astro-ph/0701629v2 17 Jun 2007

ear combination of the two GW polarizations, $h_+(t)$ and $h_\times(t)$, with the beam pattern functions, F_+ and F_\times . The beam patterns define the detector sensitivity for the two polarizations. They are determined by the angles describing the instantaneous orientation of the LISA constellation relative to the GW polarizations. As the LISA detector constellation orbits the Sun, with a one year period, F_+ and F_\times are slowly changing in time and this introduces an additional time dependence in the LISA signal. As first shown by Cutler [19], the source sky position can be determined with LISA using this modulation. In the formalism given by Cutler [19], this modulation couples time and angular dependencies in a complicated way, making the estimation of localization errors numerically costly for a large set of SMBH binary random orientations and parameters.

Using a different approach, Cornish & Rubbo [23] have derived the orbital modulation in a much simpler form, in which the angular parameter dependence and the time dependence can be decoupled. Here, starting directly from the original Cutler [19] expression, we give an independent derivation of the Cornish & Rubbo [23] formula and write it in an equivalent form, from which decoupling is more evident. We do this by expanding the LISA response function into a discrete Fourier sum of harmonics of the fundamental frequency of LISA’s orbital motion, $f_\oplus = 1 \text{ yr}^{-1}$. Since LISA’s orbit does not include high frequency features, we expect this sum to be quickly convergent. In fact, it is clear from the Cornish & Rubbo [23] result that the expansion terminates at $4f_\oplus$ and that there are no higher order harmonics due to the detector’s motion. The series coefficients in the expansion are independent of time and only depend on the relative angles at ISCO. We then develop a Fisher matrix formalism in which parameter error distributions can be mapped independently of time, while the time dependence can be computed independently of the specific SMBH binary orbital elements. A Monte Carlo simulation for random binary orientations then becomes a simple linear combination, without any integral evaluations. This greatly reduces the numerical cost of estimating parameter uncertainty distributions, even at fixed observation time (e.g. to map distributions of errors at ISCO). We use this numerical cost advantage

1. to map the distribution of localization errors for the full three dimensional grid of SMBH total mass ($M = 10^5 - 10^8 M_\odot$), redshift ($z = 0.1 - 7$) and arbitrary look-back time (t_f) before merger,
2. to study how source localization error distributions vary systematically with sky position, and
3. to discuss implications, in terms of advance warning times, for prompt electromagnetic counterpart searches with large field-of-view astronomical instruments.

We call this new approach the harmonic mode decomposition (HMD). The method verifies that the amplitude modulation, which is restricted to frequencies less than $4f_\oplus = 1.3 \times 10^{-7} \text{ Hz}$, is indeed a very slow modulation when compared to the GW frequency of LISA SMBH inspirals (0.03 mHz–1 Hz). One plausibly expects that physical parameters which

determine the amplitude modulation (like the source sky position and orbital inclination relative to the detector) can be estimated independently of the parameters which determine the GW frequency (like masses, orbital phase, time to ISCO). In the HMD method, the two sets of parameters are naturally separated and can be estimated independently. In particular, parameters related to the modulation can essentially be determined on a background of GW-cycle averaged signal. In the present work, we compute LISA inspiral localization errors with the approximation that high frequency signal parameters have strictly no cross-correlations with parameters related to the slow orbital modulation. In addition to the numerical advantages mentioned above, the HMD formalism offers a clear interpretation of the time evolution of uncertainties for the slow modulation parameters. This can be used to gain a better understanding of the general evolutionary properties of localization errors. The following questions, that we address in detail in our work, are particularly relevant.

- (i) Under what conditions do the localization uncertainties scale simply with the measured signal-to-noise ratio, and how do these uncertainties evolve during the final stages of inspiral?
- (ii) To what extent do the high and low frequency signal parameters decouple?
- (iii) What are the best determined combinations of the angular parameters?
- (iv) How and why does the shape of the 3D localization error ellipsoid change during the final week(s) of observation?

In our analysis, we neglect the ‘‘Doppler phase’’ due to LISA’s orbital motion, SMBH spin precession effects and any finite SMBH binary orbital eccentricities. These approximations are advantageous for the resulting simplicity, but the use of the HMD method is not restricted to these approximations. We also outline a generalized HMD method which remains numerically much more efficient than standard methods. We leave a numerical implementation of this general HMD method to future work. It will be particularly interesting to determine how our approximate results for the evolution of LISA localization errors are modified when spin precession effects are included, since spin precession effects were shown to improve the final localization errors by factors of 3–5 at ISCO [17, 22].

The remainder of this paper is organized as follows. In § II we define our conventions and the assumptions made in our analysis. In § III we expand the LISA GW signal in Fourier modes and obtain the conversion from actual physical parameters to corresponding Fourier amplitudes. In § IV, we incorporate these results into a Fisher matrix formalism and derive the expressions necessary to estimate correlation errors for HMD signals. In § V, we quantify the computational advantages of the HMD method. In § VI, we present results from Monte Carlo computations of the time evolution of localization errors and discuss results in terms of advance warning times for prompt electromagnetic counterpart searches. In § VII, we

develop toy models to interpret the time-dependence of LISA localization errors and to answer questions (i)–(iv) above. We summarize our results and conclude in § VIII.

II. ASSUMPTIONS AND CONVENTIONS

This section is divided into three parts. First, we list the definitions of physical quantities used in this paper, in particular the variables describing a SMBH inspiral. Second, we give the equations which determine the LISA inspiral signal. Third, we state all the assumptions made in this work.

A. Definitions

In general, an SMBH inspiral is described by a total of 17 parameters. These include 2 redshifted mass parameters, (\mathcal{M}_z, η_z) , 6 parameters related to the BH spin vectors, p_{spin} , the orbital eccentricity, e , the source luminosity distance, d_L , 2 angles locating the source in the sky, (θ_N, ϕ_N) , 2 angles that describe the relative orientation of the binary orbit, (θ_{NL}, ϕ_{NL}) , a reference time, t_{merger} , and a reference phase at ISCO, ϕ_{ISCO} , and the orbital phase, ϕ_{orb} . Throughout this work, we restrict ourselves to circular orbits by omitting the orbital eccentricity, e , and instead of the orbital phase, ϕ_{orb} , we use the look-back time before merger, t , as our evolutionary time parameter. The LISA signal for a GW inspiral is determined by the above set of parameters and two additional angular parameters describing the orientation of LISA, (Ξ, Φ) . We elaborate on the definitions of our mass and angular parameters below.

1. Mass Parameters

For component masses m_1 and m_2 , the total mass is $M = m_1 + m_2$, the reduced mass is $\mu = m_1 m_2 / M$, the symmetric mass ratio is $\eta = \mu / M$ and the chirp mass is defined as $\mathcal{M} = M \eta^{3/5}$ [24]. Throughout this work, we use geometrical units: $G \equiv c \equiv 1$. In this case, the mass can be expressed in units of time: $10^6 M_\odot \equiv 4.95 \text{ sec}$. The measured GW waveforms are insensitive to the cosmological parameters, if they are expressed in terms of the luminosity distance and the redshifted mass parameters, e.g. $m_z = (1+z)m$ (same for redshifted chirp and reduced masses).

2. Time Parameters

We write a generic look-back time (or “observation time”) before merger as t , and a generic redshifted GW frequency (or “observation frequency”) as f [37]. We use the leading order (i.e. Newtonian) approximation for the frequency evolution. Therefore, the observed frequency at look-back time t before

merger is (e.g. eq. 3.3 in ref. [25])

$$\begin{aligned} f_0(\mathcal{M}_z, t) &= \frac{5^{3/8}}{8\pi} t^{-3/8} \mathcal{M}_z^{-5/8}, \\ &= 2.7 \times 10^{-4} \text{ Hz} \left(\frac{t}{\text{day}} \right)^{-3/8} \eta_{0.25}^{-3/8} M_{6z}^{-5/8}, \end{aligned} \quad (1)$$

or equivalently

$$\begin{aligned} t_0(\mathcal{M}_z, f) &= 5(8\pi f)^{-8/3} \mathcal{M}_z^{-5/3} \\ &= 6.7 \text{ min} \left(\frac{f}{\text{Hz}} \right)^{-8/3} \eta_{0.25}^{-1} M_{6z}^{-5/3}, \end{aligned} \quad (2)$$

where M_{6z} is the redshifted total mass in units of $4 \times 10^6 M_\odot$, $\eta_{0.25} = \eta/0.25$ is the symmetric mass ratio ($\eta_{0.25} = 1$ for equal component masses, § II C), $f_c = c/R_\oplus = c/(1 \text{ AU}) = 2.00 \text{ mHz}$ is the inverse light-travel time across the radius of the LISA orbit, and the null index stands for the order of approximation. The inspiral phase extends until the innermost stable circular orbit (ISCO), at $6M$, is reached

$$\begin{aligned} f \leq f_{\text{ISCO}} &= 6^{-3/2} \pi^{-1} M_z^{-1} = 1.1 \text{ mHz} \times M_{6z}^{-1}, \\ t \geq t_{\text{ISCO}} &= 5(3/2)^4 \eta^{-1} M_z = 33 \text{ min} \times \eta_{0.25}^{-1} M_{6z}. \end{aligned} \quad (3)$$

where t_{ISCO} is the (observer-frame) look-back time before merger corresponding to the ISCO, and f_{ISCO} is the (observer-frame) frequency at ISCO.

In the present work, we fix the start of the observation (i.e. when the source first enters LISA’s frequency band) at look-back time t_i , and examine how the value of an end-of-observation time, t_f , prior to merger affects the precision on source localization. We restrict ourselves to pre-ISCO inspiral signals, corresponding to $t_f \geq t_{\text{ISCO}}$. Note that any instantaneous look-back time t associated with an observation lasting from look-back times t_i to t_f must obey $t_{\text{ISCO}} \leq t_f \leq t \leq t_i$ in our notation.

3. Angular Parameters

LISA is an equilateral triangle-shaped interferometer with an arm-length of $5 \times 10^6 \text{ km}$, orbiting around the Sun. The constellation trails 20° behind the Earth and is tilted 60° relative to the ecliptic. The detector plane precesses around the orbital axis with the same one-year period as the orbital period [26].

Following closely Refs. [19] and [17], including in notation, we define two coordinate systems. The barycentric frame is tied to the ecliptic, with $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ lying in the ecliptic plane and $\hat{\mathbf{z}}$ normal to it. The detector reference frame tied to the detector, with $\hat{\mathbf{z}}'$ normal to the detector plane, while $\hat{\mathbf{x}}', \hat{\mathbf{y}}'$ are in the plane and co-rotating with the detector so that the arms are described by time-independent vectors. We refer to the barycentric frame with normal coordinates and to the detector frame with primed coordinates. The unit vectors defining the source location on the sky, $\hat{\mathbf{N}}$, and the SMBH binary orbital angular momentum, $\hat{\mathbf{L}}$, are described by polar angles (θ_N, ϕ_N)

and (θ_L, ϕ_L) in the ecliptic frame, (θ'_N, ϕ'_N) and (θ'_L, ϕ'_L) in the detector frame:

$$\hat{\mathbf{N}}(\theta_N, \phi_N) = \hat{\mathbf{z}} \cos \theta_N + \hat{\mathbf{x}} \sin \theta_N \cos \phi_N + \hat{\mathbf{y}} \sin \theta_N \sin \phi_N, \quad (5)$$

$$\hat{\mathbf{L}}(\theta_L, \phi_L) = \hat{\mathbf{z}} \cos \theta_L + \hat{\mathbf{x}} \sin \theta_L \cos \phi_L + \hat{\mathbf{y}} \sin \theta_L \sin \phi_L. \quad (6)$$

Since we assume no SMBH spins, orbital angular momentum is conserved and the $(\theta_N, \phi_N, \theta_L, \phi_L)$ coordinates are time-independent properties of the sources.

Let (Ξ, Φ) be the two angles specifying the orientation of the LISA system in the ecliptic: Φ describes its orbital phase during its motion around the Sun, while Ξ describes the rotation of the triangle around its geometrical center. If their values at merger are written Ξ_0 and Φ_0 , then at an arbitrary look-back time t :

$$\Xi(t) = \Xi_0 - \omega_{\oplus} t, \quad (7)$$

$$\Phi(t) = \Phi_0 - \omega_{\oplus} t, \quad (8)$$

where $\omega_{\oplus} \equiv 2\pi/\text{yr}$ is the orbital angular velocity around the Sun.

The time dependence of the detector normal vector $\hat{\mathbf{z}}'$ can be expressed as

$$\hat{\mathbf{z}}' = \frac{1}{2} \hat{\mathbf{z}} - \frac{\sqrt{3}}{2} \hat{\mathbf{x}} \cos \Phi - \frac{\sqrt{3}}{2} \hat{\mathbf{y}} \sin \Phi. \quad (9)$$

The detector angles are given by

$$\cos \theta'_N = \frac{1}{2} \cos \theta_N - \frac{\sqrt{3}}{2} \sin \theta_N \cos(\Phi - \phi_N), \quad (10)$$

$$\phi'_N = \Xi + \tan^{-1} \left[\frac{\frac{\sqrt{3}}{2} \cos \theta_N + \frac{1}{2} \sin \theta_N \sin(\Phi - \phi_N)}{\sin \theta_N \sin(\Phi - \phi_N)} \right]. \quad (11)$$

Let us also define ψ' , the *polarization angle* of the GW waveform, as [17]

$$\tan \psi' = \frac{\hat{\mathbf{L}} \cdot \hat{\mathbf{z}}' - (\hat{\mathbf{L}} \cdot \hat{\mathbf{N}})(\hat{\mathbf{z}}' \cdot \hat{\mathbf{N}})}{\hat{\mathbf{N}} \cdot (\hat{\mathbf{L}} \times \hat{\mathbf{z}}')}. \quad (12)$$

Note that there are only 6 independent angular parameters $(\theta_N, \phi_N, \theta_L, \phi_L, \Xi, \Phi)$. Other detector specific quantities like $\theta'_N, \phi'_N, \theta'_L, \phi'_L$, and ψ' can be expressed in terms of these 6 independent parameters using eqs. (5–12).

Let us introduce a new set of 6 independent angles,

$$\Omega = (\theta_N, \phi_N, \theta_{NL}, \phi_{NL}, \alpha, \gamma), \quad (13)$$

with the following definitions:

- θ_{NL} is the relative latitude of $\hat{\mathbf{L}}$ and $\hat{\mathbf{N}}$ (i.e. the inclination of the binary orbit to the line of sight),
- ϕ_{NL} is the relative longitude of $\hat{\mathbf{L}}$ and $\hat{\mathbf{N}}$,
- $\alpha \equiv \Xi - \Phi + \phi_N - \frac{3\pi}{4}$,
- $\gamma(t) \equiv \Phi(t) - \phi_N$.

The explicit definitions are given in Appendix B.

Let us refer to the angles at the reference time $t = t_{\text{merger}} = 0$ as $\Omega(0)$. Although $\Phi \equiv \Phi(t)$ and $\Xi \equiv \Xi(t)$ are time-dependent, as given by (7,8) α is a time-independent combination, unlike the time-dependent $\gamma \equiv \gamma(t)$. The angles at $t = 0$ are thus given by $\Omega(0) = (\theta_N, \phi_N, \theta_{NL}, \phi_{NL}, \alpha, \gamma_0)$.

These angles have the interesting property that they possess isotropic *a priori* distributions, like the original $\Omega(0)$ variables, but the measured GW waveforms expressed in terms of these new variables are much simpler than when they are expressed in terms of the original set eqs. (5–12).

Two additional quantities which are useful to describe the sensitivity of the detector in various directions are the antenna beam patterns [19]:

$$F_{\times,+}(\Omega) = \frac{1 + \cos^2 \theta'_N}{2} \cos 2\phi'_N \cos 2\psi'_N \pm \cos \theta'_N \sin 2\phi'_N \sin 2\psi'_N, \quad (14)$$

where the sign \pm is defined to be positive for F_{\times} , and negative for F_{+} . Equation (14) and the transformation rules eqs. (5–12) define the time evolution of the antenna beam patterns for a given set of final angles $\Omega(0)$ as the LISA system orbits around the Sun. Note that the LISA system is equivalent to two independent orthogonal-arm interferometers which are rotated by 45° relative to each other [19]. Both data-streams are given by the same equations (see eq. [21] below), modulo a change of one of the angles for the second detector: $\phi''_N = \phi'_N - \pi/4$ (or equivalently $\alpha'' = \alpha' - \pi/4$ using our time-independent angular variables). Thanks to this simple relationship between the two data-streams, it is possible to carry out all the calculations for the first data-stream, and later include the second data-stream in the final expression by varying the fiducial angle α .

4. Grouping the Parameters

We group the most important parameters describing the inspiral as follows:

$$p_{\text{slow}} \equiv \{d_L, \Omega\}, \quad (15)$$

$$p_{\text{fast}} \equiv \{\mathcal{M}_z, \mu_z, t_{\text{merger}}, \phi_{\text{ISCO}}\}, \quad (16)$$

$$p_{\text{spin}} \equiv \{2 \text{ spin magnitudes}, 4 \text{ spin angles}\}. \quad (17)$$

This organization of parameters has fundamental importance in our formalism. As we show in § II B, the parameters p_{fast} and p_{spin} relate to the high frequency GW signal, while the parameters p_{slow} relate to the distinctly slow orbital modulation.

B. LISA Inspiral Signal Waveform

For a circular binary inspiral, the two polarizations of GW signal are well approximated by the restricted post-Newtonian

expressions

$$h_+(t) = 2 \frac{\mathcal{M}^{5/3} (\pi f)^{2/3}}{d_L} (1 + \cos^2 \theta_{NL}) \cos \phi_{\text{GW}}(t), \quad (18)$$

$$h_\times(t) = -4 \frac{\mathcal{M}^{5/3} (\pi f)^{2/3}}{d_L} \cos \theta_{NL} \sin \phi_{\text{GW}}(t). \quad (19)$$

The GW phase $\phi_{\text{GW}}(t) \equiv \phi_{\text{GW}}(p_{\text{fast}}, p_{\text{spin}}; t)$, which is twice the orbital phase, $\phi(t) = 2\phi_{\text{orb}}(t)$, can be expanded into the series

$$\begin{aligned} \phi_{\text{GW}}(p_{\text{fast}}, p_{\text{spin}}; t) &\approx \phi_{\text{ISCO}} + \phi_0(\mathcal{M}_z; t) + \phi_1(\mathcal{M}_z, \mu_z; t) \\ &\quad + \phi_2(\mathcal{M}_z, \mu_z, p_{\text{spin}}; t) + \dots, \end{aligned} \quad (20)$$

where $\phi_0(\mathcal{M}_z; t)$ is the leading order Newtonian solution to the phase evolution, successive terms correspond to small general relativistic corrections, ϕ_{ISCO} is the reference phase at ISCO and $\phi_n(t_{\text{ISCO}}) = 0$ for all $n \geq 0$. The instantaneous GW frequency is defined as the time derivative of the GW phase (20), i.e. $f = f(t) \equiv d\phi_{\text{GW}}/dt$, which changes very slowly compared to the GW phase itself, $\phi_{\text{GW}}(t)$. In practice we use the Newtonian approximation (1), $f_0(t) = d\phi_0/dt$. Note that equation (20) depends implicitly on the reference time, t_{merger} , since our time variable t is the look-back time before t_{merger} (see § II A 2)

The signal measured by LISA is a linear combination of the two polarizations (18), weighted by the antenna beam patterns $F_+^{\text{I,II}}$ and $F_\times^{\text{I,II}}$ for each of the two equivalent interferometers, defined by (14), resulting in the two observable data-streams

$$h^{\text{I,II}}(t) = \frac{\sqrt{3}}{2} [F_+^{\text{I,II}} h_+(t) + F_\times^{\text{I,II}} h_\times(t)], \quad (21)$$

where the factor $\sqrt{3}/2 = \sin(60^\circ)$ comes from the opening angle of the LISA arms. The beam patterns are determined by the relative orientation of the source polarizations and the detector. Their time-dependence is due to the following three main effects: LISA changes its orientation as it orbits the Sun, LISA changes its relative distance to the source as it orbits the Sun, and the orbital plane of the SMBH binary can precess because of spin-orbit coupling effects. Substituting (18) in (21) and expressing it in complex form, we get

$$h^{\text{I,II}}(t) = \frac{\mathcal{A}(\mathcal{M}_z, f)}{d_L} G^{\text{I,II}}(\Omega, f) e^{i\phi_{\text{GW}}(p_{\text{fast}}, p_{\text{spin}}; t)}, \quad (22)$$

where $\mathcal{A}(\mathcal{M}_z, f)/d_L$ defines the overall amplitude scale, with

$$\mathcal{A}(\mathcal{M}_z, f) = 2\sqrt{3} (\pi f)^{2/3} \mathcal{M}_z^{5/3}. \quad (23)$$

The $G(\Omega, f)$ factor defines the angular dependence of the signal,

$$G^{\text{I,II}}(\Omega, f) = G_A^{\text{I,II}}(\Omega) e^{i\varphi_{\text{D}}(\Omega, f)}, \quad (24)$$

where $G_A(\Omega)$, the *amplitude modulation*, captures the varying detector sensitivity with direction and polarizations of the GWs,

$$G_A^{\text{I,II}}(\Omega) = \frac{1 + \cos^2 \theta_{NL}}{2} F_+^{\text{I,II}}(\Omega) - i \cos \theta_{NL} F_\times^{\text{I,II}}(\Omega). \quad (25)$$

The additional $\varphi_{\text{D}}(\Omega, f)$ modulation is the *Doppler phase modulation*, which is the difference between the phase of the wavefront at the detector and at the barycenter [19]:

$$\varphi_{\text{D}}(\Omega, f) = 2\pi \frac{f}{f_c} \sin \theta_N \cos \gamma. \quad (26)$$

There is a non-negligible number of Doppler phase cycles only for a GW frequency satisfying $f \geq f_c$ (see definition of f_c below eq. [2] above). However, equation (3) shows that $f \leq f_{\text{ISCO}} < f_c$, hence the f_c frequency is reached only after ISCO for typical SMBH component masses of $m_1 = m_2 = 10^6 M_\odot$ and redshift $z = 1$. Even for smaller $10^5 M_\odot$ component masses, the total number of cycles, N_{pm} , remains < 1 until the final 5 hr of inspiral. Therefore the Doppler phase (26) is practically negligible for SMBH inspirals. In fact, estimating localization errors without accounting for the Doppler phase affects results by less than a factor of 10^{-3} (for $m_1 = m_2 = 10^6 M_\odot$ at $z = 1$; S. A. Hughes, private communication). Therefore, in eq. (24), we neglect $\varphi_{\text{D}}(\Omega, f)$ and restrict ourselves to the approximation

$$G^{\text{I,II}}(\Omega, f) \equiv G_A^{\text{I,II}}(\Omega). \quad (27)$$

The explicit frequency-dependence dropped out, and the time evolution of the signal G_A is now fully determined by the time evolution of the angles Ω .

Note that the amplitude modulation (25), $G_A^{\text{I,II}}(\Omega)$, is traditionally expressed in complex polar notation (e.g. [19]), where the magnitude and argument of the complex number are called *polarization amplitude* and *phase*. As we will show, the mode decomposition is simplest in the original Cartesian complex form (25), which already includes both the polarization amplitude and phase; thus, we do not distinguish these two quantities in the following. The function $G^{\text{I,II}}(\Omega, f)$ given in (24) also accounts for spin-orbit precession if the orbital orientation (θ_{NL}, ϕ_{NL}) in Ω is chosen to be time-dependent, to satisfy the equations for spin-orbit precession, and if an extra precession phase shift, $\exp(i\delta_P(\theta_{NL}, \phi_{NL}))$, is introduced (see eq. 2.14 in Lang & Hughes [22]) in addition to the Doppler phase in (24). In our calculations, we neglect spin precession but discuss how the HMD method can be extended to include that effect in § IV C.

Finally, we express the measured signal (22) as

$$h^{\text{I,II}}(p; t) = h_c(p_{\text{fast}}, p_{\text{spin}}; t) \times h_m^{\text{I,II}}(p_{\text{slow}}; t), \quad (28)$$

where h_c is the *high frequency carrier signal* and h_m is the *slow modulation*:

$$h_c(p_{\text{fast}}, p_{\text{spin}}; t) = \mathcal{A}(\mathcal{M}_z, f(t)) e^{i\phi_{\text{GW}}(p_{\text{fast}}, p_{\text{spin}}; t)} \quad (29)$$

$$h_m^{\text{I,II}}(p_{\text{slow}}; t) = \frac{G_A^{\text{I,II}}(\Omega(t))}{d_L}. \quad (30)$$

Equation (28) shows that the two sets of parameters p_{slow} and $\{p_{\text{fast}}, p_{\text{spin}}\}$ are exclusively determined by the low frequency modulation and the high frequency carrier, respectively. For this reason, we only expect a low level of cross-correlation between these sets of parameters: parameters associated with very different timescale components should essentially decouple. In Sec. VII A and Appendix A, we consider

several toy models which allow us to understand the necessary conditions, and the extent to which, parameters associated with high and low frequency components decorrelate in the course of an extended, continuous observation.

C. Simplifying Assumptions

In the present work, we make the following assumptions:

1. We assume that the amplitude modulation can be used to determine the luminosity distance and angular parameters, $p_{\text{slow}} = \{d_L, \theta_N, \phi_N, \theta_{NL}, \phi_{NL}\}$, while the other parameters, $p_{\text{fast}} = \{\mathcal{M}_z, \mu_z, t_{\text{merger}}, \phi_{\text{ISCO}}\}$, are determined from the high frequency GW phase. We assume no cross-correlations between these two sets of parameters. This is supported by the results listed in Table 1 of Hughes [7], which shows the full covariance matrix of a Monte Carlo realization of 2PN waveforms. The correlation coefficients are ~ 0.1 for the above quantities, and the absolute scale of the second set of parameters is very low in the first place. Berti et al. [12, 13] also report that the sets p_{fast} and p_{slow} are relatively uncorrelated for general relativity and even for alternative theories of gravity. In the latter case, the carrier $h_c(t)$ in the signal (28) is modified but not the slow modulation, $h_m(t)$, so that the general expectation of decoupling is maintained.
2. We assume that there are no additional errors on the detector orientations $\Phi(0)$ and $\Xi(0)$. These parameters are given by t_{merger} via eq. (8) and (7), and t_{merger} itself is determined by the high frequency carrier signal to high precision. Using the full data-stream up to ISCO, $\delta t_{\text{merger}} \sim 2 \text{ sec}$ [5, 7]. Using (8) and (7), we estimate $|\delta\Phi(0)| = |\delta\Xi(0)| \equiv \omega_{\oplus} \delta t_{\text{merger}} = 4 \times 10^{-7} \text{ rad} = 0.08''$. This is so small that we expect the errors $\delta\Phi(0)$ and $\delta\Xi(0)$ to be negligible at any relevant end-of-observation times $t_f > t_{\text{ISCO}}$, even if the t_f -dependence of these errors scale as steeply as $(S/N)^{-1}$ (see also Appendix A).
3. We use the circular, restricted post-Newtonian (PN) approximation for the GW waveform, keeping only the leading order (i.e. Newtonian) term in the signal amplitude. Higher order corrections to the GW amplitude introduce additional structure to the waveform. They improve the parameter estimation uncertainties for high mass binaries [27, 28] and introduce additional correlations between the parameters. It will be important to consider these corrections to the amplitude in future investigations. Arbitrary PN corrections to the GW phase only enter via h_c in the signal given by eq. (28). Since we neglect correlations between the sets of parameters p_{slow} and $(p_{\text{fast}}, p_{\text{spin}})$, all the restricted PN corrections to the phase drop out and become irrelevant for the p_{slow} parameter estimations.
4. We neglect the effects of Doppler phase modulation. This is plausible for SMBH binaries with component masses $m > 10^5 M_{\odot}$, since the GW wavelength in this case is generally greater than LISA's orbital diameter and $N_{\text{pm}} < 1$ (see eqs. [26] and [1]).
5. We neglect SMBH spins and, in particular, neglect the spin-orbit precession for angular determinations. This assumption is useful in simplifying our equations and in focusing on the behavior of pure angular modulation. Future studies can incorporate spin-orbit precession by convolving the angular modulation decomposition with spin-orbit effects.
6. We fix the start of LISA observations at a look-back time $t_i \equiv \min\{t_0(f_{\text{min}}), 1 \text{ yr}\}$ prior to merger. This corresponds to the time when the GW inspiral frequency f crosses the low frequency noise wall of the detector at $f_{\text{min}} = 0.03 \text{ mHz}$, but we limit the initial look-back time to a maximum of 1 yr before merger. Note that LISA's effective mission lifetime is estimated to be 3 yr. Integrated observation times longer (but also shorter) than our assumed t_i values are possible in principle, depending on source specifics. In a more elaborate treatment, one could define t_i as an a priori random variable. We fix t_i here mostly for simplicity and focus on the effects of varying the values of t_f ($< t_i$). In § VII A we show that localization errors are primarily determined by the end-of-observation time, t_f , and that values of $t_i > 1 \text{ yr}$ do not significantly change the evolution or final localization error estimates. If, however, $t_i \ll 1 \text{ yr}$ (that is, if t_{merger} is within a few months of the beginning of observation), then localization errors can become significantly worse than in our results with $t_i = 1 \text{ yr}$.
7. We neglect finite arm-length effects and we do not make use of the three independent observables of the time delay interferometry [29]. This is a valid assumption for SMBH inspirals since here $f \ll c/L = 0.01 \text{ Hz}$.
8. We neglect any finite orbital eccentricities. We note that, for eccentric orbits, higher order harmonics appear in the GW phase. In principle, since these harmonics affect the high frequency GW phase, but not the slow modulation, including finite eccentricities should not significantly affect localization error estimates. For rather eccentric orbits, high-order harmonics with $f \gg f_c$ can have a non-negligible Doppler phase (2), which would lead to an improvement in the determination of θ_N and ϕ_N . Although eccentricity is efficiently damped by gravitational radiation reaction [30], the presence of gaseous circumbinary disks could lead to non-zero eccentricities for at least some LISA inspiral events [31, 32].
9. We follow Barack & Cutler [21] in calculating the LISA root spectral noise density, $S_n(f)$, which includes the instrumental noise as well as galactic/extra-galactic backgrounds. For the instrumental noise [13], we use the effective non-angularly averaged online LISA Sensitivity Curve Generator[38], while we use the isotropic formulae for the galactic and extra-galactic background [21].

10. Our analysis focuses on statistical errors and does not account for possible systematic errors. For example, waveform templates might be inaccurate either due to the imprecision of the theory if the true waveform is not the one predicted by general relativity, or due to practical limitations from necessarily finite realizations of the large template space. Such inaccuracies can introduce new systematic errors.

III. HARMONIC MODE DECOMPOSITION

In our formalism, the angular information of the LISA inspiral signal is contained exclusively in the periodic modulation due to the detector motion around the Sun, which adds an amplitude modulation to the high frequency waveform. This modulation has a fundamental frequency, $f_{\oplus} = 1/\text{yr}$, along with upper harmonics jf_{\oplus} , where j is an integer. Although it is intuitively clear that the high frequency harmonics will tend to have a vanishing contribution, it is hard to establish this just by looking at eqs. (5–12), which define the time evolution. In this section we show that it is possible to derive surprisingly simple analytical expressions for the amplitude of each harmonic. We provide an outline of the derivation starting from the commonly used Cutler [19] formulae (5–12) and alternatively from those in Cornish & Rubbo [23]. We show that the derivation is much simpler in the latter case, in the sense that the Cornish & Rubbo [23] expression is almost already in the desired form.

A. Derivation using Cutler [19]

We expand the modulating signal (25,30) in a Fourier series

$$h_m(p_{\text{slow}}(0);t) = \frac{G_A(\Omega(t))}{d_L(z)} = \sum_{j=-\infty}^{\infty} g_j(p_{\text{slow}}(0))e^{ij\omega_{\oplus}t}, \quad (31)$$

where $g_j(p_{\text{slow}}(0))$ are the mode amplitude coefficients and $p_{\text{slow}}(0)$ are the distance and angle variables at $t = 0$ (see § II A). The coefficients can be obtained as

$$g_j(p_{\text{slow}}(0)) = \frac{1}{2\pi d_L} \int_0^{1\text{yr}} dt G_A(\Omega(t))e^{-ij\omega_{\oplus}t}. \quad (32)$$

Substituting the definition of $G_A(\Omega(t))$ from eq. (25), using the time evolution of $\Omega(t)$, eqs. (5–12), integral (32) can be evaluated.

Although conceptually simple, a direct analytical evaluation of integral (32) is overly cumbersome. Thus, for practical reasons, we follow an alternative path. We start with the original Cutler [19] formulae, given by eqs. (14) and (25). First, using general trigonometric identities, we can express $\cos 2x$ and $\sin 2x$ with $\tan(x)$ for $x = \phi'_N$ and $x = \psi'$. In the second step, we express and substitute for $\tan \phi'_N$ and $\tan \psi'_N$ with ecliptic variables using (11) and (12). In the third step, we express the trigonometric functions in complex form. After this step, each

term in the beam pattern (14) is of the form

$$\frac{\sum_n a_n e^{in\gamma}}{\sum_m b_m e^{im\gamma}}, \quad (33)$$

where the sums over n and m integers are finite, containing only a few terms, and a_n and b_n depend only on the angles $(\theta_N, \phi_{NL}, \alpha)$. In the fourth step we simplify the product of fractions. It turns out that, after combining terms, the denominators drop out exactly, leaving a formula just like (31), except that the largest element in the sum is $|j| = 4$. In the fifth step, we substitute in (25), and finally, change back from complex to trigonometric notation for the coefficients, using the half-angles $\theta_N/2$ and $\theta_{NL}/2$. Finally we arrive to the remarkably simple form:

$$h_m(p_{\text{slow}}) = d_L(z)^{-1} \sum_{j=0}^4 (LN_j D_j + L^* N_j^* D_j + L^* N_j D_j^* + LN_j^* D_j^*), \quad (34)$$

where the functions L , N , and D depend only on the angular momentum angles, sky position angles, and detector angles, respectively:

$$L(\theta_{NL}, \phi_{NL}) \equiv \sin^4 \left(\frac{\theta_{NL}}{2} \right) e^{-2i\phi_{NL}}, \quad (35)$$

$$N_j(\theta_N) \equiv w_j \cos^j \left(\frac{\theta_N}{2} \right) \sin^{4-j} \left(\frac{\theta_N}{2} \right), \quad (36)$$

$$D_j(\alpha, \gamma) \equiv i e^{2i\alpha} e^{ij\gamma}, \quad (37)$$

where $w_j = 1/16 \times (9, 12\sqrt{3}, 18, 4\sqrt{3}, 1)$ for $j = (0, 1, 2, 3, 4)$, respectively, and we have defined asterisks to refer to the following conjugates:

$$L^*(\theta_{NL}, \phi_{NL}) \equiv L(\pi - \theta_{NL}, -\phi_{NL}), \quad (38)$$

$$N_j^*(\theta_N) \equiv (-1)^j N_j(\pi - \theta_N), \quad (39)$$

$$D_j^*(\alpha, \gamma) \equiv -D_j(-\alpha, -\gamma) \equiv \overline{D_j(\alpha, \gamma)}. \quad (40)$$

Note that using these conjugate functions, only the non-negative terms $0 \leq j \leq 4$ remain in the sum (34).

Substituting the time dependence implicit in $\gamma \equiv \gamma(0) + \omega_{\oplus}t$, equation (34) becomes

$$h_m(p_{\text{slow}}(0), t) = d_L(z)^{-1} \sum_{j=-4}^4 g_j(p_{\text{slow}}(0))e^{ij\omega_{\oplus}t}, \quad (41)$$

where the coefficients are

$$g_j(p_{\text{slow}}(0)) = \begin{cases} LN_j D_{|j|}(0) + L^* N_j^* D_j(0) & \text{if } j \geq 0 \\ L^* N_{|j|} D_{|j|}^*(0) + LN_{|j|}^* D_{|j|}^*(0) & \text{if } j \leq 0 \end{cases} \quad (42)$$

and the detector functions $D_j(0)$ and $D_j^*(0)$ refer to their values at $t = 0$, $\gamma(0)$. (Note that L, N_j, L^*, N_j^* are all time-independent.) Since the decomposition (31) is unique, the coefficients (42) that we read off from our result also satisfy eq. (32).

B. Derivation using Cornish & Rubbo [23]

Our result in (34) can also be derived from the Cornish & Rubbo [23] formulae for the LISA response function. In their paper, these authors use a different set of angles, which relate to ours as follows: $\theta^{CR} = \theta_N$, $\phi^{CR} = \phi_N$, $\psi^{CR} \equiv \phi_{NL} - (\pi/2)$, $\lambda^{CR} \equiv -\alpha + \phi_N \equiv \Phi - \Xi - (3\pi/4)$, and $\alpha^{CR} \equiv \gamma + \phi_N \equiv \Phi$. Note that our set of angles is very similar to theirs, except that we measure the detector angles relative to the source, ϕ_N . This is advantageous given the rotational symmetry around the Earth orbital axis, making angles relative to the source the only ones that should have an effect on the measured GWs; we expect ϕ_N to drop out of the equations when using α and γ . Note, once again, that the variables $(\theta_N, \phi_N, \theta_{NL}, \phi_{NL}, \alpha)$ are time independent, while $\gamma \equiv \gamma(t)$. Writing the Cornish & Rubbo [23] beam patterns for low frequencies, which is fully equivalent to Cutler [19], with our angular variables[39], we get

$$F_+^{I,II} = -\frac{1}{2}[\cos(2\phi_{NL})D_+^{I,II}(t) - \sin(2\phi_{NL})D_\times^{I,II}(t)], \quad (43)$$

$$F_\times^{I,II} = -\frac{1}{2}[\sin(2\phi_{NL})D_+^{I,II}(t) + \cos(2\phi_{NL})D_\times^{I,II}(t)], \quad (44)$$

where

$$D_+ = \frac{1}{32} \{ -36 \sin^2 \theta_N \sin(2\gamma + 2\alpha) + (3 + \cos 2\theta_N) \\ \times \{ \cos(2\phi_N)[9 \sin(-2\alpha + 2\phi_N) - \sin(4\gamma + 2\alpha + 2\phi_N)] \\ + \sin(2\phi_N)[\cos(4\gamma + 2\alpha - 2\phi_N) - 9 \cos(-2\alpha + 2\phi_N)] \} \\ - 4\sqrt{3} \sin(2\theta_N)[\sin(3\gamma + 2\alpha) - 3 \sin(\gamma + 2\alpha)] \}, \quad (45)$$

and

$$D_\times = \frac{1}{8\sqrt{3}} \{ \sqrt{3} \cos \theta_N [9 \cos(-2\alpha) - \cos(4\gamma + 4\alpha)] \\ - 6 \sin \theta_N [\cos(3\gamma + 2\alpha) + 3 \cos(\gamma + 2\alpha)] \}. \quad (46)$$

One notices instantly that the time dependence here is much simpler than in the original Cutler [19] formula, as it is inscribed only in the various harmonics of γ . We can identify the highest harmonic present to be 4γ . Expanding the trigonometric functions using standard identities, we obtain

$$D_+ = -\frac{1}{32} \begin{pmatrix} 9(3 + \cos 2\theta_N) \\ -12\sqrt{3} \sin 2\theta_N \\ 36 \sin^2 \theta_N \\ 4\sqrt{3} \sin 2\theta_N \\ 3 + \cos 2\theta_N \end{pmatrix} \cdot \begin{pmatrix} \sin(2\alpha) \\ \sin(2\alpha + \gamma) \\ \sin(2\alpha + 2\gamma) \\ \sin(2\alpha + 3\gamma) \\ \sin(2\alpha + 4\gamma) \end{pmatrix}, \quad (47)$$

and

$$D_\times = -\frac{1}{8} \begin{pmatrix} -9 \cos \theta_N \\ 6\sqrt{3} \sin \theta_N \\ 0 \\ 2\sqrt{3} \sin \theta_N \\ \cos \theta_N \end{pmatrix} \cdot \begin{pmatrix} \cos(2\alpha) \\ \cos(2\alpha + \gamma) \\ \cos(2\alpha + 2\gamma) \\ \cos(2\alpha + 3\gamma) \\ \cos(2\alpha + 4\gamma) \end{pmatrix}, \quad (48)$$

where $a \cdot b = \sum_n a_n b_n$ is the usual dot product. Now, the second sets of elements carry the time dependence and the detector orientation information, while the first sets describe the

sky position. Note that the explicit ϕ_N dependence dropped out, as expected. Next, we manipulate equations (47,48), substituting complex expressions for the trigonometric ones, and substituting these into eq. (25). We finally arrive at eq. (34) after changing to half-angles $\theta_N/2$ and $\theta_{NL}/2$.

We note that eqs. (34) or (41,42) are fully general expressions, equivalent to the standard LISA inspiral signal in eqns. (14) and (25). The two data-streams are obtained by substituting $\alpha = \alpha^I$ and α^II , corresponding to the two independent LISA-equivalent Michelson interferometers (see § II A). To verify our final result, we compare numerically the signals computed using eqs. (14,25) with the signals computed using eqs. (41,42), for a large set of random choices of angles. Agreement is achieved at machine precision levels.

The main utility of eq. (34), is that it can be used to “deconstruct” parameter error histograms, i.e. to understand how the errors depend on the fiducial values of the parameters. As compared to Cornish & Rubbo [23], our result leads to an explicit decoupling of the signal angular dependence into simple products of one-dimensional functions. In particular, the dependence on sky position, angular momentum, and detector angles are separated. Using the special conjugate functions L^* , D^* , N^* , eq. (34) displays the symmetry properties of the signal. Finally, one angular variable, ϕ_N is eliminated exactly.

IV. ESTIMATING PARAMETER UNCERTAINTIES IN THE HMD FORMALISM

Parameter estimations for LISA GW inspiral signals are possible with matched filtering and the expected uncertainties can be forecast using the Fisher matrix formalism [33, 34]. In this section, we apply this approach to the LISA signal derived in § III, with an angular dependence of the signal decomposed into harmonic modes. In § IV B, we consider the simple case of a high frequency carrier signal that is modulated by a low-frequency function, without any cross-correlation between the two sets of relevant parameters. We derive a simple formula for the estimation of modulating parameter uncertainties. In § IV C, we consider a more general post-Newtonian signal and show that parameters related to source localization can still be decoupled from the time evolution and the other source parameters.

A. Fisher Matrix Formalism

Let us consider a generic real signal described by the function $h(x)$, which depends on N parameters $\{p_a\}_{a \in [1, N]}$. The measured signal is $y(x) = h(x) + n(x)$, where $n(x)$ is a realization of the noise specified by a probability distribution. Let us assume that the noise is Gaussian, is statistically stationary with respect to x , has zero mean value, $\langle n(x) \rangle = 0$, where $\langle \rangle$ represents an ensemble average, and has known variance, $\sigma(x)^2 = \langle n(x)^2 \rangle$. The parameter estimation errors for p_a can then be calculated using the Cramer-Rao bound [33]

$$\langle \delta p_a \delta p_b \rangle \geq \langle \Gamma^{-1} \rangle_{ab}, \quad (49)$$

where equality is approached for high S/N signals. Here Γ_{ab} is the Fisher matrix defined by

$$\Gamma_{ab} = \int_{x_{\min}}^{x_{\max}} \frac{\partial_a h(x) \partial_b h(x)}{\sigma^2(x)} dx, \quad (50)$$

where ∂_a is the partial derivative with respect to the parameter p_a . Note that $\sigma(x)$ here is defined as the noise per unit x . In eq. (50), x denotes time t for time-domain samples, or f for frequency-domain samples. The purpose of this work is to study how an arbitrary end of the observation, at x_{\max} (or t_f below, for time samples) affects the resultant correlation errors $\langle \delta p_a \delta p_b \rangle$, for a fixed start-of-observation at x_{\min} (or t_i below, for time samples).

An important quantity for the evolution of parameter estimation errors is the signal-to-noise ratio, S/N , defined by

$$\left(\frac{S}{N}\right)^2 = \int_{x_{\min}}^{x_{\max}} \frac{h^2(x)}{\sigma^2(x)} dx. \quad (51)$$

For LISA, the noise varies with signal frequency. In this case, the Fisher matrix can be evaluated in Fourier space [33, 34],

$$\Gamma(t_f)_{ab} = \Re \left\{ 4 \int_{f_{\min}}^{f(t_f)} \frac{\overline{\partial_a \tilde{h}(f)} \partial_b \tilde{h}(f)}{S_n^2(f)} df \right\}, \quad (52)$$

where $\tilde{h}(f)$ is the Fourier transform of $h(t)$, the GW signal (28), ∂_a is the partial derivative with respect to parameter p_a , bars denote complex conjugation, and $S_n^2(f)$ is the one-sided spectral noise density (§ II A).

B. Approximate solution

We seek an alternative equivalent form of eq. (52) specific to GW inspirals for which, as in eq. (28), the high frequency carrier signal is decoupled from the slow modulation. In case of SMBH inspirals, with a high frequency signal $h_c(t)$ changing its frequency slowly as $f_0(t)$ given in eq. (1), and further modulated by a slowly varying function $h_m(t)$ as given in eq. (28), the integral in eq. (52) can be evaluated in the stationary phase approximation, by substituting

$$\tilde{h}(f) = h_m[t_0(f)] \times \tilde{h}_c(f), \quad (53)$$

where $\tilde{h}_c(f)$ is the Fourier transform of the carrier signal and $h_m[t_0(f)]$ is the modulating function evaluated at the time when the carrier frequency is f . This can be converted to the time domain, by simply changing the integration variable to $t = t_0(f)$ using the frequency evolution in eq. (2):

$$\Gamma(t_f)_{ab} = \Re \left\{ 4 \int_{t_i}^{t_f} \frac{\overline{\partial_a \tilde{h}(t)} \partial_b \tilde{h}(t)}{S_n^2[f_0(t)]} \left| \frac{df_0(t)}{dt} \right| dt \right\}, \quad (54)$$

and

$$\tilde{h}(t) = h_m(t) \times \tilde{h}_c[f_0(t)]. \quad (55)$$

We are only interested in estimating uncertainties for the p_{slow} variables (§ II A), which are determined exclusively by $h_m(t)$. Recall from eq. (29) that $|h_c(t)| = \mathcal{A}$ so that, for the Fourier transform[40], we have $|h_c(t)|^2 = 4|\tilde{h}_c(f)|^2(df/dt)$. Using these relationships, let us define the instantaneous relative noise amplitude per unit time $\sigma(t)$ as

$$\begin{aligned} \sigma^{-2}(t) &= 4 \frac{\tilde{h}_c^2[f_0(t)]}{S_n^2[f_0(t)]} \left| \frac{df_0(t)}{dt} \right| = \frac{\mathcal{A}^2[f_0(t)]}{S_n^2[f_0(t)]} \\ &= \frac{3\sqrt{5}}{4} \frac{\mathcal{M}_z^{5/2} t^{-1/2}}{S_n^2[f_0(\mathcal{M}_z, t)]}. \end{aligned} \quad (56)$$

The last equality follows from the Newtonian waveform and frequency evolution, eqs. (23) and (1). We point out that the mass dependence is captured entirely by $\sigma(t)$ and does not appear anywhere else in what follows.

By combining eqs. (54), (55), and (56), we arrive at

$$\Gamma(t_f)_{ab} = \int_{t_i}^{t_f} \Re \left\{ \frac{\overline{\partial_a h_m(t)} \partial_b h_m(t)}{\sigma^2(t)} \right\} dt. \quad (57)$$

Equation (57) is the special case of (52), where the carrier signal-to-noise ratio and modulation, h_m , are conveniently isolated.

We are now ready to make use of the harmonic mode decomposition. Substituting eq. (31) into (57) gives

$$\Gamma(t_f)_{ab} = \Re \left\{ \sum_{j_1, j_2=-4}^4 \frac{\overline{\partial_a g_{j_1}} \partial_b g_{j_2} P_{j_2-j_1}(t_f)}{\sigma^2(t)} \right\}, \quad (58)$$

where

$$P_{j_2-j_1}(t_f) = \int_{t_i}^{t_f} \frac{e^{i(j_2-j_1)\omega \oplus t}}{\sigma^2(t)} dt. \quad (59)$$

The function $P_j(t_f)$ is shown in Figure 1 for $j = 0$, together with real and imaginary parts for the $j = 4$ case, for $m_1 = m_2 = 10^6 M_\odot$ at $z = 1$. Since the accumulated signal-to-noise ratio is $S/N = P_0(t)$, the figure shows that the instantaneous signal-to-noise ratio is $[d/dt](S/N) = [d/dt]P_0(t) \approx t^{-2}$. The extrapolated signal-to-noise blows up at ‘‘merger’’ ($t = 0$). Data analysis for such a non-stationary signal-to-noise ratio evolution has several interesting implications, which we study further with toy models in Appendix A. We find that, for such a signal-to-noise ratio evolution, specific combinations of parameters can always be measured to very high accuracy.

The time dependence in eq. (59) couples only to the combination $j = j_2 - j_1$. This allows us to rearrange the double sum on (j_1, j_2) and evaluate one of them independent of time:

$$\Gamma(p_{\text{slow}}, t_f)_{ab} = \Re \left\{ \sum_{j=-8}^8 [\mathcal{F}_j(p_{\text{slow}}(0))]_{ab} P_j(t_f) \right\}, \quad (60)$$

where

$$[\mathcal{F}_j(p_{\text{slow}}(0))]_{ab} = \sum_{j'=-8}^8 \overline{\partial_a g_{j+j'}(p_{\text{slow}}(0))} \partial_b g_{j'}(p_{\text{slow}}(0)). \quad (61)$$

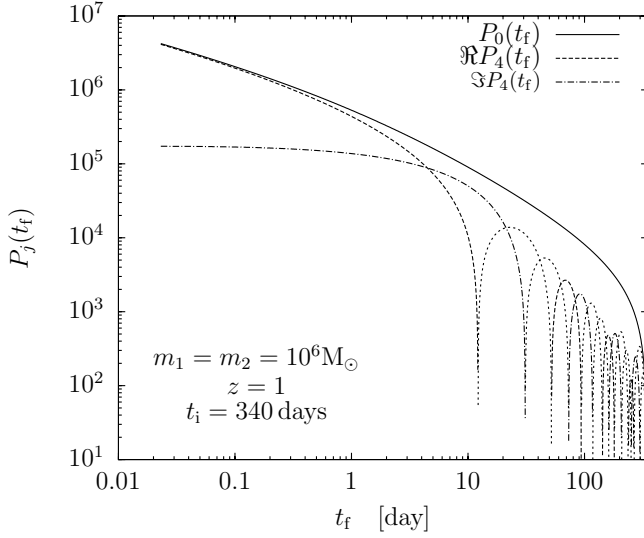


FIG. 1: The time evolution of the fundamental functions $P_j(t_f)$, used to construct the Fisher matrix to forecast localization errors by LISA. The dependence of the Fisher matrix on the look-back time t_f is obtained from the 9 fundamental functions with $0 \leq j \leq 8$. The curves show $P_0(t_f)$, as well as the real and imaginary parts of $P_4(t_f)$, for $m_1 = m_2 = 10^6 M_\odot$ and $z = 1$. Thin dotted lines represent negative values. Note that $P_0(t_f) \equiv (S/N)^2$, which is the simple scaling of inverse squared errors, neglecting correlations. The signal-to-noise ratio scales steeply, approximately as $S/N \propto t_f^{-1}$. The curve $P_4(t_f)$ illustrates how all the other similar $P_j(t_f)$ functions vary, with a relative number $|j|$ of oscillations, and $P_{-j}(t_f) \equiv \overline{P_j(t_f)}$.

Our parameters in the correlation matrix are $p_a = (d_L, \theta_N, \phi_N, \theta_{NL}, \phi_{NL})$ since we assume that the other parameters, i.e. $\{\mathcal{M}_z, \eta_z, \phi_{\text{ISCO}}, t_{\text{merger}}, \alpha, \gamma(0)\}$, are known from the high frequency carrier signal (§ II C). It is straightforward to compute the derivatives of $g_j(d_L, \Omega)$ using eq. (42) for all parameters p_a , except ϕ_N . The ϕ_N dependence in g_j in eq. (42) is implicit in $\alpha \equiv \alpha(\Xi(0), \Phi(0), \phi_N)$ and $\gamma(0) \equiv \gamma(\Phi(0), \phi_N)$ (see § II A). Since we assume that $\Xi(0)$ and $\Phi(0)$ are measured to very high precision with the high frequency carrier (§ II C), we can use the chain rule to express the ϕ_N derivative as $\partial_{\phi_N} g_j = \partial_\alpha g_j - \partial_{\gamma_0} g_j$.

Up to this point we did not make use of the fact that the LISA signal is equivalent to two orthogonal arm interferometers rotated by $\Delta\alpha = \pi/4$ with respect to each other. To account for both data-streams being measured simultaneously, the Fisher matrix is written as the sum of the two Fisher matrices corresponding to each individual interferometer. Writing out only the α dependence, we have $\Gamma_{ab}^{\text{tot}}(\alpha) = \Gamma_{ab}(\alpha) + \Gamma_{ab}(\alpha - \pi/4)$. Finally, according to eq. (49), the parameter error covariance matrix is the inverse of this total Fisher matrix:

$$\langle \delta p_a \delta p_b \rangle \geq [\Gamma_{ab}(d_L, \theta_N, \theta_{NL}, \phi_{NL}, \alpha, \gamma(0)) + \Gamma_{ab}(d_L, \theta_N, \theta_{NL}, \phi_{NL}, \alpha - \pi/4, \gamma(0))]^{-1}. \quad (62)$$

Equation (62) along with (60) is our final expression, describing the time evolution of parameter estimation uncertainties. We note that after combining both data-streams, the matrices $[\mathcal{F}_j(p_{\text{slow}}(0))]_{ab}$ for $4 < |j| \leq 8$ modes vanish exactly for all

$p_{\text{slow}}(0)$.

Let us emphasize the most important features of eq. (60):

- The parameter dependence is separated from the time dependence. The Fisher matrix, Γ_{ab} , is written as a linear combination of matrices $\mathcal{F}_j(p_{\text{slow}})$ weighted by the scalars $P_j(t)$, where $\mathcal{F}_j(p_{\text{slow}})$ is independent of time and $P_j(t)$ is independent of the parameters p_a . Evaluating $\mathcal{F}_j(p_{\text{slow}})$ requires the computation of the parameter derivatives $\partial_a g_j$.
- The evaluation of all integrals $P_j(t_{f_n})$ for different $n = 1, 2, \dots, N_{t_f}$ can be done in the same amount of time as needed for a single integration, since the t_f dependence enters only in the integration bound in eq. (59),
- Large Monte Carlo (MC) simulations can easily be performed since the time evolution is given by a small number of functions, $P_j(t)$, which can be calculated a priori and pre-saved. No integrations at all are necessary during the MC simulation for calculating distributions of correlation matrices.

In § V below, we estimate the improvement in the computation time provided by the HMD method for calculating distributions of parameter errors and their time evolution.

C. Generalization to the exact PN signal

So far, we only considered the simplest case, assuming no cross-correlation between parameters p_{slow} and $(p_{\text{fast}}, p_{\text{spin}})$, for a restricted post-Newtonian waveform. Moreover, we assumed the Doppler-phase (24) to be negligible. Including cross-correlations and the Doppler phase would allow us to examine the range of validity of our approximations, and it would allow us to extend computations to the lower component mass regime, $m < 10^5 M_\odot$, where the Doppler phase becomes important. Furthermore, including spin precession effects can modify angular localization errors by a factor of ~ 3 , at least for the final errors at $t_f \approx t_{\text{ISCO}}$ [17, 22]. While we continue to use our initial set of approximations in later sections, we outline here how the HMD formalism could be used to decouple the time-dependence from the angular parameter-dependence, even in the case of the most general (arbitrary order) restricted post-Newtonian waveform. Source sky position angles (θ_N, ϕ_N) , detector angles at ISCO $(\alpha, \gamma(0))$ and luminosity distance (d_L) can all be decoupled even if spin-orbit and spin-spin precessions are included in the waveform, which is potentially a great advantage over the traditional calculation methods (see § V for a detailed discussion).

Consider a general restricted post-Newtonian signal given by eq. (22), for which we substitute the harmonic mode expansion [41],

$$\tilde{h}^{\text{I,II}}(f) = \tilde{A} \sum_{j=-4}^4 g_j^{\text{I,II}} f^{-7/6} e^{i[j\omega_{\oplus} t(f) + \varphi_D + \phi_{\text{GW}}]}, \quad (63)$$

where $\tilde{A} \equiv \tilde{A}(p_{\text{fast}})$ is the amplitude in the frequency-domain (eq. 68 in ref. [17]), $g_j \equiv g_j(p_{\text{slow}})$ is the modulation amplitude in eq. (42), $\varphi_D \equiv \varphi_D(p_{\text{slow}}, f) = c_D(p_{\text{slow}})f$ is the Doppler phase (see eq. 26 above), $c_D(p_{\text{slow}}) = 2\pi f_c^{-1} \sin \theta_N \cos \gamma$, $\phi_{\text{GW}} \equiv \phi_{\text{GW}}(p; f)$ is the GW phase (20) (e.g. eq. 3.2 in ref. [25]),

$$\phi_{\text{GW}}(p; f) = \sum_{n=0}^N c_n^{\text{GW}}(p_{\text{fast}}, p_{\text{spin}}) u_n^{\text{GW}}(f), \quad (64)$$

and time-frequency relationships $t(f) \equiv t(p; f)$ can be written as (e.g. eq. 3.3 in ref. [25])

$$t(p; f) = \sum_{n=0}^N c_n^{\text{T}}(p_{\text{fast}}, p_{\text{spin}}) u_n^{\text{T}}(f). \quad (65)$$

Here, the c_n^{GW} and c_n^{T} coefficients are frequency-independent, while $u_n^{\text{GW}}(f)$ and $u_n^{\text{T}}(f)$ are parameter-independent. They correspond to the various post-Newtonian terms in the post-Newtonian expansion, and N corresponds to the highest order term. The frequency functions are very simple powers of f , i.e. $u_n^{\text{GW}}(f) = f^{n-(5/3)}$ and $u_n^{\text{T}}(f) = f^{n-(8/3)}$. (Note that neither the c_n^{GW} and c_n^{T} coefficients nor the $u_n^{\text{GW}}(f)$ and $u_n^{\text{T}}(f)$ functions are complex. Every term in eq. (63) is real except for the $g_j \equiv g_j(p_{\text{slow}})$ coefficients and the complex argument.)

Equations (63-65) can be combined into

$$\tilde{h}(f)^{\text{I,II}} = \sum_{j=-4}^4 A_j^{\text{I,II}}(p) f^{-7/6} e^{i\Psi_j(p; f)}, \quad (66)$$

where

$$A_j^{\text{I,II}}(p) = \tilde{A}(p_{\text{fast}}) g_j^{\text{I,II}}(p_{\text{slow}}), \quad (67)$$

and

$$\begin{aligned} \Psi_j(p; f) &= c_D f + j\omega_{\oplus} \sum_{n=0}^N c_n^{\text{T}} u_n^{\text{T}}(f) + \sum_{n=0}^N c_n^{\text{GW}} u_n^{\text{GW}}(f) \quad (68) \\ &\equiv \sum_{k=0}^{2N+1} c_{jk} u_k(f), \quad (69) \end{aligned}$$

where in the last step we introduced $u \equiv \{f, u^{\text{T}}, u^{\text{GW}}\}$ and $c_j \equiv \{c_D, j\omega_{\oplus} c^{\text{T}}, c^{\text{GW}}\}$ to collect all f -functions and coefficients in one vector and one matrix, respectively.

To compute the Fisher matrix, we need to obtain the partial derivatives of the signal with respect to the parameters:

$$\tilde{h}_{,a}(f) = \sum_{j=-4}^4 \left(A_{j,a} + i \sum_{k=0}^{2N+1} A_j c_{jk,a} u_k(f) \right) f^{-7/6} e^{i\Psi_j(f)}, \quad (70)$$

where commas in indices denote partial derivatives with respect to the parameter following the index. Note, that the parameter index a spans all parameters p_{slow} , p_{fast} , and p_{spin} , and the Fisher matrix accounts for correlations between these parameters.

Equation (70) can now be substituted in the Fisher matrix in eq. (52). We get

$$\Gamma(t_f)_{ab} = \Gamma_{ab}^{(0)}(t_f) + \Gamma_{ab}^{(1)}(t_f) + \Gamma_{ab}^{(2)}(t_f), \quad (71)$$

where

$$\Gamma_{ab}^{(0)}(t_f) = \Re \left\{ \sum_{j_1, j_2=-4}^4 A_{j_1, a} \overline{A_{j_2, b}} P_{j_1 j_2}^{(0)}(t_f) \right\}, \quad (72)$$

$$\begin{aligned} \Gamma_{ab}^{(1)}(t_f) = \Re \left\{ i \sum_{j_1, j_2=-4}^4 A_{j_1, a} \overline{A_{j_2, b}} \sum_{k=0}^{2N+1} c_{j_1 k, b} P_{j_1 j_2 k}^{(1)}(t_f) \right. \\ \left. - i \sum_{j_1, j_2=-4}^4 A_{j_1, a} \overline{A_{j_2, b}} \sum_{k=0}^{2N+1} c_{j_1 k, a} P_{j_1 j_2 k}^{(1)}(t_f) \right\}, \quad (73) \end{aligned}$$

$$\Gamma_{ab}^{(2)}(t_f) = \Re \left\{ - \sum_{j_1, j_2=-4}^4 A_{j_1, a} \overline{A_{j_2, b}} \sum_{k_1, k_2=0}^{2N+1} c_{j_2 k_1, a} c_{j_1 k_2, b} P_{j_1 j_2 k_1 k_2}^{(2)}(t_f) \right\}, \quad (74)$$

and where

$$\begin{aligned} P_{j_1 j_2}^{(0)}(t_f) &= 4 \int_{f_{\text{min}}}^{f(t_f)} \frac{f^{-7/3} e^{i(j_1-j_2)\omega_{\oplus} t(f)}}{S_n^2(f)} df, \\ P_{j_1 j_2 k}^{(1)}(t_f) &= 4 \int_{f_{\text{min}}}^{f(t_f)} \frac{f^{-7/3} u_k(f) e^{i(j_1-j_2)\omega_{\oplus} t(f)}}{S_n^2(f)} df, \quad (75) \\ P_{j_1 j_2 k_1 k_2}^{(2)}(t_f) &= 4 \int_{f_{\text{min}}}^{f(t_f)} \frac{f^{-7/3} u_{k_1}(f) u_{k_2}(f) e^{i(j_1-j_2)\omega_{\oplus} t(f)}}{S_n^2(f)} df, \end{aligned}$$

are the frequency dependent terms.

Equation (71) is our final result, where the localization parameters (i.e. angles and distance) are decoupled from all other parameters (i.e. masses, spins, reference time and phase at ISCO). The equation explicitly shows that, contrary to the traditional methods usually adopted for Monte Carlo computations of random binary orientations and sky positions [5, 7, 8, 13, 17, 19, 20, 21, 22], the localization of a LISA inspiral event and its time-dependence can be explored without the need to evaluate integrals for each realization of the fiducial angles. Note that the only approximation made to obtain eq. (71) was to neglect of spin-orbit and spin-spin precession in the general restricted post-Newtonian solution for the Fisher matrix. The time-dependence is given by the $P(t_f)$ functions in eq. (75) and the extrinsic parameter-dependence is given by the coefficients, A . The $P(t_f)$ functions in eq. (75) can be computed a priori, independently of the fiducial angles. Note that $P(t_f)$ depends implicitly on the parameters (p_{fast}, p_S) through $t(f)$ in eq. (65), and its inverse $f(t)$. Generally, there are at most $(2J_{\text{max}} + 1) \times [1 + (2N+1) + (1/2)(2N+1)(2N+2)]$ such independent functions.

From the general case, we can now deduce the special solution in eqs. (60) and (62) valid for a Newtonian evolution, no Doppler phase, and no cross-correlations between p_{slow} and $(p_{\text{fast}}, p_{\text{spin}})$. This approximation simply corresponds to the first term $\Gamma_{ab}^{(0)}(t_f)$ in eq. (71), where in eq. (75) the $t(f)$

function is computed using the Newtonian formula $t_0(f)$ given by eq. (2). Note also, that the next term in eq. (71), $\Gamma_{ab}^{(1)}(t_f)$, corresponds to the cross-correlation of the amplitude modulation with the “high frequency carrier signal” (i.e. Doppler phase and GW phase). The last term, $\Gamma_{ab}^{(2)}(t_f)$, corresponds to the cross-correlations among parameters in the high frequency carrier.

Finally, we briefly consider extensions to include spin-orbit and spin-spin precessions in the signal. Let us refer to the angular momentum angles as $p_L(t) \equiv (\theta_{NL}(t), \phi_{NL}(t))$, which are now time-dependent. As we briefly show next, in the case of spin precession, the $P^{0,1,2}(t_f)$ time-dependent integrals lose the convenient property of being independent of p_L , but nevertheless, the parameters describing the sky position and detector orientation are still time-frequency independent and they are decoupled in this prescription. Indeed, an extra precession phase $\exp[i\delta_P(p_L, p_{\text{spin}}, t)]$ has to be included in eq. (63) and $g_j^{1,II}(p_{\text{slow}})$. We now have $\Psi_j^{\text{prec}}(p; f) = \Psi_j(p; f) + \delta_P(p_L, p_{\text{spin}}; t)$ instead of eq. (68). Thus, when taking the derivatives of the signal in eq. (70), we will get additional terms proportional to $A_j^{1,II}(p)\partial_a\delta_{P,a}$ and the original $A_{j,a}$ terms will have time variation due to the p_L dependence of $g_j^{1,II}$. Finally, after these modifications, the Fisher matrix will be similar to that in eq. (71), except now the p_L terms cannot be moved outside of the frequency-integral but have to be attached to the time-varying $P^{0,1,2}(t_f)$ part [42]. The main advantage we retain is therefore that the source position and detector angles $(\theta_N, \phi_N, \alpha, \gamma(0))$ will still only be included in the coefficients $A_j^{1,II}(p)$ and $c_{jk,a}$ and the time-evolution can be still be computed independently of these parameters. In § V C we show that this indeed reduces computation times by a large factor relative to the traditional methods (e.g. [17, 22]).

We leave numerical implementations and explorations of parameter distributions and their time-dependence, in this case of a general inspiral waveform, to future work.

V. COMPUTATION TIME

One of the great advantages of introducing the HMD method is the reduction in the computational time needed to evaluate error distributions for the parameters which determine how efficiently LISA can localize SMBH binary inspiral events: sky position, angular momentum orientation and final detector orientation. In general, this is a computationally very demanding task because of the large dimensionality of the angular parameter space. Mapping the structure of the distribution of correlations in the parameter space of mock LISA measurements requires vast Monte Carlo simulations, which are presently limited by computational resources. Currently, only a small portion of this space has been explored [7, 8, 13, 17, 20, 21, 22]. In this context, it is desirable to tune methods to the specific problem at hand. The HMD method described above is specifically constructed to exploit the structure of LISA inspiral signals.

A. Approximate solution

The computational time for parameter space exploration, using the HMD method with the approximations described in § IV B, is significantly reduced for the following reasons. The standard approach for estimating parameter errors requires the evaluation of an $N_p \times N_p$ symmetric Fisher matrix, where each matrix element is an integral over the range spanned by the GW frequency during inspiral (see Refs. [33, 34]; and § IV). Here, N_p is the number of parameters describing the signal. The number of independent elements in a symmetric matrix is $(1/2)N_p(N_p + 1)$. Let us assume that the evaluation of a single integral requires to compute the waveform at N_{int} separate instances. The computation of one integral is sufficient also to trace the time evolution at N_{int} different t_f values, if one uses a single trapezoidal integral in frequency from f_{ISCO} to f_{min} and stores results at each intermediate value of f . Since the time evolution of the frequency is known independently of the angles, we can already get the integral for N_{int} different t_f values. For randomly chosen fiducial angles in a MC simulation of size N_{MC} requires the calculation to be repeated N_{MC} times. To evaluate the evolution of parameter errors as a function of t_f for N_{t_f} different instances requires N_{t_f} computations. Therefore, the standard method costs $\Delta T_{\text{standard}} = (1/2)N_p(N_p + 1)N_{\text{int}}N_{MC}$ computational time units.

In contrast, with our proposed HMD method (§ IV B), the p_{slow} parameters are decoupled from the p_{fast} parameters (§ II C) and from time. The $N_p \times N_p$ Fisher matrix can be split into two smaller matrices, with $N_{p1} \times N_{p1}$ and $N_{p0} \times N_{p0}$ components, where $N_p = N_{p1} + N_{p0}$. The $N_{p1} \times N_{p1}$ matrix determines the angular errors (which are deduced from the amplitude modulation of the signal), while the other matrix determines the remaining parameters (e.g. masses, phase and time at ISCO, using the high frequency carrier only). Since we are only interested here in parameters relating to the localization of the source, p_{slow} , it is sufficient for us to consider the corresponding $N_{p1} \times N_{p1}$ matrix only. Since it is symmetric, it has only $(1/2)N_{p1}(N_{p1} + 1)$ independent elements, but it turns out that the computation of only N_{p1} elements is sufficient for a single harmonic (we need only the N_{p1} derivatives of the g_j functions, see eq. [58]). Using $2N_J + 1$ harmonic modes and a MC simulation with N_{MC} random choices of fiducial parameters costs $N_{p1}(2N_J + 1)N_{MC}$ time units. In this method, the time dependence is decoupled, so that parameter dependencies can be taken outside of the integral (see eq. [58] and $\mathcal{F}_j(p_{\text{slow}})$ in eq. [60]). The MC sampling can therefore be evaluated independently of time. The time evolution of the signal for each harmonic is known independently of the angles, by construction ($P_j(t_f)$, eq. [59]), and this integration for each component can be evaluated *a priori*, independently of the fiducial parameter values. For each such mode, we would like to evaluate a number N_{t_f} of integrals. Fortunately, since the different integrals differ only via the lower integration bound in the time domain, all integrals can be obtained during a computation of the integral with the largest time domain, $t_f = t_{\text{ISCO}}$. Therefore for a total of $(2N_J + 1)$ modes, building the time-evolution functions $P_j(t_f)$ takes of order $(2N_J + 1)N_{t_f}$ time units. This is generally much faster than building the time-independent co-

efficients $\mathcal{F}_j(p_{\text{slow}})$. In summary, with the HMD method, one only needs $\Delta T_{\text{HMD}} = N_{p1}(2N_J + 1)N_{MC} + (2N_J + 1)N_{t_f}$ units of time.

Comparing methods, we find that the computational requirements of the HMD method is lower by a factor of

$$\frac{\Delta T_{\text{HMD}}^{\text{no spin}}}{\Delta T_{\text{standard}}^{\text{no spin}}} = \frac{N_{p1}(2N_J + 1)N_{MC} + (2N_J + 1)N_{t_f}}{(1/2)N_p(N_p + 1)N_{\text{int}}N_{MC}}. \quad (76)$$

Recall from § II A that the number of parameters for a non-spin case is $N_{p0} = 4$, $N_{p1} = 5$, so that $N_p = N_{p0} + N_{p1} = 9$. Choosing $N_{\text{int}} = 10^3$, $N_{MC} = 10^4$, $N_{t_f} = 100$, and $N_J = 4$ for the other parameters in eq. (76) as a representative example, the gain in computational efficiency is $\Delta T_{\text{standard}}/\Delta T_{\text{HMD}} = (5 \times 10^8)/(5 \times 10^5) \approx N_{\text{int}} = 1000$. Moreover, the Fisher matrix is much smaller, 5×5 , which offers an important further advantage when performing the inversion to obtain the error covariance matrix. Using the HMD method, the inversion of the Fisher matrix is computationally less expensive than generating the matrix. Note that the second term in $\Delta T_{\text{HMD}}^{\text{no spin}}$, corresponding to the $P_j(t_f)$ functions, is negligible in this case and the computation time is dominated by constructing the coefficient matrices. A calculation of the representative MC example above, with a non-optimized implementation of the HMD method, takes less than a minute on a regular workstation.

The case for substantial improvement with the HMD method becomes even more compelling when additional parameters (spins and higher order PN terms) are included, as we discuss next.

B. Post-Newtonian Signal without Spin Precession

We now consider the general HMD method outlined in § IV C, with $N_{\text{spin}} \equiv 6$ spin components. The spin parameters can be grouped as $N_{SM} \equiv 2$ independent spin magnitudes and $N_{SA} \equiv 4$ independent spin angles. Since the spins can be oriented arbitrarily, the spin angular parameters have to be randomly chosen, in addition to the other angular parameters in any Monte Carlo computation. This enlargement of the parameter space of random parameters greatly increases the computational cost, both for the standard method and the HMD method. However, we show next that the incremental cost is much less severe for the HMD method. The HMD method should be considered in future work aimed at computing time-dependent parameter estimation errors in the general case with spins. Here, we neglect the effects of spin precession, so that the angles $\Omega = (\theta_N, \phi_N, \theta_{NL}, \phi_{NL}, \alpha, \gamma(0))$ are decoupled and, since the signal does not depend on ϕ_N , there are only $N_\Omega \equiv 5$ independent (spin-unrelated) angles.

The larger the parameter space, the larger the sample size must be in a Monte Carlo computation. Let us assume that the sample size is chosen to be $N_{MC} = \left(N_{MC}^{(1)}\right)^d$, where $N_{MC}^{(1)}$ is the effective number of samples for a single parameter, and $d = N_\Omega$ when spanning the Ω -space only, $d = N_{SA}$ when spanning the spin-angle space only, and $d = N_\Omega + N_{SA}$ when spanning both.

To compute the time-independent matrices, we have to evaluate $N_p(2N_J + 1)$ independent $A_{j,a}$ coefficients for the full Ω -space and we have to compute the $N_p(2N_J + 1)(2N_{PN} + 1)$ independent $c_{jk,n}$ matrices over a $d = N_D + N_{SA}$ dimensional parameter space. (Here $N_D \equiv 2$ denotes the number of parameters on which the Doppler phase depends, (θ_N, γ) in eq. [26], using the fact that both c_{GW} and c_T also depend on all spin angles in eqs. [64,65].) In § IV C we have shown that there are $(2N_J + 1) \times (2N_{PN}^2 + 7N_{PN} + 6)$ independent integrals, where N_{PN} is the number of terms in the post-Newtonian expansion plus the Doppler phase. We have to compute these integrals for all spin angle orientations. Therefore, the computational cost scales as

$$\begin{aligned} \Delta T_{\text{HMD}}^{\text{no spin prec}} &= N_p(2N_J + 1) \left(N_{MC}^{(1)}\right)^{N_\Omega} \\ &+ (2N_J + 1)(2N_{PN}^2 + 7N_{PN} + 6)N_{t_f} \left(N_{MC}^{(1)}\right)^{N_{SA}} \\ &+ N_p(2N_J + 1)(2N_{PN} + 1) \left(N_{MC}^{(1)}\right)^{N_D + N_{SA}}. \end{aligned} \quad (77)$$

For the standard method, reiterating the argument in § V A, we get

$$\Delta T_{\text{standard}}^{\text{no spin prec}} = \frac{1}{2}N_p(N_p + 1)N_{\text{int}} \left(N_{MC}^{(1)}\right)^{N_\Omega + N_{SA}}, \quad (78)$$

where now $N_p = N_{p1} + N_{p0} + N_{\text{spin}} = 15$. Taking $N_{MC}^{(1)} = 10$, $N_{PN} = 4$, $N_J = 4$, $N_D = 2$, $N_\Omega = 5$, $N_{SA} = 4$, $N_{\text{int}} = 10^3$, $N_{t_f} = 100$ as a representative example, we find that the HMD method is computationally less expensive by a factor $\Delta T_{\text{standard}}^{\text{no spin prec}}/\Delta T_{\text{HMD}}^{\text{no spin prec}} = (1 \times 10^{14})/(2 \times 10^9) = 7 \times 10^4$, as compared to the standard method.

Note that the first term in eq. (77) corresponds to our original approximations in § IV B, i.e. no cross-correlations between p_{spin} and p_{slow} . This approximation indeed leads to much faster computations, since $\Delta T_{\text{standard}}^{\text{no spin prec}}/\Delta T_{\text{HMD}}^{\text{approximate}} = 9 \times 10^6$ for the same representative example.

C. Post-Newtonian Signal with Spin Precession

Accounting for spin precession, the $N_L = 2$ angular momentum angles, (θ_{NL}, ϕ_{NL}) , and the N_{SA} spin angles are now changing with time. In this case, one has to solve a differential equation for the evolution of these angles for each individual Monte Carlo realization. We assume that this can be computed independently of the Fisher matrices and that it would take $N_{DE}N_{t_f}$ computation time units to evaluate, at each of the N_{t_f} time instances, for each initial set of angles.

The HMD method costs

$$\begin{aligned} \Delta T_{\text{HMD}}^{\text{spin prec}} &= N_p(2N_J + 1) \left(N_{MC}^{(1)}\right)^{N_\Omega} \\ &+ N_p(2N_J + 1)(2N_{PN} + 2) \left(N_{MC}^{(1)}\right)^{N_D + N_{SA}} \\ &+ (2N_J + 1)(2N_{PN}^2 + 7N_{PN} + 6)N_{t_f} \left(N_{MC}^{(1)}\right)^{N_L + N_{SA}} \\ &+ N_{DE}N_{t_f} \left(N_{MC}^{(1)}\right)^{N_L + N_{SA}}, \end{aligned} \quad (79)$$

where the first term involves constructing the $A_{j,a}$ coefficient matrices, the second term involves constructing the $c_{jk,a}$ coefficient matrices, the third term is for computing all three time–evolution quantities $P(t_f)$ in eq. (75), and the fourth term is for solving the precession equations.

In the standard method, we need first to solve the precession evolution differential equation and then construct the Fisher matrix [22]. Following the assumptions made above, we estimate a cost

$$\Delta T_{\text{standard}}^{\text{spin prec}} = N_{DE} N_t \left(N_{MC}^{(1)} \right)^{N_L + N_{SA}} + \frac{1}{2} N_p (N_p + 1) N_{\text{int}} \left(N_{MC}^{(1)} \right)^{N_{\Omega} + N_{SA}} N_t. \quad (80)$$

Using $N_L = 2$ and all other parameters as in § V B, we find that the HMD method is computationally more efficient by a factor $\Delta T_{\text{standard}}^{\text{spin prec}} / \Delta T_{\text{HMD}}^{\text{spin prec}} = (2 \times 10^{14}) / (6 \times 10^{10}) = 2 \times 10^3$, as compared to the standard method. Since the (θ_{NL}, ϕ_{NL}) subspace could no longer be decoupled, the efficiency of the HMD method relative to the traditional method lost a factor of 30, as compared to the no spin precession case in § V B. Nevertheless, the computational advantage remains very substantial.

VI. RESULTS

Having described the HMD formalism in detail, we now apply it to build MC simulations aimed at studying how RMS source localization errors [43] evolve as a function of look–back time, t_f , before merger. The low computational cost of the HMD method allows us to survey simultaneously the dependencies on source sky position, SMBH masses and redshifts. We carry out MC calculations with 3×10^3 random samples for the angles $\cos \theta_N, \cos \theta_{NL}, \phi_{NL}, \alpha, \gamma(0)$. Several thousands values of M and z are considered, in the range $10^5 < M/M_{\odot} < 10^8$ and $0.1 \leq z \leq 7$. In addition, we ran specific MC calculations to study possible systematic effects with respect to the source sky position, by fixing θ_N and ϕ_N (on a grid of several hundred values) and varying all the other relevant angles.

In all of our computations, we calculate the error covariance matrix for $d_L, \theta_N, \phi_N, \theta_{NL}, \phi_{NL}$. Following Lang & Hughes [22], we calculated the major and minor axes of the 2D sky position uncertainty ellipsoid, $2a$ and $2b$, and the equivalent diameter, $\sqrt{4ab}$.

We have verified our HMD implementation and the general validity of our assumptions by comparing our results at ISCO with those of Lang & Hughes [22] (for $m_1 = m_2 = 10^5, 10^6, 10^7 M_{\odot}$ at $z = 1$ and $m_1 = m_2 = 10^5, 10^6 M_{\odot}$ at $z = 3$, in the no spin precession case). Depending on SMBH masses and redshifts, we found agreement at the 5–30% level for the mean errors on the luminosity distance, major axis, and minor axis. The small discrepancies may be due to differences in the set of assumptions made. Lang & Hughes [22] account for the small cross-correlations between the p_{slow} and $\{p_{\text{fast}}, p_{\text{spin}}\}$ parameters and they choose t_i to be uniformly distributed between merger time and LISA’s mission lifetime. Recently,

Lang & Hughes reported angular errors that are a factor of 2–3 lower [35], which are inconsistent with our results at this level. Nevertheless, these discrepancies are still small relative to the typical width of error distributions or to the systematic variations of mean errors with t_f , M , and z (from a factor of few to orders of magnitudes, see Fig. 2 below). This successful comparison justifies the use of the HMD method to study the dependence of localization errors on look–back time.

A. Time dependence of source localization errors

We calculate the variation with look–back time, t_f , of the distribution of marginalized parameter errors for a range of values of $M, z, \theta_N, \theta_{NL}, \phi_{NL}, \alpha, \gamma(0)$. Figure 2 shows results for random angles and $m_1 = m_2 = 10^6 M_{\odot}$, at $z = 1$.

The top panel shows the luminosity distance error, δd_L , while the bottom panel describes the equivalent diameter, $2\sqrt{ab}$, of the sky position error ellipsoid with minor and major axes a and b . The figure displays results for three separate cumulative probability distribution levels, 90%, 50%, 10%, so that 10% refers to the best 10% of all events, as sampled by the random distribution of angular parameters. The evolution of errors scales steeply with look–back time for $t_f \gtrsim 40$ days. In this regime, the improvement of errors is proportional to $(S/N)^{-1}$. For smaller look–back times, errors stop improving in the “worst” (90% level) case, improve with a much shallower slope than $(S/N)^{-1}$ for the “typical” (50% level) case, and keep improving close to the $(S/N)^{-1}$ scaling in the “best” case (10% level among the realizations of fiducial angular parameters). Although Figure 2 shows only the equivalent diameter of the 2D sky localization error ellipsoid, we have also computed the distribution of the minor and major axes. We find that $a \approx b \approx \sqrt{ab}$ initially (i.e. the ellipsoid is circular), but the geometry changes significantly during the last two weeks to merger. For example, in the typical case, the major axis a stops improving at late times, while the minor axis b maintains a steep evolution. Therefore the eccentricity of the 2D angular error ellipsoid changes quickly with look–back time. This is important because large eccentricities can play a role in assessing observational strategies for EM counterpart searches [18].

To map possible systematic effects with respect to source sky position, we carried out MC computations with random $(\cos \theta_{NL}, \phi_{NL}, \alpha)$ angles (sample size $N_{MC} = 3 \times 10^3$) but fixed source sky latitude and longitude relative to the detector (θ_N, γ) , for $m_1 = m_2 = 10^6 M_{\odot}$ and $z = 1$. We find no systematic trends with sky position for δd_L , for any value of the look–back time, t_f . Neither do we find systematic trends with sky position for the distributions of minor and major axes of the angular ellipsoid, for any value of the look–back time, t_f , as long as θ_N is not along the equator. The case of equatorial sources, with $\theta_N \approx 90^\circ$ and a short look–back time t_f before merger, is the only nontrivial one we have identified. In that case, we find a minor systematic trend with γ longitude. The error distributions shift periodically up and down, relative to the average, when changing γ from 0 to 2π .

In addition, to map dependencies with mass–redshift–look–

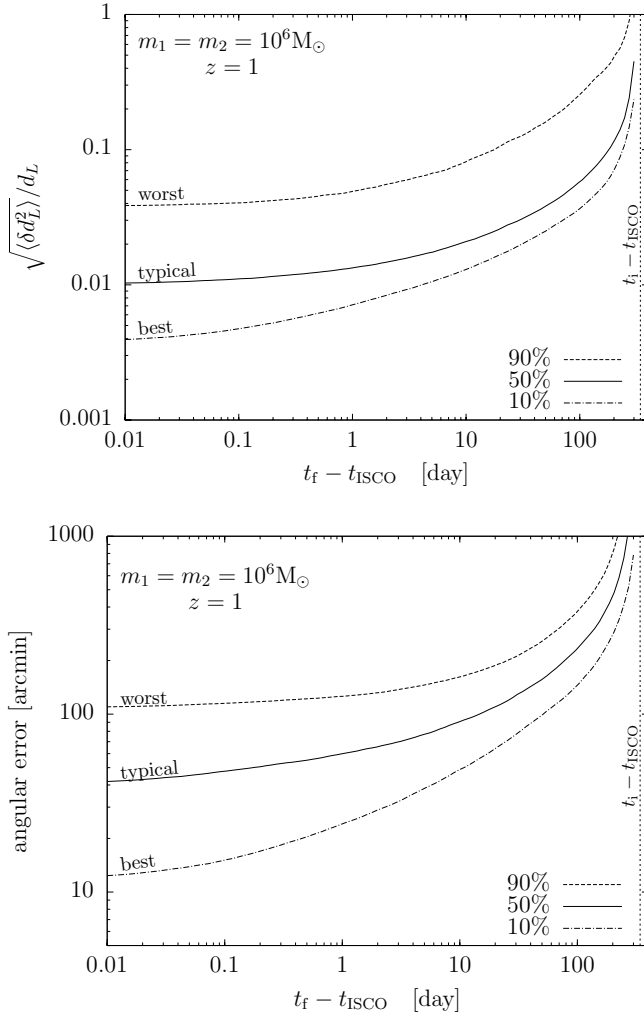


FIG. 2: Evolution with pre-ISCO look-back time, t_f , of LISA source localization errors, for $M = 2 \times 10^6 M_\odot$ and $z = 1$. The top panel shows luminosity distance errors and the bottom panel shows sky position angular errors (equivalent diameter, $2\sqrt{ab}$, of the error ellipsoid). Best, typical, and worst cases for random orientation events represent the 10%, 50%, and 90% levels of cumulative error distributions, respectively. Errors for worst case events effectively stop improving at a finite time before ISCO, even though the signal-to-noise ratio accumulates quickly at late times. Errors for best case events (especially the minor axis) follow the signal-to-noise ratio until the final few hours before merger.

back time of localization errors, we carried out MC computations with arbitrary $(\cos\theta_N, \cos\theta_{NL}, \phi_{NL}, \alpha, \gamma)$ angles, with sample size $N_{MC} = 3 \times 10^3$, for several thousand pairs of (M, z) values. We find that the evolution with look-back time of error distributions depends sensitively, and in a complicated way, on the mass-redshift parameters. Generally, localization errors increase with redshift. Firstly, the S/N is approximately proportional to the instantaneous value $\sigma(t_{\text{ISCO}}) \propto \eta^{3/4} [(1+z)M]^{5/4} / d_L(z) S_n(f_{\text{ISCO}})^{-1}$ (eq. [56]) and, secondly, the beginning-of-observation time scales as $t_i \propto \eta^{-1} [(1+z)M]^{-5/3}$ (eq. [2]). For $(1+z)M < 4 \times 10^6 M_\odot$, the total observation time can exceed one year and the second effect

is unimportant. We further describe mass-redshift dependencies below, in § VI B, in relation to advance warning times for targeted electromagnetic counterpart searches.

The results on localization errors from our extensive exploration of the parameter space of potential LISA sources can be summarized as follows:

1. Probability distributions

- The error distributions for δd_L , $2a$, and $2b$ all have long tails: 1%–99% cumulative probability levels are separated by factors of ~ 100 , while the 10%–90% levels are separated by factors of ~ 10 .
- The δd_L distribution is skewed, with a median closer to the best case, a median smaller than the mean, even on a logarithmic scale. On the other hand, sky localization error distributions are roughly symmetric on a logarithmic scale.

2. Fiducial parameter dependencies

- The δd_L errors are roughly independent of fiducial angles throughout the observation.
- For non-equatorial sources, the distribution of sky localization errors, $(2a, 2b)$, is independent of sky position, i.e. the distribution does not have a systematic dependence on θ_N and $\gamma \equiv \Phi - \phi_N$ (for random $\alpha, \theta_{NL}, \phi_{NL}$).
- There is a small systematic trend with γ for equatorial sources.
- There is a complicated dependence of sky localization errors on \mathcal{M} , z , and look-back time t_f . For $(1+z)(\eta/0.25)^{3/5} M \lesssim 4 \times 10^6$, and long observation times, errors scale with $(S/N)^{-1} \approx [(1+z)\mathcal{M}]^{5/4} d_L(z)^{-1} S_n(f_{\text{ISCO}}(M, z))^{-1} f_a(t_f)$, where $f_a(t_f)$ is the t_f -scaling shown in Fig 2. For larger redshifted masses, the scaling has a complicated structure in the M, z, t_f space that we did not analyze in detail (but see eq. (A10) in the Appendix for scalings in terms of t_i and t_f).

3. Time dependence

- Luminosity distance and sky localization errors roughly scale with $(S/N)^{-1}$ until 2 weeks before ISCO.
- For the luminosity distance δd_L and the major axis $2a$, there is little improvement within the last week before ISCO for the typical to worst cases (i.e. 50%–90% levels of cumulative error distributions).
- For the minor axis $2b$, only the worst case events stop improving within the last week. The typical to best cases continuously improve until the last hour.
- The eccentricity of the sky localization error ellipsoid changes with time during the first and last two weeks of observation. The eccentricities are

smaller in between these two time intervals. For a detailed discussion of the eccentricity and its impact on counterpart searches, see Ref. [18].

- For the luminosity distance δd_L , the relative width of error distributions does not change during observation and variations in the difference between the 90% and 10% levels of the cumulative distributions do not exceed 10%, except for the initial weeks, when the distribution is much more spread out.
- For the sky localization errors, the width of error distributions increases during the final two weeks of observation, by a factor ~ 2 for the major axis and a factor ~ 4 for the minor axis.

B. Advance warning times for EM searches

From the astronomical point of view, being able to identify with confidence, prior to merger, a small enough region in the sky where any prompt electromagnetic (EM) counterpart to a LISA inspiral event would be located, is of great interest. With sufficient “warning time,” it would then be possible to trigger efficient searches for EM counterparts as the merger proceeds and during the most energetic coalescence phase. In particular, an efficient strategy to catch such a prompt EM counterpart would be to continuously monitor with a wide-field instrument a single field-of-view (FOV), through coalescence and beyond. Astronomical strategies for EM counterpart searches are the focus of a second paper in this series [18].

Given an angular scale, θ_{FOV} , corresponding to the hypothetical FOV of a specific astronomical instrument, it is thus of considerable interest to determine the value of the look-back time t_l at which the major axis, minor axis or equivalent diameter of the sky localization error ellipsoid provided by LISA just reach the relevant θ_{FOV} scale. This would allow one to trigger an efficient search for EM counterparts, in a well defined region of the sky that can be monitored. We will hereafter refer to this time as the *advance warning time*. Note that it is important to differentiate the sizes of the major and minor axes of the angular error ellipsoid in this context because the eccentricity can be large, and thus important in assessing optimal strategies for EM counterpart searches [18].

For definiteness, we evaluate advance warning times for angular diameters $\theta_{\text{FOV}} = 1^\circ$ and 3.57° here but generalizations to other θ_{FOV} values are obviously possible. The choice of the latter figure is motivated by the 10 deg^2 FOV proposed for the future Large Synoptic Survey Telescope, or LSST [36]. Figure 3 shows advance warning times for a fixed source redshift at $z = 1$ and various values of the total SMBH mass, M . Figure 4 shows similar results for various source redshifts, at a fixed value of $M = 2 \times 10^6 M_\odot$.

In each case, we consider equal mass SMBH binaries and a maximum observation time of 1 yr (or lower if set by the GW noise frequency wall at 0.03 mHz). Each panel in Figs. 3 and 4 shows the values of advance warning times at which

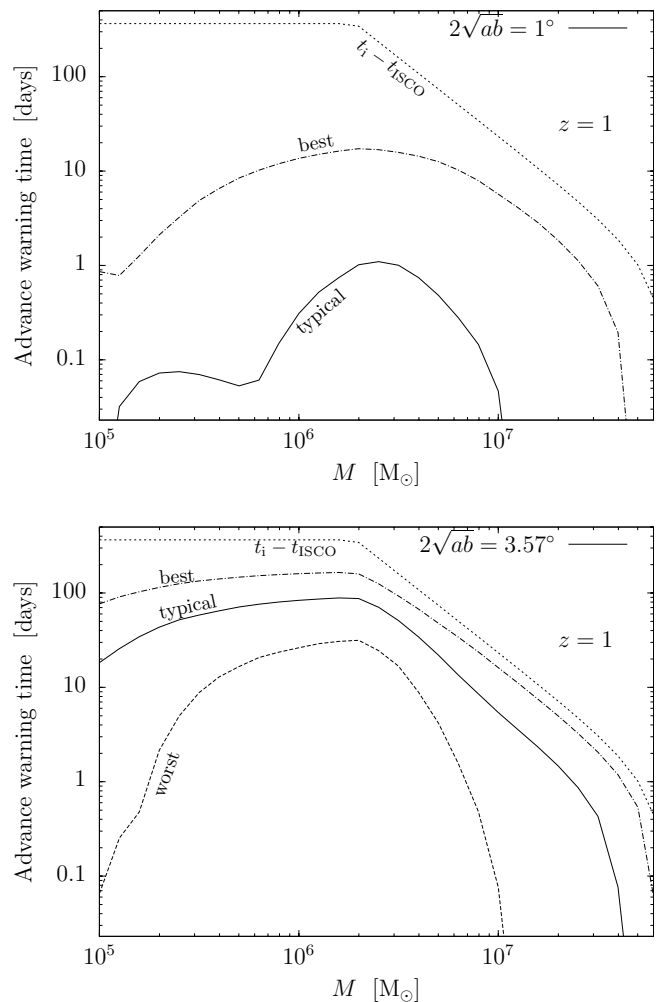


FIG. 3: Advance warning times (in days) for equal mass binary inspirals at $z = 1$, as a function of total mass, M (in solar units). Best, typical, and worst cases refer to 10%, 50%, and 90% levels of cumulative error distributions for random orientation events, as before. The advance warning times shown correspond to the values of look-back times when the equivalent diameter, $2\sqrt{ab}$, of the error ellipsoid first reaches 1° (top panel) or an LSST-equivalent field-of-view (3.57° , bottom panel). In the top panel, the worst case events are not shown because angular errors are too large even at ISCO. For the largest mass SMBHs, the maximum observation time (and thus t_l) is below one year.

the equivalent diameter $2\sqrt{ab}$ of the localization error ellipsoid drops below the reference θ_{FOV} value. For each case, we show results for cumulative error distribution levels of 10%, 50%, and 90%, labeled “best”, “typical,” and “worst” cases, as before. Figure 3 shows that LISA can localize on the sky events at $z = 1$ to within an LSST FOV at least one month ahead of merger, for 50% of events with masses $2 \times 10^5 M_\odot \leq M \leq 3 \times 10^6 M_\odot$, and at least 4 days ahead of merger for 90% of events in the same mass range. Figure 4 shows that advance warning times decrease with redshift, leaving at least 1 day ahead of merger for 50% of events with $M = 2 \times 10^6 M_\odot$, as long as $z \lesssim 1$ for $\theta_{\text{FOV}} = 1^\circ$ and as

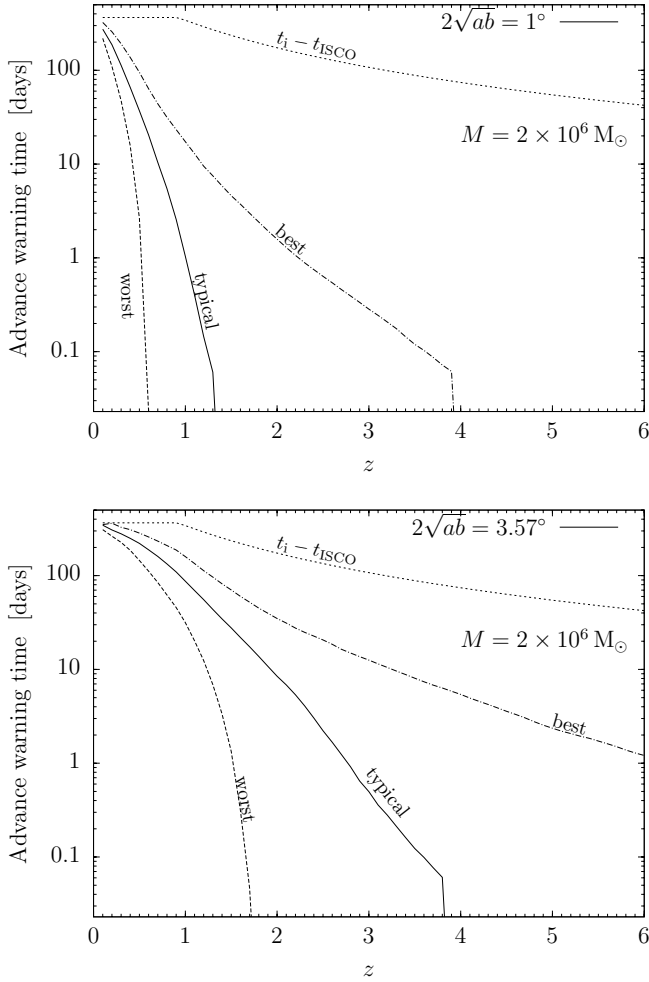


FIG. 4: Same as Fig. 3, for a fixed total mass $M = 2 \times 10^6 M_\odot$ but various values of the source redshift, z .

long as $z \lesssim 3$ for an LSST FOV. For events with this mass scale and the LSST FOV, there is a 10% chance that a 1 day advance warning is possible up to $z \sim 5$ –6.

Figures 3 and 4 display advance warning times for single one dimensional slices of the full (M, z) space of potential LISA events. With the HMD method, however, it is possible to explore the entire parameter space of SMBH inspirals by repeating the calculation on a dense grid of (M, z) values. We construct a uniform grid in the $(\log M, z)$ plane, with $\Delta z = 0.1$ and $\Delta \log M = 0.1$, and perform MC computations with 3×10^3 randomly oriented angles for each grid element. As a result, we obtain a complete description of the time evolution of sky localization errors in the large parameter space of potential LISA sources. Figure 5 displays advance warning time contours from this extensive MC calculation, for typical (50%) and best case (10%) events, adopting the LSST FOV as a reference.

Advance warning time contours are logarithmically spaced, with solid-red contours every decade and a thick red line highlighting the 10 day contour. Since advance warning times were computed on a finite mesh, contour levels for arbitrary M

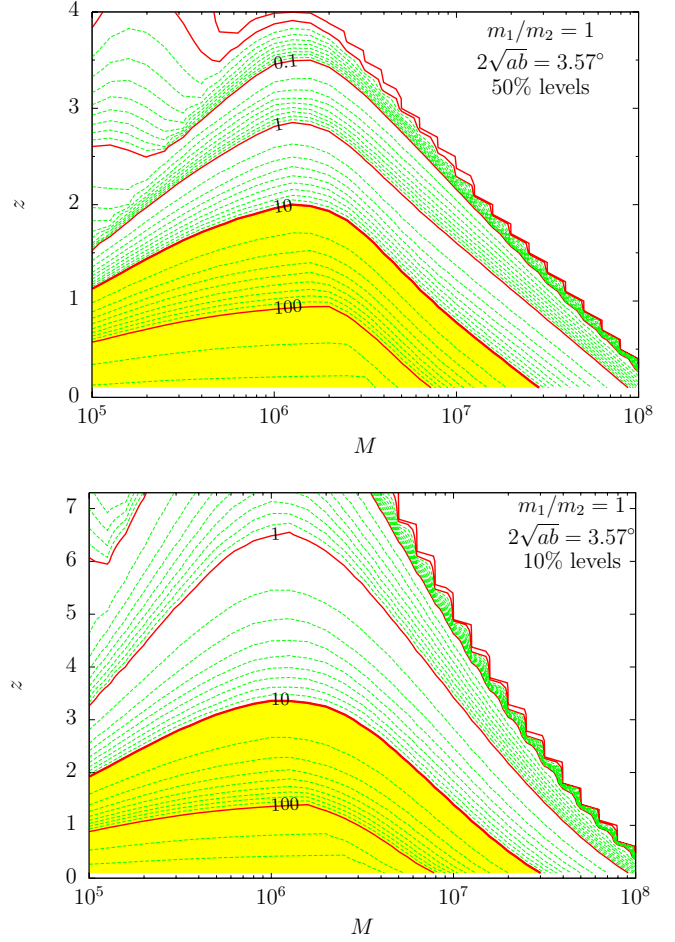


FIG. 5: Contours of advance warning times in the total mass (M) and redshift (z) plane with SMBH mass ratio $m_1/m_2 = 1$. The contours trace the look-back times at which the equivalent radius ($2\sqrt{ab}$) of the localization error ellipsoid first reach an LSST-equivalent field-of-view (3.57°). The contours correspond to the 50% (*top*) and 10% (*bottom*) level of cumulative distributions for random orientation events. The contours are logarithmically spaced in days and 10 days is highlighted with a thick line.

and z values were obtained by interpolation. Our interpolated mesh is smooth if $t_f \sim 0.1$ day, but it gets edgy for short advance warning time approaching ISCO. Figure 5 shows that a 10 day advance warning is possible with a unique LSST-type pointing for a large range of masses and source redshifts, up to $M \sim 3 \times 10^7 M_\odot$ and $z \sim 1.9$. The bottom shows how far the advance warning concept can be stretched, by focusing on the 10% best cases of random orientation events. In this case a 10 day advance warning is possible up to $z \sim 3$ for masses around $M \sim 10^6 M_\odot$. Note that, in both cases, allowing for a warning of just one day would extend considerably the range of masses and redshifts for which a unique LSST-type pointing is sufficient.

These results can also be generalized to unequal-mass SMBH binaries. At fixed total mass, M , an unequal-mass binary has an instantaneous signal-to-noise ratio that is reduced because of a lower η value, but it also has a total observation

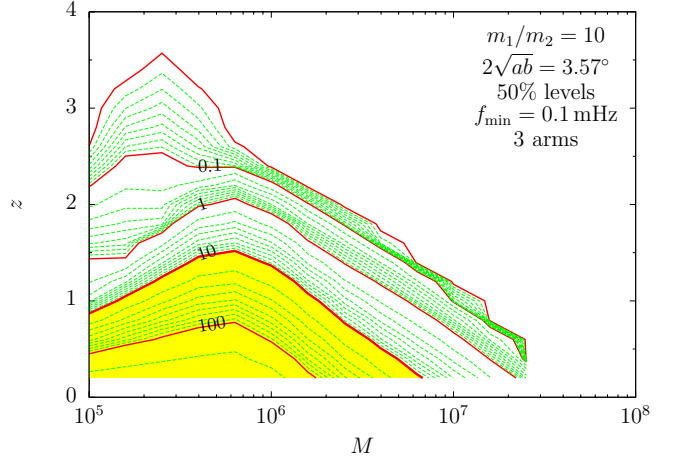
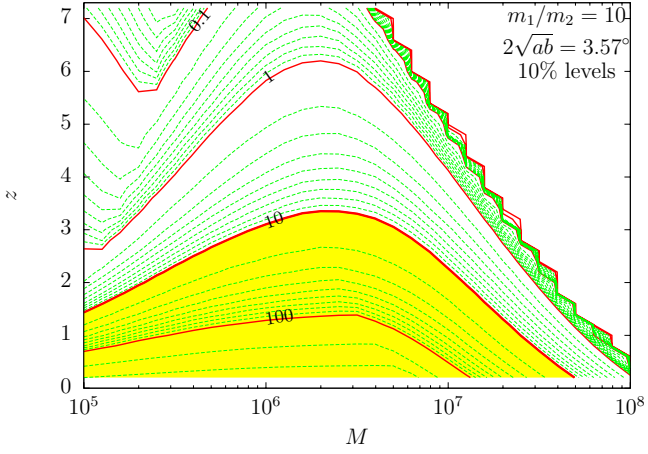
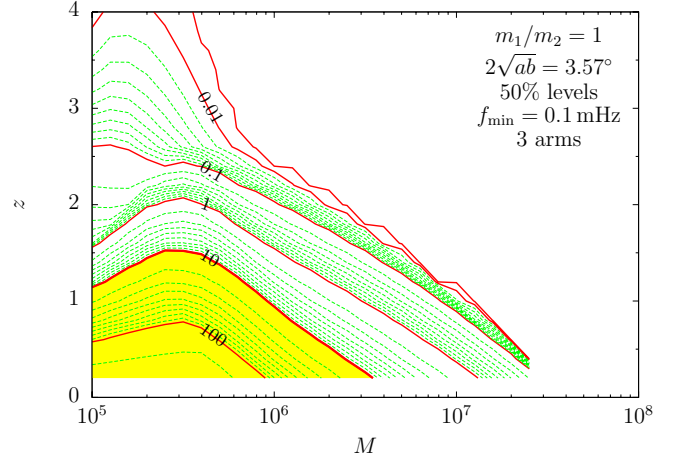
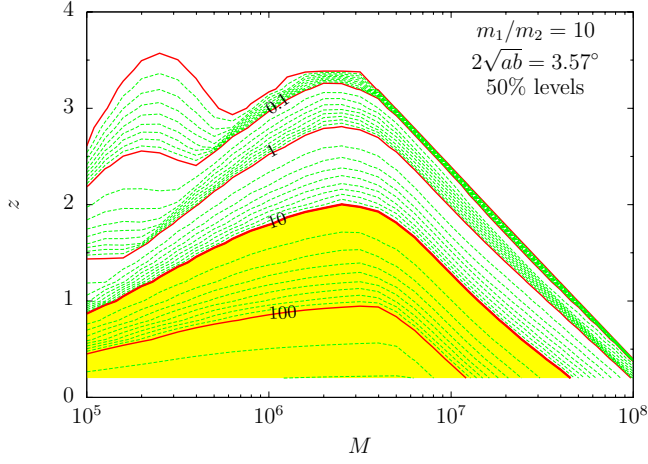


FIG. 6: Same as Fig. 5, except for a SMBH mass ratio of $m_1/m_2 = 10$.

FIG. 7: Same as the top panels of Figs. 5 and 6, except for a degraded minimum detector frequency of $f_{\min} = 0.1$ mHz.

time that is potentially longer. Localization errors for unequal-mass inspiral events with total observation times longer than a month (i.e. with $\eta_{0.25}^{2/5}(1+z)M < 1.8 \times 10^7 M_\odot$) are degraded relative to the equal-mass cases discussed so far. For larger total mass, however, the worsening of errors is mitigated, or even reverted, relative to the equal mass case, thanks to the longer observation time. The error ellipsoid also becomes less eccentric thanks to this additional observation time. Figure 6 summarizes results on advance warning times from the same MC computations as in Fig. 5, but this time for unequal-mass SMBH binaries with mass ratio $m_1/m_2 = 10$. Despite a systematic degradation in advance warning times (especially noticeable at low M values), the main effect of introducing a mass ratio $m_1/m_2 = 10$ is to shift advance warning time contours to somewhat larger values of total mass, M . Our main conclusions on advance warning times are not very strongly affected by the inequality of mass components in the population of SMBH binaries considered.

Finally, it is important to understand how sensitive the results are to the LISA detector characteristics. In particular, we examined how advance warning times are affected by increasing the minimum frequency noise wall or by losing one of the arms of the 3-arm constellation. Figure 7 displays results

for $f_{\min} = 10^{-4}$ Hz, for $m_1/m_2 = 1$ and $m_1/m_2 = 10$. Increasing f_{\min} mostly reduces the total observation time for high mass inspirals ($t_1 \sim f_{\min}^{-8/3} M_z^{-5/3}$; see eq. [2]) and reduces the signal-to-noise ratio by a small factor. As a result, the advance warning time contours primarily shift in the (M, z) plane in the direction of smaller total masses by a factor of ~ 7 , and secondly shift moderately (30–50%) to smaller redshifts. Losing one LISA arm (i.e. using only one of the two interferometers) most importantly removes the ability of the second datastream to break correlations in localization errors and also reduces the signal-to-noise by a small factor. As a result errors do not improve much during the last ~ 10 days before merger. Compared to the case with two interferometers, contours representing an advance warning of less than 10 days are shifted to significantly smaller z (especially for the minor axis of the sky localization ellipsoid), close to the 10-day contour, but warning times beyond ~ 10 days worsen only moderately. We conclude that even if $f_{\min} = 0.1$ mHz or if only one of the two interferometers is used, LISA still admits 10-day advance localizations for a broad range of masses and redshifts, between $10^5 \lesssim M \lesssim 2 \times 10^6$ and $z \lesssim 1$.

VII. DISCUSSION

We have introduced a novel technique, the HMD method, to compute time-dependent GW inspiral signals for LISA. The method relies on the fact that LISA’s orbital motion induces a modulation on timescales that are long relative to the inspiral GW frequency. Since this modulation is periodic, with a fundamental frequency of f_{\oplus} , it can be expanded in a discrete Fourier sum. In the HMD formalism, dependencies on sky position, orbital angular momentum orientation, and detector orientation in the LISA signal are inscribed in time-independent coefficients, while time-dependent basis functions are independent of these angles. This decomposition helps to reduce the computational cost of Monte Carlo simulations exploring the time-dependence of source localization errors by orders of magnitude.

Moreover, the HMD method can be used in conjunction with plausible approximations to further decrease the computational cost of explorations of the parameter space of localization errors for LISA inspiral events. In our analysis, we identified two different characteristic frequency constituents of the signal: the high frequency restricted post-Newtonian GW inspiral waveform and the low frequency amplitude modulation resulting from the detector’s orbital motion. In the HMD method, these two components separate and parameters that depend only on the low frequency modulation (such as the source position and the orbital angular momentum angle) can be estimated independently of the other source parameters determined by the high frequency carrier signal. Our working assumption was that cross-correlations among these two sets of parameters must be much smaller than parameter correlations within either set. This hypothesis is valid very generally in the no spin limit for SMBHs, as shown by full Fisher matrix calculations without such approximations for general relativity [7] and alternative theories of gravity [13].

In order to further examine the validity of our assumptions and the ultimate boundaries of our models, and to understand our results, we have constructed illustrative toy models that we now describe in some detail. These toy models show that the separation of parameters into various subsets associated with different characteristic frequencies of the signal is a rather general property, which turns out to be an efficient way of reducing the computational cost of error estimations for the LISA problem.

A. Simple toy models

In this section, we discuss very simple toy models which capture the essence of the problem posed by the time-evolution of parameter error estimations. We then use these models to answer general questions on the LISA-specific parameter estimation problem.

Our harmonic decomposition technique is based on the simple intuition that the angular information can be deduced from the slow periodic modulation of the high frequency GW waveform. In § III, we have shown that modulation harmonics with frequencies larger than $4f_{\oplus}$ vanish exactly. Here, we discuss

the general properties of such a modulation. In the case of LISA, the high frequency carrier signal has an effective, cycle-averaged signal-to-noise ratio which monotonically increases with time as SMBH binaries approach merger. To mimic such events, we also assume in all of our toy models that the instantaneous signal-to-noise ratio continuously improves throughout the observation.

We seek answers to the following questions:

1. How do mean errors evolve during the final days of observation?

On the one hand, in standard angle-averaged treatments (e.g. [5, 13, 15]), an evolution of errors with the inverse of the signal-to-noise ratio is generally assumed. This would suggest a large improvement during the last day of inspiral. On the other hand, the slow modulation picture suggests just the contrary: not much improvement is expected at late times when there is effectively very little modulation (Finn & Larson 2005, private communication).

2. Does the introduction of additional high frequency components in the signal have any effect on the estimations of low frequency parameters?

In the GW context, it is of general interest to determine under what circumstances additional high frequency signal components, such as higher order post-Newtonian corrections or spin-induced effects, remain decoupled from the determination of angular and distance information based on the signal amplitude modulation.

3. Are there combinations of signal parameters for which errors improve rapidly in the last days of observation? If so, what are these combinations? What determines how many such rapidly-improving combinations there will be?

If the distance d_L correlates with the angles, then in principle the volume of the 3D error box can be much smaller than the product of the marginalized errors $\delta\Omega \times \delta d_L$ would imply. Unfortunately, in practice, this is unlikely to help to reduce the number of false counterparts, because the δz error will be dominated by weak lensing [15].

4. How does the width of parameter error distributions evolve with time? Are the best and worst cases approaching the typical case prior to the final days of observation? How do we expect the eccentricity of localization error ellipsoid to evolve with time for LISA?

Here, we restrict our discussion to a brief summary of our findings and direct the reader to Appendix A for further details on these toy models.

The parameter estimation uncertainties are defined by the correlation error matrix. For N_p parameters, this defines an N_p -dimensional error-ellipsoid in the N_p -dimensional parameter space, where parameters are constrained at a given confidence level. Marginalized errors for a given parameter are

then related to the projection of this ellipsoid on the basis vector corresponding to that parameter. Since the principal axes of this error ellipsoid are generally not aligned with the original parameters, the marginalized errors can be substantial even if the volume of the error ellipsoid is close to zero. This happens if the ellipsoid is very “thin” but has a large size in at least one direction. Diagonal elements of the correlation matrix provide marginalized squared errors on the parameters, while eigenvalues provide squared errors along the principal axes.

We consider three versions of toy signals to understand how a particular harmonic mode contributes to the time-dependence of parameter uncertainties and to find answers to Questions 1–4 above. We start with the simplest toy model and refine this model by adding more details and complexity in the successive models. In each case, we discuss general implications for the model under consideration.

1. Basic toy model

In our basic toy model, we assume that the true signal is comprised of a constant carrier signal, which is modulated by a single known-frequency cosine, f_{\oplus} :

$$h(t) = c_0 + c_1 \cos(2\pi f_{\oplus} t), \quad (81)$$

where c_0 and c_1 are unknown parameters to be estimated. We assume that the noise level is rapidly decreasing during the observation, mimicking the gradual increase in the instantaneous signal-to-noise ratio for LISA inspiral signals. The contradictory statements made in relation to Question 1 above can be explored with this model. We find that marginalized parameter errors scale with the signal-to-noise ratio far away from merger (i.e. $t_f \gtrsim 0.1 f_{\oplus}^{-1}$) but they quickly converge to their final values at late times, even though the signal-to-noise ratio keeps accumulating. It is possible to derive analytical formulae for the evolution of parameter errors to fully characterize this behavior (see Appendix A). We find that, even though the error ellipse rapidly decreases in volume, as the inverse of the signal-to-noise ratio near merger, the error ellipse only shrinks along one of its dimensions, the semi-minor axis, so that a non-negligible residual uncertainty remains in the orthogonal subspace (e.g. along the semi-major axis). This residual uncertainty carries over to final marginalized errors for both parameters. Therefore, this first toy model verifies the second option in relation to Question 1.: there is no late improvement because there is very little effective signal modulation, making the signal-to-noise argument largely irrelevant. However, we find below that this model does not carry some essential features of the LISA signal which modify somewhat our final answer to Question 1 (see final toy model below).

2. Second toy model

In our toy second model, we modify the single frequency signal by postulating two pairs of unknown amplitudes and

phases for two different *a priori* known frequencies, satisfying $f_2 \gg f_1$, which modulate an otherwise constant signal:

$$h(t) = c_0 + s_1 \sin(2\pi f_1 t) + c_1 \cos(2\pi f_1 t) + s_{10} \sin(2\pi f_2 t) + c_{10} \cos(2\pi f_2 t). \quad (82)$$

The number of unknowns in this model is five: c_0, s_1, c_1, s_{10} and c_{10} are the coefficients of the functions 1, $\sin(2\pi f_1 t)$, $\cos(2\pi f_1 t)$, $\sin(2\pi f_2 t)$, and $\cos(2\pi f_2 t)$. Again, we assume that the noise decreases quickly with time before merger, at $t = 0$. This model is designed to answer our Question 2 above. In this case, we find that parameter errors are correlated only with unique frequency components and the constant signal, all the way to $t_f \gtrsim 0.1 f_2^{-1}$. The model thus demonstrates how components associated with very different variation timescales can decouple from each other. Moreover, as for the first toy model, we find that marginalized parameter errors effectively stop improving past a finite time before merger (Question 1), which is simply related to their respective frequencies. As a result, a nonzero residual error remains again, even though the signal-to-noise ratio continuously increases near merger.

3. Final toy model

In our final toy model, we insert a few additional features essential to a realistic LISA data-stream. Firstly, we assume 5 low-frequency harmonics, 1, $\sin(2\pi f_1 t)$, $\cos(2\pi f_1 t)$, $\sin(4\pi f_1 t)$, $\cos(4\pi f_1 t)$, with unknown amplitudes. We also include a high frequency carrier signal with known frequency, $f_2 \gg f_1$, but unknown amplitudes in $\sin(2\pi f_2 t)$ and $\cos(2\pi f_2 t)$, for a total of seven free parameters. Secondly, we note that the LISA system is equivalent to two orthogonal arm interferometers with both detectors measuring polarization phases simultaneously (which correspond to the real and imaginary parts of the amplitude modulation, § IV). Therefore, the signal is comprised of 4 simultaneous data-streams. We incorporate this feature by assuming 4 measurements (i.e. 4 corresponding Fisher matrices) of the signal with 4 given phase shifts ($\varphi_i^{s1}, \varphi_i^{c1}, \varphi_i^{s2}, \varphi_i^{c2}$; $1 \leq i \leq 4$) so that

$$h(t) = c_0 + s_1 \sin(2\pi f_1 t + \varphi_i^{s1}) + c_1 \cos(2\pi f_1 t + \varphi_i^{c1}) + s_2 \sin(2\pi f_1 t + \varphi_i^{s2}) + c_2 \cos(2\pi f_1 t + \varphi_i^{c2}) + s_{10} \sin(2\pi f_2 t) + c_{10} \cos(2\pi f_2 t), \quad (83)$$

In this case, we find that 4 principal components improve quickly at late times. As in our second toy model, the high frequency parameters decouple from the slow frequency ones, except at very late times when $t_f \gtrsim 0.1 f_2^{-1}$.

This final toy model allows us to answer all of Questions 1–4 as follows.

- Answer 1: Four out of 5 slow principal components of the error ellipsoid are quickly improving with time, while one of them stops improving at $t_f \lesssim 0.1 f_1^{-1}$. Therefore, any parameter with a large projection along this one poor principal component will stop improving,

while parameters nearly orthogonal to it will keep improving quickly. Thus, both statements made in relation to Question 1 above can in fact be correct, depending on the connection between a given parameter and the poor principal component. Typically, we expect marginalized parameter uncertainties to evolve as $(S/N)^{-1}$ for $t_f \gtrsim 0.1 f_1^{-1}$. For smaller t_f values, closer to merger, they would continue to improve, albeit with a shallower slope.

- Answer 2: We find that the introduction of additional high frequency components does not change the evolution of original parameter estimations as long as the time-to-merger is larger than a fraction of the time period of the additional high frequency components.
- Answer 3: As the signal-to-noise ratio increases quickly at late times, rapidly evolving parameter error combinations are given by the principal components of the error ellipsoid corresponding to the final situation at merger. With 4 data-streams, there are 4 such best principal components. Analogously, for the LISA amplitude modulation given by eq. (25), we expect that the 2 polarization phases for the 2 beam patterns at ISCO can be best determined: $(1 + \cos^2 \theta_{NL}) F_+^{I,II}(\Omega_{\text{ISCO}})$ and $\cos \theta_{NL} F_\times^{I,II}(\Omega_{\text{ISCO}})$. (In terms of ecliptic angular variables, these are the real and imaginary parts of the combination given by eq. (34).)
- Answer 4: The widths of error distributions for slow parameters do not change significantly as long as $t_f \gtrsim 0.1 f_1^{-1}$. During this final stretch of time before merger, however, one of the principal components stops improving and the major axis of the error ellipsoid freezes. Since the physical parameters can be considered to be randomly oriented with respect to the ellipsoid axes, distributions of marginalized errors suddenly start broadening for $t_f \lesssim 0.1 f_1^{-1}$, with a worst case relative orientation leading to very little improvement and a best case relative orientation corresponding to a scaling with $(S/N)^{-1}$.

B. Implications for LISA

These simple toy models offer a general interpretation of the time dependence of LISA’s parameter estimation errors for source localization. The LISA data stream is described by $N_{p1} = 5$ physical parameters, p_{slow} , which are not the harmonic coefficients themselves but determine these coefficients, g_j (or conversely, the mode expansion coefficients g_j determine the physical parameters p_{slow} ; see § III). Neglecting Doppler phase and spin precession effects, $2J_{\text{max}} + 1 = 9$ modes determine the signal by eqs. (41,42). In principle, any $N_{p1} = 5$ of the g_j mode amplitudes uniquely determine the physical parameters, p_{slow} . However, in the presence of noise, each of these modes are uncertain and the combination of all modes helps in reducing the estimation errors of the p_{slow} variables.

The key implication of our toy models for LISA is that the estimation of low frequency g_j modes with low $|j|$ are effectively decoupled from the high frequency signal, unless the merger is within ~ 0.1 times the cycle time of the fast-oscillating signal. We have shown that the HMD of the orbital modulation consists purely of low-order harmonics, with $|j| \leq 4$. In comparison, the high frequency GW phase has a much higher frequency, corresponding to $j > 1000$, and this high frequency signal’s cycle time is greater than the time to merger throughout $t \gtrsim t_{\text{ISCO}}$. Hence, physical parameters p_{slow} will remain decoupled from parameters p_{fast} , all the way to ISCO. This finding is independent of details of the waveform and the modulation, in agreement with the results of Ref. [13] which show that decoupling occurs independently of the details of the $h_c(t)$ signals, including the modified inspiral waveforms of alternative theories of gravity. In terms of post-Newtonian expansions, only terms above second order have cycle times as large as the cycle time of the amplitude modulation. These terms are responsible for the small cross-correlations of the two sets of parameters found by Ref. [7].

We have not considered spin precession effects, but Vecchio [17] and Lang & Hughes [22] find that spin precession effects can help improve the final localization errors by a factor of ~ 3 . Spin precession cycle times decrease continuously, become of order a few days or less during the last week prior to merger, and of order hours during the last day of inspiral. Therefore, according to our simple models, we expect spin precession effects to improve the source parameter estimation errors especially during the final two weeks before ISCO. During that period of time, in the absence of spin effects, parameter uncertainties (especially the sky position major axis and the luminosity distance) cease to improve when using only the amplitude modulation.

The best-determined parameters at ISCO are, approximately, the independent detector outputs at ISCO, i.e. the real and imaginary parts of $h_1^{I,II}(p_1)$: $d_L^{-1}(1 + \cos^2 \theta_{NL}) F_+^{I,II}(\Omega)$ and $d_L^{-1} \cos \theta_{NL} F_\times^{I,II}(\Omega)$ (see Appendix A4). These are the 4 independent combinations of 5 physical parameters p_1 which correspond to the eigenvectors of the error covariance matrix following the steep evolution $\propto (S/N)^{-1}$ all the way to ISCO. We refer to the fifth independent combination, which is orthogonal to these best eigenvectors, the “worst” eigenvector, since for this combination, the evolution ceases to improve as $(S/N)^{-1}$ within $\sim 0.1 \times$ (amplitude modulation cycle time) of merger. It is straightforward to obtain this worst combination explicitly by using the 4 other eigenvectors and Gram-Schmidt orthogonalization (but we have not done this in practice). Since the highest frequency harmonic of the slow modulation is for $j = 4$, the corresponding cycle time is $\text{yr}/4$. Thus, we expect errors will stop improving roughly 1–2 weeks prior to merger. Distributions of errors will quickly broaden during these final stages of observation before ISCO. Simply scaling errors with $(S/N)^{-1}$, as in the angle-averaged formalism (e.g. [13, 15]), is acceptable if one studies the evolution of parameter errors at $t_f \gtrsim 2$ weeks, or if one only focuses on the best case parameter combinations. In general, the exponent in the (S/N) scaling decreases as one approaches merger time depending on how close the particular combination of angles

considered is to the worst combination.

Our findings for the eccentricity evolution of LISA’s sky localization error ellipsoid can also be understood with the simple toy models. In fact, we found this behavior to be expected for any model signal with relative instantaneous signal amplitude increasing quickly with time, e.g. $t^{-\alpha}$, $\alpha \gtrsim 2$. In this case, the principal axes of the general parameter error ellipsoid separate near $t_f = 0$. There are a limited number of principal errors which rapidly decrease to zero near $t_f = 0$, while others “freeze out” at a time related to a fraction of the cycle time of the particular waveform ($\sim 0.1 T_{\text{cycle}}$ if $t_i > T_{\text{cycle}}$). For LISA, there are 5 variable parameters, $p_{\text{slow}} = (d_L, \theta_N, \phi_N, \theta_L, \phi_L)$, and estimation uncertainties of 4 combinations of these parameters, $d_L^{-1}(1 + \cos^2 \theta_{NL})F_+^{I,II}(\Omega)$, $d_L^{-1} \cos \theta_{NL}F_\times^{I,II}(\Omega)$, improve quickly with $(S/N)^{-1}$. These combinations correspond to the best 4 principal axes of the 5-dimensional error ellipsoid. The remaining 5th principal axis does not improve as $(S/N)^{-1}$, but rather stops improving at a fraction of the last modulation cycle time. The two dimensional sky position error ellipsoid is the projection of the general 5-dimensional error ellipsoid on the (θ_N, ϕ_N) plane. This plane will generally not be aligned with the principal axes of the 5-dimensional ellipsoid. In a typical case, therefore, there will be a nonzero projection on the worst principal component and the sky position ellipsoid will stop shrinking along the worst principal component. This explains why the major axis, $2a$, ceases to improve and the eccentricity increases close to merger.

According to this argument, it is somewhat surprising to find that the minor axis, $2b$, can stop improving much before ISCO. Figure 2 shows that this happens in the worst 10% of all cases for randomly chosen source angular parameters. The reason for this is that, in some cases, not all rapidly improving “best” principal components have a small absolute error at ISCO. For example, consider an edge-on binary inspiral ($\cos \theta_{NL} \approx 0$). Since two of the quickly improving parameters are simply proportional to $\cos \theta_{NL}$, the errors will be very large for these parameters. Thus, depending on the relative orientation of the detector and the source at ISCO, there can be large absolute errors in some cases even for the best combinations of parameters. In short, both axes of the sky position error ellipsoid can stop improving at late times in those cases when LISA is oriented in its least favorable direction at ISCO.

VIII. CONCLUSIONS

We have developed a new harmonic mode decomposition (HMD) method to study the feasibility of using LISA inspiral signals to locate coalescing SMBH binaries in the sky, as the mergers proceed. According to our extensive HMD survey of potential LISA sources, it will be possible to trigger large field-of-view searches for prompt electromagnetic counterparts during the final stages of inspiral and coalescence. Our results indicate, for instance, that for a typical $z \sim 1$ merger event with total mass $M \sim 10^5 - 10^7 M_\odot$, a 10-day advance notice will be available to localize the source to within a 10 deg^2 region of the sky. The advance notice to localize the source to a 10 times smaller area of 1 deg^2 is < 1

day for the typical event, suggesting that a wide-field instrument of the LSST class, with a 10 deg^2 field-of-view, may offer significant advantages over a smaller, 1 deg^2 field-of-view instrument for observational efforts to catch prompt electromagnetic counterparts to SMBH binary inspirals.

The robust identification of such electromagnetic counterparts would have multiple applications, from an alternative method to measure cosmological parameters to precise measurements of merger geometries in relation to host galaxy properties [8, 15]. If such electromagnetic counterpart searches can be implemented effectively and successfully, LISA could become an extremely valuable instrument for astrophysics and cosmology, beyond the original general relativistic measurement goals. Given the advance warning time capabilities established here, effective strategies for electromagnetic counterpart searches, including the concept of partially dedicating a $\gtrsim 10 \text{ deg}^2$ field-of-view fast survey instrument of the LSST class, are considered in detail in a separate investigation [18].

Acknowledgments

We thank Samuel Finn and Shane Larson for influential early discussions on this problem and Scott Hughes and Tom Prince for valuable comments which improved our manuscript. BK acknowledges support from a Smithsonian Astrophysical Observatory Predoctoral Fellowship and from Öveges József Fellowship. ZH acknowledges partial support by NASA through grant NNG04GI88G, by the NSF through grant AST-0307291, and by the Hungarian Ministry of Education through a György Békésy Fellowship. KM was supported in part by the National Science Foundation under Grant No. PHY05-51164 (at KITP). Z.F. acknowledges support from OTKA through grant nos. T037548, T047042, and T047244.

APPENDIX A: SIMPLE TOY MODELS

1. Single Frequency Model

First, let us consider the following simple model with two unknowns, c_0 and c_1 ,

$$h(t) = c_0 + c_1 \cos(2\pi f_\oplus t), \quad (\text{A1})$$

where $f_\oplus \equiv \text{yr}^{-1}$ is fixed and assumed to be known prior to the observation. We call t the “look-back time” before merger. Let us assume that the relative noise continuously decreases during the observation and that the differential squared signal-to-noise ratio (without modulation) is given by $\sigma^{-2}(t) = t^{-2}$ in eq. (50). Here $t = 0$ is a proxy for the “merger”. Close to merger, the signal-to-noise ratio accumulates very rapidly. We assume that $h(t)$ is measured in the time interval $t_i \geq t \geq t_f$, where t_i is the start of observation, t_f is the end of observation (i.e. $x = t_{\text{merger}} - t$, $x_{\text{min}} = t_f$, and $x_{\text{max}} = t_i$ in eq. [50]). We fix t_i and examine the dependence of parameter estimation errors as a function of t_f , assuming $t_f \ll t_i$.

Note that, for the signal (A1), the fiducial values (c_0, c_1) drop out when calculating the RMS parameter errors Δc_0 and Δc_1 using eq. (50). More generally, this is true for any signal which is a linear combination of the unknown parameters. All our toy models will have this property and the results presented in this section will be general in that respect.

First, let us substitute (A1) in (49) and (50), and evaluate the expected covariance matrix numerically. Figure 8 displays the time dependence of marginalized parameter errors and principal errors. The plots show that the parameter errors all decrease with the signal to noise ratio when the look-back time before merger is large. However if the end of the observation is within a certain critical time to merger, $t_f < t_c$, only one principal component follows the signal-to-noise ratio. Figure 8 shows that $t_c \sim 0.1$ yr. The start of the observation in Figure 8 was fixed at $t_i = 5$ yr.

It is also interesting to examine what happens for general total observation times, do errors stop improving within some time t_c before merger? If yes, how does t_c depend on the two timescales t_i and f_{\oplus}^{-1} ? We examine this question numerically, substituting (A1) in (49) and (50) and now varying both t_f/f_{\oplus}^{-1} and t_i/f_{\oplus}^{-1} . Let us define the critical end-of-observation, t_c , as the time when the marginalized squared parameter error is first within a factor of 2 of its final value. Figure 9 plots the result for the two parameters. Figure 9 shows that t_c is determined by f_{\oplus}^{-1} for large t_i , but becomes t_i -dependent for lower t_i values. In the limit $t_i \ll f_{\oplus}^{-1}$, the critical look-back time is independent of f_{\oplus}^{-1} , it becomes a constant fraction of t_i .

Note that, in the limit of an observation extending up to merger, at $t = 0$, the signal becomes $h(0) = c_0 + c_1$ and it has infinite instantaneous signal-to-noise ratio. Therefore, this is the best combination of parameters for which the scaling of errors can follow $(S/N)^{-1}$ all the way to $t = 0$. The worst combination is $c_0 - c_1$, which stops improving before $t = 0$.

For this simple model, the origin of these features can be understood by analyzing the principal errors and the marginalized errors in the error covariance matrix. For this purpose, we present an analytical algebraic solution to this problem. To simplify the equations, let us set the time-scale to $f_{\oplus}^{-1}/(2\pi)$. In this case the Fisher matrix (50) is

$$\Gamma_{ij}(t_f, t_i) = \begin{pmatrix} \int_{t_f}^{t_i} t^{-2} dt & \int_{t_f}^{t_i} \cos(t) t^{-2} dt \\ \int_{t_f}^{t_i} \cos(t) t^{-2} dt & \int_{t_f}^{t_i} \cos^2(t) t^{-2} dt \end{pmatrix}. \quad (\text{A2})$$

The integrals can be evaluated analytically,

$$\Gamma_{ij}(t_f, t_i) = \begin{pmatrix} \frac{1}{t} & \frac{\cos(t)}{t} + \text{Si}(t) \\ \frac{\cos(t)}{t} + \text{Si}(t) & \frac{\cos(2t)+1}{2t} + \text{Si}(2t) \end{pmatrix} \Bigg|_{t_i}^{t_f}, \quad (\text{A3})$$

where $\text{Si}(x) = \int_0^x \frac{\sin(x)}{x} dx$ is the sine integral.

In the next two subsections, we find the limiting behavior of marginalized and principal parameter errors in two different limits: $f_{\oplus}^{-1} \ll t_i$ and $t_i \ll f_{\oplus}^{-1}$, respectively.

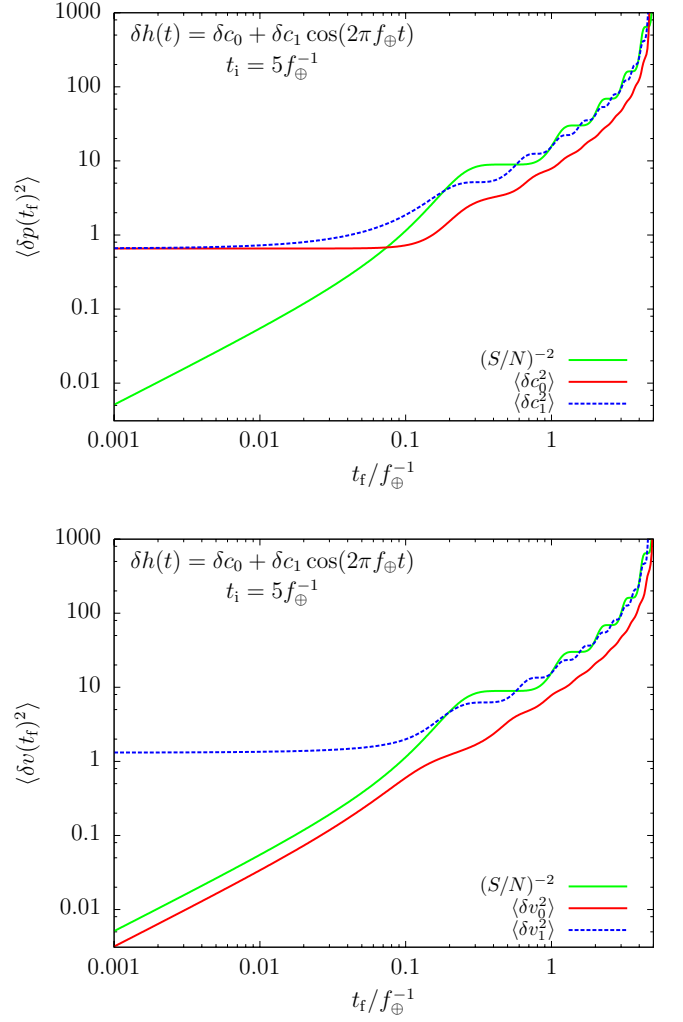


FIG. 8: Marginalized parameter errors (*top*) and principal errors (*bottom*) for the single frequency model. The green curve shows the scaling with inverse squared signal-to-noise ratio, $(S/N)^{-2}$, for reference on both plots. A total observation of $t_i = 5$ yr is assumed. Marginalized errors follow the signal-to-noise ratio for large t_f , but they stop improving within $t_f < t_c \sim 0.1$ yr from merger. Only one eigenvalue scales with the signal-to-noise ratio near merger.

a. Long Observations ($f_{\oplus}^{-1} \ll t_i$)

Here, we assume that the signal has been measured for a very long total time and we concentrate on the effects of changing the end of the observation time, t_f , near merger. Therefore, we take the limit $t_i \rightarrow \infty$, for which

$$\Gamma_{ij}(t_f) = \begin{pmatrix} \frac{1}{t_f} & \frac{\cos(t_f)}{t_f} + \text{Si}(t_f) \\ \frac{\cos(t_f)}{t_f} + \text{Si}(t_f) & \frac{\cos(2t_f)+1}{2t_f} + \text{Si}(2t_f) \end{pmatrix} - \begin{pmatrix} 0 & \pi/2 \\ \pi/2 & \pi/2 \end{pmatrix}. \quad (\text{A4})$$

We consider the case of a total observation time which is not negligible compared to a cycle time, f_{\oplus}^{-1} , i.e. $t_f \ll t_i$. We next examine two possible cases, $f_{\oplus}^{-1} \ll t_f \ll t_i$ and $t_f \ll f_{\oplus}^{-1} \ll t_i$,

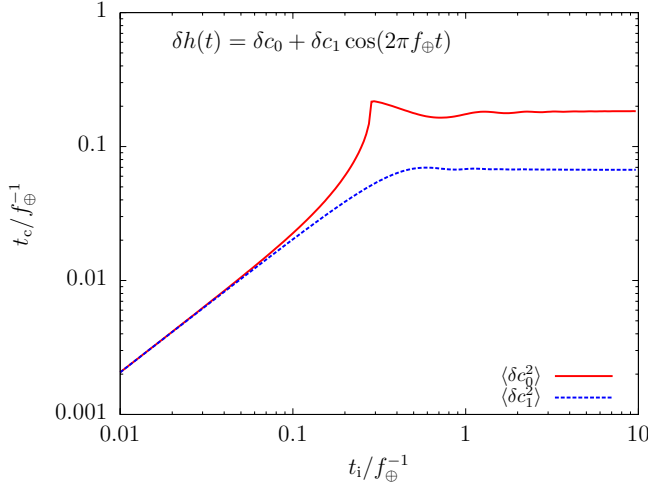


FIG. 9: Critical look-back time, t_c , at which parameter errors stop improving. Here t_c is defined as the time at which marginalized squared errors are within a factor of 2 of their final values for the first time.

separately.

First let us assume that the merger is still far away in time in units of a cycle period ($f_\oplus \ll t_f \ll t_i$). We substitute (A4) in (49) and expand $\Gamma^{-1}(t_f)$ into a t_f^{-1} series:

$$(\Gamma^{-1})_{ij} \approx \frac{t_f}{1 - \frac{\sin(2t_f)}{2t_f} + \frac{\cos(2t_f) - 1}{t_f^2}} \begin{pmatrix} 1 - \frac{\sin(2t_f)}{2t_f} & \frac{2\sin(t_f)}{t_f} \\ \frac{2\sin(t_f)}{t_f} & 2 \end{pmatrix}. \quad (\text{A5})$$

Equation (A5) gives the large t_f behavior of marginalized errors and correlations, which can be compared to Figure 8 in the appropriate regime, $t_f > 1$ yr. In this case, to leading order, *all of the squared errors* scale with t_f , which is the scaling of the inverse squared signal-to-noise ratio, $(S/N)^{-2}$, for our noise model.

Next, let us examine the case when the end-of-observation time is close to merger, i.e. $t_f \ll f_\oplus \ll t_i$. Now, taking the inverse of the matrix and expanding into a t_f series around $t_f = 0$ gives

$$(\Gamma^{-1})_{ij} \approx \frac{2}{\pi} \begin{pmatrix} 1 + \frac{2}{3\pi} t_f^3 & -1 + \frac{t_f^2}{2\pi^2} \\ -1 + \frac{t_f^2}{2\pi^2} & 1 + \frac{\pi}{2} t_f \end{pmatrix}, \quad (\text{A6})$$

which gives the short timescale behavior of marginalized errors and correlations. The eigenvalues of Γ^{-1} define the squared length of the individual principal axes of the parameter error ellipsoid, in this case

$$\begin{pmatrix} \langle \delta v_0^2 \rangle \\ \langle \delta v_1^2 \rangle \end{pmatrix} \approx \begin{pmatrix} \frac{t_f}{2} + \frac{3\pi}{16} t_f^2 \\ \frac{4}{\pi} + \frac{t_f}{2} + \left(\frac{5\pi}{16} - \frac{2}{\pi}\right) t_f^2 \end{pmatrix}. \quad (\text{A7})$$

Note that, in eqs. (A2)-(A7), time is measured in units of $f_\oplus^{-1}/(2\pi)$. In full units, the squared marginalized parameter

errors (i.e. diagonal elements) of (A6) become

$$\begin{pmatrix} \langle \delta c_0^2 \rangle \\ \langle \delta c_1^2 \rangle \end{pmatrix} = \begin{pmatrix} \frac{2}{\pi} \left[1 + \left(\frac{t_f}{\sqrt{3/(16\pi^2)} f_\oplus^{-1}} \right)^3 \right] \\ \frac{2}{\pi} \left[1 + \frac{t_f}{\pi^2 f_\oplus^{-1}} \right] \end{pmatrix}. \quad (\text{A8})$$

For the eigenvalues (A7), we get

$$\begin{pmatrix} \langle \delta v_0^2 \rangle \\ \langle \delta v_1^2 \rangle \end{pmatrix} = \begin{pmatrix} t_f / \left(\frac{1}{\pi} f_\oplus^{-1} \right) \\ \frac{4}{\pi} \left[1 + \frac{t_f}{\pi^2 f_\oplus^{-1}} \right] \end{pmatrix}. \quad (\text{A9})$$

Equation (A8) implies that the evolution of the marginalized squared error on c_0 is very flat for small t_f , when the second term is negligible, i.e. $t_f \ll \sqrt{\frac{3}{16\pi^2}} f_\oplus^{-1} = 0.267$ yr, then rises steeply ($\propto t_f^3$). The marginalized squared c_1 error is also constant near merger, for $t_f \ll \frac{1}{\pi} f_\oplus^{-1} \approx 0.1$ yr, and it increases $\propto t_f \propto (S/N)^{-2}$ for larger t_f . Equation (A9) shows that one of the principal errors has a very different time-evolution: it has no constant term proportional to t_f^0 . Therefore the semi-minor axis of the error ellipsoid can decrease continuously with the signal to noise ratio. On the other hand, the semi-major axis becomes constant for $t_f \ll \frac{4}{\pi^2} f_\oplus^{-1} = 0.4$ yr. Since the marginalized errors are nontrivial linear combinations of the principal errors, the constant principal error carries over to both marginalized errors and dominates their evolution. All of these findings are in excellent agreement with the numerical results shown in Fig. 8 for $t_f \ll 1$ yr and in Fig. 10 for $t_i/f_\oplus^{-1} > 1$.

It is worth emphasizing that, even if the total observation time had been infinite, $t_i \rightarrow \infty$, the parameters could *not* have been estimated to infinite precision in this model. It is not very surprising if one recalls that in this model we defined errors to be infinitely large at infinitely early times ($\sigma^2(t) \propto t^2$). For stationary noise, the contribution of the last cycle to the resultant RMS estimation error for a total observation of N_{cyc} cycles is $1/\sqrt{N_{\text{cyc}}}$. In contrast, *rather than the total number of cycles, the typical error during the last cycle dominates the determination of noise*, for the particular noise model used here.

The main conclusion from this toy model analysis is that errors stop improving close to merger, at $t_c \sim 0.1 f_\oplus^{-1}$. It can be extended to more general noise models, with $\sigma^2(t) = t^{-\alpha}$ and $\alpha \neq 2$. Repeating the calculations for larger α values, we find that parameter estimation errors become more and more insensitive to very early times, $t_f \ll t \sim t_i$, and that marginalized parameter estimation errors cease to improve at some t_c , which is now an α -dependent fraction of a single cycle time before merger. For $\alpha > 2$, we find that errors increase more abruptly at $t_f \gtrsim t_c$, which is consistent with the signal-to-noise ratio being a steeper function of time. On the other hand, for lower α values, parameter estimation errors become more and more sensitive to very early times, $t_f \ll t \sim t_i$. In this case, the marginalized parameter estimation errors are again very slowly changing for $0 \sim t_f < t_c$, but the approximate time t_c at which parameter errors stop decreasing will be primarily

determined by t_i , rather than by the cycle period f_{\oplus}^{-1} . The transition at $t_f \gtrsim t_c$ is not as abrupt, but extends to several cycles. The $\alpha = 0$ case corresponds to a stationary instantaneous signal-to-noise ratio, with errors scaling slowly as $1/\sqrt{t_i - t_f}$. This case is irrelevant to LISA inspiral signals, which have $\alpha \sim 2$ to a good approximation for $1\text{day} < t < t_i$ in the relevant range of SMBH masses.

b. Short observations ($t_i \ll f_{\oplus}^{-1}$)

Let us now examine the opposite limiting case, where the start of observation time is already within the final cycle before merger. This is relevant to LISA signals, since the observation time of SMBH inspirals is often below a full year, especially for $(1+z)M \geq 4 \times 10^6 M_{\odot}$.

We again restrict ourselves to the case with a total observation time that is non-negligible, i.e. $t_f \ll t_i$. Using time units of $f_{\oplus}^{-1}/(2\pi)$, expanding (A3) into a series of both t_i and t_f/t_i , we get

$$(\Gamma^{-1})_{ij} \approx \frac{120}{t_i^3(10-t_i^2)} \left[\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \frac{t_f}{t_i} \begin{pmatrix} \frac{30-10t_i^2}{10-t_i^2} & -\frac{30-5t_i^2}{10-t_i^2} \\ -\frac{30-5t_i^2}{10-t_i^2} & \frac{30-\frac{5}{2}t_i^2}{10-t_i^2} \end{pmatrix} \right] \quad (\text{A10})$$

Equation (A10) gives the parameter estimation covariance during the final stages of observation before merger for small total observation times. In this case, the final errors strongly depend on the total observation time. The errors reach their final values when the second term becomes negligible in eq. (A10). To leading order, this happens at $t_c \sim t_i/3$ for both parameters, independently of the cycle time, f_{\oplus}^{-1} . Equation (A10) approximates well the t_i dependence of t_c shown in Fig. 9 for $t_i/f_{\oplus}^{-1} < 0.2$

2. Double Frequency Model

Now consider a more elaborate model with five unknowns c_0, s_1, c_1, s_{10} , and c_{10} :

$$h(t) = c_0 + s_1 \sin(2\pi f_1 t) + c_1 \cos(2\pi f_1 t) + s_{10} \sin(2\pi f_2 t) + c_{10} \cos(2\pi f_2 t). \quad (\text{A11})$$

Here, the signal is comprised of two different characteristic frequencies, f_1 and f_2 , for which we assume $f_1 \ll f_2$. Moreover we assume that f_1 and f_2 are fixed and known prior to the measurement, e.g. we take $f_1 \equiv 1\text{yr}^{-1}$ and $f_2 \equiv 10\text{yr}^{-1}$. We again assume an observation in the look-back time interval $t_i \geq t \geq t_f$ and take the average instantaneous signal-to-noise ratio to increase as $\sigma(t)^{-2} = t^{-2}$.

Let us substitute in (49) and (50), and evaluate the expected covariance matrix numerically. Figure 10 displays the results. As in the previous model, these plots show that all parameter errors decrease with signal to noise ratio until the last cycle and all marginalized errors stop improving beyond some

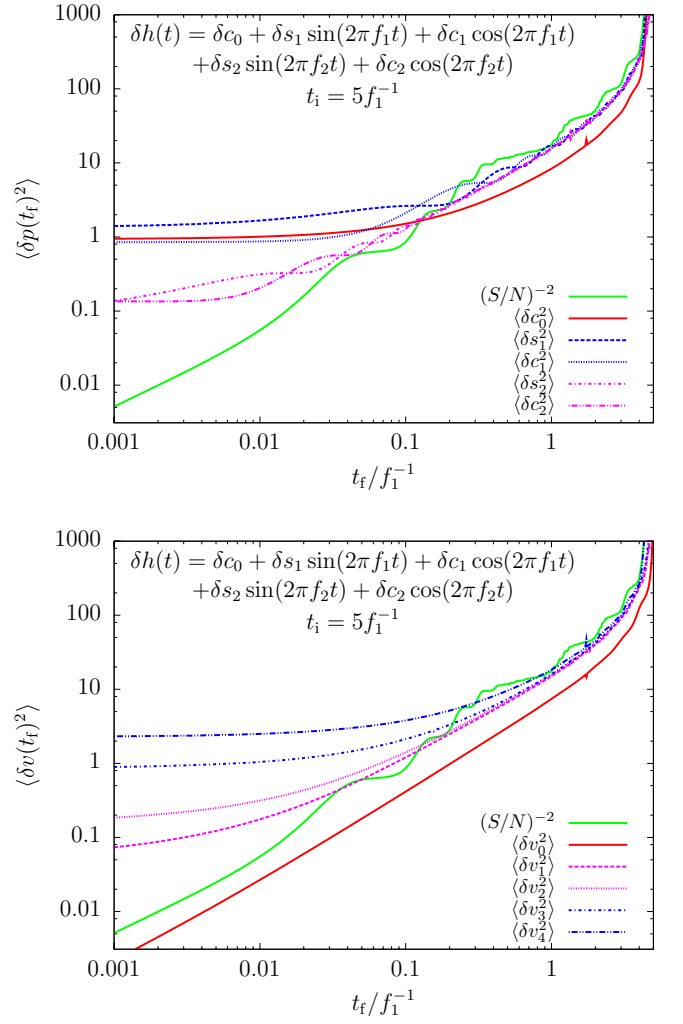


FIG. 10: Marginalized parameter errors (top) and principal errors (bottom) for the double frequency model. The green curve shows the scaling with $(S/N)^{-2}$ for reference on both plots. A total observation time $t_i = 5$ yr is assumed. Marginalized errors follow the signal-to-noise ratio for large t_f values, but they stop improving after $t_f \lesssim 0.1f$, for both frequencies. By comparing the two plots, it is clear that high frequency component errors decouple and that they are determined by two corresponding eigenvalues in the bottom panel.

nonzero residual error at late times. Thus, the general trends shown in Fig. 10 are very much similar to the ones in the previous simple model (Fig. 8). Again, contrary to the standard $1/\sqrt{N_{\text{cyc}}}$ expectation, the error during the last cycle dominates the total error of the accumulated signal. Moreover, comparing Figs. 8 and 10 shows that the presence of additional independent high frequency degrees of freedom practically does not modify the evolution of marginalized parameter errors associated with low frequency components, if $t_i > f_1^{-1}$. During the final cycle, the error ellipsoid becomes “thin” and the narrow dimension will not be aligned with any of the parameters. As a result, this bad principal error dominates each of the marginalized parameter errors at late times. (Note that the start-of-observation time in Figure 10 is $t_i = 5$ yr.)

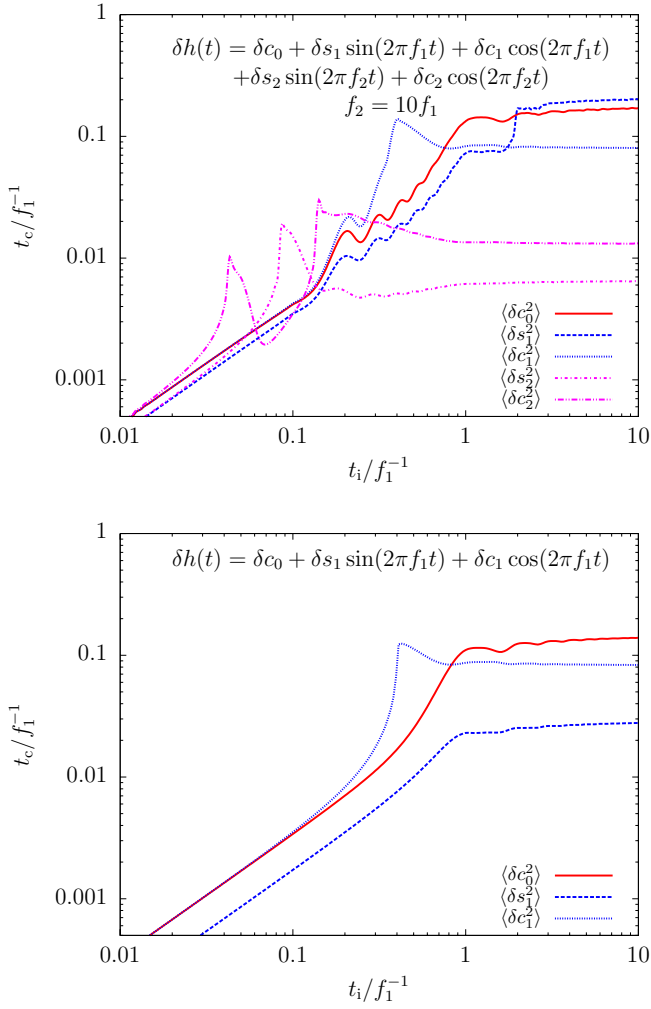


FIG. 11: Critical look-back time, t_c (as in Fig. 9), at which marginalized parameter errors stop improving. *Top*: Only (c_0, s_1, c_1) are allowed to vary, using the prior $(s_{10}, c_{10}) \equiv (0, 0)$. *Bottom*: All 5 parameters $(c_0, s_1, c_1, s_{10}, c_{10})$ are determined from the observation. For $t_i \gtrsim f_1^{-1}$, estimations of low frequency parameters (c_0, s_1, c_1) stop improving at $t_c \sim 0.1f_1^{-1}$, while improvement for high frequency parameters occurs all the way to $t_c \sim 0.1f_2^{-1}$.

The critical look-back time, t_c , at which this happens is different for the different frequency components. The top panel in Fig. 10 shows that $t_{c_i} \sim 0.1f_i$ approximately for both sets of components (s_1, c_1) and (s_{10}, c_{10}) , where f_i denotes the corresponding frequencies $f_1 = 1 \text{ yr}^{-1}$ and $f_2 = 10 \text{ yr}^{-1}$, respectively. The bottom panel in Fig. 10 shows that the principal errors separate in three groups. There is one best eigenvector that improves continuously until the end, two that stop improving near $t_{c_1} \sim 0.1f_1$ and two that stop improving at $t_{c_2} \sim 0.1f_2$. The high frequency parameters (s_{10}, c_{10}) totally decouple from the two worst principal components, (v_0, v_1) , and, as a result, decouple from the low frequency parameters (c_0, s_1, c_1) which are primarily determined by (v_0, v_1) .

As for our previous model in § A 1, the critical look-back time is generally different for different t_i values. The bottom panel in Figure 11 shows the time t_c at which the squared er-

rors first double, as a function of t_i/f_1^{-1} , as in Fig. 9. Fig. 11 justifies the rule-of-thumb scaling $t_{c_i} \sim 0.1f_i$ if the observation time is at least one cycle period, f_1^{-1} .

The central question for the present analysis is how sensitive is the time evolution of low frequency modulation errors to the presence of high frequency components. We can examine this question by computing the critical look-back time, t_c , when the high frequency terms are totally neglected. The top panel in Fig. 9 shows that, if one limits the parameters to (c_0, s_1, c_1) , and the total observation time is not smaller than the long-period cycle time, $\sim f_1^{-1}$, the resulting t_c value for parameters c_0 and c_1 is unchanged at the few percent level. However, if the high frequency components are introduced, the s_1 error evolves differently since it asymptotes already at much larger t_c values ($\sim 0.1f_1^{-1}$ rather than $\sim 0.03f_1^{-1}$). The reason is that, for small t , with a noise level decreasing quickly, the corresponding function $s_1 \sin(2\pi f_1 t) \approx 2\pi f_1 s_1 t$ is linearly independent of, and thus uncorrelated with, the functions c_0 and $c_1 \cos(2\pi f_1 t)$ which are both constant to first order. Hence, if there are no more unknowns than (c_0, s_1, c_1) , then c_0 and c_1 are correlated while s_1 is decoupled and can be determined independently of the other parameters. However, if we add any parameters which are *not* constant for $t \ll f_1^{-1}$, then s_1 becomes correlated with those. This is exactly what happens in the bottom panel of Fig. 9, when considering the high frequency modulations: the estimation on s_1 becomes limited for $t \lesssim t_{c_1} \sim 0.1f_1^{-1}$ due to the correlations with s_{10} and c_{10} . Quite similarly, if one introduces any other low-frequency function that is not constant to first order, like $s_2 \sin(4\pi f_1 t)$, then the correlations with this parameter will limit the improvement of estimation errors for s_1 at $t_c \sim 0.1f_1$, even when neglecting the high frequency components. As we shall see, this is the case for LISA: there are generally more than one sin and cos low-frequency modes. In this case, the evolution of estimation errors for low frequency parameters can be obtained with the high frequency modes (like s_{10} and c_{10}) priored out. This justifies our simple intuition: once the signal is decomposed into different time-scale components, the parameter estimation problem becomes separable and the evolution of parameter errors corresponding to different such time-scales can be estimated independently from each other.

Rather than going through an analytical derivation as in § A 1, we answer one remaining question here: what combination of the original parameters $(c_0, s_1, c_1, s_{10}, c_{10})$ corresponds to the best principal component, v_0 , which can be determined extremely accurately at late times, $t_f \rightarrow 0$? At $t = 0$, the noise drops to zero. Therefore, the quantity we can measure using the $t = 0$ information is simply $h(t = 0)$. Looking back at eq. (A11), this is $c_0 + c_1 + c_2$. It will be interesting to look for similar “best determined combinations” of physical parameters for the case of the LISA’s realistic signals.

3. Four data-stream models

For our final toy model, we insert additional features of a realistic LISA data-stream. We consider five low frequency unknowns, c_0, s_1, c_1, s_2, c_2 , and a high frequency carrier signal

with additional unknowns s_{10} , and c_{10} . Moreover we consider the simultaneous measurement of four data-streams. The signal is

$$h(t) = c_0 + s_1 \sin(2\pi f_1 t + \varphi_1^{s_1}) + c_1 \cos(2\pi f_1 t + \varphi_1^{c_1}) + s_2 \sin(2\pi f_1 t + \varphi_1^{s_2}) + c_2 \cos(2\pi f_1 t + \varphi_1^{c_2}) + s_{10} \sin(2\pi f_2 t) + c_{10} \cos(2\pi f_2 t), \quad (\text{A12})$$

where $\varphi_i^{c_1, s_1, c_2, s_2}$ ($i = 1 \dots 4$) are fixed a priori randomly chosen numbers defining the relative phases of the various modes which are being simultaneously measured. We compute independent Fisher matrices for each four set of $\varphi_i^{c_1, s_1, c_2, s_2}$. We assume that $f_1 \ll f_2$ and that f_1 and f_2 are fixed and known prior to the measurement. We choose $f_2 = 10f_1$ and find the evolution of marginalized errors and principal errors in two limits:

- (i) neglecting cross-correlations with the high frequency parameters by assuming a prior $\delta s_{10} = \delta c_{10} = 0$, and
- (ii) accounting for these high frequency parameters.

We again assume an observation in the look-back time interval $t_i \geq t \geq t_f$ and take the average instantaneous signal-to-noise ratio to increase as $\sigma(t)^{-2} = t^{-2}$.

The results for these models are shown in Figure 12. The marginalized errors (top) and principal errors (bottom) are shown for both cases (i) and (ii) above. The figures show that, in agreement with our previous model, uncertainties on the low frequency parameters are not affected by the high frequency parameters, except during the final 0.1 cycle time of the high frequency component, $0.1 f_2^{-1}$. The figures also show that the four principal components of the error ellipsoid improve quickly at late times.

Marginalized parameter errors improve quickly if they have negligible projection on the bad directions of the error ellipsoid. As a result, our expectation is that errors will typically not stop improving abruptly, but that there will be a shallower evolution in the final two weeks. In the worst case for a given parameter, if it is aligned with the bad ellipsoid principal component, it will stop improving near merger. In the best case, if the parameter is orthogonal to the bad ellipsoid principal component, it will improve quickly throughout the final days of inspiral. Therefore, we understand that the distribution of errors broadens for $t_f \ll 0.1 f_1^{-1}$.

4. Best Determined Parameters

In the previous section, we have shown that, if the noise decreases quickly like t^2 near merger (at $t = 0$), the best-determined parameters are the eigenvectors of the error covariance matrix that improve with $(S/N)^{-1}$. Near merger, these are the independent detector outputs at $t = 0$. In the case of LISA inspirals, the observation only extends down to ISCO. In this case, the best determined combination of physical parameters p_1 at ISCO are the real and imaginary parts of $h_1^{\text{I,II}}(p_1)$. To prove this, we have to show that these are uncorrelated and decrease with $(S/N)^{-1}$. The functions

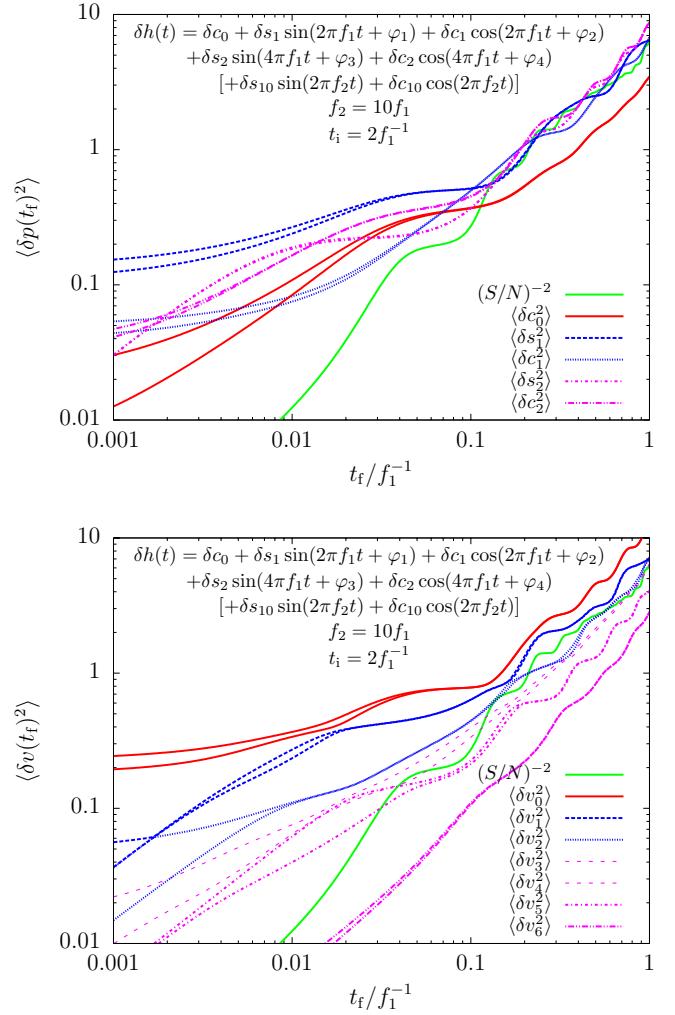


FIG. 12: Marginalized parameter errors (top) and principal errors (bottom) for the four data-stream model. Pairs of curves with the same line style show results for cases with five and seven parameters. The extra two parameters correspond to high frequency (f_2) components, which affect errors on the other parameters through correlations only slightly (factor of $\lesssim 2$) if $t_f \lesssim 0.1 f_2^{-1}$. The green curve shows the scaling with inverse squared signal-to-noise ratio, $(S/N)^{-2}$, for reference on both plots. A total observation time $t_i = 2\text{yr}$ is assumed. Marginalized errors follow the signal-to-noise ratio for large t_f values. Four principal errors scale with the signal-to-noise ratio near merger.

$h^{\text{I}}(t)$ and $h^{\text{II}}(t)$ are uncorrelated by construction, since they correspond to the two independent Michelson detector outputs (see § IIB and Cutler [19]). The real and imaginary parts of one of the detectors, $\Re h_1^{\text{I}}(t)$ and $\Im h_1^{\text{I}}(t)$, are uncorrelated since they are the coefficients of the high frequency carrier, $\sin \phi_{GW}$ and $\cos \phi_{GW}$, for which correlation over one ϕ_{GW} cycle (during which the detector noise is approximately constant) is zero. Another way to see this is to focus on the real part in the definition of the Fisher matrix (57), which is expressed as the integral of $\Re[\partial_a h_1^{\text{I,II}}(t) \partial_b h_1^{\text{I,II}}(t)]$. The term in brackets is purely imaginary for the cross correlation of

$\Re h_1^I(t)$ and $\Im h_1^I$, hence the real part is always zero. Therefore, the correlation matrix for $\Re h_1^{I,II}$, $\Im h_1^{I,II}$ is diagonal. For diagonal terms, the derivatives are 1 and the integrals become simply $\int \sigma^{-2} dt$, which is exactly $(S/N)^2$. The RMS estimation uncertainty of $\Re h_1^{I,II}(p_1)$ and $\Im h_1^{I,II}(p_1)$ follows the $(S/N)^{-1}$ all the way down to ISCO. These best combinations are $d_L^{-1}(1 + \cos^2 \theta_{NL})F_+^{I,II}(\Omega)$ and $d_L^{-1} \cos \theta_{NL}F_\times^{I,II}(\Omega)$.

The evolution of an arbitrary combination of angles will be determined by the projection of this combination on the covariance matrix eigenvectors. A linear combination of good eigenvectors leads to similarly quick improvement of errors with $(S/N)^{-1}$. However, as soon as there is a nonzero projection on the fifth eigenvector, the estimation uncertainty will stop improving at $\sim 0.1T_{\text{cycle}}$ which, for the highest $j = 4$ harmonic, is between 1–2 weeks.

APPENDIX B: ANGULAR VARIABLES

Here we define the relative angles θ_{NL} and ϕ_{NL} , using the polar angles (θ_N, ϕ_N) and (θ_L, ϕ_L) and the corresponding unit vectors $\hat{\mathbf{N}}$ and $\hat{\mathbf{L}}$.

Let us write a rotation around $\hat{\mathbf{z}}$ and $\hat{\mathbf{y}}$ as $O_z(\phi)$ and $O_y(\theta)$, respectively. Then, $\hat{\mathbf{z}} = O_y(-\theta_N)O_z(-\phi_N)\hat{\mathbf{N}}$ and we define

$$\begin{pmatrix} \sin(\theta_{NL}) \cos(\phi_{NL}) \\ \sin(\theta_{NL}) \sin(\phi_{NL}) \\ \cos(\theta_{NL}) \end{pmatrix} \equiv O_y(-\theta_N)O_z(-\phi_N)\hat{\mathbf{L}}. \quad (\text{B1})$$

This uniquely defines θ_{NL} and ϕ_{NL} , which correspond to the relative latitude and longitude, respectively. More explicitly, we get

$$\theta_{NL} = \arccos(\hat{\mathbf{N}} \cdot \hat{\mathbf{L}}) = \quad (\text{B2})$$

$$= \arccos[\sin \theta_N \sin \theta_L \cos(\phi_L - \phi_N) + \cos \theta_N \cos \theta_L],$$

$$\phi_{NL} = \begin{cases} 2\pi - \phi_0 & \text{if } (\phi_L - \phi_N)/\pi \in [-1, 0] \cup [1, 2] \\ \phi_0 & \text{otherwise} \end{cases}, \quad (\text{B3})$$

where

$$\phi_0 = \arccos\left(\frac{\cos \theta_N \sin \theta_L \cos(\phi_L - \phi_N) - \sin \theta_N \cos \theta_L}{\sin \theta_{NL}}\right). \quad (\text{B4})$$

-
- [1] O. Dreyer, B. Kelly, B. Krishnan, L. S. Finn, D. Garrison & R. Lopez-Aleman, *Class. Quantum Grav.* **21**, 787 (2004).
- [2] M. C. Miller, *ApJ* **618**, 426 (2005).
- [3] S. A. Hughes & K. Menou, *ApJ* **623**, 689 (2005).
- [4] E. Berti, V. Cardoso & C. M. Will, *Phys. Rev. D* **73**, 064030 (2006).
- [5] K. G. Arun, *Phys. Rev. D* **74**, 024025 (2006).
- [6] B. F. Schutz, *Nature* **323**, 310 (1986).
- [7] S. A. Hughes, *Mon. Not. Roy. Astron. Soc.* **331**, 805 (2002).
- [8] D. E. Holz & S. A. Hughes, *ApJ* **629**, 15 (2005).
- [9] M. C. Begelman, R. D. Blandford & M. J. Rees, *Nature* **287**, 307 (1980).
- [10] J. E. Barnes & L. Hernquist, *Ann. Rev. Astron. Astroph.* **30**, 705 (1992).
- [11] K. Menou, Z. Haiman, & V. K. Narayanan, *ApJ* **558**, 535 (2001).
- [12] E. Berti, A. Buonanno & C. M. Will, *Class. Quant. Grav.* **22** S943 (2005).
- [13] E. Berti, A. Buonanno & C. M. Will, *Phys. Rev. D* **71**, 084025 (2005).
- [14] M. Milosavljevic & E. S. Phinney, *ApJ* **622**, L93 (2005).
- [15] B. Kocsis, Z. Frei, Z. Haiman, & K. Menou, *ApJ* **637**, 27 (2006).
- [16] M. Dotti, R. Salvaterra, A. Sesana, M. Colpi & F. Haardt, *Mon. Not. Roy. Astron. Soc.* **372**, 869 (2006).
- [17] A. Vecchio, *Phys. Rev. D* **70**, 042001 (2004).
- [18] B. Kocsis, Z. Haiman, K. Menou, & Z. Frei, in prep. (2007).
- [19] C. Cutler, *Phys. Rev. D* **57**, 7089 (1998).
- [20] T. A. Moore & R. W. Hellings, *Phys. Rev. D* **65**, 062001 (2002).
- [21] L. Barack & C. Cutler, *Phys. Rev. D* **69**, 082005 (2004).
- [22] R. Lang & S. A. Hughes, *Phys. Rev. D* **74**, 122001 (2006).
- [23] R. Cornish & L. J. Rubbo, *Phys. Rev. D* **67**, 022001 (2003).
- [24] C. W. Misner, K. S. Thorne, & J. A. Wheeler, *Gravitation* (San Francisco: Freeman) (1973)
- [25] E. Poisson & C. M. Will, *Phys. Rev. D* **52**, 848 (1995).
- [26] K. Danzmann, LISA and LISA Pathfinder (Noordwijk: ESA), www.rssd.esa.int/SP/SP/docs/LISASymposium...
- .../K.Danzmann/KD_LISASymp04.pdf (2004).
- [27] C. Van Den Broeck and A. S. Sengupta, *Class.Quant.Grav.* **24** 155 (2007).
- [28] C. Van Den Broeck and A. S. Sengupta, *Class.Quant.Grav.* **24** 1089 (2007).
- [29] T. A. Prince, M. Tinto, S., L. Larson, & J. W. Armstrong, *Phys. Rev. D* **66**, 122002 (2002).
- [30] P. C. Peters, *Phys. Rev. B* **136**, 1224 (1964).
- [31] P. J. Armitage & P. Natarajan, *ApJ* **634**, 921 (2005).
- [32] J. C. Papaloizou, R. P. Nelson & F. Masset, *A&A* **366**, 263 (2001).
- [33] L. S. Finn, *Phys. Rev. D* **46**, 5236 (1992).
- [34] C. Cutler & È. E. Flanagan, *Phys. Rev. D* **49**, 2658 (1994).
- [35] R. Lang & S. A. Hughes, *Phys. Rev. D* **75**, 089902E (2007).
- [36] J. A. Tyson & the LSST collaboration, *Proc. SPIE Int. Soc. Opt. Eng.* **4836**, 10, astro-ph/0302102 (2002).
- [37] Note that, contrary to our convention for redshifted mass parameters, we drop the z index for f and t because we never consider comoving frequencies or times.
- [38] www.srl.caltech.edu/~shane/sensitivity/
- [39] Cornish & Rubbo [23] combine the $\sqrt{3}/2$ factor with the beam patterns $F_+^{CR,I,II} = \frac{\sqrt{3}}{2}F^{I,II}$, but we follow the original definition, where $\sqrt{3}/2$ appears only when taking the linear combination of GW polarizations (21).
- [40] The reason for the factor 4 is that the mean squared of $\cos(x)$ or $\sin(x)$ is $1/2$ in (18), and since we use one-sided signals in frequency domain ($f > 0$), responsible for another factor of $1/2$ in comparison.
- [41] Note that this expression includes both the polarization amplitude and the polarization phase, as both of these terms are accounted for by the complex harmonic mode coefficients $g_j^{I,II}(p_{\text{slow}})$.
- [42] Note that if p_L was intricately coupled to the other angular parameters in $A_j^{I,II}$ then it would be impossible to detach the p_L part from $A_j^{I,II}$ and attach it to $P^{(0),(1),(2)}(t_f)$. Fortunately, these terms were originally included exclusively in the coefficients $A_j^{I,II}(p)$

in eq. (67) and in a very simple way: the p_L terms are found only in the L and L^* coefficients in $g_j^{\text{III}}(p_{\text{slow}})$ (see eqs. [41,42,67]). The precession phase terms, $\partial_a \delta_{\text{P}}(p_L, p_{\text{spin}}, t)$, can also be simply attached to $P^{(0),(1),(2)}(t_f)$. Due to the precession phase, the index k now spans the range $0 \leq k \leq 2N+2$.

[43] The Fisher matrix method yields $\sqrt{\langle \delta p_i^2 \rangle}$ RMS error for each set of fiducial angles. As an approximation, we identify the distribution of errors with the distribution of RMS errors.